

**Design, Synthesis and Structural Characterisation  
of Inhibitors of  
1-Deoxy-D-xylulose-5-phosphate Synthase**

Robin Matthias Gierse





The research described in this thesis was conducted between September 2016 and August 2020 at the Stratingh Institute for Chemistry, University of Groningen, The Netherlands, at the Helmholtz Institute for Pharmaceutical research Saarland (HIPS), Germany, at the Faculty of Natural Sciences and Technology of the Saarland University, Germany and at the Bayer CropScience research facility in Frankfurt-Hoechst, Germany.

This work was financially supported by Bayer CropScience and the Netherlands Organisation for Scientific Research (LIFT Grant 731.015.414)



Bayer CropScience



university of  
 groningen



UNIVERSITÄT  
DES  
SAARLANDES

## Design, Synthesis and Structural Characterisation of Inhibitors of

### 1-Deoxy-D-xylulose-5-phosphate Synthase

#### PhD Thesis

to obtain the degree of PhD of the  
University of Groningen  
on the authority of the  
Rector Magnificus Prof. C. Wijmenga  
and in accordance with  
the decision by the College of Deans.

#### Dissertation

submitted towards the degree  
Doctor of Natural Sciences (Dr. rer. nat.)  
of the Faculty of Natural Sciences and Technology  
at Saarland University

Double PhD degree

by

**Robin Matthias Gierse**

born on 6 July 1989  
in Duisburg, Germany

Groningen/Saarbrücken  
2022

**Tag des Kolloquiums:** 15. November 2022

**Dekan:** Prof. Dr. Ludger Santen

**Berichterstatter:** Prof. Dr. Anna K.H. Hirsch  
Prof. Dr. Dirk-Jan Slotboom  
Prof. Dr. Wulf Blankenfeld

**Weitere Mitglieder:** Prof. Dr. Wesley Browne  
Prof. Dr. Matthew R. Groves  
Prof. Dr. Adriaan J. Minnaard  
Prof. Dr. Gerard Roelfes

**Akad. Mitglied:** Dr. Jennifer Herrmann

**Vorsitz:** Prof. Dr. Claus-Michael Lehr

---

---

## Short Summary

Due to the emergence of pathogenic organisms with resistance to classical antibiotics, the development of new drugs is needed. The enzyme 1-deoxy-D-xylulose-5-phosphate synthase (DXPS) is a potential target for such a new antibiotic.

DXPS is the first enzyme of the methylerythritol phosphate (MEP) pathway, one of two known pathways for the biosynthesis of essential terpene building-blocks. It is found in many bacteria and plants, whereas most other organisms, especially mammals, use the mevalonate pathway. Inhibition of the MEP pathway is therefore one way to impair the growth and survival of microorganisms.

The focus of this thesis is the protein structure of DXPS and the identification and development of DXPS inhibitors. In Chapter 1.2 an overview of the enzyme and the metabolic pathway is given, Chapter 1.3 updates on developments since 2017. Chapter 1.4 introduces our general workflow for protein-templated dynamic combinatorial chemistry (ptDCC).

The main part describes in Chapters 2.1 and 2.2 protein crystallographic work to improve the resolution of *D. radiodurans* DXPS and structural elucidation of DXPS homologous from pathogenic species. In parallel, the hit-identification strategies ligand-based virtual screening (Chapter 2.3) and ptDCC (Chapter 2.4) were applied to find DXPS inhibitors. Finally, Chapter 2.5 describes the development and crystallographic validation of bioisosters for acylhydrazone-based ptDCC hits.

---

## Kurzzusammenfassung

Aufgrund der Zunahme von antibiotika-resistenten Pathogenen ist die Entwicklung neuer Antibiotika erforderlich. Das Enzym 1-Desoxy-D-xylulose-5-phosphat-Synthase (DXPS) ist ein potenzielles Ziel für eine solche Neuentwicklung.

DXPS ist das erste Enzym des Methylerythritolphosphat (MEP)-Weges, einer von zwei Stoffwechselwegen für die Biosynthese der essentiellen Terpen bausteine. Er kommt in vielen Bakterien und Pflanzen vor, wohingegen die meisten anderen Organismen, insbesondere Säugetiere, den Mevalonatweg nutzen. Die Hemmung des MEP-Weges ist daher eine Möglichkeit, das Wachstum und Überleben von Mikroorganismen gezielt zu beeinträchtigen.

Der Schwerpunkt dieser Arbeit liegt auf der Proteinstruktur von DXPS sowie der Identifizierung und Entwicklung von DXPS-Inhibitoren. Zunächst wird ein Überblick über das Enzym, den MEP-Weg und den aktuellen Forschungsstand seit 2017 gegeben (Kapitel 1.2 und 1.3). Das Protokoll unserer Arbeitsgruppe für protein-templierte dynamische kombinatorische Chemie (ptDCC) wird anschließend in Kapitel 1.4 vorgestellt.

Der Hauptteil beschreibt in den Kapiteln 2.1 und 2.2 proteinkristallographische Arbeiten zur Verbesserung der Auflösung von *D. radiodurans* DXPS sowie zur Strukturaufklärung von DXPS-homologen von Pathogenen. Parallel dazu wurden die Hit-identifikations-Strategien ligandenbasiertes virtuelles Screening (Kapitel 2.3) und ptDCC (Kapitel 2.4) angewandt, um DXPS-Inhibitoren zu finden. Abschließend wird in Kapitel 2.5 die Entwicklung und kristallographische Validierung von Bioisosteren für Acylhydrazon-basierte ptDCC-Hits beschrieben.

---

## Publications Included in this Thesis

- Publication 1:** Robin M. Gierse, Eswar Redeem, Eleonora Diamanti, Carsten Wrenger, Matthew R. Groves\*, Anna K.H. Hirsch\*: DXS as a target for structure-based drug design; **2017**, *Future Medicinal Chemistry*, Volume 7, Issue 16, Pages 2131-2141
- Publication 2:** Alwin M. Hartman ‡, Robin M. Gierse ‡, Anna K.H. Hirsch: Protein-Templated Dynamic Combinatorial Chemistry: Brief Overview and Experimental Protocol; **2019** *European Journal of Organic Chemistry*, Issue 22, Pages 3581-3590
- Publication 3:** Robin M. Gierse ‡, Eswar R. Reddem ‡, Alaa Alhayek, Dominik Baitinger, Zhoor Hamid, Harald Jakobi, Bernd Laber, Gudrun Lange, Anna K.H. Hirsch\*, Matthew R. Groves\*: Identification of a 1-deoxy-D-xylulose-5-phosphate synthase (DXS) mutant with improved crystallographic properties; *Biochemical and Biophysical Research Communications* **2021**, Volume 539, Pages 42-46
- Publication 4:** Robin M. Gierse ‡, Rick Oerlemans ‡, Eswar R. Reddem, Victor Gawriljuk, Alaa Alhayek, Dominik Baitinger, Harald Jakobi, Bernd Laber, Gudrun Lange, Anna K.H. Hirsch\*, Matthew R. Groves\*: First crystal structures of 1-deoxy-d-xylulose 5-phosphate synthase (DXPS) from *Mycobacterium tuberculosis* indicate a distinct mechanism of intermediate stabilization; *Scientific Reports* **2022**, Volume 12, Article number 7221
- Publication 5:** Sandra Johannsen ‡, Robin M. Gierse ‡, Arne Krüger, Rachel L. Edwards, Vittoria Nanna, Anna Fontana, Zhu Di, Tiziana Masini, Lais Pessanha de Carvalho, Mael Poizat, Bart Kieftenbelt, Dana M. Hodge, Daan Bunt, Kamila Anna Meissner, Edmarcia Elisa de Souza, Melloney Dröge, Bernard van Vliet, Jack den Hartog, Michael C. Hutter, Jana Held, Audrey R. Odom John, Carsten Wrenger, Anna K. H. Hirsch: Identification of three new inhibitor classes against *Plasmodium falciparum*; *submitted*; published on the pre-print server *ChemRxiv*, 12 April 2022, DOI: 10.26434/chemrxiv-2022-npslc
- Publication 6:** Ravindra P. Jumde, Melissa Guardigni, Robin M. Gierse, Alaa Alhayek, Di Zhu, Zhoor Hamed, Sandra Johannsen, Walid A. M. Elgaher, Philipp J. Neusens, Christian Nehls, Jörg Haupenthal, Norbert Reiling, Anna K.H. Hirsch: Hit-Optimization Using Target-Directed Dynamic Combinatorial Chemistry: Development of Inhibitors of the Anti-Infective Target 1-Deoxy-D-Xylulose-5-Phosphate Synthase; *Chemical Science* **2021**, Issue 22, Pages 7775-7785
- Publication 7:** Varsha R. Jumde ‡, Milon Mondal ‡, Robin M. Gierse, M. Yagiz Unver, Francesca Magari, Roos C. W. van Lier, Andreas Heine, Gerhard Klebe and Anna K. H. Hirsch: Design and Synthesis of Bioisosteres of Acylhydrazones as Stable Inhibitors of the Aspartic Protease Endothiapepsin; **2018**, Volume 13, Issue 21, Pages 2266-2270

‡: Authors contributed equally

\*: Corresponding Authors

---

## Contribution Report

### Publication 1:

- conceived and planned the manuscript
- wrote main part of the manuscript (Review)
- created most figures

### Publication 2:

- wrote half of the manuscript (Review)

### Publication 3:

- designed and cloned the plasmid
- expressed, purified and crystallized the protein
- performed kinetic and biophysical characterization of protein
- collected X-ray data, solved, build and refined the protein model
- conceived and planned the manuscript, wrote main part
- created all figures

### Publication 4:

- designed and cloned the plasmid
- expressed, purified and crystallized the protein
- performed kinetic and biophysical characterization of protein
- collected X-ray data, solved, build and refined the protein model
- conceived and planned the manuscript, wrote main part

### Publication 5:

- performed bioinformatic analysis
- wrote parts of the manuscript
- created corresponding figures

### Publication 6:

- expressed and purified the enzymes
- performed and analyzed inhibition and mode of inhibition assays
- contributed to the writing and editing of the manuscript

### Publication 7:

- purified and crystallized the protein, soaked the protein crystals
- collected X-ray data, solved, build and refined the protein model
- analyzed the co-crystal structure
- wrote parts of the manuscript and created corresponding figures

---

## Abbreviations

<b>3-GAP</b>	glyceraldehyde 3-phosphate
<b>CD</b>	circular dichroism
<b>DCC</b>	dynamic combinatorial chemistry
<b>DMADP</b>	dimethylallyl diphosphate
<b>DOXP</b>	1-deoxy-D-xylulose-5-phosphate
<b>DXS</b>	1-deoxy-D-xylulose-5-phosphate synthase
<b>DXPS</b>	1-deoxy-D-xylulose-5-phosphate synthase
<b>HDX-MS</b>	hydrogen/deuterium exchange mass spectrometry -CoA
<b>IDP</b>	isopentenyl diphosphate
<b>IspC</b>	1-deoxy-D-xylulose-5-phosphate reductoisomerase
<b>LThDP</b>	C2 $\alpha$ -lactylThDP
<b>MEP</b>	2-C-methyl-D-erythritol 4-phosphate
<b>MIC</b>	minimum inhibitory concentration
<b>MOI</b>	mode of inhibition
<b>MSA</b>	multiple sequence alignment
<b>MVA</b>	mevalonate
<b>PDH</b>	porcine pyruvate dehydrogenase
<b>PK</b>	pharmacokinetic
<b>ptDCC</b>	protein-templated dynamic combinatorial chemistry
<b>SBDD</b>	structure-based drug design
<b>SBVS</b>	structure-based virtual screening
<b>ThDP</b>	thiamin diphosphate
<b>UMCG</b>	University Medical Center Groningen





# Contents

Short Summary . . . . .	vi
Kurzzusammenfassung . . . . .	vii
Publications Included in this Thesis . . . . .	viii
Contribution Report . . . . .	ix
Abbreviations . . . . .	x
<b>1 Introduction</b>	<b>1</b>
1.1 General Introduction . . . . .	1
1.1.1 Nomenclature of DXPS . . . . .	4
1.2 DXPS as a Target for Structure-based Drug Design . . . . .	5
1.3 Recent Developments in DXPS Research . . . . .	24
1.4 Protein-Templated Dynamic Combinatorial Chemistry . . . . .	29
1.5 Aims and Scope . . . . .	40
<b>2 Results</b>	<b>41</b>
2.1 Identification of a <i>Deinococcus radiodurans</i> DXS Mutant with Improved Crystallographic Properties . . . . .	43
2.2 First Crystal Structures of DXPS from <i>Mycobacterium tuberculosis</i> Indi- cate a Distinct Mechanism of Intermediate Stabilization. . . . .	51
2.3 Identification of Three new Inhibitor Classes Against <i>Plasmodium falciparum</i> . . . . .	67
2.4 Hit-Optimization Using Target-Directed DCC: Development of Inhibitors of the Anti-Infective Target DXS . . . . .	133
2.5 Design and Synthesis of Bioisosteres of Acylhydrazones as Stable Inhibitors of the Aspartic Protease Endothiapepsin . . . . .	147
<b>3 Supplementary Information</b>	<b>153</b>
3.1 Supplementary Information for Chapter 2.1 . . . . .	155
3.2 Supplementary Information for Chapter 2.2 . . . . .	168
3.3 Supplementary Information for Chapter 2.3 . . . . .	180
3.4 Supplementary Information for Chapter 2.4 . . . . .	211
3.5 Supplementary Information for Chapter 2.5 . . . . .	340
<b>Summary and Outlook</b>	<b>363</b>
Summary . . . . .	363
Outlook . . . . .	365
Samenvatting . . . . .	366
Zusammenfassung . . . . .	368
Publications not Included in this Thesis . . . . .	371
<b>References</b>	<b>373</b>



# Part 1

## Introduction

The discovery of new drugs is a highly interdisciplinary and collaborative research field. Building up on the work of previous PhD students in our group,<sup>27,44</sup> this thesis bears its part to discover new inhibitors targeting the 1-deoxy-D-xylulose-5-phosphate synthase (DXPS) enzyme of the 2-C-methyl-D-erythritol 4-phosphate (MEP)-pathway. Contributing to the structural and biophysical characterization of the DXPS enzyme, the findings presented herein helped paving the way to the recently started Marie Skłodowska-Curie Innovative Training Network: "Exploiting the methylerythritol phosphate pathway as a source of drug targets for novel anti-infectives".<sup>33</sup>

### 1.1 General Introduction

Characteristic for living cells is a constant stream of life-sustaining chemical and biochemical reactions. In its entirety these reactions are called the metabolism, the intermediates and products of these reactions are termed metabolites.<sup>4</sup> The metabolites are distinguished into primary and secondary metabolites. Primary metabolites are compounds that are essential for the growth, the development and the reproduction of a cell, such as amino acids, sugars, nucleotides, and lipids.

A common definition for secondary metabolites is that they are not involved in the basic metabolic process, but are instead interacting with the environment of the cell.<sup>68</sup> Secondary metabolites are, with an estimated number of more than 300.000 compounds, a much larger class of compounds than the essential primary metabolites.<sup>47</sup> About thousand new secondary metabolites are discovered every year.<sup>68</sup> Based on their biosynthetic route, the wide variety of secondary metabolites is often separated into five big classes:

- Terpenes (or isoprenes) and steroids
- Alkaloids
- Fatty acids and polyketides
- Nonribosomal polypeptides
- Enzyme cofactors

Although classified as secondary metabolites, some compounds in these groups can fulfill an essential role for the survival and proliferation of a living organism. This is the reason why a few members of these groups can also be counted to the primary metabolites. Examples for such compounds are cofactors of enzymes or the sterols cholesterol and

ergosterol, terpenes which are essential for the modulation of cell-membrane fluidity. A eukaryotic cell needs to either biosynthesize these sterols *de novo* or take them up from its environment.

With over 35,000 known compounds, the terpenes are the largest group of secondary metabolites. It includes some well known compounds, such as carotenoids, chlorophyll and ubiquinone as well as the above-mentioned sterols.<sup>8,47</sup> All terpenes originate, at least in part, from the universal C<sub>5</sub>-substrates dimethylallyl diphosphate (DMADP) and its isomer isopentenyl diphosphate (IDP). Prenyl transferases catalyze the first step of the biosynthesis of the terpenes, a "head-to-tail" 1'+ 4 condensation reaction of DMADP with IDP.<sup>40,48</sup> After the initial reaction to form the C<sub>10</sub> geranyl diphosphate, a variety of biosynthetic routes and enzymes lead to the individual metabolites, often by further additions of the C<sub>5</sub> building block IDP or its oligomers.<sup>35</sup>

Up until 1993, it was assumed that the two building blocks DMADP and IDP are biosynthesized exclusively via the mevalonate pathway. This pathway is named after the key intermediate mevalonate, which is biosynthesized by the 3-hydroxy-3-methylglutaryl(HMG)-CoA-reductase. Mevalonate was first described as important biological compound in 1956.<sup>66,73</sup> Its discovery led to the understanding of the mevalonate pathway for the biosynthesis of IDP and DMADP in the following years.<sup>13</sup>

The knowledge of this pathway was a prerequisite for the development of anti-cholesterol drugs, as it enabled A. Endo to search for compounds inhibiting the HMG-CoA-reductase. Taking inspiration from the search for antibiotics, he speculated that fungi or mold could produce a compound that inhibits this central enzyme of the mevalonate pathway as part of their self-defense mechanism against microbes.<sup>23</sup> Screening of over 3800 broths resulted in the discovery of compactin, the first HMB-CoA reductase inhibitor in 1972.<sup>22</sup> This compound was the first one of a whole class of HMG-CoA-reductase inhibitors, termed statins. Compounds of this group are, up until today, the most-used cholesterol-reducing agents.<sup>23</sup>

Although the mevalonate pathway was accepted as the only source of DMADP and IDP for a long time, some observations were not explicable with just the mevalonate pathway. Early concerns were raised in 1981 by Raman and coworkers, who studied the pattern of [1-<sup>14</sup>C]-acetate and [2-<sup>14</sup>C]-acetate incorporation into different bacteria. They observed an exclusive integration at the C-2 position of levulinic acid, chemically derived from the terpene ubiquinone, which could not originate from terpene biosynthesis via the mevalonate pathway.<sup>49</sup>

The hints towards a second, alternative biosynthesis route were discussed, but could not be treated as a proof for a second pathway, as in many experiments the degradation or metabolism of the radiolabeled acetate could not be excluded with confidence. There were also contradicting results, such as the report of the isolation of a natural compound by L. Minale and coworkers. Already in 1977 he observed the incorporation of radiolabeled acetate by the archaeobacteria *Caldarella acidophila* in a pattern that was in full agreement with terpene biosynthesis via the mevalonate pathway.<sup>56</sup>

The discussion was ended in 1993, in parallel by results published by the labs of M. Rohmer and D. Arigoni.<sup>7,53,54,61</sup> M. Rohmer reported incorporation experiments with glucose and acetate, <sup>13</sup>C-labeled at different positions, into bacteria employing either the glyoxylate and tricarboxylic acid cycle or the Entner-Doudoroff pathway for carbon uptake. Based on the isotopic enrichment in isolated polyterpenes, it was possible to assign the source of each individual carbon atom. The observed pattern was not compatible with biosynthesis of IDP and DMADP via the mevalonate pathway. This was the proof of a second pathway: the 2-C-methyl-D-erythritol 4-phosphate pathway (MEP pathway).<sup>53,54</sup>

Similar experiments on *Ginkgo biloba* cells by M. Schwarz in the lab of D. Arigoni revealed the existence of the same pathway in plants.<sup>7,61</sup>

After the initial discovery of the MEP pathway, further labeling experiments and the just developed whole-genome analysis revealed its distribution.<sup>14,50</sup> Most bacteria, some unicellular parasites such as *Plasmodium falciparum*, and the chloroplasts of plants are relying on the MEP pathway. The mevalonate pathway is present in animals, fungi, the cytoplasm of plants, archaeobacteria and some eubacteria.<sup>14,55</sup> Plants are an interesting case, as the mevalonate pathway is used for the biosynthesis of IDP and DMADP in the cytosol, whereas in the chloroplast, which is of bacterial origin, the MEP pathway is used for the biosynthesis of chlorophyll and other terpenes.

Recently, more detailed analysis of the MEP pathway distribution revealed several exceptions, in particular in bacteria.<sup>30</sup> Table 1 shows the current WHO priority list of pathogens, as interesting examples, together with the pathway used by the organism.

**Table 1.1:** WHO priority list of pathogens and the isoprenoid biosynthesis pathways used by the organism.<sup>30</sup> MEP=2-C-methyl-D-erythritol 4-phosphate-pathway; MVA=mevalonate pathway

Organism	Metabolic pathway	Gram staining
<i>Acinetobacter baumannii</i>	MEP	-
Enterobacteriaceae	MEP	-
<i>Enterococcus faecium</i>	MVA	+
<i>Staphylococcus aureus</i>	MVA	+
<i>Helicobacter pylori</i>	MEP	-
<i>Campylobacter</i> spp.	MEP	-
<i>Salmonella</i> spp.	MEP	-
<i>Neisseria gonorrhoeae</i>	MEP	-
<i>Streptococcus pneumoniae</i>	MVA	+
<i>Haemophilus influenza</i>	MEP	-
<i>Shigella</i> spp.	MEP	-

The MEP pathway was discussed shortly after its discovery as a source of potential anti-infective enzymatic targets.<sup>31,32,34,52,67</sup> The fact that many pathogenic bacteria use the MEP pathway, but the human sterole synthesis relies on the mevalonate (MVA) pathway, led to the idea of inhibiting the bacterial metabolism, while the human one would be unaffected by target-related toxicity.

Following the hypothesis that an inhibitor of the MEP pathway should be active against *E. coli* and *Bacillus subtilis*, both relying on the MEP pathway, but not against *S. aureus*, which uses the MVA pathway, Shimizu et al. searched databases for known antibiotics with this antibacterial spectrum and found fosmidomycin.<sup>37</sup> This natural compound was isolated and described already in 1979, but only after the discovery of the MEP pathway fosmidomycin could be assigned to target IspC, the second enzyme of the pathway, with an IC<sub>50</sub> of 8.2 nM.<sup>36,37</sup>

The inhibition of IspC by fosmidomycin results in the expected antibiotic activity against bacteria and also the antimicrobial effects against *Plasmodium falciparum*. Combination therapies using fosmidomycin against malaria infections have been successful, although the bad pharmacokinetic (PK) properties of fosmidomycin are preventing the drug from wide

adoption.<sup>57</sup> Research to create derivatives with better PK properties is still ongoing.<sup>19,57</sup> Independent from its success as a drug, the early availability of fosmidomycin was an important proof that inhibition of one of the MEP pathway enzymes results in antimicrobial activity, validating the pathway for the development of antibiotics with a new mode of action.

Being promising drug targets, it was of great interest to determine the protein structures of all involved enzymes. A considerable number of protein structures (see Chapter 2.2 for details) of the enzymes of the MEP pathway was determined after its discussion as a source of attractive drug targets, between 2000 to 2003. The protein structure of the first enzyme of the pathway, DXPS, was published a few years later in 2007 for the model organisms *Escherichia coli* and *Deinococcus radiodurans*.<sup>75</sup> Homologous crystal structures of DXPS from pathogenic organisms have not been published since then. The last two enzymes of the pathway, IspG and IspH, contain an [4Fe-4S] cluster and their protein crystal structures were determined in 2010 and 2008, respectively.

The flux-controlling enzyme of the MEP pathway is DXPS.<sup>5,74</sup> Moreover, its product is a branch-point metabolite for the biosynthesis of the B vitamins.<sup>5,29,42</sup> It was shown that a DXPS knockout strain of *E. coli* is not viable, and its gene is often found only as a single copy in bacteria.<sup>29</sup> All these factors make the enzyme an ideal target for the development of antibiotics with a new mode of action. The next Chapter 1.2 gives a detailed introduction of DXPS, which is the focus of this thesis.

### **1.1.1 A Brief Comment on the Nomenclature of DXPS**

The enzyme DXPS (1-deoxy-D-xylulose-5-phosphate synthase) belongs to the class of transferases and has the Enzyme Class (EC) number 2.2.1.7. In literature, the synonymous abbreviations DXP synthase, DXPS and DXS are used. Historically, DXS was most commonly used, while today the UniProt database recommends the use of DXP synthase as short enzyme name and DXPS as abbreviation.<sup>3,12</sup> The publications presented in this thesis follow this change in naming. In most articles, the abbreviation DXS was used, while in the most recent publication the naming was changed to DXPS.

## 1.2 DXPS as a Target for Structure-based Drug Design

The following chapter was published in the journal *Future Medicinal Chemistry* and is reproduced with permission of Future Science Group without any changes.

### Title

DXS as a target for structure-based drug design

### Authors

Robin M. Gierse, Eswar Redeem, Eleonora Diamanti, Carsten Wrenger,  
Matthew R. Groves\*, Anna K.H. Hirsch\*

### Contributions

Robin M. Gierse conceived the manuscript, took part in its writing and created most figures. Eswar Redeem took part in writing created some figures. Eleonora Diamanti took part in writing. Carsten Wrenger provided critical feedback. Matthew R. Groves\* and Anna K.H. Hirsch\* took part in writing and conceived and supervised the project.

### Bibliographic Details

Future Medicinal Chemistry  
Volume 7, Issue 16, Pages 2131-2141  
Published 21 June **2017**  
DOI: 10.4155/fmc-2016-0239

\*: Corresponding Authors

## DXS as a target for structure-based drug design

In this review, we analyze the enzyme DXS, the first and rate-limiting protein in the methylerythritol 4-phosphate pathway. This pathway was discovered in 1996 and is one of two known metabolic pathways for the biosynthesis of the universal building blocks for isoprenoids. It promises to offer new targets for the development of anti-infectives against the human pathogens, malaria or tuberculosis. We mapped the sequence conservation of 1-deoxy-xylulose-5-phosphate synthase on the protein structure and analyzed it in comparison with previously identified druggable pockets. We provide a recent overview of known inhibitors of the enzyme. Taken together, this sets the stage for future structure-based drug design.

First draft submitted: 20 December 2016; Accepted for publication: 28 February 2017; Published online: 21 June 2017

**Keywords:** antibiotics • anti-infectives • DXS • malaria • methylerythritol phosphate pathway • protein crystallography • structure-based drug design • tuberculosis

### Why is DXS a good drug target?

New anti-infective agents are urgently needed, as pathogens are developing that are increasingly resistant to front-line anti-infectives. Over the past decades, the number of newly introduced anti-infective drugs with a unique mode of action has strongly decreased. This can be ascribed to the high development cost for anti-infectives and to the fact that new anti-infectives are only used against multiresistant pathogens to delay the emergence of resistance. Because of this, the estimated revenue of new anti-infectives in the first year after introduction to the market is low, while a patent protection period of 20 years is starting on the day of the filing of the patent application. As a consequence, the number of major pharmaceutical companies that are developing new anti-infectives has dropped from 18 in 1990 to only four in 2010 [1]. Since 2010, a number of actions have been taken to address this problem and increase the research output on novel antibiotics that start to show small effects [1].

Nevertheless, new anti-infectives and new targets for anti-infectives are needed, and an increasing number of academic groups have initiated research in this area; in particular, the treatment of so-called 'neglected diseases' is receiving increasingly more attention.

The recently discovered methylerythritol 4-phosphate (MEP) pathway seems to offer a variety of promising drug targets [2]. The pathway is named after its second intermediate, 2C-methyl-D-erythritol 4-phosphate and is one of the two distinct metabolic pathways leading to the universal building blocks for the biosynthesis of isoprenoids. This pathway is found in most bacteria, plants and protozoa, whereas higher eukaryotic organisms and some bacteria rely on the mevalonate pathway.

Thus, in pathogenic species that use the MEP pathway, it is possible to inhibit growth by selectively withdrawing a crucial source for the biosynthesis of several important secondary metabolites, while the eukaryotic/human mevalonate pathway is not affected [3]. This

Robin Matthias Gierse<sup>1</sup>, Eswar Redeem<sup>1</sup>, Eleonora Diamanti<sup>1</sup>, Carsten Wrenger<sup>2</sup>, Matthew R Groves<sup>\*,3</sup> & Anna KH Hirsch<sup>\*,1</sup>

<sup>1</sup>Stratingh Institute for Chemistry, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

<sup>2</sup>Unit for Drug Discovery, Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil

<sup>3</sup>Structural Biology Unit, Drug Design, Department of Pharmacy, University of Groningen, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands

\*Author for correspondence:

Tel.: +31 50 363 3305

Fax +31 50 363 7582

m.r.groves@rug.nl

\*\*Author for correspondence:

Tel.: +31 50 363 4275

a.k.h.hirsch@rug.nl

offers the potential for a new class of anti-infectives with minor side effects on the human body.

The enzyme DXS catalyzes the first and rate-determining step of the MEP pathway, and is therefore of particular interest. In this review, we will investigate the DXS enzymes of the model organisms for bacteria and plants – *Escherichia coli* and *Arabidopsis thaliana*, respectively. We will also examine exemplary species for pathogenic bacteria (*Mycobacterium tuberculosis*) and for protozoan parasites (*Plasmodium falciparum*). The bacterium *Deinococcus radiodurans* is also included, of which the only crystal structure in complex with thiamine diphosphate (ThDP) is known. In addition to reviewing the recent research done on this enzyme, we also focus on the structural properties of this enzyme and the consequences for a structure-based approach on inhibitor design against DXS enzymes of various species.

### The MEP pathway is essential for some of the deadliest pathogens

#### *Plasmodium falciparum*

In recent years, substantial efforts have been made to eradicate malaria. The efforts resulted in a decreased death rate of 48% from 2000 to 2015 [4]. As a consequence of the widespread use of antimalarial drugs, the selection pressure for resistant parasites increased. The first resistant strains against artemisinin and its derivatives have emerged. Efforts to stop the spread of those resistant strains of *Plasmodium falciparum* failed and artemisinin-resistant *P. falciparum* is now widespread in East Asian countries [5,6]. In the light of the increasing spread of resistance, there is an urgent need for new antimalarial drugs with novel modes of action. In this context, the MEP pathway seems to be a rich source of promising drug targets [7]. The antibiotic fosmidomycin has been used since 2004 in the fight against malaria and was the first inhibitor targeting an enzyme of the MEP pathway, in particular, the enzyme DOXP reductoisomerase [8]. Although the drug fosmidomycin has poor pharmacokinetic properties and its use is therefore limited, it validated the MEP pathway as a good target for new anti-infectives [7].

#### *Mycobacterium tuberculosis*

The WHO lists tuberculosis as one of the ten most common causes of death worldwide [9]. Caused by *M. tuberculosis*, this infectious disease primarily affects the lungs of the patients. Around 30% of the world population is infected with tuberculosis, mainly in the latent state, which does not manifest any symptoms. The active state of tuberculosis is mostly (95%) found in developing countries.

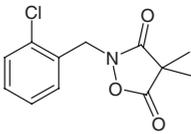
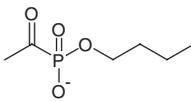
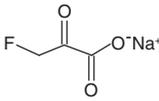
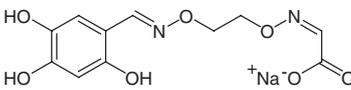
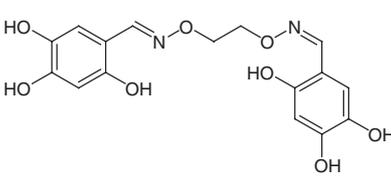
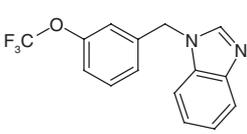
Common treatment of tuberculosis consists primarily of a combination of the antibiotics rifampicin and isoniazid, and takes several months. Through the extended and inappropriate use of antituberculosis drugs, resistances have emerged against one or multiple antibiotics. In the case of multidrug-resistant tuberculosis, treatments with second-line antibiotics are available, which suffer from a number of side effects. In 2006, the first case of a tuberculosis strain was reported in Italy, which was resistant against all known antituberculosis drugs (totally drug-resistant tuberculosis [TDR-TB]) [10]. Since then, more reports on cases of TDR-TB have appeared (e.g., in Iran and India) [11]. The development of drugs against TDR-TB with new modes of action is therefore urgently needed. As a result, *M. tuberculosis* DXS of the MEP pathway is a promising target and is worth exploring in detail.

### Herbicides

The first two enzymes of the MEP pathway are rate-limiting for the whole pathway [12]. The essential role of the MEP pathway in plants, for the synthesis of isoprenoids in the chloroplast, has attracted interest from two different fields.

First, the inhibition of the pathway results in plants that are unable to synthesize the terpene chlorophyll. This has been shown in several DXS knockout studies that show an albino phenotype owing to the inability to perform photosynthesis [13]. There is currently one known herbicide used, clomazone, which was introduced in 1986 and which targets the DXS enzyme of plants. Although it was long known that the herbicide affects the biosynthesis of diterpenes and tetraterpenes, the targeted protein could only be identified after the discovery of the MEP pathway in 1997 [14]. It took until 2013 to understand its mechanism of action. In living organisms, it is metabolized to ketoclomazone (Table 1) and undergoes ring opening before binding to DXS [15]. The molecule is, however, very small and has numerous unfavorable properties, such as high volatility, which limit its use as a herbicide.

Second, there is a strong interest in engineering the MEP pathway to increase the production of terpenoids in plants. Various natural substances are used in medicine, and their overexpression could significantly reduce the effort needed for the extraction from plant material, which often contains only a very small amount of the substance of interest [16]. However, genetically engineered enzymes are not directly aiding the discovery of new drugs, but the results obtained from enzyme function and pathway regulation should be transferable between the various homologs of DXS, owing to their high conservation of amino acids in the active site (see Figure 1 and their predicted 3D similarity [Figure 2]).

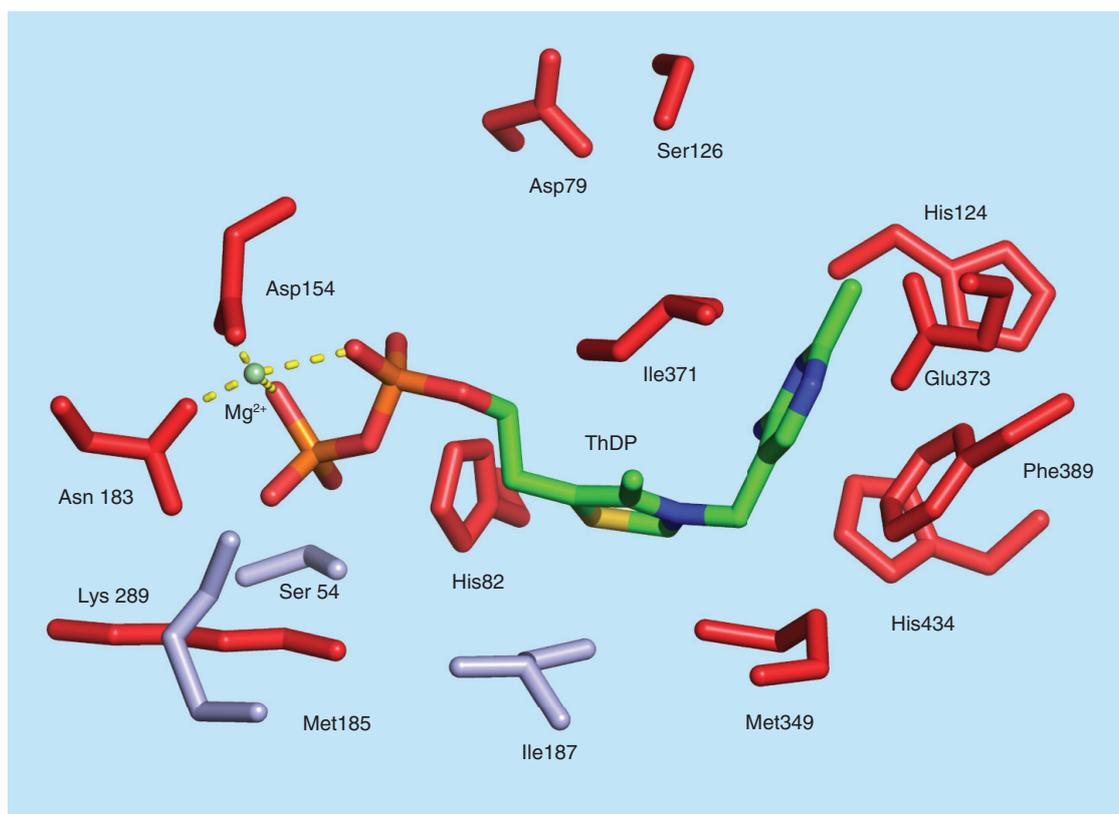
Table 1. Overview of the known inhibitors of DXS.					
No.	Inhibitor	Structure	Enzyme	$K_i$ ( $\mu\text{M}$ )	Mechanism of inhibition
1	Ketoclozazole		<i>Escherichia coli</i>	75	Uncompetitive to pyruvate
2	Butylacetylphosphonate		<i>Mycobacterium tuberculosis</i>	220	Noncompetitive to D-GAP
			<i>Yersinia pestis</i>	$4 \pm 2$	Competitive to pyruvate
			<i>Salmonella enterica</i>	$7.5 \pm 0.9$	Noncompetitive to D-GAP
			<i>E. coli</i>	$8.4 \pm 0.4$	Competitive to pyruvate
3	$\beta$ -fluoropyruvate		<i>Plasmodium vivax</i>	$35 \pm 1.7$	Competitive to pyruvate
			<i>Plasmodium falciparum</i>	$77 \pm 7.2$	Noncompetitive to D-GAP
				$43 \pm 3.8$	Competitive to pyruvate
				$131 \pm 10$	Noncompetitive to D-GAP
4	Trihydroxybenzaloximes Mixed oxime			$18.4 \pm 3.4$	Noncompetitive to pyruvate
					Competitive to D-GAP
5	Symmetrical oxime			$1.0 \pm 0.2$	Uncompetitive to pyruvate
					Competitive to D-GAP
6	1-(3-(trifluoromethoxy)benzyl)-1H-benzo[d]imidazole		<i>Deinococcus radiodurans</i>	$151 \pm 22$	
7	Bacimethrin		<i>E. coli</i>		

D-GAP: D-Glyceraldehyde-3-phosphate.

### The MEP pathway

Isoprenoids (or terpenoids) constitute a structurally diverse class of complex natural products, fulfilling a range of different functions such as protein degradation, apoptosis, influences on the cell wall, steroids, hormones, carotenoids, pigments and chlorophyll [17]. Isoprenoids are synthesized in all organisms using two common building blocks: isopentenyl diphosphate (IDP) and dimethylallyl diphosphate (DMADP).

Rohmer and Arigoni discovered that certain eubacteria, parasites and several plants use pyruvate and D-glyceraldehyde-3-phosphate (D-GAP) rather than acetyl-CoA, to produce the universal isoprenoid precursors [2,14,18]. This biosynthetic route, known as the MEP pathway, produces DMADP and IDP from a set of orthogonal enzymes nonexistent in higher organisms and mammals, which rely on the well-established mevalonate pathway [19].



**Figure 1. Mapping of the sequence conservation of the DXS enzymes from different species.**

Blue: Low conservation; Red: High conservation; Green: ThDP  
ThDP: Thiamine diphosphate.

The MEP pathway starts with decarboxylation of pyruvate and D-GAP to produce 1-deoxy-D-xylulose 5-phosphate (DOXP) catalyzed by DXS and assisted by ThDP and  $Mg^{2+}$ . DOXP is converted to MEP by IspC. The enzyme IspD catalyzes the reaction between MEP and CTP to produce 4-diphosphocytidyl-2C-methyl-D-erythritol (CDP-ME) and diphosphate.

The ATP-dependent enzyme IspE catalyzes the formation of 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate from CDP-ME. In the subsequent step, IspF catalyzes the formation of 2C-methyl-D-erythritol-2,4-cyclodiphosphate from CDP-ME. In the final steps, IspG and IspH generate of IDP and DMADP (Figure 3).

The structural information for each step was elucidated in detail except for DXS, which is the first and rate-determining enzyme in the MEP pathway [14,20,21]. In bacteria, the product of the DXS enzyme, DOXP, is a branch point in the metabolic pathway. It is not only used as a substrate for isoprenoid biosynthesis in the MEP pathway, but also acts as a precursor in the biosynthesis of pyridoxine (vitamin B6) and thiamine (vitamin B1) [22]. Moreover, bacteria rely only on the MEP pathway as demonstrated by the fact that the

DXS-knockout *E. coli* is not viable, but it can be rescued by media supplemented with vitamins B1, B6 and MEP. Vitamin B is needed because of the loss of the precursor DOXP for thiamine biosynthesis in the knockout strain. The supplement MEP is a downstream metabolite of the MEP pathway that restores the function of the rest of the pathway (Figure 3) [23].

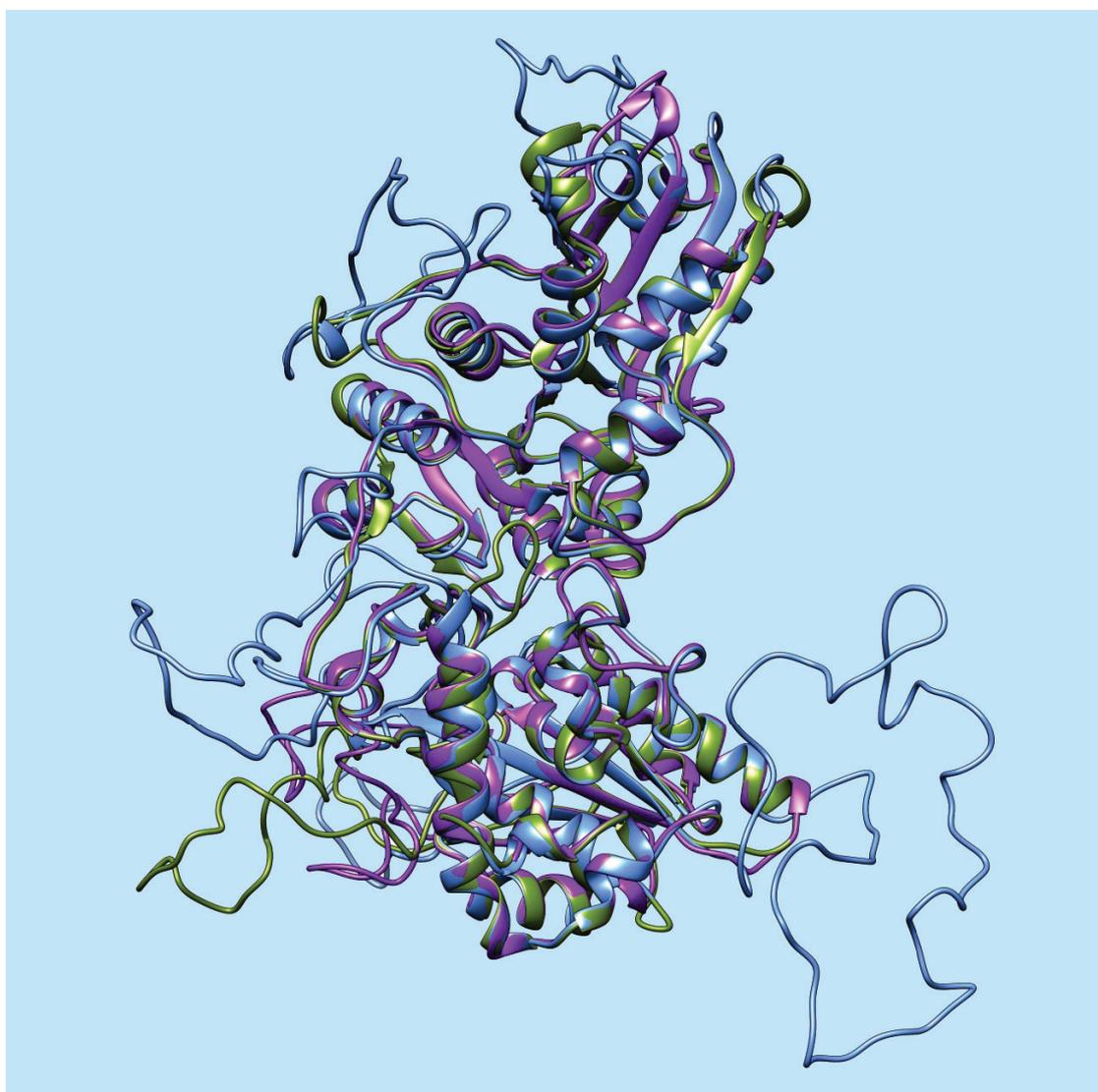
### Possible resistance development against drugs that target DXS

As the first enzyme of its pathway, DXS uses the substrates pyruvate and GAP. These metabolic intermediates are used for many different reactions in living organisms. Therefore, if the DXS enzyme is inhibited, its substrates do not accumulate. Pyruvate and GAP are instead processed in other metabolic pathways, or their production is downregulated [22]. An increase in substrate concentration is often observed for inhibition of enzymes in the middle of a pathway relying on specific substrates. The local concentrations of substrates reach high concentrations and may be able to reduce or completely revoke the effect of competitive inhibitors. Strictly, this would not be defined as a resistance, but as an error in target selection.

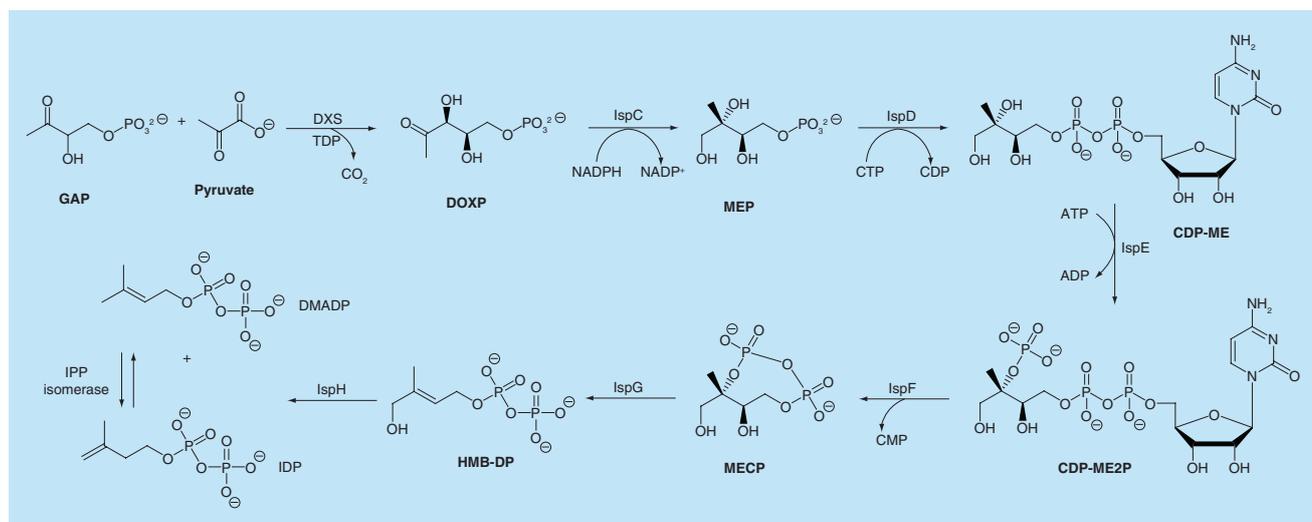
DXS is not only a promising drug target, it also seems to be robust against the most common mechanisms of resistance development. This should be relatively insensitive to intrinsic resistance arising by mutation. In most microorganisms, DXS is available as a single or double gene copy; therefore, there is a lower probability of a single-point mutation compared with targets that exist as several gene copies. In addition, the highly conserved active-site pocket gives hope to reduced endogenous resistance by mutation of the active site of the protein [24]. Often a resistance mutation bears some negative effect on the general fitness of the microorganism and is lost in the absence of selective pressure [24]. As a result,

it is not a persistent mode of drug resistance. Another intrinsic resistance mechanism of microorganisms is the overexpression of the targeted gene [25]. In the case of the DXS enzyme, the first one in its pathway, it may well be possible that the depletion of some of the downstream metabolites upregulates gene expression of DXS. Overexpression of the DXS enzyme could influence the anti-infective effect of the drug in different ways, depending on the mechanism of inhibition of the drug.

In addition to intrinsic resistance, microorganisms can also acquire or develop resistance. This can be due to several mechanisms, which can be divided into three groups: reduction of the intracellular drug



**Figure 2. Overlay of the predicted structures of *Arabidopsis thaliana* (olive), *Mycobacterium tuberculosis* (purple) and *Plasmodium falciparum* (light blue).** The domains of the protein are predicted similar to the *Deinococcus radiodurans* DXS template; the linker regions differ strongly in length and form random-coiled structures. For color figures, please see online at: <http://www.future-science.com/doi/full/10.4155/fmc-2017-0239>



**Figure 3. The methylerythritol 4-phosphate pathway for the biosynthesis of isopentenyl diphosphate and dimethylallyl diphosphate.** DMADP: Dimethylallyl diphosphate; IDP: Isopentenyl diphosphate; MEP: Methylerythritol 4-phosphate.

concentration, modification of the drug target and modification of the anti-infective drug [26]. Most of them are drug specific, but we can consider inhibitor classes such as ThDP or substrate analogs.

Reduced intracellular concentration is often achieved due to increased efflux by active transport of the drug. There are a range of thiamine transporters in nature, which could mutate to transport anti-infective thiamine analogs [27]. Most known transporters, however, act on the uptake of thiamine and might facilitate accumulation of the drug in the cell. In addition, if a microbacterial thiamine exporter would emerge as a resistance mechanism, it would be difficult to achieve a high selectivity for the drug in comparison with thiamine for the transporter, of which some are also known for their promiscuity to transport nicotinamide. The exporter would also actively transport the essential vitamin, thiamine, out of the cell, which is probably not beneficial for the viability of the microorganism [26,28]. For the second group of inhibitors, substrate analogues, active export would, in general, be possible, depending on the similarities of the inhibitor with the natural metabolites.

### The structure of DXS

The structures of the DXS enzyme (EC 2.2.1.7) from *D. radiodurans* (PDB code: 2O1X) and *E. coli* (PDB code: 2O1S) have been solved using x-ray crystallography at resolutions of 2.9 and 2.4 Å, respectively [29]. In brief, the enzyme contains three subunits I, II and III. All three domains share an  $\alpha/\beta$ -fold and consist of a central, parallel  $\beta$ -sheet, which is sandwiched by  $\alpha$ -helices and closely resembles the transketolase (TK) and pyruvate dehydrogenase (PDH) E1 subunit (Figure 4) [29].

A key feature of both structures is that during crystallization, the enzyme probably undergoes partial digestion, and both known structures show no interpretable density for an internal loop (amino acids 199–243 and 183–240 in *D. radiodurans* and *E. coli*, respectively). There is no evidence yet whether the loop is missing due to intrinsic disorder or as a result of proteolytic cleavage during the crystallization process [29]. Importantly, all structural analyses in the literature and also in this review are based on the assumption that the x-ray crystal structures represent a catalytically competent form of DXS. This assumption is supported by the crystal structure of the enzyme TK, which shares the same overall fold but is missing in this loop [31].

### Druggable pockets

We have recently reviewed the druggability of the enzyme DXS [30]. Using the program DoGSiteScorer, we identified the substrate-binding pocket and four other pockets as druggable pockets (Figure 4) [32]. DoG-SiteScorer is an algorithm to predict possible binding pockets for drugs based on the 3D structure of a protein, paying special attention to the size, compactness and physicochemical properties of the pockets.

After pocket identification, the volume and surface of the pocket are calculated by the program using ellipsoids, which are fitted into the pocket. The functional groups within 4 Å of the pocket are identified and taken into account, along with the lipophilic surface. For the druggability prediction, a machine-learning technique is used, which was trained on published successful and failed inhibitor-design projects [33].

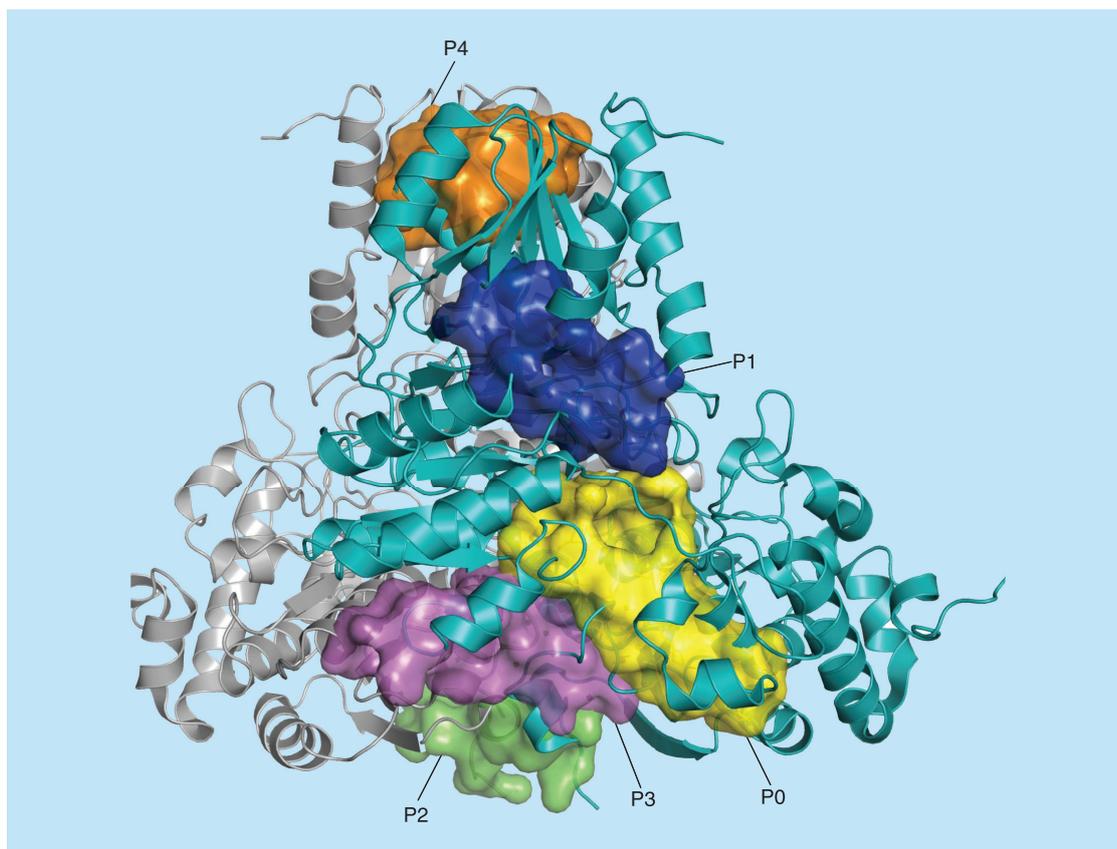
### Sequence conservation

Sequence conservation was analyzed using the multiple sequence comparison by log-expectation (MUSCLE) algorithm [34]. The sequences of the DXS enzymes of the different species were retrieved from the UNIPROT database using accession numbers – *D. radiodurans*: Q9RUB5, *E. coli*: P77488 *P. falciparum*: O96694, *M. tuberculosis*: A5U634 and *A. thaliana*: Q38854 [35]. The sequences were selected to include common model organisms as well as interesting pathogens. The alignment is shown in Figure 5 and was mapped on to the crystal structure of *D. radiodurans* DXS (PDB code: 2O1X) using the program UCSF Chimera. A color gradient from blue to red represents an increasing conservation of amino acids between the homologues.

Mapping of the sequence conservation onto the 3D structure is shown in Figure 6. We chose a visualization of the surface of the protein to focus on solvent-accessible amino acids. The strongly conserved regions of the protein are colored in red and form well-defined regions. They can be divided into a group of amino

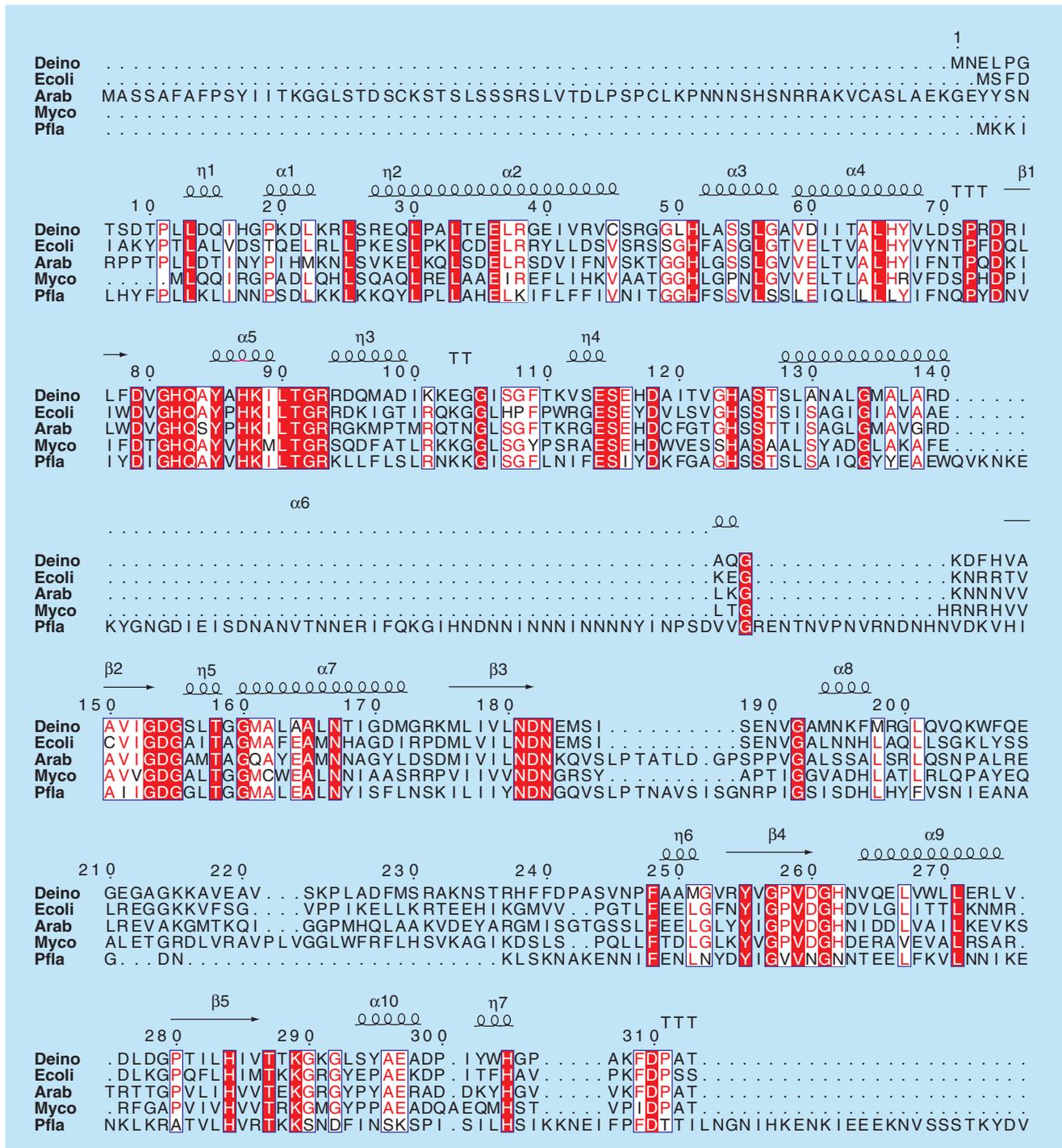
acids contributing to conserved pockets and a group of single conserved amino acids. Single conserved amino acids occur at positions, which are likely important for the correct folding of the protein, for example, Pro363, which is found in all homologues at the end of helix 12 (residues 363–355). A proline at the end of a helix is well known to destabilize the helical structure [38]. These small conserved regions corresponding to single amino acids are unlikely to be suitable as targets for the development of binders and are not further investigated.

Conserved pockets, on the other hand, tend to play an important role in the function of the enzyme. Often the binding sites for substrates, cofactors or allosteric factors are evolutionarily conserved. Depending on the pocket, they are suitable to tightly bind a small molecule and therefore a good target for the structure-based development of inhibitors. This can be seen in Figures 4 & Figure 6, where the ThDP-binding site (pocket P0) shows an expected high degree of sequence conservation. The binding site of pyruvate is also proposed to



**Figure 4. Co-crystal structure of the *Deinococcus radiodurans* DXS dimer with thiamine diphosphate (PDB code: 2O1X).** The monomers of DXS are shown in gray and cyan. The druggable pockets are highlighted as colored surfaces (P0: yellow, P1: blue, P2: green, P3: magenta, P4: orange). For further details on the pocket identification, see [30].

PDB: Protein data bank; ThDP: Thiamine diphosphate.



**Figure 5. Structure-based sequence alignment of selected 1-deoxyxylulose-5-phosphate synthase proteins.** The sequence of DXS from *Deinococcus radiodurans* (Deino; Q9RUB5) was aligned with *Escherichia coli* (*E. coli*; P77488), *Arabidopsis thaliana* (Arab; Q38854), *Mycobacterium tuberculosis* (Myco; A5U634) and *Plasmodium falciparum* (Pfla; O96694). Compared with Deino, the sequence identities of *E. coli*, Arab, Myco, Pfla are 45, 40, 40 and 33%, respectively. The sequence numbering and secondary-structure elements correspond to Deino. Conserved residues are shown in red color, while semiconserved residues are drawn in blue. The alignment was prepared with the program MUSCLE, and the figure was produced using the program ESript [36,37]. MUSCLE: Multiple Sequence Comparison by Log-Expectation.

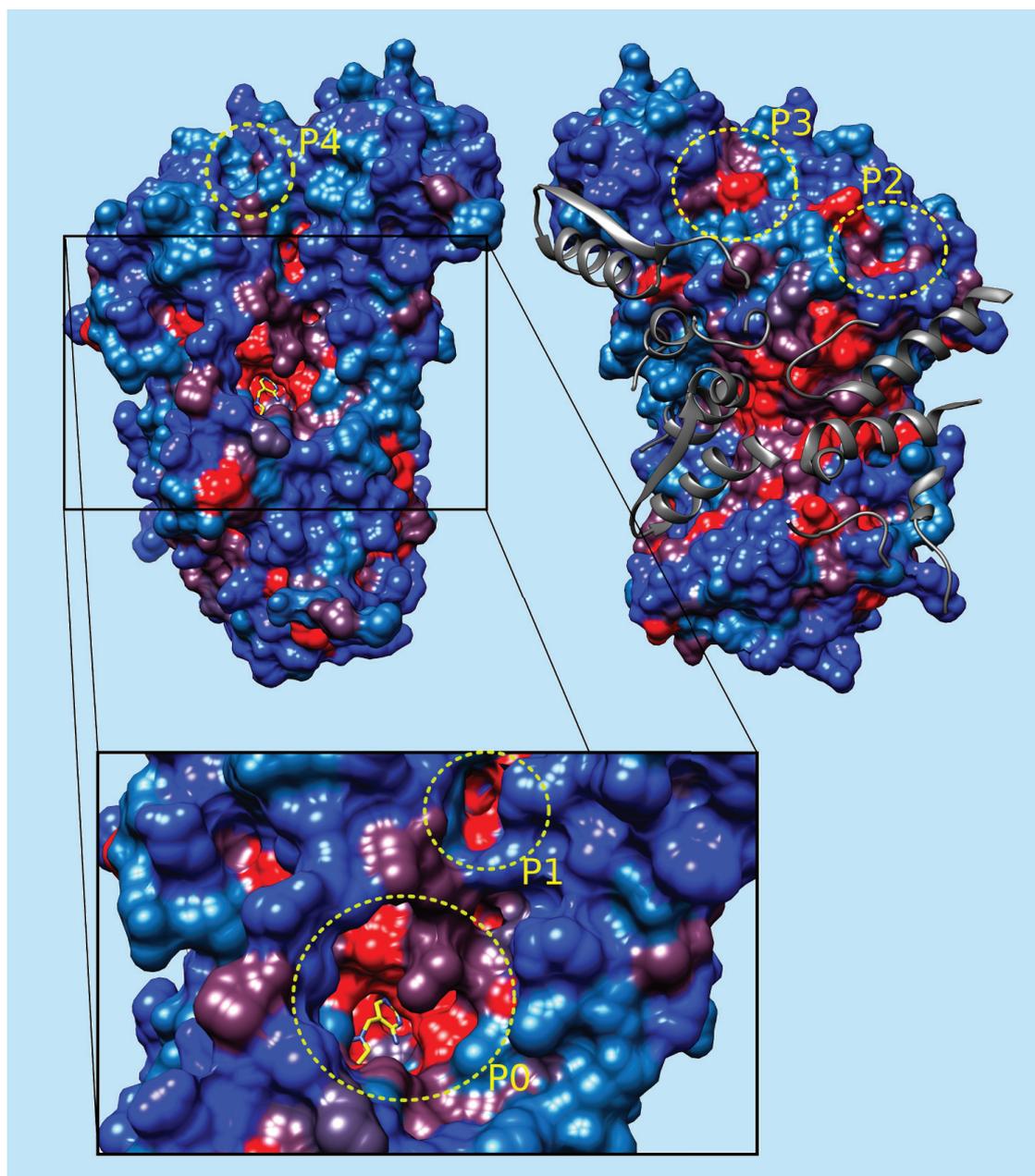
be in pocket P0, because it reacts with the C2 atom of the thiazolium ring of ThDP. The potential role of the other strongly conserved pockets is currently unknown. If the amino acids have no role in the stabilization of the protein structure, it is reasonable to propose that

they fulfill a different important function in the protein, which makes them resistant to mutation. This is likely to be the role in the reaction or regulation of the enzyme. As such, these pockets are of interest for the development of allosteric inhibitors.



**Figure 5. Structure-based sequence alignment of selected 1-deoxyxylulose-5-phosphate synthase proteins (cont. from previous spread).** The sequence of DXS from *Deinococcus radiodurans* (Deino; Q9RUB5) was aligned with *Escherichia coli* (*E. coli*; P77488), *Arabidopsis thaliana* (Arab; Q38854), *Mycobacterium tuberculosis* (Myco; A5U634) and *Plasmodium falciparum* (Pfla; O96694). Compared with Deino, the sequence identities of *E. coli*, Arab, Myco, Pfla are 45, 40, 40 and 33%, respectively. The sequence numbering and secondary-structure elements correspond to Deino. Conserved residues are shown in red color, while semiconserved residues are drawn in blue. The alignment was prepared with the program MUSCLE, and the figure was produced using the program ESript [36,37].

MUSCLE: Multiple Sequence Comparison by Log-Expectation.



**Figure 6. Mapping of the sequence conservation of the DXS enzymes from various species.** Both sides of the enzyme and a closeup of the active site with bound cofactor (ThDP) are shown. Color code – blue: low conservation; red: high conservation; and gray: closely associated parts of the second protein from the DXS homodimer. Sequence alignment was done using MUSCLE on the sequences of *Escherichia coli*, *Arabidopsis thaliana*, *Plasmodium falciparum*, *Mycobacterium tuberculosis* and *Deinococcus radiodurans* DXS [34]. Mapping was done using UCSF Chimera [39]. DXS: 1-deoxyxylulose-5-phosphatesynthase; MUSCLE: Multiple Sequence Comparison by Log-Expectation; ThDP: Thiamine diphosphate.

The ThDP-binding pocket corresponding to P0 (Figure 6), which also represents the active site of the enzyme, is highly conserved through all species. In close proximity to the thiazolium ring of ThDP, there are some less conserved amino acids, such as Ile187,

Glu184 and Met185. This pocket may be amenable to the binding of ThDP analogues as inhibitors against the whole DXS enzyme class. A minor drawback may be the difficulty to achieve selectivity against human ThDP-binding enzymes, such as TK (Figure 1,

comparison between DXS and TK). This has implications in off-target activities and may impose difficulties in the creation of a sufficiently specific DXS inhibitor for use as an anti-infective in humans.

In contrast, a small pocket in close proximity to the active site of the enzyme shows high conservation (Figure 6, the pocket on top of the ThDP-binding part). This region corresponds to the previously identified binding pocket P1 using the DoGSiteScorer algorithm. Therefore, it may be an interesting target for structure-based drug design (SBDD), as DoGSiteScorer indicates that it possesses suitable physicochemical properties to be addressed by small molecules, and our analysis here suggests that this region is strongly conserved in DXS enzymes. The human TK lacks this pocket, which may represent a possibility to address DXS selectively. The proximity to the catalytic domain of the enzyme suggests that it may have an influence on the mechanism. Furthermore, it shows a strong variability of amino acids at the entrance of the P1 pocket. If it is possible to address this pocket with a strongly binding fragment, it could be used as a scaffold for modifications, which interacts with a species-specific amino acid at the entrance of the pocket. In this way, it would be possible to develop a number of selective inhibitors against different species at the same time but targeting the same site. Considering the small size of the pocket, this should be possible without exceeding the recommended molecular weight of 500 g/mol of a drug-like molecule [40].

The predicted binding pockets P2 and P3 are close to the dimer interface. They show some sequence conservation but are more surface-exposed with a lower enclosure and a shallower depth of the pocket. In contrast to other ThDP-binding enzymes, DXS does not have its active site at the dimer interface. Therefore, these pockets at the interface are farther away from the active site and may not influence the function of the enzyme, but, instead, may play a role in the assembly of the homodimer. It is also possible that these pockets may play a role in subtle rearrangements of the monomer–monomer interaction during the catalytic cycle [41]; however, there are no data to support this. Additionally, this region of the crystal structures also possesses higher B factors than the other predicted pockets, indicative of a higher degree of mobility of the protein in this region. Therefore, calculations for this part of the enzyme may not be as accurate as for the rest because the use of a static structure for pocket identification may not be appropriate here. Thus, the pockets P2 and P3 may also be suitable for exploration, but are likely less desirable.

Pocket P4, which was identified by the DoGSiteScorer as a possible druggable pocket, shows low to

no sequence conservation. This pocket may be used to selectively address only the *D. radiodurans* DXS enzyme. However, the low conservation suggests no participation of this region of the protein in the catalysis. A low conservation is also a hint to a high possibility of random mutation and therefore resistance development against drugs that target this part of the enzyme. Further investigation of the pocket P4 seems less promising.

At the ‘back’ of the protein is the protein–protein contact surface (Figure 6, right). Several helix–helix interactions can be seen, which stabilize the binding of the homodimer and are well conserved. For a better visualization of the interface, the parts of the binding partner, which are involved in interactions, are shown in a gray cartoon visualization. This part of the enzyme nicely shows the conservation of important amino acids at the interface and the variability at the amino acids, which are solvent-exposed.

In the future, these protein–protein interactions could also be explored as possible targets for drug design. For a long time, targeting protein–protein interactions with small-molecule inhibitors has been considered difficult. Recently, an increasing number of examples of successfully targeting protein–protein interactions in SBDD have been reported [42]. It could be possible that the protein follows a cooperative mechanism, in which the interactions between the two active sites could be influenced by a binder at the interface. If the enzyme is cooperative and it is possible to separate the homodimer by a small binder, this could be a different way to inhibit the DXS enzyme, achieving high selectivity, using the unique arrangement of its domains in comparison with ThDP-binding proteins of other enzyme classes.

### Comparison of the DXS enzyme with TK

For the development of new anti-infectives, it is important to target the protein of interest with a high selectivity. ThDP is a cofactor, which is also used in the human body by several ThDP-binding proteins like TK, PDH and 2-oxoisovalerate dehydrogenase [27]. It is important to consider possible off-target effects on those human ThDP-binding enzymes during the development of new drugs. For a structural comparison, we performed a BLAST sequence analysis of the DXS enzyme and searched for human proteins with high similarity [43]. The non-DXS enzyme with the highest similarity in the results was the protein human TK and is used herein as an example for ThDP-binding proteins. The class of TK are enzymes, which transfer a ketol donor group to an accepting aldehyde or ketone group [44]. For a structural comparison, the crystal structure of human TK (PDB code: 3OOY)

was selected and compared with the structure of *D. radiodurans* DXS (PDB code: 2O1X).

The enzymes have a sequence identity of 23% and a similarity of 35% [45]. Although they have a rather different primary structure, it is remarkable that these enzymes show a very similar tertiary structure. Both enzymes consist of three domains and form a tightly associated homodimer. They show, however, significant differences in the arrangement of the three constituent domains. As a consequence, the two dimers also show a different orientation. A superimposition of the two enzymes is shown in Figure 1. The two parts of the TK and DXS dimers are colored in red and blue, and orange and light blue, respectively. They form the same overall structure however; it is possible to see four different parts of the protein due to the different arrangement of the domains.

The different arrangement has major consequences for the ThDP-binding pocket (Figures 1 & Figure 7, right). In contrast to the DXS enzyme, where the ThDP-binding domain is a deep pocket inside of each monomer, the ThDP-binding pocket of TK is located at the interface of the proteins and is formed by both chains (Figure 7, right). The linker between domains I and II of TK is with 95 amino acids longer than the 20 amino acid linker of the DXS enzyme and offers the possibility for the different arrangement of the enzyme [29]. Examining the ThDP-binding pocket at the level of the amino acids, a high similarity between the enzymes is apparent. The amino acids, which take part in binding ThDP, show a high conservation and bind ThDP in a similar strained conformation [46]. Some minor differences are observable, such as Met to Thr, Gln to His and His to Ser.

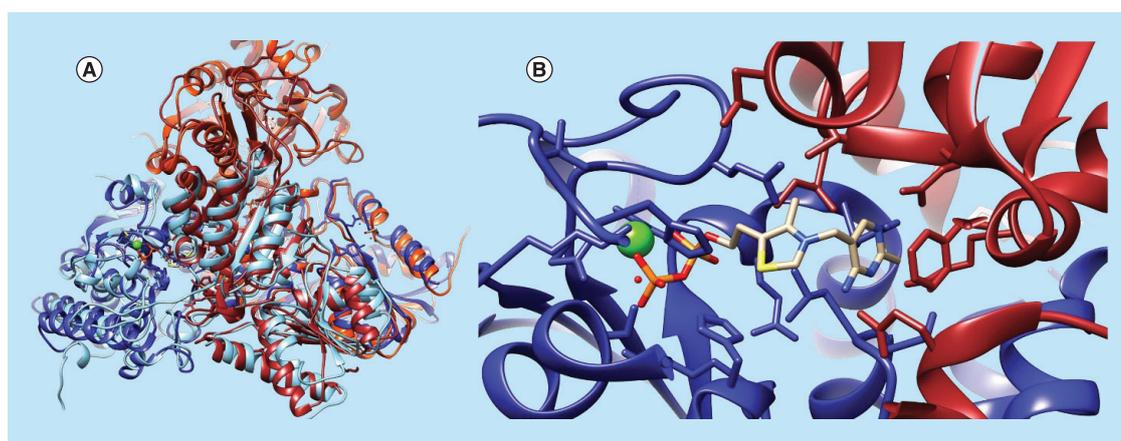
### Reaction mechanism of the DXS-catalyzed condensation

The exact reaction mechanism of the DXS enzyme is still unclear; a proposed mechanism is shown in Figure 8. In general, the enzyme catalyzes the decarboxylation and head-to-tail condensation of pyruvate and D-GAP to DOXP (Figure 3). Earlier studies have shown that the enzyme follows a classical ping-pong mechanism, where the pyruvate likely binds first covalently to the C2-atom of the thiazolium ring of ThDP, and then undergoes decarboxylation and forms a carbon-carbon bond with the C1-atom of D-GAP [47].

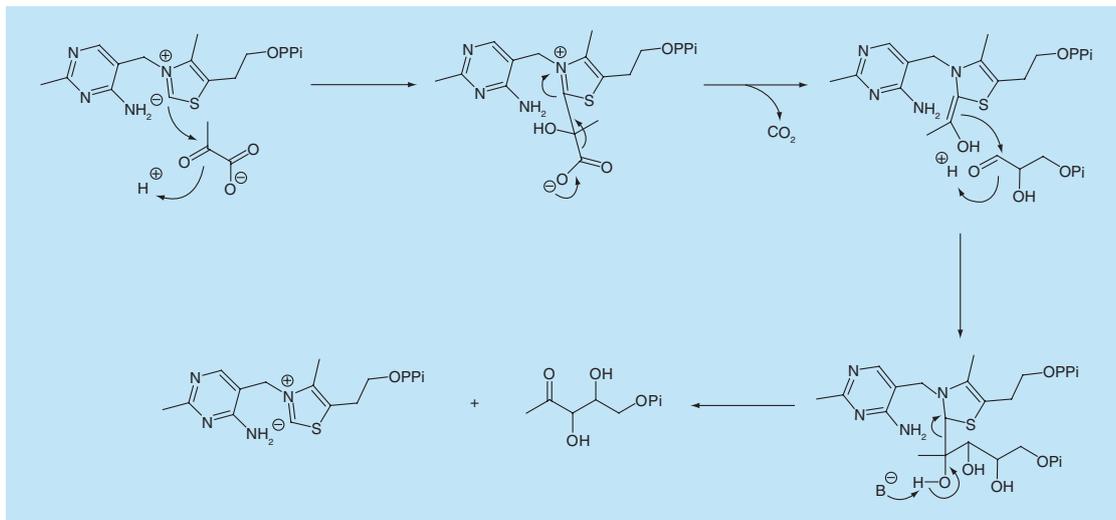
In 2010, Sisquella *et al.* proposed an ordered reaction mechanism using a single-molecule force spectroscopy nanosensor [48]. In this study, they observed a twofold binding increase of D-GAP to the enzyme-ThDP complex in the presence of pyruvate. Later in 2012, Meyers and coworkers found that pyruvate and D-GAP can bind independently to the DXS enzyme and D-GAP was found to accelerate the decarboxylation of the lactyl-ThDP intermediate by a factor of 600. They proposed that it follows a rapid equilibrium, random, sequential mechanism [49]. It is now a general consensus that the DXS reaction mechanism differs from the typical ping-pong mechanism of ThDP-binding proteins. This is important for the evaluation of DXS as a target for drug design, as this diversity in mechanism from other ThDP-dependent proteins provides an opportunity to address the DXS enzyme selectively.

### Mutational studies

To obtain more information about a reaction mechanism, site-directed mutagenesis is frequently used. By changing one specific amino acid and observing the



**Figure 7. DXS in comparison with human TK.** (A) Superposition of the *Deinococcus radiodurans* DXS enzyme with the human TK. The two chains of the human TK are colored in red and blue, and the chains of the DXS enzyme in the orange and light blue, respectively. The magnesium ion close to the ThDP is colored in green. (B) Active site of the human transketolase, which is formed at the interface of the two chains. ThDP: Thiamine diphosphate; TK: Transketolase.



**Figure 8. Proposed general reaction mechanism of the reaction catalyzed by the enzyme DXS.** ThDP is shown in its ylide state, in which it is bound and activated in the binding pocket.

ThDP: Thiamine diphosphate.

Data taken from [47].

consequence on the catalyzed reaction, the role of the mutated amino acid can often be identified.

The active site of the DXS enzyme is highly conserved, which can be seen in Figure 1, but not the entire conserved amino acids take an active part in the catalyzed reaction. Some of them may have other roles, such as binding the cofactor. An example of an essential amino acid that does not take part in the reaction is Phe398. It stabilizes the binding of the pyrimidine ring of the cofactor through  $\pi$ - $\pi$ -stacking interactions.

In 2014, Querol-Audi *et al.* carried out mutations around the active-site pocket of *E. coli* DXS enzyme to determine their importance in catalysis. Mutational studies confirmed the importance of His49, Asp427, His431 and Glu370 (corresponding to His51, Asp430, His434 and Glu373 in the alignment in Figure 5 and structure of dr-DXS shown in Figure 1); for instance, mutation of Glu370 (Gln or Ala) resulted in a completely inactive enzyme, demonstrating its crucial role in cofactor deprotonation during catalysis [50]. The His431 mutation (Gln or Ala) resulted only in a 50% decrease in the reaction rate, suggesting a supportive role in catalysis. For the mutants of His49 and Asp427 (Gln, Asn or Ala), they observed a drop to only 2–10% of the wild-type activity, suggesting that these residues are critical in abstracting a proton from the donor substrate [50].

In 2016, Woodcock and coworkers also performed mutational studies on the *E. coli* and *D. radiodurans* DXS. They identified two important histidine amino acids, His82 and His304, in a close proximity of 3.7 Å, which form a hydrogen bond to the thiazolium ring of ThDP. Only one of them is needed, as they can replace each other in stabilizing the intermediate

in a compensatory fashion [46]. They performed additional computational calculations and postulated that the activation of ThDP to an ylide in the active site proceeds via an activated water molecule instead of the previously assumed direct deprotonation by His434 [46].

### Studies in plants

There are many protein-engineering studies that focus on optimization of the MEP pathway for an overexpression of plant secondary metabolites. One of their findings is a feedback inhibition of DXS by IDP and DMADP. They bind to the ThDP pocket, but not as tightly as ThDP because they are smaller in size and cannot form as many stabilizing interactions as ThDP [51,52]. Compounds IDP and DMADP act as natural competitive inhibitors, but using them as a scaffold for drug design does not seem to be advisable because of their ubiquity in living organisms [53].

**Structural differences between the DXS homologues**  
In this section, we focus on the differences between the DXS homologues of the DXS enzymes of *E. coli*, *A. thaliana*, *M. tuberculosis*, *P. falciparum* and *D. radiodurans*.

The length of their primary amino acid sequence is between 600 and 700 amino acids. An alignment of the sequences using MUSCLE is shown in Figure 5. An exception is DXS of *P. falciparum*, which has a sequence length of 1200 amino acids, due to a significant number of asparagine inserts. In general, proteins with multiple asparagine inserts have the tendency to form insoluble aggregates. However, these inserts are characteristics of *P. falciparum*, and they do not seem

to affect its viability. The reasons for the insertions are still unknown. There are speculations that heat shock proteins play a major role in the metabolism of this organism, and the adaption may be due to the life cycle of the parasite with changing temperatures of the environment, depending on the host organism [54].

During SBDD against the DXS enzyme of *P. falciparum*, this major difference in the primary sequence in comparison with the other DXS enzymes should be considered. It may not be possible to apply the results of studies or computations from model enzymes to the DXS enzyme of *P. falciparum*. On the other hand, the high asparagine content is a rare property and may be an opportunity to achieve target selectivity.

Given the lack of structural information regarding crystallographic data, we decided to simulate the tertiary structure of the proteins of interest. The 3D structures were predicted using the SWISS-MODEL web service [55–57]. For all three proteins, the tool selected the already known structure of *D. radiodurans* as a template for prediction. It can be expected that the modeled structure shows similarities with the chosen template. The resulting models are shown in Figure 2. It can be seen that the secondary structure is mostly conserved between the homologues. The tertiary structure of the secondary motifs is also very similar. This may be because of the model used in the in the prediction algorithm, but it is also reasonable because all enzymes catalyze the same reaction and have conserved structural elements.

Most of the variability is found in the unordered regions at the surface and, in particular, in the loops that connect the three domains. Those loops are flexible regions, which are difficult to predict, and the modeled structures show only one of many possibilities. The longer amino acid sequence of *P. falciparum* manifests itself through a number of disordered loops. The differences in the random coiled regions between the homologues may be useful to target a specific species. This may be of particular interest when targeting the protein–protein interface. For further studies, experimentally determined crystal structures of the different homologues are needed.

### Known Inhibitors of DXS

#### Ketoclofazone

Ketoclofazone (2-(2-chlorobenzyl)-4,4-dimethyl-isoxazolidine-3,5-dione), a derivative of the herbicide clomazone, is able to inhibit the DXS of *Chlamydomonas* with an  $IC_{50}$  value of 0.1 mM [58]. In 2010, Kuzuyama and coworkers demonstrated that it is also able to inhibit the DOXP synthase in *E. coli* and *Haemophilus influenzae*, with MIC values of 800 and 12.5  $\mu\text{g}/\text{ml}$ , respectively. They proved its mechanism of action as an uncompetitive inhibitor with respect to pyruvate, which is able to bind only the enzyme–substrate complex (EA) and

not the free enzyme (E). No information was provided about the inhibitor-binding site. A rescue screen using 1-deoxy-D-xylulose, which is phosphorylated in the bacteria to 1-deoxy-D-xylulose 5-phosphate (DOXP), the product of DXS, demonstrated that the DXS enzyme is the intracellular target of ketoclofazone [47].

#### Alkylacylphosphonate derivatives

Meyers and coworkers were able to achieve a selective inhibition against DXS by the design of alkylacylphosphonate derivatives [59]. By comparison of the dr-DXS crystal structure with other ThDP-dependent enzymes, they found that the DXS enzyme shows some peculiarities both in a distinct domain arrangement with a larger active site and in the formation of a ternary complex during the catalysis (Figures 1 & Figure 7). Therefore, selectivity for DXS over the other ThDP-dependent enzymes may be obtained by using un-natural bisubstrate analogues. On this basis, they designed and prepared a series of competitive inhibitors bearing a mimic donor substrate (pyruvate) and an apolar unnatural acceptor substrate that forms a dead-end complex with ThDP in the active pocket of the DXS enzyme. They synthesized a set of alkylacylphosphonate derivatives such as butylacetylphosphonate or methylacetylphosphonate [60].

Interestingly, phosphonates modified at the alkyl position rather than the acyl moiety have more influence on the inhibitory activity. These bisubstrate analogues are active against DXS enzymes with potencies in the low micromolar range, but, unfortunately, display very poor cell permeability. Continuing the selection of substrates able to mimic pyruvate, Battistini *et al.* demonstrated that  $\beta$ -fluoropyruvate, similarly to methylacetylphosphonate, displays competitive inhibition against pyruvate, while exhibiting noncompetitive inhibition versus D-GAP [61].

#### Hydroxybenzaldoximes

Next, Meyers and coworkers extended the idea of unnatural bisubstrate inhibition of DXS using an oxime moiety as a linker between the acceptor substrate and the moiety that should occupy the pyruvate-binding site. From the diversely decorated library, they identified two new hits (2,4,5-trihydroxybenzaldoxime and 3,4,5-trihydroxybenzaldoxime), exhibiting a  $K_i$  value in the low micromolar range.

They display a unique mode of inhibition compared with other inhibitors of ThDP-binding enzymes and are competitive with respect to D-GAP [62].

#### Thiamine derivatives

Recently, we performed *de novo* design of fragments and validated their binding mode in solution using the powerful STI NMR methodology [63], consisting of

saturation-transfer difference NMR, transferred-NOE and the INPHARMA methodologies [64].

This study enabled us to discover a promising fragment (3-((1H-imidazol-1-yl)methyl)-5-methoxypyridine) featuring an  $IC_{50}$  value of  $1.8 \pm 0.5$  mM. Fragment growing led to a small molecule (Figure 7 & Table 1) having a potency in the micromolar range, demonstrating the power of this NMR-based approach to guide fragment growing in the absence of protein x-ray crystallography. Therefore, the lack of structural information available for DXS could explain the very small number of inhibitors of this enzyme reported so far. We then carried out a structure-based study by constructing a homology model of *M. tuberculosis* DXS [65]. We docked a series of ThDP and thiamine derivatives both into the homology model constructed and into the crystal structure of *D. radiodurans* DXS as these ortholog enzymes have a high degree of similarity (68% sequence identity in the ThDP-binding pocket of *D. radiodurans* and *M. tuberculosis* DXS [66]). The obtained experimental results displayed remarkable differences in the binding mode of the inhibitors tested between the two enzymes and, in turn, these evidences could guide a rational design of new potential inhibitor for *M. tuberculosis* DXS.

Bacimethrin is a thiamine antagonist, which undergoes conversion to 2'-methoxythiamine diphosphate (MeOThDP) by enzymes producing ThDP. The metabolic product is a chemical analog of ThDP in which the C2'-methyl group of the 4'-aminopyridine ring is replaced by a 2'-methoxyl group. This natural antibacterial product was isolated from *Streptomyces albus* for the first time in 1987, and recently Jordan and coworkers studied the role of MeOThDP on five ThDP enzymes including DXS [67]. By using different mechanistic tools, they found that MeOThDP binds in the active site of ThDP-dependent enzymes. The difference between ThDP and MeOThDP is a methoxy substituent at position 2' of ThDP. In their study, they investigated the influence of this modification on the reaction catalyzed by different ThDP-binding enzymes [68].

Using circular dichroism and fluorescence titration experiments they showed that the binding of ThDP to the enzyme is different for each enzyme class. They conclude that it should be possible to use the different ThDP-binding modes to target DXS selectively with a ThDP analog. For 2'-MeOThDP they found a different binding mode to the ThDP-binding pocket of DXS. Upon binding of 2'-MeOThDP to an inactive apoenzyme, it was possible to restore 9–14% of DXS enzyme activity. Investigation of other classes of ThDP-binding enzymes led from no activity restoration (2-oxoglutarate dehydrogenase) to up to 75% (PDH) activity restoration after binding 2'-MeOThDP to an apoenzyme. This points toward

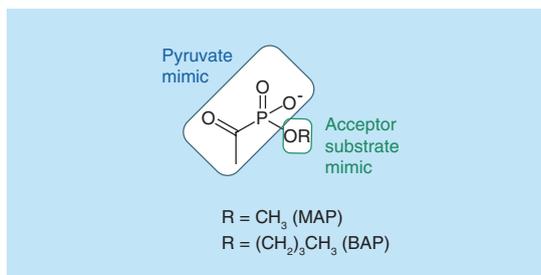
very different recognition of ThDP analogs in those enzymes (Figure 9).

### Future perspective

The MEP pathway was discovered 20 years ago and is still not fully understood. The increasing amount of research activities on the MEP pathway will deliver interesting results over the next 5–10 years. Today's focus is still on determining the structures of the constituent enzymes and on elucidating their mechanism. With this information about the structure and mechanism of the MEP pathway, targeted manipulation may be possible. This could be used to develop new drugs and herbicides, modify plant metabolism or even explore yet unknown applications.

As more information becomes available about the various targets, the number of studies on the MEP pathway will increase. As a result, the number of inhibitors will also increase. By adopting state-of-the-art medicinal-chemistry strategies, undesired side- or off-target effects and unfavorable pharmacokinetic properties should be circumvented. Upon identification of a promising scaffold, SBDD and optimization become possible. A particular problem that has to be addressed is the high polarity of the active sites of enzymes in this pathway. As a consequence, strongly binding compounds will also probably be polar, which makes them less membrane-permeable, a problem, in particular, when targeting *P. falciparum*. The development of less polar prodrugs or screening in whole-cell assays may circumvent this problem. Comparable developments are expected in the field of herbicide development.

A different field of emerging research may be the genetic engineering of organisms to overexpress and modify isoprenoids. Nature offers many complex molecules that are used for different applications, but are hard to synthesize. They are still mainly extracted from natural sources, which are sometimes hard to cultivate and often contain only small amounts of the isoprenoid of interest. The amount of plant material needed for the extraction of secondary metabolites could be reduced by several orders of magnitude. This would



**Figure 9. Designed alkylacylphosphonate derivatives as inhibitors of 1-deoxyxylulose-5-phosphate synthase.**

free resources and may even enable the use of some isoprenoids in the first place.

#### Financial & competing interests disclosure

Funding from the Dutch Ministry of Education, Culture and Science (Gravitation Program: 024.001.035) and The Netherlands' Organisation for Scientific Research (LIFT grant:

731.015.414) are gratefully acknowledged. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

#### Executive summary

##### The methylerythritol 4-phosphate pathway

- The methylerythritol 4-phosphate (MEP) pathway was discovered in 1996 and is the second known pathway for the biosynthesis of the universal building blocks of isoprenoids. This complex class of secondary metabolites fulfils numerous essential roles in living organism like protein degradation and prenylation and is the basis for the biosynthesis of steroids, carotenoids, pigments and chlorophyll.
- The pathway is found in some bacteria, the chloroplasts of plants and protozoa, and the enzymes of the MEP pathway are therefore considered as possible new targets for the development of new anti-infectives against, for instance, malaria or tuberculosis.
- The MEP pathway is used in the chloroplasts of plants to deliver the universal building blocks for the biosynthesis of the isoprenoid chlorophyll, isopentenyl diphosphate and dimethylallyl diphosphate. The mode of action of the herbicide clomazone is based on the inhibition of this pathway and validates the MEP pathway as a target for further development of herbicides.

##### The structure of 1-deoxyxylulose-5-phosphate synthase

- The inhibitor class of thiamine diphosphate (ThDP) analogues is probably robust against microbial resistance development, in particular, against acquired resistance via increased efflux due to the important metabolic role and its structural similarity with the essential vitamin B1.
- The conservation of amino acids was mapped on the protein structure of *Deinococcus radiodurans* DXS. This is a useful visualization of amino acid sequence conservation and was used to discuss previously identified binding pockets; see Figure 6.
- Pocket P1 (Figures 4 & Figure 6) is in close proximity to the active site of the DXS enzyme, and has suitable physicochemical properties and a high sequence conservation. It seems to be a good target for structure-based drug design.

##### Structural differences

- A comparison of predicted structures of the DXS enzymes of *Arabidopsis thaliana*, *Mycobacterium tuberculosis* and *Plasmodium falciparum* based on the known crystal structures from *Escherichia coli* and *Deinococcus radiodurans* shows high structural conservation of the three subdomains of the DXS enzyme, but also variability of the linkers.
- A comparison of the DXS enzyme with other ThDP-binding enzymes gives a high overall structural similarity, but also showed some differences of the active site. Based on this, it may be possible to design a ThDP analogue that targets the DXS enzymes selectively in comparison with other ThDP-binding enzymes.

#### References

Papers of special note have been highlighted as:

- of interest; •• of considerable interest

- 1 Cooper MA, Shlaes D. Fix the antibiotics pipeline. *Nature* 472(7341), 32 (2011).
- 2 Butler MS, Blaskovich MAT, Cooper MA. Antibiotics in the clinical pipeline at the end of 2015. *J. Antibiot. (Tokyo)* 70, 3–24 (2017).
- 3 Lichtenhaler HK, Rohmer M, Lichtenhaler JS, Schwender M, Lichtenhaler HK. Two independent biochemical pathways for isopentenyl diphosphate and isoprenoid biosynthesis in higher plants. *Physiol. Plant.* 101, 643–652 (1997).
- 4 Lynen F. Biosynthetic pathways from acetate to natural products. *Pure Appl. Chem.* 14 (1), 137–168 (1967).
- 5 WHO. WHO malaria report (2015). [www.who.int](http://www.who.int)
- 6 Samarasekera U. Countries race to contain resistance to key antimalarial. *Lancet* 374(9686), 277–280 (2009).
- 7 WHO. Update on artemisinin and ACT resistance – September 2015 (2015). [www.who.int](http://www.who.int)
- 8 Guggisberg AM, Amthor RE, Odom AR. Isoprenoid biosynthesis in *Plasmodium falciparum*. *Eukaryot. Cell* 13(11), 1348–1359 (2014).
- 9 Borrmann S, Adegnika AA, Matsiegui PB *et al.* Fosmidomycin–Clindamycin for *Plasmodium falciparum* infections in African children. *J. Infect. Dis.* (189), 901–908 (2004).
- 10 WHO. Tuberculosis fact sheet N°104 (2015). [www.who.int](http://www.who.int)

- 11 Migliori GB, Ortmann J, Girardi E *et al.* Extensively drug-resistant tuberculosis, Italy and Germany. *Emerg. Infect. Dis.* 13(5), 780–781 (2007).
- 12 Udawadia ZF, Amale RA, Ajbani KK, Rodrigues C. Totally drug-resistant tuberculosis in India. *Clin. Infect. Dis.* 54, 579–581 (2006).
- 13 Wright LP, Rohwer JM, Ghirardo A *et al.* Deoxyxylulose 5 phosphate synthase controls flux through the methylerythritol 4 phosphate pathway in *Arabidopsis thaliana*. *Plant Physiol.* 165(4), 1488–1504 (2014).
- 14 Duke OS, Kenyon WH, Rex NP. FMC 57020 Effects on Chloroplast Development in Pitted Morningglory (*Ipomoea lacunosa*) Cotyledons. *Weed Sci.* 33(6), 786–794 (1985).
- 15 Rohmer M, Seemann M, Horbach S, Bringer-Meyer S, Sahn H. Glycerinaldehyde 3-phosphate and pyruvate as precursors of isoprenic units in an alternative non-mevalonate pathway for terpenoid biosynthesis. *J. Am. Chem. Soc.* 118(11), 2564–2566 (1996).
- 16 Hayashi D, Kato N, Kuzuyama T, Sato Y, Ohkanda J. Antimicrobial *N*-(2-chlorobenzyl)-substituted hydroxamate is an inhibitor of 1-deoxy-D-xylulose 5-phosphate synthase. *Chem. Commun. (Camb.)* 49, 5535–5537 (2013).
- 17 Tong Y, Su P, Zhao Y *et al.* Molecular cloning and characterization of DXS and DXR genes in the terpenoid biosynthetic pathway of *Tripterygium wilfordii*. *Int. J. Mol. Sci.* 16(10), 25516–25535 (2015).
- 18 Estévez JM, Cantero A, Reindl A, Reichler S, León P. 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. *J. Biol. Chem.* 276(25), 22901–22909 (2001).
- 19 Haymond A, Johny C, Dowdy T *et al.* Kinetic characterization and allosteric inhibition of the *Yersinia pestis* 1-deoxy-d-xylulose 5-phosphate reductoisomerase (MEP synthase). *PLoS ONE* 9(8), e106243 (2014).
- 20 Goldstein JL, Brown MS. Regulation of the mevalonate pathway. *Nature* 343, 425–430 (1990).
- 21 Pokhillko A, Bou-Torrent J, Pulido P, Rodríguez-Concepción M, Ebenhöf O. Mathematical modelling of the diurnal regulation of the MEP pathway in *Arabidopsis*. *New Phytol.* 206(3), 1075–1085 (2015).
- 22 Hunter WN. The non-mevalonate pathway of isoprenoid precursor biosynthesis. *J. Biol. Chem.* 282(30), 21573–21577 (2007).
- 23 Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44(D1), D457–D462 (2016).
- 24 Hahn FM, Eubanks LM, Testa Ca, Blagg SJ, Baker Ja, Poulter CD. 1-Deoxy-D-xylulose 5-phosphate synthase, the gene product of open reading frame (ORF) 2816 and ORF 2895 in *Rhodobacter capsulatus*. *J. Bacteriol.* 183(1), 1–11 (2001).
- 25 Silver LL, Bostian KA. Discovery and development of new antibiotics: the problem of antibiotic resistance. *Antimicrob. Agents Chemother.* 37(3), 377–383 (1993).
- 26 Davies J, Davies D. Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* 74(3), 417–433 (2010).
- 27 Blair JMA, Webber MA, Baylay AJ, Ogbolu DO, Piddock LJV. Molecular mechanisms of antibiotic resistance. *Nat. Rev. Microbiol.* 13(1), 42–51 (2014).
- 28 Manzetti s, Zhang J, Van Der Spoel D. Thiamin function, metabolism, uptake, and transport. *Biochemistry* 53(5), 821–835 (2014).
- 29 Genee HJ, Bali AP, Siedler S *et al.* Functional mining of transporters using synthetic selections. *Nat. Chem. Biol.* 12(12), 1015–1022 (2016).
- 30 Wrenger C, Müller IB, Schifferdecker AJ, Jain R, Jordanova R, Groves MR. Specific inhibition of the aspartate aminotransferase of *Plasmodium falciparum*. *J. Mol. Biol.* 405(4), 956–971 (2011).
- 31 Xiang S, Usunow G, Lange G, Busch M, Tong L. Crystal structure of 1-deoxy-D-xylulose 5-phosphate synthase, a crucial enzyme for isoprenoids biosynthesis. *J. Biol. Chem.* 282(4), 2676–2682 (2007).
- **First published structures of the 1-deoxyxylulose-5-phosphatesynthase enzyme.**
- 32 Masini T, Kroezen BS, Hirsch AKH. Druggability of the enzymes of the non-mevalonate-pathway. *Drug Discov. Today* 18(23–24), 1256–1262 (2013).
- **In this paper, pockets P0–P4 of the 1-deoxyxylulose-5-phosphate synthase were first identified.**
- 33 Volkamer A, Kuhn D, Grombacher T, Rippmann F, Rarey M. Combining global and local measures for structure-based druggability predictions. *J. Chem. Inf. Model.* 52(2), 360–372 (2012).
- 34 Volkamer A, Kuhn D, Rippmann F, Rarey M. Dogsitescorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* 28(15), 2074–2075 (2012).
- 35 Li W, Cowley A, Uludag M *et al.* The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* 43, 580–584 (2015).
- 36 The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* 43, D204–D212 (2015).
- 37 Larkin MA, Blackshields G, Brown NP *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21) 2947–2948 (2007).
- 38 Gouet P, Courcelle E, Stuart DI. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15(4), 305–308 (1999).
- 39 Sankararamkrishnan R, Vishveshwara S. Geometry of proline-containing alpha-helices in proteins. *Int. J. Pept. Protein Res.* 39(4), 356–63 (1992).
- 40 Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE. Tools for integrated sequence–structure analysis with UCSF Chimera. *BMC Bioinformatics* 7, 339 (2006).
- 41 Mitschke L, Parthier C, Schröder-Tittmann K, Coy J, Lüdtkke S, Tittmann K. The crystal structure of human transketolase and new insights into its mode of action. *J. Biol. Chem.* 285(41), 31559–31570 (2010).
- 42 Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and

## Review Gierse, Redeem, Diamanti, Wrenger, Groves &amp; Hirsch

- development settings. *Adv. Drug Deliv. Rev.* 46, 3–26 (2001).
- 43 Bakail M, Ochsenbein F. Targeting protein–protein interactions, a wide open field for drug design. *C. R. Chim.* 19(1–2), 19–27 (2016).
- 44 Boratyn GM, Schäffe AA, Agarwala R *et al.* Domain enhanced lookup time accelerated BLAST. *Biol. Direct* 7(12), 1–14 (2012).
- 45 Sciences ML. Review structure, mechanism and catalytic duality of thiamine-dependent enzymes. *Cell. Mol. Life Sci.* 64, 892–905 (2007).
- 46 Cook CE, Bergman MT, Finn RD, Cochrane G, Birney E, Apweiler R. The European Bioinformatics Institute in 2016: data growth and integration. *Nucleic Acids Res.* 44, 20–26 (2016).
- 47 White JK, Handa S, Vankayala SL, Merkler DJ, Woodcock HL. Thiamin diphosphate activation in 1-deoxy-D-Xylulose 5-phosphate synthase: insights into the mechanism and underlying intermolecular interactions. *J. Phys. Chem. B* 120(37), 9922–9934 (2016).
- 48 Matsue Y, Mizuno H, Tomita T, Asami T, Nishiyama M, Kuzuyama T. The herbicide ketoclofomazine inhibits 1-deoxy-D-xylulose 5-phosphate synthase in the 2-C-methyl-D-erythritol 4-phosphate pathway and shows antibacterial activity against *Haemophilus influenzae*. *J. Antibiot. (Tokyo)* 63(10), 583–588 (2010).
- 49 Sisquella X, de Pourcq K, Alguacil J *et al.* A single-molecule force spectroscopy nanosensor for the identification of new antibiotics and antimalarials. *FASEB J.* 24(11), 4203–4217 (2010).
- 50 Brammer LA, Smith JM, Wades H, Meyers CF. 1-Deoxy-D-xylulose 5-phosphate synthase catalyzes a novel random sequential mechanism. *J. Biol. Chem.* 286(42), 36522–36531 (2011).
- 51 Querol-Audí J, Boronat A, Centelles JJ, Imperial S. Catalytically important residues in *E. coli* 1-deoxy-d-xylulose 5-phosphate synthase. *J. Biosci. Med.* 2(4), 30–35 (2014).
- 52 Rodriguez-Concepcion M. The MEP pathway as a metabolic crossroad for microbial and plant volatile organic compounds. *Plant Cell Environ.* 39(12), 2589–2591 (2016).
- 53 Banerjee A, Preiser AL, Sharkey TD. Engineering of recombinant poplar deoxy-D-xylulose-5-phosphate synthase (PtDXS) by site-directed mutagenesis improves its activity. *PLoS ONE* 11(8), 1–13 (2016).
- 54 Banerjee A, Wu Y, Banerjee R, Li Y, Yan H, Sharkey TD. Feedback inhibition of deoxy-D-xylulose-5-phosphate synthase regulates the methylerythritol 4-phosphate pathway. *J. Biol. Chem.* 288(23), 16926–16936 (2013).
- 55 Muralidharan V, Goldberg DE. Asparagine repeats in *Plasmodium falciparum* proteins: good for nothing? *PLoS Pathog.* 9(8), 8–11 (2013).
- 56 Biasini M, Bienert S, Waterhouse A *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 42, 252–258 (2014).
- 57 Kiefer F, Arnold K, Ku M, Bordoli L, Schwede T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* 37, 387–392 (2009).
- 58 Guex N, Peitsch MC, Schwede T, Workspace I. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* 30(Suppl. 1), S162–S173 (2009).
- 59 Mueller C, Schwender J, Zeidler J, Lichtenthaler HK. Properties and inhibition of the first two enzymes of the non-mevalonate pathway of isoprenoid biosynthesis. *Biochem. Soc. Trans.* 28, 792–793 (2000).
- 60 Smith JM, Vierling RJ, Meyers CF. Selective inhibition of *E. coli* 1-deoxy-D-xylulose-5-phosphate synthase by acetylphosphonates. *Medchemcomm* 3, 65–67 (2012).
- 61 Smith JM, Warrington NV, Vierling RJ *et al.* Targeting DXP synthase in human pathogens: enzyme inhibition and antimicrobial activity of butylacetylphosphonate. *J. Antibiot. (Tokyo)*. 67(1), 77–83 (2014).
- 62 Battistini MR, Shoji C, Handa S, Breydo L, Merkler DJ. Mechanistic binding insights for 1-deoxy-d-xylulose-5-phosphate synthase, the enzyme catalyzing the first reaction of isoprenoid biosynthesis in the malaria-causing protists, *Plasmodium falciparum* and *Plasmodium vivax*. *Protein Expr. Purif.* 120, 16–27 (2016).
- 63 Bartee D, Morris F, Al-khouja A, Meyers CLF, Sciences M. Hydroxybenzaldoximes are D-GAP-competitive inhibitors of *E. coli* 1-deoxy-D-xylulose-5-phosphate synthase. *Chembiochem* 16(12), 1771–1781 (2016).
- 64 Pilger J, Mazur A, Monecke P *et al.* A combination of spin diffusion methods for the determination of protein-ligand complex structural ensembles *Angew. Chem. Int. Ed. Engl.* 54 (22), 6511–6515 (2015).
- 65 Masini T, Pilger J, Kroezen BS *et al.* De novo fragment-based design of inhibitors of DXS guided by spin-diffusion-based NMR spectroscopy. *Chem. Sci.* 5, 3543–3551 (2014).
- 66 Masini T, Lacy B, Monjas L *et al.* Organic and biomolecular chemistry validation of a homology model of *Mycobacterium tuberculosis* DXS: rationalization of observed activities of thiamine derivatives as potent inhibitors of two orthologues of DXS. *Org. Biomol. Chem.* 13(46), 11263–11277 (2015).
- 67 Drautz H, Messerer W, Zahner H. Bacimethrin isolated from *Streptomyces albus*: identification, derivatives, synthesis and biological properties. *J. Antibiot. (Tokyo)*. 40(10), 1431–1439 (1987).
- 68 Nemeria NS, Shome B, Decolli AA *et al.* Competence of thiamin diphosphate-dependent enzymes with 2'-methoxythiamin diphosphate derived from bacimethrin, a naturally occurring thiamin anti-vitamin. *Biochemistry* 55, 1135–1148 (2016).

## 1.3 Recent Developments in DXPS Research

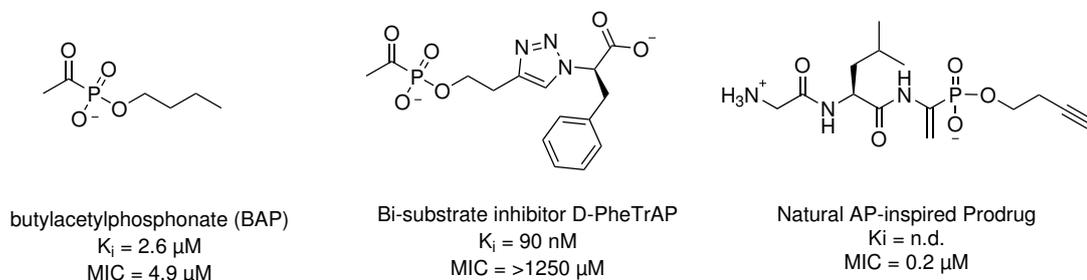
### Introduction

The review of section 1.2 was written in 2017, at the beginning of this PhD project. Over the past years, the research on DXPS has made good progress. This chapter will therefore give a brief overview on the most recent findings regarding DXPS since 2017.

### Inhibitors

Although the interest in targeting the MEP pathway is high, many projects focus on the development of inhibitors against downstream enzymes of the pathway.<sup>69</sup> To the best of our knowledge, no new class of DXPS inhibitors was published since 2017 by other groups.<sup>17</sup>

The only inhibitors currently in development are the class of covalent acetylphosphonates, initially described by Caren Freel Meyers in 2012 (Structure shown in Chapter 1.2, Table 1 and Figure 9).<sup>64</sup> They could show target engagement with DXPS and further grow the inhibitors into the glyceraldehyde 3-phosphate (3-GAP) binding site.<sup>10,60</sup> By creating bi-substrate inhibitors (Figure 1.1), also competitive to 3-GAP, selectivity over the off-target mammalian porcine pyruvate dehydrogenase (PDH) was achieved.<sup>9</sup> In 2019 Bartee et al. published a study describing enamine prodrugs of acetylphosphonates as DXPS inhibitors, enabling active uptake of the inhibitor into the bacteria via the OppA peptide permease. The prodrugs achieve a strong increase of their minimum inhibitory concentration (MIC) against Gram-negative pathogens by increasing the intracellular accumulation (Figure 1.1).<sup>11</sup> With these recent developments the alkylphosphonates, initially developed as tool compound to study the enzymatic mechanism of DXPS, are progressing into the direction of compounds with more drug-like properties.



**Figure 1.1:** Structures of BAP, the best bi-substrate inhibitor and the enamine prodrug with their activity on the drDXPS enzyme, MIC values are 90% inhibition of *E. coli* MG1665

### Enzymatic Mechanism

The exact enzymatic mechanism of DXPS has been the topic of discussion for a long time. After the discovery of DXPS, it was initially assumed the mechanism follows a classical ping-pong mechanism, characteristic for previously known ThDP-dependent enzymes.<sup>26</sup> The studies published in the following years were alternating between those arguing for an ordered sequential mechanism<sup>15,25</sup> and those in favour of the classical ping-pong mechanism.<sup>46</sup> Further complicating the discussion, Caren L. Freel Meyers and coworkers suggested 2011 a novel random sequential, preferred order reaction mechanism, which

distinguishes DXPS from other enzymes with pyruvate decarboxylase and carbonylase activity.<sup>16</sup> The mechanism proposes that both substrates are able to bind and dissociate independently of the other to the active site, with the reaction only taking place if both substrates are bound.

While this mechanism was still up to dispute in 2016, the random sequential enzyme mechanism gained more and more acceptance in recent years.<sup>72</sup> The increased acceptance was mainly due to thorough studies performed by Freel Meyers and coworkers in the last years, employing several independent biophysical methods.

As the protein crystal structures of DXPS are lacking electron density of flexible amino acid loops close to the active site (discussed in Chapter 2.1), alternative approaches to study the enzyme during catalysis were necessary. Because of this, circular dichroism (CD) and hydrogen/deuterium exchange mass spectrometry (HDX-MS) was used to study the conformational dynamics of DXPS during catalysis. Using these techniques, two distinct conformations of DXPS were observed for the first time in 2017, a closed and an open conformation.<sup>77</sup> Both conformations can be observed in the absence of substrates, while in the presence of pyruvate the closed conformation is most prominent. After addition of 3-GAP the enzyme folds preferably into the open conformation.<sup>77</sup>

A unique property of the DXPS enzyme is the stabilization of the thiamin diphosphate (ThDP) cofactor with pyruvate covalently bound, as the pre-decarboxylation C2 $\alpha$ -lactylThDP (LThDP) intermediate. In contrast, other ThDP-dependent enzymes that follow a ping-pong mechanism stabilize the post-decarboxylation ThDP adduct in its enamine state as intermediate of the first half-reaction.<sup>26</sup> Due to the stabilization of the pre-decarboxylation lactyl-intermediate, a side reaction can take place. In this side reaction O<sub>2</sub> instead of 3-GAP can act as a trigger to induce the decarboxylation of pyruvate. This was first proven by DeColli et al., who studied the reaction of DXPS in anaerobic conditions.<sup>21</sup>

## DXPS Structure

Since the determination of the first protein crystal structure of *E. coli* and *D. radiodurans* DXPS in 2007, the research on DXPS structures was silent for a long time. Only recently, in 2019, two additional structures of DXPS from *D. radiodurans* were published, as part of the Freel Meyers Lab studying the enzymatic mechanism.<sup>18</sup> The two new structures capture the drDXPS enzymes with different intermediates/dead end substrates bound and show a crystallographic proof of the previously mentioned open and closed form of the enzyme. A discussion of the conformational changes of DXPS can be found in Chapter 2.1. The two structures have the pdb-IDs 6ouw and 6ouv and improved the highest resolution of a DXPS protein model from 2.4 Å to 1.94 Å. Together with the previously published structures of drDXPS and *E. coli* DXPS (pdb-ID: 2o1x and 2o1s, respectively) a total of four structures from DXPS have been published by other labs.

No structures of DXPS from pathogenic organisms are published until now in the protein data bank. However, the calculation of homology models from *P. falciparum* and *M. tuberculosis* shows that the lack of new structures is not caused by a lack of interest in these.<sup>28,45</sup>

## DXPS-Enzyme Activity Assays

Measurement of the DXPS enzyme activity is challenging. The substrates pyruvate and 3-GAP, as well as the product 1-deoxy-D-xylulose-5-phosphate (DOXP) are polar, low

molecular-weight compounds. This makes direct measurement, for example based on LC-(MS) methods, difficult. The compounds also have no specific absorption characteristic in the UV-Vis range, which would be the method of choice to monitor enzymatic assays. As a result, most of the early research involving DXPS was performed with radio-labeled substrates, such as  $^{14}\text{C}$ -pyruvate.<sup>43</sup> To avoid the backdraws of radio isotope assays, other assays detecting DOXP were developed in the following years. One makes use of enzymatic dephosphorylation using alkaline phosphatase of the product, making it feasible for LC-analysis and detecting the 1-deoxyxylulose using a refractive index spectrometer.<sup>38</sup> A different approach was the derivatisation of DOXP using 3,5-diamino-benzoic acid, leading to an fluorescent adduct with an emission maximum at 504 nm.<sup>51</sup> Besides selective assays detecting the DOXP product of DXPS, methods are available to quantify the substrate pyruvate, either direct or using a secondary enzymatic assay such as the commercially available pyruvate oxidase kit.<sup>63</sup>

The assay with the highest prominence today is a coupled assay, initially published by Altincicek et al. in the year 2000.<sup>6</sup> The authors coupled the reaction of DXPS with the auxiliary enzyme 1-deoxy-D-xylulose-5-phosphate reductoisomerase (IspC), to follow the decrease of the IspC cofactor NADPH spectrophotometrically.<sup>6</sup> Measurement of NADPH concentration by UV-Vis absorbance at 340 nm or detection of its fluorescence emission at 460 nm made it possible to observe the reaction directly. This direct enzymatic assay enabled the study of the enzyme kinetics of the DXPS enzyme, and is, given enough supply of the auxiliary enzyme IspC, also capable of a higher throughput.

Although nowadays a wide variability of DXPS assays is available, each one comes with its own drawbacks. Time, instrument or material requirements for assays based on derivatisation are often high, whereas the auxiliary enzyme IspC of the coupled assay is still not commercially available. As a consequence, new DXPS assays are still developed today.

In 2018 a modification of the Kulka assay, adapted to measure the concentration of the product DOXP was published.<sup>62</sup> The underlying Kulka resorcinol assay was described first in 1956, and uses the reaction between the furfural intermediate with resorcinol (1,3 dihydroxybenzol) to form a colorful adduct. Shaw and Ghosh miniaturized the assay volumes and were able to decrease the concentration of HCl from 12 M down to 5.4 M, which now can be handled with automatic pipettes. As it is an end-point assay and requires sample derivatisation, it is not suitable for high-throughput applications.

A different assay was published in 2019, which measures the concentration of residual pyruvate.<sup>41</sup> The assay uses 1,2-diamino-4,5-methylenedioxybenzene as a pre-column derivatisation agent for  $\alpha$ -ketoacetic acids. Quantification is performed after HPLC separation using UV-Vis detection at 340 nm.

## Plants

The work on the DXPS enzyme is split into two separate fields with different goals. The research on bacterial enzymes is motivated by the search for new antibiotics, while the DXPS enzymes of plants are studied mainly to improve the yield of rare secondary plant products.

Much of the research performed in recent years on the MEP pathway from plants is dedicated to improve the expression levels of MEP pathway enzymes and by this, increase the biosynthesis of secondary metabolites. Since the description of DXPS as an

rate-limiting enzyme of the MEP-Pathway in 2001, this plant enzyme moved into the focus of attention of several research groups.<sup>24</sup>

Studies describing the localization and level of expression of DXPS are now published frequently for different plants, for example *Pinus massoniana*<sup>39</sup>, *Andrographis paniculata*<sup>65</sup>, *Populus trichocarpa*<sup>76</sup> and *Castilleja tenuiflora* Benth<sup>58</sup>. Just recently Rong Li et al. showed that DXPS overexpression in transgenic *A. thaliana* is able to increase the cellular levels of chlorophyll and carotenoids.<sup>39</sup> The increased expression of DXPS can have several secondary effects on the plant, caused by the higher levels of terpenes. Recently the over-expression of DXPS in *Populus trichocarpa* showed an increased resistance of the plant to stress by insects or fungal infection.<sup>71</sup>

Although performed with different goals in mind, some findings are transferable between the research field of plant and bacterial DXPS. In particular, as the microbial production of terpenoids originating from plants is considered in some cases over the optimization of whole plants. The resulting studies of the enzymatic mechanism and thermal stability of plant DXPS overlaps with the work done on bacteria. A recent example of this is a computational analysis of *B. subtilis* as a producer of terpenoids.<sup>59</sup> Performed with the intention to increase the carbon flux of the MEP pathway, amino-acids conserved in bacteria and their putative functionality were identified:

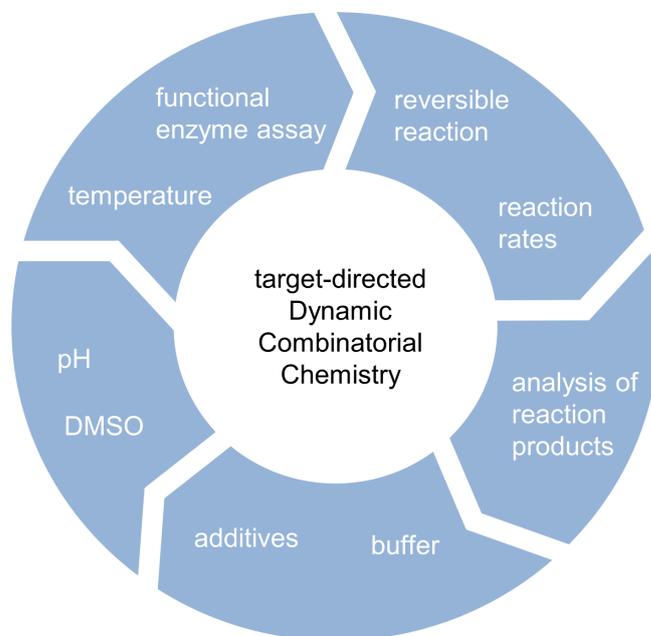
*"Analysis of the top-ranked missense mutations for the 14 hotspots through POPMUSIC and MAESTRO affirm the biological credibility of only 5 mutations (...) The P318D mutation indicates a higher chance of improving the thermostability of DXPS. Since DXPS is the most crucial enzyme to direct the carbon flux toward the biosynthesis of terpenoids in B. subtilis, the present study might be helpful to develop its functionally improved variants for improving the microbial production of terpenoid-based flavoring, fragrance, and therapeutic compounds."*

These findings could be beneficial for the design of new bacterial and plant DXPS enzymes with improved properties, like protein crystal formation or thermostability, and provide a good example for the connection of the field of plant and bacterial DXPS research.



## 1.4 Protein-Templated Dynamic Combinatorial Chemistry

The following chapter was published as an open-access article in the European Journal of Organic Chemistry under the terms of the creative commons license CC BY-NC 4.0 and is printed in this thesis without any changes.<sup>2</sup>



### Title

Protein- Templated Dynamic Combinatorial Chemistry: Brief Overview and Experimental Protocol

### Authors

Alwin M. Hartman<sup>‡</sup>, Robin M. Gierse<sup>‡</sup> and Anna K. H. Hirsch

### Contributions

A. M. Hartman conceived the manuscript and took part in writing. R. M. Gierse took part in writing of the manuscript and performed the protein stability studies. A. K. H. Hirsch was involved in writing and supervised the project.

### Bibliographic Details

European Journal of Organic Chemistry  
Volume 7, Issue 22, Pages 3581-3590  
3 May 2019  
DOI: 10.1002/ejoc.201900327

<sup>‡</sup>: Authors contributed equally

## Combinatorial Chemistry | Very Important Paper |

# VIP Protein-Templated Dynamic Combinatorial Chemistry: Brief Overview and Experimental Protocol

Alwin M. Hartman,<sup>[a,b,c][‡]</sup> Robin M. Gierse,<sup>[a,b,c][‡]</sup> and Anna K. Hirsch<sup>\*[a,b,c]</sup>

**Abstract:** Dynamic combinatorial chemistry (DCC) is a powerful tool to identify bioactive compounds. This efficient technique allows the target to select its own binders and circumvents the need for synthesis and biochemical evaluation of all individual derivatives. An ever-increasing number of publica-

tions report the use of DCC on biologically relevant target proteins. This minireview complements previous reviews by focusing on the experimental protocol and giving detailed examples of essential steps and factors that need to be considered, such as protein stability, buffer composition and cosolvents.

## Introduction

Since its dawn more than two decades ago, combinatorial chemistry approaches<sup>[1–5]</sup> have developed into target-directed dynamic combinatorial chemistry (tdDCC) and have matured as a hit-identification tool.<sup>[6–11]</sup> A growing number of groups have shown the general applicability and scope of tdDCC for the identification of modulators of targets.<sup>[6,12–20]</sup> tdDCC refers to general pharmacologically relevant targets which next to pro-

teins also include DNA and RNA, whereas protein-templated DCC (ptDCC) only refers to proteins. Several reviews and book chapters on tdDCC have been published in recent years.<sup>[21–23]</sup> This minireview covers our work on ptDCC and provides the key features of our protocol, explaining the essential steps in designing a successful ptDCC experiment.

Carefully chosen building blocks are connected in a reversible manner via covalent or noncovalent bonds to form a dynamic combinatorial library (DCL) (Figure 1). Biocompatibility,

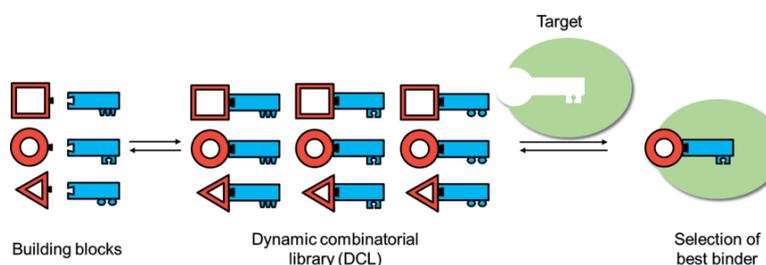


Figure 1. Schematic representation of target-directed dynamic combinatorial chemistry.

[a] Department of Drug Design and Optimization, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI),

Campus Building E8.1, 66123 Saarbrücken, Germany

E-mail: Anna.Hirsch@helmholtz-hips.de

<http://www.helmholtz-hzi.de/hirsch>

[b] Stratingh Institute for Chemistry, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

[c] Department of Pharmacy, Medicinal Chemistry, Saarland University,

Campus Building E8.1, 66123 Saarbrücken, Germany

[‡] These authors contributed equally to this work.

ORCID(s) from the author(s) for this article is/are available on the WWW under <https://doi.org/10.1002/ejoc.201900327>.

© 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

pH dependence, temperature, solubility and stability of the components are important factors, which should be taken into account. The ideal DCLs do not require cosolvents, however, it can occur that the formed products have a lower solubility than the building blocks and in order to keep all compounds in solution, a cosolvent such as dimethyl sulfoxide (DMSO) is commonly used. Precipitation of DCL components could lead to an undesired shift in the equilibrium. By contrast, a desired shift of the equilibrium can be obtained by the addition of an external stimulus, such as a protein target. There are in general two different approaches that can be followed in ptDCC: “adaptive DCC”, in which the target is present during the formation of the DCL and “pre-equilibrated DCC”, in which the target is added after the DCL is established. An advantage of pre-equilibrated DCC is that the exchange chemistry can be applied in

conditions which are not tolerated by the protein. A disadvantage is that the screening step is performed under static conditions and no amplification effects can be observed since the protein does not alter the equilibrium.

In ptDCC, the member(s) of DCLs, which bind best will be amplified, leading to an increase in their concentration com-

pared to a control reaction without the external stimulus. These binders can then be further evaluated for their biochemical properties.

To enable a comparative analysis of DCLs, a blank reaction, without the target, should be run concurrent with a templated reaction. Another approach of DCC is non-comparative, in which the hits can be analyzed in complex with the target or after being released from the target. There are different techniques that can be used to analyze the DCLs: liquid and size-exclusion chromatography coupled to mass spectrometry, NMR spectroscopy, fluorescence spectroscopy and X-ray crystallography. Figure 2 illustrates the comparative approach vs. the non-comparative approach, which can be adopted in DCC. The reaction mixture can be "frozen", in order to prevent the library from re-equilibrating during the analysis. In the case of acyl-hydrazone chemistry, this can be achieved by an increase in pH. Denaturation by heat, addition of a solvent or (ultra-fast) centrifugation ensures that all binders are released from the protein before analysis.

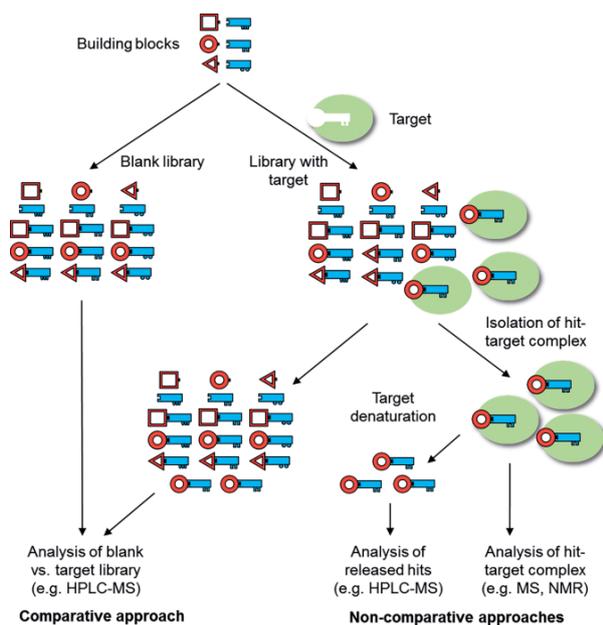


Figure 2. DCC approaches: comparative and non-comparative. In the comparative approach the library in presence of a target is compared to the library in absence of the target. In the non-comparative approaches, the hit-target complexes will be separated from the mixture and analyzed as a complex or as released hits. Adapted from Frei et al.<sup>[21]</sup>

## 1. Reversible Reaction Suitable for DCC

Only a limited number of reversible reactions have been used thus far, they are summarized in Scheme 1. One of the most frequently used reactions is the (acyl)-hydrazone formation, which combines ketone or aldehyde building blocks with (acyl)-hydrazides. This condensation reaction can take place in water, making it biocompatible.<sup>[24]</sup> The synthesis of the building blocks is generally straightforward or they may be commercially available.

At physiological conditions, neutral pH and room temperature, acylhydrazone formation and exchange are relatively slow.



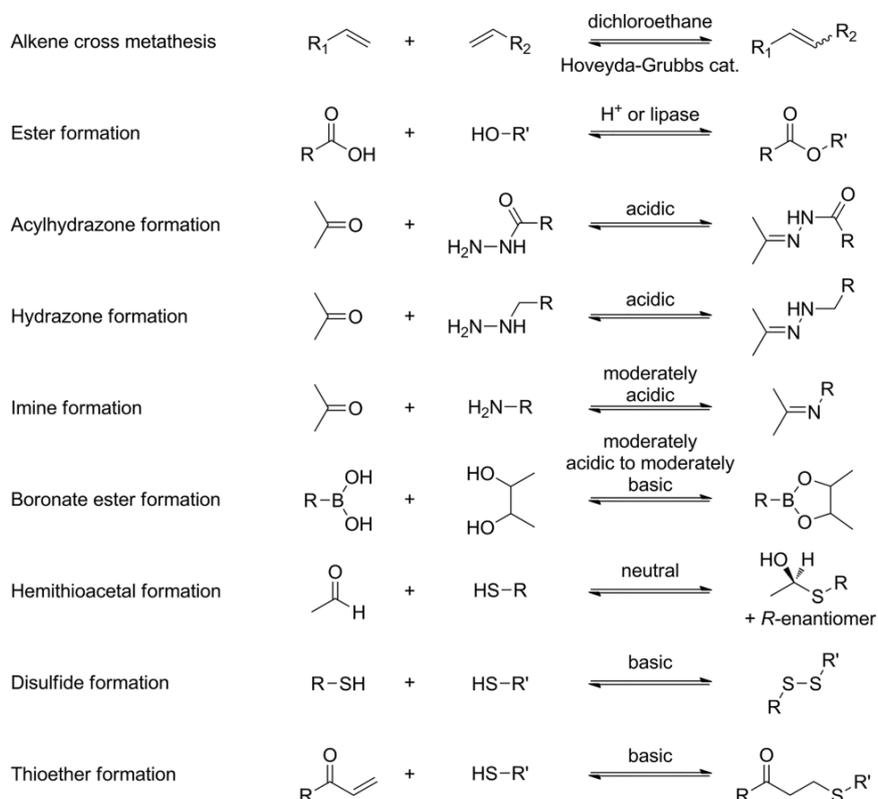
*Alwin M. Hartman studied Chemistry at the University of Groningen. In his Master's research, he synthesized inhibitors of the aspartic protease endo-thiasepsin in the Hirsch group. In September 2015, he started his PhD research in the same group, focussing on new applications of dynamic combinatorial chemistry to medicinal chemistry and chemical biology.*



*Robin M. Gierse studied Biochemistry at the University of Greifswald. He obtained his M.Sc. with a thesis on the synthesis of crosslink-active microRNAs in the bio-organic chemistry lab of Prof. S. Müller. Subsequently, he worked at the company Enzymicals as a junior scientist. He joined the Hirsch group in the fall of 2016 as a PhD student. His research focuses on the development of novel anti-infectives and includes molecular and structural biology as well as computational drug design.*



*Anna Hirsch read Natural Sciences at the University of Cambridge and developed the double conjugate addition of dithiols to propargylic carbonyl systems in the group of Prof. Steven V. Ley. She received her Ph.D. with Prof. François Diederich from ETH Zurich in 2008 on de novo design and synthesis of the first inhibitors of an anti-infective target. After a postdoc in the group of Prof. Jean-Marie Lehn in Strasbourg, she took up a position as assistant professor at the Stratingh Institute for Chemistry at the University of Groningen in 2010 and was promoted to associate professor in 2015. In 2017, she became head of the department for drug design and optimization at the HIPS. Her work focuses on rational approaches to drug design, including structure- and fragment-based drug design in combination with dynamic combinatorial chemistry and kinetic target-guided synthesis, focusing on anti-infective targets*

Scheme 1. Reversible reactions used in target-directed DCC to identify bioactive compounds. Adapted from Van der Vlag and Hirsch.<sup>[23]</sup>

At acidic pH, the equilibrium is reached rapidly. However, Greaney and co-workers have shown that the pH dependence can be influenced by the addition of a nucleophilic catalyst.

They were able to reach equilibrium reasonably fast at a comparatively high pH of 6.2 by using aniline, as a nucleophilic catalyst.<sup>[12]</sup> Previously Dawson and co-workers have shown that

**Table 1. Protein-templated DCC studies reported over the past five years, in which a target was used as a template to influence the equilibrium. Therefore, only articles using an adaptive approach are listed, pre-equilibrated DCC examples are omitted.**<sup>[29–31]</sup> The table is adapted from Frei et al. and complemented.<sup>[32] [a]</sup>

Target	Reversible reaction	Analysis	Library size	Equilibration time	Method applied for affinity measurement	Best affinity	Ref.
Wt Tau RNA	Disulfide	HPLC-MS and NMR	21	2 days	Fluorescence titration	$EC_{50} = 70 \text{ nM}$	Artigas et al. 2015 <sup>[33]</sup>
HIV F55 RNA	Disulfide	MS	12	4 days	n.a.	n.a.	McAnany et al. 2016 <sup>[34]</sup>
Vascular endothelial growth factor receptor (VEGFR) 2	Imine	HRMS	297	24 h	In vitro activity against cancer cell lines	$IC_{50} = 2.4 \text{ }\mu\text{M}$	Yang et al. 2016 <sup>[35]</sup>
Endothiapepsin	Acyldiazone	HPLC-MS	90	20 h	Inhibition assay	$IC_{50} = 54.5 \text{ nM}$ $K_i = 25.4 \text{ nM}$	Mondal et al. 2016 <sup>[36]</sup>
FimH	Acyldiazone	HPLC	8	3 days	SPR	$K_D = 273 \text{ nM}$	Frei et al. 2017 <sup>[37]</sup>
UDP-galacto-pyranose mutase	Acyldiazone	HPLC	11	24 h	Fluorescence-based assay and MIC	$K_D = 3 \text{ }\mu\text{M MIC} = 26 \text{ }\mu\text{g mL}^{-1}$	Fu et al. 2017 <sup>[38]</sup>
Myeloperoxidase (MPO)	Hydrazone	Activity assay	6	n.a.	in vivo activity assay	$IC_{50} = 79 \text{ nM}$	Soubhye et al. 2017 <sup>[39]</sup>
ecFabH	Acyldiazone	<sup>19</sup> F-NMR	5	12 h	Enzymatic assay	$IC_{50} = 3 \text{ mM}$	Ektström et al. 2018 <sup>[40]</sup>
Multi-protein strategy on AlkB oxygenases: FTO, ALKBH3 and ALKBH5	Acyldiazone	DSF and HPLC	10	5 h	HPLC-based demethylase and DSF assays	$IC_{50} = 2.6 \text{ }\mu\text{M}$	Das et al. 2018 <sup>[41]</sup>
<i>Trypanosoma cruzi</i> bromodomain-containing (TcBDF3)	Acyldiazone	HPLC-MS	30	n.a.	DSF	$IC_{50} = 13\text{--}23 \text{ }\mu\text{M}$	García et al. 2018 <sup>[42]</sup>
G-Quadruplex DNA	Imine formation	HPLC and ESI-MS	10	24 h	n.a.	n.a.	Jana et al. 2019 <sup>[43]</sup>

[a] DSF = differential scanning fluorimetry, HPLC = high-pressure liquid chromatography,  $IC_{50}$  = half maximal inhibitory concentration, ITC = isothermal titration calorimetry,  $K_D$  = dissociation constant,  $K_i$  = inhibition constant, MIC = minimum inhibitory concentration, MS = mass spectrometry, n.a. = not available, NMR = nuclear magnetic resonance, SPR = surface plasmon resonance.

aniline could serve as a catalyst for acylhydrazone formation and oxime ligation.<sup>[25,26]</sup> Derivatives of aniline, which bear substituents at the aryl ring, are even more effective catalysts.<sup>[27]</sup>

The acylhydrazone linkage is reversible but sufficiently stable to allow for direct analysis under acidic conditions and stable against hydrolysis at physiological pH values, allowing for the “freezing” of the reversible reaction upon increasing the pH.<sup>[24]</sup>

An overview of studies published over the past five years in the field of ptDCC is given in Table 1. It must be noted that much more work has been published applying DCC for the formation of diverse libraries in the drug-discovery process. For example the coupling of DCC to DNA-encoded libraries, creating so called DNA-encoded dynamic combinatorial chemical libraries (EDCCLs). Iminobiotin and homotetrameric streptavidin were used as a model system to identify a bidentate protein/ligand interaction. The addition of an external stimulus, for example a target protein, can shift the thermodynamic equilibrium and hence a DNA amplification can be observed after sequencing.<sup>[28]</sup>

## 2. A Closer Look on the Templating Protein

To obtain meaningful results from DCC experiments, the quality of the input template is critical. As the equilibrium of the library shifts by the templating effect of the added protein sample, it should consist of the target protein as close to its native state as possible. The quantity of the used template depends on the protein target, there are reported successful DCC projects with 0.1 to 1.5 equivalents of protein.<sup>[29,44]</sup> DCC experiments are also possible with a mixture of proteins, but a well-defined sample eases up downstream data analysis and reduces the number of false positives for the desired target.<sup>[41]</sup> The condition of the protein sample depends on various variables. For DCC experiments the purity, concentration, tertiary and quaternary structure of the protein, additives and contaminations, as well as the pH-value are of particular importance. During the experiment, which can take up to several days, protein degradation and precipitation could occur. The tests described herein should give an overview and help to choose suitable experimental conditions to plan new DCC experiments. In the next paragraphs, we will briefly discuss the influence of those factors and suitable analytical methods to monitor them.

### 2.1. Purity

In the case of a mixed or impure protein sample, there might be several templated reactions proceeding in parallel. It is impossible to differentiate between a small fraction of the sample showing a strong template effect and a large fraction of the protein pool showing only a weak amplification of a binder. This will result in overlapping data, which are difficult to analyze, and may result in false positives. We therefore recommend starting with the highest protein purity available.

### 2.2. Stability

Not only the initial state, but also the stability of the templating protein during the reaction should be checked by preliminary

tests before conducting a DCC experiment. The time span over which a DCC experiment, pre-equilibrated or adaptive, is monitored can vary. It depends on the reaction rate and concentration and should ideally be monitored until the library reaches an equilibrium state. Usually, the DCL reaches a new equilibrium within the first few days, depending on the reversible reaction and conditions used (Table 1). However, if the protein is stable for longer periods of time, longer equilibration times are possible, for example up to 20 days for the very stable protease endothiasepsin (see Section 2.4).<sup>[24]</sup>

It is important that the protein is not precipitating or degrading during the experiment. Precipitation of the protein will remove the template from the solution. Denaturation of the template will lead to entirely new templates, which would affect the equilibrium state of the DCL. This can lead to random and irreproducible amplification of compounds by the unordered protein and a decrease of initially already amplified best binders of the native template. If the protein target is labile, it is therefore necessary to follow the reaction over time to identify the temporary, templated equilibrium of the DCC library. In this, compounds amplified by the native state of the template can be found.

Eventually, after prolonged incubation times, nearly every protein will degrade and, by this, change the equilibrium of the DCL again. Compounds amplified in this step should be disregarded, as they were not templated by the native protein. Observation of the DCC experiment for longer timeframes than the template's stability under the specific conditions should therefore be avoided.

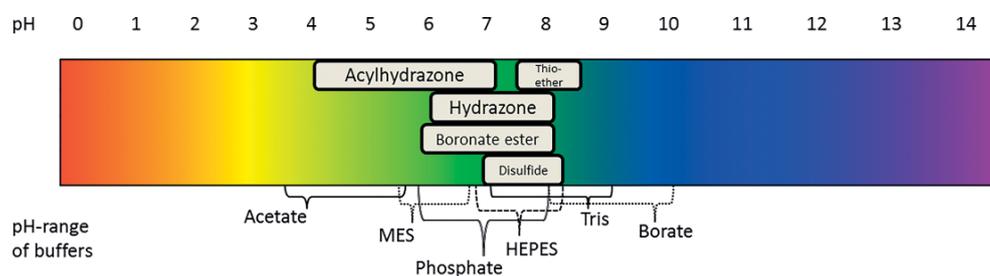
### 2.3. Buffer and pH

When choosing a buffer for DCC experiments, several different requirements have to be met. Attention should be paid to possible side reactions with the DCL or chelation effects. For example, Tris buffer could form imines with aldehyde building blocks, which might influence the formation of the DCL. Some stabilization of the protein is beneficial, but strong interactions of the buffer with the target protein should also be avoided, for instance, a phosphate buffer for a phosphate binding protein. The phosphate could compete with possible binders; possible effects of competition are discussed in more detail in Section 2.5. So far, in most cases common buffer systems have been used, which are shown in Table 2 and Scheme 2. The choice of buffer is, however, not limited to the established systems.

Table 2. Buffers commonly used in different DCC reactions. \*Tris buffer requires special attention.

Reaction	Buffer described in literature
Acylhydrazone formation <sup>[6,24,37]</sup>	Ammonium- and Sodium acetate, Phosphate, Tris*
Hydrazone formation <sup>[45,46]</sup>	Phosphate, Tris*
Disulfide <sup>[47,48]</sup>	Phosphate, Borate
Thioether <sup>[49]</sup>	Water/DMSO
Imine <sup>[13]</sup>	Water
Boronate ester <sup>[50,51]</sup>	Ammonium acetate, Water

For many protein targets, the stability at room temperature and the optimal buffer conditions are not known. We therefore



Scheme 2. Example of possible buffers and the pH ranges of reactions used in DCC experiments.

recommend determining these conditions prior to performing DCC experiments. As several interdependent factors, like pH, buffer, ionic strength and ions influence the stability of a protein, it is difficult to suggest a stepwise flow scheme for the determination of the ideal buffer composition for a given protein.<sup>[28]</sup> Not only the protein but also the exchange chemistry might be affected significantly by varying these parameters. We propose to first measure the effect of pH, buffer and ionic strength over a wide range in parallel. Afterwards, a small selection (2 to 5) of the most stabilizing combinations can be evaluated for their long-term effect on the protein. Subsequently, the best condition will then be used to determine the influence of DMSO (Section 2.6) and, if of interest, additives (Section 2.5). The selection of the initial buffers could be broadened, in case no suitable condition was found.

Two or more buffers should be screened per pH value to distinguish the influence of the buffer component and the pH value on the stability of the protein. It is also possible to use a so-called "superbuffer",<sup>[52]</sup> a mix of three or more buffer components, enabling the adjustment of a wide pH range, without changing the buffer composition or concentration.

The effect of the buffer components on a protein can be measured in a straightforward way, by determining the melting point of the target protein via a thermal-shift assay/differential scanning fluorimetry (TSA/DSF).<sup>[28]</sup> In this method, the protein is incubated together with a lipophilic dye, for example sypro orange. The dye shows an increase in fluorescence after binding to the hydrophobic parts of a protein. These are often located at the inside of a protein and become exposed during temperature-induced unfolding/melting. The temperature-dependent increase in fluorescence can be measured in a RT-PCR apparatus and yields the  $T_m$  of the protein.

Other methods, like DSC, ITC and CD (differential scanning calorimetry, isothermal titration calorimetry and circular dichroism spectroscopy) and the determination of melting points by CD could also be used to gain information on the interaction and possible stabilization of the protein with its buffer, but require a high amount of protein and/or long measurement time. The TSA, however, offers high throughput and a short assay time, together with already several published or commercially available kits in 96-well format.<sup>[53,54]</sup> These kits were originally intended to screen for optimal crystallization conditions and cover several stability-influencing conditions. When performing DCC experiments, the design of an individual 96-well plate layout, tailored to the buffers and conditions compatible with the

planned DCC reaction, might be useful. This is a short time investment, which might pay off quickly in the future, if ptDCC is used on several different targets.

After a DCC-compatible, stabilizing buffer condition has been identified, the protein should be checked for its long-term stability. To check for cleavage of the protein backbone an analysis by SDS-PAGE is of sufficient sensitivity (Figure 3). To determine if the protein folding is affected, TSA is again the method of choice, since the signal directly depends on the unfolding process of the protein. With prolonged degradation, the melting point decreases slightly. As a secondary effect of the degradation, the fluorescence curve can show bi- and multiphasic melting points and an overall decrease in signal intensity and resolution. A fully denatured enzyme will just show a decreasing fluorescence signal with no peak from protein unfold-

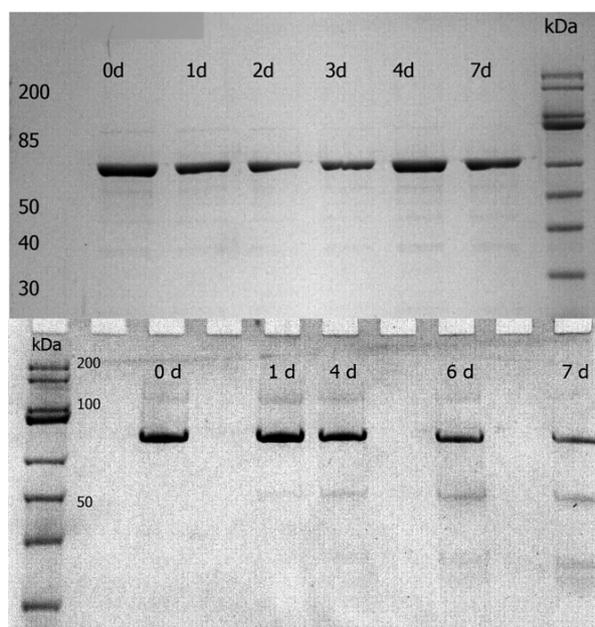


Figure 3. 12 % SDS-PAGE of different homologues of the enzyme 5-deoxyxylulose 5-phosphate synthase (DXS) after incubation at RT. The protein on the upper gel shows no sign of degradation. The second protein, shown on the lower gel, shows signs of degradation, starting already at day one with a very faint band around 50 kDa. From day 6 on a decrease of the main protein band also becomes clear. In the top left corner, a gel-label was removed using image processing software.

ing. As controls, a fresh and a heat-treated sample of the target protein should be included in the experiment.

The tendency of a protein to precipitate is concentration-dependent. Because of this, the assays determining the protein stability should be performed with the same protein concentration that is intended to be used in the DCC experiment. If this is not possible, due to limited protein availability, the first experiments might be done with less protein. However, at least for the chosen final condition, the stability assessment should be repeated with the protein concentration that will be used in the DCC experiments.

## 2.4. Functional Enzyme Assay

For enzymatic protein targets, a functional assay can be used instead of TSA and PAGE measurements for the assessment of long-term stability. The analysis of activity data of a functional assay to determine the best experimental conditions of the DCC experiments leaves less room for interpretation than the analysis of the results of a melting-point analysis. Therefore, if a functional assay is available, and the enzyme is showing catalytic activity in the desired pH range of the DCC reaction, the activity assay should be the method of choice.

In a previous study from 2014, we could monitor the activity of the target protein endothiapepsin by performing a fluorescence-based assay (Figure 4). The pH optimum of endothiapepsin is 4.5, and the enzymatic activity was not affected even after 20 days incubation at RT and a pH of 4.6. Considering this high stability, no buffer optimization was needed.<sup>[24]</sup>

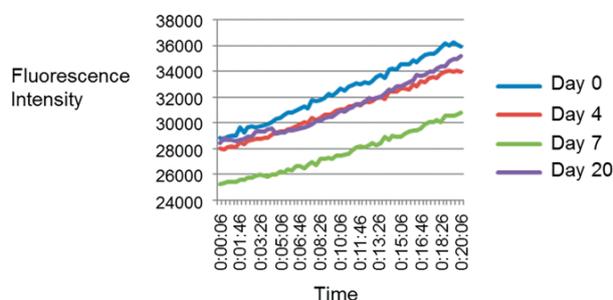


Figure 4. Activity of endothiapepsin, a pepsin-like aspartic protease, in a fluorescence-based assay at different time intervals of incubation at room temperature. Adapted from Mondal et al.<sup>[24]</sup>

## 2.5. Additives and Contaminations

During the purification, the protein might be in contact with different buffers and conditions. Some of the buffer components might remain bound to the protein, even after buffer exchange. These contaminants might influence the experiment. It is therefore recommended to critically evaluate the composition of the protein sample. Not only should the protein storage buffer be evaluated, but also the origin of the sample.

Common substances that could be found in protein samples are for example imidazole as a leftover from an IMAC (immobilized metal affinity chromatography) purification step. Protein

samples are often supplemented with reducing agents like 2-ME, DTT or TCEP (2-Mercaptoethanol, Dithiothreitol or Tris(2-carboxyethyl)phosphine) in concentrations up to 10 mM to keep the protein in a reducing environment. If disulfide formation is the reversible reaction of choice, the final reducing agent concentration should be evaluated to make sure that the formation of disulfide bonds is not inhibited.

The effect of additives and contaminations is related to the volume of the protein sample used in the individual DCC experiments. This not only determines the final concentration of protein, but also the concentrations of the contaminants. If the batch-to-batch concentration of the protein varies and its volume is adjusted to reach the same final concentration in the DCC experiment, it should be noted that the concentrations and effects of the additives in the DCC experiment might vary.

Compounds that remain in the protein sample can have an influence on the DCC reaction or on the target protein. One should critically check every buffer component on possible interference with the planned exchange chemistry. Performing a control experiment with all buffer components, in the absence of protein, can assure that no side reactions are taking place. If the exact composition of the protein sample is unknown, a small volume of the buffer might be obtained by concentration of the protein using an ultrafiltration device and using the flow-through for the control experiment.

Some agents used during protein purification, such as cryoprotectants like glycerol or detergents like Tween, will interact in a non-specific way with the protein surface. From our experience, if there is no hint that they might affect the experiment, leftover cryoprotectants and detergents can be tolerated. Special care should be taken if cofactors, coenzymes or ions are supplemented during the purification process to stabilize the enzyme. The same holds true for buffer components structurally related to those supplements. Everything that binds to the targeted binding pocket is competing with the DCC library. If a natural, tight binding cofactor is present during the experiment, it could prevent the building blocks from binding and therefore also inhibit their amplification. However, the use of tight binders can be beneficial in control experiments. If a compound with a known binding site is inhibiting the formation of some previously observed binders this can be taken as a hint that the templated binders are targeting the same protein pocket.

## 2.6. DMSO

Addition of a small percentage of DMSO to the reaction solution is a common practice in the design of enzymatic assays to improve the solubility of hydrophobic compounds. For biochemical assays, DMSO concentrations up to 10 % are regularly used.<sup>[55]</sup>

In DCC experiments, the building blocks of the library are typically dissolved in DMSO stock solutions to enable the easy assembly of a library. Depending on the library composition and number of compounds used, the final DMSO concentration would vary. To keep the reaction conditions comparable, we recommend adding DMSO up to a concentration that can be kept constant for all experiments of a project. This fixed concen-

tration should be evaluated and chosen beforehand, to ensure the protein tolerates it.

DMSO has a very broad range of effects on proteins, it can even decrease the solubility and induce precipitation.<sup>[56]</sup> Both, rate acceleration, as well as inhibition of the enzyme-catalyzed reaction by DMSO have been observed. An influence of already low percentages of DMSO on the enzymatic activity often hints to DMSO acting as an unspecific effector, interacting with the active site of the enzyme.<sup>[57]</sup> If the enzymatic activity is reduced by DMSO at higher concentrations (>10 % DMSO), it is often by influencing the overall protein conformation by displacing water molecules bound to the surface and unfolding the protein.<sup>[58]</sup> On the other hand, there are DMSO-tolerant enzymes known which show activity up to 80 % DMSO.<sup>[57]</sup> Enzyme activity assays are the method of choice to estimate the effect of DMSO on an enzyme. If no activity assay is available, the effect of DMSO could also be measured using TSA, however, interactions with the active site are difficult to detect with this method. We often observe a small effect on the  $T_m$  of a protein, but a strong effect on the enzymatic activity. Taken together, the DMSO concentration has several effects on the protein structure. The benefits of DMSO addition need to be weighed against the risk of creating an artificial enzymatic fold, which could amplify compounds that would not bind under native conditions. Therefore, the DMSO concentration should be as low as possible, in our lab up to 5 % are regularly used.

## 2.7. Temperature

To speed up the rate at which the DCL reaches equilibrium, the experiments are normally performed at room temperature. For labile proteins, a lower reaction temperature may be necessary, which can improve the stability of the proteins. At the same time, the equilibration rate is decreased, leading to a prolonged incubation time. The optimal temperature for protein stability in DCC could vary from enzyme to enzyme and thus needs to be evaluated in each individual case but room temperature is used in most cases.

## 3. Setting up a ptDCC Experiment

When crystal structures are available, or even co-crystal structures, a structure-based approach can be undertaken to design promising building blocks. In this case, also non-binders could be designed as control elements, which are not supposed to

emerge as hits. The type of reversible linkage should be carefully selected because it influences the molecular recognition by the target. For example, the acylhydrazone linkage resembles the amide functionality and features hydrogen-bond donors and acceptors. We showed that by combining DCC with de novo structure-based design, the risks associated with this attractive approach are reduced.<sup>[24]</sup>

### 3.1. Formation of the DCLs

The building blocks might have to be dissolved in DMSO, allowing them as well as the formed products to stay soluble in the final mixture. In principle, they could also be dissolved in the desired buffer, which would be most ideal. In 2014, we coupled DCC to saturation-transfer difference (STD) NMR spectroscopy, which requires lower concentrations of protein than a general DCC experiment (Table 3). STD-NMR spectroscopy enables selection of the binders from the DCL, since the intensity of these signals is stronger due to a more efficient saturation transfer. As a result, STD-NMR spectra cannot be used to determine concentrations of DCL members and therefore amplification cannot be calculated. In follow-up experiments, it is possible to determine the  $K_D$  value of a ligand via STD-NMR or other biophysical assays.<sup>[59]</sup>

The ratio of hydrazides vs. aldehydes should allow for the formation of all possible products, therefore at least one equivalent of each hydrazide per aldehyde should be used. For example, if three aldehydes are used then at least three equivalents of each hydrazide should be added, making sure that there is an excess of hydrazides. When required, a nucleophilic catalyst like aniline could be added. The most frequently used concentration of DMSO lies around 5–10 %.

Control experiments should be considered, which should clarify where binding of molecules to the protein occurs and if it is specific or unspecific. This could for example be performed by the addition of a known inhibitor. If the previously observed amplification is not observed any longer, then the hit compounds are competitive binders. Based on the work of Danieli et al., B. Ernst and co-workers propose that the use of bovine serum albumin (BSA), as a negative control template for which no amplification is expected since the binding pocket is different, is not a good control since it could influence the library composition, whilst the use of a competitive inhibitor is better. BSA has been used in DCC to show that the applied library only gives hits with the real target and that BSA would yield the same result as the blank.<sup>[32,60]</sup> BSA is commonly known for its

Table 3. General protocol for DCC and protocol for DCC coupled to <sup>1</sup>H-STD-NMR. \* Aniline or another nucleophilic catalyst could be added when required. \*\* In a control experiment, no protein is added. \*\*\* Buffer conditions to guarantee protein stability should be determined a priori.

	Final concentration in general DCC	Final concentration used in DCC coupled to <sup>1</sup> H-STD-NMR <sup>[24]</sup>
Aldehyde	0.1 mM	0.4 mM
Hydrazide	0.1–0.3 mM	1 mM (for each of the five hydrazides)
DMSO	5–10 %	5–10 %
Aniline*	10 mM	–
Protein**	10–100 μM	4 μM
Buffer***	0.1 M	Ammonium acetate in D <sub>2</sub> O (0.1 M, pH 4.6)
pH*	Acidic–neutral	pH 4.6

stability and was thought not to interfere with biological reactions, however recently DCC experiments have even been used to target BSA.<sup>[61]</sup>

### 3.2. Analysis of the DCLs

Different techniques such as fluorescence-polarization, SPR, ITC, MST, STD-NMR, crystallography and others can be used to evaluate and possibly optimize obtained hits. We and Rademann and co-workers have reviewed the analytical methods used in protein-templated dynamic combinatorial chemistry to detect hit compounds.<sup>[23,62]</sup>

A commonly applied method to analyze DCC experiments is the recording of HPLC-MS chromatograms of the libraries. As an illustrative example of the comparative approach, we drew HPLC chromatograms of a blank library and a target library (Figure 5). When we compare both chromatograms, we see that peak number five has increased in the library containing the target, whereas peaks three and six have decreased. The total amount of building blocks stays the same, only the equilibrium can be shifted towards one or more products.

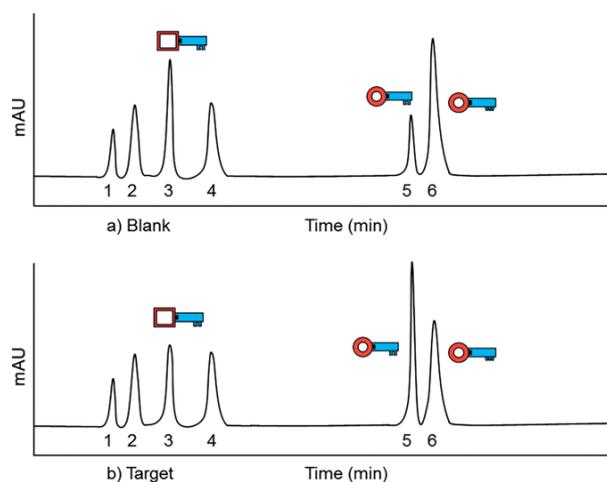


Figure 5. Schematic example of HPLC chromatograms: (a) blank library chromatogram, (b) target library chromatogram.

In order to accurately determine the amplification or decrease of peaks, their relative peak areas (RPA) should be compared. The fictional RPAs of both chromatograms in Figure 5 are given in Table 4. The amplification factor in percentage can

be calculated by Equation (1), where the amplification factor in “fold” is given by Equation (2). Using these two equations, the product at peak five has increased by 100 % or twofold. Frei et al. report on a particularly thorough analysis of a DCL using the lectin FimH as a target, using HPLC analysis with an optimized DCC protocol.<sup>[37]</sup>

$$\text{amplification factor (\%)} = \frac{RPA_{\text{target}} - RPA_{\text{blank}}}{RPA_{\text{blank}}} * 100\% \quad (1)$$

$$\text{amplification fold} = \frac{\text{New}}{\text{Old}} \quad (2)$$

### 3.3. DCL Analyzed with STD-NMR spectroscopy

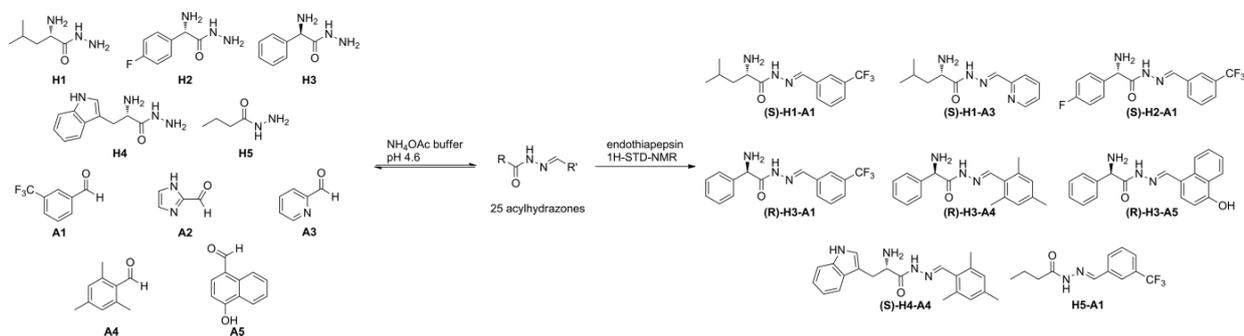
Inspired by the work of Ramström and co-workers,<sup>[20]</sup> we analyzed the formed DCLs by STD-NMR spectroscopy (Scheme 3). Five sub-libraries enabled a clear analysis. As a control with a known binder we used saquinavir ( $K_i = 48 \text{ nM}$ ), a potent peptidic inhibitor, to differentiate specific from nonspecific binding. Each sub-library contained all five hydrazides and one of the aldehyde building blocks and was allowed to equilibrate for 24 hours before adding the target. By analyzing the imine-type proton signals of the acylhydrazone products in the  $^1\text{H}$ -STD-NMR spectra (Figure 6), we identified in total eight binders. To confirm the results from STD-NMR, we performed an enzyme-inhibition assay and showed that the hits were inhibitors with  $IC_{50}$  values ranging from  $12.8 \mu\text{M}$  to  $365 \mu\text{M}$ . The high hit rate in this report may be a result of the powerful and synergistic combination of de novo structure-based drug design and DCC. In addition, it is due to use of five sublibraries in which the best binder of each library is detected, whereas in a regular ptDCC setup only the overall best binders will be discovered. In STD-NMR the protein is used as a tool to analyze the library, whereas in a ptDCC experiment the protein influences the equilibrium and hence the concentrations

#### 4.4. How to proceed after obtaining hits

Having obtained a validated hit, identified by de novo structure-based drug design in combination with DCC and STD-NMR, we have used a structure-based design approach to improve the molecular recognition by the target.<sup>[63]</sup> In this specific case, we were fortunate to have an X-ray crystal structure of the target endothiapepsin in complex with the hit. If this is not the case, optimization is still possible, relying on structure-activity relationships.

Table 4. Example of relative peak areas (RPA) obtained from HPLC chromatograms from Figure 4.

Peak number	Relative peak area in blank [%]	Relative peak area in target [%]	Amplification in %	Amplification in “fold”
1	10	10	–	1
2	15	15	–	1
3	20	16	–20 %	0.8
4	16	16	–	1
5	12	24	100 %	2
6	27	19	–30 %	0.7
Total	100 %	100 %		



Scheme 3. Formation of dynamic combinatorial library and enzymatic selection of the best binders by  $^1\text{H}$ -STD-NMR analysis. Adapted from Mondal et al.<sup>[24]</sup>

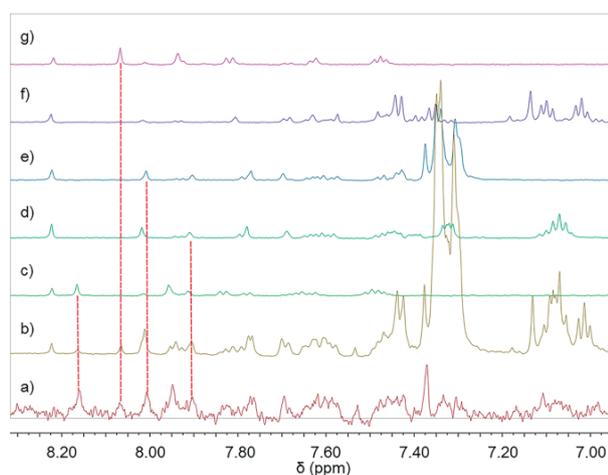


Figure 6. DCL generated from **H1-5 + A4**: (aromatic region) a)  $^1\text{H}$ -STD-NMR spectrum of **H1-5 + A4**, b)  $^1\text{H}$ -NMR spectrum of **H1-5 + A4**, c)  $^1\text{H}$ -NMR spectrum of **H3+A4**, d)  $^1\text{H}$ -NMR spectrum of **H4+A4** (2 singlets correspond to the *E/Z* isomers), e) **H1+A4**, f) **H2+H4** and g) **H5+A4**. Adapted from Mondal et al.<sup>[24]</sup>

## Conclusions

There are a number of steps, which should be carefully taken into account, in order to obtain active hits by DCC. If information on the target is available, e.g. a crystal-structure, one could consider a structure-based design when choosing the building blocks. The type of reversible linkage to be used can be chosen at this stage. Conditions necessary for the equilibration to take place should be compatible with the target. After establishing conditions, which will ensure the target remains folded, the actual DCC experiment can be started. To do so, stock solutions of building blocks, catalyst and protein should be prepared. The formed DCLs can be analyzed by different techniques such as STD-NMR or HPLC-MS. Compounds that have been selected by the target, and their biochemical properties should be evaluated and possibly optimized in further studies.

## Acknowledgments

Funding from Netherlands Organisation for Scientific Research (VIDI grant: 723.014.008; LIFT grant: 731.015.414) and from the

Helmholtz Association's Initiative and Networking Fund is gratefully acknowledged. We thank Dr. Ravindra Jumde for fruitful discussions regarding this manuscript.

**Keywords:** Target-directed dynamic combinatorial chemistry · Protein stability · Hit identification · Medicinal chemistry · Biochemical activity

- [1] K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski, R. J. Knapp, *Nature* **1991**, 354, 82–84.
- [2] R. A. Houghten, C. Pinilla, S. E. Blondelle, J. R. Appel, C. T. Dooley, J. H. Cuervo, *Nature* **1991**, 354, 84–86.
- [3] R. A. Houghten, *Proc. Natl. Acad. Sci. USA* **1985**, 82, 5131–5135.
- [4] H. M. Geysen, R. H. Meloen, S. J. Barteling, *Proc. Natl. Acad. Sci. USA* **1984**, 81, 3998 LP-4002.
- [5] R. Frank, W. Heikens, G. Heisterberg-Moutsis, H. Blöcker, *Nucl. Acid. Res.* **1983**, 11, 4365–4377.
- [6] I. Huc, J.-M. Lehn, *Proc. Natl. Acad. Sci. USA* **1997**, 94, 2106–2110.
- [7] M. H. Ohlmeyer, R. N. Swanson, L. W. Dillard, J. C. Reader, G. Asouline, R. Kobayashi, M. Wigler, W. C. Still, *Proc. Natl. Acad. Sci. USA* **1993**, 90, 10922 LP1-0926.
- [8] S. J. Rowan, P. S. Lukeman, D. J. Reynolds, J. K. M. Sanders, *New J. Chem.* **1998**, 22, 1015–1018.
- [9] V. A. Polyakov, M. I. Nelen, N. Nazarpack-Kandlousy, A. D. Ryabov, A. V. Eliseev, *J. Phys. Org. Chem.* **1999**, 12, 357–363.
- [10] A. Ganesan, *Angew. Chem. Int. Ed.* **1998**, 37, 2828–2831; *Angew. Chem.* **1998**, 110, 2989.
- [11] C. Karan, B. L. Miller, *Drug Discovery Today* **2000**, 5, 67–75.
- [12] V. T. Bhat, A. M. Caniard, T. Luksch, R. Brenk, D. J. Campopiano, M. F. Greaney, *Nat. Chem.* **2010**, 2, 490–497.
- [13] Z. Fang, W. He, X. Li, Z. Li, B. Chen, P. Ouyang, K. Guo, *Bioorg. Med. Chem. Lett.* **2013**, 23, 5174–5177.
- [14] M. Demetriades, I. K. H. Leung, R. Chowdhury, M. C. Chan, M. A. McDonough, K. K. Yeoh, Y.-M. Tian, T. D. W. Claridge, P. J. Ratcliffe, E. C. Y. Woon, et al., *Angew. Chem. Int. Ed.* **2012**, 51, 6672–6675; *Angew. Chem.* **2012**, 124, 6776.
- [15] E. C. Y. Woon, M. Demetriades, E. A. L. Bagg, W. Aik, S. M. Krylova, J. H. Y. Ma, M. Chan, L. J. Walport, D. W. Wegman, K. N. Dack, et al., *J. Med. Chem.* **2012**, 55, 2173–2184.
- [16] S. Sakai, Y. Shigemasa, T. Sasaki, *Tetrahedron Lett.* **1997**, 38, 8145–8148.
- [17] R. J. Lins, S. L. Flitsch, N. J. Turner, E. Irving, S. A. Brown, *Angew. Chem. Int. Ed.* **2002**, 41, 3405–3407; *Angew. Chem.* **2002**, 114, 3555.
- [18] R. J. Lins, S. L. Flitsch, N. J. Turner, E. Irving, S. A. Brown, *Tetrahedron* **2004**, 60, 771–780.
- [19] B. Shi, R. Stevenson, D. J. Campopiano, M. F. Greaney, *J. Am. Chem. Soc.* **2006**, 128, 8459–8467.
- [20] R. Caraballo, H. Dong, J. P. Ribeiro, J. Jiménez-Barbero, O. Ramström, *Angew. Chem. Int. Ed.* **2010**, 49, 589–593; *Angew. Chem.* **2010**, 122, 599.
- [21] P. Frei, R. Hevey, B. Ernst, *Chem. Eur. J.* **2019**, 25, 60–73.
- [22] M. Mondal, A. K. H. Hirsch, *Chem. Soc. Rev.* **2015**, 44, 2455–2488.

- [23] R. Van der Vlag, A. K. H. Hirsch, in *Compr. Supramol. Chem.* 2, Elsevier, **2017**, pp. 487–509.
- [24] M. Mondal, N. Radeva, H. Köster, A. Park, C. Potamitis, M. Zervou, G. Klebe, A. K. H. Hirsch, *Angew. Chem. Int. Ed.* **2014**, *53*, 3259–3263; *Angew. Chem.* **2014**, *126*, 3324.
- [25] A. Dirksen, S. Dirksen, T. M. Hackeng, P. E. Dawson, *J. Am. Chem. Soc.* **2006**, *128*, 15602–15603.
- [26] A. Dirksen, T. M. Hackeng, P. E. Dawson, *Angew. Chem. Int. Ed.* **2006**, *45*, 7581–7584; *Angew. Chem.* **2006**, *118*, 7743.
- [27] P. Crisalli, E. T. Kool, *J. Org. Chem.* **2013**, *78*, 1184–1189.
- [28] F. V. Reddavid, W. Lin, S. Lehnert, Y. Zhang, *Angew. Chem. Int. Ed.* **2015**, *54*, 7924–7928; *Angew. Chem.* **2015**, *127*, 8035.
- [29] L. Monjas, L. J. Y. M. Swier, I. Setyawati, D. J. Slotboom, A. K. H. Hirsch, *ChemMedChem* **2017**, *12*, 1693–1696.
- [30] F. T. Kern, K. T. Wanner, *ChemMedChem* **2015**, *10*, 396–410.
- [31] F. Kern, K. T. Wanner, *Bioorg. Med. Chem.* **2019**, *27*, 1232–1245.
- [32] see ref.<sup>[21]</sup>
- [33] G. Artigas, P. López-Senín, C. González, N. Escaja, V. Marchán, *Org. Biomol. Chem.* **2015**, *13*, 452–464.
- [34] J. D. McAnany, J. P. Reichert, B. L. Miller, *Bioorg. Med. Chem.* **2016**, *24*, 3940–3946.
- [35] Z. Yang, Z. Fang, W. He, Z. Wang, H. Gan, Q. Tian, K. Guo, *Bioorg. Med. Chem. Lett.* **2016**, *26*, 1671–1674.
- [36] M. Mondal, N. Radeva, H. Fanlo-Virgós, S. Otto, G. Klebe, A. K. H. Hirsch, *Angew. Chem. Int. Ed.* **2016**, *55*, 9422–9426; *Angew. Chem.* **2016**, *128*, 9569.
- [37] P. Frei, L. Pang, M. Silbermann, D. Eris, T. Mühlethaler, O. Schwardt, B. Ernst, *Chem. Eur. J.* **2017**, *23*, 11570–11577.
- [38] J. Fu, H. Fu, M. Dieu, I. Halloum, L. Kremer, Y. Xia, W. Pan, S. P. Vincent, *Chem. Commun.* **2017**, *53*, 10632–10635.
- [39] J. Soubhye, M. Gelbcke, P. Van Antwerpen, F. Dufrasne, M. Y. Boufadi, J. Nève, P. G. Furtmüller, C. Obinger, K. Zouaoui Boudjeltia, F. Meyer, *ACS Med. Chem. Lett.* **2017**, *8*, 206–210.
- [40] A. G. Ekström, J. T. Wang, J. Bella, D. J. Campopiano, *Org. Biomol. Chem.* **2018**, *16*, 8144–8149.
- [41] M. Das, T. Yang, J. Dong, F. Prasetya, Y. Xie, K. H. Q. Wong, A. Cheong, E. C. Y. Woon, *Chem. Asian J.* **2018**, *13*, 2854–2867.
- [42] P. García, V. L. Alonso, E. Serra, A. M. Escalante, R. L. E. Furlan, *ACS Med. Chem. Lett.* **2018**, *9*, 1002–1006.
- [43] S. Jana, D. Panda, P. Saha, G. D. Pantoş, J. Dash, *J. Med. Chem.* **2019**, *62*, 762–773.
- [44] A. J. Clipson, V. T. Bhat, I. McNae, A. M. Caniard, D. J. Campopiano, M. F. Greaney, *Chem. Eur. J.* **2012**, *18*, 10562–10570.
- [45] M. Sindelar, K. T. Wanner, *ChemMedChem* **2012**, *7*, 1678–1690.
- [46] R. Nguyen, I. Huc, *Chem. Commun.* **2003**, 942–943.
- [47] D. A. Erlanson, A. C. Braisted, D. R. Raphael, M. Randal, R. M. Stroud, E. M. Gordon, J. A. Wells, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 9367–9372.
- [48] R. F. Ludlow, S. Otto, *J. Am. Chem. Soc.* **2008**, *130*, 12218–12219.
- [49] B. Shi, M. F. Greaney, *Chem. Commun.* **2005**, 886–888.
- [50] see ref.<sup>[14]</sup>
- [51] I. K. H. Leung, T. Brown, C. J. Schofield, T. D. W. Claridge, *Medchemcomm* **2011**, *2*, 390–395.
- [52] J. Newman, *Acta Crystallogr., Sect. D* **2004**, *60*, 610–612.
- [53] L. Reinhard, H. Mayerhofer, A. Geerlof, J. Mueller-Dieckmann, M. S. Weiss, *Acta Crystallogr., Sect. F Struct. Biol. Cryst. Commun.* **2013**, *69*, 209–214.
- [54] F. H. Niesen, H. Berglund, M. Vedadi, *Nat. Protoc.* **2007**, *2*, 2212–2221.
- [55] G. S. Sittampalam, N. P. Coussens, K. Brimacomber, A. Grossman, M. Arkin, D. Auld, C. Austin, J. Baell, B. Bejcek, J. M. M. Caaveiro, et al. **2004**.
- [56] T. Arakawa, Y. Kita, S. N. Timasheff, *Biophys. Chem.* **2007**, *131*, 62–70.
- [57] D. H. Rammner, *Ann. N. Y. Acad. Sci.* **1967**, *141*, 291–299.
- [58] M. Jackson, H. H. Mantsch, *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1991**, *1078*, 231–235.
- [59] A. Viegas, J. Manso, F. L. Nobrega, E. J. Cabrita, *J. Chem. Educ.* **2011**, *88*, 990–994.
- [60] B. Danieli, A. Giardini, G. Lesma, D. Passarella, B. Peretto, A. Sacchetti, A. Silvani, G. Pratesi, F. Zunino, *J. Org. Chem.* **2006**, *71*, 2848–2853.
- [61] C. Qiu, Z. Fang, L. Zhao, W. He, Z. Yang, C. Liu, K. Guo, *React. Chem. Eng.* **2019**, *4*, 658–662.
- [62] M. Jaegle, E. L. Wong, C. Tauber, E. Nawrotzky, C. Arkona, J. Rademann, *Angew. Chem. Int. Ed.* **2017**, *56*, 7358–7378; *Angew. Chem.* **2017**, *129*, 7464.
- [63] A. M. Hartman, M. Mondal, N. Radeva, G. Klebe, A. K. Hirsch, *Int. J. Mol. Sci.* **2015**, *16*, <https://doi.org/10.3390/ijms160819184>.

Received: February 28, 2019

## 1.5 Aims and Scope

At the beginning of this PhD project, the initial structure determination of DXPS had never been repeated. With a resolution of 2.4 Å and 2.9 Å, the DXPS protein structures were on the edge of being still suitable for structure-based drug design (SBDD) projects.<sup>20,70</sup>

To obtain a more reliable starting point for our SBDD approaches, the main goal of this PhD project was therefore to reproduce and improve the resolution of the protein crystal structure of *D. radiodurans* DXPS. As a next step, we wanted to solve the DXPS structures from the target organisms of our drug design projects, such as *Mycobacterium tuberculosis* and *Plasmodium falciparum*. The results of the structural biology part are reported in Chapters 2.1 and 2.2.

The second aim of the thesis was the design of inhibitors for DXPS as a starting point for new antibiotics. To be able to work on this from the beginning of the project, we used alternative, structure-independent hit-identification methods. A ligand-based approach, searching for cofactor-competitive inhibitors, is reported in Chapter 2.3. As a second structure-independent technique dynamic combinatorial chemistry (DCC) was employed as an hit-identification method, its results are reported in Chapter 2.4.

In addition, a contribution to the establishment of DCC as a commonly used method in drug discovery is reported as proof-of-concept study in Chapter 2.5, in which we report non-toxic, stable bioisosteres, enabling quick follow-up experiments with acylhydrazone-based DCC hits.

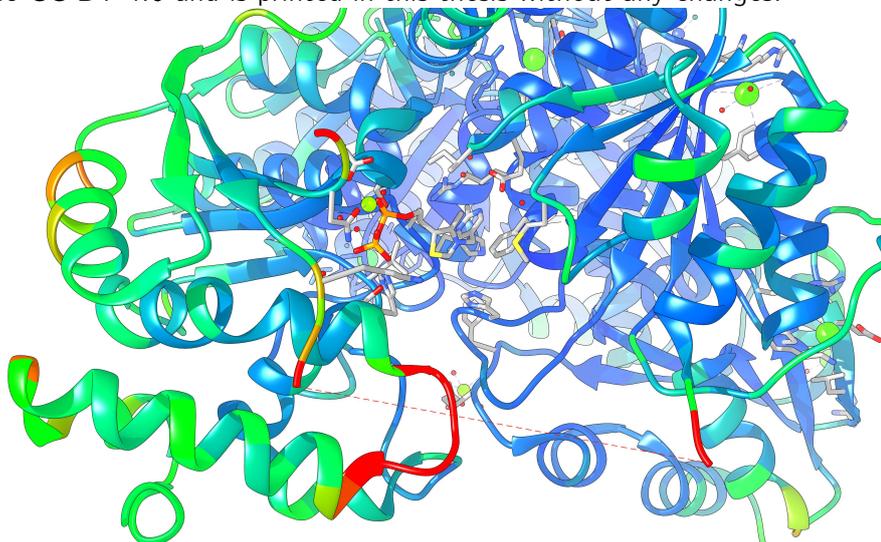
**Part 2**

**Results**



## 2.1 Identification of a *Deinococcus radiodurans* DXS Mutant with Improved Crystallographic Properties

The following chapter was published as an open-access article in the Journal Biochemical and Biophysical Research Communications under the terms of the creative commons license CC BY 4.0 and is printed in this thesis without any changes.<sup>2</sup>



### Title

Identification of a 1-deoxy-D-xylulose-5-phosphate synthase (DXS) mutant with improved crystallographic properties

### Authors

Robin M. Gierse ‡, Eswar R. Reddem ‡, Alaa Alhayek, Dominik Baitinger, Zhoor Hamid, Harald Jakobi, Bernd Laber, Gudrun Lange, Anna K.H. Hirsch\*, Matthew R. Groves\*

### Contributions

R. M. Gierse conceived the project, was involved in all experimental work, solved and refined the protein structures and drafted the manuscript. E. R. Reddem was involved in the protein design, purification and crystallization as well as data collection and interpretation. A. Alhayek, D. Baitinger and Z. Hamid were involved in protein production, purification and characterization. H. Jakobi, B. Laber and G. Lange were involved in the supervision of the project. A. K. H. Hirsch and M. R. Groves were involved in the preparation of the manuscript and conceived and supervised the project.

### Bibliographic Details

Biochemical and Biophysical Research Communications  
Volume 539, Pages 42-46  
5 February 2021  
DOI: 10.1016/j.bbrc.2020.12.069

‡: Authors contributed equally

\*: Corresponding Authors



Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: [www.elsevier.com/locate/ybbrc](http://www.elsevier.com/locate/ybbrc)

## Identification of a 1-deoxy-D-xylulose-5-phosphate synthase (DXS) mutant with improved crystallographic properties



Robin M. Gierse<sup>a, b, c, 1</sup>, Eswar R. Reddem<sup>c, d, 1</sup>, Alaa Alhayek<sup>a, b</sup>, Dominik Baitinger<sup>a</sup>, Zhoor Hamid<sup>a, b</sup>, Harald Jakobi<sup>e</sup>, Bernd Laber<sup>e</sup>, Gudrun Lange<sup>e</sup>, Anna K.H. Hirsch<sup>a, b, c, \*\*</sup>, Matthew R. Groves<sup>d, \*</sup>

<sup>a</sup> Department for Drug Design and Optimization, Helmholtz Institute for Pharmaceutical Research (HIPS) – Helmholtz Centre for Infection Research (HZI), Campus Building E 8.1, 66123, Saarbrücken, Germany

<sup>b</sup> Department of Pharmacy, Saarland University, Campus Building E8.1, 66123, Saarbrücken, Germany

<sup>c</sup> Stratingh Institute for Chemistry, University of Groningen, Nijenborgh 7, 9747, AG Groningen, Netherlands

<sup>d</sup> Pharmacy Department, Drug Design Group, University of Groningen, Antonius Deusinglaan 1, 9700, AV Groningen, Netherlands

<sup>e</sup> Research & Development Crop Science, Bayer AG, Industriepark Höchst, 65926, Frankfurt, Germany

### ARTICLE INFO

#### Article history:

Received 14 December 2020

Received in revised form

16 December 2020

Accepted 18 December 2020

#### Keywords:

MEP-Pathway

*Deinococcus radiodurans*

1-deoxy-D-xylulose-5-phosphate synthase (DXS)

Structure-based drug design

Antimicrobial resistance

### ABSTRACT

In this report, we describe a truncated *Deinococcus radiodurans* 1-deoxy-D-xylulose-5-phosphate synthase (DXS) protein that retains enzymatic activity, while slowing protein degradation and showing improved crystallization properties. With modern drug-design approaches relying heavily on the elucidation of atomic interactions of potential new drugs with their targets, the need for co-crystal structures with the compounds of interest is high. DXS itself is a promising drug target, as it catalyzes the first reaction in the 2-C-methyl-D-erythritol 4-phosphate (MEP)-pathway for the biosynthesis of the universal precursors of terpenes, which are essential secondary metabolites. In contrast to many bacteria and pathogens, which employ the MEP pathway, mammals use the distinct mevalonate-pathway for the biosynthesis of these precursors, which makes all enzymes of the MEP-pathway potential new targets for the development of anti-infectives. However, crystallization of DXS has proven to be challenging: while the first X-ray structures from *Escherichia coli* and *D. radiodurans* were solved in 2004, since then only two additions have been made in 2019 that were obtained under anoxic conditions. The presented site of truncation can potentially also be transferred to other homologues, opening up the possibility for the determination of crystal structures from pathogenic species, which until now could not be crystallized. This manuscript also provides a further example that truncation of a variable region of a protein can lead to improved structural data.

© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

### 1. Introduction

In 2015, the world health organization (WHO) published a

**Abbreviations:** AMR, Antimicrobial resistance; drDXS, *Deinococcus radiodurans* DXS protein; ΔdrDXS, Truncated *Deinococcus radiodurans* DXS protein; DXS, 1-deoxy-D-xylulose-5-phosphate synthase; MEP, 2-C-methyl-D-erythritol 4-phosphate; TSA, Thermal shift assay; WHO, World health organization.

\* Corresponding author.

\*\* Corresponding author. Department for Drug Design and Optimization, Helmholtz Institute for Pharmaceutical Research (HIPS) – Helmholtz Centre for Infection Research (HZI), Campus Building E 8.1, 66123, Saarbrücken, Germany.

E-mail addresses: [Anna.Hirsch@helmholtz-hips.de](mailto:Anna.Hirsch@helmholtz-hips.de) (A.K.H. Hirsch), [m.r.groves@rug.nl](mailto:m.r.groves@rug.nl) (M.R. Groves).

<sup>1</sup> Authors contributed equally.

global action plan on antimicrobial resistance (AMR) [1]. One of the five main objectives in the plan is an increase of research and development to fight AMR. In addition to improvements in how available antibiotics can be used, the development of innovative drugs is an essential strategy to address emerging resistance. The 2019 WHO report on antibacterial agents in clinical development defines the required innovation of a drug by the absence of cross-resistance, a new compound or target class or a new mode of inhibition. The authors estimate that in the next five years, eleven new antibiotics could be approved, but only one might be innovative and active against resistant Gram-negative bacteria - highlighting that the need for innovative antibiotics is as urgent as ever [2].

The targets of antibiotics currently on the market have been

<https://doi.org/10.1016/j.bbrc.2020.12.069>

0006-291X/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

mainly involved in mechanisms essential for the proliferation of pathogens, such as protein and cell-wall biosynthesis or DNA/RNA replication and repair [3]. Nowadays, in the search for innovative antibiotics, more and more unconventional targets are explored. One source for unconventional targets is the methylerythritol-phosphate (MEP)-pathway. Before the discovery of this pathway in 1993, its products, isopentenyl diphosphate and dimethylallyl diphosphate, were thought to be exclusively accessible via the mevalonate pathway [4–6]. With the discovery of the MEP pathway, an alternative biosynthetic route to these universal building blocks for all isoprenoids was found. Many bacteria and the chloroplasts of plants rely on this pathway, whereas humans and most Eukarya use the mevalonate pathway [5–7]. This distinction amongst species makes the MEP pathway very attractive for the development of new drugs [8,9]. As a result, the MEP pathway has been the target of several projects to develop new antibiotics [10–12].

The first enzyme of the MEP pathway, 1-deoxy-D-xylulose-5-phosphate synthase (DXS), catalyzes the rate-limiting formation of 1-deoxy-d-xylulose 5-phosphate (DOXP) [13]. Compared with all other downstream intermediates, DOXP offers the additional benefit of being also the starting material for the biosynthesis of thiamine diphosphate (vitamin B1) and pyridoxal 5-phosphate (vitamin B6) [14,15]. As DXS is also involved in multiple essential biosynthetic routes, we selected the enzyme DXS as a strategic branch point and therefore as a particularly interesting target for our own drug development [16,17].

Modern hit-identification strategies benefit greatly from structural knowledge of the enzymatic target, and drug optimization is accelerated by detailed knowledge of the atomic interactions between the compounds of interest with its target. However, while DXS has long been a target of interest, there is a relative paucity of high resolution structural information on this target: the first two crystal structures from organisms *Escherichia coli* and *Deinococcus radiodurans* were published in 2007, at highest resolution of 2.4 Å [18]. Notably, Xiang et al. reported that the DXS protein of *E. coli* was only crystallized successfully after a fungal contamination led to partial proteolysis of the enzyme [18,19]. Due to the potential of the MEP pathway for the development of new antibiotics, the demand for additional structural information of DXS homologues is high. This is illustrated by reports on the computation of homology models of pathogenic organisms, such as *Mycobacterium tuberculosis* and *Plasmodium falciparum* [20,21], and the use of orthogonal methods to gain structural insights, such as H/D exchange MS [22]. In 2019, two DXS crystal structures of *D. radiodurans* were solved to higher resolution (1.95 Å) by Drennan and coworkers [23]. These structures give further insight on the catalytic steps of DXS, but crystals must be grown under anoxic conditions, limiting the functional states of the enzyme that can be structurally characterized.

To support our own structure-based drug-design projects, we developed a truncated DXS construct that crystallizes readily under aerobic conditions and diffracts to a resolution of 2.10 Å. The truncated loop has a very low evolutionary conservation and we show that its removal had no influence on enzymatic activity. Due to the low conservation of the identified region, a similar approach should also be applicable to homologues of DXS, enabling the determination of crystal structures from pathogenic organisms in the future.

## 2. Materials and methods

### 2.1. Protein expression and purification

The truncated DXS gene was obtained commercially, cloned into

the pETM-11 expression vector and transformed into *Escherichia coli* BL21 (DE3). drDXS and ΔdrDXS were expressed and purified as described by Xiang et al. with minor modifications [18]. After IMAC and Ion-exchange chromatography, the ΔdrDXS-containing fractions were pooled and cleavage of the His-tag was performed by TEV-protease digestion at 4 °C overnight. Removal of the tag and protease was achieved by reversed IMAC chromatography. After concentration by ultrafiltration using a VivaSpin ultrafiltration device with a molecular weight cut-off of 30 kDa, the final gel filtration chromatography of ΔdrDXS was performed in 20 mM Tris-HCl (pH 7.5) 150 mM NaCl, 10 mM DTT.

### 2.2. Crystallization

The protein was concentrated in the presence of 50 μM ThDP to 23 mg/mL by ultrafiltration. The sample was centrifuged at 14,000 rpm before setting up drops. Initial screening was performed using commercial screens at RT in 96-well, SDMRC2 sitting-drop plates. For optimization, crystals were grown at RT using 2 μL hanging-drops, 1:1 mixture of protein and mother liquor. Crystals were obtained after 48 h with 0.2 M calcium acetate hydrate, 0.1 M Tris pH 8.5 and 20% PEG 4000 as the precipitant.

### 2.3. Data collection, processing and refinement

Diffraction data were collected at beamlines P11 and P13, DESY, Hamburg. Data reduction and scaling was performed using XDS [24]. Molecular replacement with Molrep was used for phasing, using 2o1x as a search model [25]. The model was further refined during several rounds of iterative manual model building and refinement in the CCP4 suite using coot and refmac [26–28]. Refinement statistics are shown in Table S1.

### 2.4. Kinetic measurements

The DXS activity was analyzed at RT as previously reported, with minor modifications [29,30]. Assay volume was 60 μL, to enable the use of 384-well plates (Greiner BioOne), buffer was 200 mM HEPES, pH 8.0. Data analysis was performed with the enzyme kinetics module of Origin pro 2019.

### 2.5. Thermal-shift assay (TSA)

Analysis was performed using an ABI StepOneplus RT-PCR instrument using white 96-well plates. A continuous heating rate of 1 °C/min from 20 °C to 95 °C was used. Sample volume was 25 μL, consisting of 20 μL TSA buffer (20 mM Tris-HCl, pH 8.0; 300 mM NaCl, 5 mM MgCl<sub>2</sub>), 2.5 μL protein solution and 2.5 μL dye (Sypro Orange, 5000 x in DMSO, Sigma-Aldrich). Optimal concentrations were experimentally determined, 1 mg/mL of protein, 50 x Sypro Orange in TSA buffer yielded the best signal-to-noise ratio.

### 2.6. LC-MS measurements

The protocol for LC-MS measurements is provided, together with its results, in the SI.

## 3. Results and discussion

### 3.1. Truncation strategy used to design a crystallizing protein construct of DXS

While reproducing the protein crystals of *D. radiodurans* DXS (drDXS), we have observed a partial proteolysis of the 67 kDa protein into fragments of 20 and 40 kDa size, as previously reported

for *E. coli* and *P. aeruginosa* DXS (Fig. S1) [18,31]. This prompted us to use the technique of limited proteolysis to optimize the DXS protein. This method is based on the observation that well-structured domains of a protein are protected against proteolytic digestion [32,33]. Analysis of such a partial digestion can lead to re-engineered proteins that contain the protected, well-folded domains and have often more suitable properties for protein crystallography.

In the case of DXS, while a digestion site is not identified in the published crystal structures, amino acids 199–244 show no electron density in the DXS structures 6ouw, 2o1x and 2o1s and only fragmented, partial density for 6ouu. As a result, we hypothesized that the 20 kDa fragment corresponds to amino acids 1–199 and the 40 kDa fragment to amino acids 240–629. The degraded protein sample was analyzed by LC-MS to identify the exact cleavage site. A ~20 kDa fragment could not be observed, but we could observe a mixture of three different species with masses of 42,890, 42,690 and 42,489 Da, respectively (Fig. S2, S3). The observed masses correspond well to the calculated masses of the drDXS protein fragments with the amino acids 232, 234 or 236 to 629, respectively (sequence following UniProt-ID: Q9RUB5). Taken together, we concluded that the loop of amino acids 199–236 of drDXS is particularly sensitive to proteases, but we cannot exclude the possibility of autocatalytic cleavage of this loop.

To answer the question if the flexible loop is a species-independent property of the DXS enzyme, we analyzed the sequence conservation of all 498 deposited and manually annotated bacterial DXS genes of the Uniprot database [34]. A simplified image of the calculated multiple sequence alignment (MSA) is shown in Fig. 1 (full MSA in SI). While the sequence homology of all 498 DXS proteins is 62.6% overall, the digested loop (200–232) displays a lower homology of 41.2%. We found that the loop also has a high variability in length, ranging from 5 to 58 amino acids in *Myxococcus xanthus* and *Kocuria rhizophila*, respectively. The mean loop length is with 45 amino acids similar to that of the 43 amino acids of *D. radiodurans*. Such variability is an indication that this loop is not essential for the catalytic reaction.

Based on the LC-MS results, the MSA and the lack of density between amino acids 200–240 in 2o1x, we designed a construct that replaces amino acids 201–243 with six glycine residues. This linker was designed to be long enough to bridge the gap of 11.7 Å between the two amino acid chains, but short enough to avoid introduction of multiple linker conformations. We expected that these modifications yield an optimized protein (ΔdrDXS), with properties more suitable for crystallization (sequence in SI).

### 3.2. Biophysical characterization

Purified ΔdrDXS protein was analyzed by LC-MS. The sample eluted as a single peak with a mass of 63,382.95 Da, which is in good agreement with the calculated mass for ΔdrDXS of 63,382.11 Da (Fig. S4). LC-MS and SDS-PAGE analysis showed the intact protein, even after a week incubation at RT, confirming the desired improvement in stability of ΔdrDXS (Fig. S1, S4).

To analyze if the truncation affects the activity of ΔdrDXS, we determined the enzyme kinetics for both substrates [31]. The results are summarized in Table 1 and shown in Fig. S6, S7. The truncated drDXS enzyme retains its catalytic activity. It shows, however, slightly lower affinities for both substrates and a reduced turnover number.

To further investigate the effects of the truncation, the melting points ( $T_m$ ) were determined using a thermal shift assay (TSA) [35], in which any increase of the  $T_m$  is a sign of improved protein stability. This is often used to screen for optimal buffer conditions or analyze the effect of mutations [36]. With a  $T_m$  of 55.2 °C, the truncated enzyme shows a nearly identical value to that of the native enzyme, which has a  $T_m$  of 55.0 °C (Fig. S8), indicating that our loop truncation had no significant effect on protein stability.

Crystallization screening of ΔdrDXS yielded several conditions, with the best crystals diffracting to a resolution of 2.1 Å. The protein structure obtained is deposited in the PDB with the code 6xxg and the collection and refinement statistics are reported in Table S1.

### 3.3. Effects of the truncation on protein folding

The truncated protein is catalytically active and no major changes in its properties could be identified using biophysical characterization methods. Since the ΔdrDXS protein yields well-diffracting protein crystals, we were also able to analyze the effect of the truncation by comparison of the obtained X-ray structure with the wild-type enzyme.

To compare the structures, the  $C_\alpha$ -RMSD of residues 8–183, 253–288 and 322–627 between truncated and the wild-type structures were calculated, and a superposition colored by individual  $C_\alpha$ -RMSD is shown in Fig. 2 [26,37]. The RMSD on C-alpha position is 0.459 Å, 0.476 Å and 0.328 Å with 2o1x, 6ouu and 6ouw, respectively. These values show that the majority of the structure is unaffected by the truncation. While comparing the structures, we could also identify two regions that are present in a novel conformation: an  $\alpha$ -helix (residues 186–200) and a  $\beta$ -hairpin motif (residues 303–320; Fig. 2).

The  $\beta$ -hairpin motif (residues 303–320) was described recently as part of a so-called “spoon”. This motif undergoes structural

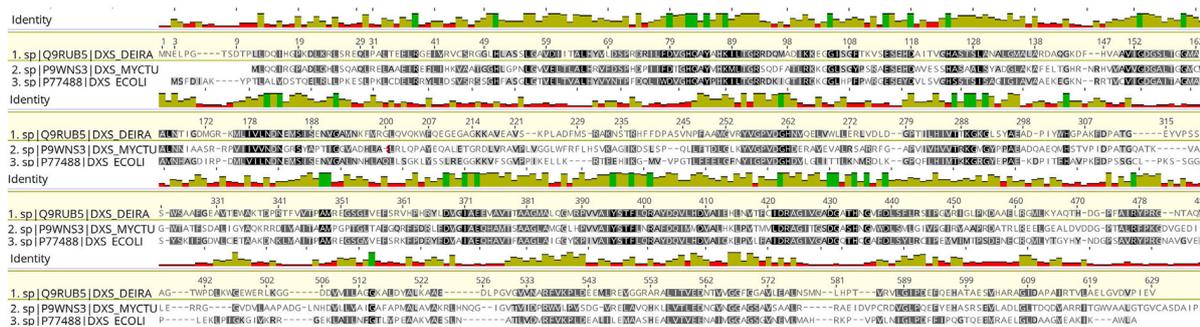
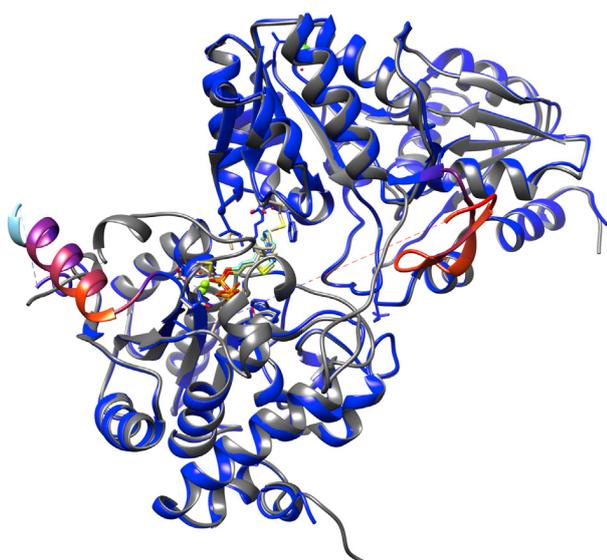


Fig. 1. MSA of DXS enzymes, including *Deinococcus radiodurans* (1, DEIRA) and *Mycobacterium tuberculosis* (2, MYCTU). For the sake of clarity, the image only shows three sequences. The identity, shown as bar graph above the sequences, was calculated using all 498 aligned sequences. Between amino acids 200 and 240, a highly variable region can be observed.

**Table 1**  
Kinetic comparison of the native and truncated drDXS enzyme.

	Pyruvate	D-GAP
drDXS	$K_m$ : $58 \pm 9 \mu\text{M}$ $v_{\text{max}}$ : $2.2 \mu\text{mol}/\text{min}$ $k_{\text{cat}}$ : $0.78 \text{ s}^{-1}$	$K_m$ : $193 \pm 23 \mu\text{M}$ $v_{\text{max}}$ : $1.8 \mu\text{mol}/\text{min}$ $k_{\text{cat}}$ : $0.64 \text{ s}^{-1}$
$\Delta\text{drDXS}$	$K_m$ : $85 \pm 9 \mu\text{M}$ $v_{\text{max}}$ : $2.6 \mu\text{mol}/\text{min}$ $k_{\text{cat}}$ : $0.46 \text{ s}^{-1}$	$K_m$ : $260 \pm 16 \mu\text{M}$ $v_{\text{max}}$ : $2.3 \mu\text{mol}/\text{min}$ $k_{\text{cat}}$ : $0.38 \text{ s}^{-1}$

**Fig. 2.** Superimposition of the drDXS structures with the code 2o1x (gray) and 6xxg (colored by  $C_{\alpha}$ -RMSD). Color coding: blue – low RMSD to red – high RMSD. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

rearrangements during the catalytic cycle of the DXS enzyme upon pyruvate binding. In our structure, the  $\beta$ -sheets of chain A adopt a conformation similar to the reported “bent spoon”-motif, while the equivalent residues of chain B are disordered. However, the “bent spoon” of our structure is distinct from than reported by Drennan and coworkers in 6ouw [23]. With this observation, the presented structure further contributes to the understanding of conformational changes in this region during catalysis.

The  $\alpha$ -helix formed by residues 186–200 is directly adjacent to the truncated amino acids 201–243. It can also be observed in the wild-type structure, but starting at residue 193. The C-terminal extension of the helix alters its orientation and moves residues 184–188 away from the ThDP binding site. The amino acids folding away from the active site do not take part in the catalyzed reaction, but form the hydrophobic surface of the ThDP binding site.

The same conformational change of this  $\alpha$ -helix can also be observed in the recently published structure 6ouw. In this structure, the amino acids from position 184 are also part of the  $\alpha$ -helix and point away from the active site. This similar folding to the truncated structure shows that this conformational change is not caused by the truncation. Jordan and coworkers have also identified the amino acids 183–199 as flexible and solvent-exposed using H/D exchange MS experiments. They show that this part of the protein adopts two distinct states, driven by substrate binding [22]. It seems that the truncation facilitates the formation of a stable  $\alpha$ -helix.

### 3.4. Crystal contacts and packing

We could observe that the residues 186–200 of the  $\alpha$ -helix form new lattice contacts. The  $\alpha$ -helix of chain A is at a distance of 8 Å and parallel to an  $\alpha$ -helix formed by the amino acids 28–46 of an adjacent DXS protein. This proximity enables salt bridges between Asn195 and Arg199 and the Glu35 of the neighboring protein in the crystal lattice with a distance of 2.7 Å and 4.5 Å, respectively (Fig. S9). In future constructs, these interactions could be strengthened by introduction of more charged amino acids.

Comparing the packing, we could identify two different forms of DXS protein crystals. The previously determined structures 2o1x and 6ouw both have a Matthew’s coefficient ( $V_M$ ) of  $2.75 \text{ \AA}^3/\text{Da}$  and a solvent content of 55%. The truncated structure has a  $V_M$  of  $2.27 \text{ \AA}^3/\text{Da}$  and a solvent content of 45%, similar to the DXS structure of 6ouw. As shown in Fig. 3, these two proteins adopt a different orientation in the crystal lattice and have a tighter packing, reducing the unit cell parameters. In the tightly packed structure 6ouw, the residues 307–319 are in the “bent-spoon” conformation, in the truncated structure a similar  $\beta$ -hairpin, but at a location 14 Å distant to the “bent spoon” motif, can be seen. Both structures have the previously described  $\alpha$ -helices of the “fork” motif (amino acids 292–306) in a disordered state with no observable electron density.

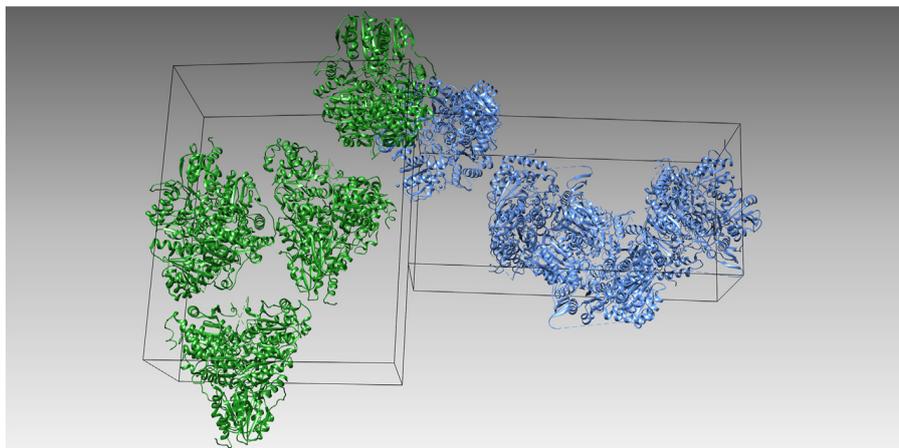
A higher density in protein crystals correlates with an increased resolution and is a desirable feature for future crystal structures in complex with ligands [38,39]. Using the truncated protein and the addition of pyruvate during crystallization seems to push the protein into the “bent spoon” conformation and might be used in the future to obtain better protein crystals. With a sample size of only four protein crystals, this hypothesis will need further evaluation as more DXS structures become available.

## 4. Conclusion

In our efforts to create a DXS enzyme for crystallographic studies with improved stability, we identified the part of the enzyme that is most susceptible to degradation. The identified loop is comprised of the residues 201–243, which show a high evolutionary variability in both length and sequence through all known bacterial homologues. We designed and expressed a mutant protein lacking the identified loop. The truncation showed only a small effect on the enzymatic activity and no effect on the biophysical properties of DXS, but a substantial improvement in the crystallographic properties of the protein. The  $\Delta\text{drDXS}$  protein has an improved tendency to form protein crystals under aerobic conditions, and diffract to a better resolution than previously published aerobic crystals of the wild-type protein.

Comparison of the obtained crystal structure with published structures showed no effect of the truncation on the remaining protein. Only two regions of the enzyme, previously known to be flexible, were identified in a different conformation and give additional evidence for the recently reported “spoon”/“fork” motif at the active site, proposed by Drennan and coworkers [23].

As we could find this loop in all species, we expect that the transfer of the identified site of truncation to other bacterial



**Fig. 3.** Left, green: Unit cell of the drDXS enzyme in the 2o1x and 6ouv crystal structures. Cell size is  $78 \times 125 \times 151 \text{ \AA}$  with a volume of  $1.47 \times 10^6 \text{ \AA}^3$ ; right, blue: Unit cell of  $\Delta\text{drDXS}$ . Unit cell parameters are with  $72 \times 87 \times 181 \text{ \AA}$  resulting in a smaller volume of  $1.13 \times 10^6 \text{ \AA}^3$ . (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

homologues will show comparable effects. The MSA that we supply in the SI can be used by other research groups interested in DXS to design similar truncated homologues proteins with improved properties. We expect that this will facilitate the determination of DXS crystal structures from relevant pathogens.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgments

The authors thank Chantal Bader and Patrick Haack for LC-MS measurements. We acknowledge DESY (Hamburg, Germany), member of the Helmholtz Association HGF, and EMBL Hamburg for the provision of experimental facilities. Parts of this research was done at PETRA III, Beamlines P11 and P13 and we thank Anja Burkhart for assistance.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bbrc.2020.12.069>.

#### Funding

This work was funded by The Netherlands Organisation for Scientific Research (NWO), LIFT Grant 731.015.414; and the Helmholtz Association's Initiative and Networking Fund.

#### References

- [1] Global Action Plan on Antimicrobial Resistance, *Microbe Mag* 10 (2015) 354–355, <https://doi.org/10.1128/microbe.10.354.1>.
- [2] W.H.O.WHO, World Health Organization, *Antibacterial Agents in Clinical Development: an Analysis of the Antibacterial Clinical Development Pipeline*, 2019, 2019.
- [3] C. Walsh, A.S. for Microbiology, *Antibiotics: Actions, Origins, Resistance*, ASM Press, 2003.
- [4] M. Rohmer, M. Seemann, S. Horbach, et al., Glycerinaldehyde 3-phosphate and pyruvate as precursors of isoprenic units in an alternative non-mevalonate pathway for terpenoid biosynthesis, *J. Am. Chem. Soc.* 118 (1996) 2564–2566, <https://doi.org/10.1021/ja9538344>.
- [5] H.K. Lichtenthaler, M. Rohmer, J.S. Lichtenthaler, et al., Two independent biochemical pathways for isopentenyl diphosphate and isoprenoid biosynthesis in higher plants, *Physiol. Plantarum* 101 (1997) 643–652.
- [6] M. Rohmer, M. Knani, P. Simonin, et al., Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate, *Biochem. J.* 295 (1993) 517–524, <https://doi.org/10.1042/bj2950517>.
- [7] W.N. Hunter, The non-mevalonate pathway of isoprenoid precursor biosynthesis, *J. Biol. Chem.* 282 (2007) 21573–21577, <https://doi.org/10.1074/jbc.R700005200>.
- [8] T. Masini, A.K.H. Hirsch, Development of inhibitors of the 2C-methyl-D-erythritol 4-phosphate (MEP) pathway enzymes as potential anti-infective agents, *J. Med. Chem.* 57 (2014) 9740–9763, <https://doi.org/10.1021/jm5010978>.
- [9] S. Heuston, M. Begley, C.G.M. Gahan, et al., Isoprenoid biosynthesis in bacterial pathogens, *Microbiol. (United Kingdom)* 158 (2012) 1389–1401, <https://doi.org/10.1099/mic.0.051599-0>.
- [10] C. Mueller, J. Schwender, J. Zeidler, et al., Properties and inhibition of the first two enzymes of the non-mevalonate pathway of isoprenoid biosynthesis, *Biochem. Soc. Trans.* 28 (2000) 792–793, <https://doi.org/10.1042/bst0280792>.
- [11] J.M. Smith, N. V. Warrington, R.J. Vierling, et al., Targeting DXP synthase in human pathogens: enzyme inhibition and antimicrobial activity of butylacetylphosphonate, *J. Antibiot. (Tokyo)* 67 (2014) 77–83, <https://doi.org/10.1038/ja.2013.105>.
- [12] D. Barte, S. Sanders, P.D. Phillips, et al., Enamide prodrugs of acetyl phosphonate deoxy- d -xylulose-5-phosphate synthase inhibitors as potent antibacterial agents, *ACS Infect. Dis.* 5 (2019) 406–417, <https://doi.org/10.1021/acinfecdis.8b00307>.
- [13] J.M. Estévez, A. Cantero, A. Reindl, et al., 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants, *J. Biol. Chem.* 276 (2001) 22901–22909, <https://doi.org/10.1074/jbc.M100854200>.
- [14] Q. Du, H. Wang, J. Xie, Thiamin (vitamin B1) biosynthesis and regulation: a rich source of antimicrobial drug targets? *Int. J. Biol. Sci.* 7 (2011) 41–52, <https://doi.org/10.7150/ijbs.7.41>.
- [15] I.B. Müller, J.E. Hyde, C. Wrenger, Vitamin B metabolism in *Plasmodium falciparum* as a source of drug targets, *Trends Parasitol.* 26 (2010) 35–43, <https://doi.org/10.1016/j.pt.2009.10.006>.
- [16] DXS as a target for structure-based drug design, *Future Med. Chem.* 9 (2017) 1277–1294, <https://doi.org/10.4155/fmc-2016-0239>.
- [17] T. Masini, B.S. Kroezen, A.K.H. Hirsch, Druggability of the enzymes of the non-mevalonate-pathway, *Drug Discov. Today* 18 (2013) 1256–1262, <https://doi.org/10.1016/j.drudis.2013.07.003>.
- [18] S. Xiang, G. Usunow, G. Lange, et al., Crystal structure of 1-Deoxy-D-xylulose 5-phosphate synthase, a crucial enzyme for isoprenoids biosynthesis, *J. Biol. Chem.* 282 (2007) 2676–2682, <https://doi.org/10.1074/jbc.M610235200>.
- [19] C.R. Mandel, D. Gebauer, H. Zhang, et al., A serendipitous discovery that in situ proteolysis is essential for the crystallization of yeast CPSF-100 crystallization communications, *Acta Crystallogr. F* 62 (2006) 1041–1045, <https://doi.org/10.1107/S1744309106038152>.
- [20] A.M. Goswami, Computational analysis, structural modeling and ligand binding site prediction of *Plasmodium falciparum* 1-deoxy-D-xylulose-5-phosphate synthase, *Comput. Biol. Chem.* 66 (2017) 1–10, <https://doi.org/10.1016/j.compbiolchem.2016.10.010>.
- [21] T. Masini, B. Lacy, L. Monjas, et al., Validation of a homology model of *Mycobacterium tuberculosis* DXS: rationalization of observed activities of thiamine derivatives as potent inhibitors of two orthologues of DXS, *Org. Biomol. Chem.* 13 (2015) 11263–11277, <https://doi.org/10.1039/c5ob01666e>.

## 2.1. IDENTIFICATION OF A *DEINOCOCCUS RADIODURANS* DXS MUTANT WITH IMPROVED CRYSTALLOGRAPHIC PROPERTIES

R.M. Gierse, E.R. Reddem, A. Alhayek et al.

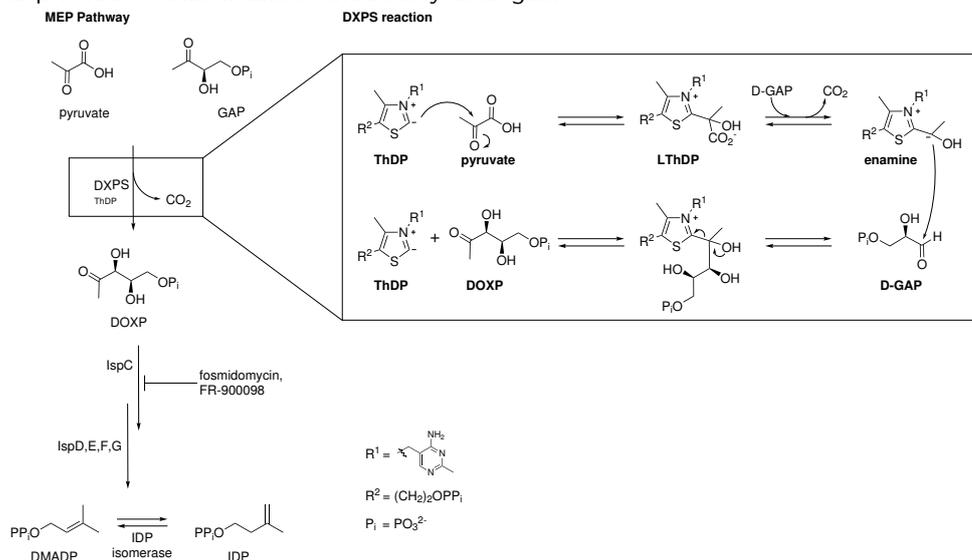
Biochemical and Biophysical Research Communications 539 (2021) 42–47

- [22] J. Zhou, L. Yang, A. DeColli, et al., Conformational dynamics of 1-deoxy-d-xylulose 5-phosphate synthase on ligand binding revealed by H/D exchange MS, *Proc. Natl. Acad. Sci. Unit. States Am.* 114 (2017) 9355–9360, <https://doi.org/10.1073/pnas.1619981114>.
- [23] P.Y.T. Chen, A.A. DeColli, C.L. Freel Meyers, et al., X-ray crystallography–based structural elucidation of enzyme-bound intermediates along the 1-deoxy-D-xylulose 5-phosphate synthase reaction coordinate, *J. Biol. Chem.* 294 (2019) 12405–12414, <https://doi.org/10.1074/jbc.RA119.009321>.
- [24] W. Kabsch, XDS, *Acta Crystallogr. D* 66 (2010) 125–132, <https://doi.org/10.1107/S0907444909047337>.
- [25] A. Vagin, A. Teplyakov, MOLREP : an automated program for molecular replacement, *J. Appl. Crystallogr.* 30 (1997) 1022–1025, <https://doi.org/10.1107/S0021889897006766>.
- [26] P. Emsley, B. Lohkamp, W.G. Scott, et al., Features and development of coot, *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66 (2010) 486–501, <https://doi.org/10.1107/S0907444910007493>.
- [27] M.D. Winn, C. Charles, K.D. Cowtan, et al., Overview of the CCP 4 Suite and Current Developments, vol. 4449, 2011, pp. 235–242, <https://doi.org/10.1107/S0907444910045749>.
- [28] G.N. Murshudov, A.A. Vagin, E.J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr. D* 53 (1997) 240–255, <https://doi.org/10.1107/S0907444996012255>.
- [29] T. Masini, J. Pilger, B.S. Kroezen, et al., De novo fragment-based design of inhibitors of DXS guided by spin-diffusion-based NMR spectroscopy, *Chem. Sci.* 5 (2014) 3543–3551, <https://doi.org/10.1039/c4sc00588k>.
- [30] S. Hecht, J. Wungsintaweekeul, F. Rohdich, et al., Biosynthesis of Terpenoids : efficient multistep biotransformation procedures affording isotope-labeled 2 C -methyl- D -erythritol 4-phosphate using recombinant 2 C -methyl- D -erythritol 4-phosphate, *Synthase* 61 (2001) 7770–7775, <https://doi.org/10.1021/jo015890v>.
- [31] B. Altincicek, M. Hintz, S. Sanderbrand, et al., Tools for discovery of inhibitors of the 1-deoxy-D-xylulose 5-phosphate (DXP) synthase and DXP reductoisomerase: an approach with enzymes from the pathogenic bacterium *Pseudomonas aeruginosa*, *FEMS Microbiol. Lett.* 190 (2000) 329–333, [https://doi.org/10.1016/S0378-1097\(00\)00357-8](https://doi.org/10.1016/S0378-1097(00)00357-8).
- [32] S.L. Cohen, B.T. Chait, A.R. Ferré-D'Amaré, et al., Probing the solution structure of the DNA-binding protein Max by a combination of proteolysis and mass spectrometry, *Protein Sci.* 4 (1995) 1088–1099, <https://doi.org/10.1002/pro.5560040607>.
- [33] A. Dong, Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA, *Nucleic Acids Res.* 29 (2001) 439–448, <https://doi.org/10.1093/nar/29.2.439>.
- [34] A. Bateman, UniProt: a worldwide hub of protein knowledge, *Nucleic Acids Res.* 47 (2019) D506–D515, <https://doi.org/10.1093/nar/gky1049>.
- [35] U.B. Ericsson, B.M. Hallberg, G.T. DeTitta, et al., Thermofluor-based high-throughput stability optimization of proteins for structural studies, *Anal. Biochem.* 357 (2006) 289–298, <https://doi.org/10.1016/j.ab.2006.07.027>.
- [36] S. Boivin, S. Kozak, R. Meijers, Optimization of protein purification and characterization using Thermofluor screens, *Protein Expr. Purif.* 91 (2013) 192–206, <https://doi.org/10.1016/j.pep.2013.08.002>.
- [37] E.F. Pettersen, T.D. Goddard, C.C. Huang, et al., UCSF chimera — a visualization system for exploratory research and analysis, *J. Comput. Chem.* 25 (2004) 1665–1612, <https://doi.org/10.1002/jcc.20084>.
- [38] K.A. Kantardjiev, Matthews Coefficient Probabilities: Improved Estimates for Unit Cell Contents of Proteins, DNA, and Protein – Nucleic Acid Complex Crystals, 2003, pp. 1865–1871, <https://doi.org/10.1110/ps.0350503.tially>.
- [39] B.W. Matthews, Solvent content of protein crystals, *J. Mol. Biol.* 33 (1968) 491–497, [https://doi.org/10.1016/0022-2836\(68\)90205-2](https://doi.org/10.1016/0022-2836(68)90205-2).



## 2.2 First Crystal Structures of DXPS from *Mycobacterium tuberculosis* Indicate a Distinct Mechanism of Intermediate Stabilization.

The following chapter was published as an open-access article in the Journal Scientific Reports under the terms of the Creative Commons Attribution 4.0 International License and is printed in this thesis without any changes.<sup>2</sup>



### Title

First crystal structures of 1-deoxy-D-xylulose 5-phosphate synthase (DXPS) from *Mycobacterium tuberculosis* indicate a distinct mechanism of intermediate stabilization.

### Authors

Robin M. Gierse ‡, Rick Oerlemans ‡, Eswar R. Reddem, Victor Gawriljuk, Alaa Alhayek, Dominik Baitinger, Harald Jakobi, Bernd Laber, Gudrun Lange, Anna K.H. Hirsch\*, Matthew R. Groves\*

### Contributions

R. M. Gierse conceived the project, was involved in most experimental work, solved and refined the protein structures and drafted the manuscript. R. Oerlemans optimized the protein expression and crystallization and was involved in the writing of the manuscript. E. R. Reddem was involved in the protein design, purification and crystallization as well as data collection and interpretation. V. Gawriljuk was involved in the bioinformatic analysis and took part in the preparation of the manuscript. A. Alhayek and D. Baitinger were involved in the protein expression, purification and characterization. H. Jakobi, B. Laber and G. Lange were involved in the supervision of the project. A. K. H. Hirsch and M. R. Groves were involved in the preparation of the manuscript and conceived and supervised the project.

### Bibliographic Details

Scientific reports

Volume 12, Article Number 7221

4 May 2022

DOI: 10.1038/s41598-022-11205-9

‡: Authors contributed equally

\*: Corresponding Authors



OPEN

# First crystal structures of 1-deoxy-D-xylulose 5-phosphate synthase (DXPS) from *Mycobacterium tuberculosis* indicate a distinct mechanism of intermediate stabilization

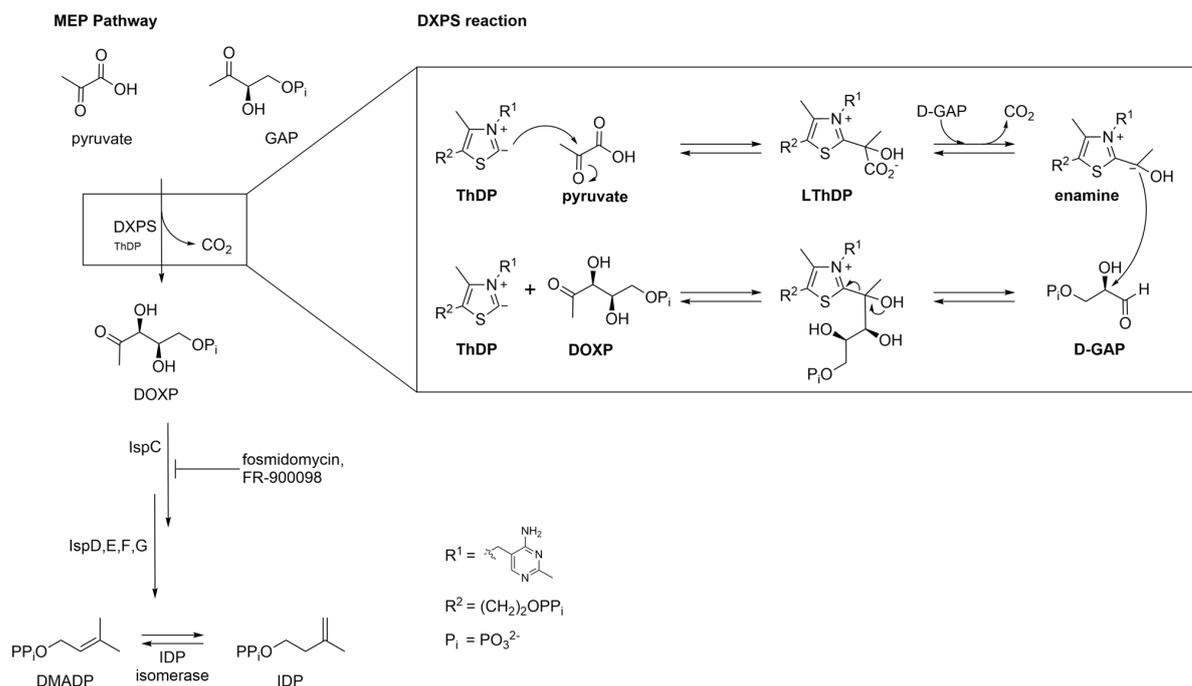
Robin M. Gierse<sup>1,2,3,7</sup>, Rick Oerlemans<sup>4,7</sup>, Eswar R. Reddem<sup>3,4</sup>, Victor O. Gawriljuk<sup>4,5</sup>, Alaa Alhayek<sup>1,2</sup>, Dominik Baitinger<sup>1</sup>, Harald Jakobi<sup>6</sup>, Bernd Laber<sup>6</sup>, Gudrun Lange<sup>6</sup>, Anna K. H. Hirsch<sup>1,2,3</sup>✉ & Matthew R. Groves<sup>4</sup>✉

The development of drug resistance by *Mycobacterium tuberculosis* and other pathogenic bacteria emphasizes the need for new antibiotics. Unlike animals, most bacteria synthesize isoprenoid precursors through the MEP pathway. 1-Deoxy-D-xylulose 5-phosphate synthase (DXPS) catalyzes the first reaction of the MEP pathway and is an attractive target for the development of new antibiotics. We report here the successful use of a loop truncation to crystallize and solve the first DXPS structures of a pathogen, namely *M. tuberculosis* (MtDXPS). The main difference found to other DXPS structures is in the active site where a highly coordinated water was found, showing a new mechanism for the enamine-intermediate stabilization. Unlike other DXPS structures, a “fork-like” motif could be identified in the enamine structure, using a different residue for the interaction with the cofactor, potentially leading to a decrease in the stability of the intermediate. In addition, electron density suggesting a phosphate group could be found close to the active site, provides new evidence for the D-GAP binding site. These results provide the opportunity to improve or develop new inhibitors specific for MtDXPS through structure-based drug design.

## Abbreviations

DMADP	Dimethylallyl diphosphate
DOXP	1-Deoxy-D-xylulose 5-phosphate
DXPS	1-Deoxy-D-xylulose 5-phosphate synthase
IDP	Isopentenyl diphosphate
IMAC	Immobilized metal ion affinity chromatography
LThDP	C2 $\alpha$ -lactylThDP
MEP-pathway	2-C-Methyl-D-erythritol 4-phosphate pathway
MSA	Multiple sequence alignment
SAR	Structure-activity relationship
TB	Tuberculosis

<sup>1</sup>Helmholtz Institute for Pharmaceutical Research Saarland (HIPS)-Helmholtz Centre for Infection Research (HZI), Campus Building E 8.1, 66123 Saarbrücken, Germany. <sup>2</sup>Department of Pharmacy, Saarland University, Campus Building E8.1, 66123 Saarbrücken, Germany. <sup>3</sup>Stratingh Institute for Chemistry, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands. <sup>4</sup>Department of Drug Design, Groningen Research Institute of Pharmacy, University of Groningen, Antonius Deusinglaan 1, 9700 AV Groningen, The Netherlands. <sup>5</sup>São Carlos Institute of Physics, University of São Paulo, Av. João Dagnone, 1100-Santa Angelina, São Carlos, SP 13563-120, Brazil. <sup>6</sup>Research & Development Crop Science, Bayer AG, Industriepark Höchst, 65926 Frankfurt, Germany. <sup>7</sup>These authors contributed equally: Robin M. Gierse and Rick Oerlemans. ✉email: Anna.Hirsch@helmholtz-hips.de; m.r.groves@rug.nl



**Figure 1.** An overview of the MEP pathway and the mechanism of the reaction catalyzed by DXS. Fosmidomycin and its derivative FR-900098 are inhibitors of the DXPS downstream enzyme IspC, validating the pathway as target for the development of new antibiotics.

TDR-TB      Totally drug-resistant TB  
 ThDP        Thiamine diphosphate  
 WHO        World Health Organization

The drug resistance of *Mycobacterium tuberculosis*, the causative agent for tuberculosis (TB), is a growing problem. Diagnosis and treatment of TB are both challenging tasks, the latter lasting for 6 months or more<sup>1</sup>. Although the WHO declared TB a global public health emergency in 1993, difficult and often incomplete treatment led to the emergence of resistant strains<sup>2,3</sup>. Over the years, *M. tuberculosis* has developed mechanisms of resistance against the known first-line “multi-drug resistant TB” and second line “extensively drug-resistant TB” antitubercular agents<sup>4</sup>. This development ultimately led to totally drug resistant TB (TDR-TB), resistant to all known first- and second-line antitubercular agents, observed first in 2008 in Iran and in 2012 in India<sup>5,6</sup>. To treat patients infected with TDR-TB, the development of drugs with new modes of action is urgently needed.

The 2-C-methyl-D-erythritol 4-phosphate (MEP)-pathway offers seven new target enzymes for the development of anti-TB drugs, which should, with a new mode of action, break the resistance of TDR-TB<sup>7,8</sup>. For many bacteria, this pathway is the only source of the terpene building blocks dimethylallyl diphosphate (DMADP) and isopentenyl diphosphate (IDP), essential for the biosynthesis of secondary metabolites (Fig. 1). The IspC inhibitor fosmidomycin is already used in combination therapy for the treatment of malaria, validating this pathway for the development of new drugs<sup>9</sup>. Several projects to develop further antibiotic drugs are currently running, targeting different enzymes of the MEP pathway<sup>10–15</sup>. The homologues of *M. tuberculosis* are also addressed. A series of lipophilic phosphonates that target the IspC enzyme from *E. coli* and *M. tuberculosis* was designed and synthesized. The corresponding co-crystal structure confirms a binding mode similar to that of the natural inhibitor fosmidomycin<sup>16</sup>. Aryl bis-sulfonamides were investigated to inhibit the downstream enzymes of the pathway: the IspF enzyme of *Plasmodium falciparum* and, with slightly lower activity, also the homologue from *M. tuberculosis*<sup>17</sup>. However, to the best of our knowledge, no qualified lead compound targeting MEP-pathway enzymes from *M. tuberculosis* was selected for further development.

In our drug design project, we chose to focus on 1-deoxy-D-xylulose 5-phosphate synthase (DXPS), the first and rate-limiting enzyme of the MEP-pathway<sup>18</sup>. DXPS catalyzes the ThDP-dependent decarboxylation of pyruvate and subsequent carbonylation with D-glyceraldehyde-3-phosphate (D-GAP), the second substrate, following a preferred-order, random-sequential reaction mechanism (Fig. 1)<sup>19,20</sup>. Targeting this enzyme has the additional benefit of not only inhibiting the biosynthesis of the terpene precursors, but also the biosynthesis of the vitamins B1 and B6, which are synthesized from the product of DXPS, the branch-point metabolite 1-deoxy-D-xylulose 5-phosphate (DOXP)<sup>21</sup>. A previous ligand-based approach of our group to design ThDP-competitive inhibitors of DXPS, using the *Deinococcus radiodurans* DXPS (drDXPS) crystal structure and an *M. tuberculosis* DXPS (MtDXPS) homology model, resulted in compounds showing activity against *D. radiodurans*, but with significantly less activity against MtDXPS<sup>22</sup>.

Enzyme	Pyruvate	D-GAP
<i>MtDXPS</i> 5 $\mu\text{mol/L}$	$K_m$ : $85 \pm 8 \mu\text{M}$ $k_{\text{cat}}$ : $4.5 \pm 0.1 \times 10^{-3} \text{ s}^{-1}$	$K_m$ : $75 \pm 7 \mu\text{M}$ $k_{\text{cat}}$ : $3.9 \pm 0.7 \times 10^{-3} \text{ s}^{-1}$
$\Delta$ <i>MtDXPS</i> 2 $\mu\text{mol/L}$	$K_m$ : $125 \pm 13 \mu\text{M}$ $k_{\text{cat}}$ : $17.3 \pm 0.4 \times 10^{-3} \text{ s}^{-1}$	$K_m$ : $112 \pm 11 \mu\text{M}$ $k_{\text{cat}}$ : $10.0 \pm 0.3 \times 10^{-3} \text{ s}^{-1}$

**Table 1.** Comparison of the enzyme kinetics from *MtDXPS* and  $\Delta$ *MtDXPS*.

The modern drug-design process benefits greatly from structural information of the enzymatic target. The knowledge of the molecular interactions between an inhibitor and its target protein sets the stage for structure-based optimization and therefore holds the potential to speed up the optimization of a hit or lead compound<sup>23,24</sup>. Most enzymes of the MEP pathway are structurally well-characterized, with structures of the enzymes IspC to IspH from multiple species available, both apo and in complex with inhibitors (Table S1), highlighting the interest in structure-based drug design studies for enzymes of this pathway. In contrast, the DXPS enzyme is less explored—with just five protein structures published to date, four from *D. radiodurans* and one from *E. coli*<sup>25–27</sup>. The low number of crystal structures may be a consequence of the enzyme's susceptibility to proteolytic degradation<sup>25,27,28</sup> or conformational changes during the catalytic cycle<sup>19,29</sup>—both properties that can make it difficult to obtain a homogeneous protein sample for crystallization. We recently published a modified drDXPS protein with improved crystallographic properties and speculated that our approach should be applicable to all DXPS homologues<sup>27</sup>.

In this report, we describe the application of our truncation strategy on the DXPS homologue of *M. tuberculosis*, as a representative pathogen in the focus of several drug-design projects<sup>30</sup>. Following the truncation approach, we were able to obtain protein crystals and report herein the first pathogenic DXPS structures: a holo structure with bound ThDP and a structure containing the enamine reaction intermediate (Fig. 1). We also docked *MtDXPS* inhibitors found in the literature and provide a structural rationale for their inhibitory activity. In addition to providing a molecular model of this important anti-infective target, this also shows that our truncation approach is transferable between organisms and has the potential to facilitate the determination of structures from other DXPS homologues in the future.

## Results

**Application of the truncation strategy.** In our recent article, we showed that the truncation of a non-conserved loop not visible in the structures of the DXPS enzyme of *D. radiodurans* (drDXPS) improved the crystallization of this target<sup>27</sup>. We hypothesized that this result could also be transferable to other homologues. The DXPS enzyme from *M. tuberculosis* (*MtDXPS*) was chosen to test this hypothesis. This homologue is of particular interest, as it is from a pathogenic organism and inhibitors targeting the enzyme have been investigated, utilizing a computed homology model due to the lack of an *MtDXPS* structure<sup>22,40</sup>.

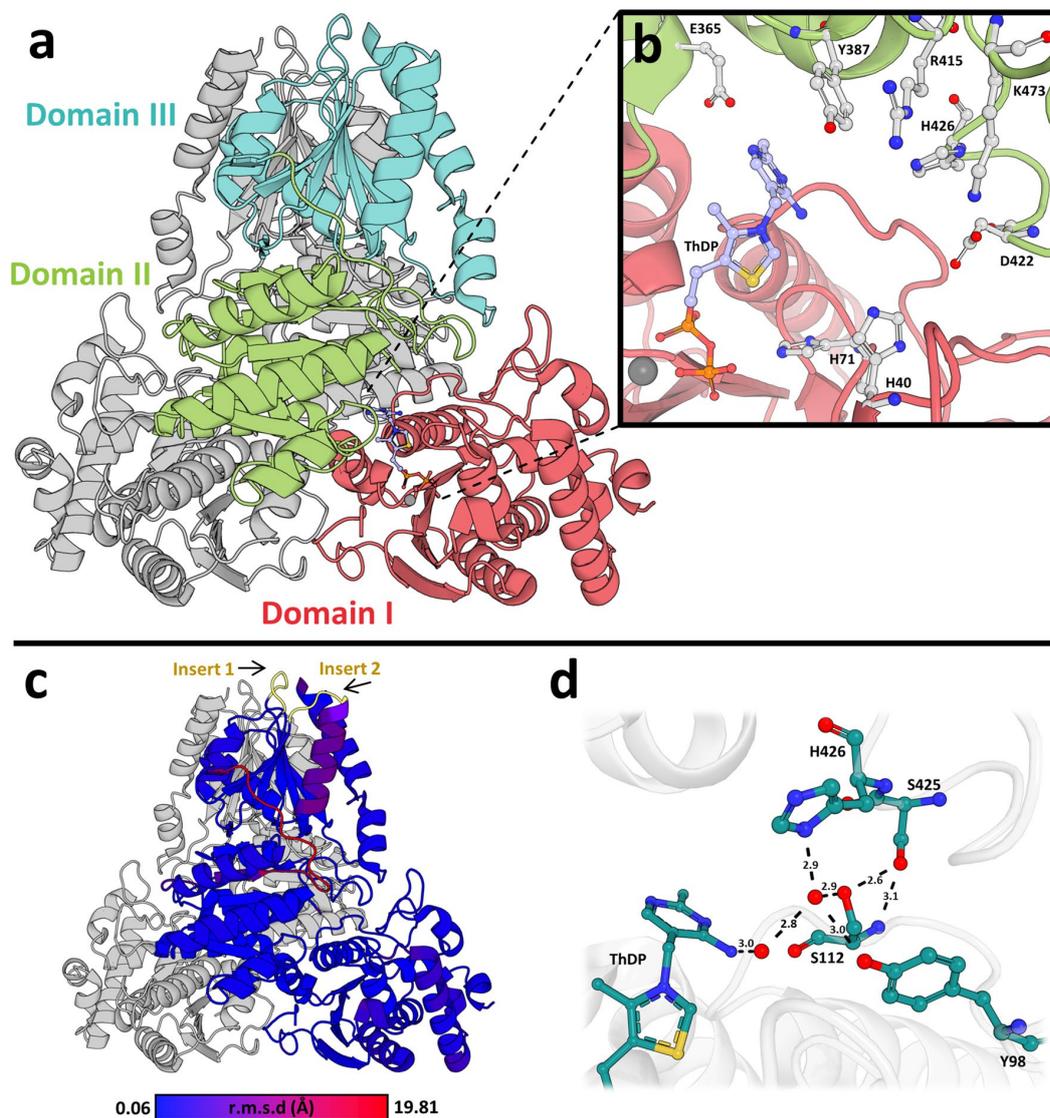
To apply the truncation strategy, we searched the previously published multiple sequence alignment (MSA) for the protein sequence of *M. tuberculosis* and compared it with the sequence of *D. radiodurans*<sup>27</sup>. The corresponding non-conserved loop from *MtDXPS* has a length of 45 amino acids, comprising amino acids 190–234 (Fig. S1). We replaced this loop with a linker of seven glycines that we visually estimated to be able to compensate for the removed amino acids, according to a homology model obtained for the *MtDXPS* homologue<sup>22</sup>. The resulting truncated sequence (**SI: Sequences**) of *MtDXPS* ( $\Delta$ *MtDXPS*) was then obtained as a synthetic gene.

**Characterization of the truncated enzyme.** The  $\Delta$ *MtDXPS* was recombinantly expressed as a soluble protein in good yields of > 50 mg/L, whereas we could express the native protein in a range of 0.5 mg/L in LB-medium. To analyze other effects of the truncation, we compared the wild-type *MtDXPS* and the truncated  $\Delta$ *MtDXPS*, using several biophysical techniques.

$\Delta$ *MtDXPS* integrity was analyzed by LC–MS. The sample showed a single protein with a high purity and a mass of 65,901.8 Da (Fig. S2). This mass is 705 Da lighter than calculated for the full length  $\Delta$ *MtDXPS* protein, and the full mass could not be observed. The weight of 705 Da corresponds to the molecular weight of the last six amino acids of the protein, suggesting either incomplete translation, a common issue in protein expression, or degradation of the sample during preparation. The terminal amino acids are often not resolved in protein crystal structures due to their flexibility, so the lack of the last six residues is unlikely to be a concern.

One of the truncation goals was to reduce the degradation of the wild-type *MtDXPS* enzyme. Over an incubation period of up to 5 days at RT, degradation of  $\Delta$ *MtDXPS* could not be detected on SDS-PAGE gels. In contrast, for full length *MtDXPS* bands corresponding to calculated degradation products of 20 and 40 kDa started to appear after 4 days (Fig. S3). This observation becomes more apparent when the time period is extended to 7 days, where a substantial decrease in band intensity of full length *MtDXPS* protein can be seen, while  $\Delta$ *MtDXPS* is still stable. Therefore, similarly to drDXPS, the loop truncation reduced the tendency for protein degradation, overall increasing the stability. As an additional method to assess protein stability, the melting point ( $T_m$ ) was determined using a thermal shift assay (TSA). The  $T_m$  obtained for both wild-type and truncated *MtDXPS* was  $45.4 \pm 0.4$  °C and  $52.7 \pm 0$  °C, respectively. The large difference in melting temperature between the two proteins ( $\Delta T_m = 7.3$  °C) strongly suggests that the truncation improved protein stability, which is often beneficial for protein crystallization<sup>31</sup>.

Similarly to drDXPS, the truncation had only small effects on the enzymatic activity of  $\Delta$ *MtDXPS*, as shown in Table 1. The  $\Delta$ *MtDXPS* enzyme showed a slightly reduced affinity for the substrates and a 2–3 times higher



**Figure 2.** Overview of the holo structure of  $\Delta MtDXPS$ . (a) Overview of the  $\Delta MtDXPS$  structure with highlighted domains. Domain I (res 1–312) in red, domain II (res 313–483) in green and domain III (484–638) in cyan. (b) Active site with important residues for catalysis shown in stick representation. (c) Structure colored according to  $C_{\alpha}$ -RMSD to the drDXPS structure (PDB ID: **2O1X**), insertion regions are colored in yellow. (d) Hydrogen-bonding network that stabilizes the different H426 conformation and key residues in the network. Water molecules are shown as red spheres and hydrogen bonds as black, dashed lines with distances indicated in Ångström.

activity, which could originate from a more accessible active site with a smaller degree of conformational variability during the catalytic cycle. Nevertheless, the kinetic parameters are still within the same magnitude and the retained enzymatic activity seems to indicate that no catalytically important residues are impacted by the truncation.

**$\Delta MtDXPS$  crystal structure.** We previously performed extensive crystallization screening of the wild-type *MtDXPS* enzyme, but were unable to obtain protein crystals. In contrast, the  $\Delta MtDXPS$  protein crystallized within 2 days in several conditions. From the best condition, we were able to determine the crystal structure of the holo protein of  $\Delta MtDXPS$  to a resolution of 1.85 Å (PDB ID: **7A9H**) (Fig. 2a). The numbering of amino acids discussed in the following text is according to the *MtDXPS* sequence with the UniProt ID: **P9WNS3** (**SI: Sequences**).

The asymmetric unit of  $\Delta MtDXPS$  contains a homodimer with each subunit consisting of three distinct domains and a ThDP molecule. Domain I (res 1–312) and Domain II (res 313–483) contribute to the active

site, where ThDP is bound. Domain III (res 484–638) makes extensive contacts at the dimer interface (Fig. 2a). Overall, the structure and domain arrangement of *MtDXPS* show high similarity to the known structures of *E. coli* and *D. radiodurans* DXPS that were described in detail by S. Xiang et al. with a C<sub>α</sub>-RMSD of 1.390 and 1.207 Å, respectively<sup>25</sup>. The main difference between the structures is the linker from amino acids 475–489, located at the surface which adopts a different conformation (Fig. 2c). From a single example of an *MtDXPS* structure we cannot conclude that this variation is not a result of the crystallization process.

Three inserts in the *MtDXPS* sequence were found according to a multiple sequence alignment (MSA) of other bacterial DXPS enzymes (Fig. S1). The first insert comprising amino acids 216–220, is in an evolutionary diverse and disordered loop, which is part of the truncated loop in  $\Delta MtDXPS$ . This region can also not be observed in the wild-type protein crystal structure of other homologues<sup>25–27</sup>. Indeed, the high sequence variability and disorder in the known crystal structures were the reasons to truncate this region for a more stable and more easily crystallizable protein<sup>25–27</sup>.

The other two inserts concern amino acids 498–502 and 523–528 and are located at the protein's solvent-accessible surface (Fig. 2c). Both insertions occur in all *Mycobacterium* strains, but are in a region of high variability with a pairwise sequence identity of 13.7% and 8.6%, compared to the MSA of 498 other bacterial DXPS, respectively<sup>25</sup>. The first is a loop extension of a  $\beta$ -sheet turn, while the second is 30 amino acids downstream and forms an  $\alpha$ -helix, both are located within domain III, responsible for the extensive contacts with the dimeric interface.

As anticipated, the DXPS active site is highly conserved, and most residues of *MtDXPS* share the same position across homologues, including key residues shown in mutational studies to be important in catalysis and substrate binding, such as Glu365, Tyr387, Arg415, Asp422, His40 and His71 (Fig. 2b)<sup>25,32,33</sup>. Three seemingly important differences, however, could be identified near the active site of *MtDXPS*: Lys473, His426 and Ser112. Lys473 plays a significant role in the catalysis and substrate binding, as mutational studies of the corresponding residue on ecDXPS have shown loss of enzymatic activity<sup>25</sup>. Most bacterial DXPS have an arginine in the corresponding position of Lys473. However, the residue is in a conformation similar to the corresponding Arg480/Arg478 of drDXS and ecDXPS structures and the observed  $K_m$  for D-GAP is in the same order of magnitude as for these, indicating that the residues could be interchanged without causing major structural changes in the protein<sup>32,34,35</sup>.

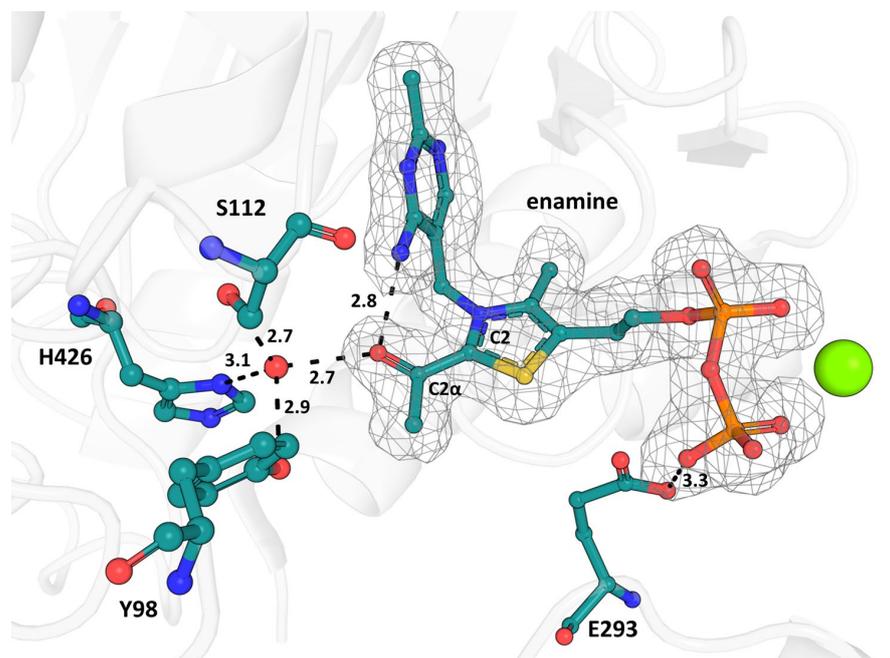
His426 is conserved in all bacterial DXPS but adopts a different conformation in the *MtDXPS* structure. Unlike most bacteria, *M. tuberculosis* has a serine at position 112 instead of a glycine or alanine. The serine at this position creates a steric restriction that prevents His426 from adopting its previously reported conformation, while at the same time enabling the coordination of a water molecule that stabilizes His426 in its new position. The water is highly coordinated, making hydrogen bonds with His426, Ser112, Tyr98 and other water molecules. This network extends to the aminopyrimidine N4' and the backbone and side chains of Ser425 (Fig. 2d). Despite the overall similarities with other homologues, the structural differences close to the active site provide new information that could be specific for *M. tuberculosis* and will be instrumental for future structure-based drug design endeavors.

**$\Delta MtDXPS$  with enamine intermediate.** In order to further elucidate the structural conformations during catalysis, we also attempted to obtain crystal structures of reaction intermediates. It has previously been shown that the lactyl-conformation of ThDP (LThDP) is stable, while the last steps of the reaction, the decarboxylation of LThDP and the addition of D-GAP, are proceeding fast after binding of D-GAP<sup>20</sup>.

A structure of  $\Delta MtDXPS$  simultaneously soaked with pyruvate and D-GAP was solved to a resolution of 1.9 Å (PDB ID: 7A9G) in space group P2<sub>1</sub> (Fig. 3). The asymmetric unit contains a homodimer, which adopts a highly similar fold to the  $\Delta MtDXPS$  holo structure, with an overall RMSD of 0.2 Å. Residues 289–295 are visible in the structure whilst they could not be modeled in the holo structure. In the active site, density could be observed near the C2 of ThDP, consistent with an enamine or acetyl intermediate adduct (Fig. 3)<sup>26</sup>. This decarboxylated form of LThDP shows well-defined electron density with a B-factor of 25 Å<sup>2</sup> (average in crystal is 22 ± 5 Å<sup>2</sup>). The B-factor is similar to that of the surrounding residues and we can assume that nearly all ThDP ligands in the crystal exist in the enamine or acetyl-form. Since we could not determine whether the enamine has converted to a less reactive acetylThDP state, we will refer to the modeled molecule as the enamine-intermediate. This is the first time the enamine-intermediate was observed in a DXPS structure without growing the crystals in an oxygen-free environment, as was necessary for the first structure of a pyruvate adduct<sup>26</sup>. This is likely the result of the very short time between soaking the crystals with pyruvate and D-GAP and the subsequent flash-cooling, capturing the reaction in an intermediate state.

We found that the enamine–ThDP intermediate is engaged in several unique interactions with  $\Delta MtDXPS$  not seen in other structures of DXPS reaction intermediates. Glu293 makes hydrogen bonds with the diphosphate moiety of ThDP and the C2 $\alpha$ -hydroxyl forms hydrogen bonds with Ser112 and His426 through the highly coordinated water also found in the holo structure (Figs. 2d and 3). In line with other reaction intermediate structures from other DXPS homologues, the N4' of the aminopyrimidine hydrogen bonds with the C2 $\alpha$ -hydroxyl.

During the catalytic cycle, ThDP reacts with pyruvate to form stabilized LThDP. Rapid decarboxylation can occur upon D-GAP binding, leading to the formation of the enamine-intermediate. In  $\Delta MtDXPS$ , the enamine intermediate seems to be stabilized via a highly coordinated water molecule also found in the holo structure. Similarly, the water molecule is coordinated to the enamine intermediate, Ser112, Tyr98 and His426, through a network that further extends to Ser425 and the backbone of Ser112. This network of hydrogen bonds might help delocalize the negative charge of the carbanion during the reaction (Fig. 4a). When the enamine is not present (holo structure, PDB ID: 7A9H), the position of the enamine-hydroxyl group is occupied by a second water atom that maintains the hydrogen-bond network (Fig. 2d).



**Figure 3.** Active site of  $\Delta MtDXPS$  with enamine-ThDP intermediate. Polder map density contoured at  $3\sigma$  in gray wire mesh for the enamine-ThDP intermediate (PDB ID: **7A9G**), calculated using Phenix. Key residues involved in the interaction with the enamine are shown and labeled. Distances are indicated in Ångström.

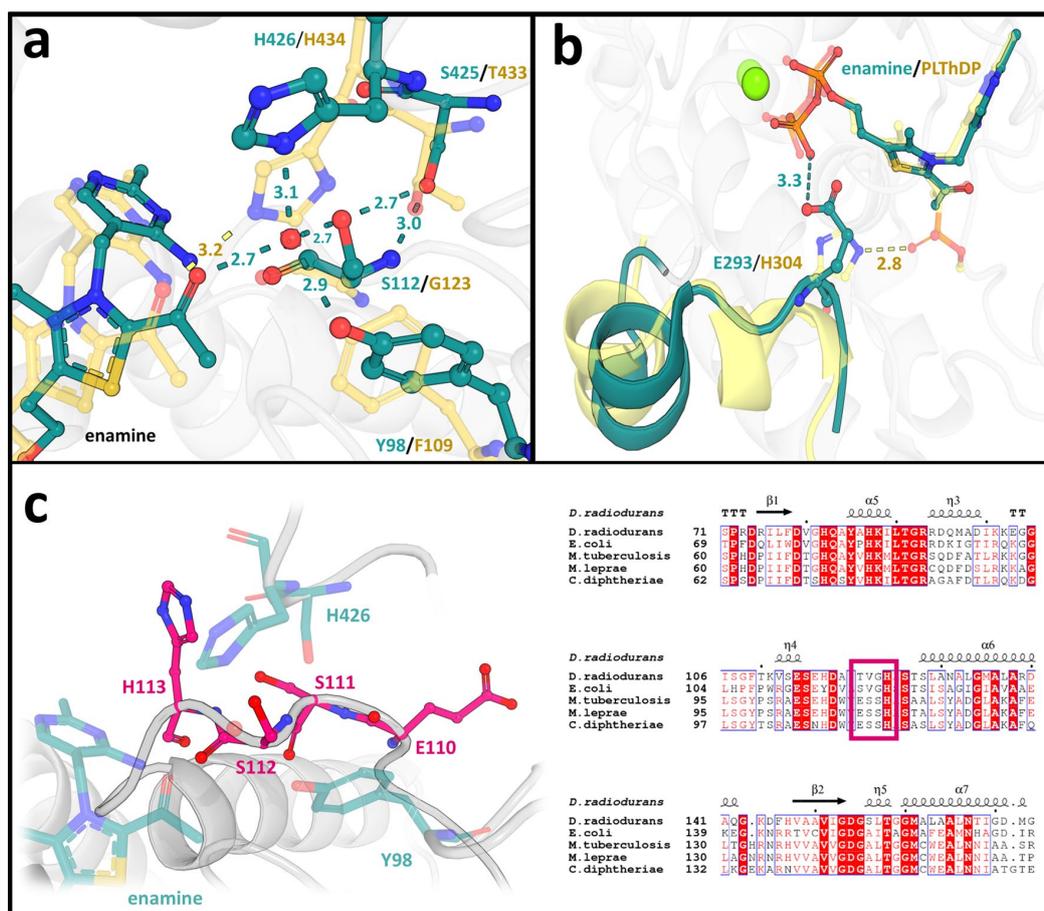
In protein structures of drDXPS and ecDXPS, this hydrogen bond (H-bond) network cannot be observed and the enamine-intermediate stabilization is directly maintained through His434 (His426 in *MtDXPS*) (Fig. 4a). As previously stated, both enzymes have a glycine instead of a serine at position 112 (121 in *E. coli*, 123 in *D. radiodurans*), and consequently no H-bond network can be formed.

The Ser112 is unique for certain bacteria, a multiple sequence alignment with 498 bacterial DXPS sequences showed that only 33 have this amino acid (Table S3). Interestingly, most bacteria that have a serine at this position also have a conserved sequence pattern. The pattern starts with Glu110, followed by Ser111, Ser112 and His113, which we called ESSH sequence motif (Fig. 4c). His 113 is essential, with 100% conservation, but the rest of the sequence is in a variable region with only 43.8% similarity. All Mycobacteria, together with the closely related Corynebacteria and a few other species, bear this sequence motif, highlighting that this different mechanism could be specific for those bacteria (Table S3).

#### Residues near the active site set *MtDXPS* apart from other DXPS and could account for its lower activity.

Two crystal structures from *D. radiodurans* were recently published by Drennan and coworkers, in which a “fork” and a “spoon” motif were introduced. These motifs adopt different conformations during the catalytic cycle<sup>26,29</sup>. The “fork” and “spoon” motifs are two loops adjacent to each other, in positions 292–306 and 307–319 (drDXPS). The corresponding amino acids in *MtDXPS* are 283–299 and 300–312, for the “fork” and “spoon”-motif, respectively. The “fork” motif in the *MtDXPS* structure seems to adopt a slightly different fold and does not include His304 (drDXS) (Fig. 4b). In both structures of *MtDXPS*, Histidine 296, which corresponds to His304 in drDXPS, could not be observed. This His304 is part of the active site and has previously been identified by mutational studies as important for the catalysis through a stabilization of both LThDP and the closed conformation of the enzyme<sup>19,26</sup>. Experiments have demonstrated a 90% reduction of DXPS activity when His304 is mutated to alanine<sup>32</sup>. In both structures of *MtDXPS*, no histidine making similar interactions was observed. In chain A of our intermediate structure (PDB ID: **7A9G**, chain A), part of the residues that would correspond to the fork motif orient in a similar, but slightly different fold, resulting in the position of catalytically important His304 in drDXPS being occupied by Glu293, which forms hydrogen bonds with the diphosphate group of ThDP (Fig. 4b). These residues are disordered in chain B and the holo structure. Since His304 (drDXPS) stabilizes the LThDP by interacting with its carboxylate group, the presence of Glu293 in that location in *MtDXPS* and its interaction with the diphosphate may lead to a lower stabilization of the LThDP intermediate and the closed conformation of the protein, potentially contributing to the lower catalytic activity.

**GAP binding site.** Observation of the second substrate D-GAP was more difficult in the protein crystals soaked with both substrates. This was expected, as we did not trap the enzyme in a catalytic state for instance, by a dead-end substrate, but instead exposed the protein crystals to a solution allowing catalytic activity. This

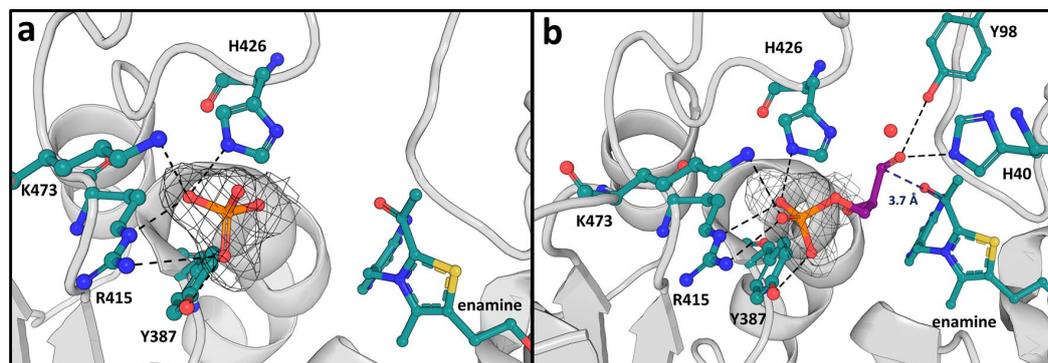


**Figure 4.** Comparison of drDXPS and MtDXPS enamine intermediate active sites. (a) Ser-network/H-bonds stabilizing the enamine intermediate. Residues from MtDXPS (PDB ID: 6OUW) are colored in cyan, while the corresponding residues from drDXPS (PDB ID: 6OUW) are colored in yellow. Hydrogen bonds are shown as dashed lines with distances in Ångström. (b) "Fork" motif. The "fork"-like motif found in the intermediate structure is highlighted in cyan, the similar motif in drDXPS is colored in yellow. The key difference is the presence of Glu293 in the MtDXPS structure. (c) ESSH sequence motif. The ESSH sequence motif found in the  $\Delta$ MtDXPS structures is shown in pink. The table shows the sequence alignment of bacterial DXPS, with several sequences bearing the ESSH sequence motif shown for representation, highlighted with a pink box.

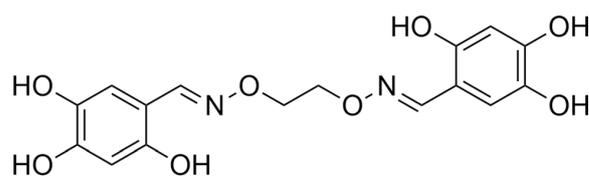
resulted in different catalytic states in the same protein crystal. However, it was possible to identify the electron density of a phosphate-like moiety and model it with an elevated B-factor of  $49.3 \text{ \AA}^2$  (Fig. 5a). While the C3-body of D-GAP cannot be built with confidence, the crystallization conditions do not contain free phosphate, indicating that the observed phosphate density could be provided by a partially disordered D-GAP molecule. The electron density for this moiety can be observed clearly in chain B of the homodimer, while it is more diffuse in chain A. The phosphate is located in close proximity to the enamine intermediate and forms hydrogen bonds with Tyr387, Arg415, His426 and Lys473 with distances between 2.8 and 3.4 Å. The interacting residues correspond to the residues identified to make up the D-GAP binding site in previous studies, determined by molecular docking or kinetic studies and alanine-scanning<sup>20,35</sup>.

Irrespective of the presence of a fully ordered D-GAP, the density of the putative phosphate group provides a reasonable starting point to model and dock the complete D-GAP substrate (Fig. 5b). The substrate molecule docks well into the binding site, with its 3-phosphate moiety superimposing with the observed density and making the same interactions with Tyr387, Arg415, His426 and Lys473 as the phosphate visible in the electron density. The carbonyl group interacts with Tyr98 and His40, which helps to orient the C3-body in an ideal position for the nucleophilic attack of the C2-carbanion of the enamine intermediate (Fig. 5b).

**Docking of MtDXPS inhibitors from the literature.** *Hydroxybenzaldoximes.* With the first protein crystal structure of a pathogenic DXPS homologue it is now possible to understand the binding of inhibitors of DXPS by in-silico docking. As an example, the class of trihydroxybenzaldoximes was chosen. This compound class is one of very few reported D-GAP-competitive inhibitors and does not compete with pyruvate<sup>36</sup>. With



**Figure 5.**  $\Delta MtDXPS$  D-GAP binding site with modeled phosphate and docked D-GAP. **(a)**  $\Delta MtDXPS$  structure (PDB ID: **7A9G**) with polder map density contoured at  $3\sigma$  in gray wire mesh for the phosphate. Key residues involved in the interaction with the phosphate are shown in stick representation and labeled. Hydrogen bonds are shown as black, dashed lines. **(b)** D-GAP molecule docked into the D-GAP binding site of  $\Delta MtDXPS$  after removal of the phosphate. The polder density map of the phosphate present in the crystal structure is shown to highlight the similar orientation and interactions of the 3-phosphate moiety of D-GAP to the modeled phosphate. The residues predicted to interact with the docked D-GAP are shown in stick representation and are labeled. Predicted hydrogen bonds are shown as black, dashed lines, while the distance between the D-GAP C3 and the enamine intermediate  $2C_{\alpha}$  are shown as blue, dashed lines.



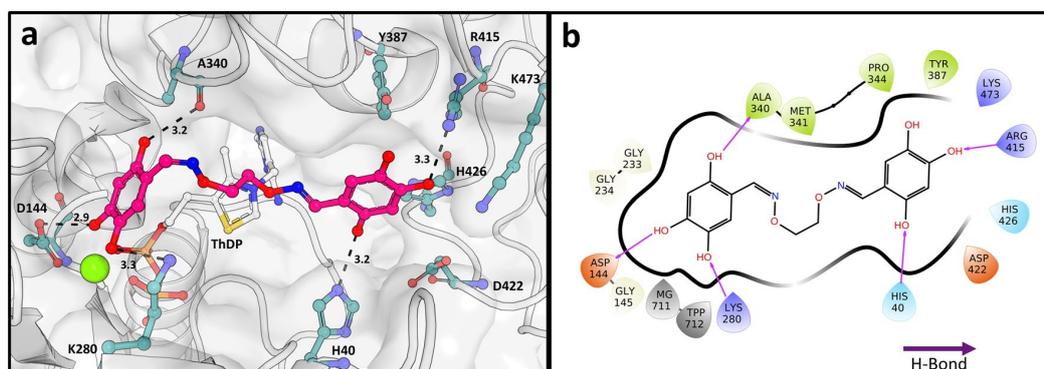
**1**

**Figure 6.** Example structure of the literature known DXS inhibitor class of hydroxybenzaldoximes **1**<sup>36</sup>.

the first crystal structure that resembles the conformation of the active site during the D-GAP addition, and the additional hint for the D-GAP position, we were interested in the interactions of this compound class with the protein.

The class of trihydroxybenzaldoximes was published first as inhibitors of DXPS by Bartee et al.<sup>36</sup>. The inhibitors were developed based on the similarity with the substrates of DXPS. One side was kept as analogue of pyruvate, the first and specific substrate of DXPS, the other side of the chain was varied with aryl residues, based on previous findings that DXPS has a high substrate promiscuity and is able to accept several different acceptor substrates<sup>37,38</sup>. During inhibitor development, Bartee et al. discovered the high activity of compound **1**, which had no pyruvate substitute, but was rather a symmetrical oxime (Fig. 6).

We docked the reported class of hydroxybenzaldoximes to our protein structure, using the docking software SeeSAR. The compound with the highest biological activity ( $K_i$   $1 \pm 0.2 \mu M$ ) against DXPS from *D. radiodurans*, compound **1**, docks well into the active site (Fig. 7). The docking resulted in 20 poses ranked using affinity scores, as calculated by HYDE. The docking pose with the highest affinity score was chosen and shows interactions of the hydroxyl groups from one part of the molecule with the amino acids of the D-GAP binding site, while the linker between the two similar warheads is spanning over the active site. Besides hydrogen bonds with the protein, the second part can form chelating interactions with the bound  $Mg^{2+}$ -ion. This metal interaction is remarkable, as it allows the compound to interact as the sixth coordination partner of the cation, while not competing with the diphosphate group of the ThDP cofactor, but instead additionally forming an H bond with the terminal  $\beta$ -phosphate. When docked to the protein structure with the bound intermediate, the inhibitor adopts a similar pose, shielding the enamine-ThDP from the solvent. This enclosure of the cofactor or intermediate is a possible explanation of the observed distinctive inhibition mode, competitive to D-GAP and noncompetitive to pyruvate<sup>36</sup>.



**Figure 7.** Docking of the hydroxybenzaloxime 1 to  $\Delta MtDXPS$ . (a) Hydroxybenzaloxime 1 docked in the active site of  $\Delta MtDXPS$  with cofactor ThDP bound (PDB ID: 7A9H). The pose of the docked molecule with the highest affinity score, as calculated by HYDE, is shown in pink and the predicted interacting residues are shown in stick presentation and are labeled. Hydrogen bonds are shown as black, dotted lines with distances indicated in Ångström. (b) 2-Dimensional ligand interaction plot of Hydroxybenzaloxime 1 with  $\Delta MtDXPS$ .

## Discussion

DXPS is one of the most important enzymes of the 2-C-methyl-D-erythritol 4-phosphate (MEP)-pathway and is a key target for the development of new anti-infective drugs to treat infections with multidrug-resistant bacteria. Here, we have applied our previously reported truncation approach and removed a disordered internal loop, to successfully crystallize the DXPS from an important anti-infective target, namely *M. tuberculosis*.

The  $\Delta MtDXPS$  structures show a high degree of similarity to known structures. However, our work proposes a different stabilization mechanism of the enamine intermediate in *MtDXPS* when compared to the enamine-intermediate structure of drDXPS. We propose that the presence of the highly coordinated water, found in both the holo and enamine-intermediate structures, interacting with Ser112, His426 and Y98, and the different conformation of H426 might help delocalize the carbanion charge, yielding a more stable intermediate. Even though mutational studies with *E. coli* and *D. radiodurans* have shown that the corresponding residues of H426 (H434 in drDXPS and H431 in ecDXPS) are not essential for catalysis and may therefore mainly contribute to substrate binding affinity, the different conformation of this residue in the  $\Delta MtDXPS$  structures suggests that it plays a different role for *M. tuberculosis* DXPS<sup>32</sup>.

Compared with other DXPS enzymes, which are reported to have  $k_{cat}$  values in the range of 0.5–25 s<sup>-1</sup>, the *MtDXPS* possesses the slowest reaction kinetics for a DXPS homologue described to date, with a  $k_{cat}$  value in the range of 0.005 s<sup>-1</sup><sup>8,39</sup>. The distinct enamine-intermediate stabilization mechanism found in  $\Delta MtDXPS$  might explain its slower reaction rate. The Ser112 present in all *Mycobacterium* could be an important factor that explains this difference in the stabilization mechanism, however, there are no catalytic data from any other homologue containing a serine in this position reported to date. Obtaining kinetic information of such homologues could support these observations. In addition, to confirm whether the serine has any impact on the observed difference in  $k_{cat}$ , the position should be mutated in the *MtDXPS*. The same holds true for the other amino acids involved in the hydrogen-bond network of the water molecule (Tyr 98, His426 and the ESSH sequence motif). Furthermore, these positions could be inserted into other, well characterized DXPS enzymes, such as drDXPS or ecDXPS, in order to obtain valuable information about their potential effect on the reaction rate.

Independent of the direct effect of the serine or the sequence motif on the catalytic activity, species with these features are more likely to have a slightly larger active site, due to the different rotameric state of His426 (Fig. 4a). By targeting this pocket or introducing a group that can interact with the H-bonding network in a similar manner to the intermediate, it might be possible to develop DXPS inhibitors selectively targeting *M. tuberculosis* over other species.

Distinct from previous studies performed on drDXPS, which found a fork-spoon motif that is positioned to open and close the active site based on substrate binding<sup>19,26</sup>, a similar but divergent structural feature was found for *MtDXPS*. This ‘fork-like’ motif utilizes a different residue for the interaction with the cofactor, potentially leading to a lesser degree of fixation of the active site conformation.

As of yet, we could not obtain a structure of *MtDXPS* with a fully ordered spoon-fork motif. This leaves the question open whether His296 interacts, in any conformation, with the active site of *MtDXPS*. If His296 is permanently located outside of the active site and the corresponding position is taken by Glu293, which makes different interactions with the diphosphate moiety of ThDP, this could also explain the lower enzymatic activity of *MtDXPS*. As mentioned, the mutation of His304 to alanine in drDXPS leads to a 90% reduction in activity<sup>32</sup>. Similar mutational studies in *MtDXPS* could be performed to evaluate their impact on the enzymatic activity, such as mutating Glu293 and His296 to alanine, as these might be crucial for the observed differences in the spoon-fork motif. Since drDXPS and ecDXPS do not have a corresponding amino acid in this position (Fig. S1), inserting it in their sequences could grant further insight into its role. A *MtDXPS* structure containing the LThDP mimic PLThDP could provide valuable insights into the conformation and interactions of the fork-like motif pre-decarboxylation and its potential effect on the catalytic activity of *MtDXPS*. Unfortunately, thus far, we were unable to obtain such a structure. However, the availability of high-resolution crystals of *MtDXPS* allows a more

complete exploration of the molecular mechanisms behind the lower DXPS activity of *MtDXPS*, which will be the subject of future research.

Furthermore, this was the first time that phosphate-like electron density could be observed near the predicted D-GAP binding site in a DXPS structure, providing new evidence of the putative binding site and residues involved in substrate stabilization. While the observed density and docking of the D-GAP substrate do not provide definitive evidence of D-GAP binding, these results fit well with expectations of D-GAP substrate binding<sup>20,35</sup>. Chain A of the intermediate structure shows a partially folded, fork-like motif with Glu293 interacting with ThDP and only weak density in the D-GAP binding site. The fork-like motif in chain B is disordered but phosphate-like density can be observed in the D-GAP binding site. It could be postulated that the enzyme can adopt both the open and closed conformation post-decarboxylation depending on the presence of D-GAP/phosphate. Bound D-GAP/phosphate would induce an open conformation whilst the fork-like motif closes in the absence of D-GAP.

As mentioned, our group attempted a ligand-based approach for the discovery of inhibitors of DXPS using a drDXPS crystal structure and a homology model of *MtDXPS* derived from that structure. This yielded compounds that showed activity against drDXPS but less so against *MtDXPS*. The homology model active site differs from the *mtDXPS* crystal structure active site in the same manner as the drDXPS active site differs from the *MtDXPS*, which can be explained by the fact that it was modeled using the drDXPS crystal structure. The conformation of the spoon-fork motif, the orientation of His426, the absence of the coordinated water and the structural position of Glu293 of the homology model match the drDXPS crystal structure, whilst the *MtDXPS* crystal structure shows that these are distinct (data not shown). Since these amino acids were proposed to form several key interactions with the compounds designed in our previous paper<sup>22</sup>, it could explain why they did not have the expected activity against *MtDXPS*. This highlights the importance of obtaining crystal structures as computationally generated models do not always reflect the real conformations of proteins, leading to inhibitor design using incorrect data.

In summary, the successful use of the loop truncation for the crystallization of DXPS from a different species, provides a solid platform that can extend the structural study of DXPS homologues. All structural differences obtained, such as the ones found for *MtDXPS*, can then assist in the development of new antibiotics with a high specificity for species that are already drug-resistant, such as *M. tuberculosis*. Additionally, our results provide a new opportunity to investigate and improve previously identified inhibitors of *MtDXPS* by docking them into the experimentally determined structures, allowing for the elucidation of SARs and facilitating the design and hit-to-lead optimization of TB-specific inhibitors. The next step will be to obtain co-crystal structures with inhibitors bound to the enzyme.

## Methods

**Expression and purification of *MtDXPS*.** The design of a pET22b plasmid hosting the native *MtDXPS* gene with an N-terminal His-Tag is reported elsewhere<sup>40</sup>. The plasmid was transformed into *E. coli* BL21 (DE3) cells, which were grown at 37 °C. At an OD<sub>600</sub> of ~0.6, protein expression was induced by the addition of isopropyl β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM, and the cells were incubated at 18 °C overnight. The bacteria were harvested by centrifugation and resuspended in lysis buffer [50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM imidazole, 5 mM β-mercaptoethanol (β-ME), 5 mM MgCl<sub>2</sub>, 2.5 U/mL benzonase, 1 tablet/200 mL Complete Protease inhibitor, EDTA free (Roche)]. The cells were lysed using a Microfluidizer M-110P, and the lysate clarified by centrifugation for 45 min at 18,000g. The lysate was filtered using a 0.45 μm syringe filter, loaded onto a metal-affinity HisTrap HP 5 mL column (GE Healthcare) and eluted using a linear gradient of imidazole from 10 to 500 mM. Protein-containing fractions were pooled and concentrated using VivaSpin ultrafiltration devices with a MWCO of 30 kDa. The concentrated sample was further purified by gel filtration on a HiLoad 16/600 Superdex 200 pg column equilibrated in 20 mM Bis-Tris-Propane, pH 7.5; 300 mM NaCl, 1 mM MgCl<sub>2</sub>, 0.5 mM TCEP-HCl.

**Expression and purification of Δ*MtDXPS*.** The truncated *MtDXPS* gene was obtained commercially as a synthetic gene and cloned into a pETM-11 plasmid using the NcoI and HindIII restriction sites. Expression, lysis and the initial IMAC purification were performed as above for full-length *MtDXPS*. The protein-containing fractions were subsequently combined and diluted with a low-salt buffer consisting of 50 mM HEPES pH 8.0, 5% glycerol, 5 mM dithiothreitol (DTT) and 100 μM MgCl<sub>2</sub> to a conductivity of 8 mS/cm. The solution was then loaded on a Resource Q anion exchange column and eluted with a linear NaCl gradient from 0 to 1 M. The protein-containing fractions were pooled and purified by gel filtration on a HiLoad 16/600 Superdex 200 pg column equilibrated with 20 mM HEPES pH 8.0, 250 mM NaCl, 5% glycerol, 5 mM DTT. The protein-containing fractions were concentrated to 5.5 mg/mL using a VivaSpin centrifugal concentrator (MWCO 30 kDa, Sartorius), and the His-tag was cleaved by TEV-protease digestion at 10 °C overnight. Removal of the tag and protease was achieved by reversed IMAC chromatography, and the protein was purified again by gel filtration on a HiLoad 16/600 Superdex 200 pg column using 20 mM MOPS pH 7.50, 200 mM NaCl, 5% Glycerol, 2 mM DTT as buffer. The purified protein was concentrated to 10 mg/mL in a VivaSpin centrifugal concentrator (MWCO 10 kDa, Sartorius).

**Protein crystallization and structure determination.** A high-throughput crystallization robot (Mosquito, SPT Labtech) was used to perform sitting-drop vapor-diffusion screenings for crystallization conditions for Δ*MtDXPS*. Prior to screening, 1 mM of ThDP and 2 mM MgCl<sub>2</sub> were added to the protein solution (10 mg/mL) and then incubated at 4 °C for 3 h. The mixture was then screened against commercially available sparse-matrix screening kits (PACT premier, JSCG plus and MORPHEUS; Molecular Dimensions) at 18 °C. 200 nL of

protein solution and 200 nL of crystallization buffer were equilibrated against 50  $\mu$ L of reservoir solution. Initial crystals were obtained after 2 days in MORPHEUS conditions B2 (0.1 M MES/imidazole pH 6.5, 10% w/v PEG 8000, 20% v/v ethylene glycol, 0.3 M sodium fluoride, 0.3 M sodium bromide and 0.3 M sodium iodide) and C2 (0.1 M MES/imidazole pH 6.5, 10% w/v PEG 8000, 20% v/v ethylene glycol, 0.3 M sodium nitrate, 0.3 M disodium hydrogen phosphate and 0.3 M ammonium sulfate) as well as PACT premier D2 (0.1 M MMT-buffer pH 5.0, 25% PEG1500). The structures described in this manuscript were obtained from crystals grown in condition PACT premier D2. The crystals were cryo-protected using reservoir solution supplemented with 30% (v/v) PEG 400 and flash-cooled in liquid nitrogen.

To obtain  $\Delta MfDXPS$  with reaction intermediate, crystals grown in the same condition were harvested, soaked for one minute in the above-mentioned cryo-protecting solution supplemented with 1 mM of D-GAP and 1 mM of pyruvate and subsequently flash-cooled in liquid nitrogen. X-ray diffraction data for the  $\Delta MfDXPS$  structures were collected on beamline P13 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany) at 100 K. Data indexing, integration and scaling was performed using XDSAPP<sup>41</sup> and AIMLESS<sup>42</sup> from the CCP4<sup>43</sup> software package. The structure was solved using the MOLREP<sup>44</sup> software in CCP4 using the previously solved drDXPS homologue (PDB ID: **2O1X**) as the reference model. The resulting models were then subjected to iterative cycles of model building and refinement with COOT<sup>45</sup> and REFMAC5<sup>46</sup>.

The structures were deposited in the PDB with accession codes **7A9H** and **7A9G**, corresponding to the holo and reaction intermediate structures, respectively. Table S2 contains the data collection and refinement statistics.

**Enzymatic assay.** The DXPS activity was analyzed at RT as previously reported, with minor modifications<sup>40,47</sup>. A continuous kinetic photometric assay was used to measure DXPS activity. NADPH depletion by the downstream IspC enzyme was determined in a microplate reader (PHERAstar, BMG Labtech) by monitoring the decrease in absorbance at 340 nm. Total assay volume was 60  $\mu$ L, containing 200 mM HEPES pH 8.0, 2 mM DTT, 1 mM MgCl<sub>2</sub>, 0.3 mM NADPH and 1.5  $\mu$ M IspC (from *E.coli*, expressed and purified in-house according to a literature procedure)<sup>48</sup>. The amount of DXPS used in the assays was determined experimentally by a dilution series of the enzyme. These were 5  $\mu$ M and 2  $\mu$ M for  $\Delta MfDXPS$  and *MfDXPS*, respectively, as they showed the highest linear reaction velocity without observable substrate depletion over a time range of 30 min. The reaction was monitored at RT for 30 min after addition of the substrate(s) and 1 min of centrifugation (2000 rpm). To determine the corresponding  $K_m$  values, the compounds ThDP, pyruvate and D-GAP were used in varying concentrations. If a substrate or cofactor was kept constant, a concentration of 0.2 mM was used for ThDP, 0.5 mM for pyruvate and 2 mM for D-GAP.

Blank correction and linear fitting of the absorption data was performed using the program Origin 2019 (OriginLab). The initial velocities obtained were plotted against the substrate concentrations, and the  $K_m$  values were determined by nonlinear curve fitting using the Michaelis–Menten model of the enzyme kinetics add-on of Origin2019.

**Thermal shift assay (TSA).** Thermal shift analyses were performed using an ABI StepOneplus RT-PCR instrument. The samples were measured in white 96-well plates. Denaturation was achieved using a continuous heating rate of 1 °C/min from 20 to 95 °C. The total sample volume was 25  $\mu$ L, consisting of 20  $\mu$ L TSA buffer (20 mM Tris–HCl, pH 8.0; 100 mM NaCl, 5 mM MgCl<sub>2</sub>), 2.5  $\mu$ L protein solution and 2.5  $\mu$ L dye (Sypro Orange, Sigma-Aldrich). The optimal concentrations were experimentally determined. A final concentration in the plate of 1.5  $\mu$ M protein and 5 $\times$  SYPRO Orange yielded the best signal-to-noise ratio. All measurements were performed in duplicate.

**LC–MS.** All ESI–MS-measurements were performed on a Dionex Ultimate 3000 RSLC system using an Aeris Widepore XB-C8, 150 $\times$ 2.1 mm, 3.6  $\mu$ m dp column (Phenomenex, USA). Separation of 1  $\mu$ L sample was achieved by a linear gradient from (A) H<sub>2</sub>O+0.1% formic acid (FA) to (B) ACN+0.1% FA at a flow rate of 300  $\mu$ L/min and 45 °C. The gradient was initiated by a 0.5 min isocratic step at 2% B, followed by an increase to 75% B in 10 min to end with a 3 min step at 75% B before re-equilibration with initial conditions. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75  $\mu$ L/min before entering the maXis 4G hr-ToF mass spectrometer (Bruker Daltonics, Bremen, Germany), using the standard Bruker ESI source. In the source region, the temperature was set to 200 °C, the capillary voltage was 4000 V, the dry-gas flow was 5.0 L/min and the nebulizer was set to 1.0 bar. Mass spectra were acquired in positive ionization mode ranging from 600 to 1800 m/z at 2.5 Hz scan rate. Protein masses were deconvoluted by using the Maximum Entropy algorithm (Copyright 1991–2004 Spectrum Square Associates, Inc.).

**Modeling and docking.** The computer program SeeSAR, version 10.3.1 from BioSolveIT was used to generate the docking poses and calculate binding affinities. The software uses the FlexX docking algorithm for the placement of ligands<sup>49</sup>. The affinities are estimated using the HYDE scoring function, which calculates the binding affinities based on the hydration differences between the bound and unbound state of the molecule<sup>50,51</sup>. The binding site was chosen around the ThDP ligand, extending to the residues Tyr387, Arg415 and Lys473, which bind the D-GAP substrate. Sequence numbering is based on the Uniprot sequence file with the code **P9WNS3**. Analysis and visualization of the results were done using the program StarDrop, version 6.6.7.25378 from Optibrium.

### Code availability

**Accession numbers** Protein structures are deposited in the PDB archive with the PDB ID: [7A9G](#) and [7A9H](#). All raw data presented in this publication are available upon request from Matthew R. Groves, email: [m.r.groves@rug.nl](mailto:m.r.groves@rug.nl).

Received: 5 October 2021; Accepted: 15 April 2022

Published online: 04 May 2022

### References

- Nahid, P., Pai, M. & Hopewell, P. C. Advances in the diagnosis and treatment of tuberculosis. *Proc. Am. Thorac. Soc.* **3**, 103–110 (2006).
- WHO. *The End TB Strategy* (WHO, 2014).
- Matteelli, A., Roggi, A. & Carvalho, A. C. C. Extensively drug-resistant tuberculosis: Epidemiology and management. *Clin. Epidemiol.* **6**, 111–118 (2014).
- Khawbung, J. L., Nath, D. & Chakraborty, S. Drug resistant tuberculosis: A review. *Comp. Immunol. Microbiol. Infect. Dis.* **74**, 101574 (2021).
- Udwadia, Z. F., Amale, R. A., Ajbani, K. K. & Rodrigues, C. Totally drug-resistant tuberculosis in India. *Clin. Infect. Dis.* **54**, 579–581 (2006).
- Velayati, A. A. *et al.* Emergence of new forms of totally drug-resistant tuberculosis bacilli: Super extensively drug-resistant tuberculosis or totally drug-resistant strains in Iran. *Chest* **136**, 420–425 (2009).
- Wang, X. & Dowd, C. S. The methylerythritol phosphate pathway: Promising drug targets in the fight against tuberculosis. *ACS Infect. Dis.* **4**, 278–290 (2018).
- Gierse, R. M. *et al.* DXS as a target for structure-based drug design. *Future Med. Chem.* **7**, 2131–2141 (2017).
- Mombo-Ngoma, G. *et al.* Fosmidomycin-piperazine as non-artemisinin-based combination for acute uncomplicated plasmodium falciparum malaria. *BMJ Glob. Heal.* **2**, A22.3–A23 (2017).
- Masini, T., Kroezen, B. S. & Hirsch, A. K. H. Druggability of the enzymes of the non-mevalonate-pathway. *Drug Discov. Today* **18**, 1256–1262 (2013).
- Courtens, C. *et al.* Amino acid based prodrugs of a fosmidomycin surrogate as antimalarial and antitubercular agents. *Bioorganic Med. Chem.* **27**, 729–747 (2019).
- Bartee, D. *et al.* Enamide prodrugs of acetyl phosphonate deoxy-D-xylulose-5-phosphate synthase inhibitors as potent antibacterial agents. *ACS Infect. Dis.* **5**, 406–417 (2019).
- Kesharwani, S. & Sundriyal, S. Non-hydroxamate inhibitors of 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR): A critical review and future perspective. *Eur. J. Med. Chem.* <https://doi.org/10.1016/j.ejmech.2020.113055> (2020).
- Balibar, C. J. Antibacterial leads targeting isoprenoid biosynthesis. In *Antibiotic Drug Discovery* (eds Firestone, S. M. & Lister, T.) 204–255 (Royal Society of Chemistry, 2017). <https://doi.org/10.1039/9781782629870-00204>.
- Mombelli, P. *et al.* Imidazole- and benzimidazole-based inhibitors of the kinase IspE: Targeting the substrate-binding site and the triphosphate-binding loop of the ATP site. *Eur. J. Org. Chem.* **2013**, 1068–1079 (2013).
- Deng, L. *et al.* Inhibition of 1-deoxy-D-xylulose-5-phosphate reductoisomerase by lipophilic phosphonates: SAR, QSAR, and crystallographic studies. *J. Med. Chem.* **54**, 4721–4734 (2011).
- Thelemann, J. *et al.* Aryl bis-sulfonamide inhibitors of IspF from *Arabidopsis thaliana* and *Plasmodium falciparum*. *ChemMedChem* **10**, 2090–2098 (2015).
- Estévez, J. M., Cantero, A., Reindl, A., Reichler, S. & León, P. 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. *J. Biol. Chem.* **276**, 22901–22909 (2001).
- Decolli, A. A., Zhang, X., Heflin, K. L., Jordan, F. & Freel Meyers, C. L. Active site histidines link conformational dynamics with catalysis on anti-infective target 1-deoxy-D-xylulose 5-phosphate synthase. *Biochemistry* **58**, 4970–4982 (2019).
- Brammer, L. A., Smith, J. M., Wades, H. & Meyers, C. F. 1-Deoxy-D-xylulose 5-phosphate synthase catalyzes a novel random sequential mechanism. *J. Biol. Chem.* **286**, 36522–36531 (2011).
- Müller, I. B., Hyde, J. E. & Wrenger, C. Vitamin B metabolism in *Plasmodium falciparum* as a source of drug targets. *Trends Parasitol.* **26**, 35–43 (2010).
- Masini, T. *et al.* Validation of a homology model of *Mycobacterium tuberculosis* DXS: Rationalization of observed activities of thiamine derivatives as potent inhibitors of two orthologues of DXS. *Org. Biomol. Chem.* **13**, 11263–11277 (2015).
- Klebe, G. Drug design: Methodology, concepts, and mode-of-action. *Drug Des. Methodol. Concepts Mode-of-Action*. <https://doi.org/10.1007/978-3-642-17907-5> (2013).
- Košak, U. *et al.* The magic of crystal structure-based inhibitor optimization: Development of a butyrylcholinesterase inhibitor with picomolar affinity and in vivo activity. *J. Med. Chem.* **61**, 119–139 (2018).
- Xiang, S., Usunow, G., Lange, G., Busch, M. & Tong, L. Crystal Structure of 1-deoxy-D-xylulose 5-phosphate synthase, a crucial enzyme for isoprenoids biosynthesis. *J. Biol. Chem.* **282**, 2676–2682 (2007).
- Chen, P. Y. T., DeColli, A. A., Freel Meyers, C. L. & Drennan, C. L. X-ray crystallography-based structural elucidation of enzyme-bound intermediates along the 1-deoxy-D-xylulose 5-phosphate synthase reaction coordinate. *J. Biol. Chem.* **294**, 12405–12414 (2019).
- Gierse, R. M. *et al.* Identification of a 1-deoxy-D-xylulose-5-phosphate synthase (DXS) mutant with improved crystallographic properties. *Biochem. Biophys. Res. Commun.* **539**, 42–47 (2021).
- Altincicek, B. *et al.* Tools for discovery of inhibitors of the 1-deoxy-D-xylulose 5-phosphate (DXP) synthase and DXP reductoisomerase: An approach with enzymes from the pathogenic bacterium *Pseudomonas aeruginosa*. *FEMS Microbiol. Lett.* **190**, 329–333 (2000).
- Zhou, J. *et al.* Conformational dynamics of 1-deoxy-D-xylulose 5-phosphate synthase on ligand binding revealed by H/D exchange MS. *Proc. Natl. Acad. Sci.* **114**, 9355–9360 (2017).
- Jumde, R. P. *et al.* Hit-optimization using target-directed dynamic combinatorial chemistry: Development of inhibitors of the anti-infective target 1-deoxy-D-xylulose-5-phosphate synthase. *Chem. Sci.* **12**, 7775–7785 (2021).
- Deller, M. C., Kong, L. & Rupp, B. Protein stability: A crystallographer's perspective. *Acta Crystallogr. Sect. Struct. Biol. Commun.* **72**, 72–95 (2016).
- Handa, S. *et al.* Mechanistic studies of 1-deoxy-D-xylulose-5-phosphate synthase from *Deinococcus radiodurans*. *Biochem. Mol. Biol. J.* **04**, 1–11 (2018).
- Querol, J., Rodríguez-Concepción, M., Boronat, A. & Imperial, S. Essential role of residue H49 for activity of *Escherichia coli* 1-deoxy-D-xylulose 5-phosphate synthase, the enzyme catalyzing the first step of the 2-C-methyl-D-erythritol 4-phosphate pathway for isoprenoid synthesis. *Biochem. Biophys. Res. Commun.* **289**, 155–160 (2001).
- White, J. K., Handa, S., Vankayala, S. L., Merkler, D. J. & Woodcock, H. L. Thiamin diphosphate activation in 1-deoxy-D-xylulose 5-phosphate synthase: insights into the mechanism and underlying intermolecular interactions. *J. Phys. Chem. B* **120**, 9922–9934 (2016).

35. BrammerBasta, L. A., Patel, H., Kakalis, L., Jordan, F. & Freel Meyers, C. L. Defining critical residues for substrate binding to 1-deoxy-D-xylulose 5-phosphate synthase—Active site substitutions stabilize the predecarboxylation intermediate C2 $\alpha$ -lactylthiamin diphosphate. *FEBS J.* **281**, 2820–2837 (2014).
36. Bartee, D., Morris, F., Al-khouja, A. & Freel Meyers, C. L. Hydroxybenzaloximes are D-GAP-competitive inhibitors of *E. coli* 1-deoxy-D-xylulose-5-phosphate synthase. *ChemBioChem* **16**, 1771–1781 (2015).
37. Brammer, L. A. & Meyers, C. F. Revealing substrate promiscuity of 1-deoxy-D-xylulose 5-phosphate synthase. *Organic letters* **11**, 4748–4751 (2009).
38. Morris, F., Vierling, R., Boucher, L., Bosch, J. & Freel Meyers, C. L. DXP synthase-catalyzed C–N bond formation: Nitroso substrate specificity studies guide selective inhibitor design. *ChemBioChem* **14**, 1309–1315 (2013).
39. Chang, A. *et al.* BRENDA, the ELIXIR core data resource in 2021: New developments and updates. *Nucleic Acids Res.* **49**, D498–D508 (2021).
40. Masini, T. *et al.* De novo fragment-based design of inhibitors of DXS guided by spin-diffusion-based NMR spectroscopy. *Chem. Sci.* **5**, 3543–3551 (2014).
41. Krug, M., Weiss, M. S., Heinemann, U. & Mueller, U. XDSAPP: A graphical user interface for the convenient processing of diffraction data using XDS. *J. Appl. Crystallogr.* **45**, 568–572 (2012).
42. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution?. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
43. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
44. Vagin, A. & Teplyakov, A. MOLREP: An automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).
45. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
46. Skubák, P., Murshudov, G. N. & Pannu, N. S. Direct incorporation of experimental phase information in model refinement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2196–2201 (2004).
47. Hecht, S. *et al.* Biosynthesis of terpenoids: Efficient multistep biotransformation procedures affording isotope-labeled 2C-methyl-D-erythritol 4-phosphate using recombinant 2C-methyl-D-erythritol 4-phosphate Synthase. *J. Org. Chem.* **66**, 7770–7775 (2001).
48. Hecht, S. *et al.* Enzyme-assisted preparation of isotope-labeled 1-deoxy-D-xylulose 5-phosphate. *J. Org. Chem.* **66**, 3948–3952 (2001).
49. Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **261**, 470–489 (1996).
50. Schneider, N., Lange, G., Hindle, S., Klein, R. & Rarey, M. A consistent description of Hydrogen bond and DEhydration energies in protein-ligand complexes: Methods behind the HYDE scoring function. *J. Comput. Aided. Mol. Des.* **27**, 15–29 (2013).
51. Reulecke, L., Lange, G., Albrecht, J., Klein, R. & Rarey, M. Towards an integrated description of hydrogen bonding and dehydration: Decreasing false positives in virtual screening with the HYDE scoring function. *ChemMedChem* **3**, 885–897 (2008).

### Acknowledgements

The authors would like to thank Chantal Bader and Patrick Haack for LC-MS measurements. We acknowledge DESY (Hamburg, Germany), member of the Helmholtz Association HGF, and EMBL Hamburg for the provision of experimental facilities. Parts of this research was done at PETRA III, Beamlines P11 and P13 and we thank Anja Burkhardt for assistance. We also thank the joint EMBL and ESRF group for access to synchrotron beamlines ID29, ID23-1 and ID23-2 (Grenoble).

### Author contributions

R.G., R.O., V.G. and M.G. wrote the main manuscript text and R.G. and V.O. prepared all figures. All authors reviewed the manuscript.

### Funding

This work was funded by The Netherlands Organisation for Scientific Research (NWO), LIFT Grant 731.015.414; the Helmholtz Association's Initiative and Networking Fund, and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 860816.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-11205-9>.

**Correspondence** and requests for materials should be addressed to A.K.H.H. or M.R.G.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

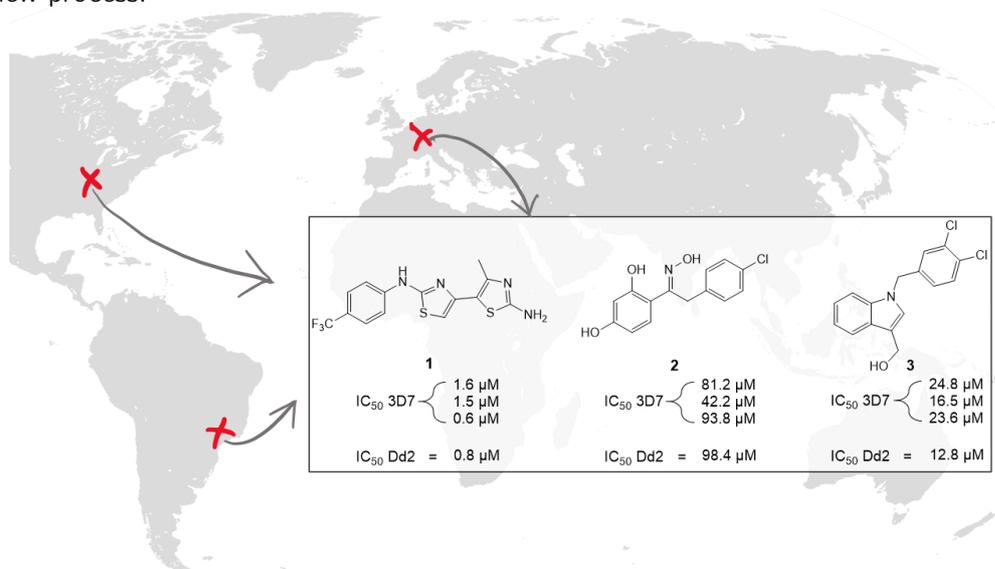
2.2. FIRST CRYSTAL STRUCTURES OF DXPS FROM *MYCOBACTERIUM TUBERCULOSIS* INDICATE A  
DISTINCT MECHANISM OF INTERMEDIATE STABILIZATION.

---



## 2.3 Identification of Three new Inhibitor Classes Against *Plasmodium falciparum*

The following chapter is published on the pre-print server ChemRxiv. It is also submitted to the Journal of Medicinal Chemistry and currently *under revision*. In this thesis the submitted manuscript is printed, which may be subject to changes in the course of the review process.



### Title

Identification of three new inhibitor classes against *Plasmodium falciparum*

### Authors

Sandra Johannsen ‡, Robin M. Gierse ‡, Arne Krüger, Rachel L. Edwards, Vittoria Nanna, Anna Fontana, Zhu Di, Tiziana Masini, Lais Pessanha de Carvalho, Mael Poizat, Bart Kieftenbelt, Dana M. Hodge, Daan Bunt, Kamila Anna Meissner, Edmarcia Elisa de Souza, Melloney Dröge, Bernard van Vliet, Jack den Hartog, Michael C. Hutter, Jana Held, Audrey R. Odom John, Carsten Wrenger, Anna K. H. Hirsch

### Contributions

S. Johannsen was involved in the synthesis, characterized the compounds and took part in drafting the manuscript. R. M. Gierse performed the bioinformatic target search and took part in drafting the manuscript. A. Krüger, K. A. Meissner, E. E. de Souza and C. Wrenger created transgenic *P. falciparum* strains and performed bioactivity assays. V. Nanna, A. Fontana, D. Bunt, M. Poizat, B. van Vliet, J. den Hartog and B. Kieftenbelt contributed to the synthesis of compounds. T. Masini and Z. Di performed the initial ligand-based screening and activity assays. M. Dröge performed in-vitro bioassays. M. C. Hutter took part in the bioinformatic analysis and calculated the homology models. L. Pessanha de Carvalho and J. Held performed bioactivity assays on *P. falciparum*. R. L. Edwards, D. M. Hodge and A. R. Odom John performed LC-MS based target validation and rescue assays. A. K. H. Hirsch was involved in editing the manuscript and supervised the project.

### Bibliographic Details

ChemRxiv

12 April 2022

DOI: 10.26434/chemrxiv-2022-npslc

‡: Authors contributed equally

## Identification of three new inhibitor classes against *Plasmodium falciparum*

Sandra Johannsen<sup>1,2,†</sup>, Robin M. Gierse<sup>1,2,3,†</sup>, Arne Krüger<sup>4</sup>, Rachel L. Edwards<sup>5</sup>, Vittoria Nanna<sup>1</sup>, Anna Fontana<sup>1</sup>, Di Zhu<sup>1,3</sup>, Tiziana Masini<sup>3</sup>, Lais Pessanha de Carvalho<sup>6</sup>, Mael Poizat<sup>7</sup>, Bart Kieftenbelt<sup>7</sup>, Dana M. Hodge<sup>8</sup>, Sophie Alvarez<sup>9</sup>, Daan Bunt<sup>3</sup>, Kamila Anna Meissner<sup>4</sup>, Edmarcia Elisa de Souza<sup>4</sup>, Melloney Dröge<sup>7</sup>, Bernard van Vliet<sup>7</sup>, Jack den Hartog<sup>7</sup>, Michael C. Hutter<sup>10</sup>, Jana Held<sup>6</sup>, Audrey R. Odom John<sup>8</sup>, Carsten Wrenger<sup>4</sup>, Anna K. H. Hirsch<sup>1,2,3,\*</sup>

1 Helmholtz Institute for Pharmaceutical Research Saarland (HIPS) – Helmholtz Centre for Infection Research (HZI), Campus Building E8.1, 66123 Saarbrücken, Germany

2 Department of Pharmacy, Saarland University, Campus Building E8.1, 66123 Saarbrücken, Germany

3 Stratingh Institute for Chemistry, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

4 Unit for Drug Discovery, Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, Av. Prof. Lineu Prestes 1374, 05508-000 São Paulo-SP, Brazil

5 Department of Pediatrics, Washington University School of Medicine, Saint Louis, Missouri 63110, United States

6 Institute of Tropical Medicine, University of Tübingen, Wilhelmstraße 27, 72074 Tübingen, Germany

7 Symeres, Kadijk 3, 9747 AT Groningen, The Netherlands

8 Department of Pediatrics, Children's Hospital of Philadelphia, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA 19104, United States

9 Proteomics & Metabolomics Facility, Center for Biotechnology, Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln , Nebraska 68588, United States

10 Center for Bioinformatics, Saarland University, Campus Building E2.1, 66123 Saarbrücken, Germany

## Abstract

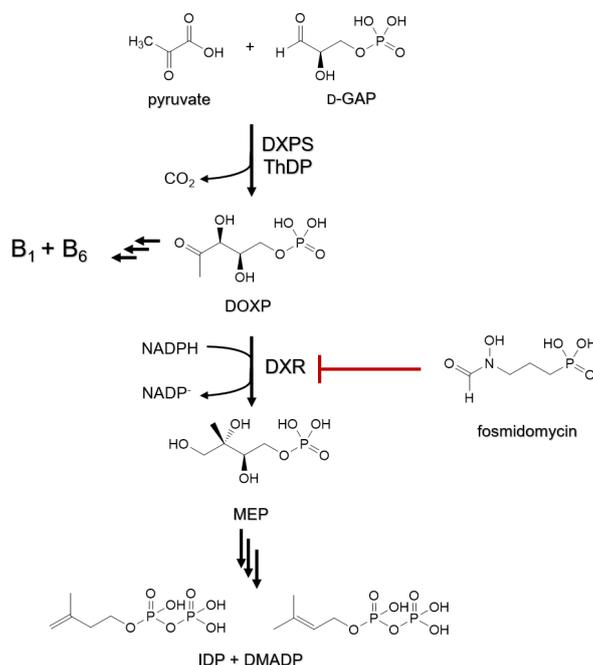
In this study, we identified three novel compound classes with potent activity against *Plasmodium falciparum*, the most dangerous human malarial parasite. Resistance of this pathogen to known drugs is increasing and compounds with different modes of action are urgently needed. One promising drug target is the enzyme 1-deoxy-D-xylulose-5-phosphate synthase (DXPS) of the methylerythritol 4-phosphate (MEP) pathway for which we have previously identified three active compound classes against *Mycobacterium tuberculosis*. The close structural similarities in the active sites of the DXPS enzymes of *P. falciparum* and *M. tuberculosis* prompted investigation of its antiparasitic action, displaying good cell-based activity for all classes. Through structure-activity relationship studies we increased their antimalarial potency, and two classes also show good metabolic stability and low toxicity against human liver cells. The most active compound **1** inhibits the growth of blood-stage *P. falciparum* with an IC<sub>50</sub> of 600 nM. The results from three different methods for target validation of compound **1** suggest intracellular polypharmacy. Similarity-based searches revealed two other possible target enzymes for this compound, which were further analyzed by docking calculations. All inhibitor classes are active against chloroquine resistant strains, confirming a new mode of action.

## Introduction

Malaria remains one of the major diseases with a high impact on health and welfare worldwide, especially in sub-tropical regions. In 2020, the World Health Organization (WHO) reported an estimated number of 627,000 deaths worldwide.<sup>1</sup> Among the six known human malaria parasites, *Plasmodium falciparum* is responsible for the majority of deaths. To treat uncomplicated *P. falciparum* malaria, artemisinin-based combination therapies (ACTs) are recommended, but the potent artemisinin derivatives must be partnered with a second drug due to their short half-life. Currently, six different ACTs are in use, but decreasing potency of artemisinin derivatives displayed by a delayed clearance phenotype is wide-spread in South-East Asia, together with resistances to the partner drugs in this combinations are threatening the efficacy of these treatments.<sup>2</sup> Therefore, finding new compounds with novel modes of action is of great importance.

A promising pool of targets is the methylerythritol 4-phosphate (MEP) pathway that is utilized by many human pathogens, such as *P. falciparum* and *Mycobacterium tuberculosis* (**Scheme 1**). The final products of the MEP pathway are isopentenyl diphosphate (IDP) and dimethylallyl diphosphate (DMADP), two precursors for the biosynthesis of isoprenoids. In malaria parasites, the MEP pathway is located in the apicoplast, a plastid-like organelle of prokaryotic origin. Removing this organelle showed its crucial role for cell survival, but also that addition of IDP or DMADP rescues the parasites. This result demonstrated the significance of the MEP pathway in *P. falciparum* and its validity as a drug target.<sup>3</sup> Further, since humans utilize a completely different pathway for isoprenoid biosynthesis, the parasite enzymes can be targeted without causing side effects on the host.<sup>4-6</sup>

The identification of fosmidomycin as a potent inhibitor of the enzyme 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR), which is investigated in several clinical trials, nicely demonstrates the effectiveness of MEP pathway inhibitors (**Scheme 1**).<sup>7</sup> Fosmidomycin mimics the substrate of DXR, 1-deoxy-D-xylulose 5-phosphate (DOXP) and it was shown that the hydroxamate group and the phosphonate group are essential for binding.<sup>8</sup> However, the high polarity of fosmidomycin greatly limits its application and leaves little room for modifications to the original structure. Its promising inhibition profile validates the effectiveness of targeting the MEP pathway and more research is urgently needed to expand the pool of potent inhibitors. Particularly, 1-deoxy-D-xylulose-5-phosphate synthase (DXPS) attracted our attention. The rate-limiting, first enzyme in the pathway catalyzes the condensation of pyruvate and glyceraldehyde 3-phosphate (D-GAP) and concomitant decarboxylation with thiamine diphosphate (ThDP) as a cofactor. A unique advantage over the other MEP enzymes is that by targeting DXPS both the production of isoprenoid precursors and the biosynthesis of the vitamins B<sub>1</sub> and B<sub>6</sub> are inhibited effectively.<sup>9-11</sup>



**Scheme 1:** Schematic illustration of the MEP pathway, highlighting the important branch point enzyme DXPS. D-GAP = glyceraldehyde 3-phosphate, DXPS = 1-deoxy-D-xylulose-5-phosphate synthase, B<sub>1</sub> = thiamine, B<sub>6</sub> = pyridoxine, DOXP = 1-deoxy-D-xylulose 5-phosphate, DXR = 1-deoxy-D-xylulose 5-phosphate reductoisomerase, MEP = 2-C-methylerythritol 4-phosphate, IDP = isopentenyl diphosphate, DMADP = dimethylallyl diphosphate.

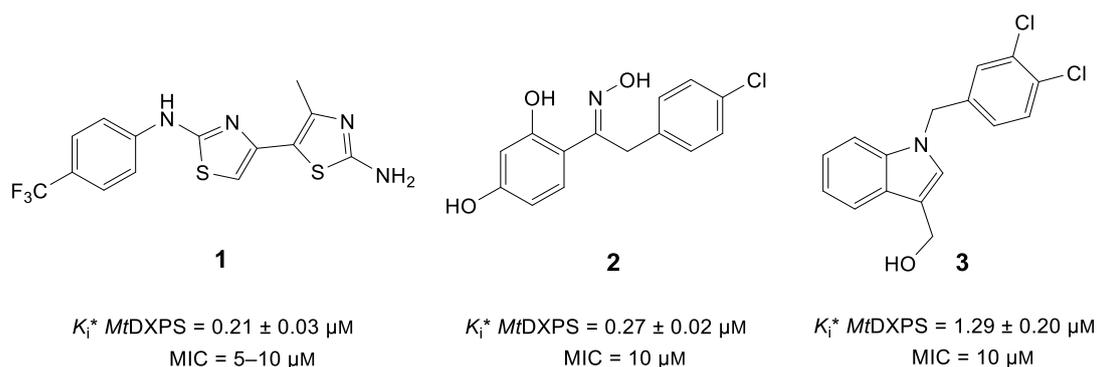
In our previous efforts to identify inhibitors of *M. tuberculosis* (*Mt*)DXPS we found three promising compound classes.<sup>12</sup> Superposition of the crystal structure of *Mt*DXPS (Protein Data Bank (PDB): 7A9H) and a homology model of *P. falciparum* (*Pf*)DXPS showed a similar structure with several loops in *Pf*DXPS that are not present in *Mt*DXPS (**Figure S 1**). However, a closer look at the active site revealed high conservation of the essential amino acids between *Mt*DXPS and *Pf*DXPS (**Figure S 2**), which suggested that active compounds against *Mt*DXPS may be effective against *Pf*DXPS.<sup>13,14</sup>

## Results and Discussion

### Structure-activity relationship (SAR) of hit compounds

In our previous work, we used ligand-based virtual screening (LBVS) as a powerful tool to identify new inhibitors based on known reference compounds for the target of interest.<sup>12</sup> LBVS relies solely on the use of descriptors of molecular structures and

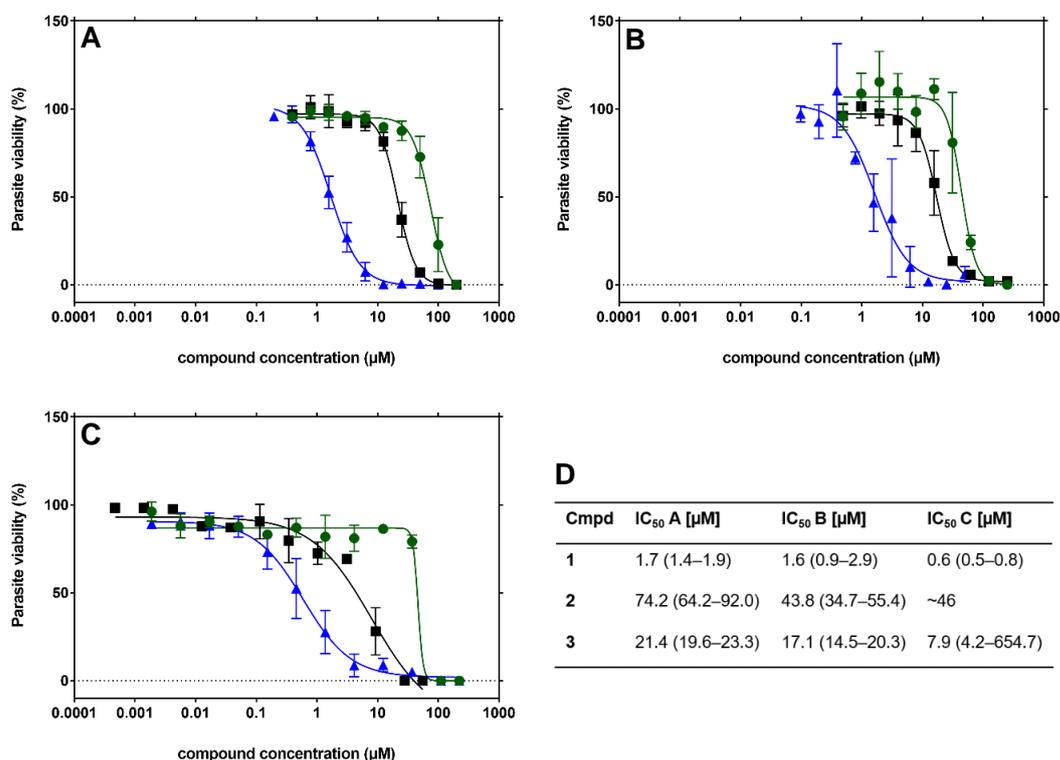
properties to compare various molecules and does not require crystallographic data.<sup>15</sup> As there were no suitable known inhibitors that could directly initialize the LBVS campaign against *Mt*DXPS, we used pseudo-inhibitors as initial ligands, with validation of key pharmacophores on the homologue *Deinococcus radiodurans* (*Dr*)DXPS, a model enzyme for *Mt*DXPS. We selected three compounds from a previous project, one ThDP analogue and two inhibitors from a *de novo* fragment design campaign.<sup>16</sup> Their 3D shape was generated and compared to all commercially available compounds from the Princeton database.<sup>17</sup> After each round of LBVS, all compounds were tested on *Dr*DXPS, as well as *Mt*DXPS. We identified three promising, structurally diverse hit classes. The most active compound in each class inhibited *Mt*DXPS in a slow-, tight-binding pattern, with submicromolar Morrison inhibition constants ( $K_i^*$ ) between 0.2–1.3  $\mu\text{M}$  and showed promising minimum inhibitory concentrations (MICs) of 5–10  $\mu\text{M}$  against *M. tuberculosis* (**Figure 1**).



**Figure 1:** Ligand-based virtual screening hits (**1**, **2**, **3**) were tested against *M. tuberculosis* DXPS. Minimum inhibitory concentrations (MICs) were determined against the *M. tuberculosis* H37Rv strain.

Despite ongoing efforts, we have no *Pf*DXPS enzyme for on-target testing available and therefore, the antimalarial activity was evaluated *ex vivo* against cultured blood-stage *P. falciparum* 3D7. Several compounds were tested in three different laboratories on three continents under varying experimental conditions, but in all cases the data were similar (Methods I–III). Compound **1** shows the most notable variation with a half-maximal inhibitory concentration ( $\text{IC}_{50}$ ) between 0.6  $\mu\text{M}$  and 1.7  $\mu\text{M}$  (**Figure 1**). Compounds **2** and **3** vary between 44–74  $\mu\text{M}$  and 8–21  $\mu\text{M}$ , respectively, which gives us great confidence in our results. All compounds were additionally tested on NF54 (Method II), a chloroquine-sensitive strain and on Dd2 (Method III), a

chloroquine-resistant strain (**Table S 5**). The differences in inhibition were small, indicating a different mode of action than chloroquine for all three compound classes, suitable for treatment of chloroquine-resistant strains. To the best of our knowledge, compounds **1–3** represent promising hits as novel antimalarials.



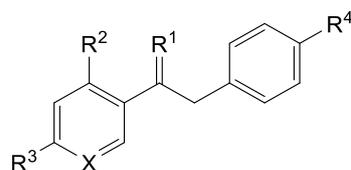
**Figure 2:** *Ex vivo* antiplasmodial activity of DXPS compounds against *P. falciparum* 3D7 from three different laboratories and assays; following (A) Method I, (B) Method II and (C) Method III. IC<sub>50</sub> curves of **1** (blue triangles), **2** (green circles) and **3** (black squares). All data was averaged from two to four independent experiments conducted in duplicate or triplicate and is shown including SD (error bars). (D) For IC<sub>50</sub> determination, data were analyzed using nonlinear regression of the log-dose-response curves and interpolated from the sigmoidal curve. 95% CI is displayed as error measure.

We synthesized several derivatives of all three classes to see if we could improve potency and achieve favorable cytotoxicity and metabolic stability properties. Here, we only discuss the measurements on 3D7 that were performed with all compounds in more detail (using Method III).

### Oximes

In derivatives for compound **2**, we maintained the two aromatic rings, connected by a two-carbon linker, but changed the substituents on the rings as well as the oxime functionality (**Table 1**). To improve solubility, we replaced the chlorine with an amino group (compounds **4**, **5** and **6**). While these modifications improve solubility two-fold to  $\sim 200 \mu\text{M}$  (**Table S 1**) for compounds **4** and **5**, the activity for all three compounds is lost. Replacing the Western aromatic ring with a pyridine neither improves the activity for the *E*- nor the *Z*-isomer (compounds **7E**, **7Z**, **8E** and **8Z**). Methylation of one of the hydroxyl-groups (compound **9**) increases the activity two-fold ( $46.8 \mu\text{M}$ ). Compound **10**, where in comparison to **2** a  $\text{NH}_2$ - is replacing a hydroxyl-group, is the most active oxime against *P. falciparum* with an  $\text{IC}_{50}$  value of  $38.2 \pm 2.2 \mu\text{M}$ . Although, we increased the activity two-fold with compound **10**, replacing the oxime moiety with an imine (**11**) or a hydrazone (**12**) improves activity. Replacing the oxime with an alcohol group (**13**) leads to a ten-fold increase ( $\text{IC}_{50} = 9.9 \pm 1.7 \mu\text{M}$ ) in comparison to the parent compound **2**. In our investigations into the mode of action of the oxime class on *MtDXPS* we identified a hydrogen bond between the hydroxyl-group of the oxime and a histidine in the active site as the critical functionality for activity. The change to the hydroxyl-group seems to improve this interaction due to the shorter chain lacking the nitrogen atom.

**Table 1:** Inhibition data for all oxime derivatives.  $\text{IC}_{50}$  measured against *P. falciparum* 3D7 (see supplementary information Method III). Original hit is compound **2** and best derivative **13** (in bold).



ID	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	R <sup>4</sup>	X	IC <sub>50</sub> [ $\mu\text{M}$ ]
<b>4</b>	NOH	OMe	OH	NH <sub>2</sub>	C	>111
<b>5</b>	NOH	OH	OH	NH <sub>2</sub>	C	>111
<b>6</b>	NOH	OMe	OMe	NH <sub>2</sub>	C	>111
<b>7Z</b>	NOH	-	OMe	Cl	N	>111
<b>8E</b>	NOH	-	OH	Cl	N	>111

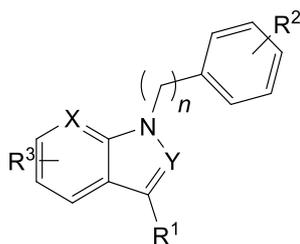
<b>8Z</b>	NOH	-	OH	Cl	N	>28
<b>2</b>	<b>NOH</b>	<b>OH</b>	<b>OH</b>	<b>Cl</b>	<b>C</b>	<b>93.8 ± 8.0</b>
<b>9</b>	NOH	OH	OMe	Cl	C	46.8 ± 15.0
<b>7E</b>	NOH	-	OMe	Cl	N	43.6 ± 9.5
<b>10</b>	NOH	OH	NH <sub>2</sub>	Cl	C	38.2 ± 11.2
<b>11</b>	NOMe	OH	OH	Cl	C	28.0 ± 8.8
<b>12</b>	NNH <sub>2</sub>	OH	OH	Cl	C	16.0 ± 0.2
<b>13</b>	<b>OH</b>	<b>OH</b>	<b>OH</b>	<b>Cl</b>	<b>C</b>	<b>9.9 ± 1.7</b>

### Indoles

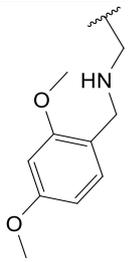
Giving the promising activity of indole **3**, we additionally explored different substitution patterns to gain insight into the SARs. The most prominent difference is observed after removal of one methylene group. Replacing the C7-carbon in the indole with a nitrogen (**14**) leads to a loss of activity. The same can be observed when attaching the phenyl ring directly to the nitrogen of the indole (compound **15**). Also, a methoxy substituent in this position is not tolerated (**16**) and a nitrile substituent in position six (**17**) results in a two-fold decrease in comparison to the original hit **3**. Replacing the core indole with an indazole (**18**) or inserting a methylene group in the R<sup>1</sup>-residue (**19**) does not result in significant changes. Removing one of the chloro-substituents gives a first interesting difference. If only the 3-chloro substituent is removed (**20**), the activity stays the same, but removal of the 4-chloro substituent (**24**) leads to a two-fold increase in activity. Interestingly, the complete removal or only removal of the hydroxyl group of the R<sup>1</sup>-substituent leads to IC<sub>50</sub> values of ~8 μM (**29**) and ~10 μM (**27**), respectively. We have shown previously that this group is essential for binding to MtDXPS, which suggests a different mode of inhibition in *P. falciparum*.<sup>12</sup> Mostly, substituents on the indole core are tolerated. While a 7-methoxy substituent is not tolerated as mentioned before, the electron-withdrawing 7-chloro group increases the activity to 13 μM (**23**). A methoxy substituent in position four is favorable (**32**), but moving it to position five (**21**) or removing one chlorine of the R<sup>2</sup>-group (**22**) reduces the activity again. A 5-nitro-substituent (**30**) or 5-bromo- (**31**) results in IC<sub>50</sub> values below 10 μM, as do a 6-fluoro (**28**) and 5-fluoro (**33**) substituent; only a fluorine in the four-position affords an activity close to the nanomolar range (**34**). Fine-tuning the compound by removing one (**25**) or two (**26**) of the R<sup>2</sup>-chlorines does not improve the activity but lowers it to ~12 μM. Since many compounds show an activity between 2 and 20 μM, these subtle changes

do not have a substantial impact. To make a more significant change, we tested a bulky substituent in the R<sup>1</sup>-position (which is essential for binding to *Mt*DXPS but not for the *P. falciparum* activity as shown earlier) and determined an IC<sub>50</sub> value of 800 ± 200 nM for compound **35**.

**Table 2:** Inhibition data for all indole derivatives. IC<sub>50</sub> values measured against *P. falciparum* 3D7 (see supplementary information Method III). Original hit is compound **3** and best derivative **35** (in bold).



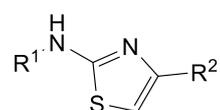
ID	R <sup>1</sup>	R <sup>2</sup>	R <sup>3</sup>	n X Y	IC <sub>50</sub> [μM]
14	-CH <sub>2</sub> OH	3,4-di-Cl	-	n = 1 X = N Y = C	> 111
15	-CH <sub>2</sub> OH	3,4-di-Cl	-	n = 0 X, Y = C	> 55.5
16	-CH <sub>2</sub> OH	3,4-di-Cl	7-OMe	n = 1 X, Y = C	> 55.5
17	-CH <sub>2</sub> OH	3,4-di-Cl	6-CN	n = 1 X, Y = C	43.8 ± 19.2
18	-CH <sub>2</sub> OH	3,4-di-Cl	-	n = 1 X = C Y = N	33.9 ± 0.2
19	-CH <sub>2</sub> CH <sub>2</sub> OH	3,4-di-Cl	-	n = 1 X, Y = C	28.9 ± 1.2
20	-CH <sub>2</sub> OH	4-Cl	-	n = 1 X, Y = C	27.7 ± 13.6
<b>3</b>	<b>-CH<sub>2</sub>OH</b>	<b>3,4-di-Cl</b>	-	<b>n = 1</b> <b>X, Y = C</b>	<b>23.6 ± 6.4</b>
21	-CH <sub>2</sub> OH	3,4-di-Cl	5-OMe	n = 1	18.5 ± 1.1

				X, Y = C	
<b>22</b>	-CH <sub>2</sub> OH	3-Cl	4-OMe	<i>n</i> = 1 X, Y = C	15.6 ± 3.3
<b>23</b>	-CH <sub>2</sub> OH	3,4-di-Cl	7-Cl	<i>n</i> = 1 X, Y = C	13.8 ± 0.9
<b>24</b>	-CH <sub>2</sub> OH	3-Cl	-	<i>n</i> = 1 X, Y = C	12.2 ± 2.8
<b>25</b>	-CH <sub>2</sub> OH	3-Cl	4-F	<i>n</i> = 1 X, Y = C	12.0 ± 1.0
<b>26</b>	-CH <sub>2</sub> OH	-	4-F	<i>n</i> = 1 X, Y = C	11.9 ± 1.5
<b>27</b>	-CH <sub>3</sub>	3,4-di-Cl	-	<i>n</i> = 1 X, Y = C	10.2 ± 3.4
<b>28</b>	-CH <sub>2</sub> OH	3,4-di-Cl	6-F	<i>n</i> = 1 X, Y = C	8.9 ± 2.4
<b>29</b>	-	3,4-di-Cl	-	<i>n</i> = 1 X, Y = C	8.1 ± 0.1
<b>30</b>	-CH <sub>2</sub> OH	3,4-di-Cl	5-NO <sub>2</sub>	<i>n</i> = 1 X, Y = C	7.2 ± 0.1
<b>31</b>	-CH <sub>2</sub> OH	-	5-Br	<i>n</i> = 1 X, Y = C	6.1 ± 1.3
<b>32</b>	-CH <sub>2</sub> OH	3,4-di-Cl	4-OMe	<i>n</i> = 1 X, Y = C	5.5 ± 1.5
<b>33</b>	-CH <sub>2</sub> OH	3,4-di-Cl	5-F	<i>n</i> = 1 X, Y = C	5.3 ± 1.3
<b>34</b>	-CH <sub>2</sub> OH	3,4-di-Cl	4-F	<i>n</i> = 1 X, Y = C	2.4 ± 0.2
<b>35</b>		<b>3,4-di-Cl</b>	-	<i>n</i> = 1 <b>X, Y = C</b>	<b>0.8 ± 0.2</b>

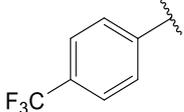
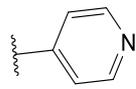
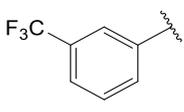
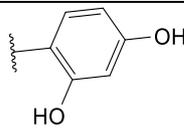
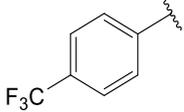
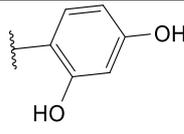
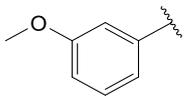
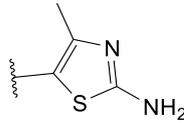
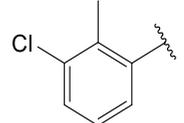
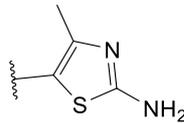
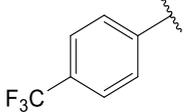
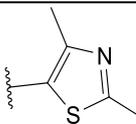
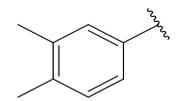
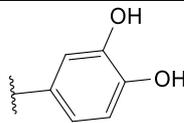
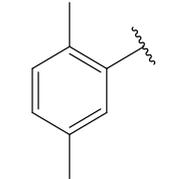
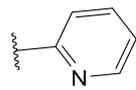
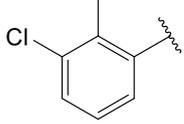
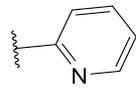
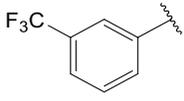
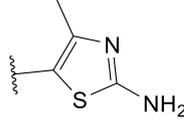
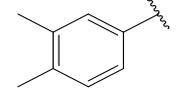
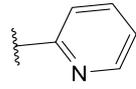
### Aminothiazoles

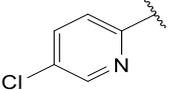
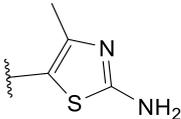
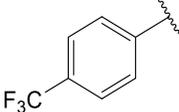
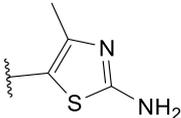
For the most promising class, the aminothiazoles, we investigated many derivatives with substantial changes on both ends of the molecule, while keeping the middle aminothiazole motif intact. We did not observe a severe reduction in activity, which indicates that the central core is essential. Changing the position of the CF<sub>3</sub>-group on the left side of the original molecule **1** results in a lower activity for 3-CF<sub>3</sub>-aminothiazole **46** (2.5 μM) and 2-CF<sub>3</sub>-aminothiazole **36** (43.0 μM). Modification of the right part of the molecule to a 2-pyridyl ring, affords single-digit micromolar activities when the left part of the molecule is either a phenyl ring with 2,5-dimethyl- (**44**), a 3,4-dimethyl- (**47**) or a 3-chloro-2-methyl- substitution (**45**). A 4-pyridyl ring does not show the same trend and the activity drops to 34 μM (**37**). A 2,4-dihydroxyphenyl ring (**38, 39**) on the right side of the molecules is not tolerated. Utilizing a 3,4-dihydroxyphenyl on the right side and a 3,4-dimethylphenyl on the left side, however, results in an IC<sub>50</sub> value of 4.4 μM (**43**). When keeping the second aminothiazole ring on the right side, modifications on the left ring with 3-methoxy- (**40**) or 3-chloro-2-methyl-substituents (**41**) do not improve the activity. Simply replacing the NH<sub>2</sub>-group on the right side with a methyl-group the activity drops ten-fold to ~10 μM (**42**), which shows the importance of this group. Another promising compound, **48** with a 4-chloro-pyridine on the right side of the molecule, has a similar potency as **1** (IC<sub>50</sub> = 1.0 μM). Overall, the original hit compound **1** was the most active with an IC<sub>50</sub> of 600 nM.

**Table 3:** Inhibition data for all aminothiazole derivatives. IC<sub>50</sub> values measured against *P. falciparum* 3D7 (see supplementary information Method III). Original hit is compound **1** (in bold).



ID	R <sup>1</sup>	R <sup>2</sup>	IC <sub>50</sub> [μM]
<b>36</b>			43.0 ± 2.4

37			$34.2 \pm 1.8$
38			$23.0 \pm 0.7$
39			$22.1 \pm 5.6$
40			$17.9 \pm 0.4$
41			$15.9 \pm 1.8$
42			$9.5 \pm 1.1$
43			$4.4 \pm 1.0$
44			$3.7 \pm 1.6$
45			$2.8 \pm 0.1$
46			$2.5 \pm 0.1$
47			$1.5 \pm 0.8$

<b>48</b>			$1.0 \pm 0.1$
<b>1</b>			$0.6 \pm 0.2$

### Cytotoxicity and metabolic stability of selected compounds

The promising *ex vivo* results motivated us to thoroughly investigate the properties of all three hit classes further. We tested them for their respective cytotoxicity and metabolic stability (**Table 4**). Indole **3** exhibits an  $IC_{50}$  at  $23.6 \pm 6.4 \mu\text{M}$  against 3D7. Evaluating its cytotoxicity against human hepatocytes (Hep G2) results in a lower  $IC_{50}$  value of  $0.8 \pm 0.2 \mu\text{M}$ , suggesting the compound lacks specificity against asexual parasites. The two most active indole derivatives on *P. falciparum*, **34** and **35**, are equally as potent against Hep G2 cells ( $0.79 \pm 0.37 \mu\text{M}$  and 90% inhibition at  $10 \mu\text{M}$ , respectively). When investigating the metabolic stability, we found that indole **3** has a half-life of only 10 min in human liver S9 fractions. More synthetic work is needed to balance these properties. While oxime **2** is active against *P. falciparum* ( $IC_{50} = 93.8 \pm 8.0 \mu\text{M}$ ), it is similarly potent against Hep G2 ( $>50 \mu\text{M}$ ). The best oxime hit **10** is two-fold more active against *P. falciparum* but equally inhibits Hep G2 cells. After removing the oxime functionality and replacing it with a hydroxyl-group, we found that compound **13** is not only the most active oxime derivative ( $IC_{50} = 9.9 \pm 1.7 \mu\text{M}$ ) but it also does not inhibit the growth of Hep G2 cells at  $100 \mu\text{M}$ . With the synthetic modifications we improved the metabolic stability from 28 min for compound **2** to 98 min for alcohol **13**. We therefore excluded the indole and oxime class from further investigations, but compound **13** emerged as a new potent inhibitor class for *P. falciparum* with an improved cytotoxicity and metabolic stability profile.

**Table 4:** Summary of *P. falciparum* activity, cytotoxicity (activity against Hep G2 in  $\mu\text{M}$  or %inhibition) and metabolic stability in human liver S9 fraction (half-life in min) of selected compounds from all three hit classes. n.d. = not determined.

Compound	IC <sub>50</sub> <i>P. falciparum</i> 3D7 ( $\mu\text{M}$ )	IC <sub>50</sub> Hep G2	t <sub>1/2</sub> (min)
<b>1</b>	0.6 $\pm$ 0.2	>50 $\mu\text{M}$	>240
<b>2</b>	93.8 $\pm$ 8.0	>50 $\mu\text{M}$	28.5 $\pm$ 0.7
<b>10</b>	38.2 $\pm$ 11.2	50–100 $\mu\text{M}$	55.8 $\pm$ 0.4
<b>13</b>	9.9 $\pm$ 1.7	> 100 $\mu\text{M}$	97.8 $\pm$ 9.9
<b>3</b>	23.6 $\pm$ 6.4	0.8 $\pm$ 0.2 $\mu\text{M}$	10.0 $\pm$ 0.4
<b>34</b>	2.4 $\pm$ 0.2	0.79 $\pm$ 0.37 $\mu\text{M}$	n.d.
<b>35</b>	0.8 $\pm$ 0.2	at 10 $\mu\text{M}$ 90 $\pm$ 2% inhibition	n.d.

Compound **1** exhibits a promising IC<sub>50</sub> value (IC<sub>50</sub> = 0.6  $\pm$  0.2  $\mu\text{M}$ ) in addition to a weaker inhibition of Hep G2 with 54.0  $\pm$  2.4  $\mu\text{M}$ . It is also the most stable compound in human liver S9 fraction (>240 min) that we tested. As such, aminothiazole **1** is a potent hit and inhibitor of *P. falciparum* growth *ex vivo*. It will be further optimized against *P. falciparum* and *M. tuberculosis* in parallel.

### Target validation

We have strong reason to believe that DXPS is the main target of all hit compounds in *M. tuberculosis*.<sup>12</sup> To confirm *PfDXPS* as a target in *P. falciparum*, we chose to further investigate compound **1**, since it showed the most promising activity. One well established method is a rescue assay with IDP, the product of the MEP pathway. It has been shown that the addition of IDP to blood-stage *Plasmodium* spp. rescued parasite survival after treatment with fosmidomycin (FSM), which inhibits the second enzyme in the MEP pathway, as well as survival of apicoplast-minus *Plasmodium* spp.<sup>3</sup> We determined the growth inhibition of blood-stage *P. falciparum* with selected compounds, including the most active compound **1**, in the presence and absence of 125  $\mu\text{M}$  IDP. This assay has never been performed with a DXPS inhibitor, but we expected a rescue effect if DXPS was the main target of compound **1**.

While, as expected, the antiparasitic activity of FSM was rescued by IDP addition, compound **1** still inhibited *P. falciparum* growth under IDP supplementation. Similar results were observed for two other active aminothiazole derivatives, **47** and **48** and

the indole derivative **34** (**Figure S 3**, **Figure S 4** and **Table S 2**). Together, these data suggest that our compounds inhibit additional pathways within *P. falciparum* apart from the MEP pathway.

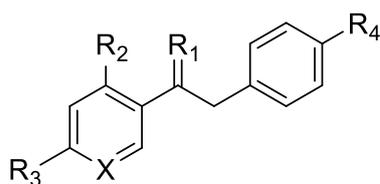
In order to elucidate whether **1** interacts with DXPS, we profiled the cellular concentration of MEP pathway intermediates by LC-MS. In this experiment, we analyzed the concentration changes of MEP pathway metabolites in the presence and absence of **1** in *P. falciparum* and *E. coli*. If a compound inhibits the MEP pathway, a reduced concentration of all metabolites downstream of the inhibited enzyme is expected. In *E. coli*, however, all downstream metabolites of DXPS are increased upon inhibitor treatment, while the pyruvate concentration drops (**Figure S 5**). This behavior has not previously been reported, but it suggests an influence on the pathway that requires further investigation. In *P. falciparum*, we observe no difference in the metabolite concentrations, but a reduction in pyruvate levels (**Figure S 6**, **Table S 3**). This decrease is consistent with a reduction in tricarboxylic acid (TCA) cycle metabolites (**Figure S 7**, **Table S 3**), but since pyruvate is tied to many other metabolic pathways, we could not determine the reason for the decrease. Although these results did not confirm our hypothesis, they indicated that a different mode of action in *P. falciparum* and *E. coli* is responsible for the anti-infective activities of **1** *ex vivo*.

To address the ambiguous result from the previous assays, we screened the three original LBVS hits against transgenic parasites overexpressing thiamine pyrophosphokinase (*Pf*TPK) and *Pf*DXPS (DOXP cell line) and compared the results to the MOCK cell line that contained only the transfected vector backbone. All three hit compounds are ThDP-competitive inhibitors as we have shown previously.<sup>12</sup> Therefore, it is possible that the compounds indiscriminately bind to *Pf*TPK as well as *Pf*DXPS. In case of inhibition, we would expect a higher IC<sub>50</sub> value for the overexpressing cell lines than for the MOCK cell line, but a statistical difference (**Table S 4**) was not detected for either *Pf*DXPS (**Figure S 8**) or *Pf*TPK (**Figure S 9**). While these results leave the compound's main target unclear, we can exclude *Pf*TPK as an off-target.

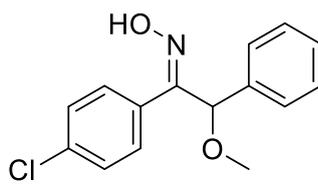
### PolyPharmacology Browser (PPB) analysis to identify alternative targets

Despite extensive efforts to experimentally verify DXPS as the molecular target, the results were not conclusive. To extend our knowledge of other possible targets and off-target proteins, we turned to an *in silico* approach using the PPB to search for other potential target enzymes. The PPB search engine employs a similarity-based approach following the idea that similar compounds should target the same proteins.<sup>18,19</sup> Several methods are used in parallel to calculate molecular fingerprints of a query compound, which are then used to search the open-access bioactivity database ChEMBL.<sup>20</sup> The search output contains compounds similar to the query molecule, associated with their biological activity. We manually analyzed the results, looking for alternative targets of our compounds.

#### Oximes



Oxime core structure



CHEMBL581471

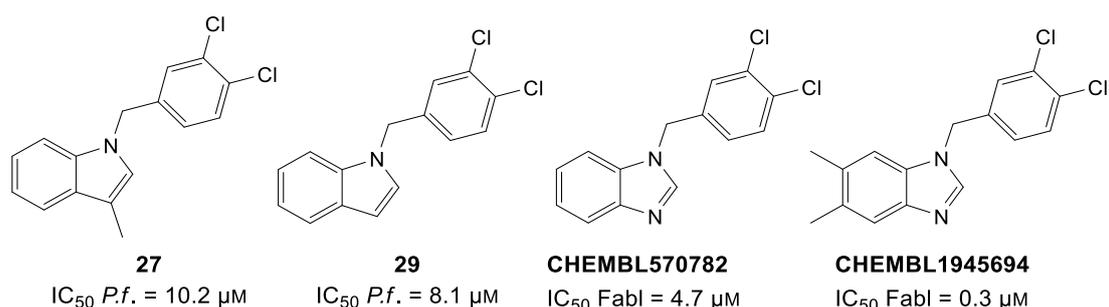
**Figure 3:** The oxime scaffold and the related compound **CHEMBL581471**. For elucidation of X and the R<sub>1</sub>-R<sub>4</sub> groups, please see **Table 1**.

Similar to our core structure, we found the related oxime **CHEMBL581471** (**Figure 3**). This compound was identified in 2008 by Novartis in a high-throughput screening (HTS) and showed activity against *P. falciparum* in a liver-stage assay. The oxime was a hit in the initial HTS, but could not be confirmed and was therefore not evaluated further. The reported EC<sub>50</sub> values are >0.912 μM in 3D7 and >1.607 μM in W2. The compound also showed good cytotoxicity behavior with >10 μM Huh7 (human hepatoma cells) inhibition, about 10 fold higher than the EC<sub>50</sub>.<sup>21</sup>

Although no molecular target is assigned to compound **CHEMBL581471**, the finding gives us additional insights for the design of further derivatives. It seems that the CH<sub>2</sub> of the linker might be a possible growth vector. Together with the replacement of the oxime functionality with an amino or hydroxyl-group, as suggested by our SAR study,

the initial oxime hit offers the potential to be developed into a different class in the future.

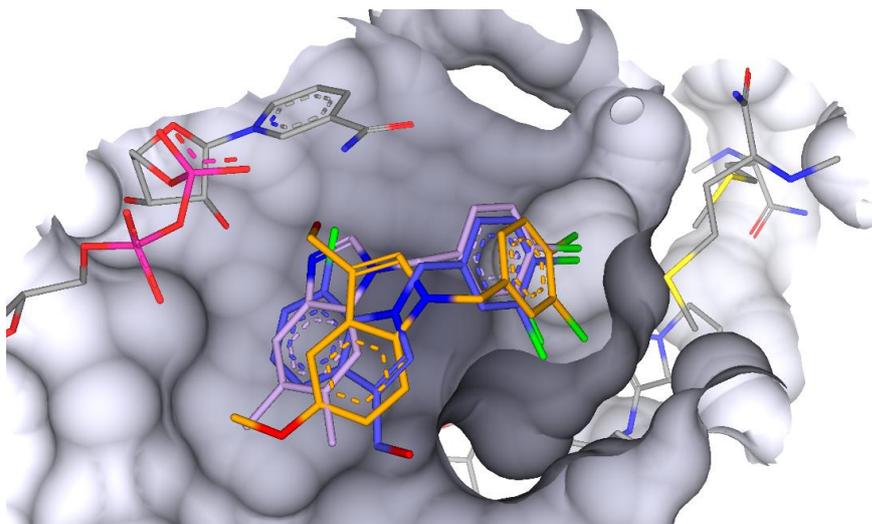
### Indoles



**Figure 4:** Compounds **29** and **31** (IC<sub>50</sub> cell-based using *P. falciparum* 3D7, Method III) shown with the related compounds inhibiting the Enoyl-[acyl-carrier-protein] reductase (FabI) (IC<sub>50</sub> based on enzymatic assay).

Our search for compounds **27** and **29** yielded the similar compounds **CHEMBL1945694** and **CHEMBL570782** (Figure 4), which are reported to inhibit the Enoyl-[acyl-carrier-protein] reductase (ENR or FabI) from *Francisella tularensis*.<sup>22</sup> The two compounds we found belong to a larger class of benzimidazoles, first identified to inhibit *F. tularensis* FabI in 2012 using a LBVS approach by Johnson and coworkers.<sup>23</sup> In follow-up studies, they reported derivatives with improved activity against the FabI enzyme of up to 14 nM and determined the crystal structure of FabI from *F. tularensis* in complex with **CHEMBL1945694**.<sup>24,25</sup>

To compare our indole-class with the benzimidazoles, we docked all compounds into the reported *F. tularensis* structure with the PDB-code 3uic and, as several protein structures of FabI from *P. falciparum* are available, also to this homologue. The *P. falciparum* structure is mainly solved in complex with triclosan and its analogues and some induced fit of the binding site is known.<sup>26</sup> Because of the induced fit, we selected the structure with the PDB-code 2op1, hosting a larger ligand of similar size as the benzimidazoles.



**Figure 5:** Docking poses of our compounds into *PfFabI* (PDB: 3uic) in complex with the co-crystallized inhibitor. The co-crystallized ligand is shown for comparison and colored purple, compound **23** in blue and compound **22** in orange. The conserved interaction of the N3 atom from the benzimidazole ring system with Tyr156 is either compensated by a hydroxyl substituent at the 3-position or substituted by flipping of the indole moiety and interaction with a Cl-substituent in 7-position.

The compounds of our indole class dock to the *F. tularensis* structure in a pose similar to the co-crystallized **CHEMBL1945694** benzimidazole and are calculated to have a very high affinity in the pico- to nanomolar range (**Table S 6**). A superposition of the best pose is shown in **Figure 5**. Docking to the FabI enzyme from *P. falciparum* also predicted nanomolar binding affinities. For both homologues, the calculated binding affinity of the best indoles presented in this study is similar to **CHEMBL1945694**, with **18** binding even slightly better (**Table S 7**).

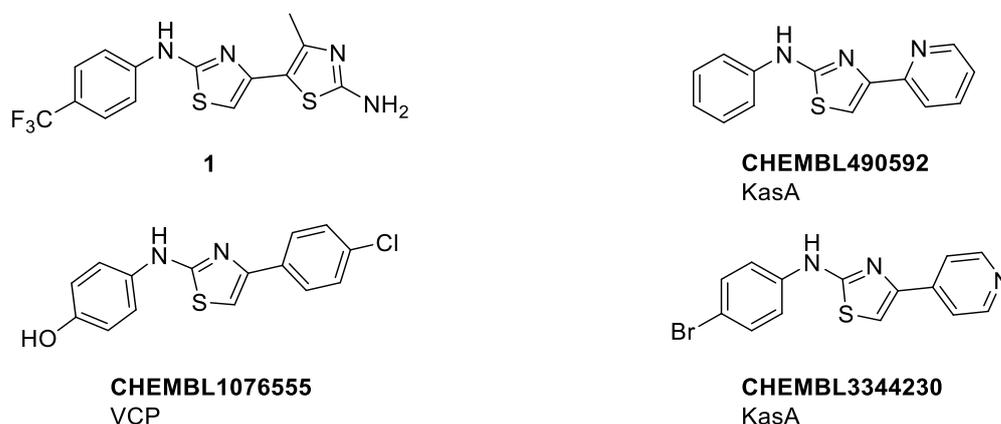
Our aminothiazole class showed binding affinities in the nano- to micromolar range, while binding of the oximes was calculated to be in the micromolar range (**Table S 8** to **Table S 11**).

The FabI enzyme is part of the fatty acid biosynthesis pathway-II (FAS-II) found in a variety of microorganisms, including *P. falciparum*. The mammalian counterpart FAS-I consists of one protein complex with low homology to the FAS-II enzymes.<sup>27–29</sup> The FabI enzyme catalyzes the final reduction step, supplying fatty acids for cell-wall biosynthesis. Enzymes of the pathway are used for the development of new antimicrobial agents.<sup>30–33</sup> It was, however, shown that the blood stage of *P. falciparum*

does not require the FAS-II pathway for proliferation.<sup>34–36</sup> Therefore, inhibition of FabI cannot explain the activity we observed in the blood stage assays. However, it could be beneficial for a new anti-malarial drug to inhibit FabI as a second target, as FabI is essential for liver stage proliferation and therefore transmission.<sup>36</sup>

Taken together, the structural similarity of the compounds (**Figure 4**) hints to FabI being an additional target of our hit classes. In particular, the indole-class is very likely to bind to FabI, as there are co-crystal structures of related compounds. Inhibition of FabI does not explain the observed *ex vivo* effects in our assays, as this pathway is only important during the liver stage of *P. falciparum* growth.<sup>28</sup>

### Aminothiazoles

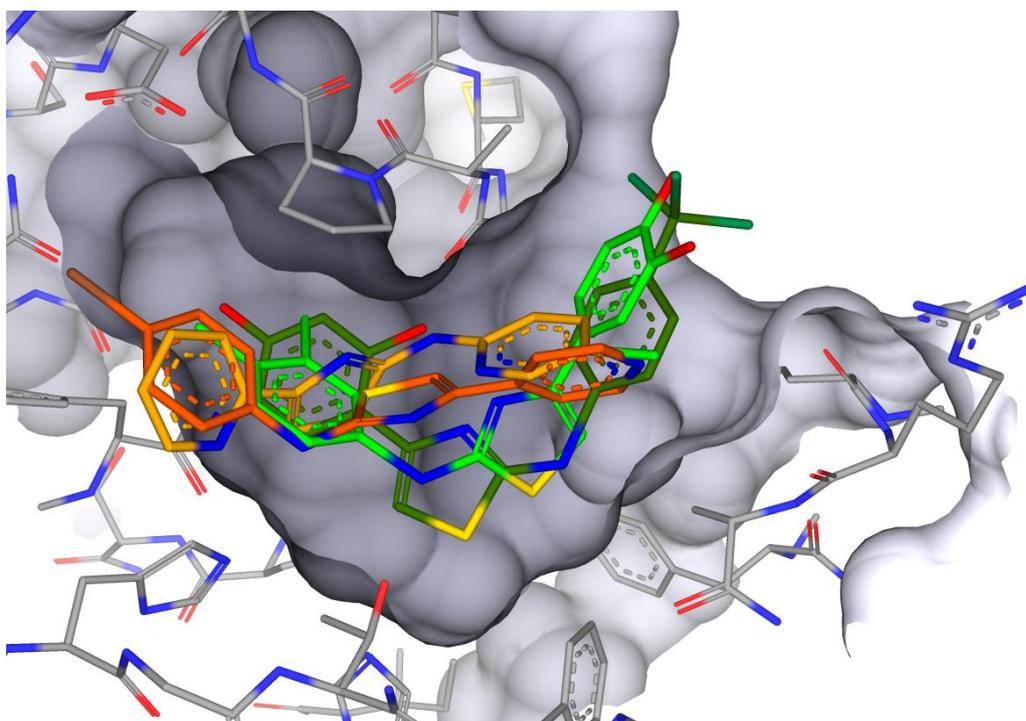


**Figure 6:** Aminothiazole hit compound **1** and similar compounds we found during the PPB search, with their reported target. KasA =  $\beta$ -ketoacyl ACP synthase; VCP = Valosin-containing protein.

For all 15 queried aminothiazole molecules we could find related compounds that are reported to target *P. falciparum*. However, the majority of the reported activities are based on cell-based assays without an assigned molecular target. Only the compounds with a known target are further analyzed below.

With the aminothiazole **CHEMBL490592** (**Figure 6**, **Figure 5**) we found a compound similar to our herein described aminothiazole class. It is part of a different group of aminothiazole-based inhibitors with activity against *M. tuberculosis*.<sup>37,12</sup> In a previous study, we could also observe anti-tubercular activity for our group of aminothiazoles. In 2014, P. Makam and T. Kannan reported that a possible target of 2-aminothiazoles

in *M. tuberculosis* is the enzyme  $\beta$ -ketoacyl ACP synthase (KasA).<sup>38</sup> The KasA protein is part of the FAS-II pathway, and involved in the biosynthesis of mycolic acid, an essential cell wall component. It is also present and active in *P. falciparum*, but similar to FabI it has been found to be dispensable for the blood stage of *P. falciparum* (please see section Indoles).<sup>34,39,40</sup> Since the protein structure of *P. falciparum* KasA has not been determined, we docked our inhibitors to the KasA homologue from *M. tuberculosis* (PDB: 2wgd), which was used to develop the aminothiazole **CHEMBL3344230**.<sup>38</sup> This enables us to directly compare our compounds with the previously published compound **CHEMBL3344230**.<sup>38</sup> A superposition of the docked compounds is shown in **Figure 7**. The calculated binding poses and affinities of our hits are in the nano- to micromolar range, similar to the reported compound (**Table S 14**), suggesting that KasA may be a target of our aminothiazole hit class.



**Figure 7:** Superposition of compounds **39** (dark green), **43** (green), **CHEMBL3344230** (red) and **CHEMBL546826** (orange) docked to the KasA enzyme from *M. tuberculosis* (PDB: 2wgd). Although not exactly in the same pose, they all dock well into the pocket.

In addition, a series of aminothiazoles have been developed as antileishmanial agents, a protozoan parasite, despite no identified target protein.<sup>41</sup> Based on the structural

similarity and docking of their reported best hit **CHEMBL546826**, it seems likely that KasA may be the target, which is also present in *Leishmania* spp. (**Figure 7, Table S 12**).

We identified compound **CHEMBL1076555** (**Figure 6**), which was included in a series of cancer inhibitors targeting the Valosin-containing protein (VCP).<sup>42</sup> Furthermore, the endoplasmic reticulum assisted degradation (ERAD) is also gaining attention as a target against protozoan pathogens, as they have only a very small network for their protein quality control system.<sup>43</sup> One of these members is VCP, which might be targeted by the reported aminothiazoles.<sup>44</sup> As a consequence of this pared down network of protein degradation, *P. falciparum* is highly susceptible to ERAD inhibitors, a VCP-knockout strain is not viable. Although VCP is also present in mammals (often termed p97), it was shown that selectivity towards VCP from *P. falciparum* can be achieved.<sup>43,45</sup> Docking into VCP was not performed, as the structure of the large hexameric protein is only solved for mammals and no co-crystal structure is available to identify the binding site. Testing our hits as well as the previously reported VCP inhibitors against *P. falciparum* might reveal VCP as an additional target, which could explain the inhibitory activity we observed in the blood-stage assays.<sup>42</sup>

### **Human off-targets**

During the search for antimicrobial targets, we found mammalian enzymes that are affected by compounds similar to our hits, including enzymes of the eicosanoid metabolism, the membrane protein KDR kinase and RNA polymerase II (**Table S 13**).<sup>46–48</sup> These early data are beneficial, as we now have the possibility to develop and test our hits for selectivity for the bacterial over the human targets. Further, if we can confirm activity against human enzymes, it is possible that some derivatives could be re-purposed. For example, we report three compounds (**3, 15, 20**) that are identical to derivatives of the Oncrasin-1 inhibitor of the RNA polymerase II, which is discussed as a new class of anti-cancer drugs.<sup>48</sup>

## **Conclusions**

We showed that our previously identified small-molecule inhibitors against *M. tuberculosis* DXPS from a LBVS campaign are also active against several *P. falciparum* strains. We successfully improved the activity of two of our three initial

hit classes. The oxime **2** was improved 2.5-fold from 93.8  $\mu\text{M}$  to 38.2  $\mu\text{M}$  by replacing one hydroxyl-group with an amino-group (compound **10**). By removing the oxime we also identified several new compounds that require further investigation. Imine **11**, hydrazone **12** or alcohol **13** are promising alternative functionalities to the oxime **2** with low cytotoxicity and improved metabolic stability in the case of compound **13**. The best hit of the indole class, compound **3** with an  $\text{IC}_{50}$  value of 23.6  $\mu\text{M}$  was improved 30-fold to 0.8  $\mu\text{M}$  by adding a bulky substituent to the hydroxyl-group (**35**). We discontinued this class due to the toxicity issues. Overall, we identified the aminothiazole **1** as a promising compound against *P. falciparum* with good activity, selectivity and excellent metabolic stability.

Using compound **1**, we investigated *PfDXPS* as a potential target. Several lines of evidence suggest compound **1** has on- and off-target effects. First, IDP failed to rescue growth after treatment. LC-MS analysis of metabolites downstream of DXPS following inhibition in both *P. falciparum* and *E. coli* was ambiguous. In *P. falciparum* we observe a reduction in pyruvate levels that is consistent with the downregulation of TCA-cycle metabolites indicating complex effects on parasite metabolism. To further investigate potential interaction with *PfDXPS* as well as *PfTPK* we used genetically modified parasites overexpressing the respective enzymes. Here, we cannot observe any significant difference between the MOCK and the overexpressing cell lines, which means DXPS is most likely not the main target of compound **1**. However, we can exclude *PfTPK* as an off-target, which suggests the compound shows no specific inhibition of thiamine-dependent proteins.

The validation of molecular targets is notoriously difficult. Using the PPB search engine, we identified three possible additional targets for our hit classes, the enzymes FabI, KasA and VCP. The independent occurrence of two enzymes from the FAS-II pathway is remarkable and offers a potential dual role of our hit compounds as drug targeting DXPS and chemoprophylaxis agent, inhibiting the FAS-II pathway and thereby the onset of parasite infection by targeting *P. falciparum* liver stage development. However, as FAS-II inhibition is not essential for proliferation in the *P. falciparum* blood stage, it cannot explain the observed asexual growth inhibition. The VCP protein could be the alternative target explaining our biological activity, but it could not be assessed further because of the lack of a protein structures from

*Plasmodium* spp. Whether our hits are dual inhibitors of DXPS and enzymes of the FAS-II pathway, will need to be determined experimentally in future studies. Antimalarial drugs on the market such as artemisinin and chloroquine also have multiple targets, which makes them highly potent, but emerging resistance increases the need for alternative treatments.<sup>49,50</sup> We showed that the discussed compounds effectively kill the chloroquine resistant *P. falciparum* strain Dd2, which makes them suitable candidates for further investigation as alternative drugs or combination partners with existing therapeutic agents.

## Experimental section

**General procedures.** NMR experiments were run on a Bruker Avance Neo 500 MHz (<sup>1</sup>H at 500.0 MHz; <sup>13</sup>C at 126.0 MHz; <sup>19</sup>F NMR at 470 MHz), equipped with a Prodigy Cryo-probe. Spectra were acquired at 298 K, using deuterated dimethylsulfoxide ((CD<sub>3</sub>)<sub>2</sub>SO, <sup>1</sup>H: 2.50 ppm, <sup>13</sup>C: 39.52 ppm), deuterated methanol (CD<sub>3</sub>OD, <sup>1</sup>H: 3.31 ppm, <sup>13</sup>C: 49.00 ppm) or deuterated chloroform (CDCl<sub>3</sub>, <sup>1</sup>H: 7.26 ppm, <sup>13</sup>C: 77.16 ppm) as solvent. Chemical shifts for <sup>1</sup>H and <sup>13</sup>C spectra were recorded in parts per million (ppm) using the residual non-deuterated solvent as the internal standard. Coupling constants (*J*) are given in Hertz (Hz). Data are reported as follows: Chemical shift, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad and combinations of these) coupling constants and integration. Mass spectrometry was performed on a SpectraSystems-MSQ LCMS system (Thermo Fisher, Dreieich, Germany). Purification of the final products was performed using preparative HPLC (Dionex UltiMate 3000 UHPLC+ focused, Thermo Scientific) on a reversed-phase column (C18 column, 5 μm, Macherey-Nagel, Germany) or flash chromatography was performed using the automated flash chromatography system CombiFlash Rf+ (Teledyne Isco, Lincoln, NE, USA) equipped with RediSepRf silica columns (Axel Semrau, Sprockhövel Germany). High-resolution mass (HRMS) of final products was determined by LCMS/MS using Thermo Scientific Q Exactive Focus Orbitrap LC-MS/MS system. The purity of the final compounds was determined using Dionex Ultimate 3000 HPLC system (Thermo Fisher Scientific). Chromatographic separation was performed on an EC 150/2 Nucleodur C18 Pyramid (3 μm particle size) analytical column (Macherey-Nagel). The mobile phase consisted of solvent A (water containing 0.1% formic acid) and solvent B (acetonitrile containing 0.1% formic acid)

with a flow rate of 0.25 mL/min. All final compounds had a purity >95%. Yields refer to analytically pure compounds and have not been optimized. All chemicals were purchased at SigmaAldrich or comparable commercial suppliers and used without further purification.

**General remarks about the analysis:** Full characterization is provided for final compounds that have not been published before or have been published in different NMR-solvents. The identity of intermediates was determined by <sup>1</sup>H-NMR, <sup>13</sup>C-NMR and <sup>19</sup>F-NMR if applicable. The <sup>13</sup>C-NMR signals are doublets in the case of six carbons in the F-substituted aromatic ring in **28**, seven carbons including -CCH<sub>2</sub>OH for **25**, eight carbons including -CCH<sub>2</sub>OH for **34** and **26**. Many indole HR-ESI-MS measurements give [M-18], due to fragmentation during ionization. The identity of *E/Z* isomers was determined by 1D-NOESY NMR for compounds **7E** and **7Z**. All other compounds with isomers were assigned based on carbon shifts. *E*-Oxime C7-shift (ppm) > *Z*-oxime C7-shift (ppm). All compounds are >95% pure by HPLC analysis.

**Synthesis and characterization of oximes.** The synthesis and characterization of **2**, **9**, **10**, and **12** were described previously by us.<sup>12</sup> 2-(4-((*tert*-Butoxycarbonyl)amino)phenyl)acetic acid (**49**) was synthesized following a literature procedure and all data were consistent with the reported values.<sup>51</sup>

**Condensation (procedure O-A).** Synthesis of the protected ketone intermediate followed a previously reported procedure.<sup>52</sup> To a Schlenk flask, methyl benzoate and phenylacetic acid (1.0 equiv.) were added and dissolved in dry dimethylformamide (DMF) under nitrogen. The yellow solution was cooled to -10 °C, and then sodium *bis*(trimethylsilyl)amide (2 M in tetrahydrofuran (THF), 4.0 equiv.) was added dropwise under stirring. After full conversion of the starting material (3–72 h) monitored by thin layer chromatography (TLC) or LC-MS, the reaction was terminated by adding saturated aqueous NH<sub>4</sub>Cl solution and concentrated *in vacuo* to remove DMF. Subsequently, the residue was extracted with ethyl acetate (3 x), and the combined organic layers were washed with saturated aqueous NaCl solution. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo*.

**Deprotection (procedure O-B).** The deprotection reaction followed a previously reported procedure.<sup>53</sup> To a solution of the condensation product from procedure **O-A**

in dry dichloromethane (9 mL) under nitrogen, boron tribromide was added (1 M in dichloromethane, 12.0 equiv.) under stirring at 25 °C. After 5 h, a saturated aqueous Na<sub>2</sub>CO<sub>3</sub> solution was added to the solution, which was extracted with dichloromethane. The organic layer was washed with water (2 x), then dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo*.

**Oxime formation (procedure O-C).** Oximation followed a previously reported procedure.<sup>54</sup> To a solution of the deprotected product in methanol, potassium acetate (3.0 equiv.) and hydroxylamine hydrochloride (1.5 equiv.) were subsequently added under stirring. The light-yellow suspension was refluxed for 2 h. Subsequently, water was added to the mixture. The organic layer was washed with saturated aqueous NaCl solution, dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo*.

***tert*-Butyl (4-(2-(2,4-dimethoxyphenyl)-2-oxoethyl)phenyl)carbamate (50).** Compound **50** was synthesized following procedure **O-A**, using **49** (650 mg, 2.59 mmol, 1.0 equiv.), methyl-2,4-dimethoxybenzoate (508 mg, 2.59 mmol, 1.0 equiv.) and sodium *bis*(trimethylsilyl)amide (10.4 mL, 10.4 mmol, 3.0 equiv.) in dry DMF (20 mL). After 72 h, no full conversion was achieved, so it was decided to terminate the reaction. Flash column chromatography (petroleum benzine/ethyl acetate 2:1 + 1% acetic acid) afforded a mixture of the product and starting material **49**. To remove the acid, the mixture was dissolved in ethyl acetate and washed with saturated aqueous NaHCO<sub>3</sub> solution (6 x). The product (106 mg, 11% yield) was obtained as a white solid. <sup>1</sup>H-NMR (500 MHz, CD<sub>3</sub>OD): δ (ppm) = 7.69 (d, *J* = 8.57 Hz, 1H), 7.29 (d, *J* = 8.37 Hz, 2H), 7.08 (d, *J* = 8.57 Hz, 2H), 6.56 (m, 2H), 4.19 (s, 2H), 3.91 (s, 3H), 3.85 (s, 3H), 1.50 (s, 9H). <sup>13</sup>C-NMR (126 MHz, CD<sub>3</sub>OD): δ (ppm) = 200.8, 166.6, 162.4, 155.4, 139.0, 133.8, 131.2, 130.9, 121.5, 119.8, 106.9, 99.2, 56.1, 56.0, 50.1, 28.70. HR-ESI-MS: calculated for C<sub>21</sub>H<sub>26</sub>NO<sub>5</sub> [*M*+H]<sup>+</sup> 372.1805, found 372.1806.

**2-(4-Aminophenyl)-1-(4-hydroxy-2-methoxyphenyl)ethan-1-one (51).** Following procedure **O-B** using **50** (90 mg, 0.24 mmol, 1.0 equiv.) and boron tribromide (1.2 mL, 1.2 mmol, 5.0 equiv.) in dry dichloromethane (2 mL), compound **51** (28 mg, 45% yield) was obtained as a white solid after flash column chromatography (CH<sub>2</sub>Cl<sub>2</sub>/CH<sub>3</sub>OH 19:1). Compound **52** was purified as a side product (4 mg, 7% yield). <sup>1</sup>H-NMR (500 MHz, CD<sub>3</sub>OD): δ (ppm) = 7.90 (d, *J* = 8.99 Hz, 1H), 7.03 (d, *J* = 8.50 Hz, 2H),

6.68 (d,  $J = 8.50$  Hz, 2H), 6.47 (dd,  $J = 2.50, 8.99$  Hz, 1H), 6.41 (d,  $J = 2.50$  Hz, 1H), 4.09 (s, 2H), 3.82 (s, 3H).  $^{13}\text{C-NMR}$  (126 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  (ppm) = 204.9, 167.7, 166.8, 147.5, 134.1, 131.0, 125.7, 116.9, 114.2, 108.3, 101.9, 56.1, 45.1. HR-ESI-MS: calculated for  $\text{C}_{15}\text{H}_{16}\text{NO}_2$   $[M+H]^+$  258.1125, found 258.1123.

**2-(4-Aminophenyl)-1-(2,4-dihydroxyphenyl)ethan-1-one (52).** To achieve full deprotection in one step, compound **50** (22 mg, 0.06 mmol, 1.0 equiv.) was heated to 110 °C in the microwave (5 min, 15 W) with pyridine hydrochloride (1 mL). The reaction was diluted with saturated aqueous  $\text{Na}_2\text{SO}_4$  solution and this aqueous solution was extracted with ethyl acetate (3 x 5 mL). The combined organic layers were dried over  $\text{Na}_2\text{SO}_4$ , filtered and concentrated *in vacuo*. The residue was co-evaporated with toluene (3 x 10 mL) to remove residual pyridine. The product (6 mg, 40% yield) was obtained as a white solid.  $^1\text{H-NMR}$  (500 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 12.67 (s, 1H), 10.60 (s, 1H), 7.90 (d,  $J = 8.89$  Hz, 1H), 6.92 (m, 2H), 6.49 (m, 2H), 6.36 (dd,  $J = 2.31, 8.89$  Hz, 1H), 6.23 (d,  $J = 2.31$  Hz, 1H), 4.95 (s, 2H), 4.02 (s, 2H).  $^{13}\text{C-NMR}$  (126 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 43.4, 102.5, 108.2, 111.9, 114.0, 121.8, 129.7, 133.7, 147.3, 164.8, 164.9, 203.1. HR-ESI-MS: calculated for  $\text{C}_{14}\text{H}_{14}\text{NO}_3$   $[M+H]^+$  244.0968, found 244.0969.

**(E)-2-(4-Aminophenyl)-1-(4-hydroxy-2-methoxyphenyl)ethan-1-one oxime (4).** Following procedure **O-C**, using compound **51** (20 mg, 0.07 mmol, 1.0 equiv.), potassium acetate (7 mg, 0.21 mmol, 3.0 equiv.) and hydroxylamine hydrochloride (7 mg, 0.1 mmol, 1.5 equiv.), the oxime (20 mg, 100% yield) was afforded as a white solid and not purified further.  $^1\text{H-NMR}$  (500 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 11.96 (s, 1H), 11.44 (s, 1H), 7.39 (d,  $J = 9.43$  Hz, 1H), 6.90 (d,  $J = 8.41$  Hz, 2H), 6.45 (m, 2H), 6.40 (m, 2H), 4.88 (s, 2H), 3.99 (s, 2H), 3.71 (s, 3H).  $^{13}\text{C-NMR}$  (126 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 160.7, 159.7, 159.4, 146.9, 129.4, 128.9, 123.4, 114.1, 111.3, 105.4, 101.5, 55.1, 28.9. HR-ESI-MS: calculated for  $\text{C}_{15}\text{H}_{17}\text{N}_2\text{O}_3$   $[M+H]^+$  273.1234, found 273.1232.

**(E)-2-(4-Aminophenyl)-1-(2,4-dihydroxyphenyl)ethan-1-one oxime (5).** Following procedure **O-C** using compound **52** (6 mg, 0.03 mmol, 1.0 equiv.), potassium acetate (7 mg, 0.07 mmol, 3.0 equiv.) and hydroxylamine hydrochloride (3 mg, 0.04 mmol, 3.0 equiv.), the oxime (5 mg, 77% yield) was afforded as a white solid after purification by flash column chromatography ( $\text{CH}_2\text{Cl}_2/\text{CH}_3\text{OH}$ , 3%  $\text{CH}_3\text{OH}$ ).  $^1\text{H-NMR}$  (500 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 11.85 (s, 1H), 11.31 (s, 1H), 9.68 (s, 1H), 7.28 (d,  $J = 8.65$  Hz,

1H), 6.90 (d,  $J = 8.40$  Hz, 2H), 6.44 (m, 2H), 6.23 (dd,  $J = 2.43, 8.65$  Hz, 1H), 6.21 (d,  $J = 2.43$  Hz, 1H), 4.87 (s, 2H), 3.95 (s, 2H).  $^{13}\text{C-NMR}$  (126 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 159.9, 159.4, 159.2, 146.9, 129.4, 128.9, 123.5, 114.0, 110.0, 106.8, 102.9, 54.9. HR-ESI-MS: calculated for  $\text{C}_{14}\text{H}_{15}\text{N}_2\text{O}_3$   $[M+H]^+$  259.1077, found 259.1075.

**1-(2,4-Dimethoxyphenyl)-2-(4-nitrophenyl)ethan-1-one (53).** Following a published procedure, 2-(4-aminophenyl)acetic acid (420 mg, 3.0 mmol) was dissolved with 1,3-dimethoxybenzene (450 mg, 3.0 mmol, 1.0 equiv.) in dichloroethane (6 mL).<sup>55</sup> Polyphosphoric acid (7 g) was added, and the reaction stirred for 3 h at 85 °C. After full conversion, the mixture was cooled to 0 °C and carefully basified with ammonia. The resulting solution was extracted with ethyl acetate (3 x 50 mL – the pH has to be over 7) and the combined organic layers were dried over  $\text{Na}_2\text{SO}_4$ , filtered and concentrated *in vacuo*. Purification by flash column chromatography ( $\text{CH}_2\text{Cl}_2$ /ethyl acetate + 1%  $\text{NH}_3$ , gradient from 0% to 40% ethyl acetate) afforded the product (200 mg, 15%) as a yellow solid.  $^1\text{H-NMR}$  (500 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  (ppm) = 7.66 (d,  $J = 8.67$  Hz, 1H), 6.92 (d,  $J = 2.83$  Hz, 2H), 6.64 (d,  $J = 2.83$  Hz, 2H), 6.57 (d,  $J = 2.28$  Hz, 1H), 6.54 (dd,  $J = 2.28, 8.67$  Hz, 1H), 4.11 (s, 2H), 3.90 (s, 3H), 3.84 (s, 3H).  $^{13}\text{C-NMR}$  (126 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta = 201.6, 166.4, 162.3, 147.0, 133.8, 131.2, 126.5, 121.7, 116.7, 106.8, 99.2, 56.1, 56.0, 50.0$ . HR-ESI-MS: calculated for  $\text{C}_{16}\text{H}_{18}\text{NO}_3$   $[M+H]^+$  272.1281, found 272.1281.

**(E)-2-(4-Aminophenyl)-1-(2,4-dimethoxyphenyl)ethan-1-one oxime (6).** Following procedure **O-C**, compound **53** (125 mg, 0.5 mmol, 1.0 equiv.), potassium acetate (136 mg, 1.5 mmol, 3.0 equiv.) and hydroxylamine hydrochloride (48 mg, 0.7 mmol, 1.5 equiv.) were dissolved in methanol (10 mL) and the oxime was formed. Purification by preparative HPLC ( $\text{H}_2\text{O}/\text{CH}_3\text{CN}$  + 0.1% formic acid, gradient 5% to 100%  $\text{CH}_3\text{CN}$ ) afforded product **6** (35 mg, 27%) as a white solid.  $^1\text{H-NMR}$  (500 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 3.72 (s, 3H), 3.76 (s, 3H), 3.81 (s, 2H), 4.78 (s, 2H), 6.35 (d,  $J = 8.40$  Hz, 2H), 6.39 (dd,  $J = 2.35, 8.36$  Hz, 1H), 6.51 (d,  $J = 2.30$  Hz, 1H), 6.67 (d,  $J = 8.35$  Hz, 2H), 6.88 (d,  $J = 8.30$  Hz, 1H), 10.90 (s, 1H).  $^{13}\text{C-NMR}$  (126 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 160.7, 158.1, 157.0, 146.5, 131.0, 129.3, 124.1, 118.8, 113.8, 104.5, 98.4, 55.4, 55.2, 32.6. HR-ESI-MS: calculated for  $\text{C}_{16}\text{H}_{19}\text{N}_2\text{O}_3$   $[M+H]^+$  287.1390, found 287.1384.

**2-(4-Chlorophenyl)-1-(6-methoxypyridin-3-yl)ethan-1-one (54).** Following procedure **O-A**, using commercially available methyl 6-methoxynicotinate (1.0 g,

5.9 mmol, 1.0 equiv.) and 4-chlorophenylacetic acid (1.0 g, 5.9 mmol, 1.0 equiv.), the pure product (1.1 g, 72%) was obtained after flash column chromatography (petroleum benzine/ethyl acetate, gradient from 0% to 100% ethyl acetate) as a white solid.  $^1\text{H-NMR}$  (500 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 8.76 (s, 1H), 8.08 (d,  $J$  = 8.80 Hz, 2H), 7.23 (m, 2H), 7.12 (m, 2H), 6.72 (d,  $J$  = 8.80 Hz, 1H), 4.12 (s, 2H), 3.93 (s, 3H).  $^{13}\text{C-NMR}$  (126 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 195.0, 167.0, 149.7, 138.7, 133.2, 132.7, 130.9, 129.1, 126.2, 111.6, 54.3, 44.8. HR-ESI-MS: calculated for  $\text{C}_{14}\text{H}_{13}\text{ClNO}_2$  [ $M+\text{H}$ ] $^+$  262.0629 ( $^{35}\text{Cl}$ ), 264.0600 ( $^{37}\text{Cl}$ ), found 262.0609 (100%), 264.0577 (30%).

**2-(4-Chlorophenyl)-1-(6-methoxypyridin-3-yl)ethan-1-one oxime (7E and 7Z).**

Starting from compound **54** (500 mg, 1.9 mmol, 1.0 equiv.) following procedure **O-C**, using potassium acetate (563 mg, 5.8 mmol, 3.0 equiv.) and hydroxylamine hydrochloride (200 mg, 2.7 mmol, 1.5 equiv.) in methanol (33 mL), the products (**7E**: 370 mg, 70%, **7Z**: 45 mg, 9%) were obtained after flash column chromatography (cyclohexane/ethyl acetate, gradient 0% to 30% ethyl acetate) as white solids. Note: 1D-NOESY experiments were performed on both compounds, irradiating the oxime-hydroxy proton. The isomers could be identified unambiguously. **7E**:  $^1\text{H-NMR}$  (500 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  (ppm) = 8.30 (dd,  $J$  = 0.51, 2.49 Hz, 1H), 7.96 (dd,  $J$  = 2.49, 8.79 Hz, 1H), 7.23 (s, 4H), 6.74 (dd,  $J$  = 0.51, 8.79 Hz, 1H), 4.16 (s, 2H), 3.88 (s, 3H).  $^{13}\text{C-NMR}$  (126 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  (ppm) = 165.8, 154.4, 145.9, 138.2, 137.2, 133.1, 131.3, 129.6, 126.7, 111.5, 54.2, 31.2. 1D-NOESY (500 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 11.56 (irradiation point), 8.40, 7.99, 7.31, 7.24, 6.80, 4.14. HR-ESI-MS: calculated for  $\text{C}_{14}\text{H}_{14}\text{ClN}_2\text{O}_2$  [ $M+\text{H}$ ] $^+$  277.0738 ( $^{35}\text{Cl}$ ), 279.0709 ( $^{37}\text{Cl}$ ), found 277.0738 (100%), 279.0705 (30%). **7Z**:  $^1\text{H-NMR}$  (500 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 11.09 (s, 1H), 8.36 (d,  $J$  = 2.32 Hz, 1H), 7.90 (dd,  $J$  = 2.32, 8.71 Hz, 1H), 7.30 (d,  $J$  = 8.40 Hz, 2H), 7.20 (d,  $J$  = 8.40 Hz, 2H), 6.78 (d,  $J$  = 8.71 Hz, 1H), 3.89 (s, 2H), 3.82 (s, 3H).  $^{13}\text{C-NMR}$  (126 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 162.9, 151.0, 147.1, 139.6, 136.7, 131.0, 130.6, 128.4, 122.1, 109.7, 53.2, 40.0. 1D-NOESY (500 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 11.09 (irradiation point), 8.36, 7.90, 7.20, 6.78, 3.89. HR-ESI-MS: calculated for  $\text{C}_{14}\text{H}_{14}\text{ClN}_2\text{O}_2$  [ $M+\text{H}$ ] $^+$  277.0738 ( $^{35}\text{Cl}$ ), 279.0709 ( $^{37}\text{Cl}$ ), found 277.0740 (100%), 279.0709 (30%).

**2-(4-Chlorophenyl)-1-(6-hydroxypyridin-3-yl)ethan-1-one (55).** Compound **54** (0.2 g, 0.38 mmol), LiCl (0.16 g, 1.9 mmol, 5.0 equiv.) and *p*-toluenesulfonic acid

(0.33 g, 1.9 mmol, 5.0 equiv.) were dissolved in dry DMF (15 mL) and heated to 150 °C for 24 h. The reaction was diluted with H<sub>2</sub>O (40 mL), and the mixture was extracted with ethyl acetate (3 x 30 mL). The combined organic layers were washed with saturated aqueous NaCl solution (30 mL), dried over MgSO<sub>4</sub>, filtered and concentrated *in vacuo*. Purification by flash column chromatography (petroleum benzene/ethyl acetate + 1% acetic acid, gradient from 0% or 100% ethyl acetate) afforded the product (54 mg, 29%) as a yellow solid. Note: The deprotection as described in procedure **O-B** does not yield any of the desired product. <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 12.25 (s, 1H), 8.38 (d, *J* = 2.56 Hz, 1H), 7.88 (dd, *J* = 2.74, 9.70 Hz, 1H), 7.37 (d, *J* = 2.80 Hz, 2H), 7.25 (d, *J* = 8.39 Hz, 2H), 6.38 (d, *J* = 9.67 Hz, 1H), 4.21 (s, 2H). <sup>13</sup>C-NMR (126 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 193.1, 162.5, 141.6, 138.5, 134.4, 131.7, 131.3, 128.2, 119.8, 116.0, 42.7. HR-ESI-MS: calculated for C<sub>13</sub>H<sub>11</sub>ClNO<sub>2</sub> [*M*+H]<sup>+</sup> 248.0473 (<sup>35</sup>Cl), 250.0443 (<sup>37</sup>Cl), found 248.0472 (100%), 250.0440 (30%).

**2-(4-Chlorophenyl)-1-(6-hydroxypyridin-3-yl)ethan-1-one oxime (8E and 8Z).**

Starting from compound **55** (40 mg, 0.2 mmol, 1.0 equiv.) following procedure **O-C**, using potassium acetate (48 mg, 0.5 mmol, 3.0 equiv.) and hydroxylamine hydrochloride (16 mg, 0.2 mmol, 1.5 equiv.), the products (**8E**: 10 mg, 24%, **8Z**: 3 mg, 8%) were afforded after preparative HPLC (H<sub>2</sub>O/CH<sub>3</sub>CN + 1% formic acid, gradient from 5% to 100% CH<sub>3</sub>CN) as white solids. **8E**: <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 11.71 (s, 1H), 11.36 (s, 1H), 7.83 (dd, *J* = 2.55, 9.75 Hz, 1H), 7.53 (d, *J* = 2.55 Hz, 1H), 7.33 (d, *J* = 8.42 Hz, 2H), 7.22 (d, *J* = 8.42 Hz, 2H), 6.33 (d, *J* = 9.75 Hz, 1H), 4.01 (s, 2H). <sup>13</sup>C-NMR (126 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 162.1, 151.4, 137.9, 136.3, 133.6, 130.8, 130.2, 128.5, 120.4, 113.8, 28.7. HR-ESI-MS: calculated for C<sub>13</sub>H<sub>12</sub>ClN<sub>2</sub>O<sub>2</sub> [*M*+H]<sup>+</sup> 263.0582 (<sup>35</sup>Cl), 265.0552 (<sup>37</sup>Cl), found 263.0581 (100%), 265.0549 (30%). **8Z**: <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 11.78 (s, 1H), 11.17 (s, 1H), 7.93 (d, *J* = 2.52 Hz, 1H), 7.70 (dd, *J* = 2.52, 9.68 Hz, 1H), 7.33 (d, *J* = 8.42 Hz, 2H), 7.21 (d, *J* = 8.42 Hz, 2H), 6.25 (d, *J* = 9.68 Hz, 1H), 3.81 (s, 2H). <sup>13</sup>C-NMR (126 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 161.7, 149.7, 141.8, 138.1, 137.5, 131.4, 130.9, 128.9, 119.3, 110.8, 38.8. HR-ESI-MS: calculated for C<sub>13</sub>H<sub>12</sub>ClN<sub>2</sub>O<sub>2</sub> [*M*+H]<sup>+</sup> 263.0582 (<sup>35</sup>Cl), 265.0552 (<sup>37</sup>Cl), found 263.0581 (100%), 265.0549 (30%).

**(E)-2-(4-Chlorophenyl)-1-(2,4-dihydroxyphenyl)ethan-1-one O-methyl oxime (11).** 2-(4-Chlorophenyl)-1-(2,4-dihydroxyphenyl)ethan-1-one (100 mg, 0.38 mmol,

1.0 equiv) and *O*-methylhydroxylamine•HCl (64 mg, 0.76 mmol, 2.0 equiv) were dissolved in methanol/pyridine (10:1, 4 mL) and added to a flask under N<sub>2</sub> atmosphere in solution.<sup>12</sup> Sodium sulfate (135 mg, 0.95 mmol, 2.5 equiv) was added and the reaction mixture was heated to 95 °C under reflux. The reaction mixture was refluxed for 18 h and then allowed to cool to room temperature. H<sub>2</sub>O (15 mL) was added followed by ethyl acetate (15 mL). The layers were mixed and separated and the aqueous layer was extracted with ethyl acetate (2x, 10 mL). The combined organic layers were washed with brine, dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated to give a pale yellow sticky solid which was purified by flash column chromatography (heptane:ethyl acetate, gradient from 0% to 100% ethyl acetate) to give the product (80 mg, 72%) as a white solid. <sup>1</sup>H-NMR (500 MHz, CD<sub>3</sub>OD): δ = 7.25 (m, 5H), 6.31 (d, *J* = 2.40 Hz, 1H), 6.28 (q, *J* = 3.71 Hz, 1H), 4.17 (s, 2H), 3.98 (s, 3H). <sup>13</sup>C-NMR (126 MHz, CD<sub>3</sub>OD): δ = 161.3, 161.1, 161.0, 136.8, 133.2, 131.0, 130.7, 129.6, 110.9, 108.3, 104.3, 62.7, 31.3. HR-ESI-MS: calculated for C<sub>15</sub>H<sub>15</sub>ClNO<sub>3</sub> [*M*+H]<sup>+</sup> 292.0735 (<sup>35</sup>Cl), 294.0705 (<sup>37</sup>Cl), found 292.0706 (100%), 294.0674 (30%).

**4-(2-(4-Chlorophenyl)-1-hydroxyethyl)benzene-1,3-diol (13).** 2-(4-Chlorophenyl)-1-(2,4-dihydroxyphenyl)ethan-1-one (100 mg, 0.38 mmol, 1.0 equiv.) was dissolved in methanol and added to a solution of sodium borohydride (22 mg, 0.58 mmol, 1.5 equiv.) in methanol (5 mL).<sup>12</sup> NaBH<sub>4</sub> was added two more times (16 mg, 0.42 mmol, 1.1 equiv. and 40 mg, 1.1 mmol, 2.8 equiv.), and the reaction was stirred for 1 h after each addition. After full conversion, the solvent was removed under reduced pressure and the solids were dissolved in a mixture of dichloromethane (20 mL), saturated aqueous NaCl solution (5 mL), and water (5 mL). The mixture was acidified to pH 5, and the organic layer was separated. The aqueous layer was further extracted with dichloromethane (2 x 35 mL), dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated. The product was purified by reverse-phase flash column chromatography (H<sub>2</sub>O/CH<sub>3</sub>OH gradient from 5% to 100% CH<sub>3</sub>OH) to give the product as a white solid (21 mg, 21%). <sup>1</sup>H-NMR (500 MHz, CD<sub>3</sub>OD): δ (ppm) = 7.20 (d, *J* = 2.81 Hz, 2H), 7.12 (d, *J* = 2.81 Hz, 2H), 6.94 (d, *J* = 7.99 Hz, 1H), 6.23 (m, 1H), 5.04 (d, *J* = 5.36, 7.70 Hz, 1H), 2.95 (ddd, *J* = 5.36, 7.70, 13.56, 54.73 Hz, 2H). <sup>13</sup>C-NMR (126 MHz, CD<sub>3</sub>OD): δ (ppm) = 158.5, 156.6, 139.4, 132.7, 132.2, 128.9, 128.5, 122.5, 107.3, 103.4, 72.0, 44.6. HR-ESI-MS: calculated for C<sub>14</sub>H<sub>12</sub>ClO<sub>3</sub> [*M*+H]<sup>-</sup> 263.0480 (<sup>35</sup>Cl), 265.0451 (<sup>37</sup>Cl), found 263.0457 (100%), 265.0426 (30%).

**Synthesis and characterization of indoles.** The synthesis and characterization of **3**, **16**, **17**, **20**, **24**, **33** were described previously.<sup>12,48</sup>

**Aldehyde formation (procedure I-A).** When the indole-carbaldehydes were not commercially available they were synthesized following a literature procedure.<sup>56</sup> POCl<sub>3</sub> (1.3 equiv.) was stirred in DMF for 15 minutes. A solution of 1-(4-chlorophenyl)-1*H*-indole (1.0 equiv.), 1-(4-chlorophenyl)-1*H*-indole (1.0 equiv.) or 1*H*-indole (1.0 equiv.) in DMF was added, and the reaction mixture was stirred at 80 °C for 15 min. Then aqueous NaOH solution (2 M) was added and stirred at 110 °C for 45 min. The reaction mixture was diluted with *tert*-butyl methyl ether (TBME) and H<sub>2</sub>O. The organic layer was separated, and the aqueous layer was extracted with TBME (2 x). The combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo* to obtain the crude product that was used without further purification unless stated otherwise.

**S<sub>N</sub>2 substitution (procedure I-B).** The 1-substituted-1*H*-indole-3-carbaldehyde products were synthesized by following a previously reported procedure.<sup>57</sup> Sodium hydride (1.8 equiv.) was suspended in DMF, and the suspension was cooled to 0 °C under nitrogen atmosphere. 1*H*-Indole-3-carbaldehyde (1.0 equiv.) was added. The mixture was stirred at 25 °C for 30 min, after which 3,4-dichlorobenzylbromide (1.2 equiv.) was added. The resulting mixture was stirred for 16 h. Water and ethyl acetate were added, and the layers were separated. The aqueous layer was extracted with ethyl acetate (2 x). The combined organic layers were washed with water (3 x) and saturated aqueous NaCl solution, dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo*. The crude product was directly used without further purification for the next reaction unless stated otherwise.

**Reduction (procedure I-C).** The (1-substituted-1*H*-indol-3-yl)methanol products were synthesized by following a previously reported procedure.<sup>58</sup> To a solution of 1-substituted-1*H*-indole-3-carbaldehyde (1.0 equiv.) in methanol, sodium borohydride (3.2 equiv.) was added portion-wise, and the reaction mixture was stirred for 1 h at 25 °C. Water was added with care to the reaction, and the mixture was extracted with dichloromethane (2 x). The combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo*.

**(1-(3,4-Dichlorobenzyl)-1H-pyrrolo[2,3-b]pyridin-3-yl)methanol (14).** 1H-Pyrrolo[2,3-b]pyridine-3-carbaldehyde (500 mg, 3.4 mmol, 1.0 equiv.) and 3,4-dichlorobenzyl-bromide (1.2 g, 4.9 mmol, 1.2 equiv.) were added as described in procedure **I-B** to a solution of NaH (150 mg, 60% dispersion in mineral oil, 3.8 mmol, 1.1 equiv.) in DMF (14 mL), and the alkylation product (600 mg, 58%) was purified by flash column chromatography (CH<sub>2</sub>Cl/CH<sub>3</sub>OH 99.5:0.5). Following procedure **I-C** the alkylation product (600 mg, 1.97 mmol, 1.0 equiv.) was reduced to the alcohol with NaBH<sub>4</sub> (240 mg, 6.3 mmol, 3.2 equiv.) in methanol (250 mL), and pure product **14** (507 mg, 84% yield) was obtained without further purification as a white solid. <sup>1</sup>H-NMR (500 MHz, CDCl<sub>3</sub>): δ (ppm) = 8.35 (dd, *J* = 1.54, 4.77 Hz, 1H), 8.06 (dd, *J* = 1.54, 7.89 Hz, 1H), 7.36 (d, *J* = 8.27 Hz, 1H), 7.31 (d, *J* = 2.05 Hz, 1H), 7.16 (s, 1H), 7.13 (dd, *J* = 4.77, 7.89 Hz, 1H), 7.05 (dd, *J* = 2.05, 8.27 Hz, 1H), 5.41 (s, 2H), 4.85 (s, 2H). <sup>13</sup>C-NMR (126 MHz, CDCl<sub>3</sub>): δ (ppm) = 147.9, 143.7, 138.0, 132.9, 131.9, 130.8, 129.5, 128.1, 126.9, 125.9, 119.5, 116.3, 114.7, 57.3, 46.8. HR-ESI-MS: calculated for C<sub>15</sub>H<sub>13</sub>Cl<sub>2</sub>N<sub>2</sub>O [*M*+H]<sup>+</sup> 307.0399 (<sup>35</sup>Cl), 309.0370 (<sup>37</sup>Cl), found 307.0394 (100%), 309.0361 (60%).

**(1-(3,4-Dichlorophenyl)-1H-indol-3-yl)methanol (15).** In contrast to procedure **I-B**, the suspension of indole-3-carbaldehyde (500 mg, 3.5 mmol, 1.0 equiv.) in DMF (10 mL), 3,4-dichlorofluorobenzene (0.5 mL, 4.2 mmol, 1.2 equiv.) and NaH (350 mg, 8.6 mmol, 2.4 equiv.) was stirred for 24 h at 190 °C.<sup>59</sup> After full conversion, the reaction mixture was diluted with TBME (10 mL) and washed with saturated aqueous NaCl solution (3 x 10 mL). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo* to obtain product **56** (960 mg, 3.3 mmol) that was used without further purification. Compound **15** was synthesized following procedure **I-C**, using **56** (960 mg, 3.3 mmol, 1.0 equiv.) and NaBH<sub>4</sub> (430 mg, 11.34 mmol, 3.4 equiv.) in methanol (200 mL). The reaction mixture was concentrated, the residue was dissolved in TBME (50 mL) and washed with saturated aqueous NaCl solution (100 mL) and water (100 mL). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo*. Purification by reversed phase column chromatography (H<sub>2</sub>O/CH<sub>3</sub>CN, gradient 20% to 100% CH<sub>3</sub>CN) afforded **15** (20% over two steps) as a yellow sticky solid. <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 7.72 (d, *J* = 7.74 Hz, 1H), 7.58 (m, 4H), 7.22 (t, *J* = 6.91 Hz, 1H), 7.15 (t, *J* = 6.91 Hz, 1H), 4.98 (t, *J* = 5.33 Hz, 1H), 4.70 (d, *J* = 5.33 Hz, 2H). <sup>13</sup>C-NMR (126 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 138.0, 135.4, 131.4,

129.8, 129.7, 128.2, 126.1, 125.1, 125.1, 122.8, 120.2, 119.9, 118.8, 110.32 55.2. HR-ESI-MS: calculated for  $C_{15}H_{10}Cl_2N^+$  [ $M-OH$ ] $^+$  274.0190 ( $^{35}Cl$ ), 276.0161 ( $^{37}Cl$ ), found 274.0184 (100%), 276.0153 (70%).

**1-(3,4-Dichlorobenzyl)-1*H*-indazol-3-yl)methanol (18).** Following procedure **I-B**, 3-iodo-1*H*-indazole (1.5 g, 6.1 mmol, 1.0 equiv.) was dissolved in a suspension of NaH (270 mg, 6.7 mmol, 1.1 equiv.) in DMF (40 mL). 3,4-Dichlorobenzylbromide (1 mL, 7.36 mmol, 1.2 equiv.) was added. The crude material was purified by column chromatography (heptane/ethyl acetate 97:3) affording 1-(3,4-dichlorobenzyl)-3-iodo-1*H*-indazole (**57**) as a white solid (2.1 g, 85% yield). A round-bottomed flask was charged with *i*-PrMgCl (2 M in THF, 3.0 mL, 6.0 mmol, 1.5 equiv.) and dry THF (5 mL) and was cooled to 0 °C.<sup>60</sup> A solution of compound **57** (1.6 g, 4.0 mmol) in dry THF (20 mL) was added dropwise, and the resulting mixture was stirred at 0 °C for 1 h. DMF (1.2 mL, 16.0 mmol, 4.0 equiv.) was added, and the mixture was stirred for 5 h. After full conversion, aqueous HCl solution (1 M 10 mL) and toluene (10 mL) were added to the reaction mixture. The layers were separated, and the organic layer was washed with saturated aqueous NaHCO<sub>3</sub> solution (5 mL) and concentrated *in vacuo*. The crude material was purified using column chromatography (heptane/ethyl acetate 95:5) affording 1-(3,4-dichlorobenzyl)-1*H*-indazole-3-carbaldehyde (**58**) as a white solid (73 mg, 6% yield). Indole **18** was synthesized following procedure **I-C**, using **58** (70 mg, 0.23 mmol, 1.0 equiv.) and NaBH<sub>4</sub> (9 mg, 0.23 mmol, 1.0 equiv.) in ethanol (10 mL). The crude product was purified using flash column chromatography (heptane/ethyl acetate 8:2) to provide **18** as a white solid (35 mg, 50%). <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 7.86 (d, *J* = 8.09 Hz, 1H), 7.70 (d, *J* = 8.49 Hz, 1H), 7.58 (d, *J* = 8.30 Hz, 1H), 7.52 (d, *J* = 1.96 Hz, 1H), 7.39 (m, 1H), 7.15 (m, 2H), 5.62 (s, 2H), 5.32 (t, *J* = 5.81 Hz, 1H), 4.77 (d, *J* = 5.81 Hz, 2H). <sup>13</sup>C-NMR (126 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 145.8, 140.4, 138.9, 131.1, 130.9, 130.2, 129.4, 127.8, 126.6, 122.3, 121.1, 120.3, 109.6, 56.6, 50.2. HR-ESI-MS: calculated for  $C_{15}H_{13}Cl_2N_2O$  [ $M+H$ ] $^+$  307.0399 ( $^{35}Cl$ ), 309.0370 ( $^{37}Cl$ ), found 307.0400 (100%), 309.0368 (70%), 289.0293 (20%).

**2-(1-(3,4-Dichlorobenzyl)-1*H*-indol-3-yl)ethan-1-ol (19).** Tryptophol (0.59 g, 3.7 mmol), *tert*-butyldimethylsilyl (TBDMS) chloride (0.86 g, 5.7 mmol, 1.5 equiv.), and imidazole (0.38 g, 5.6 mmol, 1.5 equiv.) were dissolved in CH<sub>2</sub>Cl<sub>2</sub> (50 mL).<sup>61</sup> The

reaction mixture was stirred at 25 °C for 24 h. Then it was diluted with CH<sub>2</sub>Cl<sub>2</sub> (50 mL), washed with saturated aqueous NaCl solution (100 mL), and water (100 mL). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo* to obtain crude 3-(2-((*tert*-butyldimethylsilyl)oxy)ethyl)-1*H*-indole (**59**) (1.03 g, 3.75 mmol) as a thick red oil, which was used without further purification in the next step. Compound **59** (1.03 g, 3.75 mmol, 1.0 equiv.) and 3,4-dichlorobenzyl bromide (1.08 g, 4.5 mmol, 1.2 equiv.) were mixed as described in procedure **I-B** in DMF (10 mL) with NaH (355 mg, 8.9 mmol, 2.4 equiv.). After workup, crude 3-(2-((*tert*-butyldimethylsilyl)oxy)ethyl)-1-(3,4-dichlorobenzyl)-1*H*-indole (**60**) was obtained as a yellow-green oil. Purification by normal phase column chromatography (heptane/EtOAc, gradient 0% to 10% EtOAc) afforded compound **60** (0.86 g, 205 μmol). To a solution of compound **60** (0.86 g, 2.05 mmol, 1.0 equiv.) in THF (50 mL), *tetra*-butylammonium fluoride (TBAF) in THF (1 M, 4.1 mL, 4.1 mmol, 2.0 equiv.) was added. The reaction mixture was stirred at 90 °C for 24 h. The reaction mixture was concentrated and purified by reverse phase column chromatography (H<sub>2</sub>O/CH<sub>3</sub>CN + 0.1% formic acid, gradient 20% to 100% CH<sub>3</sub>CN), affording compound **19** (0.32 g, 0.99 mmol, 26% over three steps) as a yellowish sticky oil. <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 7.56 (d, *J* = 8.31 Hz, 1H), 7.54 (d, *J* = 7.89 Hz, 1H), 7.48 (d, *J* = 1.96 Hz, 1H), 7.40 (d, *J* = 8.19 Hz, 1H), 7.32 (s, 1H), 7.10 (m, 2H), 7.01 (m, 1H), 5.37 (s, 2H), 4.67 (t, *J* = 5.32 Hz, 1H), 3.65 (m, 2H), 2.84 (t, *J* = 7.21 Hz, 2H). <sup>13</sup>C-NMR (126 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 139.8, 135.8, 131.1, 130.8, 129.9, 129.1, 128.1, 127.4, 126.6, 121.4, 118.9, 118.7, 112.1, 109.9, 61.6, 47.6, 28.7. HR-ESI-MS: calculated for C<sub>17</sub>H<sub>16</sub>Cl<sub>2</sub>NO [*M*+H]<sup>+</sup> 320.0603 (<sup>35</sup>Cl), 322.0574 (<sup>37</sup>Cl), found 320.0605 (100%), 322.0573 (70%).

**(1-(3,4-Dichlorobenzyl)-5-methoxy-1*H*-indol-3-yl)methanol (21).** 5-Methoxy-1*H*-indole-3-carbaldehyde (500 mg, 2.9 mmol, 1.0 equiv.) and 3,4-dichlorobenzylbromide (823 mg, 3.4 mmol, 1.2 equiv.) were mixed as described in procedure **I-B** in a suspension of NaH (125 mg, 60% dispersion in mineral oil, 3.1 mmol, 1.1 equiv.) in DMF (14 mL). The alkylation product (950 mg, 99%) was used without further purification. Following procedure **I-C**, the aldehyde (358 mg, 1.07 mmol, 1.0 equiv.) was reduced to the alcohol with NaBH<sub>4</sub> (130 mg, 3.42 mmol, 3.2 equiv.) in methanol (100 mL) and the pure product **21** (300 mg, 83%) was obtained without further purification as a yellowish solid. <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 7.56 (d,

$J = 8.30$  Hz, 1H), 7.47 (d,  $J = 1.95$  Hz, 1H), 7.38 (s, 1H), 7.31 (d,  $J = 8.90$  Hz, 1H), 7.12 (m, 2H), 6.75 (dd,  $J = 2.50, 8.90$  Hz, 1H), 5.34 (s, 2H), 4.83 (t,  $J = 5.45$  Hz, 1H), 4.61 (d,  $J = 5.45$  Hz, 2H), 3.75 (s, 3H).  $^{13}\text{C-NMR}$  (126 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 153.4, 139.8, 131.3, 131.1, 130.7, 129.9, 129.0, 127.7, 127.5, 127.4, 115.9, 111.6, 110.7, 101.2, 55.3, 55.3, 47.8. HR-ESI-MS: calculated for  $\text{C}_{17}\text{H}_{14}\text{Cl}_2\text{NO}^+$  [ $M\text{-OH}$ ] $^+$  318.0452 ( $^{35}\text{Cl}$ ), 320.0423 ( $^{37}\text{Cl}$ ), found 318.0415 (100%), 320.0383 (60%).

**(1-(3-Chlorobenzyl)-4-methoxy-1*H*-indol-3-yl)methanol (22).** 4-Methoxy-1*H*-indole-3-carbaldehyde (200 mg, 1.14 mmol, 1.0 equiv.) was synthesized following procedure **I-A**, using 4-methoxy-1*H*-indole (1.0 g, 6.8 mmol, 1.0 equiv.) as the starting material with  $\text{POCl}_3$  (1.3 g, 8.2 mmol, 1.2 equiv.) and DMF (2.5 g, 34 mmol, 5.0 equiv.) in NaOH (2 M, 40 mL). 3-Chlorobenzylbromide (282 mg, 1.37 mmol, 1.1 equiv.) was added as described in procedure **I-B** with NaH (82 mg, 60% in mineral oil, 2.05 mmol, 1.8 equiv.) in DMF (10 mL). Product **22** was synthesized by reduction of 1-(3-chlorobenzyl)-4-methoxy-1*H*-indole-3-carbaldehyde (380 mg, 1.27 mmol, 1.0 equiv.) following procedure **I-C**, using  $\text{NaBH}_4$  (154 mg, 4.06 mmol, 3.2 equiv.). The product (195 mg, 57% yield over three steps) was obtained after flash column chromatography (petroleum benzine/ethyl acetate, gradient from 0% to 100% ethyl acetate) as a sticky yellow solid.  $^1\text{H-NMR}$  (500 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 7.32 (m, 2H), 7.24 (m, 1H), 7.22 (m, 1H), 7.12 (m, 1H), 7.00 (d,  $J = 0.99$  Hz, 1H), 6.99 (s, 1H), 6.49 (m, 1H), 5.35 (s, 2H), 4.75 (d,  $J = 2.87$  Hz, 2H), 4.65 (t,  $J = 5.54$  Hz, 1H), 3.82 (s, 3H).  $^{13}\text{C-NMR}$  (126 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 154.1, 141.1, 137.7, 133.1, 130.5, 127.3, 126.8, 125.8, 125.2, 122.4, 116.7, 116.3, 103.4, 99.5, 56.9, 55.2, 48.4. HR-ESI-MS: calculated for  $\text{C}_{17}\text{H}_{15}\text{ClNO}^+$  [ $M\text{-OH}$ ] $^+$  284.0842 ( $^{35}\text{Cl}$ ), 286.0813 ( $^{37}\text{Cl}$ ), found 284.0831 (100%), 286.0800 (70%).

**(7-Chloro-1-(3,4-dichlorobenzyl)-1*H*-indol-3-yl)methanol (23).** 7-Chloro-1*H*-indole-3-carbaldehyde (500 mg, 2.8 mmol, 1.0 equiv.) and 3,4-dichlorobenzylbromide (804 mg, 3.4 mmol, 1.2 equiv.) were added as described in procedure **I-B** to a suspension of NaH (123 mg, 60% dispersion in mineral oil, 3.1 mmol, 1.1 equiv.) in DMF (14 mL). The alkylation product (577 mg, 61%) was purified by flash column chromatography (dichloromethane/heptane 3:1). Following procedure **I-C** the aldehyde (387 mg, 1.14 mmol, 1.0 equiv.) was reduced to the alcohol with  $\text{NaBH}_4$  (140 mg, 3.68 mmol, 3.2 equiv.) in methanol (100 mL), and pure product **23** (360 mg,

93%) was obtained without purification as a white solid.  $^1\text{H-NMR}$  (500 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 7.62 (dd,  $J = 0.91$ , 7.85 Hz, 1H), 7.56 (d,  $J = 8.34$  Hz, 1H), 7.49 (s, 1H), 7.29 (d,  $J = 2.01$  Hz, 1H), 7.13 (dd,  $J = 0.91$ , 7.50 Hz, 1H), 7.03 (t,  $J = 7.72$  Hz, 1H), 6.90 (dd,  $J = 2.01$ , 8.34 Hz, 1H), 5.73 (s, 2H), 4.98 (t,  $J = 5.38$  Hz, 1H), 4.65 (d,  $J = 5.38$  Hz, 2H).  $^{13}\text{C-NMR}$  (126 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 141.1, 131.2, 130.9, 130.9, 130.6, 130.2, 129.7, 127.9, 126.2, 123.2, 120.3, 118.8, 117.1, 115.4, 55.0, 49.7. HR-ESI-MS: calculated for  $\text{C}_{16}\text{H}_{11}\text{Cl}_3\text{N}^-$  [ $M\text{-OH}$ ] $^-$  321.9957 ( $^{35}\text{Cl}$ ), 323.9928 ( $^{37}\text{Cl}$ ), found 321.9953 (100%), 323.9921 (95%), 325.9889 (30%).

**(1-(3-Chlorobenzyl)-4-fluoro-1H-indol-3-yl)methanol (25).** 4-Fluoro-1H-indole-3-carbaldehyde (200 mg, 1.23 mmol, 1.0 equiv.) was synthesized following procedure **I-A** using 4-fluoro-1H-indole (1.0 g, 7.4 mmol, 1.0 equiv.) as starting material with  $\text{POCl}_3$  (1.4 g, 8.9 mmol, 1.2 equiv.) and DMF (2.7 g, 37 mmol, 5.0 equiv.) in NaOH (2 M, 40 mL). From there, 3-chlorobenzylbromide (302 mg, 1.47 mmol, 1.2 equiv.) was added as described in procedure **I-B** with NaH (88 mg, 60% mineral oil, 2.21 mmol, 1.8 equiv.) in DMF (10 mL). The final indole **25** was synthesized by reduction of 1-benzyl-4-fluoro-1H-indole-3-carbaldehyde (420 mg, 1.46 mmol, 1.0 equiv.) following procedure **I-C**, using  $\text{NaBH}_4$  (175 mg, 4.67 mmol, 3.2 equiv.) in methanol (10 mL). The product **25** (144 mg, 41% over three steps) was obtained after flash column chromatography (petroleum benzene/ethyl acetate, gradient from 0% to 100% ethyl acetate) as a sticky orange oil.  $^1\text{H-NMR}$  (500 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 7.45 (s, 1H), 7.33 (m, 2H), 7.29 (m, 2H), 7.16 (m, 1H), 7.07 (td,  $J = 5.13$ , 7.79 Hz, 1H), 6.77 (dd,  $J = 7.82$ , 11.17 Hz, 1H), 5.41 (s, 2H), 4.93 (t,  $J = 5.25$  Hz, 1H), 4.69 (d,  $J = 5.25$  Hz, 2H).  $^{13}\text{C-NMR}$  (126 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 156.4 (d,  $J = 245.28$  Hz), 140.6, 139.0 (d,  $J = 12.18$  Hz), 133.2, 130.6, 127.6, 127.5, 127.0, 125.9, 122.1 (d,  $J = 7.74$  Hz), 115.2 (d,  $J = 21.04$  Hz), 114.7 (d,  $J = 3.39$  Hz), 106.7 (d,  $J = 3.00$  Hz), 104.2 (d,  $J = 19.07$  Hz), 56.0, 48.5.  $^{19}\text{F-NMR}$  (470 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = -122.62 (q,  $J = 5.27$  Hz). HR-ESI-MS: calculated for  $\text{C}_{16}\text{H}_{12}\text{ClFN}^-$  [ $M\text{-OH}$ ] $^-$  272.0642 ( $^{35}\text{Cl}$ ), 274.0613 ( $^{37}\text{Cl}$ ), found 272.0630 (100%), 274.0600 (30%).

**(1-Benzyl-4-fluoro-1H-indol-3-yl)methanol (26).** 4-Fluoro-1H-indole-3-carbaldehyde (200 mg, 1.23 mmol, 1.0 equiv.) was synthesized following procedure **I-A**, using 4-fluoro-1H-indole (1.0 g, 7.4 mmol, 1.0 equiv.) as starting material with  $\text{POCl}_3$  (1.4 g, 8.9 mmol, 1.2 equiv.) and DMF (2.7 g, 37 mmol, 5.0 equiv.) in NaOH (2 M,

40 mL). From there, benzylbromide (252 mg, 1.5 mmol, 1.2 equiv.) was added as described in procedure **I-B**, using NaH (88 mg, 2.21 mmol, 1.8 equiv.) in DMF (10 mL). The final indole **26** was synthesized by reduction of the aldehyde (330 mg, 1.3 mmol, 1.0 equiv.) following procedure **I-C** using NaBH<sub>4</sub> (158 mg, 4.16 mmol, 3.2 equiv.) in methanol (10 mL). The product (203 mg, 67% over three steps) was obtained after flash column chromatography (petroleum benzene/ethyl acetate, gradient 0% to 100% ethyl acetate) as a sticky orange oil. <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 7.42 (s, 1H), 7.28 (m, 4H), 7.21 (d, *J* = 7.12 Hz, 2H), 7.05 (td, *J* = 5.33, 8.06 Hz, 1H), 6.75 (dd, *J* = 7.82, 11.32 Hz, 1H), 5.39 (s, 2H), 4.90 (t, *J* = 5.21 Hz, 1H), 4.69 (d, *J* = 5.21 Hz, 2H). <sup>13</sup>C-NMR (126 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 156.4 (d, *J* = 245.10 Hz), 139.0 (d, *J* = 12.12 Hz), 138.0, 128.6, 127.7, 127.5, 127.2, 121.9 (d, *J* = 7.76 Hz), 115.2 (d, *J* = 21.04 Hz), 114.4 (d, *J* = 3.37 Hz), 106.8 (d, *J* = 3.12 Hz), 104.0 (d, *J* = 19.11 Hz), 56.0 (d, *J* = 1.18 Hz), 49.2. <sup>19</sup>F-NMR (470 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = -122.76 (q, *J* = 5.42 Hz). HR-ESI-MS: calculated for C<sub>16</sub>H<sub>13</sub>FN [M-OH]<sup>+</sup> 238.1032, found 238.1021.

**1-(3,4-Dichlorobenzyl)-3-methyl-1H-indole (27)**. 3-Methyl-1H-indole (498 mg, 3.8 mmol, 1.0 equiv.) and 3,4-dichlorobenzylbromide (1.0 g, 4.2 mmol, 1.1 equiv.) were added as described in procedure **I-B** to a suspension of NaH (339 mg, 8.5 mmol, 2.2 equiv.) in DMF (10 mL), and pure product **27** (210 mg, 19%) was obtained after purification by flash column chromatography (heptane/ethyl acetate, gradient 0% to 100% ethyl acetate) followed by reversed-phase flash column chromatography (H<sub>2</sub>O/CH<sub>3</sub>CN, gradient 20% to 100% CH<sub>3</sub>CN) as an off-white solid. <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 7.55 (d, *J* = 8.30 Hz, 1H), 7.50 (d, *J* = 7.82 Hz, 1H), 7.47 (d, *J* = 1.99 Hz, 1H), 7.41 (d, *J* = 8.20 Hz, 1H), 7.27 (d, *J* = 0.87 Hz, 1H), 7.10 (m, 2H), 7.01 (td, *J* = 0.87, 7.46 Hz, 1H), 5.36 (s, 2H), 2.26 (d, *J* = 0.94 Hz, 3H). <sup>13</sup>C-NMR (126 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 139.8, 135.9, 131.1, 130.8, 129.9, 129.0, 128.6, 127.4, 126.5, 121.4, 118.8, 118.7, 109.8, 109.8, 47.6, 9.5. HR-ESI-MS: calculated for C<sub>16</sub>H<sub>14</sub>Cl<sub>2</sub>N [M+H]<sup>+</sup> 290.0498 (<sup>35</sup>Cl), 292.0468 (<sup>37</sup>Cl), found 290.0497 (100%), 292.0466 (70%).

**(1-(3,4-Dichlorobenzyl)-6-fluoro-1H-indol-3-yl)methanol (28)**. 6-Fluoro-1H-indole-3-carbaldehyde (500 mg, 3.1 mmol, 1.0 equiv.) and 3,4-dichlorobenzylbromide (883 mg, 3.7 mmol, 1.2 equiv.) were added as described in procedure **I-B** in a

suspension of NaH (135 mg, 60% dispersion in mineral oil, 3.4 mmol, 1.1 equiv.) in DMF (14 mL). The alkylation product (1.0 g, 98%) was used without further purification. Following procedure **I-C**, the aldehyde (1.0 g, 3.0 mmol, 1.0 equiv.) was reduced to the alcohol with NaBH<sub>4</sub> (385 mg, 10.2 mmol, 3.4 equiv.) in methanol (130 mL), and product **28** (473 mg, 48% yield over two steps) was obtained after purification by flash column chromatography (dichloromethane) as a white solid. <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 7.59 (m, 2H), 7.55 (d, *J* = 1.95 Hz, 1H), 7.43 (s, 1H), 7.38 (dd, *J* = 2.29, 10.49 Hz, 1H), 7.18 (dd, *J* = 1.95, 8.33 Hz, 1H), 6.89 (m, 1H), 5.36 (s, 2H), 4.90 (t, *J* = 5.40 Hz, 1H), 4.61 (d, *J* = 5.40 Hz, 2H). <sup>13</sup>C-NMR (126 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 159.2 (d, *J* = 235.08 Hz), 139.3, 136.2 (d, *J* = 12.27 Hz), 131.1, 130.9, 130.1, 129.3, 127.6, 127.3 (d, *J* = 3.45 Hz), 123.9, 120.6 (d, *J* = 10.28 Hz), 116.7, 107.4 (d, *J* = 24.49 Hz), 96.5 (d, *J* = 26.48 Hz), 55.2, 47.7. <sup>19</sup>F-NMR (470 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = -120.84 (m). HR-ESI-MS: calculated for C<sub>16</sub>H<sub>11</sub>Cl<sub>2</sub>FN<sup>+</sup> [*M*-OH]<sup>+</sup> 306.0253 (<sup>35</sup>Cl), 308.0223 (<sup>37</sup>Cl), found 306.0248 (100%), 308.0215 (70%).

**1-(3,4-Dichlorobenzyl)-1*H*-indole (29).** Indole (509 mg, 4.35 mmol, 1.0 equiv.) and 3,4-dichlorobenzylbromide (0.87 mL, 5.12 mmol, 1.2 equiv.) were added as described in procedure **I-B** to a suspension of NaH (381 mg, 9.5 mmol, 2.2 equiv.) in DMF (20 mL). The obtained crude material was purified by flash column chromatography (heptane/ethyl acetate, gradient 0% to 20% ethyl acetate) followed by reversed-phase column chromatography (H<sub>2</sub>O:CH<sub>3</sub>CN + 0.1% formic acid, gradient 20% to 100% CH<sub>3</sub>CN), affording compound **29** (63 mg, 5% yield) as a pale yellow sticky oil. <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 7.57 (d, *J* = 8.26 Hz, 1H), 7.56 (d, *J* = 7.83 Hz, 1H), 7.54 (d, *J* = 3.16 Hz, 1H), 7.48 (dd, *J* = 0.68, 8.26 Hz, 1H), 7.45 (q, *J* = 2.98 Hz, 1H), 7.11 (m, 2H), 7.02 (td, *J* = 0.94, 7.47 Hz, 1H), 6.50 (dd, *J* = 0.78, 3.16 Hz, 1H), 5.44 (s, 2H). <sup>13</sup>C-NMR (126 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 139.62, 135.6, 131.1, 130.8, 130.0, 129.1, 129.0, 128.3, 127.4, 121.4, 120.6, 119.4, 110.1, 101.4, 47.8. HR-ESI-MS: calculated for C<sub>15</sub>H<sub>12</sub>Cl<sub>2</sub>N [*M*+H]<sup>+</sup> 276.0341 (<sup>35</sup>Cl), 278.0312 (<sup>37</sup>Cl), found 276.0339 (100%), 278.0309 (70%).

**(1-(3,4-Dichlorobenzyl)-5-nitro-1*H*-indol-3-yl)methanol (30).** 5-Nitro-1*H*-indole-3-carbaldehyde (1.0 g, 5.3 mmol, 1.0 equiv.) and 3,4-dichlorobenzylbromide (1.5 g, 6.3 mmol, 1.2 equiv.) were added as described in procedure **I-B** to a suspension of NaH (231 mg, 60% dispersion in mineral oil, 5.8 mmol, 1.1 equiv.) in DMF (40 mL).

The alkylation product (1.5 g, 98%) was obtained without further purification. Following procedure **I-C**, the aldehyde (2.0 g, 5.7 mmol, 1.0 equiv.) was reduced to the alcohol with NaBH<sub>4</sub> (693 mg, 18.3 mmol, 3.2 equiv.) in methanol (600 mL), and pure product **30** (800 mg, 40% over two steps) was obtained after purification by flash column chromatography (dichloromethane/ethyl acetate 95:5) as dark yellow crystals. <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 8.63 (d, *J* = 2.28 Hz, 1H), 8.03 (dd, *J* = 2.28, 9.16 Hz, 1H), 7.71 (d, *J* = 9.16 Hz, 2H), 7.59 (m, 2H), 7.18 (dd, *J* = 2.11, 8.30 Hz, 1H), 5.50 (s, 2H), 5.14 (t, *J* = 5.46 Hz, 1H), 4.70 (d, *J* = 5.46 Hz, 2H). <sup>13</sup>C-NMR (126 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 140.7, 139.1, 138.7, 131.3, 131.0, 130.4, 130.4, 129.4, 127.6, 126.5, 119.3, 117.0, 116.7, 110.7, 54.9, 48.0. HR-ESI-MS: calculated for C<sub>16</sub>H<sub>11</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>2</sub><sup>-</sup> [*M*-OH]<sup>-</sup> 333.0198 (<sup>35</sup>Cl), 335.0168 (<sup>37</sup>Cl), found 333.0193 (100%), 335.0161 (60%).

**(1-Benzyl-5-bromo-1*H*-indol-3-yl)methanol (31).** 1-Benzyl-5-bromo-1*H*-indole (100 mg, 0.35 mmol, 1.0 equiv.) was added to a solution of POCl<sub>3</sub> (40 μL, 0.43 mmol, 1.2 equiv.) in DMF (3 mL) as described in procedure **I-A**. Crude 1-benzyl-5-bromo-1*H*-indole-3-carbaldehyde (**61**) was used without further purification in the next step. Compound **61** was reduced to the alcohol **31** following procedure **I-C**, using an excess of NaBH<sub>4</sub> in methanol (50 mL). The pure product (65 mg, 60% over two steps) was obtained without further purification as a white solid. <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 7.79 (d, *J* = 1.89 Hz, 1H), 7.46 (s, 1H), 7.42 (d, *J* = 8.71 Hz, 1H), 7.30 (m, 2H), 7.22 (m, 4H), 5.38 (s, 2H), 4.91 (t, *J* = 5.48 Hz, 1H), 4.60 (d, *J* = 5.48 Hz, 2H). <sup>13</sup>C-NMR (126 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 138.1, 135.0, 129.0, 128.6, 128.5, 127.5, 127.1, 123.7, 121.7, 115.8, 112.3, 111.5, 55.1, 49.1. HR-ESI-MS: calculated for C<sub>16</sub>H<sub>13</sub>BrN<sup>-</sup> [*M*-OH]<sup>-</sup> 298.0232 (<sup>35</sup>Cl), 300.0211 (<sup>37</sup>Cl), found 298.0226 (100%), 300.0202 (95%).

**(1-(3,4-Dichlorobenzyl)-4-methoxy-1*H*-indol-3-yl)methanol (32).** 4-Methoxy-1*H*-indole-3-carbaldehyde (500 mg, 2.9 mmol, 1.0 equiv.) and 3,4-dichlorobenzylbromide (820 mg, 3.4 mmol, 1.2 equiv.) were added as described in procedure **I-B** to a suspension of NaH (123 mg, 60% dispersion in mineral oil, 3.1 mmol, 1.1 equiv.) in DMF (14 mL). The alkylation product (200 mg, 21%) was purified by flash column chromatography (dichloromethane). Following procedure **I-C**, the aldehyde (200 mg, 0.6 mmol, 1.0 equiv.) was reduced to the alcohol using NaBH<sub>4</sub> (113 mg, 3.0 mmol,

5.0 equiv.) in methanol (100 mL), and pure product **32** (150 mg, 74% over two steps) was obtained without further purification as a pinkish solid.  $^1\text{H-NMR}$  (500 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 7.57 (d,  $J$  = 8.30 Hz, 1H), 7.45 (d,  $J$  = 1.94 Hz, 1H), 7.25 (s, 1H), 7.10 (dd,  $J$  = 1.94, 8.30 Hz, 1H), 7.00 (m, 2H), 6.49 (dd,  $J$  = 2.94, 5.54 Hz, 1H), 5.35 (s, 2H), 4.75 (d,  $J$  = 5.33 Hz, 2H), 4.66 (t,  $J$  = 5.54 Hz, 1H), 3.82 (s, 3H).  $^{13}\text{C-NMR}$  (126 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 154.1, 139.7, 137.6, 131.1, 130.8, 129.9, 129.0, 127.4, 125.1, 122.5, 116.8, 116.4, 103.4, 99.6, 56.9, 55.2, 47.8. HR-ESI-MS: calculated for  $\text{C}_{17}\text{H}_{14}\text{Cl}_2\text{NO}^+ [\text{M-OH}]^+$  318.0452 ( $^{35}\text{Cl}$ ), 320.0423 ( $^{37}\text{Cl}$ ), found 318.0448 (100%), 320.0416 (70%).

**(1-(3,4-Dichlorobenzyl)-4-fluoro-1H-indol-3-yl)methanol (34)**. The aldehyde **62** (900 mg, 91%) was synthesized as a white solid, following procedure **I-B** using 4-fluoro-1H-indole-3-carbaldehyde (500 mg, 3.1 mmol, 1.0 equiv.) and 3,4-dichlorobenzylbromide (0.55 mL, 3.7 mmol, 1.2 equiv.) in a suspension of NaH (134 mg, 3.4 mmol, 1.1 equiv.) in DMF (40 mL) and purified by flash column chromatography (dichloromethane). Indole **34** was synthesized following procedure **I-C**, using **62** (100 mg, 0.3 mmol, 1.0 equiv.) and  $\text{NaBH}_4$  (12 mg, 0.3 mmol, 1.0 equiv.) in THF (25 mL). The crude product was purified using flash column chromatography (heptane/ethyl acetate 8:2) to provide **34** (25 mg, 26%).  $^1\text{H-NMR}$  (500 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 7.64 (d,  $J$  = 8.30 Hz, 1H), 7.59 (d,  $J$  = 2.02 Hz, 1H), 7.52 (s, 1H), 7.35 (d,  $J$  = 8.25 Hz, 1H), 7.21 (dd,  $J$  = 2.02, 8.30 Hz, 1H), 7.13 (td,  $J$  = 5.25, 8.01 Hz, 1H), 6.83 (dd,  $J$  = 6.27 Hz, 1H), 5.47 (s, 2H), 4.98 (t,  $J$  = 5.27 Hz, 1H), 4.75 (d,  $J$  = 5.20 Hz, 2H).  $^{13}\text{C-NMR}$  (126 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 156.8 (d,  $J$  = 245.14 Hz), 139.7, 139.4 (d,  $J$  = 12.06 Hz), 131.6, 131.3, 130.54, 129.6, 128.0 (d,  $J$  = 3.17 Hz), 122.7 (d,  $J$  = 7.59 Hz), 115.7, 115.5, 115.3 (d,  $J$  = 3.46 Hz), 107.1 (d,  $J$  = 3.30 Hz), 104.7 (d,  $J$  = 18.98 Hz), 56.4 (d,  $J$  = 1.38 Hz), 48.4.  $^{19}\text{F-NMR}$  (470 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = -122.57 (dd,  $J$  = 5.33 Hz). HR-ESI-MS: calculated for  $\text{C}_{16}\text{H}_{11}\text{Cl}_2\text{FN}^+ [\text{M-OH}]^+$  306.0253 ( $^{35}\text{Cl}$ ), 308.0223 ( $^{37}\text{Cl}$ ), found 306.0248 (100%), 308.0215 (70%).

**1-(1-(3,4-Dichlorobenzyl)-1H-indol-3-yl)-N-(2,4-dimethoxybenzyl)methanamine (35)**. 1-(3,4-Dichlorobenzyl)-1H-indole-3-carbaldehyde was prepared according to procedure **I-B**. To a solution of 1-(3,4-dichlorobenzyl)-1H-indole-3-carbaldehyde (0.5 g, 1.6 mmol, 1.0 equiv.) in dichloroethane (33 mL) sodium triacetoxyborohydride (1.05 g, 4.9 mmol, 3.0 equiv.), 2,4-dimethoxy benzylamine (0.74 mL, 4.9 mmol,

3.0 equiv.) and acetic acid (0.09 mL, 1.6 mmol, 1.0 equiv.) were added. The reaction mixture was stirred for 23 h and terminated by the addition of water and a saturated aqueous NaHCO<sub>3</sub>-solution (10 mL). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo*. The crude material was purified by non-pressurized flash column chromatography (CH<sub>2</sub>Cl<sub>2</sub>:CH<sub>3</sub>OH 96:4). The product was a yellow oil (17 mg, 2%). Note: Several attempts with different pressurized column chromatography methods did not provide the clean compound. <sup>1</sup>H-NMR (500 MHz, CD<sub>3</sub>OD): δ (ppm) = 7.56 (d, *J* = 7.90 Hz, 1H), 7.39 (d, *J* = 8.30 Hz, 1H), 7.31 (t, *J* = 5.77 Hz, 2H), 7.24 (d, *J* = 1.95 Hz, 1H), 7.14 (m, 3H), 7.03 (dd, *J* = 3.40, 8.33 Hz, 1H), 6.54 (d, *J* = 2.30 Hz, 1H), 6.48 (dd, *J* = 3.53, 8.34 Hz, 1H), 5.35 (s, 2H), 4.05 (s, 2H), 3.85 (s, 2H), 3.78 (s, 3H), 3.71 (s, 3H). <sup>13</sup>C-NMR (126 MHz, CD<sub>3</sub>OD): δ (ppm) = 162.7, 160.3, 140.4, 137.9, 133.5, 132.5, 132.3, 131.8, 129.8, 129.4, 129.1, 127.8, 123.4, 120.9, 119.6, 118.0, 111.8, 111.1, 105.4, 99.4, 55.8, 55.8, 49.5, 48.3, 43.4. HR-ESI-MS: calculated for C<sub>25</sub>H<sub>25</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>2</sub> [*M*+H]<sup>+</sup> 455.1288 (<sup>35</sup>Cl), 457.1258 (<sup>37</sup>Cl), found 455.1290 (100%), 457.1257 (60%).

**Synthesis and characterization of aminothiazoles.** We previously reported the synthesis and characterization of compounds **1**, **36**, **38**, **46**, and **48**.<sup>12</sup> 1-(2-Amino-4-methylthiazol-5-yl)-2-bromoethan-1-one (**63**), 2-bromo-1-(2,4-dimethylthiazol-5-yl)ethan-1-one (**64**), and 2-bromo-1-(pyridin-2-yl)ethan-1-one (**65**) were synthesized following literature procedures and all data were consistent with the reported values.<sup>62,58</sup>

**Aminothiazole formation (procedure A-A).** Synthesis of aminothiazoles followed a previously reported procedure.<sup>62</sup> The impure mixture of α-bromoketones (1.0 equiv.) and substituted thiourea (0.95 equiv.) were dissolved in absolute ethanol. Then, *N,N*-diisopropylethylamine (DIPEA) (1.1 equiv.) was added, and the mixture stirred for up to three days. TLC analysis showed that the product spot turned red after irradiation with UV-light. After completion of the reaction, the solvent was evaporated, the residue diluted with ethyl acetate and filtered through Celite. The filtrate was washed with water (3 x) and saturated aqueous NaCl solution (2 x), dried over MgSO<sub>4</sub>, filtered and concentrated *in vacuo*.

**5-(Pyridin-4-yl)-*N*-(4-(trifluoromethyl)phenyl)thiazol-2-amine (37).** Compounds 2-bromo-1-(pyridin-4-yl)ethan-1-one hydrochloride (150 mg, 0.5 mmol, 1.0 equiv.) and

1-(4-(trifluoromethyl)phenyl) thiourea (117 mg, 0.5 mmol, 1.0 equiv.) were mixed as described in procedure **A-A** in ethanol (15 mL). The pure product (172 mg, 100%) was afforded as a yellow solid without further purification.  $^1\text{H-NMR}$  (500 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  (ppm) = 8.80 (d,  $J$  = 6.90 Hz, 2H), 8.56 (d,  $J$  = 6.90 Hz, 2H), 8.18 (s, 1H), 7.95 (d,  $J$  = 8.65 Hz, 2H), 7.66 (d,  $J$  = 8.65 Hz, 2H).  $^{13}\text{C-NMR}$  (126 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 163.4, 148.5, 145.5, 143.9, 142.5, 126.5 (d,  $J$  = 3.66 Hz), 124.6 (q,  $J$  = 270.96 Hz), 122.4, 121.6 (q,  $J$  = 31.99 Hz), 117.0, 116.6.  $^{19}\text{F-NMR}$  (470 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = -59.96. HR-ESI-MS: calculated for  $\text{C}_{15}\text{H}_{11}\text{F}_3\text{N}_3\text{S}$  [ $M+\text{H}$ ] $^+$  322.0620, found 322.0613.

**4-(2-((4-(Trifluoromethyl)phenyl)amino)thiazol-5-yl)benzene-1,3-diol (39).** This compound was ordered from Princeton Biomolecular Research and analyzed by NMR and MS prior to  $\text{IC}_{50}$  determination.  $^1\text{H-NMR}$  (500 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 10.63 (s, 1H), 10.46 (s, 1H), 9.50 (s, 1H), 8.18 (s, 1H), 7.76 (m, 2H), 7.58 (t,  $J$  = 8.03 Hz, 1H), 7.30 (d,  $J$  = 7.44 Hz, 1H), 7.25 (s, 1H), 6.37 (m, 1H), 6.31 (d,  $J$  = 8.46 Hz, 1H).  $^{13}\text{C-NMR}$  (126 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 161.5, 158.1, 156.3, 147.4, 141.6, 130.3, 129.8 (q,  $J$  = 31.54 Hz), 128.9, 124.2 (q,  $J$  = 272.27 Hz), 120.4, 117.4 (q,  $J$  = 3.52 Hz), 112.8 (q,  $J$  = 4.03 Hz), 111.9, 107.0, 102.9, 102.3.  $^{19}\text{F-NMR}$  (470 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = -61.41. HR-ESI-MS: calculated for  $\text{C}_{16}\text{H}_{12}\text{F}_3\text{N}_2\text{O}_2\text{S}$  [ $M+\text{H}$ ] $^+$  353.0566, found 353.0559.

***N*<sup>2</sup>-(3-Methoxyphenyl)-4'-methyl-[5,5'-bithiazole]-2,2'-diamine (40).** Compounds **63** (401 mg, 1.7 mmol, 1.0 equiv.) and 1-(3-methoxyphenyl)thiourea (310 mg, 1.7 mmol, 1.0 equiv.) were mixed as described in procedure **A-A** in ethanol (15 mL). Purification by preparative HPLC ( $\text{H}_2\text{O}:\text{CH}_3\text{OH}$  + 0.05% formic acid, gradient 5% to 100%  $\text{CH}_3\text{OH}$ ) afforded the product (27 mg, 7%) as a light-pink solid.  $^1\text{H-NMR}$  (500 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 10.23 (s, 1H), 8.13 (s, 1H, formic acid), 7.54 (t,  $J$  = 2.21 Hz, 1H), 7.19 (t,  $J$  = 8.12 Hz, 1H), 7.01 (dd,  $J$  = 1.57, 8.12 Hz, 1H), 6.96 (s, 2H), 6.60 (s, 1H), 6.52 (dd,  $J$  = 2.21, 8.25 Hz, 1H), 3.76 (s, 3H), 2.32 (s, 3H).  $^{13}\text{C-NMR}$  (126 MHz,  $(\text{CD}_3)_2\text{SO}$ ):  $\delta$  (ppm) = 165.8, 163.1 (formic acid), 162.1, 159.9, 144.2, 143.6, 142.3, 129.6, 114.3, 109.2, 106.8, 102.5, 99.8, 55.0, 17.1. HR-ESI-MS: calculated for  $\text{C}_{14}\text{H}_{15}\text{N}_4\text{OS}_2$  [ $M+\text{H}$ ] $^+$  319.0682, found 319.0674.

***N*<sup>2</sup>-(3-Chloro-2-methylphenyl)-4'-methyl-[5,5'-bithiazole]-2,2'-diamine (41).** Compounds **63** (752 mg, 3.1 mmol, 1.0 equiv.) and 1-(3-chloro-2-methylphenyl)thiourea (619 mg, 3.1 mmol, 1.0 equiv.) were mixed as described in

procedure **A-A** in ethanol (31 mL). Purification by preparative HPLC (H<sub>2</sub>O/CH<sub>3</sub>OH + 0.05% formic acid, gradient from 5% to 100% CH<sub>3</sub>OH) afforded the product (240 mg, 23%) as a pink solid. <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 9.51 (s, 1H), 8.13 (s, 1H, formic acid), 7.87 (q, *J* = 3.10 Hz, 1H), 7.18 (q, *J* = 5.32 Hz, 2H), 6.93 (s, 2H), 6.58 (s, 1H), 2.31 (s, 3H), 2.27 (s, 3H). <sup>13</sup>C-NMR (126 MHz, CD<sub>3</sub>OD): δ (ppm) = 165.8, 164.3, 163.3 (formic acid), 144.3, 143.5, 141.0, 134.0, 127.2, 127.1, 124.1, 120.1, 114.2, 100.5, 17.0, 15.0. HR-ESI-MS: calculated for C<sub>14</sub>H<sub>14</sub>ClN<sub>4</sub>S<sub>2</sub> [*M*+H]<sup>+</sup> 337.0343 (<sup>35</sup>Cl), 339.0313 (<sup>37</sup>Cl), found 337.0334 (100%), 339.0302 (30%).

**2',4'-Dimethyl-N-(4-(trifluoromethyl)phenyl)-[5,5'-bithiazol]-2-amine (42).**

Compounds **64** (100 mg, 0.4 mmol, 1.0 equiv.) and 1-(4-(trifluoromethyl)phenyl)thiourea (79 mg, 0.4 mmol, 1.0 equiv.) were mixed as described in procedure **A-A** in ethanol (4 mL). Purification by flash column chromatography (petroleum benzene/ethyl acetate, gradient 0% to 100% ethyl acetate) afforded the product (112 mg, 73%) as a yellow solid. <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 10.76 (s, 1H), 7.85 (d, *J* = 8.50 Hz, 2H), 7.69 (d, *J* = 8.50 Hz, 2H), 7.09 (s, 1H), 2.61 (s, 3H), 2.54 (s, 3H). <sup>13</sup>C-NMR (126 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 162.4, 162.1, 147.3, 144.2, 142.3, 126.4 (q, *J* = 3.59 Hz), 126.4, 124.7 (q, *J* = 270.86 Hz), 121.2 (q, *J* = 31.94 Hz), 116.6, 104.8, 18.7, 17.0. <sup>19</sup>F-NMR (470 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = -59.88. HR-ESI-MS: calculated for C<sub>15</sub>H<sub>13</sub>F<sub>3</sub>N<sub>3</sub>S<sub>2</sub> [*M*+H]<sup>+</sup> 356.0497, found 356.0491.

**4-(2-((3,4-Dimethylphenyl)amino)thiazol-4-yl)benzene-1,2-diol (43).**

This compound was ordered from Princeton Biomolecular Research and analyzed by NMR and MS prior to IC<sub>50</sub> determination. <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 9.97 (s, 1H), 9.08 (s, 1H), 9.04 (s, 1H), 7.46 (dd, *J* = 2.13, 8.28 Hz, 1H), 7.39 (d, *J* = 2.13 Hz, 1H), 7.30 (d, *J* = 2.05 Hz, 1H), 7.17 (dd, *J* = 2.05, 8.28 Hz, 1H), 7.07 (d, *J* = 8.20 Hz, 1H), 6.91 (s, 1H), 6.76 (d, *J* = 8.20 Hz, 1H), 2.22 (s, 3H), 2.16 (s, 3H). <sup>13</sup>C-NMR (126 MHz, (CD<sub>3</sub>)<sub>2</sub>SO): δ (ppm) = 163.2, 150.7, 145.5, 145.3, 139.4, 136.8, 130.1, 129.1, 126.7, 118.5, 117.3, 115.8, 114.6, 113.6, 99.6, 20.0, 18.9. HR-ESI-MS: calculated for C<sub>17</sub>H<sub>17</sub>N<sub>2</sub>O<sub>2</sub>S [*M*+H]<sup>+</sup> 313.1005, found 313.0992.

**N-Phenyl-5-(pyridin-2-yl)thiazol-2-amine (44).** Compounds **65** (156 mg, 0.6 mmol, 1.0 equiv.) and 1-(2,5-dimethylphenyl)thiourea (100 mg, 0.6 mmol, 1.0 equiv.) were mixed as described in procedure **A-A** in ethanol (5 mL). The pure product (157 mg, 95%) was obtained after workup as an orange-brown solid. <sup>1</sup>H-NMR (500 MHz,

(CD<sub>3</sub>)<sub>2</sub>SO):  $\delta$  (ppm) = 9.37 (s, 1H), 8.57 (dt,  $J$  = 1.27, 4.78 Hz, 1H), 7.87 (m, 2H), 7.78 (m, 1H), 7.47 (s, 1H), 7.30 (ddd,  $J$  = 2.24, 5.57, 6.63 Hz, 1H), 7.11 (d,  $J$  = 7.55 Hz, 1H), 6.85 (dd,  $J$  = 1.27, 7.55 Hz, 1H), 2.30 (s, 3H), 2.23 (s, 3H). <sup>13</sup>C-NMR (126 MHz, (CD<sub>3</sub>)<sub>2</sub>SO):  $\delta$  (ppm) = 165.9, 152.2, 150.1, 149.4, 139.2, 137.4, 135.6, 130.6, 126.1, 124.4, 122.6, 122.0, 120.2, 106.7, 21.0, 17.7. HR-ESI-MS: calculated for C<sub>16</sub>H<sub>16</sub>N<sub>3</sub>S [ $M+H$ ]<sup>+</sup> 282.1059, found 282.1054.

**N-(3-Chloro-2-methylphenyl)-5-(pyridin-2-yl)thiazol-2-amine (45).** Compounds **65** (20 mg, 0.1 mmol, 1.0 equiv.) and 1-(3-chloro-2-methylphenyl)thiourea (20 mg, 0.1 mmol, 1.0 equiv.) were mixed as described in procedure **A-A** in ethanol (5 mL). The pure product (20 mg, 66%) was obtained after workup as an orange solid. <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO):  $\delta$  (ppm) = 9.60 (s, 1H), 8.57 (d,  $J$  = 4.65 Hz, 1H), 8.01 (d,  $J$  = 8.02 Hz, 1H), 7.86 (m, 2H), 7.53 (s, 1H), 7.29 (ddd,  $J$  = 1.72, 5.80, 6.99 Hz, 1H), 7.25 (d,  $J$  = 8.02 Hz, 1H), 7.20 (d,  $J$  = 7.44 Hz, 1H), 2.34 (s, 3H). <sup>13</sup>C-NMR (126 MHz, (CD<sub>3</sub>)<sub>2</sub>SO):  $\delta$  (ppm) = 165.2, 152.1, 150.1, 149.4, 140.9, 137.3, 134.0, 127.4, 126.9, 124.0, 122.6, 120.3, 120.0, 107.5, 15.0. HR-ESI-MS: calculated for C<sub>15</sub>H<sub>13</sub>ClN<sub>3</sub>S [ $M+H$ ]<sup>+</sup> 302.0513 (<sup>35</sup>Cl), 304.0484 (<sup>37</sup>Cl), found 302.0490 (100%), 304.0456 (30%).

**N-(3,4-Dimethylphenyl)-4-(pyridin-2-yl)thiazol-2-amine (47).** Compounds **65** (390 mg, 1.39 mmol, 1.0 equiv.) and 1-(3,4-dimethylphenyl)thiourea (250 mg, 1.39 mmol, 1.0 equiv.) were mixed as described in procedure **A-A** in ethanol (10 mL). The pure product (388 mg, 95%) was obtained after workup as a yellow-brown solid. <sup>1</sup>H-NMR (500 MHz, (CD<sub>3</sub>)<sub>2</sub>SO):  $\delta$  = 10.13 (s, 1H), 8.55 (m, 1H), 7.97 (d,  $J$  = 7.85 Hz, 1H), 7.90 (td,  $J$  = 1.89, 7.57 Hz, 1H), 7.48 (m, 2H), 7.42 (d,  $J$  = 1.89 Hz, 1H), 7.31 (ddd,  $J$  = 1.07, 4.80, 7.46 Hz, 1H), 7.10 (d,  $J$  = 8.24 Hz, 1H), 2.23 (s, 3H), 2.18 (s, 3H). <sup>13</sup>C-NMR (126 MHz, (CD<sub>3</sub>)<sub>2</sub>SO):  $\delta$  = 163.8, 152.2, 150.3, 149.4, 139.0, 137.5, 136.7, 130.0, 129.2, 122.7, 120.4, 118.5, 114.7, 106.5, 19.9, 18.8. HR-ESI-MS: calculated for C<sub>16</sub>H<sub>16</sub>N<sub>3</sub>S [ $M+H$ ]<sup>+</sup> 282.1059, found 282.1047.

**Method I: Cell culture and growth-inhibition assay of *P. falciparum*.** *P. falciparum* 3D7 parasites (Wellcome Trust Dundee) and all derived transgenic cell lines (see below) were maintained in continuous culture at 37 °C and an atmosphere consisting of 90% N<sub>2</sub>, 5% O<sub>2</sub>, and 5% CO<sub>2</sub> as described previously<sup>63</sup> with modifications.<sup>64</sup> Parasites were maintained in 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) and 11.9 mM sodium bicarbonate buffered RPMI 1640 medium

supplemented with D-glucose (11 mM), hypoxanthine, Albumax-II (0.5% w/v), and 10 µg/mL gentamicin. Fresh 0+ blood was provided by the Brazilian blood bank ProSangue (Brazil) and in agreement with the ethics committee at ICB-USP. The antiplasmodial effect of the *de novo* synthesized compounds was validated against *P. falciparum* 3D7 strain conducting SYBR Green I (Invitrogen) drug assays as reported.<sup>65,66</sup> Briefly, two-fold serial dilutions of compounds were prepared in 96-well plates in a range of 200 µM to 0.4 µM in triplicate and incubated for 96 h under normal growth conditions using an initial parasitemia of 0.5% and a hematocrit of 2% in a volume of 100 µL per well. Parasite proliferation was determined by measuring DNA load via fluorescence in the wells through addition of 100 µL lysis buffer supplemented with SYBR Green I (0.02% v/v) and incubation for 1 h at room temperature in the dark. Fluorescence was quantified using a CLARIOstar plate reader (BMG Labtech, Germany) at excitation and emission wavelength bands of 485 (± 9) and 530 (± 12) nm, respectively. Focal and gain adjustment was performed using the non-treated controls (highest expected fluorescence signal). Data was acquired via the CLARIOstar (V5.20) and MARS software, manually normalized, and plotted using the nonlinear regression curve fit implemented in GraphPad Prism as described below in more detail (version 7.00 for Windows, GraphPad Software, La Jolla California USA, [www.graphpad.com](http://www.graphpad.com)). Non-treated parasites, highest solvent concentration on parasites, and highest drug concentration in medium were used as controls for maximal growth, solvent control, and native drug fluorescence, respectively.

**Method II: Cell culture and growth-inhibition assay of *P. falciparum*.** *P. falciparum* strain 3D7 was obtained through the MR4 as part of the BEI Resources Repository, NIAID, NIH ([www.mr4.org](http://www.mr4.org)) and strain NF54 was generously supplied by D.A. Fidock (Columbia University). Parasites were cultured in a 2% suspension of human erythrocytes and RPMI 1640 (Sigma) medium supplemented with 27 mM sodium bicarbonate, 11 mM glucose, 5 mM HEPES, 1 mM sodium pyruvate, 0.37 mM hypoxanthine, 0.01 mM thymidine, 10 µg/mL gentamicin, and 0.5% Albumax (Gibco) at 37°C, 5% O<sub>2</sub>/5% CO<sub>2</sub>/90% N<sub>2</sub> atmosphere as previously described.<sup>67,68</sup> Asynchronous cultures of *P. falciparum* 3D7 and NF54 were diluted to 1% parasitemia and treated with DXPS inhibitors at concentrations ranging from 97.7 nM–250 µM. Assays were performed in opaque 96-well plates in 100 µL culture volume. After three days, parasite growth was quantified by measuring DNA content using PicoGreen (Life

Technologies) as described.<sup>69</sup> Fluorescence was measured on a FLUOstar Omega microplate reader (BMG Labtech) at 485 nm excitation and 528 nm emission.

**Method III: Cell culture and growth inhibition assay of *P. falciparum*.** Two laboratory strains of *P. falciparum*, the chloroquine-sensitive 3D7 and the multi-resistant Dd2 (obtained from MR4) were kept in complete culture medium (RPMI 1640, 25 mM HEPES, 2 mM L-glutamine, 50 µg/mL gentamicin and 0.5% w/v AlbuMAX) at 37 °C, 5% CO<sub>2</sub> and 5% oxygen at 2.5% hematocrit with daily change of medium.<sup>67</sup> All compounds were dissolved in DMSO at stock solutions between 25 and 100 mM; the reference drug chloroquine diphosphate (MW: 515.86 g/mol) was diluted in distilled water. Further dilutions were prepared in complete culture medium so that final concentrations of solvent did not interfere with parasite growth. Antiplasmodial activity of the different compounds was tested in a drug-sensitivity assay against the two laboratory strains using the histidine-rich protein 2 (HRP2) assay as described previously.<sup>70,71</sup> In brief: 96-well plates were pre-coated with the different compounds in a threefold dilution before ring-stage parasites were added in complete culture medium at a hematocrit of 1.5% and a parasitemia of 0.05% in a total volume of 225 µL per well. After three days of incubation, plates were frozen until analyzed by HRP2-ELISA. All compounds were evaluated in duplicate in at least two independent experiments. Statistics: The 50% IC<sub>50</sub> was determined by analyzing the nonlinear regression of log concentration–response curves using the drc-package v0.9.0 of R v2.6.1.<sup>72</sup>

**Hep G2 cell culture and viability assay as counter screens.** HepG2, A549 or HEK293 cells (2 x 10<sup>5</sup> cells per well) were seeded in 24-well, flat-bottomed plates. Culturing of cells, incubations and OD measurements were performed as described previously with small modifications.<sup>73</sup> 24 h after seeding the cells, the incubation was started by the addition of compounds in a final DMSO concentration of 1%. The metabolic activity of the living cell mass was determined after 48 h. At least three independent measurements were performed for each compound. The IC<sub>50</sub> values were determined during logarithmic growth using GraphPad Prism software. All experiments were performed at least in triplicate and data reported represent the mean ± SD.

**Metabolic Stability Tests in Human Liver S9 Fraction.** For the evaluation of combined phase I and phase II metabolic stability, the compound (1  $\mu\text{M}$ ) was incubated with 1 mg/mL pooled liver S9 fraction (Xenotech), 2 mM NADPH, 1 mM UDPGA, 10 mM  $\text{MgCl}_2$ , 5 mM GSH and 0.1 mM PAPS at 37 °C for 0, 5, 15, 30 and 60 min. The metabolic stability of testosterone (1  $\mu\text{M}$ ), verapamil (1  $\mu\text{M}$ ) and propranolol (1  $\mu\text{M}$ ) were determined in parallel to confirm the enzymatic activity of the S9 fraction. The incubation was stopped by precipitation of S9 enzymes with two volumes of cold acetonitrile containing internal standard (150 nM diphenhydramine). Samples were stored on ice for 10 min, and precipitated protein was removed by centrifugation (15 min, 4 °C, 4,000 rpm). Concentration of the remaining test compound at the different time points was analyzed by LC-MS/MS (TSQ Quantum Access MAX, Thermo Fisher, Dreieich, Germany) and used to determine half-life ( $t_{1/2}$ ).

**Kinetic solubility determination.** The desired compounds were sequentially diluted in DMSO in a 96-well plate. 3  $\mu\text{L}$  of each well were transferred into another 96-well plate and mixed with 147  $\mu\text{L}$  of PBS. Plates were shaken for 5 min at 600 rpm at room temperature (r.t.), and the absorbance at 620 nm was measured. Absorbance values were normalized by blank subtraction and plotted using GraphPad Prism 8.4.2 (GraphPad Software, San Diego, CA, USA). Solubility (S) was determined based on the First X value of AUC function using a threshold of 0.005.

**IDP rescue assay in *P. falciparum*.** *P. falciparum* cultures were treated as described in **Method II**. For IDP (Isoprenoids.com) rescue experiments, 125  $\mu\text{M}$  IDP was added to the appropriate wells for the duration of the experiment. The well-described DXR inhibitor FSM was used as a positive control.  $\text{IC}_{50}$  values were calculated by nonlinear regression analysis using GraphPad Prism software. All experiments were performed at least in triplicate and data reported represent the mean  $\pm$  SD.

***P. falciparum* sample preparation for mass spectrometry analysis.** *P. falciparum* strain 3D7 was cultured at 37 °C in 30 mL volumes in 100 mm tissue culture dishes (Techno Plastic Products) at 4% hematocrit until >6.5% parasitemia. Cultures were synchronized until >75% of parasites were in ring stage growth, and then treated for 12 h with or without compound **1** at 7.65  $\mu\text{M}$  (5 x the 3D7  $\text{IC}_{50}$ ) in triplicate. Cultures were lysed with 5% saponin, the parasite pellets washed with 1x phosphate-buffered saline (PBS; Gibco), and the pellets stored at -80 °C.

***E. coli* sample preparation for mass spectrometry analysis.** Overnight cultures of *E. coli*  $\Delta$ ToIC were diluted 1:1000 in LB media and grown at 37 °C until reaching mid-logarithmic phase ( $OD_{600} = 0.67\text{--}7.2$ ). Cultures were then treated with or without compound **1** at 48  $\mu\text{M}$  (10 x the  $IC_{50}$ ) in triplicate for 2 h while shaking at 37 °C. For normalization, the  $OD_{600}$  was determined after 2 h of treatment with the inhibitor. Cells were pelleted by centrifugation for 5 min at 3000 x g at 4 °C. The supernatants were removed, and cells were washed twice with 1x PBS. The supernatants were removed, and the pellets stored at  $-80^{\circ}\text{C}$  until analysis.

**LC-MS/MS analysis.** Metabolites were extracted via the addition of glass beads (212–300  $\mu$ ) and 600  $\mu\text{L}$  chilled  $\text{H}_2\text{O}$ : chloroform: methanol (3:5:12 v/v) spiked with PIPES (piperazine-N,N'-bis(2-ethanesulfonic acid) as internal standard. The cells were disrupted with the TissueLyser II instrument (Qiagen) using a microcentrifuge tubes adaptor set pre-chilled for 2 min at 20 Hz. The samples were then centrifuged at 16,000 g at 4 °C, the supernatants collected, and pellet extraction repeated once more. The supernatants were pooled and 300  $\mu\text{L}$  chloroform and 450  $\mu\text{L}$  of chilled water were added to the supernatants. The tubes were vortexed and centrifuged. For *E. coli* samples, the upper layer was transferred to a new tube and dried using a speed-vac. For *P. falciparum* samples, the upper layer was transferred to a 2-mL tube PVDF filter (ThermoFisher, F2520-5) and centrifuged for 5 min at 4,000 x g at 4 °C. The samples were then transferred to new tubes and dried using a speed-vac. Both *E. coli* and *P. falciparum* pellets were re-dissolved in 100  $\mu\text{L}$  of 50% acetonitrile.

For pyruvate analysis from the *E. coli* samples, the LC separation was done on the Shimadzu Nexera II using the Poroshell 120 (Agilent, 2.7  $\mu\text{m}$ , 150 X 2.1 mm) flowing at 0.5 mL/min. The gradient of the mobile phases A (20 mM ammonium acetate, pH 9.8, 5% ACN) and B (100% acetonitrile) was as follow: 85% B for 1 min, to 40% B in 9 min, hold at 40% B for 2 min, then back to 85% B in 0.5 min. For the TCA/glycolysis/pentose phosphate pathway metabolites from the *P. falciparum* samples, the same mobile phases were used on a Luna-NH2 column (3  $\mu\text{m}$ , 150 x 2 mm, Phenomenex) at a flow rate of 1 mL/min. The gradient was as follow: 80% B for 1 min, to 30% B in 6 min, hold at 30% B for 5 min, then back to 80% B in 0.5 min. Finally, for the MEP metabolites, same column and mobile phases were used as the latter, except the gradient was as follow: 60% B for 1 min, to 6% B in 3 min, hold at

6% B for 5 min, then back to 60% B in 0.5 min. The LC system was interfaced with a Sciex QTRAP 6500+ mass spectrometer equipped with a TurbolonSpray (TIS) electrospray ion source. Analyst software (version 1.6.3) was used to control sample acquisition and data analysis. The QTRAP 6500+ mass spectrometer was tuned and calibrated according to the manufacturer's recommendations. Metabolites were detected using MRM transitions that were previously optimized using standards. The instrument was set-up to acquire in negative mode. For quantification, an external standard curve was prepared using a series of standard samples containing different concentrations of metabolites and fixed concentration of the internal standard. The limit of detection for deoxyxylulose 5-phosphate (DOXP), methylerythritol phosphate (MEP), cytidine diphosphate methylerythritol (CDP-ME), and methylerythritol cyclophosphate (MEcPP) was 0.0064  $\mu\text{M}$  for a 10  $\mu\text{L}$  injection volume. The limits of detection for a 5  $\mu\text{L}$  injection volume for the TCA cycle, glycolytic and pentose phosphate metabolites were as follows: aconitate, malate, succinate = 0.31  $\mu\text{M}$ ; glucose-6-phosphate and glycerol-3-phosphate = 0.78  $\mu\text{M}$ ; citrate, glucose-1-phosphate and fructose-6-phosphate = 1.56  $\mu\text{M}$ ; ribose-5-phosphate and ribulose-5-phosphate = 2.34  $\mu\text{M}$ ; 2-phospho glyceric acid, 3-phospho glyceric acid and lactate = 3.12  $\mu\text{M}$ ; fumarate, pyruvate, phosphoenolpyruvate and sedoheptulose-7-phosphate = 6.25  $\mu\text{M}$ . T-tests were used to test for significance between untreated (UNT) and drug-treated bacteria (Prism).

**Generation of transgenic parasite lines.** The open reading frame (ORF) encoding 1-deoxy-D-xylulose-5-phosphate synthase (*PfDXPS*; PF3D7\_1337200) was amplified from genomic DNA isolated from unsynchronized *P. falciparum* 3D7 cultures using the Platinum PCR SuperMix High Fidelity (Invitrogen). Forward and reverse primers contained restriction sites for *KpnI* and *AvrII* in sense and antisense orientation, respectively (Oligonucleotides *pfDXPS-Kpn1-S*: GAGAGGTACCATGATTTTTAATTATGTGTTTTTAAGAAC; *pfDXPSmyc-Avr2-AS*: GAGACCTAGGTTACAGGTCCTCCTCTGAGATCAGCTTCTGCTCGCCTGTAGGAT TATTTTTAAGATAATTTTTAATTCTATTGAC). PCR products and the transfection vector pARL1a-hDHFR *PfARG*-GFP were digested with *KpnI* and *AvrII*, purified (Gel extraction Kit and PCR purification kit; QIAGEN), and cloned into the transfection vector yielding the construct for overexpression pARL-DXS-strep. XL 10-Gold *E. coli* ultra-competent cells (Agilent Technologies) were transformed with the generated

construct to amplify the plasmid and colonies checked using restriction analysis and sequencing of the plasmid. Bacterial clones carrying the overexpression construct and the empty pARL1a vector (MOCK plasmid) were amplified in over-night cultures, isolated (Plasmid Maxi Kit; QIAGEN) and subsequently used to transfect *P. falciparum* 3D7 ring-stage parasites, as already described.<sup>74</sup> Briefly, 120 µg plasmid DNA was centrifuged, air dried, the pellet resuspended in TE-buffer and cytomix reagent and mixed with the *P. falciparum* culture. Parasites were transfected using electroporation and selected with WR 99210. The generation and characterization of the *Pf*TPK overexpression cell line has been reported before.<sup>75</sup>

**Quantification of DXPS overexpression in *P. falciparum* 3D7 transgenic cell lines.** Total RNA was isolated from *P. falciparum* using TRIzol Reagent (Invitrogen). First-strand cDNA synthesis was prepared using a RevertAid™H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) from a total of 1 µg of RNA. Real-time PCR was performed with 1 µL of cDNA, 7.5 µL SYBR green fast master mix (Applied Biosystems), 0.45 µL (10 µM) each of the forward and reverse primers and 5.6 µL DEPC-treated water, on an Applied Biosystems® QuantStudio® 3 Real-Time PCR System (Thermo Fisher Scientific). After cycling, melting curve analysis was performed. The relative transcription levels of MOCK (control) and the DXS (DXS-pARL) were determined by the  $\Delta\Delta CT$  method.<sup>76</sup> Target transcription levels were normalized to the housekeeping gene, *Pf*Aldolase, as endogenous control reference as reported before.<sup>75</sup> Sequences of primers were as follows: *Pf*Aldolase forward: tgtaccaccagccttaccag; reverse: ttcttgccatgtgttcaat; DXS-pARL forward: tcagtggagagggtgaaggt; reverse: gttggccatggaacaggtag. Three technical replicates from three biological replicates were performed for each experiment. Expression was found to be six times higher in comparison to the MOCK control cell line.

**Growth-inhibition assay of transgenic *P. falciparum* cell lines.** *P. falciparum* cultures were treated as described in **Method I**. For target specification towards *Pf*TPK and *Pf*DXPS, compounds were tested in comparison against overexpressing cell lines and the respective MOCK line containing only the transfected vector backbone. Parasite viability/proliferation was determined. Analysis of the IC<sub>50</sub> values and interpretation of the curves was performed as described in **Growth inhibition and IDP rescue assays of *E. coli***.

**Nonlinear regression fit and analysis of dose-response drug assays of compound 1, 2 and 3.** Nonlinear regression as implemented in GraphPad Prism 7.00 (log(inhibitor) vs. response – Variable slope (four parameters)) was used to fit the measured data to interpolate the IC<sub>50</sub> value from the curve. No specific model was applied. Data was pre-processed by normalizing according to the following formula

$$y_{normalized} = \frac{y - y_{minimum}}{y_{maximum} - y_{minimum}}$$

Where  $y$  is the fluorescence signal in each well,  $y_{minimum}$  the background fluorescence, and  $y_{maximum}$  the highest measured fluorescence signal in the untreated wells. Drug concentrations (in  $\mu\text{M}$ ) were transformed to the  $\log(10)$  of the values. Means of each independent experiment were plotted as individual values, and the SD of the mean from the means shown as error bars. The 95% CI is indicated as measure of error for the calculated IC<sub>50</sub>s. In case of transgenic cell lines, the built-in comparison Akaike's Information Criteria (AICc) method was applied to the Log IC<sub>50</sub>s of the different cell lines. The test calculates a percentage probability of the simpler model "Log IC<sub>50</sub> is the same for datasets" being correct. The test for *homoscedasticity* was performed to confirm if no weighting of values was appropriate.

**Computational Methods, homology model of PfDXPS.** As template for the homology model of PfDXPS the crystal structure of DXPS from *D. radiodurans* (2o1x, chain B) was used because the corresponding region in the active center of the likewise related DXPS from *E. coli* (2o1s) strongly deviates in comparison to the DXPS from *M. tuberculosis*. Subsequently, the actual homology model was generated by the SWISS-MODEL web service, yielding a QMEAN4 value of -6.43.<sup>77,78</sup> The magnesium atom was added manually using the coordinates from the template.

## Acknowledgments

We thank Selina Wolter, Jeannine Jung, Dennis Jener, Andreas M. Kany, and Jörg Hauptenthal at the HIPS for performing the cytotoxicity, solubility, and metabolic stability assays and Ravindra P. Jumde for proofreading of the manuscript.

## Associated Content

Supporting Information Available: Additional figures of homology model, more detailed assay results and tables with structures, inhibition data, t-tests and docking scores. NMR, HRMS and HPLC spectra of all tested compounds. *PfDXPS* homology model based on structure 2O1X. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## Author information

Corresponding author:

\*anna.hirsch@helmholtz-hips.de

Author contributions:

†These authors contributed equally.

Notes:

The authors declare no competing financial interest.

This work has been supported by The Netherlands Organisation for Scientific Research (LIFT grant: 731.015.414, A. K. H. H. and R. M. G.) and the Helmholtz Association's Initiative and Networking fund. D. Z. acknowledges support from the CSC Fellowship. A.O.J. is supported by NIH/NIAID R01-AI103280, R21-AI123808, and R21-AI130584, and A.O.J. is an Investigator in the Pathogenesis of Infectious Diseases (PATH) of the Burroughs Wellcome Fund. Further, the authors acknowledge the financial support provided by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) via the grants 2015/26722-8, 2016/24790-9, 2017/03966-4, 2018/08820-0, 2020/12277-0.

## Abbreviations used

ACTs, artemisinin-based combination therapies; DIPEA, *N,N*-diisopropylethylamine; DMADP, dimethylallyl diphosphate; DOXP, 1-deoxy-D-xylulose 5-phosphate; DXPS,

1-deoxy-d-xylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; ERAD, endoplasmic reticulum assisted degradation; FabI, Enoyl-[acyl-carrier-protein] reductase; FAS, fatty acid biosynthesis pathway; FSM, fosmidomycin; GAP, glyceraldehyde 3-phosphate; Hep G2, human hepatocytes; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; HRP2, histidine-rich protein 2; HTS, high-throughput screening; Huh7, human hepatoma cells; IDP, isopentenyl diphosphate; KasA,  $\beta$ -ketoacyl ACP synthase; LBVS, ligand-based virtual screening; MEP, methylerythritol 4-phosphate; NaHMDS, sodium bis(trimethylsilyl)amide; OAc, acetate; ORF, open reading frame; PBMN, polymyxin B nonapeptide; PPB, polypharmacology browser; TBDMS, *tert*-butyldimethylsilyl chloride; TBME, *tert*-butyl methyl ether; TCA, tricarboxylic acid; ThDP, thiamine diphosphate; TPK, thiamine pyrophosphokinase; VCP, Valosin-containing protein; WHO, World Health Organization.

## References

- (1) World Health Organization. *World Malaria Report 2021*, **2021**.
- (2) World Health Organization. *Status report on artemisinin resistance and ACT*, **2018**.
- (3) Yeh, E.; DeRisi, J. L. Chemical rescue of malaria parasites lacking an apicoplast defines organelle function in blood-stage *Plasmodium falciparum*. *PLoS biology* **2011**, *9* (8), e1001138.
- (4) Frank, A.; Groll, M. The methylerythritol phosphate pathway to isoprenoids. *Chem. Rev.* **2017**, *117* (8), 5675–5703.
- (5) Gierse, R. M.; Redeem, E.; Diamanti, E.; Wrenger, C.; Groves, M. R.; Hirsch, A. K. DXS as a target for structure-based drug design. *Future Med. Chem.* **2017**, *9* (11), 1277–1294.
- (6) Hale, I.; O'Neill, P. M.; Berry, N. G.; Odom, A.; Sharma, R. The MEP pathway and the development of inhibitors as potential anti-infective agents. *Med. Chem. Commun.* **2012**, *3* (4), 418.
- (7) Fernandes, J. F.; Lell, B.; Agnandji, S. T.; Obiang, R. M.; Bassat, Q.; Kremsner, P. G.; Mordmüller, B.; Grobusch, M. P. Fosmidomycin as an antimalarial drug: a meta-analysis of clinical trials. *Future Microbiol.* **2015**, *10* (8), 1375–1390.
- (8) Umeda, T.; Tanaka, N.; Kusakabe, Y.; Nakanishi, M.; Kitade, Y.; Nakamura, K. T. Molecular basis of fosmidomycin's action on the human malaria parasite *Plasmodium falciparum*. *Sci. Rep.* **2011**, *1*, 9.
- (9) Du, Q.; Wang, H.; Xie, J. Thiamin (vitamin B1) biosynthesis and regulation: a rich source of antimicrobial drug targets? *Int. J. Biol. Sci.* **2011**, *7* (1), 41–52.
- (10) Hill, R. E.; Himmeldirk, K.; Kennedy, I. A.; Pauloski, R. M.; Sayer, B. G.; Wolf, E.; Spenser, I. D. The biogenetic anatomy of vitamin B6. A <sup>13</sup>C NMR investigation of the biosynthesis of pyridoxol in *Escherichia coli*. *J. Biol. Chem.* **1996**, *271* (48), 30426–30435.
- (11) Müller, I. B.; Hyde, J. E.; Wrenger, C. Vitamin B metabolism in *Plasmodium falciparum* as a source of drug targets. *Trends. Mol. Med.* **2010**, *26* (1), 35–43.
- (12) Di, Z.; Johannsen, S.; Masini, T.; Simonin, C.; Haupenthal, J.; Andreas, A.; Awale, M.; Gierse, R. M.; van der Laan, T.; van der Vlag, R.; Nasti, R.; Poizat, M.;

Buhler, E.; Reiling, N.; Müller, R.; Fischer, M.; Reymond, J.-L.; Hirsch, A. K. H. Discovery of novel drug-like antitubercular hits targeting the MEP pathway enzyme DXPS by strategic application of ligand-based virtual screening. *ChemRxiv* **2022**.

(13) Goswami, A. M. Computational analysis, structural modeling and ligand binding site prediction of Plasmodium falciparum 1-deoxy-d-xylulose-5-phosphate synthase. *Comput. Biol. Chem* **2017**, *66*, 1–10.

(14) Masini, T.; Lacy, B.; Monjas, L.; Hawksley, D.; Voogd, A. R. de; Illarionov, B.; Iqbal, A.; Leeper, F. J.; Fischer, M.; Kontoyianni, M.; Hirsch, A. K. H. Validation of a homology model of Mycobacterium tuberculosis DXS: rationalization of observed activities of thiamine derivatives as potent inhibitors of two orthologues of DXS. *Org. Biomol. Chem.* **2015**, *13* (46), 11263–11277.

(15) Bajorath, J. Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. *J. Chem. Inf. Model.* **2001**, *41* (2), 233–245.

(16) Masini, T.; Pilger, J.; Kroezen, B. S.; Illarionov, B.; Lottmann, P.; Fischer, M.; Griesinger, C.; Hirsch, A. K. H. De novo fragment-based design of inhibitors of DXS guided by spin-diffusion-based NMR spectroscopy. *Chem. Sci.* **2014**, *5* (9), 3543–3551.

(17) Princeton Biomolecular Research. *Express Collection*. [http://www.princetonbio.com/products/diverse\\_targeted\\_screening\\_compounds](http://www.princetonbio.com/products/diverse_targeted_screening_compounds) (accessed 2015).

(18) Awale, M.; Reymond, J.-L. The polypharmacology browser: a web-based multi-fingerprint target prediction tool using ChEMBL bioactivity data. *J. Cheminformatics* **2017**, *9* (1), 11.

(19) Awale, M.; Reymond, J.-L. *The Polypharmacology Browser PPB2: Target Prediction Combining Nearest Neighbors with Machine Learning*, **2018**.

(20) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; Veij, M. de; Félix, E.; Magariños, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; Gordillo-Marañón, M.; Hunter, F.; Junco, L.; Mugumbate, G.; Rodriguez-Lopez, M.; Atkinson, F.; Bosc, N.; Radoux, C. J.; Segura-Cabrera, A.; Hersey, A.; Leach, A. R. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* **2019**, *47* (D1), D930-D940.

(21) Plouffe, D.; Brinker, A.; McNamara, C.; Henson, K.; Kato, N.; Kuhlen, K.; Nagle, A.; Adrián, F.; Matzen, J. T.; Anderson, P.; Nam, T.-G.; Gray, N. S.; Chatterjee, A.; Janes, J.; Yan, S. F.; Trager, R.; Caldwell, J. S.; Schultz, P. G.; Zhou, Y.; Winzeler, E. A. In silico activity profiling reveals the mechanism of action of antimalarials discovered in a high-throughput screen. *PNAS* **2008**, *105* (26), 9059–9064.

(22) Wilkinson, C.; McPhillie, M. J.; Zhou, Y.; Woods, S.; Afanador, G. A.; Rawson, S.; Khaliq, F.; Prigge, S. T.; Roberts, C. W.; Rice, D. W.; McLeod, R.; Fishwick, C. W.; Muench, S. P. The benzimidazole based drugs show good activity against *T. gondii* but poor activity against its proposed enoyl reductase enzyme target. *Bioorg. Med. Chem. Lett.* **2014**, *24* (3), 911–916.

(23) Mehboob, S.; Hevener, K. E.; Truong, K.; Boci, T.; Santarsiero, B. D.; Johnson, M. E. Structural and enzymatic analyses reveal the binding mode of a novel series of *Francisella tularensis* enoyl reductase (FabI) inhibitors. *J. Med. Chem.* **2012**, *55* (12), 5933–5941.

(24) Ren, J.; Mistry, T. L.; Su, P.-C.; Mehboob, S.; Demissie, R.; Fung, L. W.-M.; Ghosh, A. K.; Johnson, M. E. Determination of absolute configuration and binding efficacy of benzimidazole-based FabI inhibitors through the support of electronic circular dichroism and MM-GBSA techniques. *Bioorg. Med. Chem. Lett.* **2018**, *28* (11), 2074–2079.

(25) Mehboob, S.; Song, J.; Hevener, K. E.; Su, P.-C.; Boci, T.; Brubaker, L.; Truong, L.; Mistry, T.; Deng, J.; Cook, J. L.; Santarsiero, B. D.; Ghosh, A. K.; Johnson, M. E. Structural and biological evaluation of a novel series of benzimidazole inhibitors of *Francisella tularensis* enoyl-ACP reductase (FabI). *Bioorg. Med. Chem. Lett.* **2015**, *25* (6), 1292–1296.

(26) Freundlich, J. S.; Wang, F.; Tsai, H.-C.; Kuo, M.; Shieh, H.-M.; Anderson, J. W.; Nkrumah, L. J.; Valderramos, J.-C.; Yu, M.; Kumar, T. R. S.; Valderramos, S. G.; Jacobs, W. R.; Schiehsler, G. A.; Jacobus, D. P.; Fidock, D. A.; Sacchettini, J. C. X-ray structural analysis of *Plasmodium falciparum* enoyl acyl carrier protein reductase as a pathway toward the optimization of triclosan antimalarial efficacy. *J. Biol. Chem.* **2007**, *282* (35), 25436–25444.

(27) Hevener, K. E.; Mehboob, S.; Su, P.-C.; Truong, K.; Boci, T.; Deng, J.; Ghassemi, M.; Cook, J. L.; Johnson, M. E. Discovery of a novel and potent class of *F. tularensis*

enoyl-reductase (FabI) inhibitors by molecular shape and electrostatic matching. *J. Med. Chem.* **2012**, *55* (1), 268–279.

(28) Rana, P.; Ghouse, S. M.; Akunuri, R.; Madhavi, Y. V.; Chopra, S.; Nanduri, S. FabI (enoyl acyl carrier protein reductase) - A potential broad spectrum therapeutic target and its inhibitors. *Eur. J. Med. Chem.* **2020**, *208*, 112757.

(29) Wickramasinghe, S. R.; Inglis, K. A.; Urch, J. E.; Müller, S.; van Aalten, D. M. F.; Fairlamb, A. H. Kinetic, inhibition and structural studies on 3-oxoacyl-ACP reductase from *Plasmodium falciparum*, a key enzyme in fatty acid biosynthesis. *Biochem.* **2006**, *393* (Pt 2), 447–457.

(30) Fernández de Luco, J.; Recio-Balsells, A. I.; Ghiano, D. G.; Bortolotti, A.; Belardinelli, J. M.; Liu, N.; Hoffmann, P.; Lherbet, C.; Tonge, P. J.; Tekwani, B.; Morbidoni, H. R.; Labadie, G. R. Exploring the chemical space of 1,2,3-triazolyl triclosan analogs for discovery of new antileishmanial chemotherapeutic agents. *RSC Med. Chem.* **2021**, *12* (1), 120–128.

(31) Lu, X.; Huang, K.; You, Q. Enoyl acyl carrier protein reductase inhibitors: a patent review (2006 - 2010). *Expert Opin. Ther. Pat.* **2011**, *21* (7), 1007–1022.

(32) Marreddy, R. K. R.; Wu, X.; Sapkota, M.; Prior, A. M.; Jones, J. A.; Sun, D.; Hevener, K. E.; Hurdle, J. G. The fatty acid synthesis protein enoyl-ACP reductase II (FabK) is a target for narrow-spectrum antibacterials for *Clostridium difficile* infection. *ACS Infect. Dis.* **2019**, *5* (2), 208–217.

(33) Zitko, J.; Doležal, M. Enoyl acyl carrier protein reductase inhibitors: an updated patent review (2011 - 2015). *Expert Opin. Ther. Pat.* **2016**, *26* (9), 1079–1094.

(34) Shears, M. J.; Botté, C. Y.; McFadden, G. I. Fatty acid metabolism in the *Plasmodium* apicoplast: Drugs, doubts and knockouts. *Mol. Biochem. Parasitol.* **2015**, *199* (1-2), 34–50.

(35) van Schaijk, B. C. L.; Kumar, T. R. S.; Vos, M. W.; Richman, A.; van Gemert, G.-J.; Li, T.; Eappen, A. G.; Williamson, K. C.; Morahan, B. J.; Fishbaugher, M.; Kennedy, M.; Camargo, N.; Khan, S. M.; Janse, C. J.; Sim, K. L.; Hoffman, S. L.; Kappe, S. H. I.; Sauerwein, R. W.; Fidock, D. A.; Vaughan, A. M. Type II fatty acid biosynthesis is essential for *Plasmodium falciparum* sporozoite development in the midgut of *Anopheles* mosquitoes. *Eukaryot. Cell* **2014**, *13* (5), 550–559.

- (36) Vaughan, A. M.; O'Neill, M. T.; Tarun, A. S.; Camargo, N.; Phuong, T. M.; Aly, A. S. I.; Cowman, A. F.; Kappe, S. H. I. Type II fatty acid synthesis is essential only for malaria parasite late liver stage development. *Cell. Microbiol.* **2009**, *11* (3), 506–520.
- (37) Meissner, A.; Boshoff, H. I.; Vasan, M.; Duckworth, B. P.; Barry, C. E.; Aldrich, C. C. Structure-activity relationships of 2-aminothiazoles effective against *Mycobacterium tuberculosis*. *Bioorg. Med. Chem.* **2013**, *21* (21), 6385–6397.
- (38) Makam, P.; Kannan, T. 2-Aminothiazole derivatives as antimycobacterial agents: Synthesis, characterization, in vitro and in silico studies. *Eur. J. Med. Chem.* **2014**, *87*, 643–656.
- (39) Al-Balas, Q.; Anthony, N. G.; Al-Jaidi, B.; Alnimr, A.; Abbott, G.; Brown, A. K.; Taylor, R. C.; Besra, G. S.; McHugh, T. D.; Gillespie, S. H.; Johnston, B. F.; Mackay, S. P.; Coxon, G. D. Identification of 2-aminothiazole-4-carboxylate derivatives active against *Mycobacterium tuberculosis* H37Rv and the beta-ketoacyl-ACP synthase mtFabH. *PLoS one* **2009**, *4* (5), e5617.
- (40) Waters, N. Functional characterization of the acyl carrier protein (PfACP) and beta-ketoacyl ACP synthase III (PfKASIII) from *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **2002**, *123* (2), 85–94.
- (41) Bhuniya, D.; Mukkavilli, R.; Shivahare, R.; Launay, D.; Dere, R. T.; Deshpande, A.; Verma, A.; Vishwakarma, P.; Moger, M.; Pradhan, A.; Pati, H.; Gopinath, V. S.; Gupta, S.; Puri, S. K.; Martin, D. Aminothiazoles: Hit to lead development to identify antileishmanial agents. *Eur. J. Med. Chem.* **2015**, *102*, 582–593.
- (42) Bursavich, M. G.; Parker, D. P.; Willardsen, J. A.; Gao, Z.-H.; Davis, T.; Ostanin, K.; Robinson, R.; Peterson, A.; Cimborá, D. M.; Zhu, J.-F.; Richards, B. 2-Anilino-4-aryl-1,3-thiazole inhibitors of valosin-containing protein (VCP or p97). *Bioorg. Med. Chem. Lett.* **2010**, *20* (5), 1677–1679.
- (43) Harbut, M. B.; Patel, B. A.; Yeung, B. K. S.; McNamara, C. W.; Bright, A. T.; Ballard, J.; Supek, F.; Golde, T. E.; Winzeler, E. A.; Diagana, T. T.; Greenbaum, D. C. Targeting the ERAD pathway via inhibition of signal peptide peptidase for antiparasitic therapeutic design. *PNAS* **2012**, *109* (52), 21486–21491.

- (44) Chung, D.-W. D.; Ponts, N.; Prudhomme, J.; Rodrigues, E. M.; Le Roch, K. G. Characterization of the ubiquitylating components of the human malaria parasite's protein degradation pathway. *PLoS one* **2012**, *7* (8), e43477.
- (45) Chou, T.-F.; Brown, S. J.; Minond, D.; Nordin, B. E.; Li, K.; Jones, A. C.; Chase, P.; Porubsky, P. R.; Stoltz, B. M.; Schoenen, F. J.; Patricelli, M. P.; Hodder, P.; Rosen, H.; Deshaies, R. J. Reversible inhibitor of p97, DBeQ, impairs both ubiquitin-dependent and autophagic protein clearance pathways. *PNAS* **2011**, *108* (12), 4834–4839.
- (46) Bilodeau, M. T.; Rodman, L. D.; McGaughey, G. B.; Coll, K. E.; Koester, T. J.; Hoffman, W. F.; Hungate, R. W.; Kendall, R. L.; McFall, R. C.; Rickert, K. W.; Rutledge, R. Z.; Thomas, K. A. The discovery of N-(1,3-thiazol-2-yl)pyridin-2-amines as potent inhibitors of KDR kinase. *Bioorg. Med. Chem. Lett.* **2004**, *14* (11), 2941–2945.
- (47) Rödl, C. B.; Vogt, D.; Kretschmer, S. B. M.; Ihlefeld, K.; Barzen, S.; Brüggerhoff, A.; Achenbach, J.; Proschak, E.; Steinhilber, D.; Stark, H.; Hofmann, B. Multi-dimensional target profiling of N,4-diaryl-1,3-thiazole-2-amines as potent inhibitors of eicosanoid metabolism. *Eur. J. Med. Chem.* **2014**, *84*, 302–311.
- (48) Wu, S.; Wang, L.; Guo, W.; Liu, X.; Liu, J.; Wei, X.; Fang, B. Analogues and derivatives of oncrasin-1, a novel inhibitor of the C-terminal domain of RNA polymerase II and their antitumor activities. *J. Med. Chem.* **2011**, *54* (8), 2668–2679.
- (49) Bridgford, J. L.; Xie, S. C.; Cobbold, S. A.; Pasaje, C. F. A.; Herrmann, S.; Yang, T.; Gillett, D. L.; Dick, L. R.; Ralph, S. A.; Dogovski, C.; Spillman, N. J.; Tilley, L. Artemisinin kills malaria parasites by damaging proteins and inhibiting the proteasome. *Nat. Commun.* **2018**, *9* (1), 3801.
- (50) Shibeshi, M. A.; Kifle, Z. D.; Atnafie, S. A. Antimalarial drug resistance and novel targets for antimalarial drug discovery. *Infect. Drug. Resist.* **2020**, *13*, 4047–4060.
- (51) Urgin, K.; Jida, M.; Ehrhardt, K.; Müller, T.; Lanzer, M.; Maes, L.; Elhabiri, M.; Davioud-Charvet, E. Pharmacomodulation of the antimalarial plasmodione: synthesis of biaryl- and N-arylalkylamine analogues, antimalarial activities and physicochemical properties. *Molecules (Basel, Switzerland)* **2017**, *22* (1), 161.
- (52) Wu, G.; Yin, W.; Shen, H. C.; Huang, Y. One-pot synthesis of useful heterocycles in medicinal chemistry using a cascade strategy. *Green Chem.* **2012**, *14* (3), 580.

- (53) Snyder, S. A.; Zografos, A. L.; Lin, Y. Total synthesis of resveratrol-based natural products: a chemoselective solution. *Angew. Chem. Int. Ed Engl.* **2007**, *46* (43), 8186–8191.
- (54) To, T. A.; Vo, Y. H.; Nguyen, A. T.; Phan, A. N. Q.; Truong, T.; Phan, N. T. S. A new route to substituted furocoumarins via copper-catalyzed cyclization between 4-hydroxycoumarins and ketoximes. *Org. Biomol. Chem.* **2018**, *16* (28), 5086–5089.
- (55) Ivanov, I.; Nikolova, S.; Statkova-Abeghe, S. Efficient one-pot Friedel–Crafts Acylation of benzene and its derivatives with unprotected aminocarboxylic acids in polyphosphoric acid. *Synth. Commun.* **2006**, *36* (10), 1405–1411.
- (56) Wang, S.; Shen, Y.-B.; Li, L.-F.; Qiu, B.; Yu, L.; Liu, Q.; Xiao, J. N-Alkylation-initiated redox-neutral 5 + 2 annulation of 3-alkylindoles with o-aminobenzaldehydes: access to indole-1,2-fused 1,4-benzodiazepines. *Org. Lett.* **2019**, *21* (22), 8904–8908.
- (57) Giraud, F.; Loge, C.; Pagniez, F.; Crepin, D.; Barres, S.; Picot, C.; Le Pape, P.; Le Borgne, M. Design, synthesis and evaluation of 3-(imidazol-1-ylmethyl)indoles as antileishmanial agents. Part II. *J. Enzyme Inhib. Med. Chem.* **2009**, *24* (5), 1067–1075.
- (58) Somei, M.; Yamada, F.; Kobayashi, K.; Shimizu, A.; Aoki, N. A synthesis method of indole-3-methanamine and/or gramine from indole-3-carboxaldehyde, and its application for the synthesis of brassinin, its 4-substituted analogus, and 1,3,4,5-tetrahydropyrrolo[4,3,2-de]quinoline. *Heterocycles* **1993**, *36* (12), 2783.
- (59) Kulkarni, A.; Soni, I.; Kelkar, D. S.; Dharmaraja, A. T.; Sankar, R. K.; Beniwal, G.; Rajendran, A.; Tamhankar, S.; Chopra, S.; Kamat, S. S.; Chakrapani, H. Chemoproteomics of an indole-based quinone epoxide identifies druggable vulnerabilities in vancomycin-resistant *Staphylococcus aureus*. *J. Med. Chem.* **2019**, *62* (14), 6785–6795.
- (60) Caracciolo Torchiarolo, G.; Iacoangeli, T.; Furlotti, G.; (None). Process for the preparation of 1-benzyl-3-hydroxymethyl-1H-indazole and its derivatives and required magnesium intermediates. WO2010EP60937 20100728.
- (61) Matsumoto, K.; Arai, S.; Nishida, A. Formal synthesis of (±) —quebrachamine through regio- and stereoselective hydrocyanation of arylallene. *Tetrahedron* **2018**, *74* (23), 2865–2870.

- (62) Kesicki, E. A.; Bailey, M. A.; Ovechkina, Y.; Early, J. V.; Alling, T.; Bowman, J.; Zuniga, E. S.; Dalai, S.; Kumar, N.; Masquelin, T.; Hipskind, P. A.; Odingo, J. O.; Parish, T. Synthesis and evaluation of the 2-aminothiazoles as anti-tubercular agents. *PloS one* **2016**, *11* (5), e0155209.
- (63) Collins, W. E.; Moss, D.; Chin, W. The Continuous Cultivation of Plasmodium Fragile by the Method of Trager-Jensen. *Am. J. Trop. Med. Hyg.* **1979**, *28* (3), 591–592.
- (64) Das Gupta, R.; Krause-Ihle, T.; Bergmann, B.; Müller, I. B.; Khomutov, A. R.; Müller, S.; Walter, R. D.; Lüersen, K. 3-Aminoxy-1-aminopropane and derivatives have an antiproliferative effect on cultured Plasmodium falciparum by decreasing intracellular polyamine concentrations. *Antimicrob. Agents Chemother.* **2005**, *49* (7), 2857–2864.
- (65) Smilkstein, M.; Sriwilaijaroen, N.; Kelly, J. X.; Wilairat, P.; Riscoe, M. Simple and inexpensive fluorescence-based technique for high-throughput antimalarial drug screening. *Antimicrob. Agents Chemother.* **2004**, *48* (5), 1803–1806.
- (66) Meissner, K. A.; Kronenberger, T.; Maltarollo, V. G.; Trossini, G. H. G.; Wrenger, C. Targeting the Plasmodium falciparum plasmepsin V by ligand-based virtual screening. *Chem. Biol. Drug Des.* **2019**, *93* (3), 300–312.
- (67) Trager, W.; Jensen, J. B. Human malaria parasites in continuous culture. *Science (New York, N.Y.)* **1976**, *193* (4254), 673–675.
- (68) Zhang, B.; Watts, K. M.; Hodge, D.; Kemp, L. M.; Hunstad, D. A.; Hicks, L. M.; Odom, A. R. A second target of the antimalarial and antibacterial agent fosmidomycin revealed by cellular metabolic profiling. *Biochemistry* **2011**, *50* (17), 3570–3577.
- (69) Coley, P. D.; Corbett, Y.; Cubilla, L.; Ortega-Barria, E.; Herrera, L.; Capson, T. L.; Gonzalez, J.; Kursar, T. A.; Romero, L. I.U.Z. A Novel DNA-Based Microfluorimetric Method to Evaluate Antimalarial Drug Activity. *Am. J. Trop. Med. Hyg.* **2004**, *70* (2), 119–124.
- (70) Noedl, H.; Bronnert, J.; Yingyuen, K.; Attlmayr, B.; Kollaritsch, H.; Fukuda, M. Simple histidine-rich protein 2 double-site sandwich enzyme-linked immunosorbent assay for use in malaria drug sensitivity testing. *Antimicrob. Agents Chemother.* **2005**, *49* (8), 3575–3577.

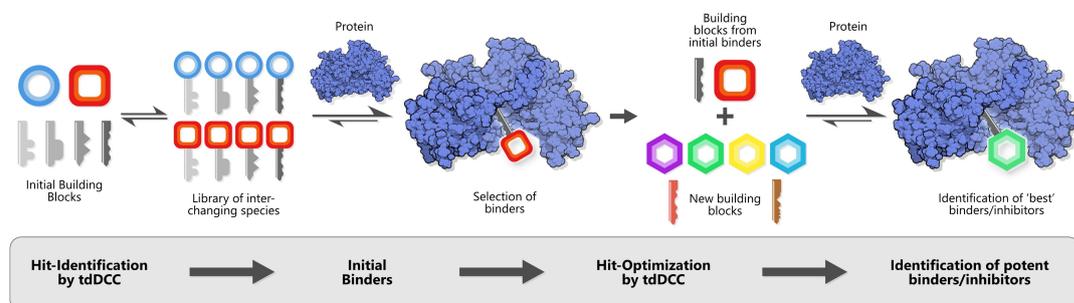
- (71) Carvalho, L. P. de; Sandri, T. L.; José Tenório de Melo, E.; Fendel, R.; Kremsner, P. G.; Mordmüller, B.; Held, J. Ivermectin Impairs the Development of Sexual and Asexual Stages of Plasmodium falciparum In Vitro. *Antimicrob. Agents Chemother. (Bethesda)* **2019**, 63 (8).
- (72) Vienna Austria R Foundation for Statical Computing. *A language and environment for statistical computing*, Vol. 1, **2008**.
- (73) Hauptenthal, J.; Baehr, C.; Zeuzem, S.; Piiper, A. RNase A-like enzymes in serum inhibit the anti-neoplastic activity of siRNA targeting polo-like kinase 1. *Int. J. Cancer* **2007**, 121 (1), 206–210.
- (74) Müller, I. B.; Knöckel, J.; Eschbach, M.-L.; Bergmann, B.; Walter, R. D.; Wrenger, C. Secretion of an acid phosphatase provides a possible mechanism to acquire host nutrients by Plasmodium falciparum. *Cell. Microbiol.* **2010**, 12 (5), 677–691.
- (75) Chan, X. W. A.; Wrenger, C.; Stahl, K.; Bergmann, B.; Winterberg, M.; Müller, I. B.; Saliba, K. J. Chemical and genetic validation of thiamine utilization as an antimalarial drug target. *Nat. Commun.* **2013**, 4 (1), 2060.
- (76) Schmittgen, T. D.; Livak, K. J. Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*, 3(6), 1101-1108. *Nat. Protoc.* **2008**, 3 (6), 1101–1108.
- (77) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; Beer, T. A. P. de; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, 46 (W1), 296-303.
- (78) Benkert, P.; Biasini, M.; Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **2011**, 27 (3), 343–350.

### 2.3. IDENTIFICATION OF THREE NEW INHIBITOR CLASSES AGAINST *PLASMODIUM FALCIPARUM*



## 2.4 Hit-Optimization Using Target-Directed DCC: Development of Inhibitors of the Anti-Infective Target DXS

The following chapter was published in RSC Chemical Science under the terms of the creative commons license CC BY-NC 3.0 and is printed in this thesis without any changes.<sup>1</sup>



### Title

Hit-Optimization Using Target-Directed Dynamic Combinatorial Chemistry: Development of Inhibitors of the Anti-Infective Target 1-Deoxy-D-Xylulose-5-Phosphate Synthase

### Authors

Ravindra P. Jumde, Melissa Guardigni, Robin M. Gierse, Alaa Alhayek, Di Zhu, Zhoor Hamed, Sandra Johannsen, Walid A. M. Elgaher, Philip J. Neusens, Christian Nehls, Jörg Haupenthal, Norbert Reiling and Anna K.H. Hirsch

### Contributions

R. P. Jumde conceived the study, drafted the manuscript and was involved in most of the experimental work. M. Guardigni was involved in the tdDCC experiments and synthesis of hit compounds. R. M. Gierse and A. Alhayek expressed and purified the enzymes and performed, together with Z. Hamid, enzyme activity assays. D. Zhu and R. M. Gierse performed the MOI study. S. Johannsen performed the binding affinity assays by SPR and W. A. M. Elgaher analyzed the SPR data. J. Haupenthal coordinated the tests for antibacterial activity. C. Nehls performed the DLS experiments. N. Reiling conducted and evaluated the antiTB activity assays. P. J. Neusens synthesized bioisostere 41. A. K. H. Hirsch conceived the project and supervised the research. All authors contributed to the writing and editing of the manuscript.

### Bibliographic Details

RSC Chemical Science  
Issue 22, Pages 7775 - 7785  
22 July 2021  
DOI: 10.1039/D1SC00330E

## EDGE ARTICLE

Cite this: *Chem. Sci.*, 2021, 12, 7775

All publication charges for this article have been paid for by the Royal Society of Chemistry

## Hit-optimization using target-directed dynamic combinatorial chemistry: development of inhibitors of the anti-infective target 1-deoxy-D-xylulose-5-phosphate synthase†

Ravindra P. Jumde,<sup>a</sup> Melissa Guardigni,<sup>ad</sup> Robin M. Gierse,<sup>abc</sup> Alaa Alhayek,<sup>ab</sup> Di Zhu,<sup>abc</sup> Zhoor Hamid,<sup>ab</sup> Sandra Johannsen,<sup>ab</sup> Walid A. M. Elgaher,<sup>a</sup> Philipp J. Neusens,<sup>ab</sup> Christian Nehls,<sup>e</sup> Jörg Haupenthal,<sup>a</sup> Norbert Reiling<sup>fg</sup> and Anna K. Hirsch<sup>id\*ab</sup>

Target-directed dynamic combinatorial chemistry (tdDCC) enables identification, as well as optimization of ligands for un(der)explored targets such as the anti-infective target 1-deoxy-D-xylulose-5-phosphate synthase (DXPS). We report the use of tdDCC to first identify and subsequently optimize binders/inhibitors of the anti-infective target DXPS. The initial hits were also optimized for their antibacterial activity against *E. coli* and *M. tuberculosis* during subsequent tdDCC runs. Using tdDCC, we were able to generate acylhydrazone-based inhibitors of DXPS. The tailored tdDCC runs also provided insights into the structure–activity relationship of this novel class of DXPS inhibitors. The competition tdDCC runs provided important information about the mode of inhibition of acylhydrazone-based inhibitors. This approach holds the potential to expedite the drug-discovery process and should be applicable to a range of biological targets.

Received 18th January 2021  
Accepted 21st April 2021

DOI: 10.1039/d1sc00330e

rsc.li/chemical-science

### Introduction

In dynamic combinatorial chemistry (DCC), compound libraries can be generated under thermodynamic control from a range of appropriate building blocks. The addition of external stimuli or targets can alter this thermodynamic equilibrium and change the composition of these libraries. In target-directed dynamic combinatorial chemistry (tdDCC), the target protein selects its binders and favors their formation over non-binders, resulting in their amplification in the library pool. Since the first report of tdDCC by Lehn and coworkers,<sup>1,2</sup> it has

emerged as a powerful tool to identify new ligands of biological targets.<sup>3–9</sup> This self-screening approach reduces the synthetic efforts given that only the amplified hits need to be synthesized. Ultimately, it results in acceleration of the discovery of hit compounds in the early stages of drug discovery.

Discovered in 1993, the 2C-methyl-D-erythritol 4-phosphate (MEP) pathway is an alternative biosynthetic route to generate the universal isoprenoid precursors and corresponding essential metabolites for cell survival.<sup>10</sup> The enzyme 1-deoxy-D-xylulose-5-phosphate synthase (DXPS) catalyzes the first, rate-limiting step for the production of isoprenoid precursors of this pathway. It is also involved in the biosynthesis of thiamine (vitamin B1) and pyridoxal (vitamin B6) in bacteria.<sup>11,12</sup> The absence of the MEP pathway in humans makes it a promising new target for the development of novel medicines against many life-threatening diseases like tuberculosis and malaria. Despite substantial efforts dedicated to the discovery of inhibitors for DXPS, to date, very few active compounds are known, which fulfill the requirements as an ideal candidate for further development.<sup>13–17</sup> The scarcity of inhibitors and X-ray crystal structures of the anti-infective target DXPS makes tdDCC a particularly attractive approach for the discovery of new inhibitors.

To date, the use of tdDCC in drug discovery has been limited to the initial hit-identification process. We decided to explore the possibility of using tdDCC both for hit-identification and hit-optimization. In the traditional medicinal-chemistry approach

<sup>a</sup>Department of Drug Design and Optimization, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS) – Helmholtz Centre for Infection Research (HZI), Campus Building E8.1, 66123 Saarbrücken, Germany. E-mail: Anna.Hirsch@helmholtz-hips.de

<sup>b</sup>Department of Pharmacy, Saarland University, Campus Building E8.1, 66123 Saarbrücken, Germany

<sup>c</sup>Stratingh Institute for Chemistry, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

<sup>d</sup>D3-PharmaChemistry, Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genoa, Italy

<sup>e</sup>RG Biophysics, Research Center Borstel, Leibniz Lung Center, Borstel, Germany

<sup>f</sup>RG Microbial Interface Biology, Research Center Borstel, Leibniz Lung Center, Borstel, Germany

<sup>g</sup>German Center for Infection Research (DZIF), Partner site Hamburg-Lübeck-Borstel-Riems, Borstel, Germany

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d1sc00330e

of hit-optimization, iterative cycles of ‘design–synthesis–test–analysis’ are carried out until the suitable candidate for further development has emerged.<sup>18</sup> This process involves cumbersome synthetic campaigns, followed by purification, characterization, and biological evaluation of every candidate. In tdDCC, the target selects and amplifies its best binders/inhibitors from the library. Thus, it can be used for further optimization of the initial hits in subsequently tailored libraries inspired by the initial hits. Moreover, as only amplified hits need to be synthesized and analyzed for their biological activities, this approach has the potential to reduce the time and cost associated with the traditional workflow of hit-optimization.

## Results and discussion

For this study, we used acylhydrazone formation as a reversible reaction of choice and DXPS enzyme from *Deinococcus radiodurans* (drDXPS) as a target protein. Considering the stability issues of the DXPS protein at lower pH and the slow acylhydrazone formation at neutral pH, the use of aniline as a nucleophilic catalyst was required.<sup>8,19</sup> The template effect of drDXPS on the dynamic combinatorial library (DCL) was observed by comparing it with the blank DCC experiment (without protein) using LC-MS/MS analysis.

### Design of the initial dynamic combinatorial library

The choice of the first DCL was based on structural similarities of building blocks with the cofactor thiamine diphosphate (ThDP) of the enzyme DXPS (Fig. 1a). We selected three different aldehydes containing pyrimidine and pyridine moieties and seven different hydrazides mostly featuring various heterocyclic cores. The docking study of the 21 possible acylhydrazones products from DCL-1 using LeadIT<sup>20</sup> and SeeSAR<sup>21</sup> showed that they are accommodated in the ThDP binding pocket and occupy a part of the substrate binding pocket of the enzyme drDXPS (Fig. 1 and ESI, Fig. 1–6†).

### tdDCC-1

We performed the first DCC experiment by reacting three aldehydes (A1–A3) with seven hydrazides (H1–H7) in phosphate

buffer (pH 6.25) with an excess of aniline and 5% DMSO (Fig. 2a). We evaluated the composition of DCL-1 in the blank reaction after it had reached equilibrium; after 6 h, no considerable changes in the relative peak areas (RPAs) of products were visible (Fig. 2b). A comparison of an adaptive DCC experiment in presence of drDXPS protein (40 mol%) with the blank DCL-1 revealed the amplification of five acylhydrazones (1–5) (Fig. 2c and ESI, Fig. 8†). Subsequently, we evaluated the effect of protein concentration on the amplification of these hits in an adaptive DCL, where 20 mol% of drDXPS were used, and the effect of pre-equilibration, where drDXPS (20 mol%) was added to the pre-equilibrated library (after 6 h). The systematic effect of protein concentration on the amplification of these hits is rarely studied,<sup>22,23</sup> as the use of a large amount of protein is often considered necessary for tdDCC. Comparison of the composition of these DCLs with the blank DCLs still revealed the template effect, although the amplification of all five hits (1–5) was substantially reduced when 20 mol% of drDXPS were used as compared to 40 mol% of drDXPS (Fig. 2e and ESI, Fig. 9†). When drDXPS (20 mol%) was added to the pre-equilibrated library, and the composition of the DCL was checked after an additional 6 h, the amplification of these hits was further reduced. Nonetheless, in these three experiments, all five acylhydrazones (1–5) were amplified (Fig. 2e and ESI, Fig. 10†). These data suggest that even a minute amount of the protein drDXPS can alter the equilibrium of the library and induce the template effect. These findings are particularly important for future DCC experiments with precious proteins.

### Biological evaluation of hits from tdDCC-1

We synthesized all five hits (1–5) from the tdDCC-1 and evaluated their binding affinities for drDXPS and the truncated homolog from *Mycobacterium tuberculosis* ( $\Delta$ mtDXPS)<sup>24</sup> using surface plasmon resonance (SPR, Table 1). The five hits showed moderate binding affinity ( $K_D$ ), ranging from 55–270  $\mu$ M for drDXPS and 40–250  $\mu$ M for  $\Delta$ mtDXPS. Interestingly, the trend of binding affinities of these hits mostly correlates with their amplification in the tdDCC. Furthermore, we evaluated the inhibitory activity of these hits against drDXPS and the homologous target in *M. tuberculosis* mtDXPS, which showed a similar trend, the most amplified hit was the most active one

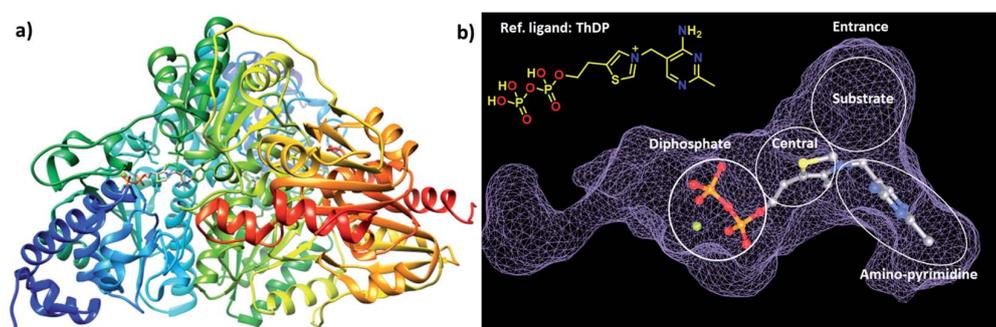


Fig. 1 (a) Crystal structure of drDXPS (PDB code: 2O1X) this image is created using UCSF Chimera.<sup>30</sup> (b) active site of drDXPS with bound cofactor thiamine diphosphate (ThDP) and the unoccupied substrate binding pocket, this image is created using online DoGSiteScorer tool.<sup>31</sup>

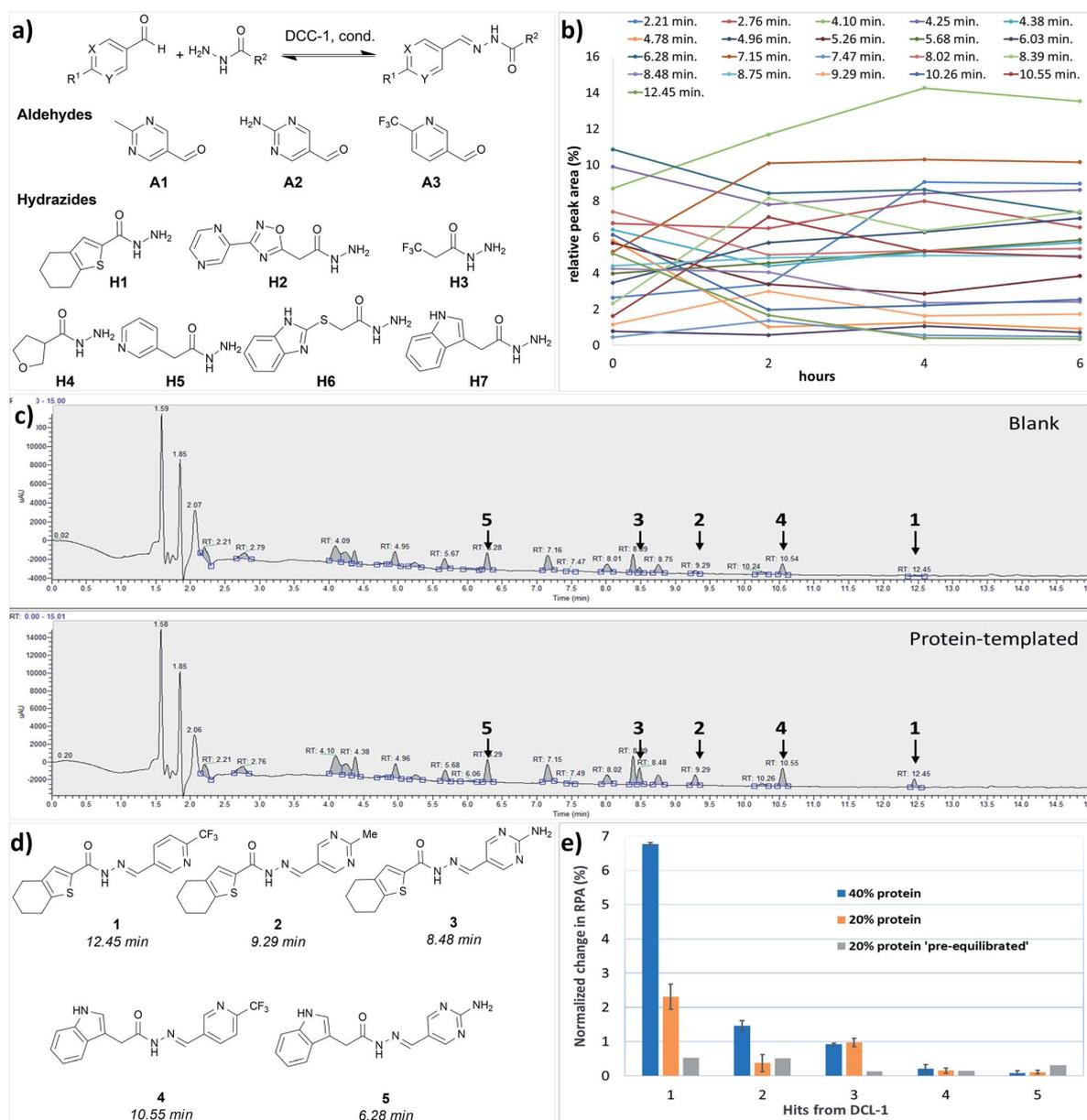


Fig. 2 (a) Dynamic combinatorial library-1, conditions: experiment was run in phosphate buffer (pH 6.25) and 5% DMSO, aldehydes (100  $\mu\text{M}$  each, in DMSO), hydrazides (300  $\mu\text{M}$  each, in DMSO), aniline (10 mM in DMSO), drDXPS protein (20–40  $\mu\text{M}$  in phosphate buffer); (b) evaluation of the equilibrium state of DCL-1, lines represent the formation of 21 acylhydrazone products over time; (c) UV-chromatogram of the blank and protein-templated adaptive DCL (40 mol% protein), see ESI, Fig. 8d and e† for peaks associated with different acylhydrazones; (d) amplified acylhydrazone hits (1–5) in DCL-1 and other experiments; (e) effect of protein concentration in the adaptive DCL run in duplicate and pre-equilibration on amplification of hits.

(1,  $\text{IC}_{50}$ : drDXPS =  $51 \pm 3 \mu\text{M}$ , Table 1, entry 1). Unfortunately, none of these hits showed considerable antibacterial activity against *Escherichia coli* TolC (<25% inhibition at 50  $\mu\text{M}$ ), which may be ascribed to permeability issues. Nonetheless, using tdDCC, we were able to find a new class of inhibitors of the anti-infective target DXPS, which provides a good starting point for further optimization. Instead of the traditional medicinal-

chemistry approach, we employed tdDCC for this purpose, letting the target protein DXPS choose its best binders.

#### tdDCC-2

We assumed that using the common structural motifs from the first hits as building blocks in the second tdDCC, along with new aldehyde and/or hydrazide counterparts would provide

Table 1 Biological evaluation of hits from tdDCC-1<sup>a</sup>

Entry	Compound	$K_D$ ( $\mu\text{M}$ )		$\text{IC}_{50}$ ( $\mu\text{M}$ )/inhib. (%)		MIC ( $\mu\text{M}$ )/inhib. (%) <i>E. coli</i> TolC
		drDXPS	$\Delta\text{mtDXPS}$	drDXPS	mtDXPS	
1	1	70 $\pm$ 5	40 $\pm$ 10	$\text{IC}_{50}$ : 51 $\pm$ 3	$\text{IC}_{50}$ : 78 $\pm$ 3	n.i.
2	2	55 $\pm$ 5	40 $\pm$ 10	Inhib.: 31 $\pm$ 4	$\text{IC}_{50}$ : 71 $\pm$ 1	Inhib.: 24 $\pm$ 0
3	3	64 $\pm$ 2	40 $\pm$ 10	Inhib.: 15 $\pm$ 9	n.d.	n.i.
4	4	n.d.	80 $\pm$ 10	n.i.	Inhib.: 25 $\pm$ 5	n.d.
5	5	270 $\pm$ 40	250 $\pm$ 20	Inhib.: 14 $\pm$ 23	Inhib.: 47	Inhib.: 18 $\pm$ 12

<sup>a</sup> Inhibition of drDXPS was determined at a concentration of 120  $\mu\text{M}$ . Inhibition of mtDXPS was determined at 200  $\mu\text{M}$ . n.d. = not determined. n.i. = no inhibition. Percent (%) growth inhibition of *E. coli* TolC was determined at 50  $\mu\text{M}$ .

better chances of identifying improved hits. From the initial hits (1–5), tetrahydrobenzothiophene and indole emerged as two privileged structural motifs, which are present in hits 1–3 and hits 4 and 5, respectively. We wanted to investigate the effect of inverting the position of these two groups in the acylhydrazone. To do so, we included their corresponding aldehydes **A4** and **A5** in DCL-2 instead of hydrazides **H1** and **H7**. In DCL-2, along with **A4** and **A5**, we included **A6** as an additional aldehyde to extend the structural diversity of the possible hits, and four new hydrazides **H8–H11** along with the three old ones (**H2**, **H5**, and **H6**, Fig. 3a). To improve the protein stability (see ESI, Table 1<sup>†</sup>) in the tdDCC conditions, we decided to use near-neutral pH from the second tdDCC experiments onwards. Analysis of DCL-2 revealed the amplification of five new hits (6–10, Fig. 3b and ESI, Fig. 12<sup>†</sup>).

#### Biological evaluation of hits from tdDCC-2

We synthesized all five new hits (6–10) and evaluated their binding affinities for drDXPS and  $\Delta\text{mtDXPS}$  by SPR (Table 2). The  $K_D$  values of these five hits were in the range of 7–150  $\mu\text{M}$  for drDXPS and 13–230  $\mu\text{M}$  for  $\Delta\text{mtDXPS}$ , a significant increase compared to the first five hits from tdDCC-1. However, when

tested against drDXPS and mtDXPS, these new hits (6–10) showed no significant improvement in their inhibitory activity. The most amplified hit **6** was most active against drDXPS ( $\text{IC}_{50}$  = 101  $\pm$  15  $\mu\text{M}$ ) and hit **7** was the most active against mtDXPS (65  $\pm$  4% inhibition at 200  $\mu\text{M}$ ). Interestingly, hits (7–10) from DCC-2 showed an improvement in antibacterial activity against *E. coli* TolC (21–100% growth inhibition at 50  $\mu\text{M}$ ) compared to the hits from DCC-1. The most active compound **9** showed an MIC value of 19  $\pm$  7  $\mu\text{M}$ . These improvements in binding affinity and antibacterial activity confirmed our hypothesis that tdDCC can be used for the optimization of inhibitors/hits.

#### tdDCC-3 & -4

To further support our hypothesis, we performed another two tailored tdDCC runs using the common structural motifs from the potent hits from DCL-2. We selected thiophenyl from hit **7** and 2,4-dichlorophenyl from hit **9** and included the corresponding hydrazides **H11** and **H12** in the next tdDCC-3 (Fig. 4). We also sought to unravel the influence of flexible linkers on the activity of the hits and thus included **H12**. In tdDCC-3 along with hydrazides **H11** and **H12**, we selected 22 commercially

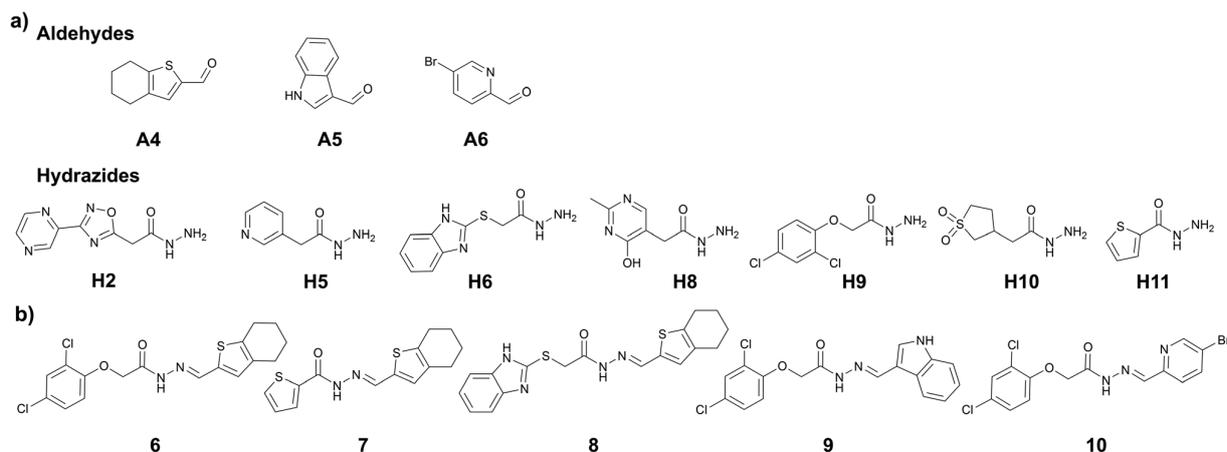


Fig. 3 (a) DCC-2, building blocks, conditions: experiment was run in phosphate buffer (pH 7.04) and 5% DMSO, aldehyde (100  $\mu\text{M}$  each, in DMSO), hydrazides (300  $\mu\text{M}$  each, in DMSO), aniline (10 mM in DMSO), drDXPS protein (40  $\mu\text{M}$  in phosphate buffer); (b) amplified acylhydrazone hits in DCL-2.

Table 2 Biological evaluation of hits from tdDCC-2<sup>a</sup>

Entry	Compound	$K_D$ ( $\mu\text{M}$ )		$\text{IC}_{50}$ ( $\mu\text{M}$ )/inhib. (%)		MIC ( $\mu\text{M}$ )/inhib. (%) <i>E. coli</i> ToIC
		drDXPS	$\Delta\text{mtDXPS}$	drDXPS	mtDXPS	
1	<b>6</b>	n.d.	21 $\pm$ 3	$\text{IC}_{50}$ : 101 $\pm$ 15	Inhib: 46 $\pm$ 3*	n.i. <sup>#</sup>
2	<b>7</b>	40 $\pm$ 10	13 $\pm$ 3	Inhib: 37 $\pm$ 9	Inhib: 65 $\pm$ 4*	Inhib: 57 $\pm$ 2
3	<b>8</b>	n.d.	230 $\pm$ 20	Inhib: 13 $\pm$ 8	Inhib: 44 $\pm$ 2*	Inhib: 73 $\pm$ 6
4	<b>9</b>	150 $\pm$ 30	140 $\pm$ 50	Inhib: 38 $\pm$ 4	Inhib: 15 $\pm$ 3	MIC: 19 $\pm$ 7
5	<b>10</b>	7 $\pm$ 1	n.d.	Inhib: 14 $\pm$ 12	Inhib: 32 $\pm$ 6	Inhib: 21 $\pm$ 7

<sup>a</sup> Inhibition of drDXPS was determined at a concentration of 120  $\mu\text{M}$ . Inhibition of mtDXPS was determined at 200  $\mu\text{M}$ . \* = inhibition of mtDXPS was determined at 120  $\mu\text{M}$ . n.d. = not determined. Percent (%) growth inhibition of *E. coli* ToIC was determined at 50  $\mu\text{M}$ . MIC values were determined only for the best compounds. <sup>#</sup> = growth inhibition of *E. coli* ToIC was determined at 25  $\mu\text{M}$ .

available aldehydes (**A1**, **A2**, **A7–A26**, see Fig. 4a and b). For operational simplicity, we divided tdDCC-3 into two groups tdDCC-3a and tdDCC-3b, where two hydrazides **H11** and **H12** were reacted with two separate groups of eleven aldehydes each (Fig. 4a and b).

For tdDCC-4, another common structural motif (indolyl) from the active hits of DCC-1 and DCC-2 was included as a building block (Fig. 4c). Here, we used 1*H*-indole-6-carbaldehyde (**A23**) instead of 1*H*-indole-3-carbaldehyde (**A5**) to understand the effect of substitution patterns of ligands on the binding to the protein and eventually in the activity of the acylhydrazones. Similarly, to understand the effect of inverting the position of the thiophenyl moiety in the acylhydrazone, we included the corresponding aldehyde **A7**. Along with these two aldehydes we also included 1*H*-pyrrole-2-carbaldehyde (**A20**) and eight hydrazides, including five new hydrazides (**H13–H17**).

Here, using DCC, and including a range of heterocyclic building blocks to cover a wide chemical space with good structural diversity, we aimed at studying to some extent the traditional medicinal chemistry approach driven by structure–activity relationships. Contrary to the traditional medicinal chemistry approach, where all possible 68 acylhydrazone products from these three DCLs should be synthesized and tested for their biological activity, we let the protein select its best binders and synthesized only the amplified derivatives.

The analysis of these target-directed DCLs revealed the amplification of ten hits each for DCL 3a and 3b and nine hits for DCL 4 (Fig. 5, and ESI, Fig. 14, 16, and 18). Carefully inspecting the hits from all three DCLs revealed that the 2,4-dichlorophenyl motif and five-membered heterocycles (thiophene, furan, imidazole, and pyrrole) are favored. As hypothesized, the use of 1*H*-indole-6-carbaldehyde instead of 1*H*-indole-

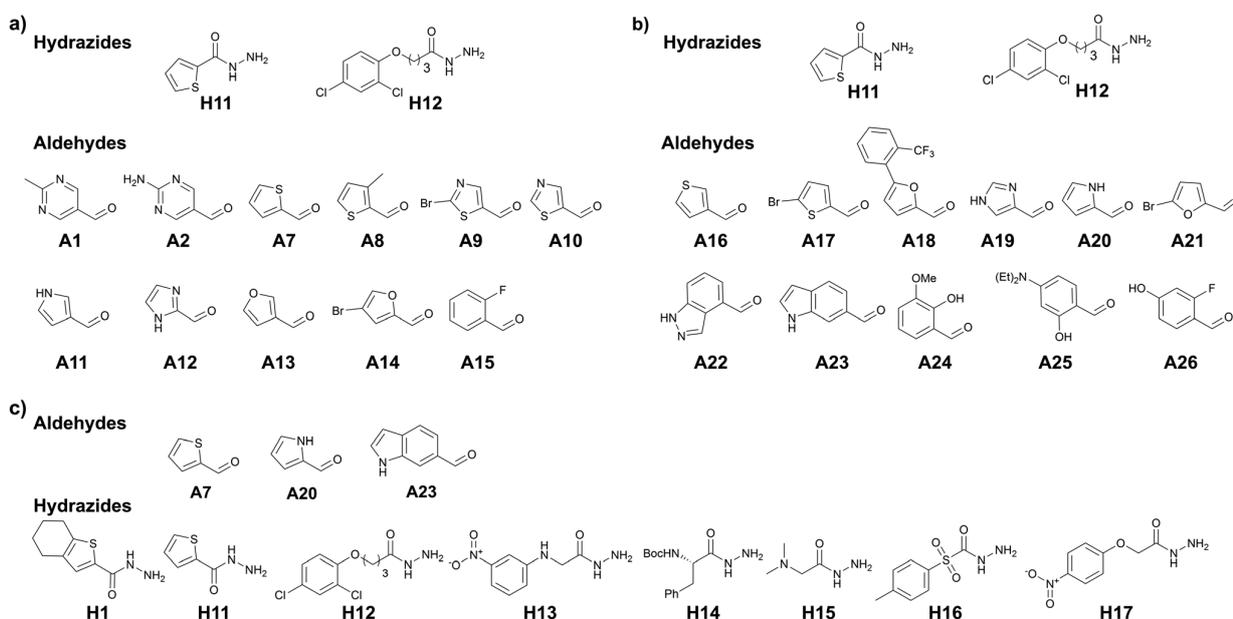


Fig. 4 (a) DCC-3a, building blocks; (b) DCC-3b, building blocks, conditions: experiment was run in phosphate buffer (pH 7.04) and 5% DMSO, aldehyde (100  $\mu\text{M}$  each, in DMSO), hydrazides (2000  $\mu\text{M}$  each, in DMSO), aniline (10 mM in DMSO), drDXPS protein (40  $\mu\text{M}$  in phosphate buffer); (c) DCC-4, building blocks, conditions: experiment was run in phosphate buffer (pH 7.04) and 5% DMSO, aldehyde (100  $\mu\text{M}$  each, in DMSO), hydrazides (300  $\mu\text{M}$  each, in DMSO), aniline (10 mM in DMSO), drDXPS protein (40  $\mu\text{M}$  in phosphate buffer).

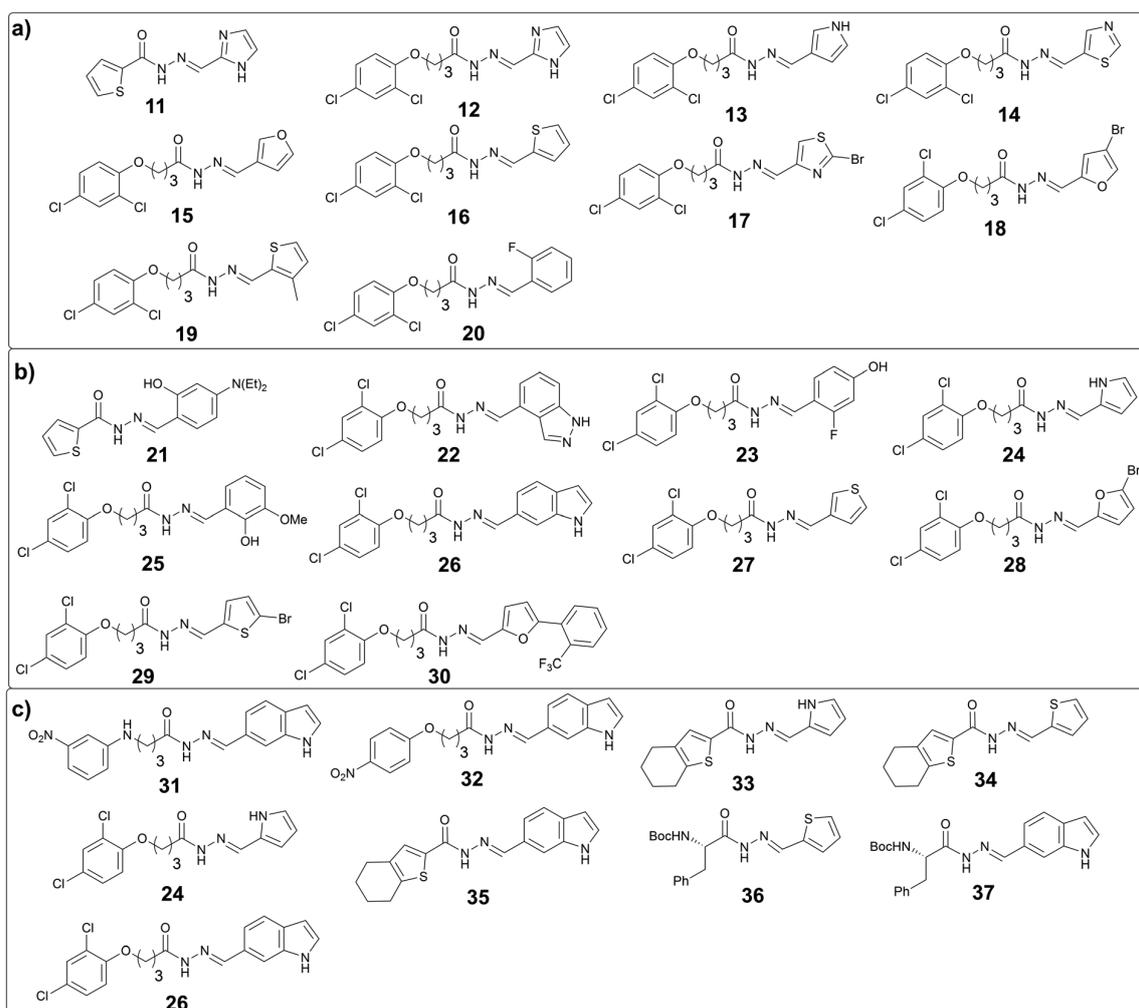


Fig. 5 (a) Amplified acylhydrazone hits in DCL-3a; (b) in DCL-3b; (c) in DCL-4.

3-carbaldehyde was well-tolerated given that five hits containing this structural motif were amplified in these DCLs (26, 31, 32, 35, and 37). Moreover, several hits containing different phenyl rings (20, 21, 23, 25, and 30) were also amplified.

#### Biological evaluation of hits from tDCC-3a, -3b and -4

We decided to synthesize a total of eleven out of 29 hits, based on their representative class (five-membered and bicyclic heterocycles, functionalized phenyl rings, and chiral compounds) and/or their amplification in the DCL and evaluated their binding affinity for drDXPS and  $\Delta$ mtDXPS by SPR (Table 3). Most of the hits show substantial improvements in binding affinity compared to the hits from the first and second round of tDCC (Table 3). Hits 11, 12, 21, 26, and 37 showed single-digit micromolar affinities (2–8  $\mu$ M) and hits 22, and 35 displayed binding affinity of  $15 \pm 3 \mu$ M and  $32 \pm 3 \mu$ M against drDXPS, respectively. Similarly, the binding affinities of these hits were also improved for  $\Delta$ mtDXPS, compound 22 which showed a single-digit micromolar affinity ( $K_D = 6 \pm 1 \mu$ M), and compound 21, 25, and 37 showed binding affinities in the range

of 18–25  $\mu$ M. When we evaluated these eleven hits for their inhibitory activities, six compounds (22, 23, 25, 26, 35, and 37) emerged as moderate inhibitors of drDXPS (40–62% activity inhibition at 120  $\mu$ M) similar to the hits from DCL-1 and -2, out of which compound 23 showed the best  $IC_{50}$  value of  $34 \pm 4 \mu$ M. In general, these hits turn out to be less potent inhibitors of mtDXPS and compound 35 showed the highest inhibition value of  $30 \pm 1\%$  at 120  $\mu$ M in the series. However, five hits (21, 23, 25, 26, and 35) showed improved antibacterial activities against *E. coli* (TolC), out of which compound 26 and 35 featured the best MIC values (MIC =  $14 \pm 4 \mu$ M and MIC =  $16 \pm 5 \mu$ M, respectively). Overall, the hits from the last three rationally designed DCLs showed significant improvements in the binding affinity and antibacterial activity compared to the hits from tDCC-1 and tDCC-2 and compound 23 emerged as the most potent inhibitor of drDXPS.

#### Antitubercular activity of DCC hits

We further evaluated the antitubercular (anti-TB) activity of 14 selected hit compounds from the three rounds of DCC (Fig. 6).

Table 3 Biological evaluation of hits from tdDCC-2<sup>a</sup>

Entry	Compound	$K_D$ ( $\mu\text{M}$ )		$\text{IC}_{50}$ ( $\mu\text{M}$ )/inhib. (%)		MIC ( $\mu\text{M}$ )/inhib. (%) <i>E. coli</i> TolC
		drDXPS	$\Delta\text{mtDXPS}$	drDXPS	mtDXPS	
1	11	8 $\pm$ 2	90 $\pm$ 30	Inhib.: 24 $\pm$ 10	Inhib.: 16 $\pm$ 4	n.i.
2	12	2 $\pm$ 1	60 $\pm$ 20	Inhib.: 33 $\pm$ 16	Inhib.: 9 $\pm$ 11	Inhib.: 30 $\pm$ 17
3	16	n.d.	n.d.	Inhib.: 4 $\pm$ 9	n.i.	Inhib.: 28 $\pm$ 17
4	21	3 $\pm$ 1	25 $\pm$ 8	Inhib.: 6 $\pm$ 14	Inhib.: 11 $\pm$ 9	MIC: 70 $\pm$ 14
5	22	15 $\pm$ 3	6 $\pm$ 1	Inhib.: 47 $\pm$ 10	n.i.	n.i.
6	23	160 $\pm$ 50	160 $\pm$ 40	$\text{IC}_{50}$ : 34 $\pm$ 4	n.i.	MIC: 45 $\pm$ 0
7	24	260 $\pm$ 80	250 $\pm$ 60	Inhib.: 14 $\pm$ 14	n.i.	Inhib.: 67 $\pm$ 23
8	25	90 $\pm$ 20	20 $\pm$ 4	Inhib.: 45 $\pm$ 10	n.i.	MIC: 60 $\pm$ 14
9	26	5 $\pm$ 2	110 $\pm$ 40	Inhib.: 58	Inhib.: 5 $\pm$ 1	MIC: 14 $\pm$ 4
10	35	32 $\pm$ 8	60 $\pm$ 20	Inhib.: 62 $\pm$ 6	Inhib.: 30 $\pm$ 1	MIC: 16 $\pm$ 5
11	37	3 $\pm$ 1	18 $\pm$ 3	Inhib.: 40 $\pm$ 10	Inhib.: 3 $\pm$ 6	Inhib.: 60 $\pm$ 24

<sup>a</sup> Inhibition of drDXPS was determined at a concentration of 120  $\mu\text{M}$  concentration of compounds. Percent (%) growth inhibition of *E. coli* TolC was determined at 50  $\mu\text{M}$ . MIC values were determined only for the best compounds. n.d. = not determined. n.i. = no inhibition.

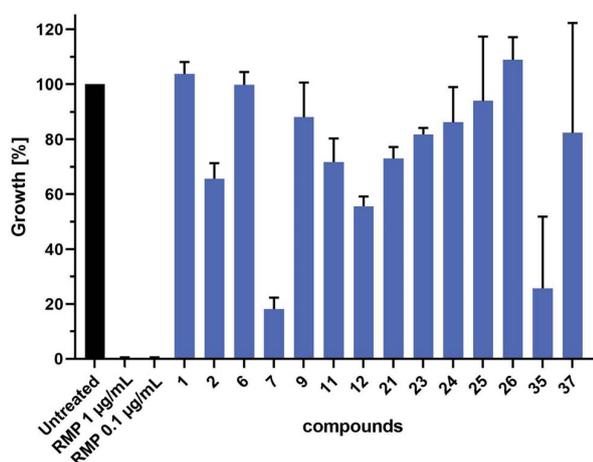


Fig. 6 Head-to-head comparative analysis of the anti-TB activity of DCC hits. mCherry-expressing *M. tuberculosis* H37Rv bacteria were cultured in the absence or presence of DCC hits (16  $\mu\text{M}$ ) or rifampicin (RMP, pos. ctrl). Shown are data of two independent experiments performed in triplicates.

These hit compounds were selected based on their inhibitory potency (1, 2, 6, 7, 23), binding affinities (11, 12, 21, 22, 26, 37), and antibacterial activity against *E. coli* TolC (9, 23–26, 35). Prior to anti-TB activity tests, dynamic light scattering (DLS) analyses were performed to establish whether the addition of the compounds to complex test culture media would lead to aggregate formation and subsequent precipitation, which would prevent functional analyses in biological systems. The DLS analyses revealed that a total of six compounds showed a concentration-dependent and up to 20–25 fold increase in the derived count rate, which was used as a parameter to measure aggregate formation in our test culture media. This allowed the determination of the highest sample concentration at which the respective compound was not yet present with an increased degree of aggregate formation (ESI, Fig. 194<sup>†</sup>). Based on these data we performed a head-to-head comparative analysis of anti-

TB activity for all 14 compounds at 16  $\mu\text{M}$ . These analyses identified compounds 7 and 35 from the second and third round of DCC, respectively, as the most active of the 14 compounds tested (Fig. 6). As both compounds did not induce aggregate formation when added at higher concentrations, MIC values could be determined to be 64  $\mu\text{M}$  for 35 and slightly above 64  $\mu\text{M}$  for 7 (ESI, Fig. 195 and 196<sup>†</sup>).

As the DLS data suggested that six compounds showed some degree of aggregation, we decided to use Triton X-100 or Tween20 in the enzyme-activity assay and binding assay respectively, to rule out any non-specific inhibition of the enzyme activity or binding to the protein by colloidal aggregates.

#### Mode-of-inhibition studies

To obtain information about the binding site of the inhibitors, we selected representative compound 1 for the mode-of-inhibition studies. We performed competition enzyme activity assays, where inhibition of drDXPS was measured with varying substrate and cofactor concentrations (curves are shown in ESI, Fig. 191–193<sup>†</sup>).

The concentration of cofactor ThDP had a strong influence on the inhibitory activity, shifting the  $\text{IC}_{50}$  value of compound 1 from 35  $\pm$  10  $\mu\text{M}$  at 0.3  $\mu\text{M}$  ThDP to only 20% inhibition of drDXPS by compound 1 at 5  $\mu\text{M}$  ThDP. Pyruvate concentrations showed a similar effect, increasing the  $\text{IC}_{50}$  value from 19  $\pm$  5  $\mu\text{M}$  at 0.1 mM pyruvate to 86  $\pm$  15  $\mu\text{M}$  at 0.4 mM and just 32% inhibition at the highest concentration of 0.6 mM. Concerning D/L-GAP, no influence of the substrate concentration on inhibition could be observed, all measured  $\text{IC}_{50}$  values are in the range of 50  $\mu\text{M}$ . These first findings show that compound 1 is ThDP- and pyruvate-competitive and non-competitive with respect to D/L-GAP.

In the first step of the reaction catalyzed by DXPS, ThDP and pyruvate are forming a covalent intermediate in the active site, while the binding of D/L-GAP to its charged pocket is accelerated 600-fold after the formation of this covalent intermediate.<sup>25</sup> A compound binding to the ThDP binding site is likely to also

occupy parts of the pyruvate pocket, as both are located adjacent to each other. To bind to the cationic binding site of D/L-GAP, compounds with negatively charged functional groups, enabling ionic interactions, would be beneficial. The type of inhibition characteristics observed for compound **1** fits well with the initial docking study of acylhydrazones products from DCL-1 in the ThDP binding site (ESI, Fig. 1–6†). In particular, the docked compound **1** is accommodated well in the ThDP binding pocket and occupies a part of the substrate pocket (Fig. 7).

The reported inhibitor butylacetylphosphonate (BAP), containing an acetylphosphonate mimic of pyruvate,<sup>26</sup> binds in the active site of DXPS and forms a covalent complex with the cofactor ThDP and extends into the substrate pocket of the active site (for the active site of drDXPS, see Fig. 1b).<sup>27</sup> The presence of ThDP and/or BAP in the protein-templated DCL can influence the amplification of acylhydrazones negatively if they are ThDP- and substrate-competitive inhibitors.

To provide support for this notion, two competition DCC experiments using DCL-1 were performed in the presence of ThDP (DCC-5) and ThDP + BAP (DCC-6). The composition of tdDCC-5 and -6 was analyzed (ESI, Fig. 19 and 20†) and compared with the composition of tdDCC-1 (Fig. 8). The addition of ThDP leads to a considerable decrease in the amplification of compounds **1**, **2**, **4**, and **5**, while compound **3** showed no change in the amplification (tdDCC-5, Fig. 8). Comparing the results from tdDCC-6 (ThDP + BAP), with tdDCC-1 and tdDCC-5, we observed a further decrease in amplification of compounds **1** and **2**, while compound **3** and **5** showed a minor or no decrease in amplification, and compound **4** showed a slight increase in amplification (Fig. 8). These findings suggest that most of the identified hits from DCL-1 compete with ThDP and substrate(s) and supports the results from the MOI study for compound **1**.

#### Bioisosteric replacement of the acylhydrazone moiety

Next, we designed and synthesized amide and 1,3,4-oxadiazoles as potential bioisosteres of the acylhydrazones **1** and **23**. As compared to acylhydrazones, amides are known to be less prone

to hydrolysis and subsequent liberation of potentially toxic hydrazides.<sup>28</sup> Moreover, the milder and simple conditions required for the synthesis of amides make them ideal candidates as bioisosteres of acylhydrazones. Similarly, structurally more rigid and hydrolytically stable 1,3,4-oxadiazoles can directly be synthesized from the building blocks of acylhydrazone formation. We synthesized amide bioisosteres **38** and **39** in moderate yield, using peptide-coupling conditions (see ESI†). The bioisosteres **40** and **41** featuring a 1,3,4-oxadiazole linker were prepared by condensation of hydrazide with the corresponding carboxylic acid followed by *in situ* dehydrative cyclizations of the resulting diacyl hydrazide intermediate in moderate to good yields (see ESI†).<sup>29</sup>

The biological evaluation of these four bioisosteres (Table 4) reveals that amide **38** retains the affinity of the parent acylhydrazone **1** for drDXPS ( $K_D = 40 \pm 10 \mu\text{M}$  vs.  $70 \pm 5 \mu\text{M}$ ) and  $\Delta\text{mtDXPS}$  ( $K_D = 60 \pm 20 \mu\text{M}$  vs.  $40 \pm 10 \mu\text{M}$ ). The amide bioisostere **39** also showed binding to drDXPS ( $K_D = 290 \pm 30 \mu\text{M}$ ) and  $\Delta\text{mtDXPS}$  ( $K_D = 260 \pm 20 \mu\text{M}$ ), although with an almost 2- and 1.5-fold decrease in binding affinities, respectively as compared to the parent acylhydrazone **23** (drDXPS:  $K_D = 160 \pm 50 \mu\text{M}$  and  $\Delta\text{mtDXPS}$ :  $K_D = 160 \pm 40 \mu\text{M}$ ). Regarding the enzymatic activity, both amide bioisosteres **38** and **39** showed a reduction in activity (Table 4, entries 1 and 2) compared to the parent acylhydrazones **1** and **23**, respectively. The 1,3,4-oxadiazole bioisostere **40** lost affinity and activity completely for DXPS enzymes (Table 4, entry 3), which could be attributed to a more rigid structure as compared to the parent acylhydrazone **1**. However, this dramatic loss of activity was not observed for a more flexible 1,3,4-oxadiazole bioisostere **41**, which shows a similar affinity as the amide bioisostere **39** (Table 4, entry 4). Interestingly, the 1,3,4-oxadiazole bioisostere **41** showed better inhibitory potency against drDXPS as compared to its amide counterpart (inhib.: 48% at  $60 \mu\text{M}$  vs. 37% at  $120 \mu\text{M}$ ) and a three-fold decrease in the  $\text{IC}_{50}$  value of the parent acylhydrazone **23** ( $\text{IC}_{50}$  (estimated) =  $99 \pm 2 \mu\text{M}$  vs.  $34 \pm 4 \mu\text{M}$ ). Concerning the antibacterial activity against *E. coli* TolC, the amide bioisostere **38** showed a slight gain in activity compared

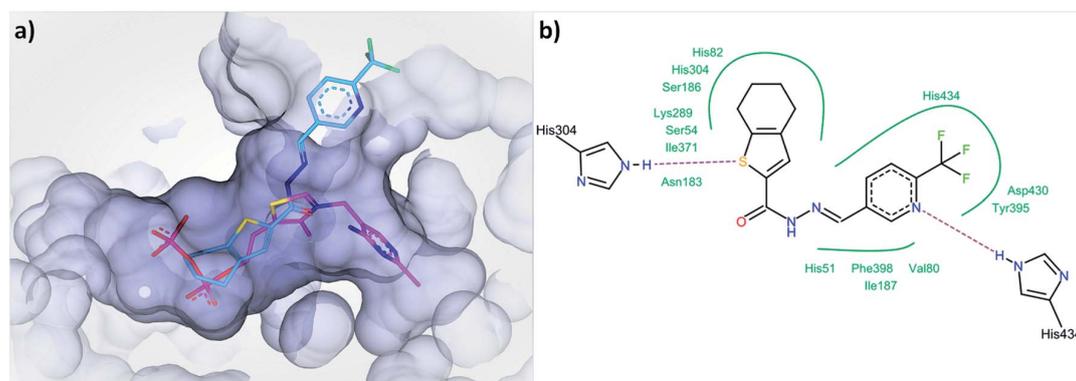


Fig. 7 (a) Docked binding mode of compound **1** (cyan) compared to ThDP (pink) in the active site of drDXPS (PDB code: 2O1X), this figure is generated using SeeSAR 10.1 (BioSolveIT),<sup>21</sup> (b) interaction of compound **1** with the residues lining the active site of drDXPS, this figure is generated with Pose View as implemented in the LeadIT suite.<sup>20</sup>

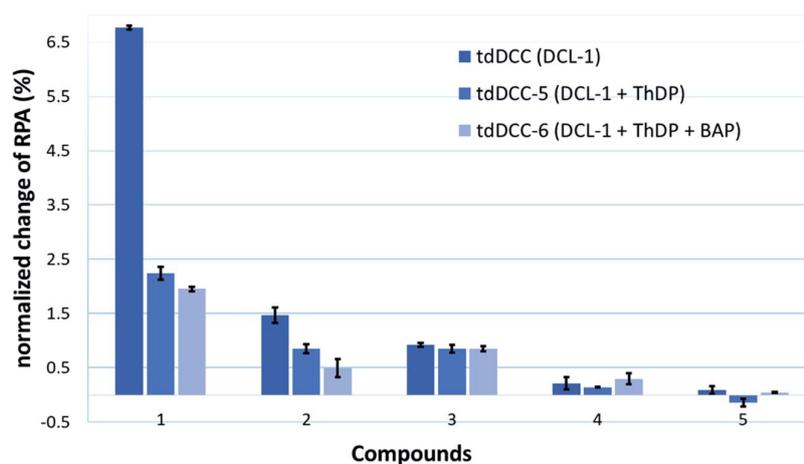
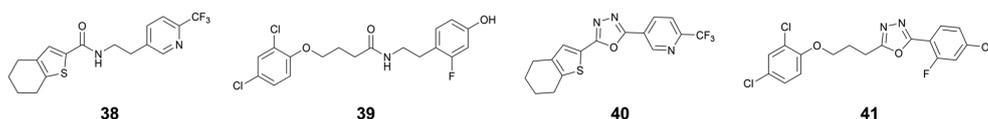


Fig. 8 Comparison of amplification of acylhydrazones in tdDCC-1, -5, and -6.

Table 4 Biological evaluation of bioisosteres<sup>a</sup>



Entry	Compound	$K_D$ ( $\mu\text{M}$ )		$\text{IC}_{50}$ ( $\mu\text{M}$ )/inhib. (%)		MIC ( $\mu\text{M}$ )/inhib. (%) <i>E. coli</i> TolC
		drDXPS	$\Delta\text{mtDXPS}$	drDXPS		
1	<b>38</b>	40 $\pm$ 10	60 $\pm$ 20	Inhib.: 25.8 $\pm$ 2.9		Inhib.: 19 $\pm$ 7
2	<b>39</b>	290 $\pm$ 30	260 $\pm$ 20	Inhib.: 37.3 $\pm$ 3.8		Inhib.: 83 $\pm$ 27
3	<b>40</b>	n.b.	n.b.	n.i.		n.i.
4	<b>41</b>	290 $\pm$ 40	210 $\pm$ 50	Inhib.: 48 $\pm$ 3 at 60 $\mu\text{M}$ ( $\text{IC}_{50}^{\#}$ : 99 $\pm$ 2)		Inhib.: 47 $\pm$ 1

<sup>a</sup> Inhibition of drDXPS was determined at 120  $\mu\text{M}$ . \* = inhibition of drDXPS at 60  $\mu\text{M}$ . n.b. = no binding. Percent (%) growth inhibition of *E. coli* TolC was determined at 50  $\mu\text{M}$ . <sup>#</sup> = estimated  $\text{IC}_{50}$ .

to the parent acylhydrazone **1** (inhib.: 19% vs. no inhib. at 50  $\mu\text{M}$ ), while 1,3,4-oxadiazole bioisostere **40** showed no inhibition. The amide bioisostere **39** largely retains the activity of the parent acylhydrazone **23** (83% inhib. at 50  $\mu\text{M}$  vs. 45  $\mu\text{M}$  MIC), but 1,3,4-oxadiazole bioisostere **41** showed almost a two-fold decrease in the antibacterial activity against *E. coli* TolC (47% inhib. at 50  $\mu\text{M}$ ).

This study reveals that the acylhydrazone motif can be replaced by a more stable amide bioisostere without losing the affinity of the parent acylhydrazone, at least in the case of compound **38**. Moreover, the 1,3,4-oxadiazole bioisostere **41**, to some extent, was able to retain the enzymatic activity of the parent acylhydrazone **23**.

## Conclusions

Our study demonstrates that tdDCC can be effectively used as a hit-optimization technique along with its well-established application for hit-identification. The process exhibits the

potential of identifying potent hits in less time compared to the traditional medicinal-chemistry approach. Importantly, using tdDCC, no prior knowledge of inhibitors is required, which allows generating new classes of hit compounds for important drug targets including underexplored ones like DXPS with little or no structural information. In the first round, using tdDCC, we were able to identify a new class of hits for the anti-infective target DXPS displaying good binding affinity (hits **1–3**) and enzymatic-activity (hits **1** and **2**). We simultaneously carried out multiparameter hit optimization using tdDCC to optimize the initial hits for their binding affinity ( $K_D \leq 5 \mu\text{M}$ ), enzymatic activity ( $\text{IC}_{50} \leq 50 \mu\text{M}$ ), and antibacterial activity ( $\text{MIC} \leq 20 \mu\text{M}$ ). These tailored tdDCC experiments also allowed to shed light on the structure–activity relationships of this new class, highlighting structural requirements for enhanced biological activity. Compounds bearing dichlorophenoxy (**12**, **26**) thiophenyl (**21**), and *N*-Boc phenylalanine (**37**) moieties and indolyl, 1*H*-imidazolyl, and substituted phenyl moieties in the hydrazide and aldehyde part, respectively, showed significant

improvement in the binding affinities. Compound **23** bearing dichlorophenoxy and 3-fluorophenol moieties in the hydrazide and aldehyde part, respectively, showed a slight increase in inhibitory potency. Compounds **9**, **26**, and **35** bearing a dichlorophenoxy and tetrahydrobenzothiophyl moiety in the hydrazide part and an indolyl in the aldehyde part showed a substantial improvement in the antibacterial activity against *E. coli* TolC. Moreover, the initial hits from tdDCC-1 were also optimized for their anti-TB activity using DCC, as compounds **7** and **35** from the second and third round of DCC, respectively, inhibited the growth of *M. tuberculosis* and thereby qualified for further optimization studies. Additionally, a docking study, competition enzyme activity assays, and competition DCC experiments provided important evidence regarding the mode-of-inhibition of compound **1**. We were also able to replace the acylhydrazone from hits **1** and **23** with more stable amide (**38**, **39**) and 1,3,4-oxadiazolyl (**40**, **41**) bioisosteres. Out of these four bioisosteres, compound **38** was able to retain the affinity of the parent acylhydrazone **1**, and compounds **41** and **39** were able to largely retain the enzymatic, and antibacterial activity of parent acylhydrazone **23**, respectively. Nevertheless, a more focused study to find the ideal bioisostere, which can retain binding, enzymatic and antibacterial activity is needed. Altogether, this proof-of-concept study demonstrates a translational potential of tdDCC in the drug-discovery process.

## Experimental

Detailed experimental section including materials and methods have been described in ESI.†

## Author contributions

R. P. J. and M. G. designed and performed the tdDCC experiments and synthesized the hits. R. M. G. expressed and purified the mtDXPS and ΔmtDXPS enzymes, and A. A. together with R. M. G. drDXPS enzyme. R. M. G., A. A., and Z. H. performed enzyme activity assays. D. Z. and R. M. G. performed the competition enzyme activity assays. R. P. J. performed the competition tdDCC experiments. S. J. and R. P. J. performed the binding affinity assays by SPR. W. A. M. E. analyzed the SPR data. J. H. coordinated the tests for antibacterial activity (*E. coli* TolC) and carried out the evaluation. C. N. performed the DLS experiments. N. R. conducted the antiTB activity assays and carried out the evaluation. P. J. N. synthesized the bioisostere **41**. A. K. H. H. and R.P.J. conceived the study and A. K. H. H. supervised the research. All authors contributed to the writing and editing of the manuscript.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

The European Commission is acknowledged for an Intra-European Marie Skłodowska-Curie actions fellowship under

Horizon-2020 (796089-NovInDXS, R. P. J.). Work by Z. H. was (partially) supported by the Schlumberger Foundation Faculty for the Future program. A. K. H. H. gratefully acknowledges funding from the European Research Council (ERC Starting grant 757913), the Netherlands Organisation for Scientific Research (LIFT grant: 731.015.414 and VIDI grant: 723.014.008), and from the Helmholtz Association's Initiative and Networking Fund. We gratefully acknowledge Dr Boris Illarionov from the University of Hamburg for reconfirming the enzymatic inhibition data of the bioisosteres. We gratefully acknowledge Jannine Jung and Dennis Jener from HIPS-HZI, Lisa Niwinski, and Jasmin Scharnberg from the Research Center Borstel for their expert technical support. We gratefully acknowledge Antoine Lacour for designing a TOC image.

## Notes and references

- I. Huc and J.-M. Lehn, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 2106–2110.
- J.-M. Lehn, *Angew. Chem., Int. Ed.*, 2015, **54**, 3276–3289.
- M. Mondal and A. K. H. Hirsch, *Chem. Soc. Rev.*, 2015, **44**, 2455–2488.
- R. Huang and I. K. Leung, *Molecules*, 2016, **21**, 910.
- A. Canal-Martín, J. Sastre, M. J. Sánchez-Barrena, A. Canales, S. Baldominos, N. Pascual, L. Martínez-González, D. Molero, M. E. Fernández-Valle, E. Sáez, P. Blanco-Gabella, E. Gómez-Rubio, S. Martín-Santamaría, A. Sáiz, A. Mansilla, F. J. Cañada, J. Jiménez-Barbero, A. Martínez and R. Pérez-Fernández, *Nat. Commun.*, 2019, **10**, 2798.
- A. M. Hartman, W. A. M. Elgaher, N. Hertrich, S. A. Andrei, C. Ottmann and A. K. H. Hirsch, *ACS Med. Chem. Lett.*, 2020, **11**, 1041–1046.
- O. Ramström, T. Bunyapaiboonsri, S. Lohmann and J. M. Lehn, *Biochim. Biophys. Acta*, 2002, **1572**, 178–186.
- P. Frei, R. Hevey and B. Ernst, *Chem.-Eur. J.*, 2019, **25**, 60–73.
- A. Canal-Martín and R. Pérez-Fernández, *ACS Omega*, 2020, **5**, 26307–26315.
- M. Rohmer, M. Knani, P. Simonin, B. Sutter and H. Sahn, *Biochem. J.*, 1993, **295**, 517–524.
- Q. Du, H. Wang and J. Xie, *Int. J. Biol. Sci.*, 2011, **7**, 41–52.
- R. E. Hill, K. Himmeldirk, I. A. Kennedy, R. M. Pauloski, B. G. Sayer, E. Wolf and I. D. Spenser, *J. Biol. Chem.*, 1996, **271**, 30426–30435.
- I. Hale, P. M. O'Neill, N. G. Berry, A. Odom and R. Sharma, *MedChemComm*, 2012, **3**, 418–433.
- J. M. Smith, N. V. Warrington, R. J. Vierling, M. L. Kuhn, W. F. Anderson, A. T. Koppisch and C. L. Freel Meyers, *J. Antibiot.*, 2014, **67**, 77–83.
- D. Bartee and C. L. Freel Meyers, *Biochem*, 2018, **57**, 4349–4356.
- D. Bartee, S. Sanders, P. D. Phillips, M. J. Harrison, A. T. Koppisch and C. L. Freel Meyers, *ACS Infect. Dis.*, 2019, **5**, 406–417.
- T. Masini, B. Lacy, L. Monjas, D. Hawksley, A. R. de Voogd, B. Illarionov, A. Iqbal, F. J. Leeper, M. Fischer, M. Kontoyianni and A. K. H. Hirsch, *Org. Biomol. Chem.*, 2015, **13**, 11263–11277.

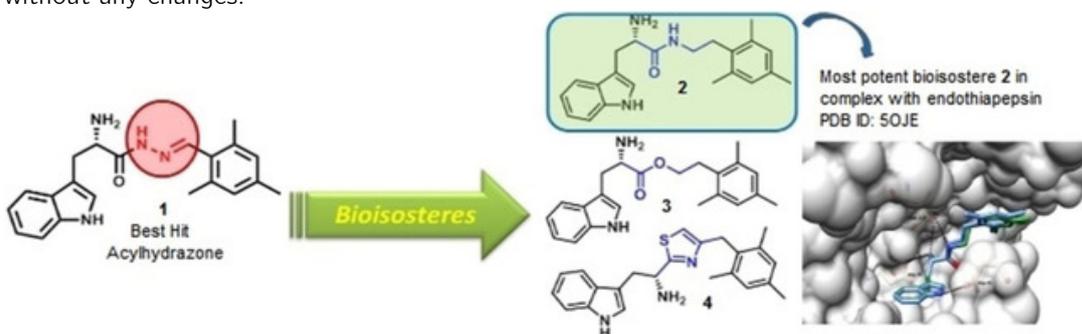
- 18 S. S. Wesolowski and D. G. Brown, in *Lead Generation*, 2016, pp. 487–512.
- 19 A. M. Hartman, R. M. Gierse and A. K. H. Hirsch, *Eur. J. Org. Chem.*, 2019, **2019**, 3581–3590.
- 20 *LeadIT version 2.3.2*, BioSolveIT GmbH, Sankt Augustin, Germany, [www.biosolveit.de/LeadIT](http://www.biosolveit.de/LeadIT), 2017.
- 21 *SeeSAR version 10.1*, BioSolveIT GmbH, Sankt Augustin, Germany, [www.biosolveit.de/SeeSAR](http://www.biosolveit.de/SeeSAR), 2020.
- 22 P. Frei, L. Pang, M. Silbermann, D. Eriş, T. Mühlethaler, O. Schwardt and B. Ernst, *Chem.–Eur. J.*, 2017, **23**, 11570–11577.
- 23 P. Frei, M. Silbermann, T. Muehlethaler, X. Jiang, O. Schwardt, R. Hevey and B. Ernst, *ARKIVOC*, 2019, **2019**, 143.
- 24 A publication on the details of the truncation and enzymatic properties of mtDXS is in preparation, the protein structure of truncated mtDXS will be published with PDB codes 7A9G and 7A9H.
- 25 H. Patel, N. S. Nemeria, L. A. Brammer, C. L. Freil Meyers and F. Jordan, *J. Am. Chem. Soc.*, 2012, **134**, 18374–18379.
- 26 T. A. O'Brien, R. Kluger, D. C. Pike and R. B. Gennis, *Biochim. Biophys. Acta*, 1980, **613**, 10–17.
- 27 J. M. Smith, R. J. Vierling and C. F. Meyers, *MedChemComm*, 2012, **3**, 65–67.
- 28 V. R. Jumde, M. Mondal, R. M. Gierse, M. Y. Unver, F. Magari, R. C. W. van Lier, A. Heine, G. Klebe and A. K. H. Hirsch, *ChemMedChem*, 2018, **13**, 2266–2270.
- 29 N. N. Farshori, A. Rauf, M. A. Siddiqui, E. S. Al-Sheddi and M. M. Al-Oqail, *Arabian J. Chem.*, 2017, **10**, S2853–S2861.
- 30 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
- 31 A. Volkamer, D. Kuhn, F. Rippmann and M. Rarey, *Bioinformatics*, 2012, **28**, 2074–2075.





## 2.5 Design and Synthesis of Bioisosteres of Acylhydrazones as Stable Inhibitors of the Aspartic Protease Endothiapepsin

The following chapter was published as an open-access article in the Journal ChemMedChem under the terms of the creative commons license CC BY-NC 4.0 and is printed in this thesis without any changes.<sup>2</sup>



### Title

Design and Synthesis of Bioisosteres of Acylhydrazones as Stable Inhibitors of the Aspartic Protease Endothiapepsin

### Authors

Varsha R. Jumde<sup>‡</sup>, Milon Mondal<sup>‡</sup>, Robin M. Gierse, M. Yagiz Unver, Francesca Magari, Roos C. W. van Lier, Andreas Heine, Gerhard Klebe and Anna K. H. Hirsch

### Contributions

V. R. Jumde conceived the project and drafted the manuscript, synthesized compounds and performed bioassays, M. Mondal designed the experiments and performed the molecular modeling. R. M. Gierse contributed to the manuscript and performed together with F. Magari the crystallization, determination and analysis of the crystal structure. M. Y. Unver was involved in the synthesis of compounds and bioassays, R. C. W. van Lier contributed to the synthesis of compounds, A. Heine and G. Klebe contributed to the analysis and interpretation of the crystallographic data and were involved in the supervision of the project, A.K.H. Hirsch conceived and supervised the project and edited the manuscript.

### Bibliographic Details

ChemMedChem  
Volume 13, Issue 21, Pages 2266-2270  
3 September 2018  
DOI: 10.1002/cmdc.201800446

<sup>‡</sup>: Authors contributed equally

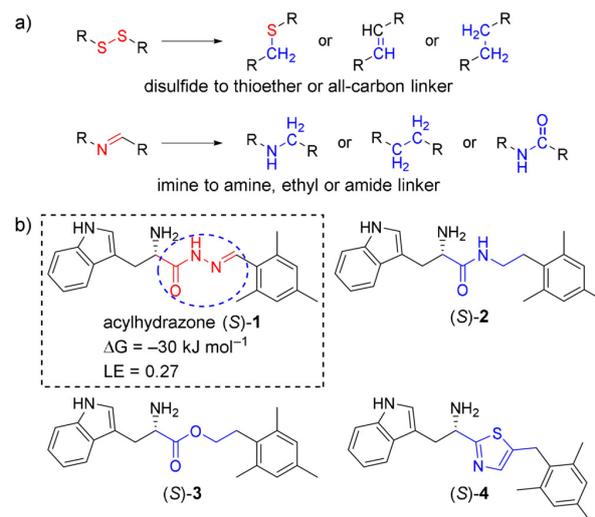
# Design and Synthesis of Bioisosteres of Acylhydrazones as Stable Inhibitors of the Aspartic Protease Endothiapepsin

Varsha R. Jumde<sup>+, [a, b]</sup> Milon Mondal<sup>+, [a]</sup> Robin M. Gierse<sup>+, [a, b, c]</sup> M. Yagiz Unver<sup>+, [a]</sup> Francesca Magari<sup>+, [d]</sup> Roos C. W. van Lier<sup>+, [a]</sup> Andreas Heine<sup>+, [d]</sup> Gerhard Klebe<sup>+, [d]</sup> and Anna K. Hirsch<sup>\*, [a, b, c]</sup>

Acylhydrazone-based dynamic combinatorial chemistry (DCC) is a powerful strategy for the rapid identification of novel hits. Even though acylhydrazones are important structural motifs in medicinal chemistry, their further progression in development may be hampered by major instability and potential toxicity under physiological conditions. It is therefore of paramount importance to identify stable replacements for acylhydrazone linkers. Herein, we present the first report on the design and synthesis of stable bioisosteres of acylhydrazone-based inhibitors of the aspartic protease endothiapepsin as a follow-up to a DCC study. The most successful bioisostere is equipotent, bears an amide linker, and we confirmed its binding mode by X-ray crystallography. Having some validated bioisosteres of acylhydrazones readily available might accelerate hit-to-lead optimization in future acylhydrazone-based DCC projects.

Dynamic combinatorial chemistry (DCC) enables rapid screening of functionally diverse compounds against a target, circumventing the need for individual synthesis, purification and characterization.<sup>[1–7]</sup> Among many other prominent examples of DCC, reversible disulfide-bond formation was first introduced in DCC by the groups of Still,<sup>[8]</sup> Sanders,<sup>[5]</sup> and Lehn<sup>[9]</sup> in the late 1990s. Later on, in 1997, the group of Lehn first applied

DCC to a protein target using imine formation/exchange.<sup>[11]</sup> Since then, its scope and wider applicability were demonstrated on a range of biological targets. Replacement of the reversible disulfide bond with thioether ( $-S-CH_2-$ )<sup>[10]</sup> or all-carbon (olefin,  $-CH_2-CH_2-$ )<sup>[11–15]</sup> and of the imine moiety with amines,<sup>[11]</sup> an ethyl linker<sup>[16]</sup> or with an amide linker<sup>[17]</sup> provides stable bioisosteres with potentially preserved binding mode, making DCC an enabling tool for medicinal chemistry and drug discovery (Figure 1 a).



**Figure 1.** a) Previous examples of bioisosteres and b) proposed bioisosteres (2–4) of the acylhydrazone 1 as stable inhibitors of endothiapepsin.

We chose the target protein endothiapepsin, belonging to the family of pepsin-like aspartic proteases, which play a causative role in numerous diseases such as malaria, Alzheimer's disease, hypertension, and HIV-1.<sup>[18]</sup> Endothiapepsin is used as a representative enzyme due to its robustness, immense stability and similarity to the drug targets of the class of aspartic proteases. Moreover, it has been used as a model enzyme for mechanistic studies,<sup>[19–21]</sup> as it is readily available in large quantity and crystallizes easily and importantly remains active at room temperature for more than 20 days.

We previously discovered acylhydrazone-based inhibitors of endothiapepsin using DCC in combination with de novo structure-based drug design, which display a promising inhibitory

[a] Dr. V. R. Jumde,<sup>+</sup> Dr. M. Mondal,<sup>+</sup> R. M. Gierse, Dr. M. Y. Unver, R. C. W. van Lier, Prof. Dr. A. K. H. Hirsch  
Chemical Biology, Strating Institute for Chemistry, University of Groningen, Nijenborgh 7, 9747 AG Groningen (The Netherlands)  
E-mail: anna.hirsch@helmholtz-hips.de

[b] Dr. V. R. Jumde,<sup>+</sup> R. M. Gierse, Prof. Dr. A. K. H. Hirsch  
Helmholtz Institute for Pharmaceutical Research Saarland (HIPS)—Helmholtz Centre for Infection Research (HZI), Department of Drug Design and Optimization (DDOP), Campus Building E8.1, 66123 Saarbrücken (Germany)

[c] R. M. Gierse, Prof. Dr. A. K. H. Hirsch  
Department of Pharmacy, Saarland University, Saarbrücken, Campus Building E8.1, 66123 Saarbrücken (Germany)

[d] F. Magari, Prof. Dr. A. Heine, Prof. Dr. G. Klebe  
Drug Design Group AG Klebe, Institute of Pharmaceutical Chemistry, Marbacher Weg 6, 35032 Marburg (Germany)

[†] These authors contributed equally to this work.

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:  
<https://doi.org/10.1002/cmdc.201800446>.

© 2018 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

profile ( $IC_{50} = 12.8 \pm 0.4 \mu\text{M}$ ).<sup>[22]</sup> Acylhydrazones are considered to be important structural motifs in medicinal chemistry, as they hold the potential to interact with a range of biological targets in antiviral, anticancer and antibacterial drug discovery.

Nevertheless, there are problems associated with acylhydrazones. This class of compounds is considered by some as a member of the pan-assay interference compounds (PAINS).<sup>[23]</sup> They undergo photoinduced *E/Z* isomerization.<sup>[24]</sup> In addition, it is important to consider also the behavior of acylhydrazones *in vivo*. The major setback of acylhydrazones is their lack of stability due to hydrolysis into an aldehyde and a hydrazide under acidic pH. In spite of that, hydrazone and acylhydrazone linkages are used to develop pH-degradable drug-delivery systems for site-specific targeting.<sup>[25]</sup> Furthermore, some acylhydrazones, like PAC-1, are in clinical trials as a treatment for cancer.<sup>[26,27]</sup> Nevertheless, it is highly desirable to replace the labile acylhydrazone linker with stable and chemically benign analogues while maintaining the key interactions in the active site of the protein without significant changes in chemical structure.

Surprisingly, to the best of our knowledge, there are only few examples of bioisosteres of acylhydrazones,<sup>[16]</sup> but no report as a direct follow-up of a DCC experiment. In most cases, the binding mode of the bioisostere is not confirmed experimentally. Having suitable bioisosteres in hand, will establish 'acylhydrazone-based DCC' as a powerful hit/lead-identification strategy with the potential for further optimization.

Bioisosteres have been introduced as a fundamental strategy to improve the biocompatibility of the parent hit or lead compounds. As such, bioisosteres contribute to the field of medicinal chemistry, in terms of improving potency, enhancing selectivity, altering physicochemical properties, reducing or redirecting metabolism, eliminating or modifying toxicophores and acquiring novel intellectual property.<sup>[28]</sup> Herein, we describe the design, synthesis, and biochemical activity of three bioisosteres of the acylhydrazone (S)-1, the first acylhydrazone inhibitor of endothiapepsin. Importantly, unlike the parent acylhydrazone, bioisosteres (S)-2 and (S)-4 are not prone to hydrolysis, and all three do not liberate potentially toxic hydrazides.

We chose the X-ray crystal structure of endothiapepsin in complex with acylhydrazone (S)-1 (PDB ID: 4KUP)<sup>[22]</sup> as a starting point for the design of stable bioisosteres of the labile acylhydrazone moiety. Hit (S)-1 displays an  $IC_{50}$  value of  $12.8 \mu\text{M}$  and a ligand efficiency (LE) of 0.27. It interacts with the catalytic dyad using H-bonding interactions (Asp35 (2.8 Å, 3.2 Å) and Asp219 (2.9 Å)) through its  $\alpha$ -amino group.

We designed bioisosteres using two different design approaches, namely Recore in the LeadIT suite<sup>[29]</sup> and the molecular modeling software Moloc<sup>[30]</sup> for molecular modeling and computation of the dipole moments. In Recore, a defined moiety of a molecule (the core) is replaced by fragments from a 3D database whilst keeping the rest of the molecule intact. To restrict the number of solutions, defined ligand-based pharmacophore constraints can be assigned. This modeling and docking resulted in various compounds displaying heterocyclic, ester or amide linkages. Among the various heterocycles (e.g.,

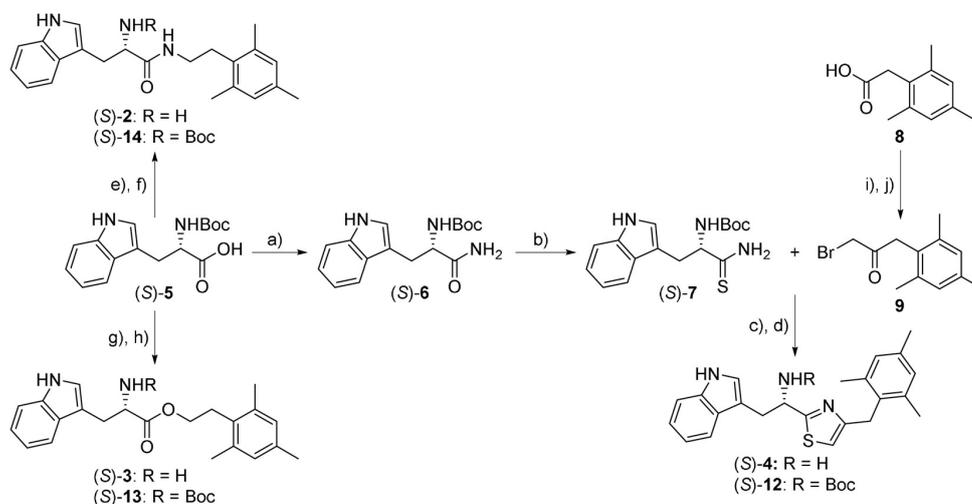
triazole, tetrazole, oxazole (Supporting Information Figure S8)), we chose the best three compounds (Figure 1b) based on their dipole moments, their calculated  $\Delta G$ , and predicted binding modes, which are similar to those of the parent acylhydrazone (S)-1 and synthesized them as a proof-of-concept study. The predicted binding modes of three representative bioisosteres in the active pocket of endothiapepsin are shown in Figure S4 (Supporting Information).

Inspection of the soaked crystal structure of endothiapepsin with acylhydrazone (S)-1 in the active site shows that the aromatic parts of the compound such as indolyl and/or mesityl moieties are able to form  $\pi$ - $\pi$ -stacking interactions with the amino acid residues of the protein backbone. In all of the structures (Figure 1b), the binding modes of the indolyl and mesityl moieties are preserved. It was computationally observed that the  $\alpha$ -amino groups of all bioisosteres (S)-2-(S)-4 form charge-assisted H bonds to the catalytic dyad (Asp35 and Asp 219) as well as additional H-bonding interactions with Asp81, and Gly221. The indolyl NH forms H bonds either with Asp81 or Asp33, the NH group of the amide donates an H bond to Gly221 in (S)-2. In addition to these, the thiazolyl ring of (S)-4 is involved in several hydrophobic interactions with the protein backbone. The main building blocks required for the synthesis of bioisosteres (S)-2-(S)-4, are *N*- $\alpha$ -Boc-L-tryptophan (5) and the 2-mesitylene-derived compounds (S)-5, 8 and 11 (see Schemes S1–S4 in the Supporting Information and Scheme 1).

Very mild peptide-coupling conditions afforded the bioisostere (S)-2 with the amide linker, followed by deprotection of the Boc group. Starting from *N*-Boc-L-tryptophan (5) and 2-mesitylethanamine hydrochloride (10) in presence of the weak base carbonyldiimidazole, furnished the corresponding amide (S)-14 in 80% yield, and after deprotection with TFA, the test compound (S)-2 in quantitative yield. The ester (S)-3 was accessible through the Steglich esterification.<sup>[31]</sup> We synthesized the bioisostere (S)-4 from the building blocks thioamide (S)-7 and ketobromide 9, which can be both accessed in two steps from *N*- $\alpha$ -Boc-L-tryptophan (5) and mesitylacetic acid (8),<sup>[32,33]</sup> respectively.

Subsequent deprotection of the Boc group of compound (S)-12 afforded bioisostere (S)-4 in quantitative yield. The first step to obtain thioamide (S)-7 consists of the synthesis of amide (S)-6 followed by thionation using Lawesson's reagent. On the other hand, using modified Arndt–Eistert reaction conditions, starting from mesitylacetic acid (8), afforded intermediate 9. To investigate the biochemical activity of the designed bioisosteres (S)-2 to (S)-4, we performed a fluorescence-based assay adapted from the HIV-protease assay (see Figures S1–S3 for the  $IC_{50}$  curves, Supporting Information).<sup>[34]</sup>

The three designed bioisosteres inhibit the activity of endothiapepsin to a different extent. The most potent inhibitor, the amide bioisostere (S)-2, displays a  $K_i$  value of  $6.1 \mu\text{M}$ , very similar to the parent acylhydrazone (S)-1 ( $K_i = 6.0 \mu\text{M}$ , Table 1). We calculated the  $K_i$  values from experimental  $IC_{50}$  values using the Cheng–Prusoff equation.<sup>[35]</sup> To verify the predicted binding mode of the bioisosteres, we soaked crystals of endothiapepsin with the most potent bioisostere (S)-2 and determined the

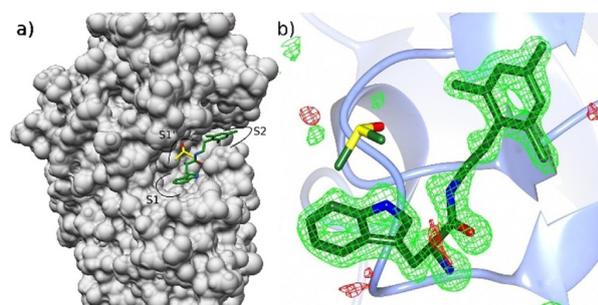


**Scheme 1.** Synthesis of bisoesters: a)  $\text{ClCO}_2\text{Et}$ ,  $\text{Et}_3\text{N}$ , dry THF, aq.  $\text{NH}_3$ ; b) Lawesson's reagent, dry  $\text{CH}_2\text{Cl}_2$ ; c)  $\text{EtOH}$ , reflux, 4 h; d) TFA,  $\text{CH}_2\text{Cl}_2$ ; e) 2-mesitylthalamine hydrochloride (**10**), 1,1'-carbonyldiimidazole, THF, RT, 15 h; f) TFA,  $\text{CH}_2\text{Cl}_2$ ,  $0^\circ\text{C} \rightarrow \text{RT}$ , 1.5 h; g) 2-mesitylethanol (**11**), DCC, DMAP (5%)  $\text{CH}_2\text{Cl}_2$ , 8 h; h)  $\text{HCl}/\text{Et}_2\text{O}$  1 M, 24 h; i)  $\text{SOCl}_2$ , dry toluene, reflux, 3 h; j) a) TMS-diazomethane,  $\text{Et}_2\text{O}$ , b) 47.5% aq. HBr.

Table 1. Biochemical evaluation of acylhydrazone (S)-1 and its bisoesters (S)-2–(S)-4. Each experiment was carried out in duplicate.				
Inhibitor	$\text{IC}_{50}$ [ $\mu\text{M}$ ] <sup>[a]</sup>	$K_i$ [ $\mu\text{M}$ ] <sup>[b]</sup>	$\Delta G_{\text{EXPT}}$ [ $\text{kJ mol}^{-1}$ ] <sup>[b]</sup>	$\Delta G_{\text{HYDE}}$ [ $\text{kJ mol}^{-1}$ ] <sup>[c]</sup>
(S)-1	$12.8 \pm 0.4$	$6.0 \pm 0.2$	−30	−32
(S)-2	$12.9 \pm 0.7$	$6.1 \pm 0.4$	−30	−27
(S)-3	$28.7 \pm 4.1$	$13.5 \pm 1.9$	−28	−28
(S)-4	$193.7 \pm 11.4$	$91.2 \pm 5.4$	−23	−31

[a] Eleven different concentrations of inhibitor were used; errors are given in standard deviations (SD). [b] Values indicate the inhibition constant ( $K_i$ ) and the Gibbs free energy of binding ( $\Delta G$ ) derived from  $\text{IC}_{50}$  values using the Cheng–Prusoff equation.<sup>[35]</sup> [c] Values indicate the calculated Gibbs free energy of binding ( $\Delta G_{\text{HYDE}}$ ; calculated by the HYDE scoring function in the LeadIT suite).

crystal structure of (S)-2 in complex with endothiapepsin at 1.58 Å resolution (PDB ID: 5OJE). The structure features clear electron density for the ligand, as shown in Figure 2 b).



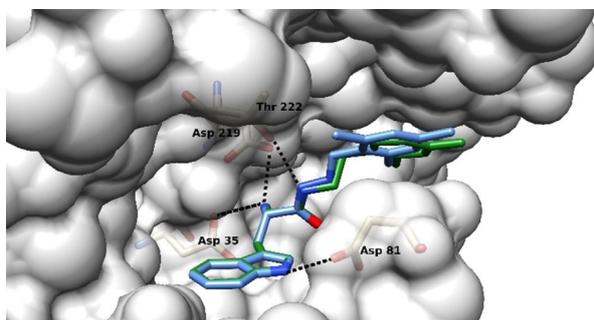
**Figure 2.** a) Zoomed-out view of the protein shown as surface. b) Electron density omit-map of the crystal structure of endothiapepsin in complex with compound (S)-2 and a coordinated DMSO molecule.  $F_o - F_c$  map contoured at  $3.3 \sigma$  (color code: protein cartoon: light blue, C: green, O: red, N: blue, S: yellow).

Upon closer examination, the location of the ligand is similar to the docked pose shown in Figure S4 (See Supporting Information). The amino group of the ligand forms two H bonds with Asp35 (2.9 Å) and Asp219 (3.0 Å). The indolyl nitrogen atom forms an H bond with Asp81 (3.2 Å). The hydrophobic part of the indolyl moiety is engaged in hydrophobic interactions with Phe116, Leu125, Tyr79 and Gly221. The mesityl substituent is involved in hydrophobic interactions with Ile300, Ile304, Tyr226, Gly80 and Asp81. The oxygen atom of the amide linkage forms water-mediated H bonds to the carbonyl oxygen of Gly37 and the amide nitrogen of Gly80. The mediating water molecules are conserved between the crystal structures in complex with (S)-1 and (S)-2 (PDB IDs: 4KUP and 5OJE, respectively, Supporting Information Figure S7).

The only difference compared to the docked pose is at the amide linkage. In contradiction to the computational modeling, the nitrogen atom of the amide does not form an H bond with the oxygen atom of Gly221, the distance is 4.2 Å. Instead, the hydroxy group of Thr222 acts as an H-bond acceptor and forms an H bond (2.9 Å) with the amide nitrogen atom of the ligand, which is also shown in Figure 3.

Due to the slightly bent shape of the coordinated ligand, both aromatic groups are able to form hydrophobic interactions with one DMSO molecule, shown in Figure 2. This DMSO molecule is well-coordinated and seems to displace several water molecules. This may be important for the stabilization of the ligand bound to the protein. A similar DMSO molecule can be observed in previous crystal structures (e.g., PDB ID: 4KUP).<sup>[22]</sup>

The single bond connecting the mesityl unit to the rest of the acylhydrazone (S)-1 is part of a conjugated system and prefers a planar orientation. It is twisted out of planarity to an unfavorable angle of  $34.4^\circ$  compared to the more favored angle of  $107.0^\circ$  as in bisoestere (S)-2 (Supporting Information Figure S6).



**Figure 3.** Superimposition of the acylhydrazone inhibitor (S)-1 (cyan) and the amide bioisostere (S)-2 (green). H bonds below 3.0 Å are shown as black dashed lines (color code: protein backbone: C: gray, O: red, N: blue, S-1: C: cyan and (S)-2: C: green).

The bioisostere (S)-2, however, contains a peptidic bond in the linker, which also prefers planarity. This forces the C–N bond, its third bond, counting from the mesityl substituent, into an unfavorable torsional angle of 122° compared to the preferred 170° of the acylhydrazone (Figure S6). In conclusion, both ligands have to adopt a slightly unfavorable conformation to bind in the pocket of the enzyme, which is reflected in their binding affinities. Based on our observations, it might be difficult to design a linker with improved binding affinity, which would need to be more flexible with respect to the torsional angles, while the H-bond donor and –acceptor functions of the peptidic nitrogen and oxygen atoms should ideally be preserved.

We report the successful replacement of the acid-sensitive and hydrolyzable acylhydrazone linker of parent hit (S)-1, affording stable and equipotent inhibitors of endothiapepsin. We designed and synthesized three bioisosteres and evaluated them for their inhibitory potency against endothiapepsin. Compounds (S)-2 and (S)-3, possessing amide and ester linkers, respectively display similar  $K_i$  values as the parent hit (S)-1, while compound (S)-4 is an order of magnitude weaker than the parent hit. The crystal structure of amide (S)-2 ( $K_i = 6.1 \mu\text{M}$ ) in complex with endothiapepsin validates the predicted binding mode. In this proof-of-concept study, we identified molecular interactions that should be taken into consideration if further modifications are done to achieve a more druglike replacement for the acylhydrazone linker. Taken together, we demonstrate that acylhydrazones can be replaced without affecting the binding mode and whilst preserving the activity, demonstrating that acylhydrazone-based DCC is a powerful tool to identify hits, which can then be optimized to stable lead compounds in a straightforward manner.

## Experimental Section

Full experimental details are provided in the Supporting Information.

## Acknowledgements

A.K.H.H. gratefully acknowledges funding from the Netherlands Organisation for Scientific Research (VIDI and LIFT grants), the Dutch Ministry of Education, Culture and Science (Gravitation Program 024.001.035), the European Research Council (ERC starting grant 757913), and the Helmholtz Association's Initiative and Networking Fund. Molecular graphics and analyses were performed with the UCSF Chimera package and ccp4mg.<sup>[36,37]</sup> The presented work was also funded by the European Union's Framework Programme for Research and Innovation Horizon 2020 (2014–2020) under the Marie Skłodowska-Curie Grant Agreement No. 675555, Accelerated Early stage drug discovery (AEGIS).

## Conflict of interest

The authors declare no conflict of interest.

**Keywords:** acylhydrazones · aspartic proteases · bioisosteres · drug design · dynamic combinatorial chemistry

- [1] I. Huc, J.-M. Lehn, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2106–2110.
- [2] A. Herrmann, *Chem. Soc. Rev.* **2014**, *43*, 1899–1933.
- [3] C. Karan, B. L. Miller, *Drug Discovery Today* **2000**, *5*, 67–75.
- [4] O. Ramström, J.-M. Lehn, *Nat. Rev. Drug Discovery* **2002**, *1*, 26–36.
- [5] S. Otto, R. L. E. Furlan, J. K. M. Sanders, *J. Am. Chem. Soc.* **2000**, *122*, 12063–12064.
- [6] S. Ladame, *Org. Biomol. Chem.* **2008**, *6*, 219–226.
- [7] M. Mondal, A. K. H. Hirsch, *Chem. Soc. Rev.* **2015**, *44*, 2455–2488.
- [8] H. Hioki, W. C. Still, *J. Org. Chem.* **1998**, *63*, 904–905.
- [9] O. Ramström, J.-M. Lehn, *ChemBioChem* **2000**, *1*, 41–48.
- [10] N. Fotouhi, P. Joshi, J. W. Tilley, K. Rowan, V. Schwinge, B. Wolitzky, *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1167–1169.
- [11] P. B. Palde, L. O. Ofori, P. C. Gareiss, J. Lerea, B. L. Miller, *J. Med. Chem.* **2010**, *53*, 6018–6027.
- [12] J. L. Stymiest, B. F. Mitchell, S. Wong, J. C. Vederas, *Org. Lett.* **2003**, *5*, 47–49.
- [13] I. Berezowska, N. N. Chung, C. Lemieux, B. C. Wilkes, P. W. Schiller, *J. Med. Chem.* **2007**, *50*, 1414–1417.
- [14] A. Mollica, G. Guardiani, P. Davis, S. W. Ma, F. Porreca, J. Lai, L. Mannina, A. P. Sobolev, V. J. Hruby, *J. Med. Chem.* **2007**, *50*, 3138–3142.
- [15] K. C. Nicolaou, R. Hughes, S. Y. Cho, N. Winssinger, C. Smethurst, H. Labischinski, R. Endermann, *Angew. Chem. Int. Ed.* **2000**, *39*, 3823–3828; *Angew. Chem.* **2000**, *112*, 3981–3986.
- [16] J. L. Catrow, Y. Zhang, M. Zhang, H. Ji, *J. Med. Chem.* **2015**, *58*, 4678–4692.
- [17] A. Valade, D. Urban, J. M. Beau, *ChemBioChem* **2006**, *7*, 1023–1027.
- [18] J. B. Cooper, *Curr. Drug Targets* **2002**, *3*, 155–173.
- [19] L. Coates, P. T. Erskine, S. Mall, R. Gill, S. P. Wood, D. A. A. Myles, J. B. Cooper, *Eur. Biophys. J.* **2006**, *35*, 559–566.
- [20] L. Coates, P. T. Erskine, S. P. Wood, D. A. A. Myles, J. B. Cooper, *Biochemistry* **2001**, *40*, 13149–13157.
- [21] L. Coates, H.-F. Tuan, S. Tomanicek, A. Kovalevsky, M. Mustyakimov, P. Erskine, J. Cooper, *J. Am. Chem. Soc.* **2008**, *130*, 7235–7237.
- [22] M. Mondal, N. Radeva, H. Köster, A. Park, C. Potamitis, M. Zervou, G. Klebe, A. K. H. Hirsch, *Angew. Chem. Int. Ed.* **2014**, *53*, 3259–3263; *Angew. Chem.* **2014**, *126*, 3324–3328.
- [23] J. B. Baell, L. Ferrins, H. Falk, G. Nikolakopoulos, *Aust. J. Chem.* **2013**, *66*, 1483–1494.
- [24] D. J. Van Dijken, P. Kovaříček, S. P. Ihrig, S. Hecht, *J. Am. Chem. Soc.* **2015**, *137*, 14982–14991.
- [25] S. J. Sonawane, R. S. Kalhapure, T. Govender, *Eur. J. Pharm. Sci.* **2017**, *99*, 45–65.

- [26] R. C. Botham, T. M. Fan, I. Im, L. B. Borst, L. Dirikolu, P. J. Hergenrother, *J. Am. Chem. Soc.* **2014**, *136*, 1312–1319.
- [27] H. S. Roth, R. C. Botham, S. C. Schmid, T. M. Fan, L. Dirikolu, P. J. Hergenrother, *J. Med. Chem.* **2015**, *58*, 4046–4065.
- [28] N. A. Meanwell, *J. Med. Chem.* **2011**, *54*, 2529–2591.
- [29] BioSolveIT GmbH, Sankt Augustin. <http://www.biosolveit.de>, LeadIT, version 2.1.3.
- [30] P. R. Gerber, K. Müller, *J. Comput.-Aided Mol. Des.* **1995**, *9*, 251–268.
- [31] B. Neises, W. Steglich, *Angew. Chem. Int. Ed. Engl.* **1978**, *17*, 522–524; *Angew. Chem.* **1978**, *90*, 556–557.
- [32] A. Gangjee, J. Yang, M. A. Ihnat, S. Kamat, *Bioorg. Med. Chem.* **2003**, *11*, 5155–5170.
- [33] M. E. Kavanagh, J. L. Gray, S. H. Gilbert, A. G. Coyne, K. J. McLean, H. J. Davis, A. W. Munro, C. Abell, *ChemMedChem* **2016**, *11*, 1924–1935.
- [34] M. V. Toth, G. R. Marshall, *Int. J. Pept. Protein Res.* **1990**, *36*, 544–550.
- [35] H. C. Cheng, *J. Pharmacol. Toxicol. Methods* **2001**, *46*, 61–71.
- [36] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- [37] S. McNicholas, E. Potterton, K. S. Wilson, M. E. M. Noble, *Acta Crystallogr. Sect. D* **2011**, *67*, 386–394.

---

Manuscript received: July 4, 2018

Accepted manuscript online: September 3, 2018

Version of record online: October 9, 2018

**Part 3**

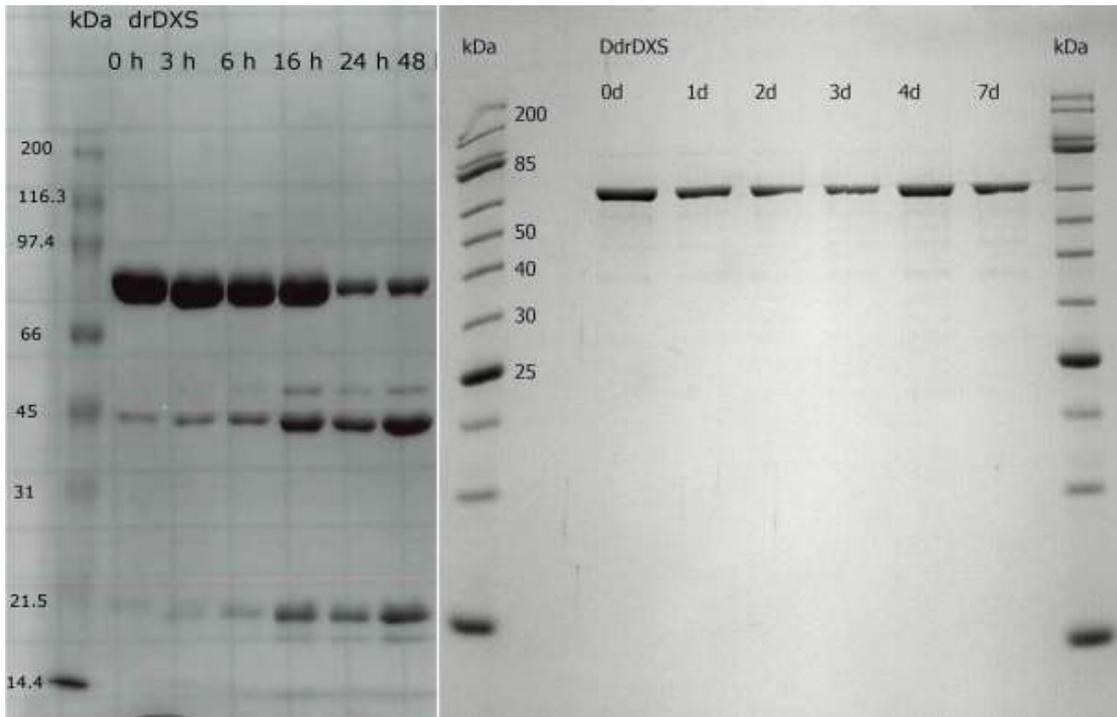
## **Supplementary Information**



## 3.1 Supplementary Information for Chapter 2.1

Supporting information:

Identification of a 1-deoxy-D-xylulose-5-phosphate synthase (DXS) mutant with improved crystallographic properties

SDS-PAGE of drDXS and  $\Delta$ drDXS

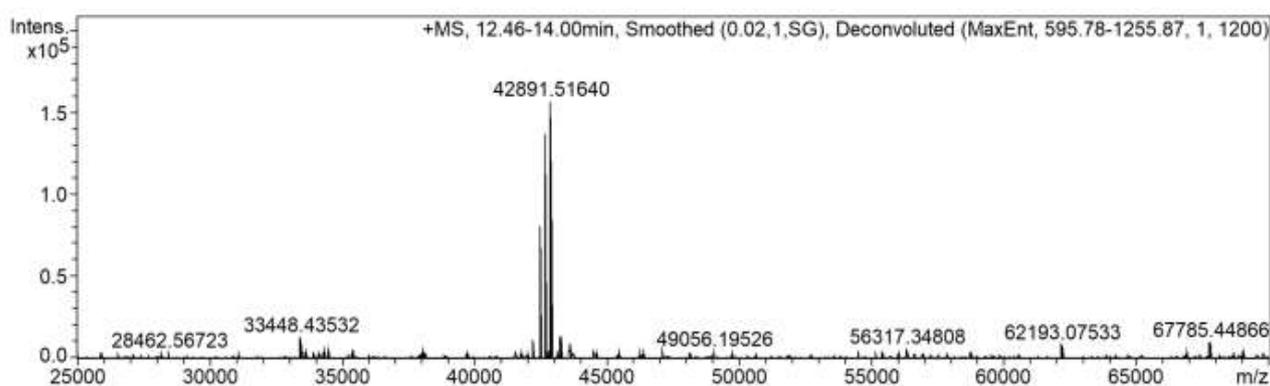
**Fig. S1:** 12% SDS-PAGE analysis of drDXS and  $\Delta$ drDXS over time. Degradation of the native enzyme was observed over a time period of 48 h. The appearance of the cleaving fragments of 40 and 20 kDa is visible. The truncated enzyme (Gel with DdrDXS label) shown no sign of truncation, even after an extended observation period of seven days. The degradation occurs most likely due to the flexible parts of the enzyme seeming to be very susceptible to minor protease contaminations.

## LC/MS measurements

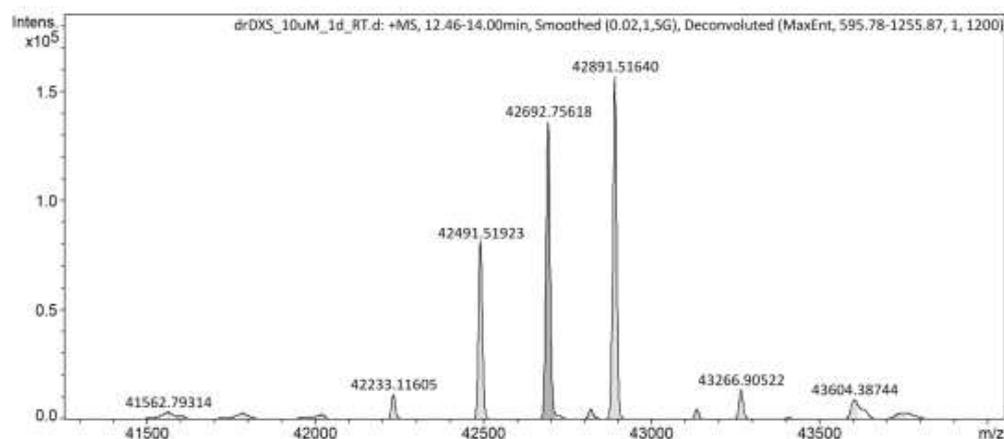
### Method

All ESI-MS-measurements were performed on a Dionex Ultimate 3000 RSLC system using an Aeris Widepore XB-C8, 150 x 2.1 mm, 3.6  $\mu\text{m}$  dp column (Phenomenex, USA). Separation of 1  $\mu\text{L}$  sample was achieved by a linear gradient from (A)  $\text{H}_2\text{O}$  + 0.1% FA to (B) ACN + 0.1% FA at a flow rate of 300  $\mu\text{L}/\text{min}$  and 45  $^\circ\text{C}$ . The gradient was initiated by a 0.5 min isocratic step at 2% B, followed by an increase to 75% B in 10 min to end up with a 3 min step at 75% B before re-equilibration with initial conditions. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75  $\mu\text{L}/\text{min}$  before entering the maXis 4G hr-ToF mass spectrometer (Bruker Daltonics, Bremen, Germany) using the standard Bruker ESI source. In the source region, the temperature was set to 200  $^\circ\text{C}$ , the capillary voltage was 4000 V, the dry-gas flow was 5.0 L/min and the nebulizer was set to 1.0 bar. Mass spectra were acquired in positive ionization mode ranging from 600 – 1800 m/z at 2.5 Hz scan rate. Protein masses were deconvoluted by using the Maximum Entropy algorithm (Copyright 1991-2004 Spectrum Square Associates, Inc.).

### LC/MS of drDXS

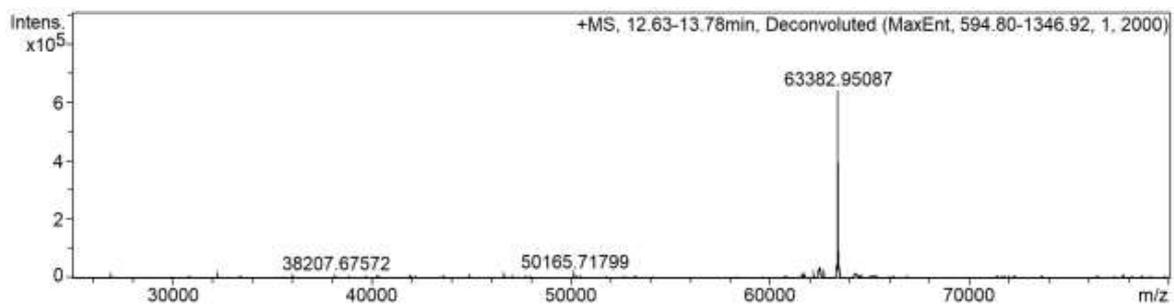


**Fig. S2:** Mass spectra of drDXS after one day of incubation at RT



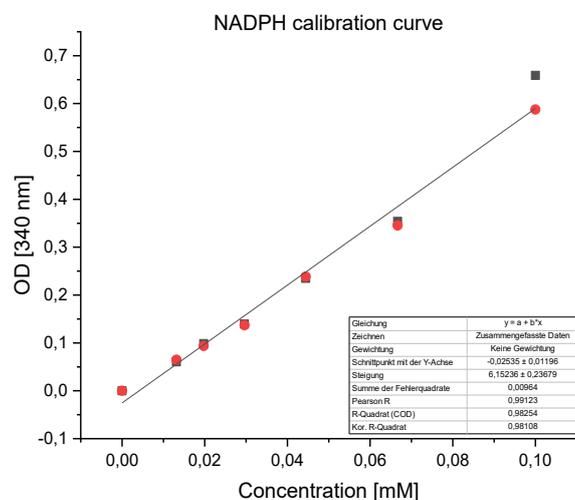
**Fig. S3:** Zoom into the main peak area of spectra shown in figure S2.

LC/MS of  $\Delta$ drDXS

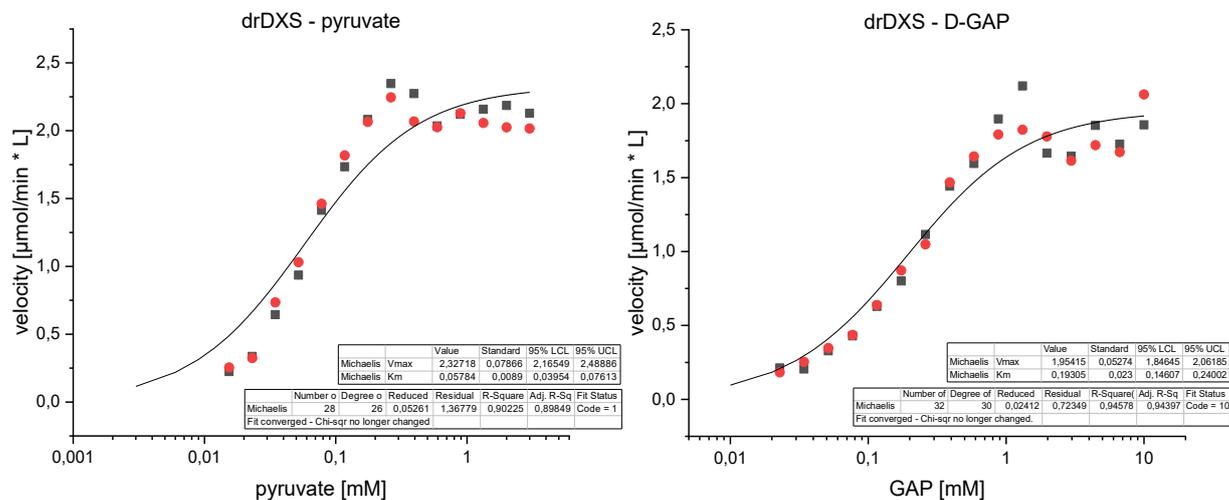


**Fig. S4:** Mass spectra of  $\Delta$ drDXS after one day of incubation at RT.

## Enzyme kinetic data



**Fig. S5:** NADPH calibration curve. Experiment was performed in duplicates and a concatenated linear fit was performed to determine the molar extinction coefficient. The determined extinction coefficient of  $6.15 \pm 0.23 \text{ L} \cdot \text{mmol}^{-1} \cdot \text{cm}^{-1}$  is close to the literature value of  $6.3 \cdot 10^3 \text{ L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$  [1].



**Fig. S6:**  $K_m$  determination for drDXS. Enzyme concentration was 50 nmol/L.

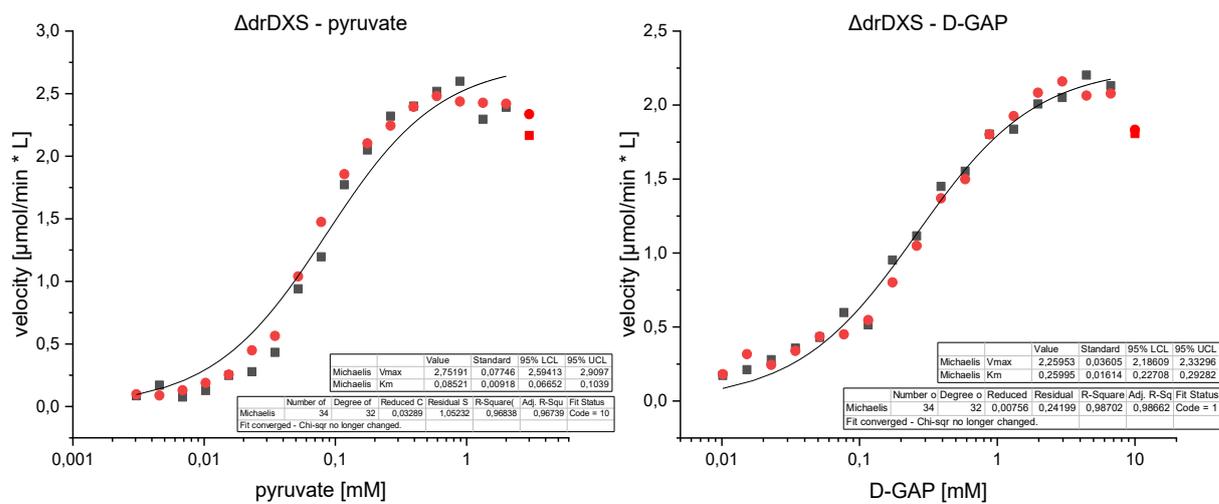
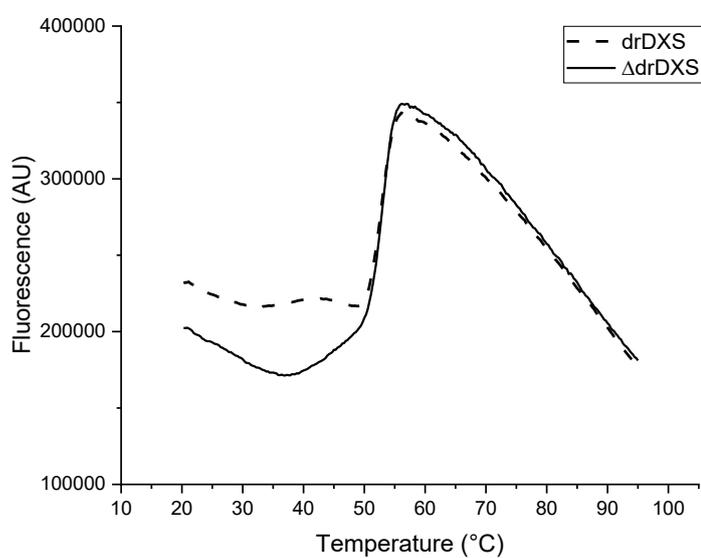
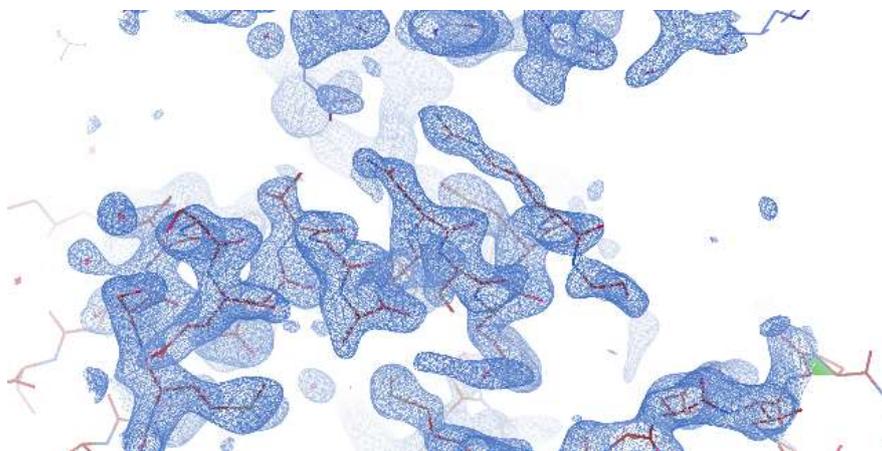


Fig. S7:  $K_m$  determination for  $\Delta drDXS$ . Enzyme concentration was 100 nmol/L.

## TSA measurements



**Fig. S8:** Thermal shift assay. Comparison of the truncated with the native DXS protein.



**Fig. S9:** Protein-protein interactions in the protein crystal of  $\Delta$ drDXS. Shown is the protein model and the electron density map 2fc-fo contoured at  $\sigma = 1.0$ .

## Data collection and refinement statistics

**Table S1:** Data collection and refinement statistics.

PDB code	6xxg
Wavelength	1.033
Resolution range	47.5 - 2.1 (2.175 - 2.1)
Space group	P 21 21 21
Unit cell	72.321 87.367 181.865 90 90 90
Total reflections	135671 (13425)
Unique reflections	67987 (6677)
Multiplicity	2.0 (2.0)
Completeness (%)	99.08 (99.48)
Mean I/sigma(I)	5.87 (0.86)
Wilson B-factor	25.64
R-merge	0.1792 (0.6797)
R-meas	0.2534 (0.9612)
R-pim	0.1792 (0.6797)
CC1/2	0.913 (0.444)
CC*	0.977 (0.784)
Reflections used in refinement	67468 (6678)
Reflections used for R-free	3424 (332)
R-work	0.1908 (0.2685)
R-free	0.2361 (0.2982)
CC(work)	0.889 (0.613)
CC(free)	0.938 (0.542)

Number of non-hydrogen atoms	8859
macromolecules	8316
ligands	63
solvent	480
Protein residues	1099
RMS(bonds)	0.013
RMS(angles)	1.79
Ramachandran favored (%)	97.98
Ramachandran allowed (%)	1.93
Ramachandran outliers (%)	0.09
Rotamer outliers (%)	2.80
Clashscore	2.52
Average B-factor	31.28
macromolecules	31.04
ligands	48.39
solvent	33.05

Statistics for the highest-resolution shell are shown in parentheses.

## Sequences

>truncated\_dr\_DXS\_GENE\_Construct

MGNELPGTSDTPLLDQIHGPKDLKRLSREQLPALTEELRGEIVRVC SRGGLHLASSLGAVDIITALHYVLDSPDRILFD  
VGHQAYAHKILTGRRDQMADIKKEGGISGFTKVSESEHDAITVGHASTSLANALGMALARDAQKDFHVAAVIGDGS LTG  
GMALAAALNTIGDMGRKMLIVLNDNEMSIENVGAMNKFMRGGGGGGSVNPF AAMGVRYVGPVDGHNVQELVWLLERLVD  
LDGPTILHIVTTKGKLSYAEADPIYWHGPAKFDPATGEYVPS SAYSWSAAFGEAVTEWAKTDPRTFVVT PAMREGSGLV  
EFSRVHPHYLDVGI AEEVAVTTAAGMALQGMRPVVAIYSTFLQRAYDQVLHDVAIEHLNVTFCIDRAGIVGADGATHNG  
VFDLSFLRSIPGVRIGLPKDAAE LRGMKYAQTHDGPFAIRYPRGNTAQVPAGTWPDLKWGEWERLKGDDVILAGGKA  
LDYALKA AEDLPGVGVNARFVKPLDEEMLREVGGRRALITVEDNTVVGFGGAVLEALNSMNLHPTVRVLGIPDEFQE  
HATAESVHARAGIDAPAIRTVLAELGVDVPIEV

>native\_drDXS\_construct

MGSSHHHHHSSGLVPRGSHMMNELPGTSDTPLLDQIHGPKDLKRLSREQLPALTEELRGEIVRVC SRGGLHLASSLGAV  
DIITALHYVLDSPDRILFDVGHQAYAHKILTGRRDQMADIKKEGGISGFTKVSESEHDAITVGHASTSLANALGMALAR  
DAQKDFHVAAVIGDGS LTGGMALAAALNTIGDMGRKMLIVLNDNEMSIENVGAMNKFMRGLQVQKWFQEGEGAGKKAVE  
AVSKPLADFMSRAKNSTRHFFDPASVNPF AAMGVRYVGPVDGHNVQELVWLLERLVDLDGPTILHIVTTKGKLSYAEAD  
PIYWHGPAKFDPATGEYVPS SAYSWSAAFGEAVTEWAKTDPRTFVVT PAMREGSGLVEFSRVHPHYLDVGI AEEVAVTT  
AAGMALQGMRPVVAIYSTFLQRAYDQVLHDVAIEHLNVTFCIDRAGIVGADGATHNGVFDLSFLRSIPGVRIGLPKDAE  
LRGMKYAQTHDGPFAIRYPRGNTAQVPAGTWPDLKWGEWERLKGDDVILAGGKALDYALKA AEDLPGVGVNARFVK  
PLDEEMLREVGGRRALITVEDNTVVGFGGAVLEALNSMNLHPTVRVLGIPDEFQEHATAESVHARAGIDAPAIRTVLA  
ELGVDVPIEV

### Multiple sequence alignment

Alignment is given in .FASTA –format. Please copy/paste into viewer of your choice.

In the published supporting information, this position shows a multiple sequence alignment of all (498) annotated bacterial DXS homologues. For the sake of clarity, the 150 pages of the alignment are not shown in this thesis.

The original SI is accessible at <https://doi.org/10.1016/j.bbrc.2020.12.069>

## Literature

- [1] H.U. Bergmeyer, New values for the molar extinction coefficients of NADH and NADPH for the use in routine laboratories (author's transl), *Z. Klin. Chem. Klin. Biochem.* 13 (1975) 507–8.  
<http://www.ncbi.nlm.nih.gov/pubmed/3038> (accessed November 30, 2017).

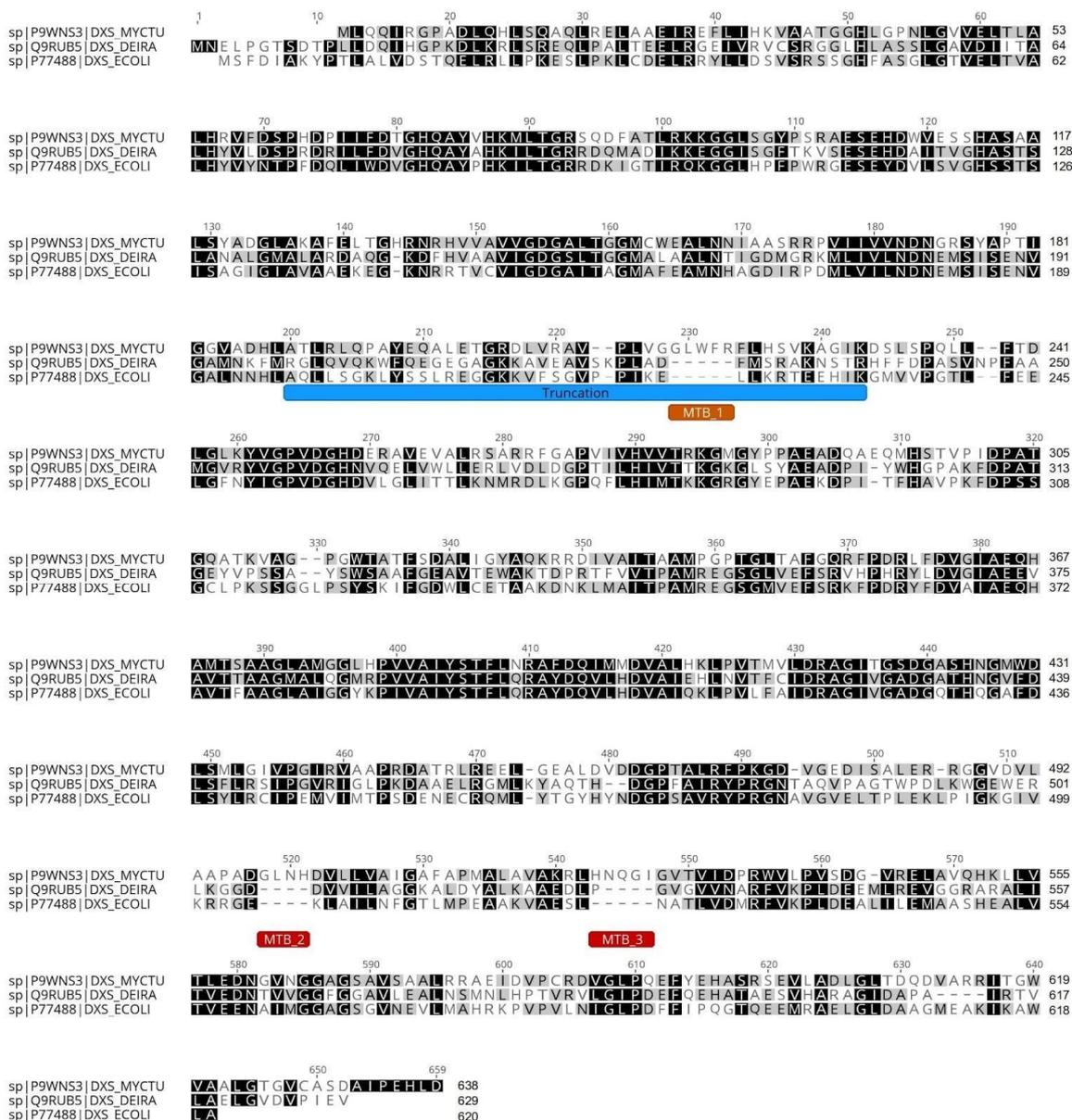


## 3.2 Supplementary Information for Chapter 2.2

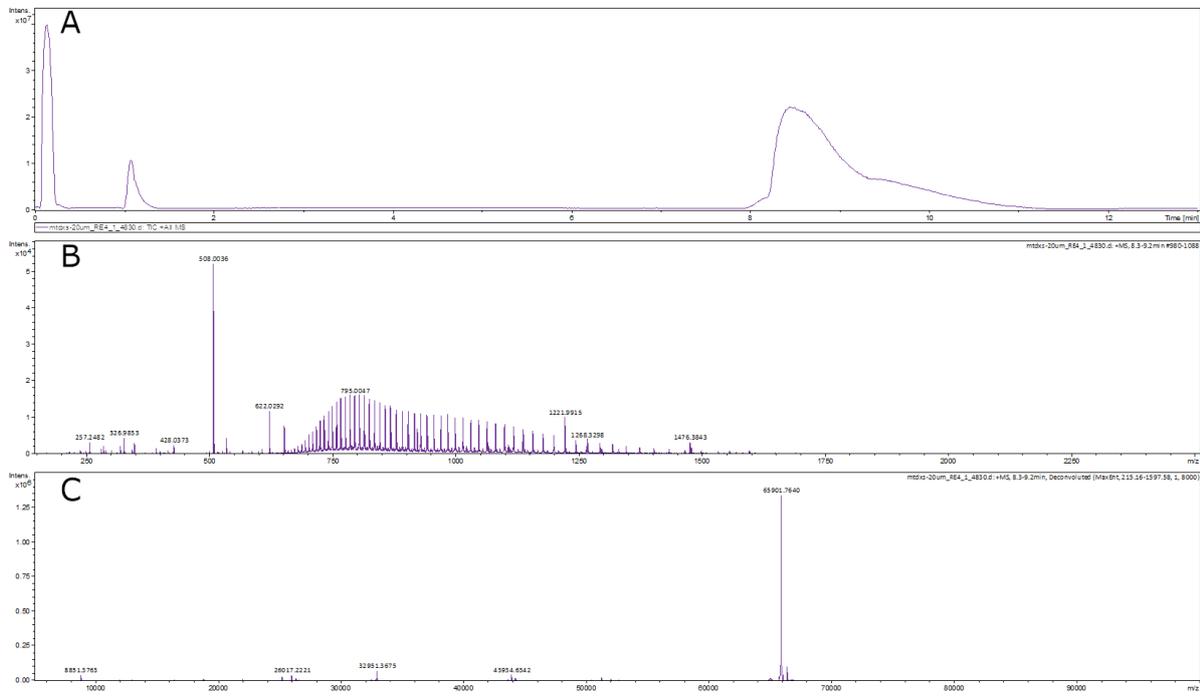
**Supplemental Information: First crystal structures of 1-deoxy-D-xylulose 5-phosphate synthase (DXPS) from *Mycobacterium tuberculosis* indicate a distinct mechanism of intermediate stabilization.**

Robin M. Gierse <sup>a,b,c</sup>, Rick Oerlemans <sup>d</sup>, Victor O. Gawriljuk <sup>e</sup>, Eswar Reddem <sup>c,d</sup>, Alaa Alhayek <sup>a,b</sup>, Dominik Baitinger <sup>a</sup>, Harald Jakobi <sup>f</sup>, Bernd Laber <sup>f</sup>, Gudrun Lange <sup>f</sup>, Anna K.H. Hirsch\* <sup>a,b,c</sup> and Matthew R. Groves\* <sup>d</sup>

Sequence alignment of DXPS from <i>D. radiodurans</i> , <i>E. coli</i> and <i>M. tuberculosis</i>	2
LC-MS analysis of $\Delta$ mtDXPS	3
Enzyme stability	4
Enzyme kinetics	5
MEP pathway enzymes	6
Collection and refinement statistics	7
Sequences	10

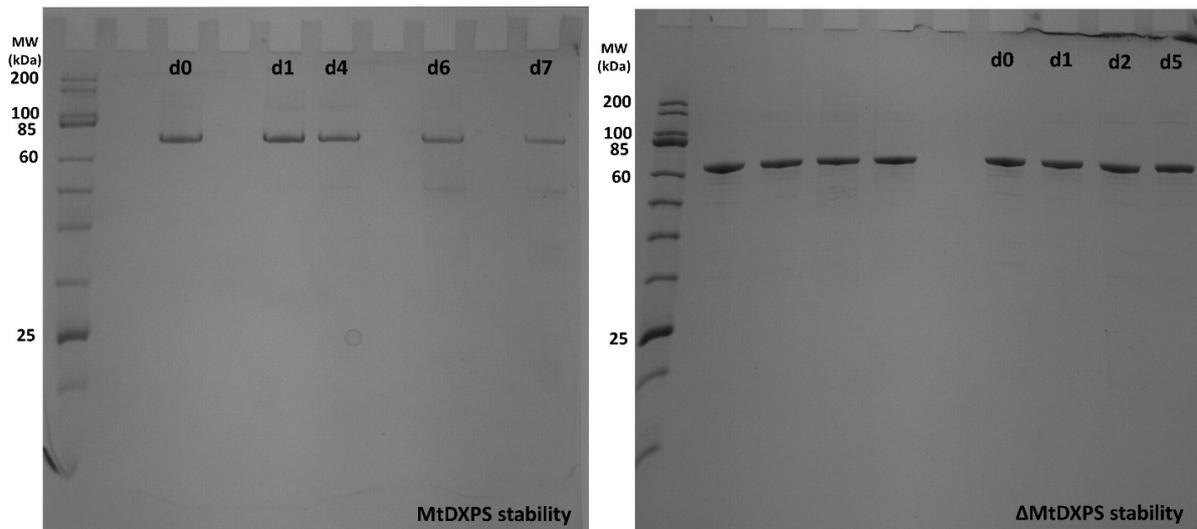
Sequence alignment of DXPS from *D. radiodurans*, *E. coli* and *M. tuberculosis*

**Figure S1:** Sequence alignment of the three crystallized DXPS homologues from *D. radiodurans*, *E. coli* and *M. tuberculosis*. Sequence inserts of *M. tuberculosis* are annotated in red, the site of truncation in blue.

LC-MS analysis of  $\Delta$ mtDXPS

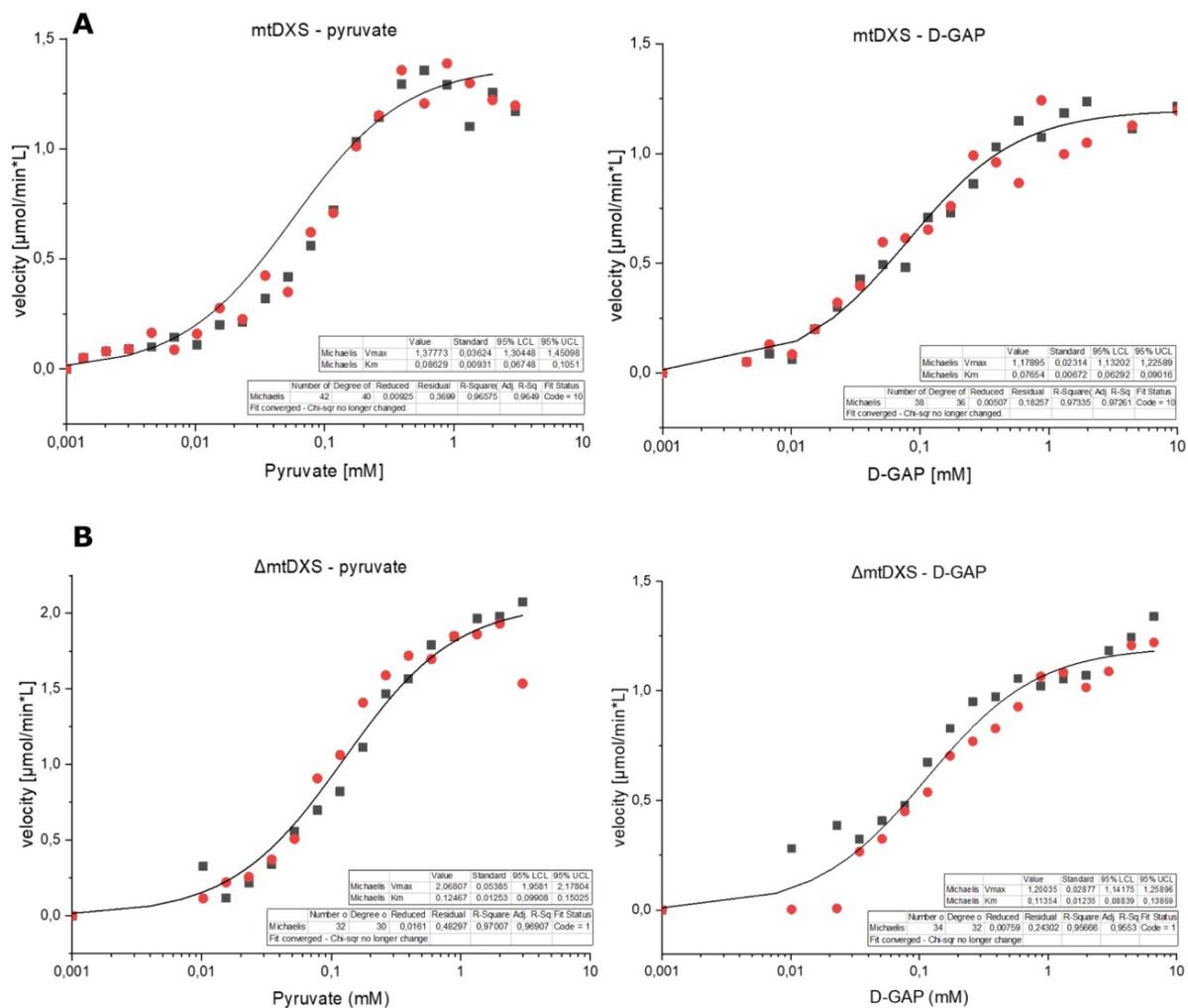
**Figure S2:** LC-MS analysis of a 20  $\mu$ M sample of  $\Delta$  mtDXPS. **A:** Total ion count (TIC) spectra of the sample. **B:** LC-MS spectrum of the peak at minute 8.3 to 9.2. **C:** Deconvoluted mass spectrum of the peak from minute 8.3 to 9.2. The full mass of the mtDXPS protein can be observed with a m/z of 75901.67.

### Enzyme stability



**Figure S3:** SDS-PAGE analysis of MtDXPS (left gel) and  $\Delta$ MtDXPS (right gel), incubated at RT. The native enzyme is showing bands of degradation products, starting from day one and increasing in intensity over time. Reduction of the main band intensity can be observed from day four on. For the truncated enzyme (lanes 5 to 8), no change of band intensity can be observed up until day five.

## Enzyme kinetics



**Figure S4:**  $K_m$  determination for the mtDXPS enzymes. **A:** mtDXPS, enzyme concentration used: 5  $\mu\text{mol/L}$  **B:**  $\Delta\text{mtDXPS}$ , enzyme concentration used: 2  $\mu\text{mol/L}$ . Calculation of  $K_{cat}$  was done using the following equation:  $V_{max} / ([\text{Enz}] * 60) = K_{cat} * s^{-1}$

**MEP pathway enzymes****Table S1:** Overview of the MEP-pathway enzymes, the corresponding number of published protein crystal structures and their source organisms. The search was conducted in the PDB archive in February 2021.

<i>Enzyme</i>	<i>Number of structures</i>	<i>Species</i>
<i>DXPS</i>	5	<i>Escherichia coli</i> , <i>Deinococcus radiodurans</i>
<i>IspC</i>	74	<i>Acinetobacter baumannii</i> , <i>E.coli</i> , <i>Moraxella catarrhalis</i> , <i>M. tuberculosis</i> , <i>Plasmodium falciparum</i> , <i>Staphylococcus schleiferi</i> , <i>Thermotoga maritima</i> , <i>Vibrio vulnificus</i> , <i>Yersinia pestis</i> , <i>Yersinia pseudotuberculosis</i> , <i>Zymomonas mobilis</i>
<i>IspD</i>	43	<i>Anaerococcus prevotii</i> , <i>Arabidopsis thaliana</i> , <i>Bacillus subtilis</i> , <i>Burkholderia thailandensis</i> , <i>Campylobacter jejuni</i> , <i>E. coli</i> , <i>Homo sapiens</i> , <i>Listeria monocytogenes</i> , <i>Mycobacterium avium</i> , <i>M. tuberculosis</i> , <i>Neisseria gonorrhoeae</i> , <i>T. maritima</i> , <i>Thermus thermophilus</i>
<i>IspE</i>	16	<i>Aquifex aeolicus</i> , <i>M. tuberculosis</i> , <i>T. thermophilus</i>
<i>IspF</i>	70	<i>A. thaliana</i> , <i>B. subtilis</i> , <i>Burkholderia cenocepacia</i> , <i>Burkholderia pseudomallei</i> , <i>C. jejuni</i> , <i>E. coli</i> , <i>Francisella tularensis</i> , <i>Haemophilus influenza</i> , <i>Mycobacterium smegmatis</i> , <i>P. falciparum</i> , <i>Plasmodium vivax</i> , <i>Salmonella enterica</i> , <i>Shewanella oneidensis</i> , <i>Thermus thermophiles</i> , <i>Y. pestis</i>
<i>IspG</i>	13	<i>Aquifex aeolicus</i> , <i>Bacillus anthracis</i> , <i>T. thermophilus</i>
<i>IspH</i>	12	<i>E. coli</i> , <i>P. falciparum</i>

**Collection and refinement statistics****Table S2:** Collection and refinement statistics

<b>PDB-ID:</b>	<b>7A9H</b>	<b>7A9G</b>
Wavelength	1.033	1.033
Resolution range	43.79 - 1.85 (1.916 - 1.85)	48.5 - 1.9 (1.968 - 1.9)
Space group	P 1	P 1 21 1
Unit cell	62.987 76.1 79.148 108.956 106.53 107.594	62.865 126.17 79.031 90 106.285 90
Total reflections	165951 (10815)	175878 (17114)
Unique reflections	93233 (6412)	92339 (9196)
Multiplicity	1.8 (1.7)	1.9 (1.9)
Completeness (%)	91.00 (62.53)	99.39 (99.38)
Mean I/sigma(I)	20.99 (6.54)	12.40 (2.52)
Wilson B-factor	18.62	22.12
R-merge	0.02334 (0.08525)	0.04858 (0.3653)
R-meas	0.03298 (0.1205)	0.0687 (0.5167)
R-pim	0.0233 (0.08508)	0.04858 (0.3653)
CC1/2	0.999 (0.979)	0.997 (0.779)
CC*	1 (0.995)	0.999 (0.936)
Reflections used in refinement	93233 (6412)	92339 (9195)
Reflections used for R-free	2100 (144)	971 (97)
R-work	0.1506 (0.1870)	0.1538 (0.2398)
R-free	0.1736 (0.2225)	0.1903 (0.2823)
CC(work)	0.960 (0.904)	0.956 (0.720)
CC(free)	0.947 (0.916)	0.958 (0.684)
Number of non-hydrogen atoms	9028	8767

macromolecules	8308	8147
ligands	54	220
solvent	666	400
Protein residues	1091	1086
RMS(bonds)	0.014	0.014
RMS(angles)	1.72	a1.75
Ramachandran favored (%)	97.96	97.85
Ramachandran allowed (%)	2.04	2.15
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	0.70	0.48
Clashscore	2.39	2.33
Average B-factor	20.85	24.13
macromolecules	20.47	23.50
ligands	14.64	39.85
solvent	26.03	28.30

Statistics for the highest-resolution shell are shown in parentheses.

**Table S3:** Uniprot code and Species name of all sequences from the MSA [1] containing the ESSH motif. Duplicates have not been removed.

Uniprot Sequence code	Species
Q4JVB5 DXS_CORJK	<i>Corynebacterium jeikeium</i>
Q6NGV3 DXS_CORDI	<i>Corynebacterium diphtheriae</i>
Q8FPI2 DXS_COREF	<i>Corynebacterium efficiens</i>

A4QEQ9 DXS_CORGB	<i>Corynebacterium glutamicum (strain R)</i>
Q8NPB2 DXS_CORGL	<i>Corynebacterium glutamicum</i>
B0RC26 DXS_CLAMS	<i>Clavibacter michiganensis subsp. sepedonicus</i>
Q6AFD5 DXS_LEIXX	<i>Leifsonia xyli subsp. xyli (strain CTCB07)</i>
B2GJ56 DXS_KOCRD	<i>Kocuria rhizophila (strain ATCC 9341 / DSM 348)</i>
A9WRA9 DXS_RENSM	<i>Renibacterium salmoninarum</i>
A1R5N7 DXS_PAEAT	<i>Paenarthrobacter aurescens</i>
B8HH36 DXS_PSECP	<i>Pseudarthrobacter chlorophenolicus</i>
A0JVG9 DXS_ARTS2	<i>Arthrobacter sp. (strain FB24)</i>
Q5YTA2 DXS_NOCFA	<i>Nocardia farcinica</i>
C0ZYV9 DXS_RHOE4	<i>Rhodococcus erythropolis</i>
Q0S1H1 DXS_RHOJR	<i>Rhodococcus jostii</i>
P0A555 DXS_MYCBO	<i>Mycobacterium bovis</i>
C1AFE1 DXS_MYCBT	<i>Mycobacterium bovis</i>
A1KM20 DXS_MYCBP	<i>Mycobacterium bovis (strain BCG / Pasteur 1173P2)</i>
A5U634 DXS_MYCTA	<i>Mycobacterium tuberculosis (strain ATCC 25177 / H37Ra)</i>
A0PT40 DXS_MYCUA	<i>Mycobacterium ulcerans (strain Agy99)</i>
A0QIL6 DXS_MYCA1	<i>Mycobacterium avium (strain 104)</i>
B8ZQW9 DXS_MYCLB	<i>Mycobacterium leprae (strain Br4923)</i>

Q50000 DXS_MYCLE	<i>Mycobacterium leprae (strain TN)</i>
A1UF44 DXS_MYCSK	<i>Mycobacterium sp. (strain KMS)</i>
Q1B9W8 DXS_MYCSS	<i>Mycobacterium sp. (strain MCS)</i>
A3PYK6 DXS_MYCSJ	<i>Mycobacterium sp. (strain JLS)</i>
A1T7Z0 DXS_MYCVP	<i>Mycolicibacterium vanbaalenii</i>
A4TCS5 DXS_MYCGI	<i>Mycolicibacterium gilvum</i>
A0QW19 DXS_MYCS2	<i>Mycolicibacterium smegmatis</i>
Q73W57 DXS_MYCPA	<i>Mycolicibacterium paratuberculosis</i>
B1MCU7 DXS_MYCA9	<i>Mycobacteroides abscessus</i>
P9WNS3 DXS_MYCTU	<i>Mycobacterium tuberculosis</i>
P9WNS2 DXS_MYCTO	<i>Mycobacterium tuberculosis</i>

## Sequences

>sp|P9WNS3|DXS\_MYCTU 1-deoxy-D-xylulose-5-phosphate synthase OS=*Mycobacterium tuberculosis* (strain ATCC 25618 / H37Rv) OX=83332 GN=dxs PE=1 SV=1

MLQQIRGPADLQHLSQLRELA AEIREFLIHKVAATGGHLGPNLGVVELTLALHRVFDSPHDPIIFDTGHQAYVHKML  
TGRSQDFATLRKKGGLSGYPSRAESEHDWVESSHASAALSADGLAKAFELTGHRNRHVAVVGDGALTGGMCWEA  
LNNIAASRRPVIIIVNDNGRSYAPTIGGVADHLATLRLQPAYEQALETGRDLVRAVPLVGGWFRFLHSVKAGIKDSLSP  
QLLFTDLGLKYVGPVDGHERAVEVALRSARRFGAPVIVHVTRKGMGYPPAEADQAEQMHSSTVPIDPATGQATKVA  
GPGWTATFS DALIGYAQKRRDIVAITAAMPGPTGLTAFGQRFDPDLFDVGLAEQHAMTSAAGLAMGGLHPVVAIYSTF  
LNRAFDQIMMDVALHKLPTMVLDRAGITGSDGASHNGMWDLSMLGIVPGIRVAAPRDATRLREELGEALDVEDDGP  
TALRFPKGDVGEDISALERRGGVDVLAAPADGLNHDVLLVAIGAFAPMALAVAKRLHNQIGVTVIDPRWVLPVSDGV  
RELAVQHKLVTLEDNGVNGGAGSAVSAALRRAEIDVPCRVDVGLPQEFYEHASRSEVLADLGLTDQDVARRITGWVAA  
LGTGVCASDAIPEHLD

**> $\Delta$ MtDXPS**

MKHHHHHPMSDYDIPTTENLYFQGAMGMLQQIRGPADLQHLSQAQLRELA AEIREFLIHKVAATGGHLGPNLGVV  
ELTLALHRVFDSPHDPIIFDTGHQAYVHKMLTGRSQDFATLRKKGGLSGYPSRAESEHDWVESSHASAALSADGLAKA  
FELTGHRNRHVAVVVGDGALTGGMCWEALN NIAASRRPVIIIVNDNGRSYAPTIGGVADHLAGGGGGGGPQLLFTD  
LGLKYVGPVDGHD ERAVEVALRSARRFGAPVIVHVVTRKGMGYPPAEADQAEQMHSTVPIDPATGQATKVAGPGWT  
ATFSDALIGYAQRRDIVAITAAMPGPTGLTAFGQRFDPDLFDVGIAEQHAMTSAAGLAMGGLHPVVAIYSTFLNRAF  
DQIMMDVALHKLPVTMVLDRAGITGSDGASHNGMWDLSMLGIVPGIRVAAPRDATRLREELGEALD VDDGPTALRF  
PKGDVGEDISALERRGGVDVLAAPADGLNHDVLLVAIGAFAPMALAVAKRLHNQIGVTVIDPRWVLPVSDGVRELA  
VQHKLVTLEDNGVNGGAGSAVSAALRRAEIDVPCR DVGLPQEFYEHASRSEVLADLGLTDQDVARRITGWVAALGTG  
VCASDAIPEHLD

1. Gierse, R. M., Reddem, E. R., Alhayek, A., Baitinger, D., Hamid, Z., Jakobi, H., Laber, B., Lange, G., Hirsch, A. K. H., and Groves, M. R. (2021) Identification of a 1-deoxy-D-xylulose-5-phosphate synthase (DXS) mutant with improved crystallographic properties. *Biochem. Biophys. Res. Commun.* **539**, 42–47



### 3.3 Supplementary Information for Chapter 2.3

## Supplementary Information

### Identification of three new inhibitor classes against *Plasmodium falciparum*

Sandra Johannsen<sup>1,2,+</sup>, Robin M. Gierse<sup>1,2,3,+</sup>, Arne Krüger<sup>4</sup>, Rachel L. Edwards<sup>5</sup>, Vittoria Nanna<sup>1</sup>, Anna Fontana<sup>1</sup>, Zhu Di<sup>1,3</sup>, Tiziana Masini<sup>3</sup>, Lais Pessanha de Carvalho<sup>6</sup>, Mael Poizat<sup>7</sup>, Bart Kieftenbelt<sup>7</sup>, Dana M. Hodge<sup>8</sup>, Sophie Alvarez<sup>9</sup>, Daan Bunt<sup>3</sup>, Kamila Anna Meissner<sup>4</sup>, Edmarcia Elisa de Souza<sup>4</sup>, Melloney Dröge<sup>7</sup>, Bernard van Vliet<sup>7</sup>, Jack den Hartog<sup>7</sup>, Michael C. Hutter<sup>10</sup>, Jana Held<sup>6</sup>, Audrey R. Odom John<sup>8</sup>, Carsten Wrenger<sup>4</sup>, Anna K. H. Hirsch<sup>1,2,3,\*</sup>

1 Helmholtz Institute for Pharmaceutical Research Saarland (HIPS) – Helmholtz Centre for Infection Research (HZI), Campus Building E8.1, 66123 Saarbrücken, Germany

2 Department of Pharmacy, Saarland University, Campus Building E8.1, 66123 Saarbrücken, Germany

3 Stratingh Institute for Chemistry, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

4 Unit for Drug Discovery, Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, Av. Prof. Lineu Prestes 1374, 05508-000 São Paulo-SP, Brazil

5 Department of Pediatrics, Washington University School of Medicine, Saint Louis, Missouri 63110, United States

6 Institute of Tropical Medicine, University of Tübingen, Wilhelmstraße 27, 72074 Tübingen, Germany

7 Symeres, Kadijk 3, 9747 AT Groningen, The Netherlands

8 Department of Pediatrics, Children's Hospital of Philadelphia, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA 19104, United States

9 Proteomics & Metabolomics Facility, Center for Biotechnology, Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, Nebraska 68588, United States

10 Center for Bioinformatics, Saarland University, Campus Building E2.1, 66123 Saarbrücken, Germany

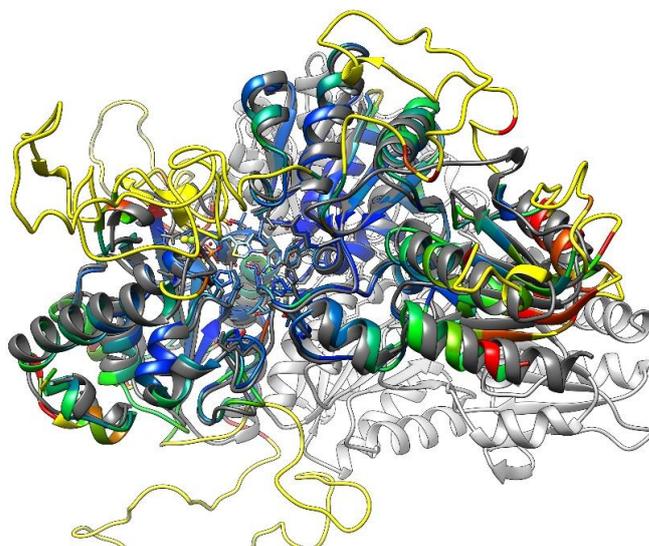
## Table of content

Solubility test of oximes.....	2
Comparison of actives sites of <i>MtDXPS</i> and <i>PfDXPS</i> .....	3
Results IDP rescue assay against <i>P. falciparum</i> 3D7 .....	4
Results LC-MS assay.....	7
Results of IC <sub>50</sub> determination against <i>PfDXPS</i> and <i>PfTPK</i> overexpressing <i>P. falciparum</i> 3D7 .....	11
Inhibition data for all compounds.....	12
Docking and scoring.....	22
Human off-traget enzymes.....	26
Table of compounds as SMILES with biological activity and PAINS count. ....	27
References.....	29

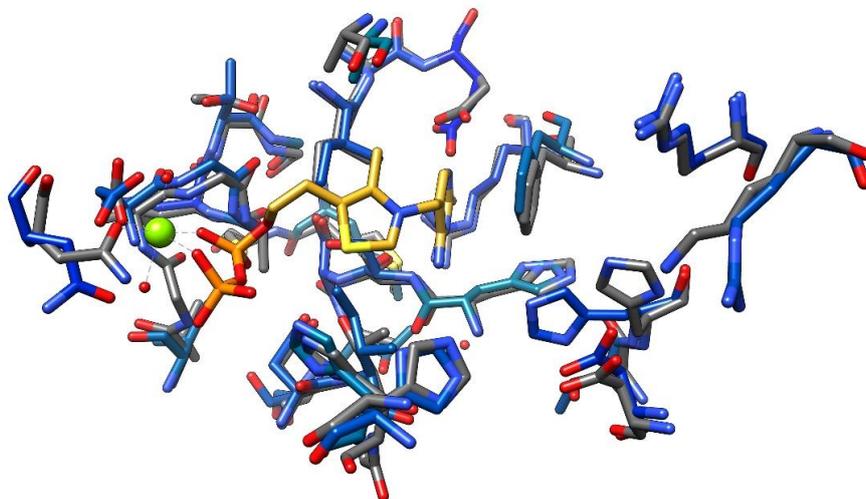
## Solubility test of oximes

**Table S 1:** Solubility in PBS buffer and 2% DMSO of new oxime derivatives in comparison to the original hit **2**.

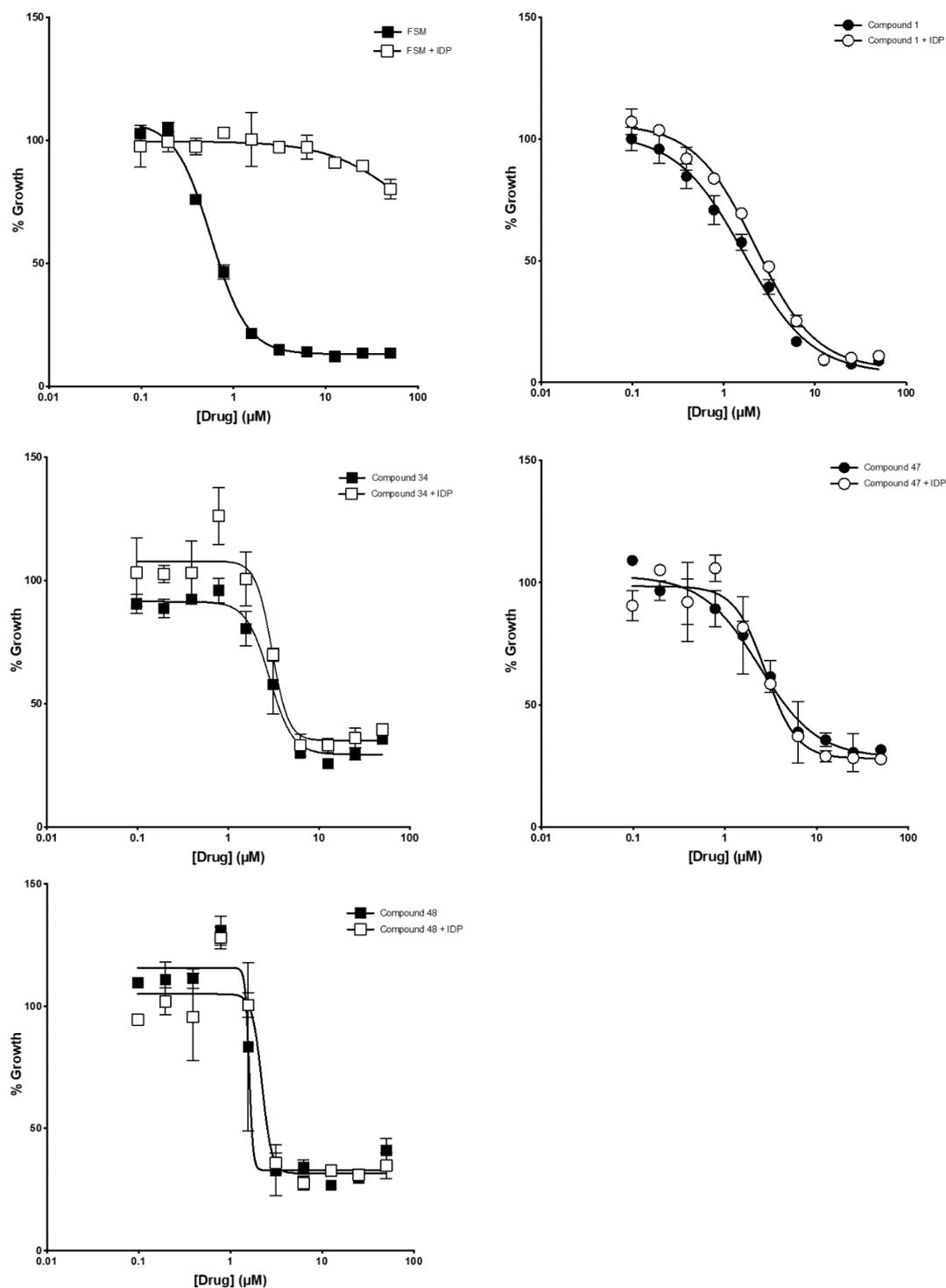
Compound	Max. solubility [ $\mu$ M]
<b>2</b>	114.2 $\pm$ 6.7
<b>4</b>	202.6 $\pm$ 1.6
<b>5</b>	193.6 $\pm$ 95.7
<b>6</b>	119.4 $\pm$ 27.4

Comparison of active sites of *Mt*DXPS and *Pf*DXPS

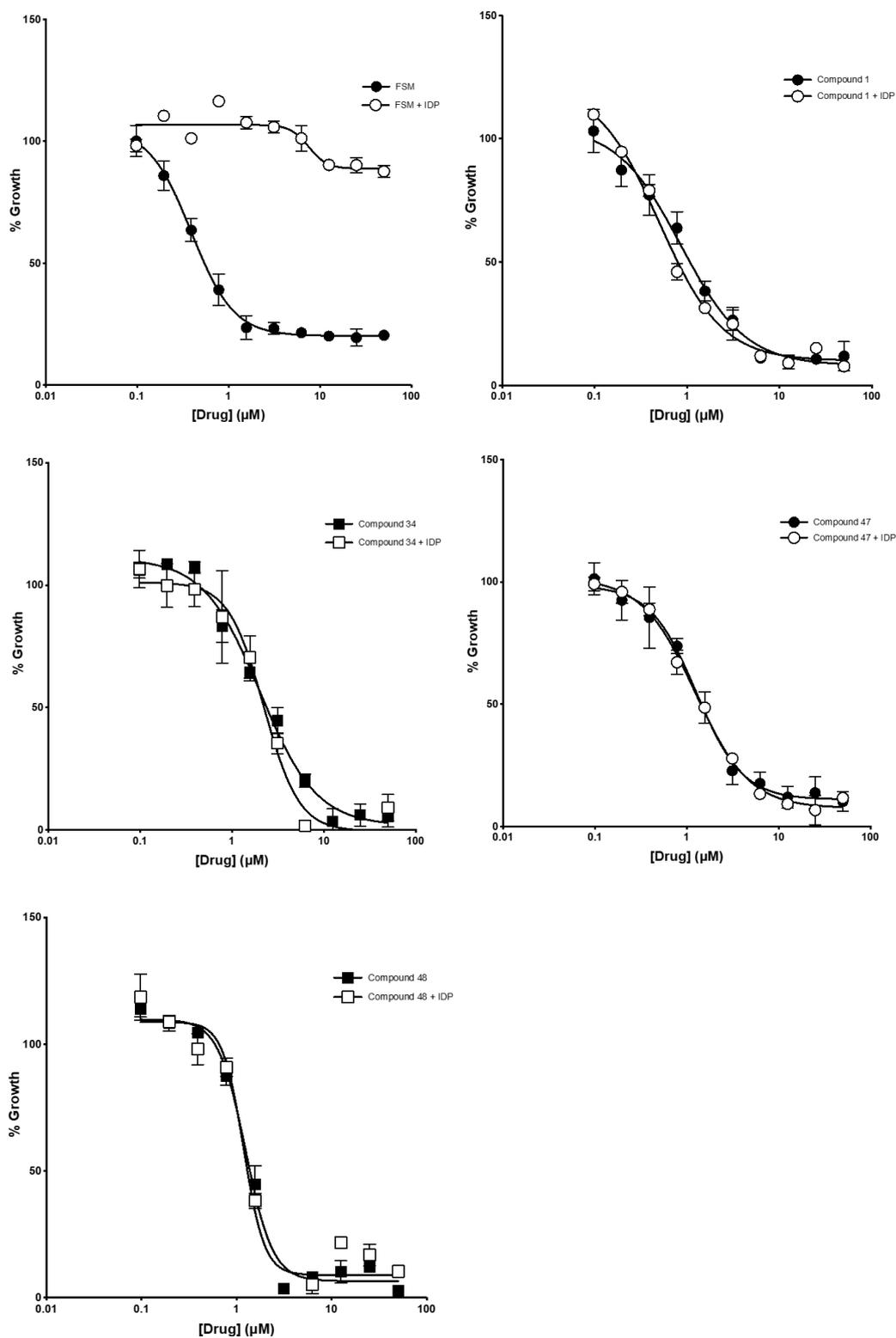
**Figure S 1:** Superposition of *Mt*DXPS crystal structure (grey, PDB: 6a9h) with the *Pf*DXPS homology model based on *dt*DXPS structure 2o1x. The core of the enzyme shows a high similarity with the *Mt*DXPS structure. *Pf*DXPS is coloured by C $\alpha$ -RMSD to the *Mt*DXPS structure, from blue = low RMSD via green to red = high RMSD; yellow: no corresponding amino acids in *Mt*DXPS structure. Typical for *P. falciparum*, its structure shows many additional loops on the surface of the protein.



**Figure S 2:** Superposition of the active site of *Mt*DXPS and *Pf*DXPS. The *Pf*DXPS amino acids are coloured by C $\alpha$ -RMSD to the *Mt*DXPS structure, from blue = low RMSD via green to red = high RMSD; the carbon atoms of the ThDP ligand are shown in gold; heteroatoms are coloured in red for oxygen, blue for nitrogen and yellow for sulphur atoms; the green sphere represents a Mg<sup>2+</sup>-ion. Most amino acids are found in identical arrangements, except His941 (corresponding to His416 of *Mt*DXPS) at the bottom right, which is predicted to be in a different rotameric state.

Results IDP rescue assay against *P. falciparum* 3D7

**Figure S 3:** IDP rescue assay: Representative graphs of growth inhibition of *P. falciparum* 3D7 of compounds **1**, **34**, **47**, **48** and fosmidomycin (FSM) as positive control.

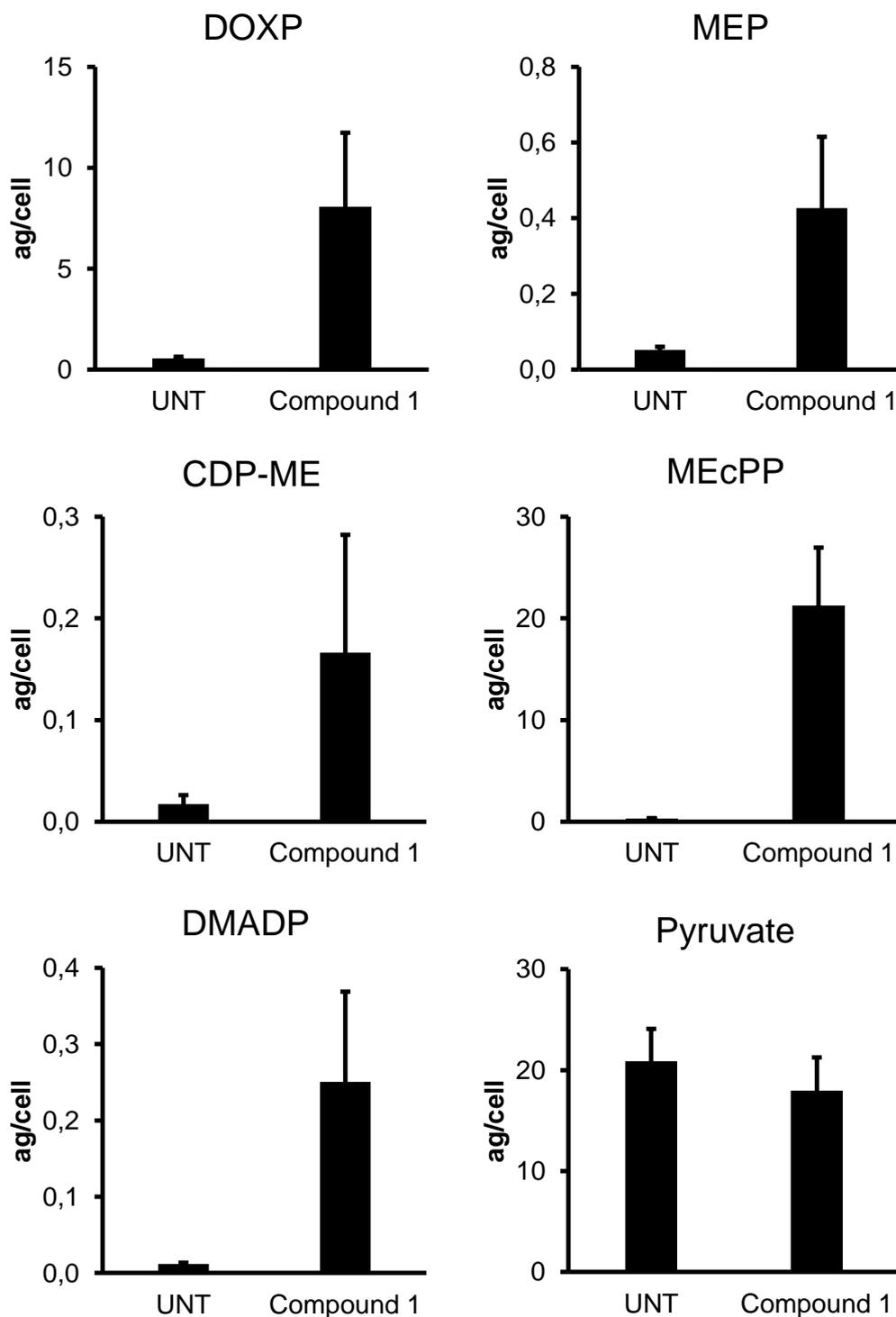


**Figure S 4:** IDP rescue assay: Representative graphs of growth inhibition of *P. falciparum* NF54 of compounds **1**, **34**, **47**, **48** and fosmidomycin (FSM) as positive control.

**Table S 2:** IDP rescue assay: IC<sub>50</sub> values with and without IDP against *P. falciparum* 3D7 and NF54. SD is displayed as error measure when more than one experiment was performed. a = mean of triplicates, b = mean of duplicates.

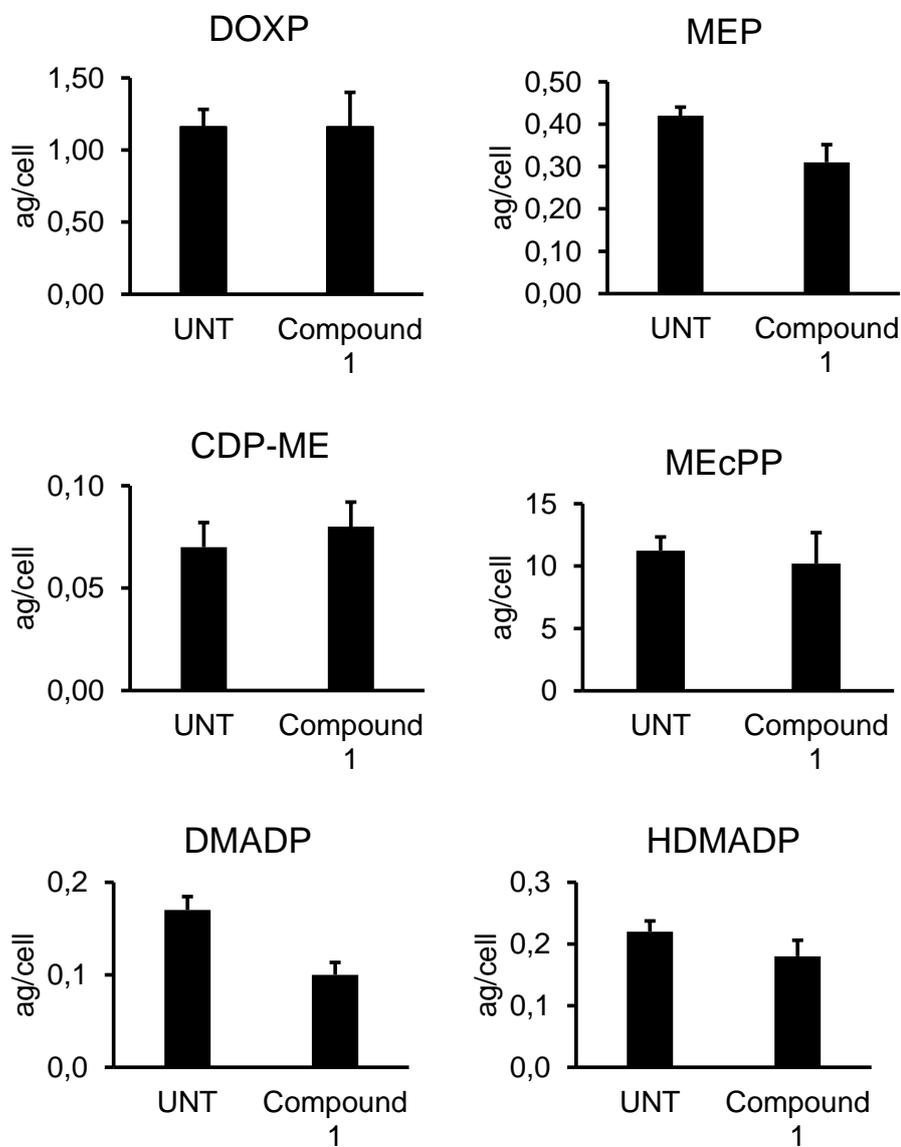
<b>Cmpd</b>	<b>IC<sub>50</sub> 3D7 [μM]</b>	<b>IC<sub>50</sub> NF54 [μM]</b>
FSM	0.9 ± 0.4 <sup>a</sup>	0.9 ± 0.5 <sup>b</sup>
FSM + IDP	37.4 ± 18.2 <sup>a</sup>	7.8 ± 0.3 <sup>b</sup>
<b>1</b>	1.9 ± 0.2 <sup>b</sup>	0.8 ± 0.1
<b>1 + IDP</b>	2.4 ± 0.2 <sup>b</sup>	0.5
<b>34</b>	2.7 ± 0.3	2.4 ± 0.6
<b>34 + IDP</b>	3.1	2.2
<b>47</b>	2.3 ± 0.4	2.1 ± 0.8
<b>47 + IDP</b>	2.8	1.2
<b>48</b>	2.0 ± 0.7	1.6 ± 0.3
<b>48 + IDP</b>	2.2	1.2

## Results LC-MS assay

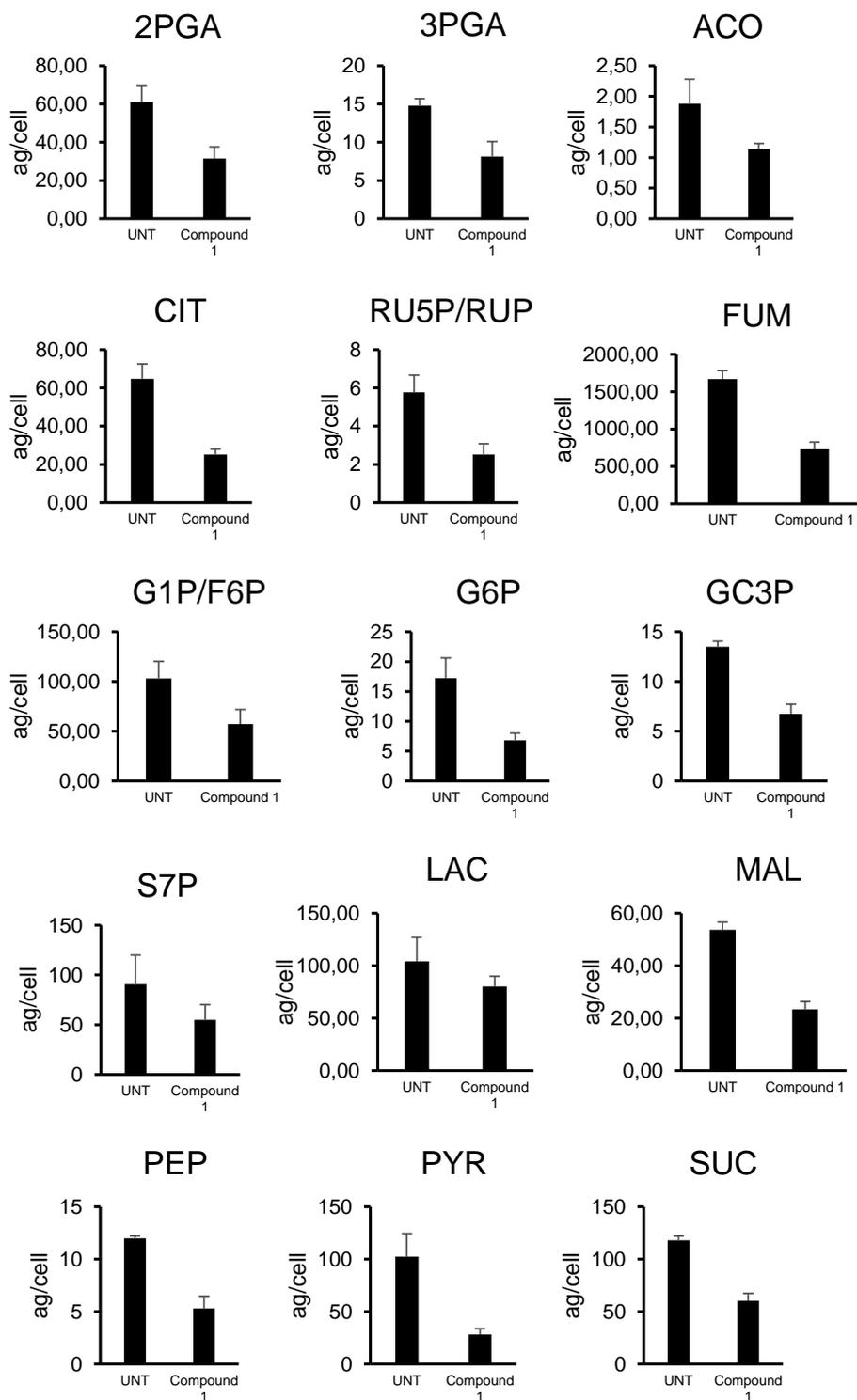


**Figure S 5:** LC-MS data of MEP pathway metabolites in *E. coli* after treatment with compound 1 at ten times the  $IC_{50}$ . Pyr = pyruvate, DOXP = 1-deoxy-D-xylulose-5-phosphate, MEP = methylerythritol phosphate, CDP-ME = 4-diphosphocytidyl-2-C-

methyl-D-erythritol, MEcDP = methylerythritol cyclodiphosphate, DMADP = dimethylallyl diphosphate, UNT = untreated.



**Figure S 6:** LC-MS data of MEP pathway metabolites in *P. falciparum* with and without treatment with compound 1 at five times the  $IC_{50}$ . Pyr = pyruvate, DOXP = 1-deoxy-D-xylulose-5-phosphate, MEP = methylerythritol phosphate, CDP-ME = 4-diphosphocytidyl-2-C-methyl-D-erythritol, MEcDP = methylerythritol cyclodiphosphate, DMADP = dimethylallyl diphosphate, HDMADP = 1-hydroxy-2-methyl-2-buten-4-yl 4-diphosphate, UNT = untreated.



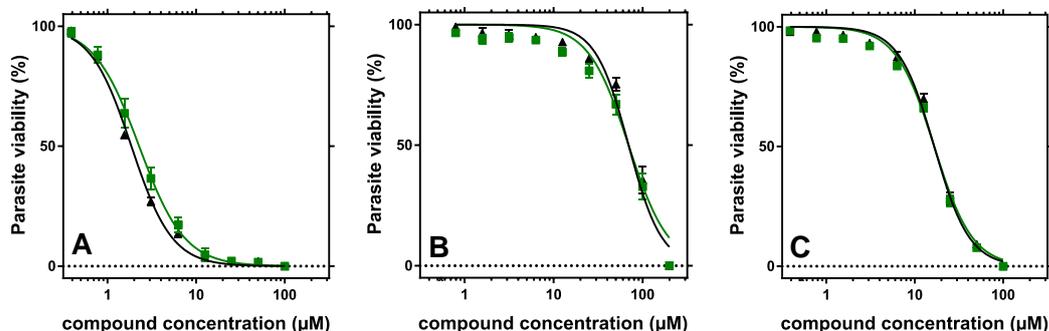
**Figure S 7:** LC-MS analysis of TCA cycle, glycolytic and pentose phosphate pathway metabolites in *P. falciparum* with and without treatment of compound 1. 2PGA = 2-phospho glyceric acid, 3PGA = 3-phospho glyceric acid, G1P = glucose-1-phosphate, F6P = fructose-6-phosphate, G6P = glucose-6-phosphate, R5P = ribose-5-phosphate,

Ru5P = ribulose-5-phosphate, PEP = phosphoenolpyruvate, GC3P = glycerol-3-phosphate, S7P = sedoheptulose-7-phosphate, ACO = aconitate, LAC = lactate, SUC = succinate, FUM = fumarate, MAL = malate, CIT = citrate, PYR = pyruvate, UNT = untreated.

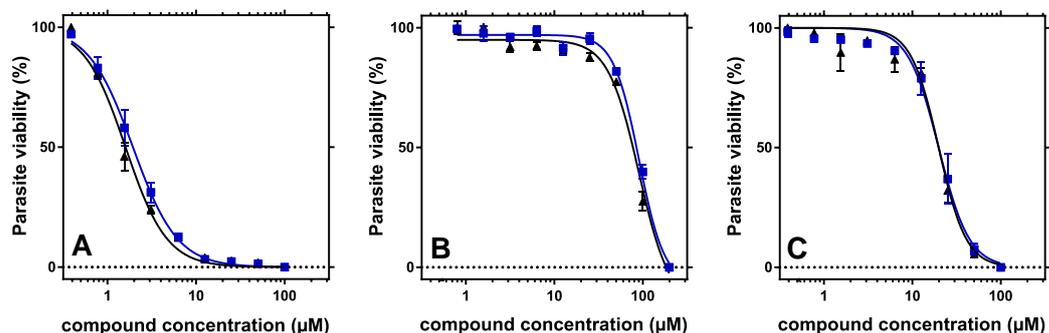
**Table S 3:** T-test analysis of LC-MS data for MEP pathway and TCA cycle metabolites. UNT = untreated, DOXP = 1-deoxy-D-xylulose-5-phosphate, MEP = methylerythritol phosphate, CDP-ME = 4-diphosphocytidyl-2-C-methyl-D-erythritol, MEcDP = methylerythritol cyclodiphosphate, DMADP = dimethylallyl diphosphate, HDMADP = 1-hydroxy-2-methyl-2-buten-4-yl 4-diphosphate, 2PGA = 2-phospho glyceric acid, 3PGA = 3-phospho glyceric acid, ACO = aconitate, CIT = citrate, FUM = fumarate, G1P = glucose-1-phosphate, F6P = fructose-6-phosphate, G6P = glucose-6-phosphate, GC3P = glycerol-3-phosphate, LAC = lactate, MAL = malate, PEP = phosphoenolpyruvate, PYR = pyruvate, R5P = ribose-5-phosphate, Ru5P = ribulose-5-phosphate, S7P = Sedoheptulose-7-phosphate, SUC = succinate.

Cmpd	Discovery? P value	Mean of UNT	Mean of + 1	Diff.	SE of diff.
DOXP	No >0,99999	1,163	1,163	0,000	0,2689
MEP	No 0,070645	0,4233	0,3100	0,1133	0,04631
CDP-ME	No 0,587914	0,06667	0,07667	-0,010	0,01700
MEcDP	No 0,720090	11,24	10,20	1,043	2,713
DMADP	No 0,032540	0,1667	0,1033	0,0633	0,01972
HDMADP	No 0,306022	0,2200	0,1833	0,0367	0,03127
2PGA	No 0,050194	61,07	31,49	29,58	10,67
3PGA	No 0,037010	14,78	8,127	6,657	2,163
ACO	No 0,144422	1,877	1,140	0,7367	0,4068
CIT	No 0,008553	64,80	25,15	39,65	8,234
FUM	No 0,003295	1668	727,7	940,6	149,9
G1P/F6P	No 0,112890	103,0	57,13	45,85	22,64
G6P	No 0,044056	17,23	6,793	10,43	3,596
GC3P	No 0,003896	13,48	6,757	6,720	1,121
LAC	No 0,388380	104,2	80,23	23,97	24,80
MAL	No 0,001940	53,67	23,30	30,37	4,200
PEP	No 0,005029	11,98	5,297	6,687	1,196
PYR	No 0,030535	102,5	28,19	74,26	22,65
RU5P/RUP	No 0,038874	5,757	2,520	3,237	1,069
S7P	No 0,340281	90,64	54,93	35,71	33,01
SUC	No 0,002161	118,0	60,23	57,74	8,218

## Results of $IC_{50}$ determination against *PfDXPS* and *PfTPK* overexpressing *P. falciparum* 3D7



**Figure S 8: Target verification using *PfDXPS* overexpressing parasites (3D7-DXPS+).** Antiplasmodial activity of compounds (**A**: compound 1, **B**: compound 2, **C**: compound 3) against 3D7-DXPS<sup>+</sup> overexpressing parasites (green squares) in comparison to the MOCK cell line (black triangles). All data were averaged from three independent experiments conducted in triplicate and is shown including SEM (error bars). For  $IC_{50}$  determination, data was analyzed using nonlinear regression of the log-concentration-response curves and interpolated from the sigmoidal curve.



**Figure S 9: Target verification using *PfTPK* overexpressing parasites (3D7-DXPS+).** Antiplasmodial activity of compounds (**A**: compound 1, **B**: compound 2, **C**: compound 3) against *PfTPK* overexpressing parasites (blue squares) in comparison to the MOCK cell line (black triangles). Data represent the result from one experiment with technical triplicates (error bars indicate SEM). For  $IC_{50}$  determination, data was analyzed using nonlinear regression of the log-concentration-response curves and interpolated from the sigmoidal curve.

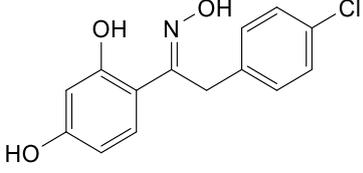
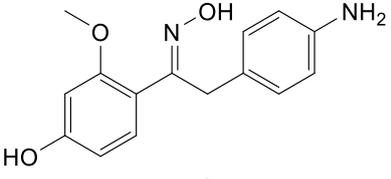
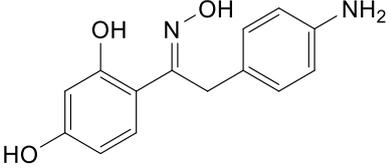
**Table S 4: Compound evaluation against *P. falciparum* 3D7-TPK<sup>+</sup> and *P. falciparum* 3D7-DXPS<sup>+</sup> in comparison to MOCK cell line.** The table displays the IC<sub>50</sub> values from antiplasmodial screens using *P. falciparum* 3D7-PfTPK<sup>+</sup> (3D7-TPK<sup>+</sup>) and *P. falciparum* 3D7-PfDXPS (3D7-DXPS<sup>+</sup>) overexpression cell lines, as interpolated from nonlinear regression curves. 95% CI is displayed as error measure. Percentages indicate statistical probability of the simpler model “LogIC<sub>50</sub> same for all data sets” being correct.

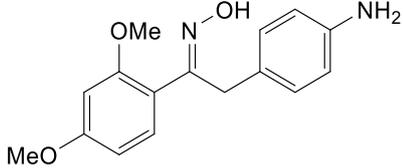
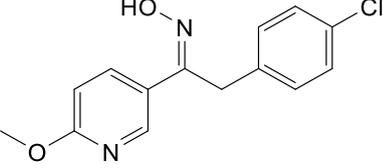
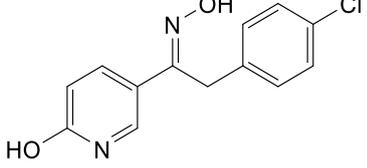
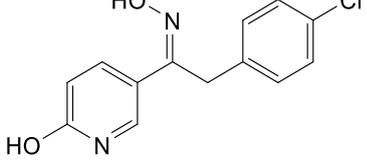
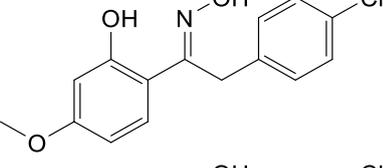
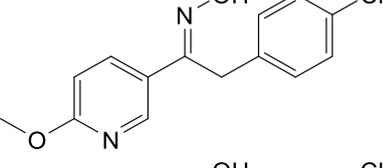
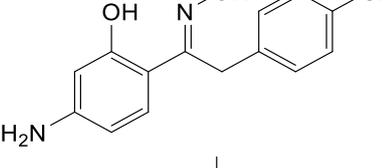
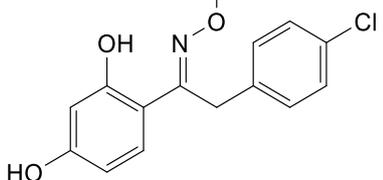
Cmpd	PfDXPS IC <sub>50</sub> (μM)	MOCK IC <sub>50</sub> (μM)	Difference (yes/no)	PfTPK <sup>[a]</sup> IC <sub>50</sub> (μM)	MOCK <sup>[a]</sup> IC <sub>50</sub> (μM)	Difference (yes/no)
2	109.9 (89.7–167.6)	109.9 (89.7–167.6)	78.49%	90.4 (78.6–116.9)	90.4 (78.6–116.9)	No 91.69%
3	17.9 (17.1–18.8)	17.9 (17.1–18.8)	77.18%	21.1 (19.4–23.1)	23.8	No 83.76%
1	2.08 (1.66–2.48)	1.61 (1.45–1.76)	Yes 15.25%	1.76 (1.36–2.1)	1.18 (0.87–1.41)	Yes 7.50%

<sup>[a]</sup>Single measurement.

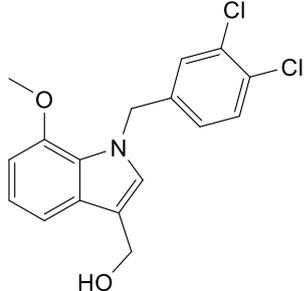
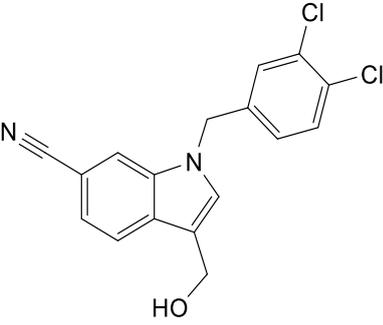
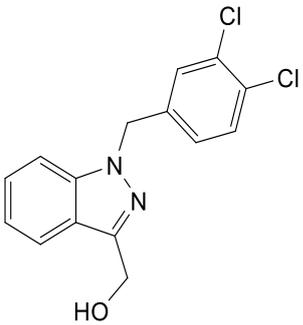
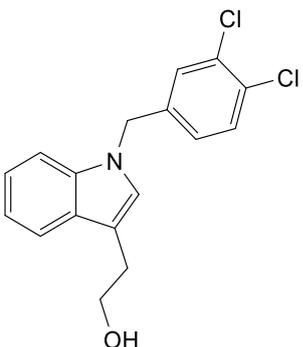
## Inhibition data for all compounds

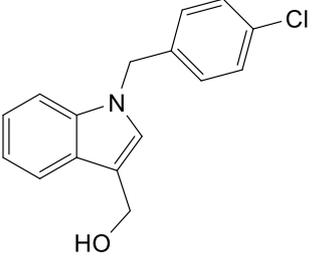
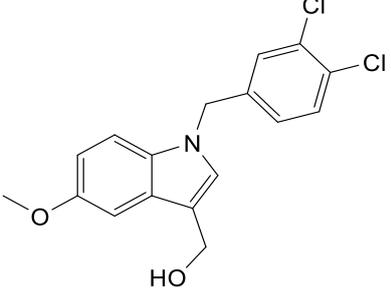
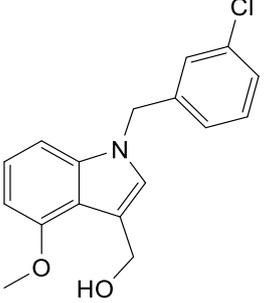
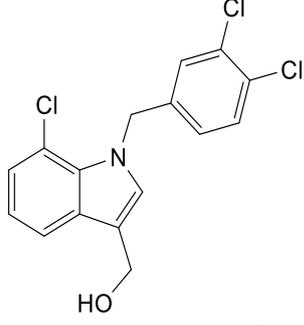
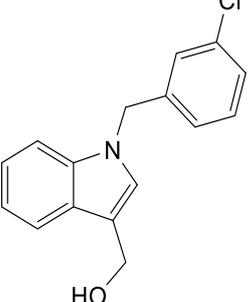
**Table S 5: Summary of determined IC<sub>50</sub> values of all compounds.** The numbers I-III are indicating the used method.

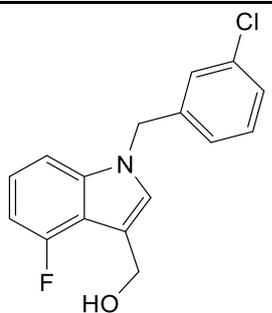
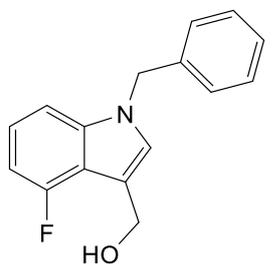
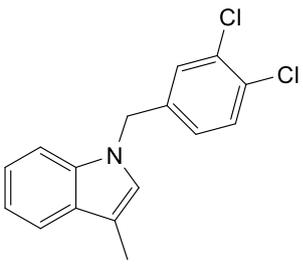
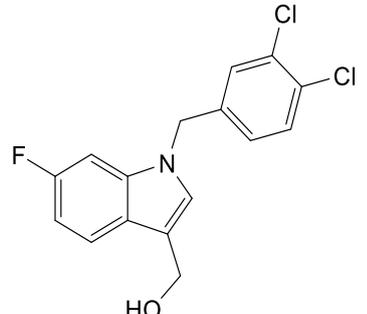
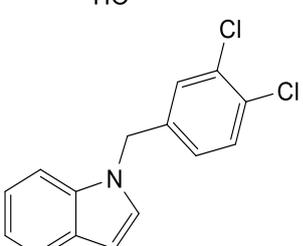
ID	Structure	IC <sub>50</sub> 3D7-III [μM]	IC <sub>50</sub> Dd2-III [μM]	IC <sub>50</sub> 3D7-II [μM]	IC <sub>50</sub> NF54-II [μM]	IC <sub>50</sub> 3D7-I [μM]
2		93.8 ± 8.0	89.4 ± 30.6	42.4 ± 4.4	41.4 ± 9.7	81.2 ± 3.3
4		>111	93.6 ± 16.1			
5		>111	>111			

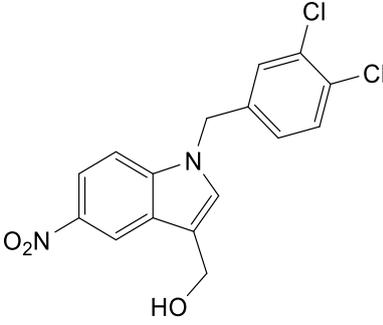
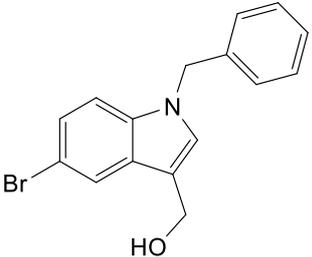
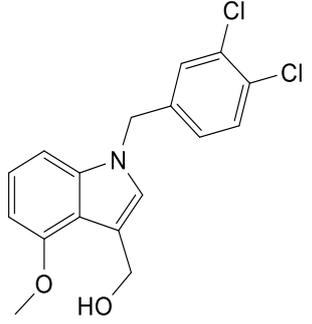
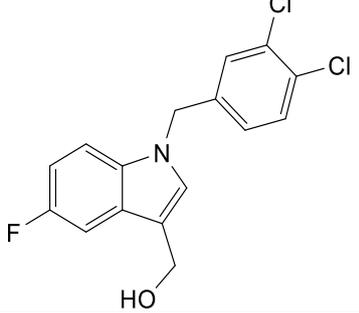
ID	Structure	IC <sub>50</sub> 3D7- III [μM]	IC <sub>50</sub> Dd2- III [μM]	IC <sub>50</sub> 3D7-II [μM]	IC <sub>50</sub> NF54 -II [μM]	IC <sub>50</sub> 3D7-I [μM]
6		>111	>111			
7Z		>111	>111			
8E		>111	>111			
8Z		>28	>28			
9		46.8 ± 15.0	24.1 ± 5.5			
7E		43.6 ± 9.5	>55			
10		38.2 ± 11.2	24.3 ± 3.6			
11		28.0 ± 8.8	19.1 ± 3.4			

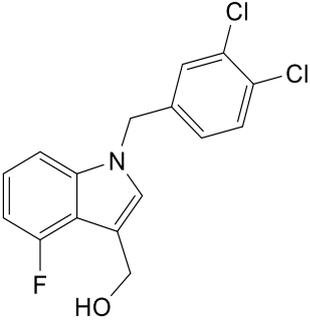
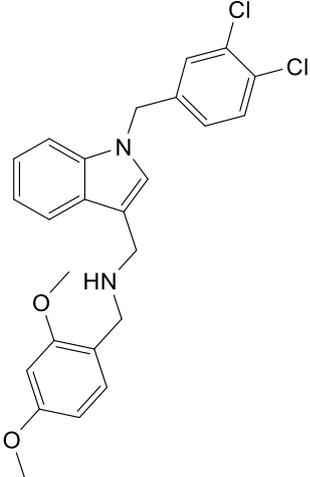
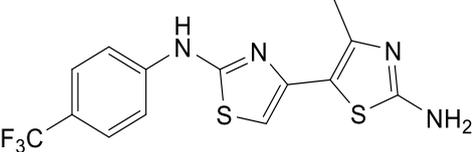
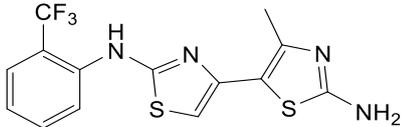
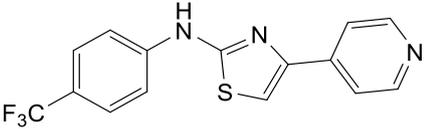
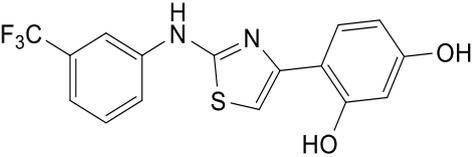
ID	Structure	IC <sub>50</sub> 3D7-III [μM]	IC <sub>50</sub> Dd2-III [μM]	IC <sub>50</sub> 3D7-II [μM]	IC <sub>50</sub> NF54-II [μM]	IC <sub>50</sub> 3D7-I [μM]
12		16.0 ± 0.2	14.4 ± 5.0	45.7 ± 3.2	45.0 ± 7.0	
13		9.9 ± 1.7	27.9 ± 12.5			
3		23.6 ± 6.4	12.8 ± 9.2	16.5 ± 2.3	14.8 ± 2.2	24.8 ± 2.9
14		> 111	91.1 ± 18.1			
15		> 55.5	54.3 ± 5.8			

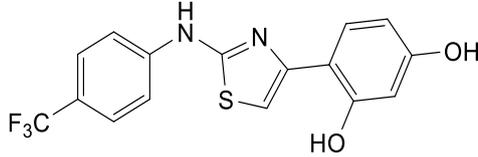
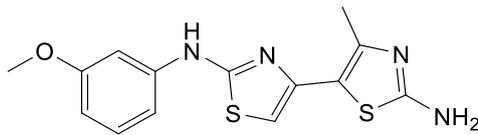
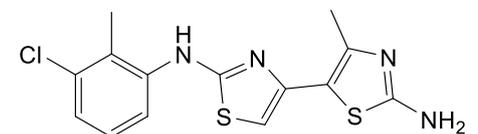
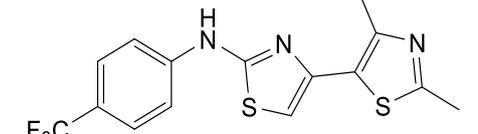
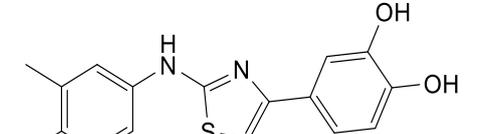
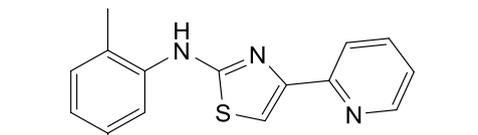
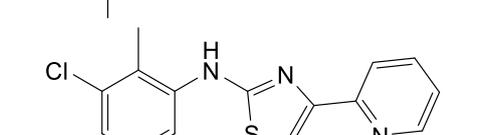
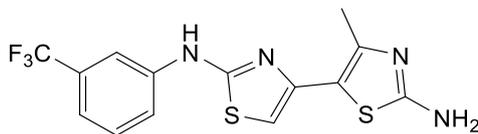
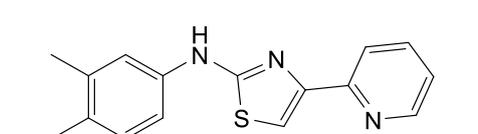
ID	Structure	IC <sub>50</sub> 3D7- III [μM]	IC <sub>50</sub> Dd2- III [μM]	IC <sub>50</sub> 3D7-II [μM]	IC <sub>50</sub> NF54 -II [μM]	IC <sub>50</sub> 3D7-I [μM]
16		> 55.5	47.5 ± 12.5			
17		43.8 ± 19.2	44.4 ± 8.7			
18		33.9 ± 0.2	69.2 ± 21.1			
19		28.9 ± 1.2	12.2 ± 2.9			

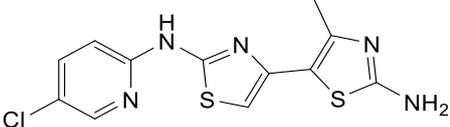
ID	Structure	IC <sub>50</sub> 3D7- III [μM]	IC <sub>50</sub> Dd2- III [μM]	IC <sub>50</sub> 3D7-II [μM]	IC <sub>50</sub> NF54 -II [μM]	IC <sub>50</sub> 3D7-I [μM]
20		27.7 ± 13.6	21.5 ± 4.8			
21		18.5 ± 1.1	53.7 ± 6.3			
22		15.6 ± 3.3	13.5 ± 3.5			
23		13.8 ± 0.9	35.5 ± 16.1			
24		12.2 ± 2.8	8.4 ± 1.2			

ID	Structure	IC <sub>50</sub> 3D7- III [μM]	IC <sub>50</sub> Dd2- III [μM]	IC <sub>50</sub> 3D7-II [μM]	IC <sub>50</sub> NF54 -II [μM]	IC <sub>50</sub> 3D7-I [μM]
25		12.0 ± 1.0	9.9 ± 0.9			
26		11.9 ± 1.5	9.2 ± 0.4			
27		10.2 ± 3.4	11.1 ± 5.1			
28		8.9 ± 2.4	10.4 ± 4.7			
29		8.1 ± 0.1	17.8 ± 6.8			

ID	Structure	IC <sub>50</sub> 3D7- III [μM]	IC <sub>50</sub> Dd2- III [μM]	IC <sub>50</sub> 3D7-II [μM]	IC <sub>50</sub> NF54 -II [μM]	IC <sub>50</sub> 3D7-I [μM]
30		7.2 ± 0.1	4.2 ± 0.7			
31		6.1 ± 1.3	5.3 ± 3.3			
32		5.5 ± 1.5	4.9 ± 2.4			
33		5.3 ± 1.3	7.5 ± 2.2			

ID	Structure	IC <sub>50</sub> 3D7- III [μM]	IC <sub>50</sub> Dd2- III [μM]	IC <sub>50</sub> 3D7-II [μM]	IC <sub>50</sub> NF54 -II [μM]	IC <sub>50</sub> 3D7-I [μM]
34		2.4 ± 0.2	2.8 ± 0.1	2.7 ± 0.3	2.4 ± 0.6	
35		46.8 ± 15.0	21.1 ± 5.5			
1		0.6 ± 0.2	0.8 ± 0.4	1.7 ± 0.6	0.8 ± 0.1	1.57 ± 0.6
36		43.0 ± 2.4	47.5 ± 12.3			
37		34.2 ± 1.8	24.5 ± 2.1			
38		23.0 ± 0.7	19.4 ± 2.3			

ID	Structure	IC <sub>50</sub> 3D7- III [μM]	IC <sub>50</sub> Dd2- III [μM]	IC <sub>50</sub> 3D7-II [μM]	IC <sub>50</sub> NF54 -II [μM]	IC <sub>50</sub> 3D7-I [μM]
39		22.1 ± 5.6	24.8 ± 5.0			
40		17.9 ± 0.4	43.0 ± 23.6			
41		15.9 ± 1.8	17.1 ± 10.1			
42		9.5 ± 1.1	5.4 ± 0.3			
43		4.4 ± 1.0	6.8 ± 0.5			
44		3.7 ± 1.6	5.0 ± 1.3			
45		2.8 ± 0.1	1.7 ± 0.7			
46		2.5 ± 0.1	4.6 ± 0.3			
47		1.5 ± 0.8	1.7 ± 0.4	2.3 ± 0.4	2.1 ± 0.8	

ID	Structure	IC <sub>50</sub> 3D7- III [μM]	IC <sub>50</sub> Dd2- III [μM]	IC <sub>50</sub> 3D7-II [μM]	IC <sub>50</sub> NF54 -II [μM]	IC <sub>50</sub> 3D7-I [μM]
48		1.0 ± 0.1	1.6 ± 0.2	2.0 ± 0.7	1.6 ± 0.3	

## Docking and scoring

The computer program SeeSAR, version 10.2 from BioSolveIT was used to dock compounds and calculate binding affinities. The program uses the FlexX docking function for the placement of ligands.<sup>1</sup> The affinities are estimated using the HYDE scoring function, which calculates the binding affinities based on the hydration differences between the bound and unbound state of the molecule.<sup>3; 2</sup> The binding site was chosen based on co-crystal structures with a ligand or substrate. In the case of more than one binding site in a protein structure, the binding site with the highest calculated affinity to the ligand was selected. Analysis and visualisation of the results was done using the program StarDrop, version 6.6.7.25378 from Optibrium.

**Table S 6:** Results of the docking of the indole-class and benzimidazole CHEMBL\_1945694 to the FabI protein crystal structure from *F. tularensis* with the PDB-code **3uic**.

ID	SMILES	HYDE_LOWER [nM]	HYDE_UPPER [nM]
16	<chem>COc1cccc2c1n(cc2CO)Cc3ccc(c(c3)Cl)Cl</chem>	0.2834	28.15
<b>CHEMBL_1945694</b>			
23	<chem>Cc1cc2c(cc1C)n(cn2)Cc3ccc(c(c3)Cl)Cl</chem>	0.3163	31.43
22	<chem>c1cc2c(cn(c2c(c1)Cl)Cc3ccc(c(c3)Cl)Cl)CO</chem>	0.3265	32.44
34	<chem>COc1ccc2c(c1)c(cn2Cc3ccc(c(c3)Cl)Cl)CO</chem>	0.5139	51.05
33	<chem>c1cc2c(c(c1)F)c(cn2Cc3ccc(c(c3)Cl)Cl)CO</chem>	0.6834	67.9
19	<chem>c1cc(c(cc1Cn2cc(c3c2ccc(c3)F)CO)Cl)Cl</chem>	0.7494	74.45
27	<chem>c1ccc2c(c1)c(cn2Cc3ccc(c(c3)Cl)Cl)CCO</chem>	0.8641	85.85
3	<chem>Cc1cn(c2c1cccc2)Cc3ccc(c(c3)Cl)Cl</chem>	1.011	100.5
24	<chem>c1ccc2c(c1)c(cn2Cc3ccc(c(c3)Cl)Cl)CO</chem>	1.022	101.6
32	<chem>c1ccc2c(c1)c(cn2Cc3cccc(c3)Cl)CO</chem>	1.397	138.8
25	<chem>COc1cccc2c1c(cn2Cc3ccc(c(c3)Cl)Cl)CO</chem>	1.732	172.1
29	<chem>c1cc(cc1)Cn2cc(c3c2ccc(c3)F)CO</chem>	2.157	214.3
14	<chem>c1ccc2c(c1)ccn2Cc3ccc(c(c3)Cl)Cl</chem>	3.489	346.7
26	<chem>c1cc2c(cn(c2nc1)Cc3ccc(c(c3)Cl)Cl)CO</chem>	4.207	418
17	<chem>c1ccc(cc1)Cn2cc(c3c2ccc(c3)F)CO</chem>	6.774	673
35	<chem>c1cc(c(cc1Cn2cc(c3c2ccc(cc3)C#N)CO)Cl)Cl</chem>	7.01	696.5
20	<chem>COc1ccc(c(c1)OC)CNCc2cn(c3c2cccc3)Cc4ccc(c(c4)Cl)Cl</chem>	7.292	724.5
15	<chem>c1ccc2c(c1)c(cn2Cc3ccc(cc3)Cl)CO</chem>	9.881	981.7
28	<chem>c1ccc2c(c1)c(cn2c3ccc(c(c3)Cl)Cl)CO</chem>	10.88	1081
31	<chem>c1cc(c(cc1Cn2cc(c3c2ccc(cc3)F)CO)Cl)Cl</chem>	12.14	1206
18	<chem>c1ccc(cc1)Cn2cc(c3c2ccc(c3)Br)CO</chem>	13.57	1348
30	<chem>c1ccc2c(c1)c(nn2Cc3ccc(c(c3)Cl)Cl)CO</chem>	29.95	2975
	<chem>c1cc(c(cc1Cn2cc(c3c2ccc(c3)N(=O)=O)CO)Cl)Cl</chem>	30.33	3014

**Table S 7:** Results of the docking of the indole-class and benzimidazole CHEMBL\_1945694 to the FabI protein crystal structure from *P. falciparum* with the PDB-code **2op1**.

ID	SMILES	HYDE_LOWER [nM]	HYDE_UPPER [nM]
<b>CHEMBL1945694</b>			
<b>694</b>	<chem>Cc1cc2c(cc1C)n(cn2)Cc3ccc(c(c3)Cl)Cl</chem>	0.3163	31.43
<b>23</b>	<chem>c1cc2c(cn(c2c(c1)Cl)Cc3ccc(c(c3)Cl)Cl)CO</chem>	0.3623	36
<b>22</b>	<chem>COc1ccc2c(c1)c(cn2Cc3ccc(c(c3)Cl)Cl)CO</chem>	0.5139	51.05
<b>33</b>	<chem>c1cc(c(cc1Cn2cc(c3c2ccc(c3)F)CO)Cl)Cl</chem>	0.7494	74.45
<b>34</b>	<chem>c1cc2c(c(c1)F)c(cn2Cc3ccc(c(c3)Cl)Cl)CO</chem>	0.805	79.98
<b>16</b>	<chem>COc1cccc2c1n(cc2CO)Cc3ccc(c(c3)Cl)Cl</chem>	0.8209	81.56
<b>19</b>	<chem>c1ccc2c(c1)c(cn2Cc3ccc(c(c3)Cl)Cl)CCO</chem>	0.8641	85.85
<b>3</b>	<chem>c1ccc2c(c1)c(cn2Cc3ccc(c(c3)Cl)Cl)CO</chem>	1.022	101.6
<b>24</b>	<chem>c1ccc2c(c1)c(cn2Cc3cccc(c3)Cl)CO</chem>	1.397	138.8
<b>32</b>	<chem>COc1cccc2c1c(cn2Cc3ccc(c(c3)Cl)Cl)CO</chem>	1.732	172.1
<b>25</b>	<chem>c1cc(cc(c1)Cl)Cn2cc(c3c2ccc(c3)F)CO</chem>	2.277	226.3
<b>27</b>	<chem>Cc1cn(c2c1cccc2)Cc3ccc(c(c3)Cl)Cl</chem>	2.718	270.1
<b>29</b>	<chem>c1ccc2c(c1)ccn2Cc3ccc(c(c3)Cl)Cl</chem>	3.489	346.7
<b>14</b>	<chem>c1cc2c(cn(c2nc1)Cc3ccc(c(c3)Cl)Cl)CO</chem>	4.207	418
<b>17</b>	<chem>c1cc(c(cc1Cn2cc(c3c2cc(cc3)C#N)CO)Cl)Cl</chem>	7.01	696.5
<b>20</b>	<chem>c1ccc2c(c1)c(cn2Cc3ccc(cc3)Cl)CO</chem>	10.1	1003
<b>26</b>	<chem>c1ccc(cc1)Cn2cc(c3c2ccc(c3)F)CO</chem>	11.96	1188
<b>28</b>	<chem>c1cc(c(cc1Cn2cc(c3c2cc(cc3)F)CO)Cl)Cl</chem>	12.14	1206
<b>31</b>	<chem>c1ccc(cc1)Cn2cc(c3c2ccc(c3)Br)CO</chem>	14.66	1456
<b>15</b>	<chem>c1ccc2c(c1)c(cn2c3ccc(c(c3)Cl)Cl)CO</chem>	17.17	1706
<b>18</b>	<chem>c1ccc2c(c1)c(cn2Cc3ccc(c(c3)Cl)Cl)CO</chem>	29.95	2975
<b>30</b>	<chem>c1cc(c(cc1Cn2cc(c3c2ccc(c3)N(=O)=O)CO)Cl)Cl</chem>	30.33	3014
<b>35</b>	<chem>COc1ccc(c(c1)OC)CNCc2cn(c3c2cccc3)Cc4ccc(c4)Cl</chem>	36.14	3591

**Table S 8:** Docking to of the oxime class to the FabI protein crystal structure from *F. tularensis* with the PDB-code **3uic**.

ID	SMILES	HYDE_LOWER [nM]	HYDE_UPPER [nM]
<b>CHEMBL_1945694_1_009</b>			
<b>5694_1_009</b>	<chem>Cc1cc2c(cc1C)n(cn2)Cc3ccc(c(c3)Cl)Cl</chem>	0.3193	31.73
<b>11</b>	<chem>CON=C(Cc1ccc(cc1)Cl)c2ccc(cc2O)O</chem>	157.9	1.569e+04
<b>13</b>	<chem>c1cc(ccc1CC(c2ccc(cc2O)O)O)Cl</chem>	204.5	2.031e+04
<b>4</b>	<chem>COc1cc(ccc1C(=NO)Cc2ccc(cc2)N)O</chem>	247.3	2.457e+04
<b>10</b>	<chem>c1cc(ccc1CC(=NO)c2ccc(cc2O)N)Cl</chem>	398.2	3.957e+04
<b>5</b>	<chem>c1cc(ccc1CC(=NO)c2ccc(cc2O)O)N</chem>	777.4	7.724e+04
<b>2</b>	<chem>c1cc(ccc1CC(=NO)c2ccc(cc2O)O)Cl</chem>	1082	1.075e+05
<b>12</b>	<chem>c1cc(ccc1CC(=NN)c2ccc(cc2O)O)Cl</chem>	3304	3.282e+05
<b>9</b>	<chem>COc1ccc(c(c1)O)C(=NO)Cc2ccc(cc2)Cl</chem>	6711	6.668e+05
<b>7E</b>	<chem>COc1ccc(cn1)C(=NO)Cc2ccc(cc2)Cl</chem>	1.09e+04	1.083e+06

**Table S 9:** Docking to of the aminothiazole class to the FabI protein crystal structure from *F. tularensis* with the PDB-code **3uic**.

ID	SMILES	HYDE_LOWER [nM]	HYDE_UPPER [nM]
<b>CHEMBL_194</b>			
<b>5694_1_009</b>	<chem>Cc1cc2c(cc1C)n(cn2)Cc3ccc(c(c3)Cl)Cl</chem>	0.3193	31.73
<b>39</b>	<chem>c1cc(ccc1C(F)(F)F)Nc2nc(cs2)c3ccc(cc3O)O</chem>	26.75	2657
<b>1</b>	<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3ccc(cc3)C(F)(F)F</chem>	55.32	5496
<b>45</b>	<chem>Cc1c(cccc1Cl)Nc2nc(cs2)c3cccn3</chem>	57.78	5741
<b>46</b>	<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3cccc(c3)C(F)(F)F</chem>	62.49	6209
<b>40</b>	<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3cccc(c3)OC</chem>	66.37	6595
<b>41</b>	<chem>Cc1c(cccc1Cl)Nc2nc(cs2)c3c(nc(s3)N)C</chem>	77.89	7739
<b>36</b>	<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3cccc3C(F)(F)F</chem>	132.8	1.319e+04
<b>38</b>	<chem>c1cc(cc(c1)Nc2nc(cs2)c3ccc(cc3O)O)C(F)(F)F</chem>	350.1	3.479e+04
<b>43</b>	<chem>Cc1ccc(cc1C)Nc2nc(cs2)c3ccc(c(c3)O)O</chem>	499.9	4.967e+04
<b>37</b>	<chem>c1cc(ccc1C(F)(F)F)Nc2nc(cs2)c3ccncc3</chem>	633.6	6.295e+04
<b>47</b>	<chem>Cc1ccc(cc1C)Nc2nc(cs2)c3cccn3</chem>	1811	1.799e+05
<b>48</b>	<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3ccc(cn3)Cl</chem>	2766	2.748e+05
<b>44</b>	<chem>c1ccc(cc1)Nc2nc(cs2)c3cccn3</chem>	2979	2.959e+05
<b>42</b>	<chem>Cc1c(sc(n1)C)c2csc(n2)Nc3ccc(cc3)C(F)(F)F</chem>	2.082e+04	2.069e+06

**Table S 10:** Docking to of the oxime class to the FabI protein crystal structure from *P. falciparum* with the PDB-code **2op1**.

ID	SMILES	HYDE_LOWER [nM]	HYDE_UPPER [nM]
<b>CHEMBL_19</b>			
<b>45694</b>	<chem>Cc1cc2c(cc1C)n(cn2)Cc3ccc(c(c3)Cl)Cl</chem>	0.3193	31.73
<b>11</b>	<chem>CON=C(Cc1ccc(cc1)Cl)c2ccc(cc2O)O</chem>	157.9	1.569e+04
<b>13</b>	<chem>c1cc(ccc1CC(c2ccc(cc2O)O)O)Cl</chem>	204.5	2.031e+04
<b>4</b>	<chem>COc1cc(ccc1C(=NO)C)c2ccc(cc2)N)O</chem>	247.3	2.457e+04
<b>10</b>	<chem>c1cc(ccc1CC(=NO)c2ccc(cc2O)N)Cl</chem>	398.2	3.957e+04
<b>5</b>	<chem>c1cc(ccc1CC(=NO)c2ccc(cc2O)O)N</chem>	777.4	7.724e+04
<b>2</b>	<chem>c1cc(ccc1CC(=NO)c2ccc(cc2O)O)Cl</chem>	1082	1.075e+05
<b>12</b>	<chem>c1cc(ccc1CC(=NN)c2ccc(cc2O)O)Cl</chem>	3304	3.282e+05
<b>9</b>	<chem>COc1ccc(c(c1)O)C(=NO)C2ccc(cc2)Cl</chem>	6711	6.668e+05
<b>7E</b>	<chem>COc1ccc(cn1)C(=NO)C2ccc(cc2)Cl</chem>	1.09e+04	1.083e+06

**Table S 11:** Docking to of the aminothiazole class to the FabI protein crystal structure from *P. falciparum* with the PDB-code **2op1**.

ID	SMILES	HYDE_LOWER [nM]	HYDE_UPPER [nM]
<b>09T_I_26</b>			
<b>2_1_009</b>	<chem>Cc1cc2c(cc1C)n(cn2)Cc3ccc(c(c3)Cl)Cl</chem>	0.3193	31.73
<b>39</b>	<chem>c1cc(ccc1C(F)(F)F)Nc2nc(cs2)c3ccc(cc3O)O</chem>	26.75	2657
<b>1</b>	<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3ccc(cc3)C(F)(F)F</chem>	55.32	5496
<b>45</b>	<chem>Cc1c(cccc1Cl)Nc2nc(cs2)c3cccn3</chem>	57.78	5741
<b>46</b>	<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3cccc(c3)C(F)(F)F</chem>	62.49	6209

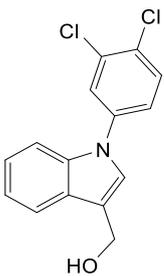
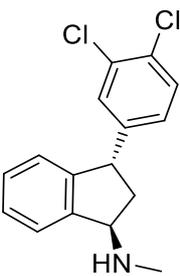
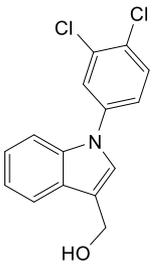
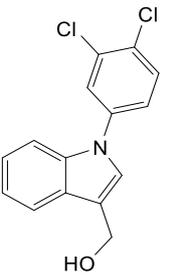
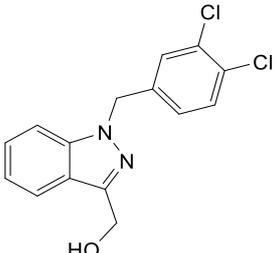
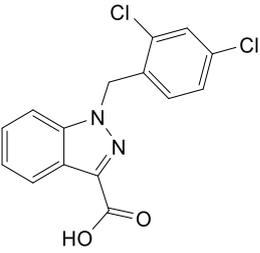
ID	SMILES	HYDE_LOWER [nM]	HYDE_UPPER [nM]
40	<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3cccc(c3)OC</chem>	66.37	6595
41	<chem>Cc1c(cccc1Cl)Nc2nc(cs2)c3c(nc(s3)N)C</chem>	77.89	7739
36	<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3cccc3C(F)(F)F</chem>	132.8	1.319e+04
38	<chem>c1cc(cc(c1)Nc2nc(cs2)c3ccc(cc3O)O)C(F)(F)F</chem>	350.1	3.479e+04
43	<chem>Cc1ccc(cc1C)Nc2nc(cs2)c3ccc(c(c3)O)O</chem>	499.9	4.967e+04
37	<chem>c1cc(ccc1C(F)(F)F)Nc2nc(cs2)c3ccncc3</chem>	633.6	6.295e+04
47	<chem>Cc1ccc(cc1C)Nc2nc(cs2)c3cccn3</chem>	1811	1.799e+05
48	<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3ccc(cn3)Cl</chem>	2766	2.748e+05
44	<chem>c1ccc(cc1)Nc2nc(cs2)c3cccn3</chem>	2979	2.959e+05
42	<chem>Cc1c(sc(n1)C)c2csc(n2)Nc3ccc(cc3)C(F)(F)F</chem>	2.082e+04	2.069e+06

**Table S 12:** Docking of the aminothiazole class, the co-crystallized inhibitor TLM and compounds CHEMBL3344230 and CHEMBL546826 to the KasA protein crystal structure from *M. tuberculosis* with the PDB-code **2wgd**.

ID	SMILES	HYDE_LOWER [nM]	HYDE_UPPER [nM]
42	<chem>Cc1c(sc(n1)C)c2csc(n2)Nc3ccc(cc3)C(F)(F)F</chem>	37.1	3686
39	<chem>c1cc(ccc1C(F)(F)F)Nc2nc(cs2)c3ccc(cc3O)O</chem>	48.11	4780
46	<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3cccc(c3)C(F)(F)F</chem>	51.6	5126
Ligand TLM	<chem>CC1=C(C(SC1=O)(C)C=C(C)C=C)O</chem>	319.1	3.171e+04
36	<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3cccc3C(F)(F)F</chem>	383.7	3.812e+04
CHEMBL3344 230	<chem>c1cc(ccc1Nc2nc(cs2)c3ccncc3)Br</chem>	498	4.948e+04
43	<chem>Cc1ccc(cc1C)Nc2nc(cs2)c3ccc(c(c3)O)O</chem>	593.5	5.897e+04
CHEMBL5468 26	<chem>c1ccnc(c1)c2csc(n2)Nc3ccc(cn3)Cl</chem>	869.3	8.637e+04
45	<chem>Cc1c(cccc1Cl)Nc2nc(cs2)c3cccn3</chem>	923.9	9.18e+04
38	<chem>c1cc(cc(c1)Nc2nc(cs2)c3ccc(cc3O)O)C(F)(F)F</chem>	1517	1.507e+05
37	<chem>c1cc(ccc1C(F)(F)F)Nc2nc(cs2)c3ccncc3</chem>	1558	1.548e+05
47	<chem>Cc1ccc(cc1C)Nc2nc(cs2)c3cccn3</chem>	5896	5.858e+05
44	<chem>c1ccc(cc1)Nc2nc(cs2)c3cccn3</chem>	7891	7.841e+05
1	<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3ccc(cc3)C(F)(F)F</chem>	8079	8.027e+05
41	<chem>Cc1c(cccc1Cl)Nc2nc(cs2)c3c(nc(s3)N)C</chem>	2.626e+04	2.609e+06
40	<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3cccc(c3)OC</chem>	2.89e+04	2.871e+06
48	<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3ccc(cn3)Cl</chem>	4.236e+05	4.208e+07

## Human off-target enzymes

**Table S 13:** Summary of possible human off-target enzymes encountered during the search for bacterial targets. The molecule drawn in “search hit” is reported in the cited source to be targeting the enzyme or pathway given in “target enzyme”. Only one hit molecule is shown, the cited source often reports many more derivatives. No ranking and in-depth analysis of likelihood to be a target was performed.

Search molecule	Search hit	Target enzyme(s)	Reference
	 Indatraline	non-selective monoamine transporter inhibitor	4; 5
	 Oncrasin-1	RNA polymerase II	6, 7
	 Lonidamine	Inhibitor of aerobic glycolysis in cancer cells, in particular hexokinase	8; 9

Search molecule	Search hit	Target enzyme(s)	Reference
		cortisol biosynthesis, CYP11B1	10
		KDR kinase	11
		Eicosanoid metabolism	12
		Several human targets reported	13
		$\alpha$ -amylase	14

Table of compounds as SMILES with biological activity and PAINS count.

Structure	ID	MW	IC <sub>50</sub> [ $\mu$ M]	IC <sub>50</sub> [ $\mu$ M] StdDev	PAINS count*
<chem>c1ccc2c(c1)c(cn2Cc3ccc(c(c3)Cl)Cl)CO</chem>	<b>3</b>	306.2	23.6	6.4	0
<chem>c1ccc2c(c1)c(cn2Cc3cccc(c3)Cl)CO</chem>	<b>24</b>	271.7	12.2	2.8	0
<chem>c1ccc2c(c1)c(cn2Cc3ccc(cc3)Cl)CO</chem>	<b>20</b>	271.7	27.2	13.6	0
<chem>c1cc2c(cn(c2c(c1)Cl)Cc3ccc(c(c3)Cl)Cl)CO</chem>	<b>23</b>	340.6	13.8	0.9	0
<chem>COc1cccc2c1n(cc2CO)Cc3ccc(c(c3)Cl)Cl</chem>	<b>16</b>	336.2	55.5	0	0
<chem>c1ccc2c(c1)ccn2Cc3ccc(c(c3)Cl)Cl</chem>	<b>29</b>	276.2	8.1	0.1	0

Structure	ID	MW	IC <sub>50</sub> [μM]	IC <sub>50</sub> [μM] StdDev	PAINS count*
<chem>COc1cccc2c1c(cn2Cc3ccc(c(c3)Cl)Cl)CO</chem>	<b>32</b>	336.2	5.5	1.5	0
<chem>Cc1cn(c2c1cccc2)Cc3ccc(c(c3)Cl)Cl</chem>	<b>27</b>	290.2	10.2	3.4	0
<chem>c1cc(c(cc1Cn2cc(c3c2cc(cc3)F)CO)Cl)Cl</chem>	<b>28</b>	324.2	8.9	2.4	0
<chem>c1ccc2c(c1)c(cn2Cc3ccc(c(c3)Cl)Cl)CCO</chem>	<b>19</b>	320.2	28.9	1.2	0
<chem>c1cc(c(cc1Cn2cc(c3c2cc(cc3)C#N)CO)Cl)Cl</chem>	<b>17</b>	331.2	43.8	19.2	0
<chem>c1cc(c(cc1Cn2cc(c3c2ccc(c3)N(=O)=O)CO)Cl)Cl</chem>	<b>30</b>	351.2	7.2	0.1	0
<chem>c1cc(c(cc1Cn2cc(c3c2ccc(c3)F)CO)Cl)Cl</chem>	<b>33</b>	324.2	5.3	1.3	0
<chem>COc1ccc2c(c1)c(cn2Cc3ccc(c(c3)Cl)Cl)CO</chem>	<b>21</b>	336.2	18.5	1.1	0
<chem>c1cc2c(cn(c2nc1)Cc3ccc(c(c3)Cl)Cl)CO</chem>	<b>14</b>	307.2	>111	0	0
<chem>c1ccc(cc1)Cn2cc(c3c2ccc(c3)Br)CO</chem>	<b>31</b>	316.2	6.1	1.3	0
<chem>c1ccc2c(c1)c(cn2Cc3ccc(c(c3)Cl)Cl)CO</chem>	<b>15</b>	292.2	55.5	0	0
<chem>COc1ccc(c(c1)OC)CNCc2cn(c3c2cccc3)Cc4ccc(c(c4)Cl)Cl</chem>	<b>35</b>	455.4	0.8	0.2	0
<chem>c1cc2c(c(c1)F)c(cn2Cc3ccc(c(c3)Cl)Cl)CO</chem>	<b>34</b>	324.2	2.4	0.2	0
<chem>c1ccc2c(c1)c(cn2Cc3ccc(c(c3)Cl)Cl)CO</chem>	<b>18</b>	307.2	33.9	0.2	0
<chem>c1cc(cc(c1)Cl)Cn2cc(c3c2ccc(c3)F)CO</chem>	<b>25</b>	289.7	12	0.1	0
<chem>COc1ccc2c(c1)c(cn2Cc3ccc(c(c3)Cl)Cl)CO</chem>	<b>22</b>	336.2	15.6	3.3	0
<chem>c1ccc(cc1)Cn2cc(c3c2ccc(c3)F)CO</chem>	<b>26</b>	255.3	11.9	1.5	0
<chem>c1cc(ccc1CC(=NO)c2ccc(cc2O)O)N</chem>	<b>5</b>	258.3	>111	0	0
<chem>COc1ccc(c(c1)OC)/C(=N/O)/Cc2ccc(cc2)N</chem>	<b>6</b>	286.3	>111	0	0
<chem>COc1cc(ccc1C(=NO)Cc2ccc(cc2)N)O</chem>	<b>4</b>	272.3	>111	0	0
<chem>c1cc(ccc1CC(c2ccc(cc2O)O)O)Cl</chem>	<b>13</b>	264.7	9.9	1.7	0
<chem>c1cc(ccc1CC(=NN)c2ccc(cc2O)O)Cl</chem>	<b>12</b>	276.7	16	0.2	0
<chem>CON=C(Cc1ccc(cc1)Cl)c2ccc(cc2O)O</chem>	<b>11</b>	291.7	28	8.8	0
<chem>c1cc(ccc1CC(=NO)c2ccc(cc2O)N)Cl</chem>	<b>10</b>	276.7	38.2	11.2	0
<chem>COc1ccc(c(c1)O)C(=NO)Cc2ccc(cc2)Cl</chem>	<b>9</b>	291.7	46.8	1.5	0
<chem>c1cc(ccc1CC(=NO)c2ccc(cc2O)O)Cl</chem>	<b>2</b>	277.7	93.8	0.8	0
<chem>c1cc(ccc1C/C(=N/O)/c2ccc(nc2)O)Cl</chem>	<b>8Z</b>	262.7	>28	0	0
<chem>c1cc(ccc1C/C(=N/O)/c2ccc(nc2)O)Cl</chem>	<b>8E</b>	262.7	>111	0	0
<chem>COc1ccc(cn1)C(=NO)Cc2ccc(cc2)Cl</chem>	<b>7Z</b>	276.7	>111	0	0
<chem>COc1ccc(cn1)C(=NO)Cc2ccc(cc2)Cl</chem>	<b>7E</b>	276.7	43.6	9.5	0
<chem>Cc1c(cccc1Cl)Nc2nc(cs2)c3ccccc3</chem>	<b>45</b>	301.8	2.8	0.1	0
<chem>Cc1c(sc(n1)C)c2csc(n2)Nc3ccc(cc3)C(F)(F)F</chem>	<b>42</b>	355.4	9.5	1.1	0
<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3ccccc3C(F)(F)F</chem>	<b>36</b>	356.4	43	2.4	0
<chem>c1cc(ccc1C(F)(F)F)Nc2nc(cs2)c3ccncc3</chem>	<b>37</b>	321.3	34.2	1.8	0
<chem>c1cc(ccc1C(F)(F)F)Nc2nc(cs2)c3ccc(cc3O)O</chem>	<b>39</b>	352.3	22.1	5.6	0
<chem>c1cc(cc(c1)Nc2nc(cs2)c3ccc(cc3O)O)C(F)(F)F</chem>	<b>38</b>	352.3	23	0.7	0
<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3cccc(c3)OC</chem>	<b>40</b>	318.4	17.9	0.4	0
<chem>Cc1c(cccc1Cl)Nc2nc(cs2)c3c(nc(s3)N)C</chem>	<b>41</b>	336.9	15.9	1.8	0
<chem>Cc1ccc(cc1C)Nc2nc(cs2)c3ccc(c(c3)O)O</chem>	<b>43</b>	312.4	4.4	0.1	0
<chem>c1ccc(cc1)Nc2nc(cs2)c3ccccc3</chem>	<b>44</b>	253.3	3.7	1.6	0
<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3cccc(c3)C(F)(F)F</chem>	<b>46</b>	356.4	2.5	0.1	0
<chem>Cc1ccc(cc1C)Nc2nc(cs2)c3ccccc3</chem>	<b>47</b>	281.4	1.5	0.8	0
<chem>Cc1c(sc(n1)N)c2csc(n2)Nc3ccc(cn3)Cl</chem>	<b>48</b>	323.8	1	0.1	0

Structure	ID	MW	IC <sub>50</sub> [μM]	IC <sub>50</sub> [μM] StdDev	PAINS count*
Cc1c(sc(n1)N)c2csc(n2)Nc3ccc(cc3)C(F)(F)F	1	356.4	0.6	0.2	0

\*Compounds were checked for PAINS motives using the PAINS filter of the software StarDrop, which searches for functional groups defined in the publication by J.Baell and G. Holloway.<sup>15</sup>

## References

- (1) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
- (2) Schneider, N.; Lange, G.; Hindle, S.; Klein, R.; Rarey, M. A consistent description of HYdrogen bond and DEhydration energies in protein-ligand complexes: methods behind the HYDE scoring function. *J. Comput. Aided Mol. Des.* **2013**, *27*, 15–29.
- (3) Reulecke, I.; Lange, G.; Albrecht, J.; Klein, R.; Rarey, M. Towards an integrated description of hydrogen bonding and dehydration: decreasing false positives in virtual screening with the HYDE scoring function. *ChemMedChem* **2008**, *3*, 885–897.
- (4) Bøgesø, K. P.; Christensen, A. V.; Hyttel, J.; Liljefors, T. 3-Phenyl-1-indanamines. Potential antidepressant activity and potent inhibition of dopamine, norepinephrine, and serotonin uptake. *J. Med. Chem.* **1985**, *28*, 1817–1828.
- (5) EMBL-EBI. ChEMBL296602, accessed 19.02.2021. [https://www.ebi.ac.uk/chembl/compound\\_report\\_card/ChEMBL296602/](https://www.ebi.ac.uk/chembl/compound_report_card/ChEMBL296602/). Published Online: accessed February 19, 2021.
- (6) Mehboob, S.; Hevener, K. E.; Truong, K.; Boci, T.; Santarsiero, B. D.; Johnson, M. E. Structural and enzymatic analyses reveal the binding mode of a novel series of *Francisella tularensis* enoyl reductase (FabI) inhibitors. *J. Med. Chem.* **2012**, *55*, 5933–5941.
- (7) Wu, S.; Wang, L.; Guo, W.; Liu, X.; Liu, J.; Wei, X.; Fang, B. Analogues and derivatives of oncrasin-1, a novel inhibitor of the C-terminal domain of RNA polymerase II and their antitumor activities. *J. Med. Chem.* **2011**, *54*, 2668–2679.
- (8) EMBL-EBI. ChEMBL1257030, accessed 19.02.2021. [https://www.ebi.ac.uk/chembl/compound\\_report\\_card/ChEMBL1257030/](https://www.ebi.ac.uk/chembl/compound_report_card/ChEMBL1257030/). Published Online: accessed February 19, 2021.
- (9) Floridi, A.; Paggi, M. G.; Marcante, M. L.; Silvestrini, B.; Caputo, A.; Martino, C. de. Lonidamine, a selective inhibitor of aerobic glycolysis of murine tumor cells. *J. Natl. Cancer Inst.* **1981**.
- (10) Hille, U. E.; Zimmer, C.; Vock, C. A.; Hartmann, R. W. First selective CYP11B1 inhibitors for the treatment of cortisol-dependent diseases. *ACS Med. Chem. Lett.* **2011**, *2*, 2–6.
- (11) Bilodeau, M. T.; Rodman, L. D.; McGaughey, G. B.; Coll, K. E.; Koester, T. J.; Hoffman, W. F.; Hungate, R. W.; Kendall, R. L.; McFall, R. C.; Rickert, K. W.; Rutledge, R. Z.; Thomas, K. A. The discovery of N-(1,3-thiazol-2-yl)pyridin-2-amines as potent inhibitors of KDR kinase. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2941–2945.

(12) Rödl, C. B.; Vogt, D.; Kretschmer, S. B. M.; Ihlefeld, K.; Barzen, S.; Brüggerhoff, A.; Achenbach, J.; Proschak, E.; Steinhilber, D.; Stark, H.; Hofmann, B. Multi-dimensional target profiling of N,4-diaryl-1,3-thiazole-2-amines as potent inhibitors of eicosanoid metabolism. *Eur. J. Med. Chem.* **2014**, *84*, 302–311.

(13) EMBL-EBI. CHEMBL1384007, accessed 19.02.2021. [https://www.ebi.ac.uk/chembl/compound\\_report\\_card/CHEMBL1384007/](https://www.ebi.ac.uk/chembl/compound_report_card/CHEMBL1384007/). Published Online: accessed February 19, 2021.

(14) Al-Asri, J.; Fazekas, E.; Lehoczki, G.; Perdih, A.; Görick, C.; Melzig, M. F.; Gyémánt, G.; Wolber, G.; Mortier, J. From carbohydrates to drug-like fragments: Rational development of novel  $\alpha$ -amylase inhibitors. *Bioorg. Med. Chem.* **2015**, *23*, 6725–6732.

(15) Baell, J. B.; Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **2010**, *53*, 2719–2740.



## 3.4 Supplementary Information for Chapter 2.4

### Hit-optimization using target-directed dynamic combinatorial chemistry: Development of inhibitors of the anti-infective target 1-deoxy-D-xylulose-5-phosphate synthase

Ravindra P. Jumde,<sup>a</sup> Melissa Guardigni,<sup>a,d</sup> Robin M. Gierse,<sup>a,b,c</sup> Alaa Alhayek,<sup>a,b</sup> Di Zhu,<sup>a,b,c</sup> Zhoor Hamid,<sup>a,b</sup> Sandra Johannsen,<sup>a,b</sup> Walid A. M. Elgaher,<sup>a</sup> Philipp J. Neusens,<sup>a,b</sup> Christian Nehls,<sup>e</sup> Jörg Hauptenthal,<sup>a</sup> Norbert Reiling,<sup>f,g</sup> Anna K. H. Hirsch<sup>a,b\*</sup>

<sup>a</sup> Department of Drug Design and Optimization, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS) – Helmholtz Centre for Infection Research (HZI), Campus Building E8.1, 66123 Saarbrücken, Germany

<sup>b</sup> Department of Pharmacy, Saarland University, Campus Building E8.1, 66123 Saarbrücken, Germany

<sup>c</sup> Stratingh Institute for Chemistry, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

<sup>d</sup> D3-PharmaChemistry, Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genoa, Italy

<sup>e</sup> RG Biophysics, Research Center Borstel, Leibniz Lung Center, Borstel, Germany.

<sup>f</sup> RG Microbial Interface Biology, Research Center Borstel, Leibniz Lung Center, Borstel, Germany.

<sup>g</sup> German Center for Infection Research (DZIF), Partner site Hamburg-Lübeck-Borstel-Riems, Borstel, Germany.

\*Corresponding author: Anna.Hirsch@helmholtz-hips.de

**TABLE OF CONTENTS**

Supplementary methods .....	3
General information .....	3
Chemicals .....	3
General procedure for DCC experiments .....	4
General procedure for molecular docking studies (GP-7): .....	5
General procedure for acylhydrazone formation (GP-8): .....	5
General procedure for amide bioisostere formation (GP-9): .....	6
General procedure for oxadiazole bioisostere formation (GP-10): .....	6
DCC-experiments .....	7
DCC-1: .....	7
DCC-2: .....	8
DCC-3: .....	8
DCC-4 .....	9
DCC-5 (Competition experiment with ThDP): .....	9
DCC-6 (Competition experiment with ThDP and BAP): .....	10
Characterization of acylhydrazones .....	10
Identification of different conformers of acylhydrazone .....	10
Determination of IC <sub>50</sub> values .....	19
Determination of antibacterial activity .....	19
DLS measurements .....	20
<i>Mycobacterium tuberculosis</i> growth analysis in liquid culture .....	20
Binding affinity ( $K_D$ ) determinations by surface plasmon resonance (SPR) .....	20
Mode-of-inhibition (MOI) study .....	21
Protein stability study by thermal shift assay .....	22
Supplementary Figures .....	23
Supplementary references .....	128

## SUPPLEMENTARY METHODS

### General information

All reactions using oxygen- and/or moisture-sensitive materials were carried out in dry solvents (*vide infra*) under a nitrogen atmosphere using oven-dried glassware. Reactions were monitored by a Liquid chromatography-mass spectrometry (LC-MS) system equipped with Dionex UltiMate 3000 pump, autosampler, column compartment, detector, and ESI quadrupole MS (MSQ Plus or ISQ EC) from Thermo Fisher Scientific, Dreieich, Germany. Purification of the final products, when necessary, was performed using preparative HPLC (Dionex UltiMate 3000 UHPLC+ focused, Thermo Scientific) on a reversed-phase column (C18 column, 5  $\mu$ m, Macherey-Nagel, Germany). The solvents used for the chromatography were water (0.1% formic acid) and MeCN (0.1% formic acid). High-resolution mass (HRMS) of final products was determined by LCMS/MS using a Thermo Scientific Q Exactive Focus Orbitrap LC-MS/MS system. NMR data were collected on a Bruker Avance Neo 500 MHz ( $^1\text{H}$  at 500.0 MHz;  $^{13}\text{C}$  at 126.0 MHz;  $^{19}\text{F}$  NMR at 470 MHz), equipped with a Prodigy Cryo-probe. Chemical shifts are reported in parts per million (ppm) relative to residual solvent peak (DMSO- $d_6$ ,  $^1\text{H}$ : 2.54 ppm;  $^{13}\text{C}$ : 39.9 ppm). Coupling constants are reported in Hertz (Hz). Multiplicity is reported with the usual abbreviations (s: singlet, br s: broad singlet, d: doublet, dd: doublet of doublets, ddd: doublet of doublet of doublets, t: triplet, dt: doublet of triplets, q: quartet, p: pentet, dp: doublet of pentets, m: multiplet).

The periodic progress and analysis of DCC were monitored by UPLC-MS (ThermoScientific Dionex Ultimate 3000 UHPLC System coupled to a ThermoScientific Q Exactive Focus with an electrospray ion source) using an Acquity Waters Column (BEH, C8 1.7  $\mu$ m, 2.1  $\times$  150 mm, Waters, Germany) at a flow rate of 0.250 mL/min with detection set at 210, 254, 290, and 310 nm, and the mass spectrum recorded in a positive mode in the range of 100–700 m/z. The solvent system was 0.1% formic acid in H<sub>2</sub>O (Solvent-A) and 0.1% formic acid in MeCN (Solvent-B). The gradient program began with 5% of Solvent-B for 1 min and was then increased to 95% of Solvent-B over 17 min and held for 2 min, followed by a decrease of Solvent-B to 5% over 0.1 min, where it was held for 2 min.

### Chemicals

Unless indicated otherwise, reagents and substrates were purchased from commercial sources and used as received. Solvents not required to be dry were purchased as technical grade and used as received. Dry solvents were purchased from commercial sources in Sure/Seal™ bottles and used as received and stored under a dry inert gas (N<sub>2</sub> or Ar). Inert atmosphere experiments were performed with standard Schlenk techniques with dried (P<sub>2</sub>O<sub>5</sub>) nitrogen gas. Unless indicated otherwise, the acylhydrazone products were purified by precipitation of the product at lower temperature followed by filtration and successive washing with cold solvents to remove soluble impurities (for details, see purification method-1). All reported compounds were characterized by  $^1\text{H}$  and  $^{13}\text{C}$  NMR and compared with literature data. All new compounds were fully characterized by  $^1\text{H}$  and  $^{13}\text{C}$  NMR and HRMS techniques. The purity of the final products was determined by LC-MS and found to be >95%.

## General procedure for DCC experiments

**DCL preparation (GP-1):** To a 1.5 mL Eppendorf Tube® containing phosphate buffer (pH 6.25 or 7.04), were added hydrazides (300–2000  $\mu\text{M}$  each, in DMSO), aldehydes (100  $\mu\text{M}$  each, in DMSO), aniline (10 mM, in DMSO), and an additional amount of DMSO to reach a final concentration of 5% in the DCL with 1 mL of end-volume. The DCL was allowed to gently mix on a rotating wheel (7 rpm) at room temperature and was frequently monitored via UPLC-MS. For analysis, 10  $\mu\text{L}$  of the corresponding library was mixed with 90  $\mu\text{L}$  acetonitrile and 2  $\mu\text{L}$  of NaOH (2 M), the mixture was centrifuged, and the supernatant was used for the analysis.

**Protein-templated DCL preparation (GP-2):** To a 1.5 mL Eppendorf Tube® containing phosphate buffer (pH 6.25 or 7.04), were added hydrazides (300–2000  $\mu\text{M}$  each, in DMSO), aldehydes (100  $\mu\text{M}$  each, in DMSO), aniline (10 mM, in DMSO), the protein drDXPS (20–40  $\mu\text{M}$  in phosphate buffer at pH 6.25 or 7.04), and an additional amount of DMSO to reach a final concentration of 5% in the DCL with 1 mL of end-volume. The DCL with the protein was allowed to gently mix on a rotating wheel (7 rpm) at room temperature and was frequently monitored via UPLC-MS, and the traces were compared with the blank composition. For analysis, 10  $\mu\text{L}$  of the corresponding library were mixed with 90  $\mu\text{L}$  acetonitrile and 2  $\mu\text{L}$  of NaOH (2 M), the mixture was centrifuged and the supernatant was used for the analysis.

Note: The protein-templated DCL-2 and -3 were run as duplicate.

**DCL preparation (pre-equilibrated) (GP-3):** To a 1.5 mL Eppendorf Tube® containing phosphate buffer (pH 6.25, 512.5  $\mu\text{L}$ ), were added hydrazides (300  $\mu\text{M}$  each, in DMSO), aldehydes (100  $\mu\text{M}$  each, in DMSO), aniline (10 mM, in DMSO), and an additional amount of DMSO to reach a final concentration of 5% in the DCL. The DCL was allowed to gently mix on a rotating wheel (7 rpm) at room temperature for six hours and was frequently monitored via UPLC-MS. The additional amount of phosphate buffer (pH 6.25, 437.5  $\mu\text{L}$ ) was added after six hours to reach 1 mL of end-volume. For analysis, 10  $\mu\text{L}$  of the corresponding library was mixed with 90  $\mu\text{L}$  acetonitrile and 2  $\mu\text{L}$  of NaOH (2 M), the mixture was centrifuged, and the supernatant was used for the analysis.

**Protein-templated DCL preparation (pre-equilibrated) (GP-4):** To a 1.5 mL Eppendorf Tube® containing phosphate buffer (pH 6.25, 512.5  $\mu\text{L}$ ), were added hydrazides (300  $\mu\text{M}$  each, in DMSO), aldehydes (100  $\mu\text{M}$  each, in DMSO), aniline (10 mM, in DMSO), and an additional amount of DMSO to reach a final concentration of 5% in the DCL. The DCL was allowed to gently mix on a rotating wheel (7 rpm) at room temperature for six hours and was frequently monitored via UPLC-MS. After six hours, the protein drDXPS (437.5  $\mu\text{L}$ , 20  $\mu\text{M}$  in phosphate buffer at pH 6.25) was added to the pre-equilibrated library. The DCL with the protein was allowed to gently mix on a rotating wheel (7 rpm) at room temperature and was frequently monitored for the next 6 h via UPLC-MS and the traces were compared with the blank composition. For analysis, 10  $\mu\text{L}$  of the corresponding library was mixed with 90  $\mu\text{L}$  acetonitrile and 2  $\mu\text{L}$  of NaOH (2 M), the mixture was centrifuged, and the supernatant was used for the analysis.

**Protein-templated DCL preparation for competition experiment with ThDP (GP-5):** To a 1.5 mL Eppendorf Tube® containing phosphate buffer (pH 6.25), were added hydrazides (300  $\mu\text{M}$  each, in DMSO), aldehydes (100  $\mu\text{M}$  each, in DMSO), aniline (10 mM, in DMSO), an additional amount of DMSO to reach a final concentration of 5%, and a pre-incubated mixture of drDXPS protein (40  $\mu\text{M}$  in phosphate buffer at pH 6.25), ThDP (100  $\mu\text{M}$  in water), and  $\text{MgCl}_2$  (100  $\mu\text{M}$  in water) in the DCL with 1 mL of end-volume. The DCL with the protein was allowed to gently mix on a rotating wheel (7 rpm) at room temperature and was frequently monitored via UPLC-MS, and the traces were compared with the blank composition. For analysis, 10  $\mu\text{L}$  of the corresponding library was mixed with 90  $\mu\text{L}$  acetonitrile and 2  $\mu\text{L}$  of NaOH (2 M), the mixture was centrifuged, and the supernatant was used for the analysis.

Note: The blank-DCL composition for competition experiments was similar except for the added protein. The protein- templated DCL was run as duplicate.

**Protein-templated DCL preparation for competition experiment with ThDP and BAP (GP-6):** To a 1.5 mL Eppendorf Tube® containing phosphate buffer (pH 6.25), were added hydrazides (300  $\mu\text{M}$  each, in DMSO), aldehydes (100  $\mu\text{M}$  each, in DMSO), aniline (10 mM, in DMSO), an additional amount of DMSO to reach a final

concentration of 5%, and a pre-incubated mixture of drDXPS protein (40  $\mu\text{M}$  in phosphate buffer at pH 6.25), ThDP (100  $\mu\text{M}$  in water),  $\text{MgCl}_2$  (100  $\mu\text{M}$  in water), and BAP (100  $\mu\text{M}$  in water) in the DCL with 1 mL of end-volume. The DCL with the protein was allowed to gently mix on a rotating wheel (7 rpm) at room temperature and was frequently monitored via UPLC-MS, and the traces were compared with the blank composition. For analysis, 10  $\mu\text{L}$  of the corresponding library was mixed with 90  $\mu\text{L}$  acetonitrile and 2  $\mu\text{L}$  of NaOH (2 M) to raise pH > 8. The mixture was centrifuged, and the supernatant was used for the analysis.

*Note:* The blank-DCL composition for the competition experiments was the same except for the added protein. The protein-templated DCL was run as duplicate.

**Assessment of DCL composition:** The library composition was assessed after the equilibrium of acylhydrazone formation was reached as reported in the literature.<sup>1,2</sup> All the DCLs in this study reached an equilibrium between four to six hours. The “amplification factor” was determined with the relative peak area (RPA), which is the percent of each peak when the sum of all peak areas was set to 100%. The “normalized RPA” was used for the final assessment of amplification of the acylhydrazone products in the DCL.

$$\text{amplification factor} = \frac{RPA_{(\text{templated})}}{RPA_{(\text{blank})}}$$

$$\text{normalized change of RPA} = \frac{(RPA_{(\text{templated})} - RPA_{(\text{blank})})}{RPA_{(\text{blank})}}$$

### General procedure for molecular docking studies (GP-7):

The X-ray crystal structure of *Deinococcus radiodurans* (drDXPS) (PDB code: 2O1X) was used for our docking studies.<sup>3</sup> The binding site of drDXPS was defined to include residues within a 6.5 Å radius around the cofactor thiaminediphosphate in the LeadIT program. The binding site includes 43 amino acid residues and a magnesium ion. All 21 possible acylhydrazones products from DCL-1 were used as ligands for docking using the FlexX docking module in the LeadIT suite v2.3.2.<sup>4</sup> Top-30 (FlexX)-scored solutions were retained and subsequently post-scored with the HYDE module in LeadIT v2.3.2. After careful visualization to exclude poses with significant inter- or intramolecular clash terms or unfavorable conformations, the resulting solutions were subsequently ranked according to their binding energies. The top-ranked pose from each compound was chosen and used for the comparison. The 2D-figures were generated with Pose View<sup>5</sup> as implemented in the LeadIT suite. The program SeeSAR 10.0 was used to visualize the acylhydrazone hits from the DCL-1 inside the active site of enzyme drDXPS and to generate the figures.<sup>6</sup>

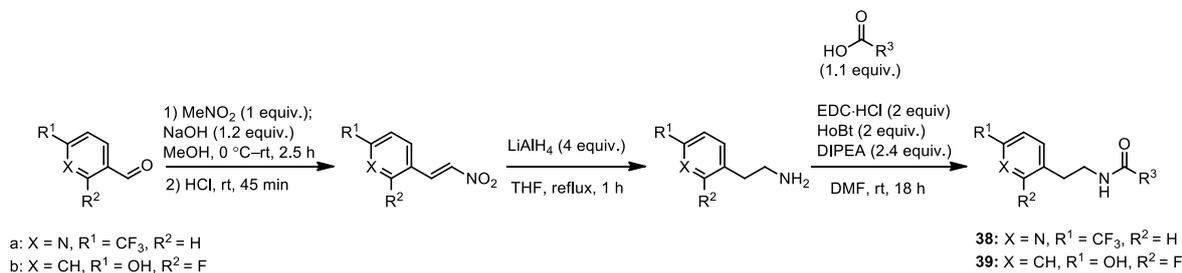
### General procedure for acylhydrazone formation (GP-8):

To a heat-dried Schlenk tube equipped with a magnetic stirring bar, the hydrazide (1–1.2 equiv.) and the corresponding aldehyde (1 equiv.) were dissolved/suspended in MeOH under nitrogen atmosphere. The reaction mixture was stirred at 65 °C until completion. After cooling the reaction to room temperature, the reaction mixture was precipitated by cooling the Schlenk tube to 0 °C in an ice bath. The resulting crude was purified either by filtration and washing sequences (Purification method-1) or by preparative HPLC, affording the corresponding acylhydrazone products in 21–94% isolated yields.

**Purification method-1:** The precipitated reaction mixture was transferred to an Eppendorf Tube® and centrifuged for 2 min, the supernatant liquid was removed, and 1 mL ice-cold MeOH was added to the residue, which was resuspended by vigorous agitation by a vortex mixer. The cold suspension was centrifuged again for 2 min, followed by removal of supernatant liquid. This process was repeated at least three times or until the sufficiently pure product (residue) was isolated, which was monitored on LC-MS. The solvents were removed from the final residue under reduced pressure to obtain the pure acylhydrazone product.

**Purification method-2:** After cooling to room temperature, the reaction mixture was concentrated *in vacuo* and purified by preparative HPLC on a reversed-phase column (C18 column, 5  $\mu\text{m}$ , Macherey-Nagel, Germany), using 15–100% gradient (MeCN (0.1% formic acid) in water (0.1% formic acid)). After removing water and acetonitrile from the pure fractions *in vacuo* followed by lyophilization, the pure acylhydrazone products were obtained.

## General procedure for amide bioisostere formation (GP-9):



**Step 1:** To a 50 mL round-bottomed flask were added aldehyde (1 equiv.), nitromethane (1 equiv.) and MeOH (2 mL). The reaction mixture was cooled to 0 °C, and an aqueous solution of NaOH (1.2 equiv. In 1.5 mL water) was added. The reaction was kept stirring for 2.5 hours while slowly warming to rt. After, the reaction mixture was poured into an aqueous solution of HCl (0.6 mL in 1 mL of water) and kept stirring at rt for an additional 45 minutes. The reaction progress was monitored on LCMS. After full conversion, water (5 mL) was added to the reaction mixture and extracted with DCM (3 × 20 mL). The combined organic phases were dried over  $\text{Na}_2\text{SO}_4$ , filtered, and volatiles were removed on a rotary evaporator. The crude nitrostyrene derivative obtained was used in the next step without further purification.

**Step 2:** The crude nitrostyrene derivative (approx, 1 mmol) from the first step was added to a two-necked round-bottomed flask, and the flask was purged using vacuum-argon cycles. The dry-THF (16 mL) was added to the reaction flask followed by portion-wise addition of  $\text{LiAlH}_4$  (4 equiv.), and the reaction mixture was kept stirring at rt for one hour. After the disappearance of starting material by LCMS, 1 M NaOH (10 mL) was slowly added to the reaction flask, and the resulting reaction mixture was filtered through Celite. The Celite pad was washed with EtOAc (3 × 20 mL), and the organic phase was separated, the aqueous phase was extracted with EtOAc and combined organic phases were treated with brine. After separation, the organic phases were dried over  $\text{Na}_2\text{SO}_4$ , filtered, and volatiles were removed on a rotary evaporator. The crude ethylamine derivative obtained was used in the next step without further purification.

**Step 3:** To a 50 mL round-bottomed flask, were added ethylamine (1equiv.), carboxylic acid (1.1 equiv.), HOBT. $\text{H}_2\text{O}$  (2 equiv.), and 2 mL DMF. To the resulting solution were added EDC-HCl (2 equiv.), and DIPEA (2.5 equiv.). The reaction was kept stirring at rt for 18 hours before treating it with water (5 mL) and extracting it with EtOAc (3 × 10 mL). The combined organic phases were dried over  $\text{Na}_2\text{SO}_4$ , filtered and volatiles were removed on a rotary evaporator. The crude obtained was purified by preparative HPLC ( $\text{H}_2\text{O}:\text{MeCN} + 0.1\%$  formic acid, gradient 5% to 100% MeCN) to afford the amide product.

## General procedure for oxadiazole bioisostere formation (GP-10):

Oxadiazole compounds were prepared according to the reported procedure of Farshori *et al.* (2017) under microwave conditions.<sup>7</sup> The original procedure was adapted as follows:

The hydrazide (1 equiv.), corresponding carboxylic acid (1 equiv.), and neutral active aluminum oxide (100 mg / 50 mg of hydrazide) were ground to a fine powder in a mortar. This powder was transferred into a microwave (MW) glass tube and phosphorous oxytrichloride was added in excess (14 equiv.). The MW reaction tube was capped under air, and the reaction mixture was heated in a laboratory microwave oven at 90 °C for 30 min. For purification, the crude was washed with ice-cold distilled water and ice-cold concentrated sodium bicarbonate solution. The residue was further washed with cold MeOH to provide the final pure product. Further purification by preparative HPLC was needed for compound **41**.

## DCC-experiments

### DCC-1:

This experiment library consists of three aldehydes (**A1–A3**) and seven hydrazides (**H1–H7**).

**DCL-1 (40 mol% protein):** The DCC-experiment was carried out according to the GP-1 (blank) and GP-2 (protein-templated) in phosphate buffer at pH 6.25 and 5% DMSO.

The DCL-composition is as below:

Entry	Blank		Protein-templated (I)		Protein-templated (II)	
	<i>amount</i>	<i>Final conc. in DCL</i>	<i>amount</i>	<i>Final conc. in DCL</i>	<i>amount</i>	<i>Final conc. in DCL</i>
Phosphate buffer	950 $\mu$ L	-	76 $\mu$ L	-	15.2 $\mu$ L	-
Hydrazide (100 mM)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)	7 $\times$ 0.6 $\mu$ L <sup>a</sup> (4.2 $\mu$ L)	(7 $\times$ 300 $\mu$ M)
Aldehyde (100 mM)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(3 $\times$ 100 $\mu$ M)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(3 $\times$ 100 $\mu$ M)	7 $\times$ 0.2 $\mu$ L <sup>b</sup> (0.6 $\mu$ L)	(3 $\times$ 100 $\mu$ M)
Aniline (1 M)	10 $\mu$ L	10 mM	10 $\mu$ L	10 mM	2 $\mu$ L	10 mM
DMSO	16 $\mu$ L	-	16 $\mu$ L	-	3.2 $\mu$ L	-
drDXPS (45.7 $\mu$ M)*	0	-	874 $\mu$ L	40 $\mu$ M	174.8 $\mu$ L	40 $\mu$ M

\*drDXPS, 45.7  $\mu$ M in phosphate buffer at pH 6.25. <sup>a</sup> to minimize the error, the 21  $\mu$ L stock solution was prepared by mixing 3  $\mu$ L of each hydrazide from the initial 100 mM stock, and 4.2  $\mu$ L were taken from this stock. <sup>b</sup> to minimize the error, the 6  $\mu$ L stock solution was prepared by mixing 2  $\mu$ L of each aldehyde from the initial 100 mM stock, and 0.6  $\mu$ L were taken from this stock.

**DCL-1 (20 mol% protein):** The DCC-experiment was carried out according to the GP-1 (blank), GP-2 (protein-templated) in phosphate buffer at pH 6.25 and 5% DMSO.

The DCL-composition is as below:

Entry	Blank		Protein-templated (I)		Protein-templated (II)	
	<i>amount</i>	<i>Final conc. in DCL</i>	<i>amount</i>	<i>Final conc. in DCL</i>	<i>amount</i>	<i>Final conc. in DCL</i>
Phosphate buffer	950 $\mu$ L	-	512.5 $\mu$ L	-	102.5 $\mu$ L	-
Hydrazide (100 mM)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)	7 $\times$ 0.6 $\mu$ L <sup>a</sup> (4.2 $\mu$ L)	(7 $\times$ 300 $\mu$ M)
Aldehyde (100 mM)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(3 $\times$ 100 $\mu$ M)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(3 $\times$ 100 $\mu$ M)	7 $\times$ 0.2 $\mu$ L <sup>b</sup> (0.6 $\mu$ L)	(3 $\times$ 100 $\mu$ M)
Aniline (1 M)	10 $\mu$ L	10 mM	10 $\mu$ L	10 mM	2 $\mu$ L	10 mM
DMSO	16 $\mu$ L	-	16 $\mu$ L	-	3.2 $\mu$ L	-
drDXPS (45.7 $\mu$ M)*	0	-	437.5 $\mu$ L	20 $\mu$ M	87.5 $\mu$ L	20 $\mu$ M

\* drDXPS, 45.7  $\mu$ M in phosphate buffer at pH 6.25. <sup>a</sup> to minimize the error, the 21  $\mu$ L stock solution was prepared by mixing 3  $\mu$ L of each hydrazide from the initial 100 mM stock, and 4.2  $\mu$ L were taken from this stock. <sup>b</sup> to minimize the error, the 6  $\mu$ L stock solution was prepared by mixing 2  $\mu$ L of each aldehyde from the initial 100 mM stock, and 0.6  $\mu$ L were taken from this stock.

**DCL-1 (20 mol% protein 'pre-equilibrated')**: The DCC-experiment was carried out according to GP-3 (blank) and GP-4 (protein-templated) in phosphate buffer at pH 6.25 and 5% DMSO.

The DCL-composition is as below:

Entry	Blank		Protein-templated	
	<i>amount</i>	<i>Final conc. in DCL</i>	<i>amount</i>	<i>Final conc. in DCL</i>
Phosphate buffer	512.5 + 437.5 $\mu$ L	-	512.5 $\mu$ L	-
Hydrazide (100 mM)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)
Aldehyde (100 mM)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(3 $\times$ 100 $\mu$ M)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(3 $\times$ 100 $\mu$ M)
Aniline (1 M)	10 $\mu$ L	10 mM	10 $\mu$ L	10 mM
DMSO	16 $\mu$ L	-	16 $\mu$ L	-
drDXPS (45.7 $\mu$ M)*	0	-	437.5 $\mu$ L	20 $\mu$ M

\* drDXPS, 45.7  $\mu$ M in phosphate buffer at pH 6.25

### DCC-2:

This experiment's library consists of three aldehydes (**A4–A6**) and seven hydrazides (**H2, H5, H6, H8–H11**).

This DCC-experiment was carried out according to the GP-1 (blank) and GP-2 (protein-templated) in phosphate buffer at pH 7.04 and 5% DMSO.

The DCL-composition is as below:

Entry	Blank		Protein-templated (I)		Protein-templated (II)	
	<i>amount</i>	<i>Final conc. in DCL</i>	<i>amount</i>	<i>Final conc. in DCL</i>	<i>amount</i>	<i>Final conc. in DCL</i>
Phosphate buffer	950	-	524.5 $\mu$ L	-	524.5 $\mu$ L	-
Hydrazide (100 mM)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)
Aldehyde (100 mM)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(3 $\times$ 100 $\mu$ M)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(3 $\times$ 100 $\mu$ M)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(3 $\times$ 100 $\mu$ M)
Aniline (1 M)	10 $\mu$ L	10 mM	10 $\mu$ L	10 mM	10 $\mu$ L	10 mM
DMSO	16 $\mu$ L	-	16 $\mu$ L	-	16 $\mu$ L	-
drDXPS (94 $\mu$ M)*	0	-	425.5 $\mu$ L	40 $\mu$ M	425.5 $\mu$ L	40 $\mu$ M

\* drDXPS, 94  $\mu$ M in phosphate buffer at pH 7.04

### DCC-3:

**DCC-3a**: This experiment's library consists of eleven aldehydes (**A1, A2, A7–A15**) and two hydrazides (**H11** and **H12**).

This DCC-experiment was carried out according to the GP-1 (blank) and GP-2 (Protein-templated) in phosphate buffer at pH 7.04 and 5% DMSO.

The DCL-composition is as below:

Entry	Blank		Protein-templated (I)		Protein-templated (II)	
	<i>amount</i>	<i>Final conc. in DCL</i>	<i>amount</i>	<i>Final conc. in DCL</i>	<i>amount</i>	<i>Final conc. in DCL</i>
Phosphate buffer	950	-	590 $\mu$ L	-	590 $\mu$ L	-
Hydrazide (200 mM)	2 $\times$ 10 $\mu$ L (20 $\mu$ L)	(2 $\times$ 2000 $\mu$ M)	2 $\times$ 10 $\mu$ L (20 $\mu$ L)	(2 $\times$ 2000 $\mu$ M)	2 $\times$ 10 $\mu$ L (20 $\mu$ L)	(2 $\times$ 2000 $\mu$ M)
Aldehyde (100 mM)	11 $\times$ 1 $\mu$ L (11 $\mu$ L)	(11 $\times$ 100 $\mu$ M)	11 $\times$ 1 $\mu$ L (11 $\mu$ L)	(11 $\times$ 100 $\mu$ M)	11 $\times$ 1 $\mu$ L (11 $\mu$ L)	(11 $\times$ 100 $\mu$ M)
Aniline (1 M)	10 $\mu$ L	10 mM	10 $\mu$ L	10 mM	10 $\mu$ L	10 mM
DMSO	9 $\mu$ L	-	9 $\mu$ L	-	9 $\mu$ L	-
drDXPS (111 $\mu$ M)*	0	-	360 $\mu$ L	40 $\mu$ M	360 $\mu$ L	40 $\mu$ M

\* drDXPS, 111  $\mu$ M in phosphate buffer at pH 7.04

**DCC-3b:** This experiment library consists of eleven aldehydes (**A16 –A26**) and two hydrazides (**H11** and **H12**). This DCC-experiment was carried out according to the GP-1 (blank) and GP-2 (Protein-templated) in phosphate buffer at pH 7.04 and 5% DMSO.

The DCL-composition is as below:

Entry	Blank		Protein-templated (I)		Protein-templated (II)	
	<i>amount</i>	<i>Final conc. in DCL</i>	<i>amount</i>	<i>Final conc. in DCL</i>	<i>amount</i>	<i>Final conc. in DCL</i>
Phosphate buffer	950 $\mu$ L	-	590 $\mu$ L	-	590 $\mu$ L	-
Hydrazide (200 mM)	2 $\times$ 10 $\mu$ L (20 $\mu$ L)	(2 $\times$ 2000 $\mu$ M)	2 $\times$ 10 $\mu$ L (20 $\mu$ L)	(2 $\times$ 2000 $\mu$ M)	2 $\times$ 10 $\mu$ L (20 $\mu$ L)	(2 $\times$ 2000 $\mu$ M)
Aldehyde (100 mM)	11 $\times$ 1 $\mu$ L (11 $\mu$ L)	(11 $\times$ 100 $\mu$ M)	11 $\times$ 1 $\mu$ L (11 $\mu$ L)	(11 $\times$ 100 $\mu$ M)	11 $\times$ 1 $\mu$ L (11 $\mu$ L)	(11 $\times$ 100 $\mu$ M)
Aniline (1 M)	10 $\mu$ L	10 mM	10 $\mu$ L	10 mM	10 $\mu$ L	10 mM
DMSO	9 $\mu$ L	-	9 $\mu$ L	-	9 $\mu$ L	-
drDXPS (111 $\mu$ M)*	0	-	360 $\mu$ L	40 $\mu$ M	360 $\mu$ L	40 $\mu$ M

\* drDXPS, 111  $\mu$ M in phosphate buffer at pH 7.04

#### DCC-4

This DCC-experiment was carried out according to the GP-1 (blank) and GP-2 (Protein-templated) in phosphate buffer at pH 7.04 and 5% DMSO.

The DCL-composition is as below:

Entry	Blank		Protein-templated (I)		Protein-templated (II)	
	<i>amount</i>	<i>Final conc. in DCL</i>	<i>amount</i>	<i>Final conc. in DCL</i>	<i>amount</i>	<i>Final conc. in DCL</i>
Phosphate buffer	950	-	590 $\mu$ L	-	590 $\mu$ L	-
Hydrazide (100 mM)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)
Aldehyde (100 mM)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(11 $\times$ 100 $\mu$ M)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(11 $\times$ 100 $\mu$ M)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(11 $\times$ 100 $\mu$ M)
Aniline (1 M)	10 $\mu$ L	10 mM	10 $\mu$ L	10 mM	10 $\mu$ L	10 mM
DMSO	9 $\mu$ L	-	16 $\mu$ L	-	9 $\mu$ L	-
drDXPS (111 $\mu$ M)*	0	-	360 $\mu$ L	40 $\mu$ M	360 $\mu$ L	40 $\mu$ M

\* drDXPS, 111  $\mu$ M in phosphate buffer at pH 7.04

#### DCC-5 (Competition experiment with ThDP):

This DCC-experiment was carried out according to the GP-5 (blank and protein-templated) in phosphate buffer at pH 6.25 and 5% DMSO using the DCL-1.

The DCL-composition is as below:

Entry	Blank		Protein-templated (I)		Protein-templated (II)	
	<i>amount</i>	<i>Final conc. in DCL</i>	<i>amount</i>	<i>Final conc. in DCL</i>	<i>amount</i>	<i>Final conc. in DCL</i>
Phosphate buffer	948 $\mu$ L	-	587.6 $\mu$ L	-	587.6 $\mu$ L	-
Hydrazide (100 mM)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)
Aldehyde (100 mM)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(3 $\times$ 100 $\mu$ M)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(3 $\times$ 100 $\mu$ M)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(3 $\times$ 100 $\mu$ M)
Aniline (1 M)	10 $\mu$ L	10 mM	10 $\mu$ L	10 mM	10 $\mu$ L	10 mM
DMSO	16 $\mu$ L	-	16 $\mu$ L	-	16 $\mu$ L	-
ThDP (100 mM)	1 $\mu$ L	100 $\mu$ M	1 $\mu$ L	100 $\mu$ M	1 $\mu$ L	100 $\mu$ M
MgCl <sub>2</sub> (100 mM)	1 $\mu$ L	100 $\mu$ M	1 $\mu$ L	100 $\mu$ M	1 $\mu$ L	100 $\mu$ M
drDXPS (111 $\mu$ M)*	0	-	360.4 $\mu$ L	40 $\mu$ M	360.4 $\mu$ L	40 $\mu$ M

\* drDXPS, 111  $\mu$ M in phosphate buffer at pH 6.25, ThDP = Thiamine diphosphate.

**DCC-6 (Competition experiment with ThDP and BAP):**

This DCC-experiment was carried out according to the GP-6 (blank and protein-templated) in phosphate buffer at pH 6.25 and 5% DMSO using the DCL-1.

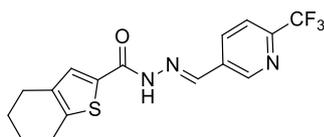
The DCL-composition is as below:

Entry	Blank		Protein-templated (I)		Protein-templated (II)	
	amount	Final conc. in DCL	amount	Final conc. in DCL	amount	Final conc. in DCL
Phosphate buffer	947 $\mu$ L	-	586.6 $\mu$ L	-	586.6 $\mu$ L	-
Hydrazide (100 mM)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)	7 $\times$ 3 $\mu$ L (21 $\mu$ L)	(7 $\times$ 300 $\mu$ M)
Aldehyde (100 mM)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(3 $\times$ 100 $\mu$ M)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(3 $\times$ 100 $\mu$ M)	3 $\times$ 1 $\mu$ L (3 $\mu$ L)	(3 $\times$ 100 $\mu$ M)
Aniline (1 M)	10 $\mu$ L	10 mM	10 $\mu$ L	10 mM	10 $\mu$ L	10 mM
DMSO	16 $\mu$ L	-	16 $\mu$ L	-	16 $\mu$ L	-
ThDP (100 mM)	1 $\mu$ L	100 $\mu$ M	1 $\mu$ L	100 $\mu$ M	1 $\mu$ L	100 $\mu$ M
MgCl <sub>2</sub> (100 mM)	1 $\mu$ L	100 $\mu$ M	1 $\mu$ L	100 $\mu$ M	1 $\mu$ L	100 $\mu$ M
BAP	1 $\mu$ L	100 $\mu$ M	1 $\mu$ L	100 $\mu$ M	1 $\mu$ L	100 $\mu$ M
drDXPS ((111 $\mu$ M)*)	0	-	360.4 $\mu$ L	40 $\mu$ M	360.4 $\mu$ L	40 $\mu$ M

\* drDXPS, 111  $\mu$ M in phosphate buffer at pH 6.25, ThDP = Thiamine diphosphate, BAP = butylacetylphosphonate

**Characterization of acylhydrazones****Identification of different conformers of acylhydrazone**

Analysis of <sup>1</sup>H NMR spectra of acylhydrazone products reported in this study gave two sets of signals. The two-dimensional Nuclear Overhauser Effect Spectroscopy (2D NOESY) experiments revealed (See Figure 21) that acylhydrazone **24** presents the relative configuration *E* of imine double bond (C=N) and the splitting of signals can be attributed to the existence of *cis/trans*-amide conformers.

**(*E*)-*N*'-((6-(Trifluoromethyl)pyridin-3-yl)methylene)-4,5,6,7-tetrahydrobenzo[*b*]thiophene-2-carbohydrazone (1):**

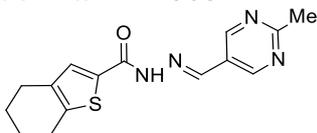
Compound **1** was synthesized according to GP-8, starting with aldehyde **A3** (87.6 mg, 0.5 mmol) and hydrazide **H1** (100 mg, 0.51 mmol). The crude was purified by using purification method-1 to afford **1** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 50:50) as a pale yellow solid (133.7 mg, 0.38 mmol, 75.7%).

<sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>)  $\delta$  12.10 (s, 2H, 1*trans* and 1*cis*), 9.08 (d, *J* = 36.6 Hz, 2H, 1*trans* and 1*cis*), 8.51 (s, 1H, *trans*), 8.40 (d, *J* = 7.3 Hz, 2H, 1*trans* and 1*cis*), 8.19 (s, 1H, *cis*), 8.01 (br s, 2H, 1*trans* and 1*cis*), 7.69 (d, *J* = 38.2 Hz, 2H, 1*trans* and 1*cis*), 2.79 (br s, 4H, 2*trans* and 2*cis*), 2.62 (t, *J* = 5.8 Hz, 4H, 2*trans* and 2*cis*), 1.95 – 1.66 (m, 8H, 4*trans* and 4*cis*).

<sup>13</sup>C NMR (126 MHz, DMSO-*d*<sub>6</sub>)  $\delta$  162.1, 158.5, 149.2, 147.1, 145.8, 143.0, 139.9, 136.4, 136.0, 135.3, 134.1, 130.5, 128.6, 125.32 123.1, 121.5, 121.0, 118.8, 25.1, 23.2, 22.6.

<sup>19</sup>F NMR (470 MHz, DMSO)  $\delta$  -66.4.

HRMS (ESI<sup>+</sup>): *m/z* calcd. for C<sub>16</sub>H<sub>15</sub>F<sub>3</sub>N<sub>3</sub>OS ([M+H]<sup>+</sup>) 354.08879, measured 354.08773.

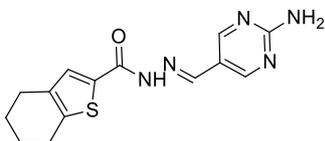
**(E)-N'-((2-Methylpyrimidin-5-yl)methylene)-4,5,6,7-tetrahydrobenzo[b]thiophene-2-carbohydrazide (2):**

Compound **2** was synthesized according to GP-8, starting with aldehyde **A1** (61.2 mg, 0.5 mmol) and hydrazide **H1** (100 mg, 0.51 mmol). The crude was purified by using purification method-1 to afford **1** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 50:50) as a pale yellow solid (79.0 mg, 0.38 mmol, 52.6%).

$^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  11.98 (s, 2H, 1*trans* and 1*cis*), 9.00 (d,  $J$  = 28.0 Hz, 4H, 2*trans* and 2*cis*), 8.40 (s, 1H, *trans*), 8.07 (s, 1H, *cis*), 7.67 (d,  $J$  = 45.9 Hz, 2H, 1*trans* and 1*cis*), 2.78 (br s, 4H, 2*trans* and 2*cis*), 2.66 (s, 6H, 3*trans* and 3*cis*), 2.61 (t,  $J$  = 5.8 Hz, 4H, 2*trans* and 2*cis*), 1.85 – 1.69 (m, 8H, 4*trans* and 4*cis*).

$^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  168.3, 168.1, 162.0, 158.4, 155.6, 145.6, 142.6, 142.2, 138.7, 136.6, 136.2, 135.2, 134.0, 130.3, 128.7, 125.9, 26.1, 25.4, 25.2, 25.0, 23.2, 22.7.

HRMS (ESI $^+$ ):  $m/z$  calcd. for  $\text{C}_{15}\text{H}_{17}\text{N}_4\text{OS}$  ( $[\text{M}+\text{H}]^+$ ) 301.11231, measured 301.11126.

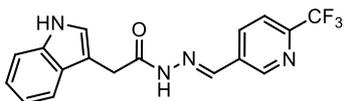
**(E)-N'-((2-Aminopyrimidin-5-yl)methylene)-4,5,6,7-tetrahydrobenzo[b]thiophene-2-carbohydrazide (3):**

Compound **3** was synthesized according to GP-8, starting with aldehyde **A2** (61.5 mg, 0.5 mmol) and hydrazide **H1** (100 mg, 0.51 mmol). The crude was purified by using purification method-1 to afford **3** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 50:50) as a pale yellow solid (136.1 mg, 0.45 mmol, 90.1%).

$^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  11.61 (d,  $J$  = 11.9 Hz, 2H, 1*trans* and 1*cis*), 8.56 (d,  $J$  = 29.9 Hz, 4H, 2*trans* and 2*cis*), 8.21 (s, 1H, *trans*), 7.88 (s, 1H, *cis*), 7.68 (s, 1H, *trans*), 7.56 (s, 1H, *cis*), 7.17 (s, 4H, 2*trans* and 2*cis*), 2.84 – 2.67 (m, 4H, 2*trans* and 2*cis*), 2.58 (t,  $J$  = 5.7 Hz, 4H, 2*trans* and 2*cis*), 1.88 – 1.58 (m, 8H, 4*trans* and 4*cis*).

$^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  164.2, 164.0, 161.6, 158.1, 157.6, 157.5, 145.2, 143.8, 141.9, 140.1, 136.5, 135.7, 135.0, 134.5, 129.7, 129.2, 117.6, 25.4, 25.1, 24.9, 23.3, 22.7, 22.6.

HRMS (ESI $^+$ ):  $m/z$  calcd. for  $\text{C}_{14}\text{H}_{16}\text{N}_5\text{OS}$  ( $[\text{M}+\text{H}]^+$ ) 302.10756, measured 302.10656.

**(E)-2-(1H-Indol-3-yl)-N'-((6-(trifluoromethyl)pyridin-3-yl)methylene)acetohydrazide (4):**

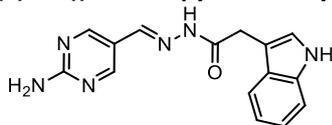
Compound **4** was synthesized according to GP-8, starting with aldehyde **A3** (92.63 mg, 0.53 mmol) and hydrazide **H7** (100 mg, 0.53 mmol). The crude was purified by using purification method-1 to afford **4** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 59:41) as a white solid (71.9 mg, 0.21 mmol, 40.0 %).

$^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  11.91 (br s, 1H, *trans*), 11.69 (br s, 1H, *cis*), 10.97 (s, 1H, *trans*), 10.91 (s, 1H, *cis*), 9.07 (d,  $J$  = 2.0 Hz, 1H, *cis*), 9.00 (d,  $J$  = 2.0 Hz, 1H, *trans*), 8.41 (dd,  $J$  = 8.3, 2.1 Hz, 1H, *cis*), 8.37 (s, 1H, *trans*), 8.34 (dd,  $J$  = 8.2, 2.1 Hz, 1H, *trans*), 8.10 (s, 1H, *cis*), 7.96 (d,  $J$  = 8.0 Hz, 2H, 1*trans* and 1*cis*), 7.59 (t,  $J$  = 7.2 Hz, 2H, 1*trans* and 1*cis*), 7.42 – 7.32 (m, 2H, 1*trans* and 1*cis*), 7.27 (dd,  $J$  = 8.1, 2.4 Hz, 2H, 1*trans* and 1*cis*), 7.14 – 7.03 (m, 2H, 1*trans* and 1*cis*), 7.03 – 6.93 (m, 2H, 1*trans* and 1*cis*), 4.11 (s, 2H, *cis*), 3.69 (s, 2H, *trans*).

$^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  173.6, 168.1, 149.2, 149.0, 147.1, 147.1, 146.8, 146.8, 146.6, 146.5, 146.3, 142.3, 138.6, 136.6, 136.5, 135.9, 135.8, 134.15, 134.1, 127.8, 127.6, 125.4, 125.3, 123.2, 123.15, 121.0, 121.0, 118.8, 111.9, 111.8, 108.3, 32.2, 29.7.

$^{19}\text{F}$  NMR (470 MHz, DMSO)  $\delta$  -66.4, -66.4.

HRMS (ESI $^+$ ):  $m/z$  calcd. for  $\text{C}_{17}\text{H}_{14}\text{F}_3\text{N}_4\text{O}$  ( $[\text{M}+\text{H}]^+$ ) 347.11197, measured 347.11080.

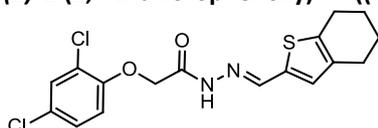
**(E)-N'-((2-Aminopyrimidin-5-yl)methylene)-2-(1H-indol-3-yl)acetohydrazide (5):**

Compound **5** was synthesized according to GP-8, starting with aldehyde **A2** (135.34 mg, 1.1 mmol) and hydrazide **H7** (228.1 mg, 1.2 mmol). The crude was purified by using purification method-1 to afford **5** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 53:47) as a white solid (132.0 mg, 0.45 mmol, 40.8%).

<sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>) δ 11.42 (s, 1H, *trans*), 11.17 (s, 1H, *cis*), 10.91 (s, 1H, *trans*), 10.86 (s, 1H, *cis*), 8.55 (s, 2H, *cis*), 8.49 (s, 2H, *trans*), 8.04 (s, 1H, *trans*), 7.80 (s, 1H, *cis*), 7.56 (t, *J* = 7.5 Hz, 2H, 1*trans* and 1*cis*), 7.33 (t, *J* = 8.4 Hz, 2H, 1*trans* and 1*cis*), 7.21 (dd, *J* = 12.3, 2.4 Hz, 2H, 1*trans* and 1*cis*), 7.12 (d, *J* = 10.2 Hz, 4H, 2*trans* and 2*cis*), 7.10 – 7.00 (m, 2H, 1*trans* and 1*cis*), 7.00 – 6.91 (m, 2H, 1*trans* and 1*cis*), 4.00 (s, 2H, *cis*), 3.59 (s, 2H, *trans*).

<sup>13</sup>C NMR (126 MHz, DMSO-*d*<sub>6</sub>) δ 162.2, 158.5, 149.2, 147.1, 146.6, 145.8, 142.9, 139.9, 136.7, 136.4, 136.0, 135.3, 134.1, 133.9, 130.5, 128.6, 125.3, 123.2, 121.5, 121.0, 118.8, 25.2, 23.2, 22.6.

HRMS (ESI<sup>+</sup>): *m/z* calcd. for C<sub>15</sub>H<sub>15</sub>N<sub>6</sub>O ([M+H]<sup>+</sup>) 295.13073, measured 295.12972.

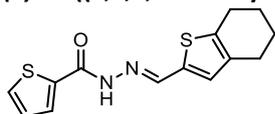
**(E)-2-(2,4-Dichlorophenoxy)-N'-((4,5,6,7-tetrahydrobenzo[*b*]thiophen-2-yl)methylene)acetohydrazide (6):**

Compound **6** was synthesized according to GP-8, starting with aldehyde **A4** (25.5 mg, 0.15 mmol) and hydrazide **H9** (30.0 mg, 0.13 mmol). The crude was purified by using purification method-1 to afford **6** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 40:60) as a pale yellow solid (40.0 mg, 0.10 mmol, 80.0%).

<sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>) δ 11.52 (d, *J* = 32.5 Hz, 2H, 1*trans* and 1*cis*), 8.35 (s, 1H, *trans*), 8.07 (s, 1H, *cis*), 7.61 (d, *J* = 2.6 Hz, 1H, *trans*), 7.57 (d, *J* = 2.6 Hz, 1H, *cis*), 7.38 (dd, *J* = 8.9, 2.6 Hz, 1H, *trans*), 7.32 (dd, *J* = 8.9, 2.6 Hz, 1H, *cis*), 7.22 – 6.94 (m, 4H, 2*trans* and 2*cis*), 5.15 (s, 2H, *cis*), 4.75 (s, 2H, *trans*), 2.78 – 2.65 (m, 4H, 2*trans* and 2*cis*), 2.59 – 2.50 (m, 4H, 2*trans* and 2*cis*), 1.84 – 1.61 (m, 8H, 4*trans* and 4*cis*).

<sup>13</sup>C NMR (126 MHz, DMSO-*d*<sub>6</sub>) δ 167.7, 163.1, 152.8, 152.6, 143.2, 139.5, 138.9, 138.4, 135.7, 134.7, 134.6, 132.2, 131.4, 129.4, 129.3, 128.1, 127.9, 125.1, 124.5, 122.5, 122.2, 115.4, 115.1, 67.1, 65.4, 24.84, 24.79, 24.7, 22.8, 22.2.

HRMS (ESI<sup>+</sup>): *m/z* calcd. for C<sub>17</sub>H<sub>17</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>2</sub>S ([M+H]<sup>+</sup>) 383.03878, measured 383.03655.

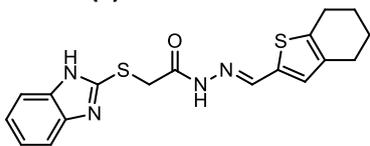
**(E)-N'-((4,5,6,7-Tetrahydrobenzo[*b*]thiophen-2-yl)methylene)thiophene-2-carbohydrazide (7):**

Compound **7** was synthesized according to GP-8, starting with aldehyde **A4** (42.0 mg, 0.25 mmol) and hydrazide **H11** (30.0 mg, 0.21 mmol). The crude was purified by using purification method-1 to afford **7** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 54:46) as a pale yellow solid (25.0 mg, 0.086 mmol, 41.0%).

<sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>) δ 11.72 (d, *J* = 4.3 Hz, 2H, 1*trans* and 1*cis*), 8.53 (s, 1H, *trans*), 8.17 (s, 1H, *cis*), 8.02 (d, *J* = 3.7 Hz, 1H, *trans*), 7.95 (d, *J* = 5.0 Hz, 1H, *trans*), 7.85 (d, *J* = 4.5 Hz, 2H, 1*trans* and 1*cis*), 7.20 (t, *J* = 4.5 Hz, 2H, 1*trans* and 1*cis*), 7.14 (d, *J* = 8.3 Hz, 2H, 1*trans* and 1*cis*), 2.80 – 2.68 (m, 4H, 2*trans* and 2*cis*), 2.59 – 2.51 (m, 4H, 2*trans* and 2*cis*), 1.87 – 1.61 (m, 8H, 4*trans* and 4*cis*).

<sup>13</sup>C NMR (126 MHz, DMSO-*d*<sub>6</sub>) δ 160.9, 157.5, 143.0, 138.8, 138.3, 135.8, 135.7, 135.0, 134.9, 134.5, 133.4, 132.0, 131.7, 128.7, 128.1, 126.6, 24.9, 24.8, 22.9, 22.2.

HRMS (ESI<sup>+</sup>): *m/z* calcd. for C<sub>14</sub>H<sub>15</sub>N<sub>2</sub>OS<sub>2</sub> ([M+H]<sup>+</sup>) 291.06258, measured 291.06025.

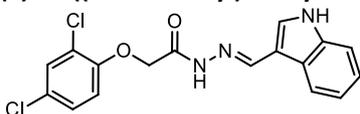
**(E)-2-((1H-Benzo[d]imidazole-2-yl)thio)-N'-((4,5,6,7-tetrahydrobenzo[b]thiophen-2-yl)methylene)acetohydrazide (8):**

Compound **8** was synthesized according to GP-8, starting with aldehyde **A4** (26.0 mg, 0.16 mmol) and hydrazide **H6** (30.0 mg, 0.13 mmol). The crude was purified by using purification method-1 to afford **8** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 54:46) as a white solid (16.0 mg, 0.04 mmol, 33.2%).

$^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  12.64 (br s, 1H, *trans*), 12.57 (br s, 1H, *cis*), 11.70 (s, 1H, *trans*), 11.52 (s, 1H, *cis*), 8.30 (s, 1H, *trans*), 8.09 (s, 1H, *cis*), 7.49 (t,  $J = 7.4$  Hz, 2H, 1*trans* and 1*cis*), 7.37 (t,  $J = 8.8$  Hz, 2H, 1*trans* and 1*cis*), 7.19 – 7.03 (m, 6H, 3*trans* and 3*cis*), 4.46 (s, 2H, 1*trans* and 1*cis*), 4.12 (s, 2H, 1*trans* and 1*cis*), 2.70 (q,  $J = 6.6$  Hz, 4H, 2*trans* and 2*cis*), 2.54 (t,  $J = 6.3$  Hz, 4H, 2*trans* and 2*cis*), 1.84 – 1.63 (m, 8H, 4*trans* and 4*cis*).

$^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  169.0, 164.0, 150.4, 150.1, 142.9, 139.5, 139.3, 138.8, 136.2, 135.2, 135.2, 132.6, 131.8, 121.9, 117.8, 110.6, 34.7, 34.0, 25.3, 25.3, 25.24, 25.19, 23.3, 22.7, 22.6.

HRMS (ESI $^+$ ):  $m/z$  calcd. for  $\text{C}_{18}\text{H}_{19}\text{N}_4\text{OS}_2$  ([ $\text{M}+\text{H}$ ] $^+$ ) 371.10003, measured 371.09790.

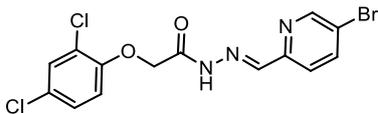
**(E)-N'-((1H-Indol-3-yl)methylene)-2-(2,4-dichlorophenoxy)acetohydrazide (9):**

Compound **9** was synthesized according to GP-8, starting with aldehyde **A5** (22.1 mg, 0.15 mmol) and hydrazide **H9** (30.0 mg, 0.13 mmol). The crude was purified by using purification method-1 to afford **9** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 76:24) as a white solid (20.0 mg, 0.055 mmol, 42.5%).

$^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  11.69 – 11.47 (m, 2H, 1*trans* and 1*cis*), 11.33 (d,  $J = 49.8$  Hz, 2H, 1*trans* and 1*cis*), 8.42 (s, 1H, *trans*), 8.21 (s, 1H, *trans*), 8.20 (s, 1H, *cis*), 8.11 (d,  $J = 7.9$  Hz, 1H, *cis*), 7.82 (d,  $J = 2.2$  Hz, 2H, 1*trans* and 1*cis*), 7.63 (d,  $J = 2.6$  Hz, 1H, *trans*), 7.60 (d,  $J = 2.6$  Hz, 1H, *cis*), 7.44 (d,  $J = 8.1$  Hz, 2H, 1*trans* and 1*cis*), 7.40 (dd,  $J = 8.9, 2.6$  Hz, 1H, *trans*), 7.35 (dd,  $J = 8.9, 2.6$  Hz, 1H, *cis*), 7.24 – 7.17 (m, 2H, 1*trans* and 1*cis*), 7.17 – 7.02 (m, 4H, 2*trans* and 2*cis*), 5.35 (s, 4H, *trans*), 4.78 (s, 4H, *cis*).

$^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  167.8, 163.0, 153.4, 153.2, 145.4, 142.0, 137.5, 131.2, 129.9, 129.7, 128.6, 128.4, 125.5, 124.8, 124.7, 124.5, 123.1, 123.0, 122.6, 122.4, 122.3, 121.1, 120.9, 115.9, 115.5, 112.3, 111.8, 111.7, 67.8, 66.3.

HRMS (ESI $^+$ ):  $m/z$  calcd. for  $\text{C}_{17}\text{H}_{14}\text{Cl}_2\text{N}_3\text{O}_2$  ([ $\text{M}+\text{H}$ ] $^+$ ) 362.04631, measured 362.04400.

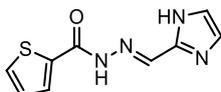
**(E)-N'-((5-Bromopyridin-2-yl)methylene)-2-(2,4-dichlorophenoxy)acetohydrazide (10):**

Compound **10** was synthesized according to GP-8, starting with aldehyde **A6** (22.1 mg, 0.15 mmol) and hydrazide **H9** (30.0 mg, 0.13 mmol). The crude was purified by using purification method-1 to afford **10** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 77:23) as a white solid (33.0 mg, 0.08 mmol, 63.0%).

$^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  11.94 (d,  $J = 12.1$  Hz, 2H, 1*trans* and 1*cis*), 8.81 – 8.66 (m, 2H, 1*trans* and 1*cis*), 8.24 (s, 1H, *trans*), 8.17 – 8.09 (m, 2H, 1*trans* and 1*cis*), 8.00 (s, 1H, *cis*), 7.95 (d,  $J = 8.5$  Hz, 1H, *cis*), 7.86 (d,  $J = 8.5$  Hz, 1H, *trans*), 7.62 (d,  $J = 2.6$  Hz, 1H, *trans*), 7.58 (d,  $J = 2.6$  Hz, 1H, *cis*), 7.38 (dd,  $J = 8.9, 2.6$  Hz, 1H, *trans*), 7.33 (dd,  $J = 8.9, 2.6$  Hz, 1H, *cis*), 7.11 (dd,  $J = 8.9, 4.1$  Hz, 2H, 1*trans* and 1*cis*), 5.34 (s, 2H, *cis*), 4.84 (s, 2H, *trans*).

$^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  168.6, 164.0, 152.8, 152.6, 151.8, 151.6, 150.3, 146.8, 143.2, 139.7, 139.5, 129.5, 129.3, 128.1, 127.9, 125.2, 124.6, 122.5, 122.2, 121.6, 121.5, 121.0, 120.8, 115.4, 115.2, 67.1, 65.7.

HRMS (ESI $^+$ ):  $m/z$  calcd. for  $\text{C}_{14}\text{H}_{11}\text{BrCl}_2\text{N}_3\text{O}_2$  ([ $\text{M}+\text{H}$ ] $^+$ ) 401.94117, measured 401.93877

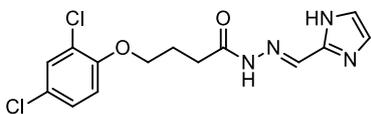
**(E)-N'-((1H-Imidazol-2-yl)methylene)thiophene-2-carbohydrazide (11)**

Compound **11** was synthesized according to GP-8, starting with aldehyde **A12** (50.0 mg, 0.52 mmol) and hydrazide **H11** (73.98 mg, 0.52 mmol). The crude was purified by using purification method-2 (Prep-HPLC) to afford **11** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 28:72) as a white solid (25 mg, 0.11 mmol, 21.8%).

$^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  14.60 (br s, 1H, *trans*), 13.57 (br s, 1H, *cis*), 13.09 (s, 2H, 1*trans* and 1*cis*), 8.12 (br s, 1H, *cis*), 7.95 (d,  $J$  = 5.0 Hz, 2H, 1*trans* and 1*cis*), 7.78 (br s, 1H, *trans*), 7.66 – 7.31 (m, 6H, 3*trans* and 3*cis*), 7.33 – 7.18 (m, 2H, 1*trans* and 1*cis*).

$^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  162.2, 158.3, 141.9, 141.4, 138.3, 135.6, 133.1, 132.8, 130.1, 130.0, 129.3, 129.0, 127.4, 127.0 119.4.

HRMS (ESI $^+$ ):  $m/z$  calcd. for  $\text{C}_9\text{H}_9\text{N}_4\text{OS}$  ([ $\text{M}+\text{H}$ ] $^+$ ) 221.04971, measured 221.04888.

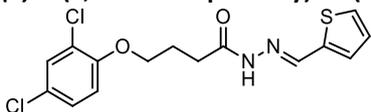
**(E)-N'-((1H-Imidazol-2-yl)methylene)-4-(2,4-dichlorophenoxy)butanehydrazide (12):**

Compound **12** was synthesized according to GP-8, starting with aldehyde **A12** (54.8 mg, 0.21 mmol) and hydrazide **H12** (20 mg, 0.21 mmol). The crude was purified by using purification method-1 (Prep-HPLC) to afford **12** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 56:44) as a white solid (53.9 mg, 0.16 mmol, 76.2%).

$^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  12.76 (br s, 1H, *trans*), 12.51 (br s, 1H, *cis*), 11.45 (s, 1H, *trans*), 11.32 (s, 1H, *cis*), 8.06 (s, 1H, *trans*), 7.88 (s, 1H, *cis*), 7.64 – 7.50 (m, 2H, 1*trans* and 1*cis*), 7.43 – 7.33 (m, 2H, 1*trans* and 1*cis*), 7.24 (br s, 1H, *trans*), 7.23 – 7.14 (m, 2H, 1*trans* and 1*cis*), 7.04 (br s, 1H, *cis*), 4.12 (dt,  $J$  = 13.6, 6.3 Hz, 4H, 2*trans* and 2*cis*), 2.83 (t,  $J$  = 7.3 Hz, 2H, *cis*), 2.43 (t,  $J$  = 7.4 Hz, 2H, *trans*), 2.05 (dp,  $J$  = 13.7, 6.8, 6.4 Hz, 4H, 2*trans* and 2*cis*).

$^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  174.2, 168.4, 153.4, 143.0, 142.9, 138.3, 134.8, 130.3, 130.1, 129.8, 128.6, 124.9, 124.78, 122.83, 119.8, 119.3, 115.5, 115.4, 68.9, 30.9, 28.6, 24.7, 24.1.

HRMS (LC-MS/MS):  $m/z$  calcd. for  $\text{C}_{14}\text{H}_{15}\text{Cl}_2\text{N}_4\text{O}_2$  ([ $\text{M}+\text{H}$ ] $^+$ ) 341.05721, measured 341.05618.

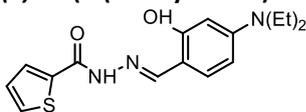
**(E)-4-(2,4-Dichlorophenoxy)-N'-(thiophen-2-ylmethylene)butanehydrazide (16):**

Compound **16** was synthesized according to GP-8, starting with aldehyde **A7** (54.8 mg, 0.21 mmol) and hydrazide **H12** (23.3 mg, 0.21 mmol). The crude was purified by using purification method-1 to afford **16** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 58:42) as a white solid (51.9 mg, 0.14 mmol, 69.4 %).

$^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  11.39 (s, 1H, *trans*), 11.30 (s, 1H, *cis*), 8.37 (s, 1H, *trans*), 8.15 (s, 1H, *cis*), 7.64 (d,  $J$  = 5.1 Hz, 1H, *trans*), 7.60 (d,  $J$  = 5.0 Hz, 1H, *cis*), 7.57 (dd,  $J$  = 4.5, 2.6 Hz, 2H, 1*trans* and 1*cis*), 7.47 – 7.31 (m, 4H, 2*trans* and 2*cis*), 7.19 (dd,  $J$  = 8.9, 1.5 Hz, 2H, 1*trans* and 1*cis*), 7.15 – 7.06 (m, 2H, 1*trans* and 1*cis*), 4.12 (dt,  $J$  = 15.4, 6.3 Hz, 4H, 2*trans* and 2*cis*), 2.76 (t,  $J$  = 7.3 Hz, 2H, 1*trans* and 1*cis*), 2.40 (t,  $J$  = 7.4 Hz, 2H, 1*trans* and 1*cis*), 2.12 – 1.91 (m, 4H, 2*trans* and 2*cis*).

$^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  173.9, 168.3, 153.4, 141.5, 139.6, 139.5, 138.3, 131.1, 130.6, 129.7, 129.1, 128.6, 128.6, 128.3, 128.3, 124.8, 124.8, 122.9, 122.8, 115.5, 68.8, 68.8, 30.8, 28.6, 24.7, 24.0.

HRMS (ESI $^+$ ):  $m/z$  calcd. For  $\text{C}_{15}\text{H}_{15}\text{Cl}_2\text{N}_2\text{O}_2\text{S}$  ([ $\text{M}+\text{H}$ ] $^+$ ) 357.02313, measured 357.02189.

**(E)-N'-(4-(Diethylamino)-2-hydroxybenzylidene)thiophene-2-carbohydrazide (21):**

Compound **21** was synthesized according to GP-8, starting with aldehyde **A25** (135.8 mg, 0.70 mmol) and hydrazide **H11** (100.0 mg, 0.70 mmol). The crude was purified by using purification

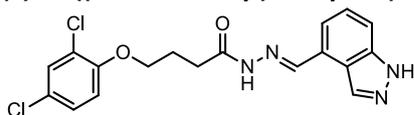
method-1 to afford **21** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 81:19) as a white solid (115.3mg, 0.36 mmol, 51.9%).

$^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  11.85 (s, 1H, *trans*), 11.47 (s, 1H, *cis*), 11.30 (s, 1H, *trans*), 9.72 (s, 1H, *cis*), 8.41 (s, 1H, *trans*), 8.26 (s, 1H, *cis*), 8.00 (d,  $J = 2.9$  Hz, 1H, *cis*), 7.92 (d,  $J = 5.0$  Hz, 1H, *cis*), 7.90 – 7.82 (m, 2H, 1*trans*, 1*trans*), 7.67 (d,  $J = 8.9$  Hz, 1H, *cis*), 7.29 – 7.13 (m, 3H, 1*trans*; 1*trans* and 1*cis*), 6.36 – 6.23 (m, 2H, 1*trans* and 1*cis*), 6.13 (d,  $J = 2.4$  Hz, 2H, 1*trans* and 1*cis*), 3.42 – 3.30 (m, 8H, 4*trans* and 4*cis*), 1.11 (t,  $J = 7.0$  Hz, 12H, 6*trans* and 6*cis*).

$^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  161.0, 160.1, 158.6, 157.4, 150.64, 150.29, 150.0, 142.6, 138.6, 134.8, 134.5, 134.1, 132.1, 131.9, 129.1, 128.6, 128.2, 127.0, 108.3, 106.9, 104.8, 104.1, 97.9, 97.6, 44.34, 44.28, 13.0.

HRMS (ESI $^+$ ):  $m/z$  calcd. For  $\text{C}_{16}\text{H}_{20}\text{N}_3\text{O}_2\text{S}$  ( $[\text{M}+\text{H}]^+$ ) 318.1276, measured 318.12665.

**(E)-N'-((1H-Indazol-4-yl)methylene)-4-(2,4-dichlorophenoxy)butanehydrazide (22):**



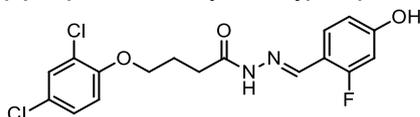
Compound **22** was synthesized according to GP-8, starting with aldehyde **A22** (6.8 mg, 0.046 mmol) and hydrazide **H12** (12.2 mg, 0.046 mmol). The crude was purified by using purification method-1 to afford **22** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 71:29) as a white solid (7.6 mg, 0.0019 mmol, 41.9%).

$^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  13.31 (s, 2H, 1*trans* and 1*cis*), 11.59 (s, 1H, *trans*), 11.43 (s, 1H, *cis*), 8.60 (d,  $J = 1.0$  Hz, 1H, *trans*), 8.49 (d,  $J = 1.0$  Hz, 1H, *cis*), 8.35 (s, 1H, *trans*), 8.21 (s, 1H, *cis*), 7.61 (d,  $J = 8.4$  Hz, 2H, 1*trans* and 1*cis*), 7.59 (d,  $J = 2.6$  Hz, 1H, *trans*), 7.56 (d,  $J = 2.6$  Hz, 1H, *cis*), 7.43 – 7.33 (m, 4H, 2*trans* and 2*cis*), 7.30 (d,  $J = 7.0$  Hz, 1H, *trans*), 7.25 (d,  $J = 7.0$  Hz, 1H, 1*cis*), 7.24 – 7.19 (m, 2H, 1*trans* and 1*cis*), 4.20 (t,  $J = 6.3$  Hz, 2H, 1*cis* and 1*trans*), 4.14 (t,  $J = 6.3$  Hz, 2H, 1*trans* and 1*cis*), 2.97 (t,  $J = 7.3$  Hz, 2H, 1*trans* and 1*cis*), 2.47 (t,  $J = 7.3$  Hz, 2H, 1*trans* and 1*cis*), 2.10 (dt,  $J = 12.2, 6.8$  Hz, 4H, 2*trans* and 2*cis*).

$^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  174.1, 168.6, 153.5, 146.1, 143.9, 140.7, 135.0, 134.4, 129.7, 128.6, 128.6, 128.0, 127.6, 126.4, 126.3, 124.8, 124.8, 123.7, 123.4, 123.0, 122.8, 119.0, 118.6, 115.6, 112.6, 68.9, 68.8, 30.8, 28.8, 24.7, 23.9.

HRMS (ESI $^+$ ):  $m/z$  calcd. For  $\text{C}_{18}\text{H}_{17}\text{Cl}_2\text{N}_4\text{O}_2$  ( $[\text{M}+\text{H}]^+$ ) 391.07286, measured 391.07179.

**(E)-4-(2,4-Dichlorophenoxy)-N'-(2-fluoro-4-hydroxybenzylidene)butanehydrazide (23):**



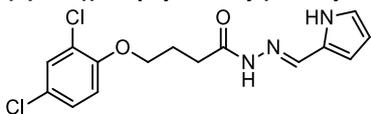
Compound **23** was synthesized according to GP-8, starting with aldehyde **A26** (35.0 mg, 0.25 mmol) and hydrazide **H12** (65.8 mg, 0.25 mmol). The crude was purified by using purification method-1 to afford **23** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 65:35) as a white solid (50.3 mg, 0.13 mmol, 55.2%).

$^1\text{H}$  NMR (500 MHz, DMSO- $d_6$ )  $\delta$  11.36 (s, 1H, *trans*), 11.24 (s, 1H, *cis*), 10.41 (br s, 2H, 1*trans* and 1*cis*), 8.26 (s, 1H, *trans*), 8.07 (s, 1H, *cis*), 7.67 (dt,  $J = 15.7, 8.7$  Hz, 2H, 1*trans* and 1*cis*), 7.57 (dd,  $J = 4.1, 2.6$  Hz, 2H, 1*trans* and 1*cis*), 7.36 (ddd,  $J = 8.8, 7.5, 2.6$  Hz, 2H, 1*trans* and 1*cis*), 7.19 (dd,  $J = 8.9, 1.8$  Hz, 2H, 1*trans* and 1*cis*), 6.70 (dd,  $J = 8.7, 2.3$  Hz, 1H, *trans*), 6.66 (dd,  $J = 8.7, 2.3$  Hz, 1H, *cis*), 6.60 (ddd,  $J = 12.7, 4.9, 2.3$  Hz, 2H, 1*trans* and 1*cis*), 4.12 (dt,  $J = 15.5, 6.3$  Hz, 4H, 2*trans* and 2*cis*), 2.80 (t,  $J = 7.4$  Hz, 2H, *cis*), 2.39 (t,  $J = 7.3$  Hz, 2H, *trans*), 2.11 – 1.96 (m, 4H, 2*trans* and 2*cis*).

$^{13}\text{C}$  NMR (126 MHz, DMSO- $d_6$ )  $\delta$  174.0, 168.2, 163.2, 163.0, 161.3, 161.2, 161.1, 161.1, 161.0, 153.4, 139.5, 139.5, 136.4, 136.4, 129.7, 128.6, 127.7, 127.6, 127.4, 127.3, 124.8, 124.7, 122.8, 115.5, 115.4, 113.2, 113.1, 103.06, 102.9, 68.8, 30.8, 28.8, 24.7, 24.2.

$^{19}\text{F}$  NMR (470 MHz, DMSO)  $\delta$  -119.91, -119.99.

HRMS (ESI $^+$ ):  $m/z$  calcd. for  $\text{C}_{17}\text{H}_{16}\text{Cl}_2\text{FN}_2\text{O}_3$  ( $[\text{M}+\text{H}]^+$ ) 385.05220, measured 385.05120.

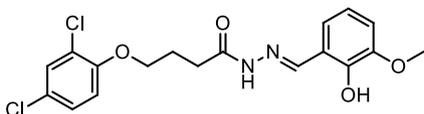
**(E)-N'-((1H-Indol-2-yl)methylene)-4-(2,4-dichlorophenoxy)butanehydrazide (24):**

Compound **24** was synthesized according to GP-8, starting with aldehyde **A20** (19.8 mg, 0.21 mmol) and hydrazide **H12** (54.8 mg, 0.21 mmol). The crude was purified by using purification method-1 to afford **24** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 53:47) as a white solid (52.5 mg, 0.15 mmol, 73.7%).

<sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>) δ 11.43 (s, 1H, *trans*), 11.24 (s, 1H, *cis*), 11.07 (s, 1H, *trans*), 10.97 (s, 1H, *cis*), 8.00 (s, 1H, *trans*), 7.82 (s, 1H, *cis*), 7.57 (dd, *J* = 4.0, 2.5 Hz, 2H, *1trans* and *1cis*), 7.36 (ddd, *J* = 9.5, 7.2, 2.6 Hz, 2H, *1trans* and *1cis*), 7.24 – 7.14 (m, 2H, *1trans* and *1cis*), 6.88 (dq, *J* = 9.8, 2.3 Hz, 2H, *1trans* and *1cis*), 6.47 – 6.35 (m, 2H, *1trans* and *1cis*), 6.18 – 6.03 (m, 2H, *1trans* and *1cis*), 4.14 (t, *J* = 6.5 Hz, 2H, *cis*), 4.10 (t, *J* = 6.2 Hz, 2H, *trans*), 2.78 (t, *J* = 7.3 Hz, 2H, *cis*), 2.38 (t, *J* = 7.4 Hz, 2H, *trans*), 2.09 – 1.97 (m, 4H, *2trans* and *2cis*).

<sup>13</sup>C NMR (126 MHz, DMSO-*d*<sub>6</sub>) δ 173.7, 167.8, 153.5, 139.6, 136.1, 129.7, 128.6, 127.7, 127.5, 124.9, 124.8, 122.9, 122.7, 122.1, 115.6, 115.5, 113.4, 112.3, 109.6, 69.0, 68.9, 30.8, 28.7, 24.9, 24.2.

HRMS (ESI<sup>+</sup>): *m/z* calcd. for C<sub>15</sub>H<sub>16</sub>Cl<sub>2</sub>N<sub>3</sub>O<sub>2</sub> ([M+H]<sup>+</sup>) 340.06196, measured 340.06112.

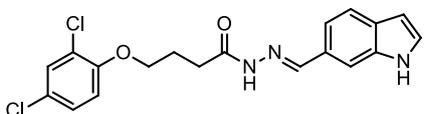
**(E)-4-(2,4-Dichlorophenoxy)-N'-(2-hydroxy-3-methoxybenzylidene)butanehydrazide (25):**

Compound **25** was synthesized according to GP-8, starting with aldehyde **A24** (38.0 mg, 0.25 mmol) and hydrazide **H12** (65.8 mg, 0.25 mmol). The crude was purified by using purification method-1 to afford **25** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 45:55) as a white solid (49.4 mg, 0.12 mmol, 49.7%).

<sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>) δ 11.64 (s, 1H, *trans*), 11.28 (s, 1H, *cis*), 10.90 (s, 1H, *trans*), 9.45 (s, 1H, *cis*), 8.35 (s, 1H, *trans*), 8.29 (s, 1H, *cis*), 7.57 (t, *J* = 2.8 Hz, 2H, *1trans* and *1cis*), 7.36 (td, *J* = 8.6, 2.6 Hz, 2H, *1trans* and *1cis*), 7.22 (dd, *J* = 8.0, 1.4 Hz, 1H, *cis*), 7.20 (d, *J* = 8.9 Hz, 2H, *1trans* and *1cis*), 7.10 (dd, *J* = 7.9, 1.5 Hz, 1H, *trans*), 7.01 (dd, *J* = 8.1, 1.4 Hz, 1H, *trans*), 6.98 (dd, *J* = 8.1, 1.5 Hz, 1H, *cis*), 6.84 (t, *J* = 7.9 Hz, 1H, *trans*), 6.79 (t, *J* = 8.0 Hz, 1H, *cis*), 4.13 (dt, *J* = 12.2, 6.3 Hz, 4H, *2trans* and *2cis*), 3.81 (d, *J* = 4.9 Hz, 6H, *3trans* and *3cis*), 2.79 (t, *J* = 7.4 Hz, 2H, *cis*), 2.44 (t, *J* = 7.3 Hz, 2H, *trans*), 2.16 – 1.95 (m, 4H, *2trans* and *2cis*).

<sup>13</sup>C NMR (126 MHz, DMSO-*d*<sub>6</sub>) δ 173.9, 168.3, 153.44, 148.4, 148.3, 147.4, 146.8, 146.3, 140.9, 129.7, 128.6, 124.8, 124.75, 122.8, 121.2, 121.0, 119.6, 119.4, 119.3, 118.2, 115.5, 115.5, 114.0, 113.2, 68.8, 56.3, 56.2, 30.7, 28.8, 24.6, 24.1.

HRMS (ESI<sup>+</sup>): *m/z* calcd. for C<sub>18</sub>H<sub>19</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>4</sub> ([M+H]<sup>+</sup>) 397.07219, measured 397.07097.

**(E)-N'-((1H-Indol-6-yl)methylene)-4-(2,4-dichlorophenoxy)butanehydrazide (26):**

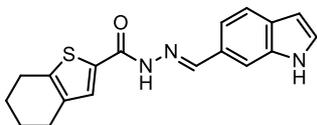
Compound **26** was synthesized according to GP-8, starting with aldehyde **A23** (36.3 mg, 0.25 mmol) and hydrazide **H12** (65.8 mg, 0.25 mmol). The crude was purified by using purification method-1 to afford **26** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 62:38) as a white solid (40.3 mg, 0.10 mmol, 41.3%).

<sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>) δ 11.29 (br s, 1H, *trans*), 11.27 (s, 1H, *trans*), 11.25 (br s, 1H, *cis*), 11.15 (s, 1H, *cis*), 8.22 (s, 1H, *trans*), 8.06 (s, 1H, *cis*), 7.69 (s, 1H, *trans*), 7.61 (s, 1H, *cis*), 7.59 – 7.51 (m, 4H, *2trans* and *2cis*), 7.46 (t, *J* = 2.8 Hz, 1H, *trans*), 7.44 (t, *J* = 2.7 Hz, 1H, *cis*), 7.41 – 7.32 (m, 4H, *2trans* and *2cis*), 7.20 (dd, *J* = 9.0, 2.3 Hz, 2H, *1trans* and *1cis*), 6.46 (q, *J* = 2.4 Hz, 2H, *1trans* and *1cis*), 4.16 (t, *J* = 6.4 Hz, 2H, *cis*), 4.12 (t, *J* = 6.3 Hz, 2H, *trans*), 2.85 (t, *J* = 7.4 Hz, 2H, *cis*), 2.42 (t, *J* = 7.3 Hz, 2H, *trans*), 2.14 – 2.00 (m, 4H, *2trans* and *2cis*).

<sup>13</sup>C NMR (126 MHz, DMSO-*d*<sub>6</sub>) δ 173.9, 168.1, 153.5, 148.0, 144.9, 136.3, 136.22, 129.7, 129.5, 128.6, 127.9, 127.8, 124.8, 124.8, 122.9, 120.8, 120.7, 118.4, 117.6, 115.5, 115.47, 111.49, 111.28, 102.0, 102.0, 68.8, 30.8, 28.8, 24.8, 24.1.

HRMS (ESI<sup>+</sup>): m/z calcd. for C<sub>19</sub>H<sub>18</sub>Cl<sub>2</sub>N<sub>3</sub>O<sub>2</sub> ([M+H]<sup>+</sup>) 390.07761, measured 390.07634.

**(E)-N'-((1H-Indol-6-yl)methylene)-4,5,6,7-tetrahydrobenzo[b]thiophene-2-carbohydrazide (35):**



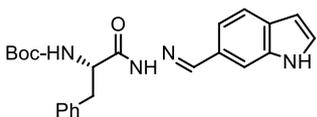
Compound **35** was synthesized according to GP-8, starting with aldehyde **A23** (37.0 mg, 0.25 mmol) and hydrazide **H1** (50 mg, 0.25 mmol). The crude was purified by using purification method-1 to afford **35** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* s = 49:51) as a white solid (76.7 mg, 0.24 mmol, 94.9%).

<sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>) δ 11.59 (d, *J* = 32.2 Hz, 2H, 1*trans* and 1*cis*), 11.35 (d, *J* = 24.0 Hz, 2H, 1*trans* and 1*cis*), 8.46 (s, 1H, *trans*), 8.15 (s, 1H, *cis*), 7.74 (br s, 3H, 1*trans*; 1*trans* and 1*cis*), 7.68 – 7.55 (m, 3H, 1*cis*; 1*trans* and 1*cis*), 7.51 (d, *J* = 8.3 Hz, 1H, *trans*), 7.48 (t, *J* = 2.7 Hz, 2H, 1*trans* and 1*cis*), 7.40 (d, *J* = 8.2 Hz, 1H, *cis*), 6.48 (br s, 2H, 1*trans* and 1*cis*), 2.80 (dt, *J* = 31.7, 6.0 Hz, 4H, 2*trans* and 2*cis*), 2.62 (q, *J* = 6.5, 5.9 Hz, 4H, 2*trans* and 2*cis*), 1.97 – 1.56 (m, 8H, 4*trans* and 4*cis*).

<sup>13</sup>C NMR (126 MHz, DMSO-*d*<sub>6</sub>) δ 161.7, 158.4, 149.2, 146.0, 145.1, 141.8, 136.48, 136.3, 135.6, 135.0, 134.7, 129.8, 129.7, 129.5, 128.1, 127.8, 127.7, 121.0, 120.8, 118.6, 118.2, 112.0, 111.3, 102.1, 102.0, 25.4, 25.2, 25.1, 23.3, 23.3, 22.8, 22.6.

HRMS (ESI<sup>+</sup>): m/z calcd. for C<sub>18</sub>H<sub>18</sub>N<sub>3</sub>OS ([M+H]<sup>+</sup>) 324.11706, measured 324.11588.

**(S,E)-tert-butyl (1-(2-((1H-indol-6-yl)methylene)hydrazinyl)-1-oxo-3-phenylpropan-2-yl)carbamate (37):**



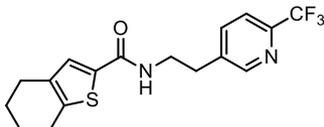
Compound **37** was synthesized according to GP-8, starting with aldehyde **A23** (30.19 mg, 0.21 mmol) and hydrazide **H14** (58.1 mg, 0.21 mmol). The crude was purified by using purification method-1 to afford **37** as a mixture of *cis* and *trans* conformers of amide (*cis:trans* = 53:47) as a white solid (34.1 mg, 0.08 mmol, 39.9%).

<sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>) δ 11.43 (s, 1H, *trans*), 11.37 (s, 1H, *cis*), 11.34 (s, 1H, *trans*), 11.25 (s, 1H, *cis*), 8.25 (s, 1H, *trans*), 8.09 (s, 1H, *cis*), 7.71 (s, 2H, 1*trans* and 1*cis*), 7.60 (dd, *J* = 16.9, 8.3 Hz, 2H, 1*trans* and 1*cis*), 7.48 (q, *J* = 2.5 Hz, 2H, 1*trans* and 1*cis*), 7.44 (dd, *J* = 8.2, 1.4 Hz, 1H, *trans*), 7.39 (dd, *J* = 8.3, 1.4 Hz, 1H, *cis*), 7.37 – 7.26 (m, 8H, 4*trans* and 4*cis*), 7.26 – 7.16 (m, 3H, 1*trans*; 1*trans* and 1*cis*), 7.10 (d, *J* = 8.8 Hz, 1H, *cis*), 6.48 (dt, *J* = 7.6, 2.4 Hz, 2H, 1*trans* and 1*cis*), 5.11 (ddd, *J* = 10.5, 8.8, 3.6 Hz, 1H, *cis*), 4.22 (ddd, *J* = 10.1, 8.3, 4.8 Hz, 1H, *trans*), 3.09 – 2.91 (m, 2H, 1*trans* and 1*cis*), 2.90 – 2.71 (m, 2H, 1*trans* and 1*cis*), 1.32 (d, *J* = 4.7 Hz, 18H, 9*trans* and 9*cis*).

<sup>13</sup>C NMR (126 MHz, DMSO-*d*<sub>6</sub>) δ 173.4, 168.6, 156.0, 155.9, 149.0, 145.7, 139.1, 138.5, 136.3, 129.8, 129.7, 129.7, 129.7, 128.7, 128.6, 128.0, 128.0, 127.7, 127.6, 126.8, 126.7, 120.9, 120.8, 118.4, 117.9, 111.5, 111.4, 102.1, 102.0, 78.5, 78.4, 55.6, 53.6, 37.8, 36.9, 28.6, 28.3.

HRMS (ESI<sup>+</sup>): m/z calcd. for C<sub>23</sub>H<sub>27</sub>N<sub>4</sub>O<sub>3</sub> ([M+H]<sup>+</sup>) 407.20832, measured 407.20696.

**N-(2-(6-(Trifluoromethyl)pyridin-3-yl)ethyl)-4,5,6,7-tetrahydrobenzo[b]thiophene-2-carboxamide (38):**



Compound **38** was synthesized according to GP-9. The first step product (*E*)-5-(2-nitrovinyl)-2-(trifluoromethyl)pyridine was prepared by starting with 6-(trifluoromethyl)nicotinaldehyde (175.1 mg, 1 mmol, 1 equiv.), nitromethane (61.1 mg, 1 mmol, 1 equiv.), and MeOH (2 mL) followed by treating with HCl (0.6 mL in 1 mL of water). After workup, the obtained crude was treated with LiAlH<sub>4</sub> (151.8 mg, 4 mmol) in dry-THF (16 mL) to afford the second step product 2-(6-(trifluoromethyl)pyridin-3-yl)ethanamine. This crude (127.5 mg, approx. 0.67 mmol, 1 equiv.) was reacted with 4,5,6,7-tetrahydrobenzo[b]thiophene-2-carboxylic acid (133.8 mg, 0.74 mmol, 1.1 equiv.), HOBT·H<sub>2</sub>O (204.7 mg, 1.34 mmol, 2 equiv.), EDC·HCl (256.27 mg, 1.34 mmol, 2 equiv.), and DIPEA (279.4 μL, 1.67 mmol, 2.5 equiv.) in

17

DMF (2 mL). The crude obtained after the workup was purified by preparative HPLC (H<sub>2</sub>O:MeCN + 0.1% formic acid, gradient 5% to 100% MeCN) to afford **38** as a pale yellow solid (75 mg, 0.21 mmol, 21%).

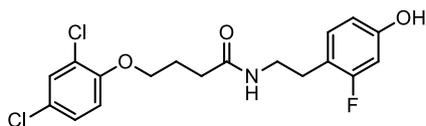
<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 8.51 (d, *J* = 2.3 Hz, 1H), 7.67 (dd, *J* = 8.1, 2.2 Hz, 1H), 7.57 (d, *J* = 8.0 Hz, 1H), 7.07 (s, 1H), 5.97 (bs, 1H), 3.60 (q, *J* = 6.7 Hz, 2H), 2.95 (t, *J* = 7.0 Hz, 2H), 2.69 (td, *J* = 5.9, 1.8 Hz, 2H), 2.54 – 2.47 (m, 2H), 1.82 – 1.67 (m, 4H).

<sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>) δ 162.5, 150.4, 146.6 (q, *J* = 34.8 Hz), 141.7, 138.0, 137.7, 136.4, 133.6, 129.4, 121.6 (q, *J* = 273.6 Hz), 120.4, 40.6, 32.9, 25.4, 25.3, 23.2, 22.6.

<sup>19</sup>F NMR (470 MHz, CDCl<sub>3</sub>) δ -67.73.

HRMS (ESI<sup>+</sup>): *m/z* calcd. for C<sub>17</sub>H<sub>18</sub>F<sub>3</sub>N<sub>2</sub>OS ([M+H]<sup>+</sup>) 355.10919, measured 355.10776

#### 4-(2,4-Dichlorophenoxy)-N-(2-fluoro-4-hydroxyphenethyl)butanamide (**39**):



Compound **39** was synthesized according to GP-9. The first step product (*E*)-3-fluoro-4-(2-nitrovinyl)phenol was prepared by starting with 2-fluoro-4-hydroxybenzaldehyde (140.1 mg, 1 mmol, 1 equiv.), nitromethane (61.1 mg, 1 mmol, 1 equiv.), and MeOH (2 mL) followed by treating with HCl (0.6 mL in 1 mL of water). After workup, the obtained crude was treated with LiAlH<sub>4</sub> (151.8 mg, 4 mmol) in dry-THF (16 mL) to afford the second step product 4-(2-aminoethyl)-3-fluorophenol. This crude (78 mg, approx. 0.5 mmol, 1equiv.) was reacted with 4-(2,4-dichlorophenoxy)butanoic acid (73 mg, 0.55 mmol, 1.1 equiv.), HOBt.H<sub>2</sub>O (168.5 mg, 1.1 mmol, 2 equiv.), EDC·HCl (210.9 mg, 1.1 mmol, 2 equiv.), and DIPEA (239 μL, 1.37 mmol, 2.5 equiv.) in DMF (2 mL). The crude obtained after the workup was purified by preparative HPLC (H<sub>2</sub>O:MeCN + 0.1% formic acid, gradient 5% to 100% MeCN) to afford **39** as a pale yellow solid (35 mg, 0.09 mmol, 9%).

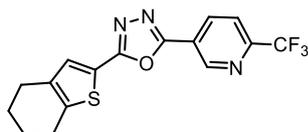
<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 7.26 (d, *J* = 2.5 Hz, 1H), 7.08 (dd, *J* = 8.8, 2.5 Hz, 1H), 6.85 (t, *J* = 8.6 Hz, 1H), 6.71 (d, *J* = 8.8 Hz, 1H), 6.50 – 6.42 (m, 2H), 5.83 (bt, *J* = 5.8 Hz, 1H), 3.90 (t, *J* = 5.8 Hz, 2H), 3.41 (q, *J* = 6.6, 5.9 Hz, 2H), 2.65 (t, *J* = 6.8 Hz, 2H), 2.35 (t, *J* = 7.1 Hz, 2H), 2.04 (p, *J* = 6.7 Hz, 2H).

<sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>) δ 173.2, 162.6, 160.6, 156.8, 156.7, 153.0, 131.2, 131.2, 129.9, 127.7, 125.8, 123.5, 116.3, 116.2, 114.1, 111.4, 111.4, 103.3, 103.1, 68.0, 40.0, 32.8, 28.5, 24.9.

<sup>19</sup>F NMR (470 MHz, CDCl<sub>3</sub>) δ -116.90.

HRMS (ESI<sup>+</sup>): *m/z* calcd. for Chemical Formula: C<sub>18</sub>H<sub>19</sub>Cl<sub>2</sub>FNO<sub>3</sub> ([M+H]<sup>+</sup>) 386.07260, measured 386.07151

#### 2-(4,5,6,7-tetrahydrobenzo[*b*]thiophen-2-yl)-5-(6-(trifluoromethyl)pyridin-3-yl)-1,3,4-oxadiazole (**40**):



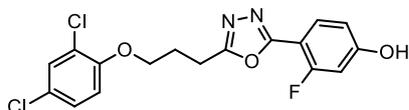
Compound **40** was synthesized according to GP-10, starting with 6-(trifluoromethyl)nicotinic acid (101.5 mg, 0.53 mmol, 1 equiv.), 4,5,6,7-tetrahydrobenzo[*b*]thiophene-2-carbohydrazide (104.1 mg, 0.53 mmol, 1 equiv.), aluminum oxide (200 mg), and phosphorous oxytrichloride (1.14 g, 7.44 mmol, 14 equiv.). The crude purified by washing with ice-cold distilled water and ice-cold concentrated sodium bicarbonate solution. The residue was further washed with cold MeOH to provide the **40** as a white solid (115.5 mg, 0.33 mmol, 62%).

<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 9.34 (d, *J* = 1.8 Hz, 1H), 8.51 (dd, *J* = 8.2, 1.7 Hz, 1H), 7.80 (d, *J* = 8.2 Hz, 1H), 7.49 (s, 1H), 2.79 (t, *J* = 6.1 Hz, 2H), 2.63 (t, *J* = 6.2 Hz, 2H), 1.81-1.70 (m, 4H).

<sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>) δ 162.2, 160.5, 150.1 (q, *J* = 35.7 Hz), 147.8, 143.2, 137.2, 135.5, 131.49, 123.0, 120.8, 121.1 (q, *J* = 270.1 Hz) 120.8, 120.0, 25.4, 25.4, 23.1, 22.5.

<sup>19</sup>F NMR (470 MHz, CDCl<sub>3</sub>) δ -68.1.

HRMS (ESI<sup>+</sup>): *m/z* calcd. for Chemical Formula: C<sub>16</sub>H<sub>13</sub>F<sub>3</sub>N<sub>3</sub>OS ([M+H]<sup>+</sup>) 352.07314, measured 352.07090

**4-(5-(3-(2,4-dichlorophenoxy)propyl)-1,3,4-oxadiazol-2-yl)-3-fluorophenol (41):**

Compound **41** was synthesized according to GP-10, starting with 2-fluoro-4-hydroxybenzoic acid (32.4 mg, 0.19 mmol, 1 mmol), 4-(2,4-dichlorophenoxy)butanehydrazide (32.4 mg, 0.19 mmol, 1 equiv.), aluminum oxide (121 mg), and phosphorous oxytrichloride (408 mg, 2.66 mmol, 14 equiv.). The crude purified by washing with ice-cold distilled water and ice-cold concentrated sodium bicarbonate solution. The residue was further washed with cold MeOH and the crude was purified by preparative HPLC (H<sub>2</sub>O:MeCN + 0.1% formic acid, gradient 20% to 100% MeCN) to afford **41** as a white solid (9.1 mg, 0.024 mmol, 12.5%).

<sup>1</sup>H-NMR (500 MHz, MeOD) δ 7.76 (t, 1H, *J* = 8.4 Hz), 7.34 (d, 1H, *J* = 2.6 Hz), 7.24 (dd, 1H, *J* = 8.7 Hz, 2.4 Hz), 7.06 (d, 1H, *J* = 8.9 Hz), 6.75 (dd, 1H, *J* = 9.0 Hz, 2.2 Hz), 6.69 (dd, 1H, *J* = 12.5 Hz, 2.4 Hz), 4.20 (t, 2H, *J* = 5.7 Hz), 3.20 (t, 2H, *J* = 7.3 Hz), 2.38 (m, 2H).

<sup>13</sup>C-NMR (500 MHz, MeOD) δ 167.8, 164.3, 164.2, 163.8, 163.4, 161.8, 154.6, 131.6, 130.7, 123.0, 126.8, 124.6, 115.5, 113.5, 104.8, 104.6, 104.1, 104.0, 69.2, 27.0, 23.2.

<sup>19</sup>F NMR (470 MHz, MeOD) δ -108.62.

HRMS (ESI<sup>+</sup>): *m/z* calcd. for Chemical Formula: C<sub>17</sub>H<sub>14</sub>Cl<sub>2</sub>FN<sub>2</sub>O<sub>3</sub> ([M+H]<sup>+</sup>) 383.03655, measured 383.03428

**Determination of IC<sub>50</sub> values**

The inhibition of the DXPS enzyme was analyzed in a previously reported photometric assay using the NADPH consumption of the downstream enzyme IspC as readout with minor modifications.<sup>9,10</sup> The assay was conducted in transparent, flat bottom 384 well-plates (Greiner BioOne) and the absorbance measured using a microplate reader (PHERAstar, BMG Labtech). The assay mixture contained 100 mM Tris-HCl with a pH of 7.6, 2 mM dithiothreitol (DTT), 5 mM MnCl<sub>2</sub>, 0.5 mM NADPH, and 1.5 μM IspC (*E. coli*, expressed and purified in-house according to literature report).<sup>9</sup> The amount of DXPS used in the assays was determined experimentally by a dilution series of the enzyme. The concentration, which showed a reaction velocity in the range of -0.1 to -0.2 OD over a time range of 10 mins was chosen for further experiments. For compounds that were known aggregators, 0.01% v/v of Triton X-100 detergent was added to the assay.

The assay was prepared using two buffers, buffer A containing Tris-HCl and most of the components of the reaction in 2x the final concentration and buffer B, consisting of only 100 mM Tris-HCl, pH 7.6 and the substrates, 0.5 mM pyruvate and 0.5 mM D/L glyceraldehyde 3-phosphate. After the addition of 60 μL of buffer A to the plate with a pre-made inhibitor dilution series in 6 μL DMSO the reaction was started by the addition of 60 μL buffer B. The plate was centrifuged for 1 min at 2000 rpm and 25 °C to remove possible air bubbles. The plate was then immediately supplied to the microplate reader and the absorbance measured at 340 nm using the mode slow kinetics with a cycle time of 30 sec and 60 cycles at RT.

Blank correction and linear fitting of the raw data was performed using the program Origin2019. The obtained initial velocities were converted to percent inhibition and plotted against the inhibitor concentration. The IC<sub>50</sub> values were determined by nonlinear curve fitting using Origin2019.

**Determination of antibacterial activity**

Experiments to determine the minimum inhibitory concentration (MIC) have been carried out as described recently.<sup>10</sup> Briefly, our studies were performed with the efflux-pump deficient *E. coli* knockout strain ΔtolC. In case MIC values could not be measured for activity reasons, the percentage (%) inhibition was determined at 100 μM (or lower, depending on the solubility of the compounds).

## DLS measurements

Compounds were analyzed for aggregate formation using dynamic light scattering (DLS) on a Zetasizer Nano ZS90 (Malvern Instruments Ltd, Worcestershire, UK). Disposable cuvettes (3.2 mL, 67.758, Sarstedt AG & Co, Nümbrecht, Germany) were filled with a total sample volume of 50  $\mu\text{L}$  and equilibrated to 25 °C for 60 sec prior to the experiments. Analyses were performed in the absence or presence of compounds using 7H9 medium supplemented with 10% OADC and 1% DMSO. Count rate data for all samples were determined from three measurements of at least three cycles of 10 sec each. The count rate at high scattering intensities is reduced by an automatic attenuator, so the derived count rate, which is corrected with the attenuation factor, was used as a measure for the degree of aggregate formation. Derived count rate values of the compounds were compared with the derived count rate of the medium at the concentrations as indicated. This allowed the determination of the highest sample concentration at which the respective compound was not yet present with an increased degree of aggregate formation.

## *Mycobacterium tuberculosis* growth analysis in liquid culture

### Bacteria

*M. tuberculosis* (Mtb) strain H37Rv (ATCC 25618) carrying a mCherry-expressing plasmid (pCherry10)<sup>11</sup> was cultured in 7H9 complete medium (BD Difco; Becton Dickinson) supplemented with oleic acid-albumin-dextrose-catalase (OADC, 10%; BD), 0.2% glycerol, and 0.05% Tween80 as previously described.<sup>12</sup> At mid-log phase ( $\text{OD}_{600} = 0.4$ ) cultures were harvested and frozen in aliquots at -80 °C.

Frozen aliquots of mCherry-Mtb H37Rv were thawed and centrifuged ( $3700 \times g$ , 10 min). Supernatants were discarded and bacteria were thoroughly resuspended in 7H9 medium (10% OADC) in the absence of glycerol and Tween80 by use of a syringe and a 26-gauge syringe needle. The bacterial suspension was passed in and out of the syringe about 10 times. Non-precipitating compounds - identified and determined by DLS - were tested in triplicates ( $2 \times 10^5$  bacteria, volume 100  $\mu\text{L}$ ) for their anti-tubercular activity in 2-fold serial dilutions starting from 64  $\mu\text{M}$  using 96-well flat clear bottom black polystyrene microplates (Corning® CellBIND®, New York, USA). Each plate was prepared with rifampicin (National Reference Center, Borstel) as a reference compound. Plates were sealed with an air-permeable membrane (Porvair Sciences, Wrexham, UK) in a 37 °C incubator with mild agitation (TiMix5, Edmund Bühler, Germany), as previously described.<sup>13</sup> Bacterial growth was measured as relative light units (RLU) from the fluorescence intensity obtained at an excitation wavelength of 575 nm and an emission wavelength of 635 nm (microplate reader, Synergy 2, BioTek Instruments, Vermont, USA) after 9 days (Figure 9, in the main text) and 10 days of culture (Supplementary Figures 164 and 165). Obtained values were normalized to RLU values of the solvent control (1% DMSO-treated bacteria set to 100%).

## Binding affinity ( $K_D$ ) determinations by surface plasmon resonance (SPR)

The SPR binding studies were performed using a Reichert SR7500DC surface plasmon resonance spectrometer (Reichert Technologies, Depew, NY, USA), and medium density carboxymethyl dextran hydrogel CMD500M sensor chips (XanTec Bioanalytics, Düsseldorf, Germany). Milli-Q water was used as the running buffer for immobilization. The *Deinococcus radiodurans* DXPS (drDXPS, 69.96 kDa) or truncated *Mycobacterium tuberculosis* DXPS ( $\Delta\text{mtDXPS}$ , 66.61 kDa) was immobilized in one of the two flow cells according to reported amine-coupling protocols.<sup>14-16</sup> The other flow cell was left blank to serve as a reference. The system was initially primed with borate buffer 100 mM (pH 9.0), then the carboxymethyl dextran matrix was activated by a 1:1 mixture of *N*-ethyl-*N'*-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC)

100 mM and *N*-hydroxysuccinimide (NHS) 100 mM at a flow rate of 10  $\mu\text{L}/\text{min}$  for 7 min. The drDXPS or  $\Delta\text{mtDXPS}$  was diluted to a final concentration of 70  $\mu\text{g}/\text{mL}$  in 10 mM sodium acetate buffer (pH 5.5 or 5.0), respectively, and was injected at a flow rate of 5  $\mu\text{L}/\text{min}$  for 10 min. The non-reacted surface was quenched by 1 M ethanolamine hydrochloride (pH 8.5) at a flow rate of 25  $\mu\text{L}/\text{min}$  for 3 min. A series of 10 buffer injections was run initially on both reference and active surfaces to equilibrate the system resulting in a stable immobilization level of approximately 5000 (13000 for  $\Delta\text{mtDXPS}$ )  $\mu$  refractive index unit ( $\mu\text{RIU}$ ) (Supplementary Figures 109 and 127), respectively. Phosphate buffered saline (PBS) buffer (10 mM  $\text{Na}_2\text{HPO}_4$ , 1.8 mM  $\text{KH}_2\text{PO}_4$ , 137 mM NaCl, 2.7 mM KCl, 0.05% v/v Tween 20, pH 7.4) or HEPES buffer (50 mM HEPES, 150 mM NaCl, 0.05% v/v Tween 20, 1 mM  $\text{MgCl}_2$ , pH 8.0) containing 5% v/v DMSO was used as the running buffer for binding studies with drDXPS and  $\Delta\text{mtDXPS}$ , respectively. All running buffers were filtered and degassed prior to use. Binding experiments were performed at 20  $^\circ\text{C}$ . Compounds dissolved in DMSO were diluted with the running buffer (final DMSO concentration of 5% v/v) and were injected at a flow rate of 30  $\mu\text{L}/\text{min}$ . Single-cycle kinetics were applied for  $K_D$  determination. The association time was set to 60 s, and the dissociation phase was recorded for 120 sec. Ethylene glycol 80% in the running buffer was used for regeneration of the surface. Differences in the bulk refractive index due to DMSO were corrected by a calibration curve (nine concentrations: 3–7% v/v DMSO in the running buffer). Data processing and analysis were performed by Scrubber software (Version 2.0c, 2008, BioLogic Software). Sensorgrams were calculated by sequential subtractions of the corresponding curves obtained from the reference flow cell and the running buffer (blank). SPR responses are expressed in the resonance unit (RU). The  $K_D$  values were calculated by the fitting of the steady-state binding responses to a 1:1 Langmuir interaction model (Hill equation was used for compound **3**).

## Mode-of-inhibition (MOI) study

MOI study followed a previously reported protocol and was performed against *D. radiodurans* DXPS.<sup>17</sup> The experiments were performed in duplicates.

Inhibitor competition with ThDP: The Michaelis–Menten constant  $K_m^{\text{ThDP}}$  was determined with an assay mixture containing 100 mM Tris-HCl (pH 7.6), 2.0 mM  $\text{MgCl}_2$ , 2.0 mM DTT, 0.5 mM NADPH, 0.5 mM sodium pyruvate, 0.5 mM D-GAP, 8.3  $\mu\text{M}$  *E. coli* IspC, 0.4  $\mu\text{M}$  *D. radiodurans* DXPS and 5% DMSO. Buffer A contained 100 mM Tris-HCl (pH 7.6), 4.0 mM  $\text{MgCl}_2$ , 4.0 mM DTT, 1.0 mM NADPH, 16.6  $\mu\text{M}$  *E. coli* IspC, 0.8  $\mu\text{M}$  *D. radiodurans* DXPS. Buffer B contained 100 mM Tris-HCl (pH 7.6), 1.0 mM sodium pyruvate, and 1.0 mM D-GAP. ThDP was titrated in the 0.1–200  $\mu\text{M}$  range, and  $K_m^{\text{ThDP}}$  was calculated using the program Dynafit.<sup>18</sup> Four ThDP concentrations were taken for  $\text{IC}_{50}$  determination, spanning from 1x  $K_m^{\text{ThDP}}$  to 16x  $K_m^{\text{ThDP}}$ . Assay mixtures contained ThDP concentrations 0.3  $\mu\text{M}$ , 1  $\mu\text{M}$ , 2  $\mu\text{M}$ , and 5  $\mu\text{M}$ , respectively. Buffer A1–A4 contained ThDP concentrations 0.6  $\mu\text{M}$ , 2  $\mu\text{M}$ , 4  $\mu\text{M}$ , and 10  $\mu\text{M}$  respectively. The rest of the procedure followed the protocol of  $\text{IC}_{50}$  determination.

Inhibitor competition with pyruvate: The Michaelis–Menten constant  $K_m^{\text{pyruvate}}$  was determined with assay mixture containing 100 mM Tris-HCl (pH 7.6), 2.0 mM  $\text{MgCl}_2$ , 2.0 mM DTT, 0.5 mM NADPH, 100  $\mu\text{M}$  ThDP, 0.5 mM D-GAP, 8.3  $\mu\text{M}$  *E. coli* IspC, 0.4  $\mu\text{M}$  *D. radiodurans* DXPS and 5% DMSO. Buffer A contained 100 mM Tris-HCl (pH 7.6), 4.0 mM  $\text{MgCl}_2$ , 4.0 mM DTT, 1.0 mM NADPH, 200  $\mu\text{M}$  ThDP, 16.6  $\mu\text{M}$  *E. coli* IspC, 0.8  $\mu\text{M}$  *D. radiodurans* DXPS. Buffer B contained 100 mM Tris-HCl (pH 7.6) and 1.0 mM D-GAP. Pyruvate was titrated in 1–500  $\mu\text{M}$  range, and  $K_m^{\text{pyruvate}}$  was calculated using the program Dynafit. Four pyruvate concentrations were taken for  $\text{IC}_{50}$  determination, spanning from 1x  $K_m^{\text{pyruvate}}$  to 5x  $K_m^{\text{pyruvate}}$ . Assay mixtures contained pyruvate concentrations 0.1 mM, 0.2 mM, 0.4 mM, and 0.6 mM respectively. Buffer B1–B4 contained pyruvate concentrations 0.2 mM, 0.4 mM, 0.8 mM, and 1.2 mM, respectively. The rest of the procedure followed the protocol of  $\text{IC}_{50}$  determination.

Inhibitor competition with D-GAP: The Michaelis–Menten constant  $K_m^{\text{D-GAP}}$  was determined with assay mixture containing 100 mM Tris-HCl (pH 7.6), 2.0 mM  $\text{MgCl}_2$ , 2.0 mM DTT, 0.5 mM NADPH, 100  $\mu\text{M}$  ThDP, 0.2 mM sodium pyruvate, 8.3  $\mu\text{M}$  *E. coli* IspC, 0.4  $\mu\text{M}$  *D. radiodurans* DXPS and 5% DMSO. Buffer A contained 100 mM Tris-HCl (pH 7.6), 4.0 mM  $\text{MgCl}_2$ , 4.0 mM DTT, 1.0 mM NADPH, 200  $\mu\text{M}$  ThDP, 16.6  $\mu\text{M}$  *E. coli* IspC, 0.8  $\mu\text{M}$  *D. radiodurans* DXPS. Buffer B contained 100 mM Tris-HCl (pH 7.6) and 0.4 mM sodium pyruvate. D-GAP

was titrated in 0.1–0.8 mM range, and  $K_m^{D-GAP}$  was calculated using the program Dynafit. Four D-GAP concentrations were taken for  $IC_{50}$  determination, spanning from  $1 \times K_m^{D-GAP}$  to  $7 \times K_m^{D-GAP}$ . Assay mixtures contained D-GAP concentration 0.1 mM, 0.2 mM, 0.4 mM, and 0.8 mM, respectively. Buffer B1–B4 contained D-GAP concentration 0.2 mM, 0.4 mM, 0.8 mM, and 1.6 mM, respectively, without significant substrate inhibition. The rest of the procedure followed the protocol of  $IC_{50}$  determination.

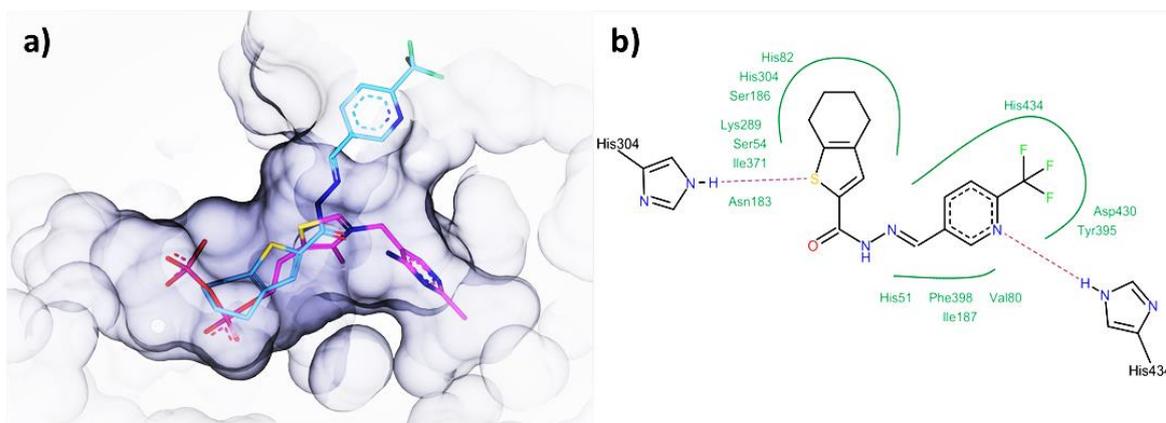
### Protein stability study by thermal shift assay

Analysis was performed using an ABI StepOneplus real-time PCR instrument. The protein samples were measured in white 96-well plates and denatured using a continuous heating rate of 0.5°C/min from 25°C up to 95 °C. The total sample volume was 25 µL of which 20 µL were buffer (100 mM phosphate buffer of pH ranges 5.8, 6.2, 7.0, and 7.5), 2.5 µL protein (2 µg), and 2.5 µL dye (Sypro Orange 50x). The melting curves were analyzed using Protein Thermal Shift 1.3 software. The most stabilizing buffer pH for drDXS, which showed the highest  $T_m$  was then used in our following experiments.

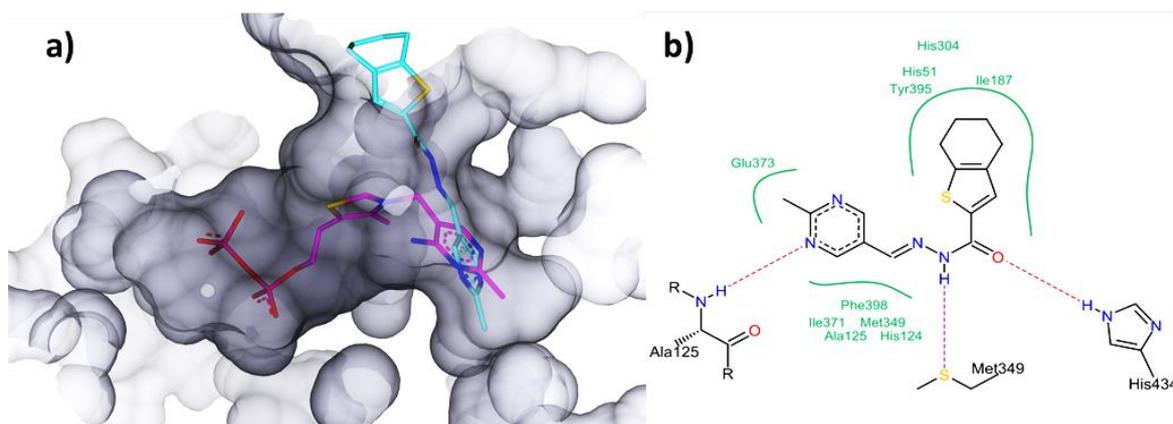
Time/hr	Melting temperature of drDXS (°C)			
	pH 7.5	pH 7.0	pH 6.2	pH 5.8
<b>0</b>	51.2	51.8	48.4	41.4
<b>3</b>	51.5	51.8	48.4	42.3
<b>6</b>	51.7	51.8	48.4	45.2
<b>24</b>	51.5	51.7	48.4	44.9

**Supplementary Table 1** Melting temperature of drDXS protein at different pH over time

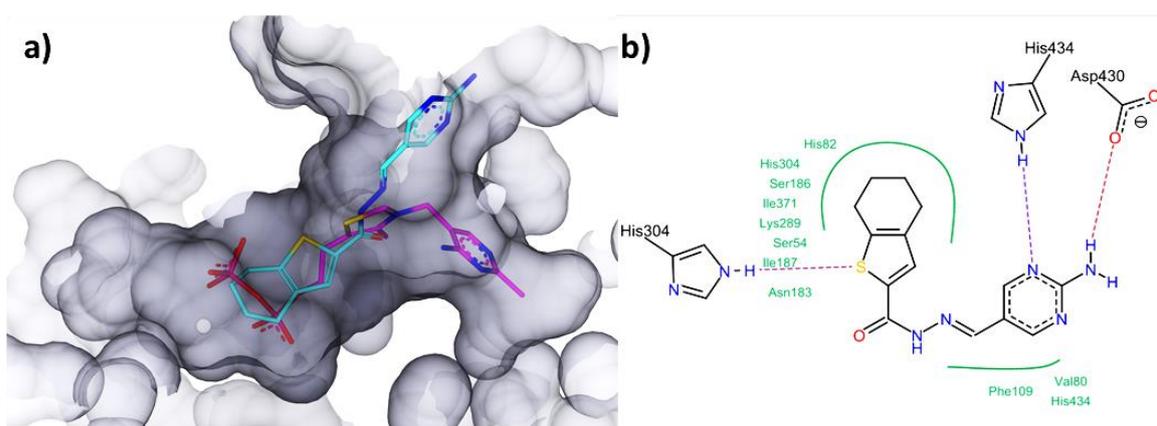
## SUPPLEMENTARY FIGURES



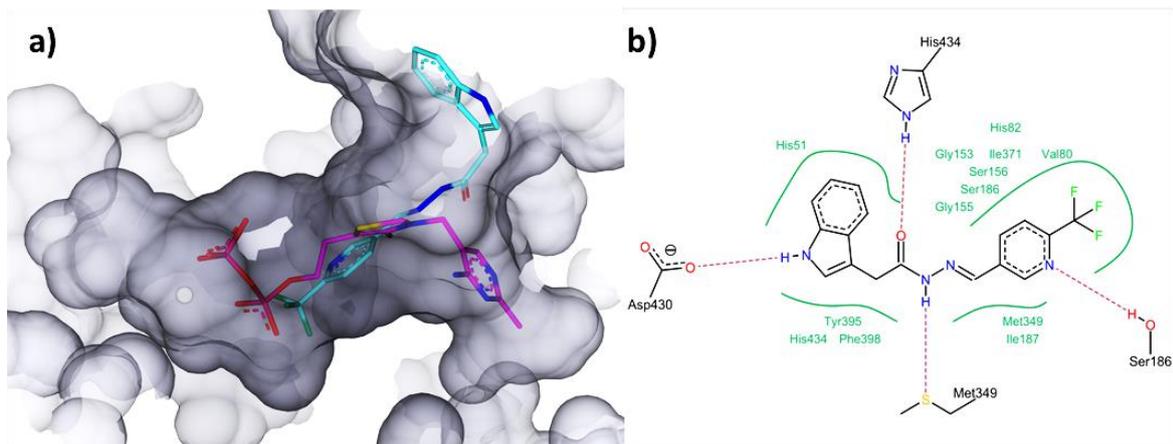
**Supplementary Figure 1** a) Binding mode of compound 1 (cyan) compared to the ThDP (pink) in the active site of drDXPS, b) Interaction of compound 1 with the residues in the active site of drDXPS.



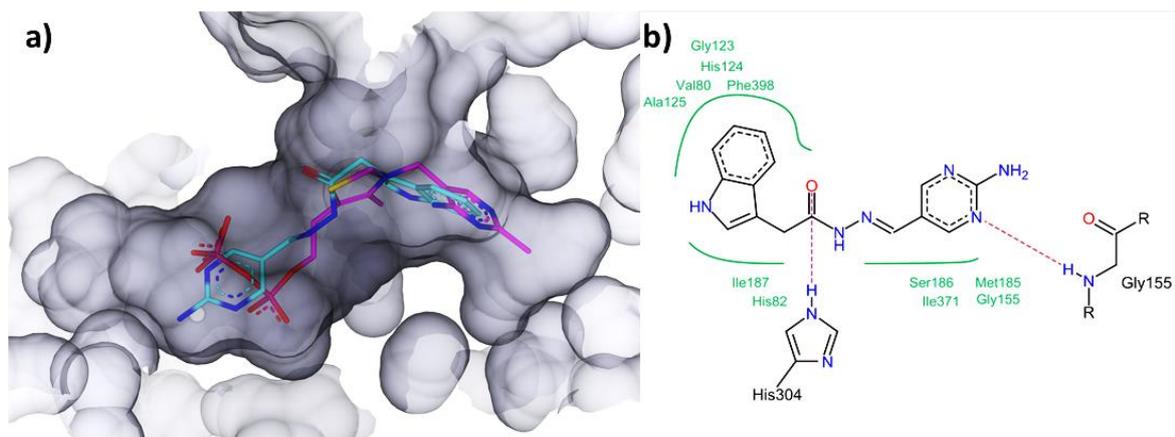
**Supplementary Figure 2** a) Binding mode of compound 2 (cyan) compared to the ThDP (pink) in the active site of drDXPS, b) Interaction of compound 2 with the residues in the active site of drDXPS.



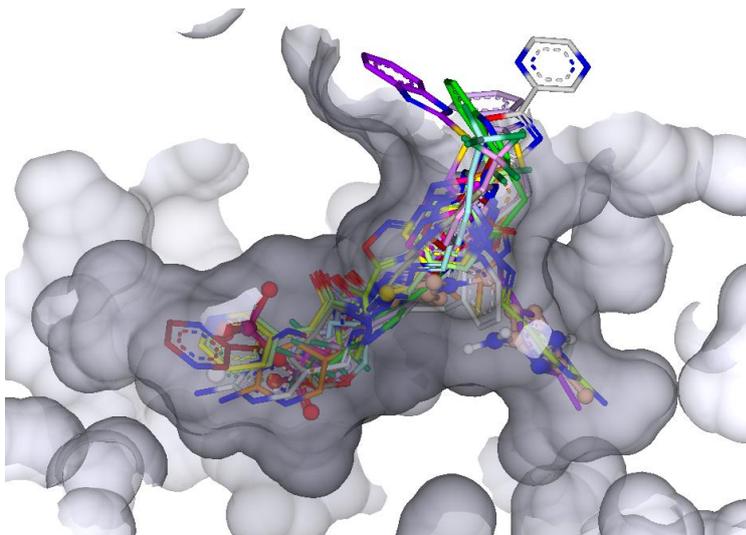
**Supplementary Figure 3** a) Binding mode of compound 3 (cyan) compared to the ThDP (pink) in the active site of drDXPS, b) Interaction of compound 3 with the residues in the active site of drDXPS.



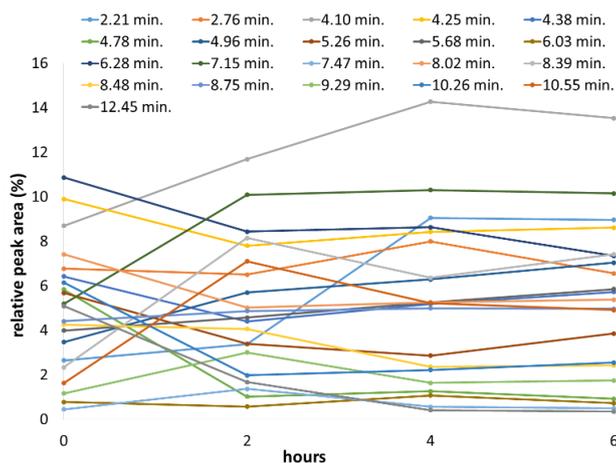
**Supplementary Figure 4** a) Binding mode of compound 4 (cyan) compared to the ThDP (pink) in the active site of drDXPS, b) Interaction of compound 4 with the residues in the active site of drDXPS.



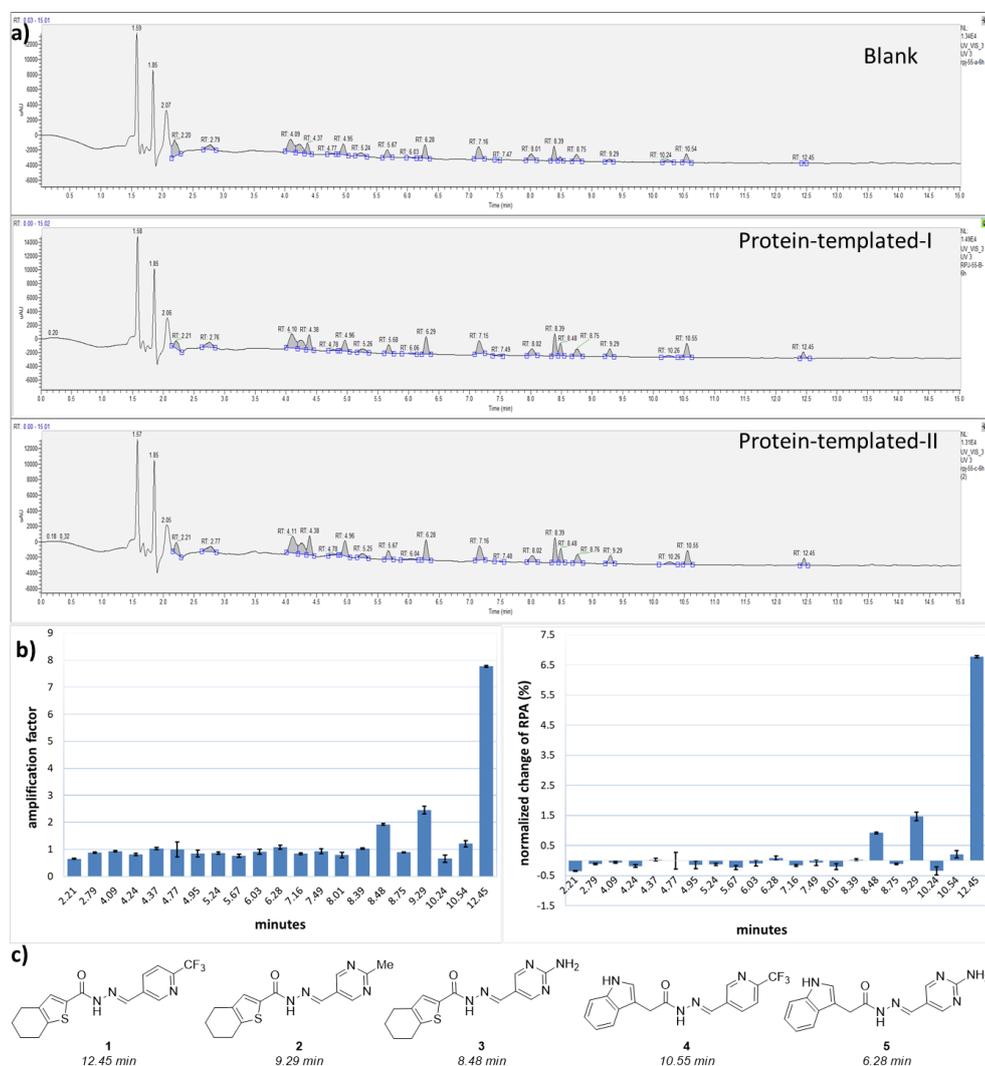
**Supplementary Figure 5** a) Binding mode of compound 5 (cyan) compared to the ThDP (pink) in the active site of drDXPS, b) Interaction of compound 5 with the residues in the active site of drDXPS.



**Supplementary Figure 6** Binding mode comparison of other 16 possible acylhydrazones to the ThDP in the active site of drDXPS

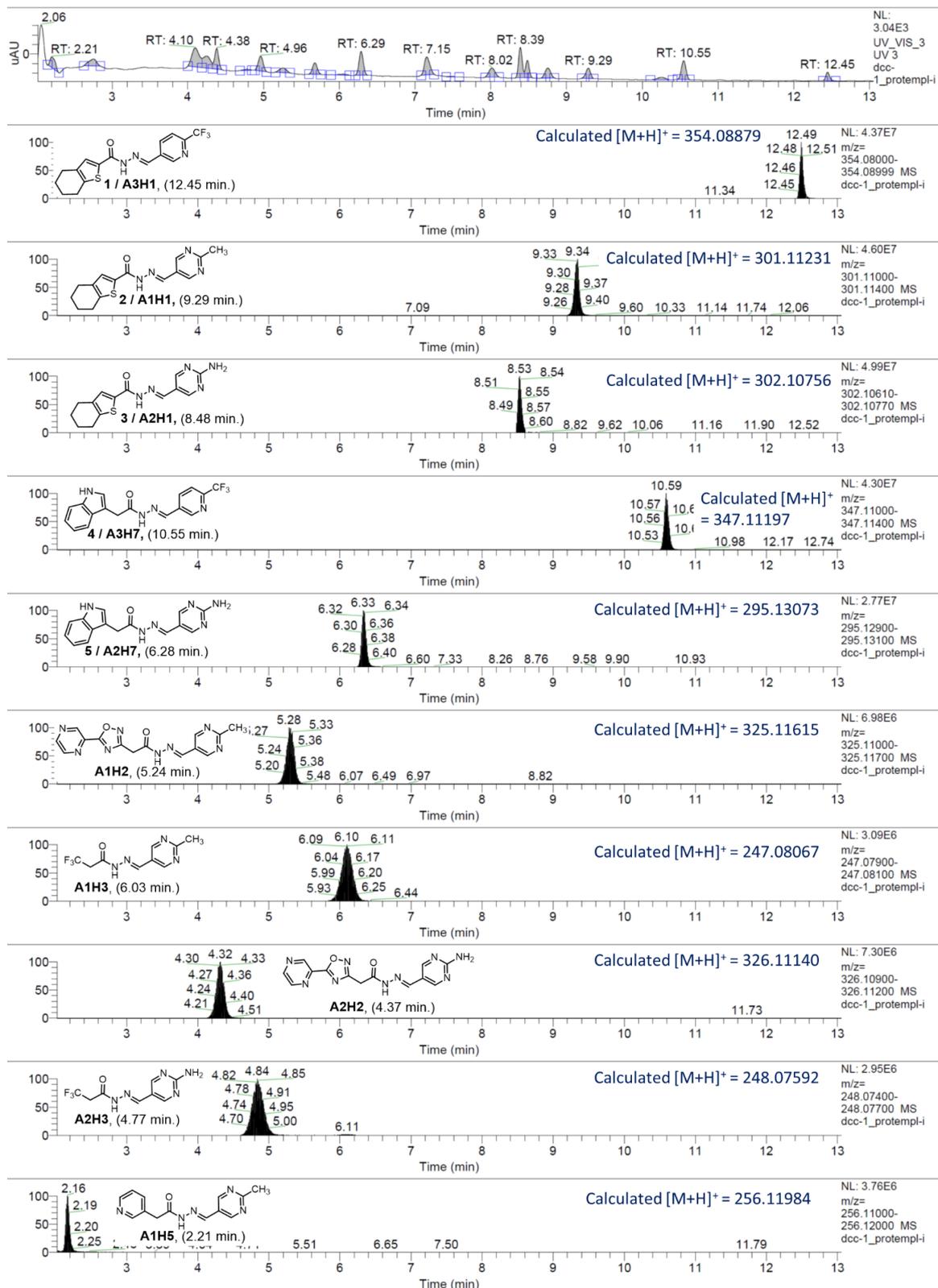


**Supplementary Figure 7** Evaluation of the equilibrium state of acylhydrazone formation in blank DCL-1 by comparing relative peak areas of products formed over time.



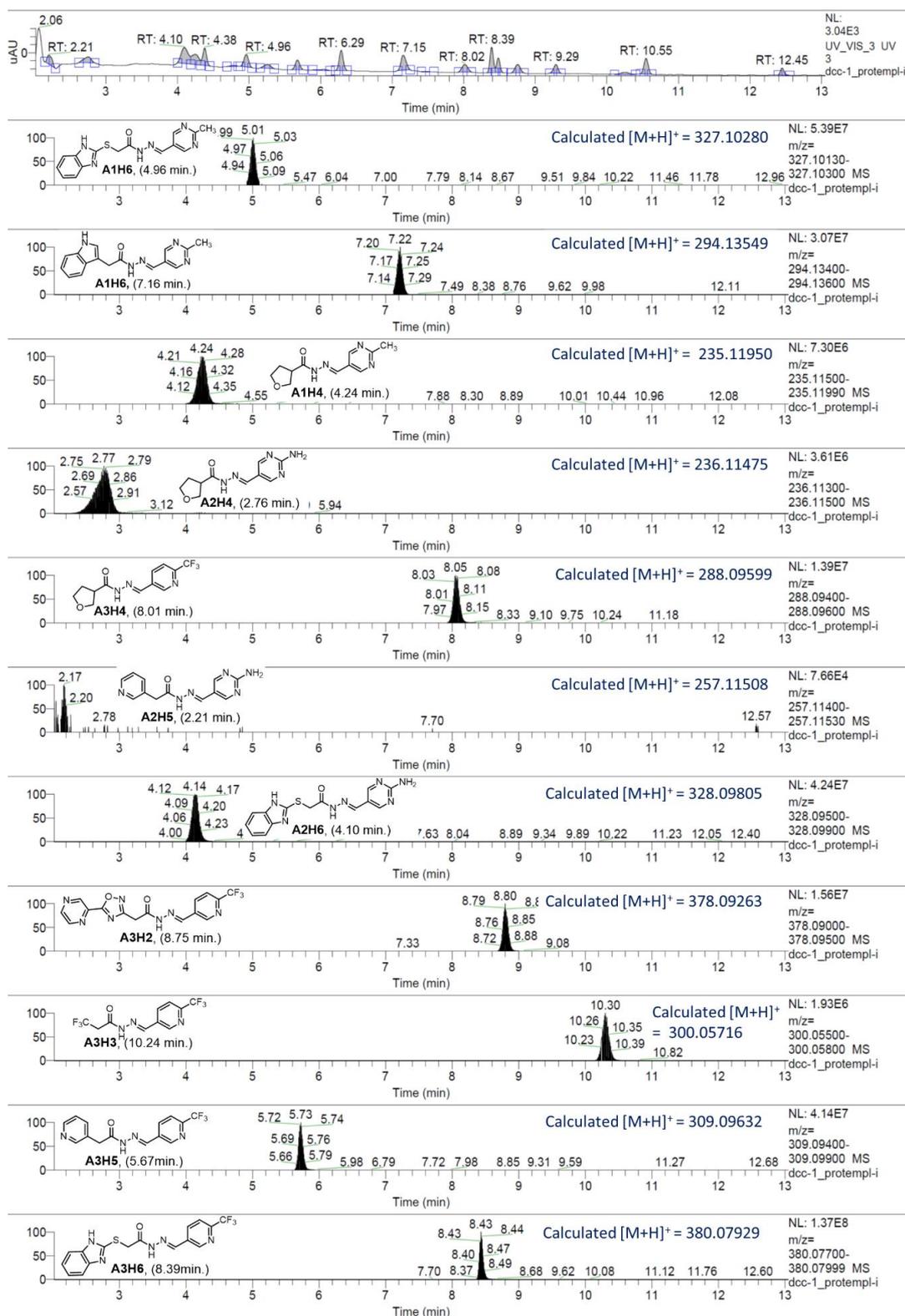
**Supplementary Figure 8** a) Comparison between blank and protein-templated DCC-1 (40 mol% protein) at 6 h, b) amplification factor and normalized change of RPA of products, c) amplified acylhydrazone products in protein-templated DCC.

PART 3. SUPPLEMENTARY INFORMATION

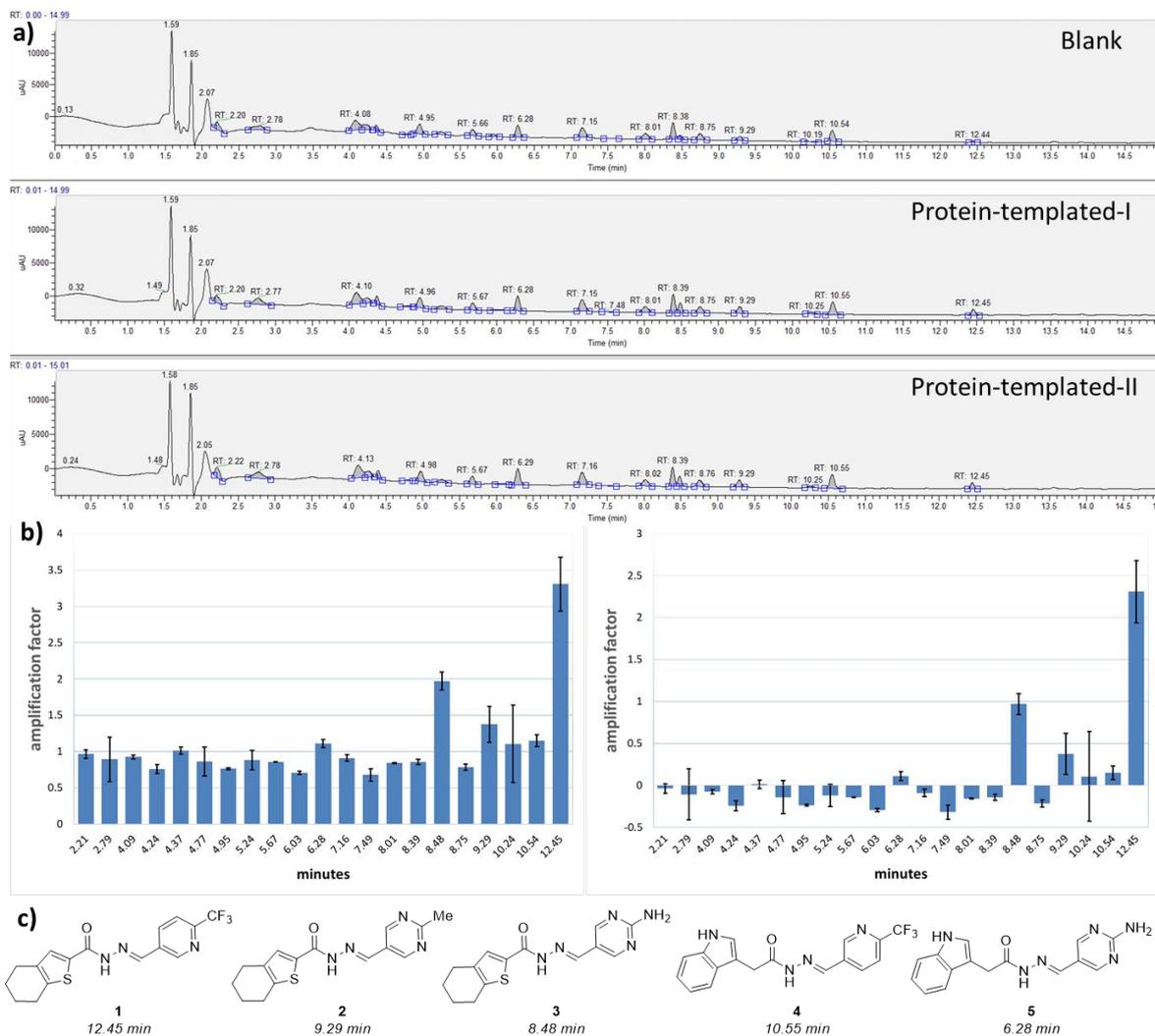


Supplementary Figure 8 d) Different acylhydrazones formed in DCC-1, their retention times, and observed mass.

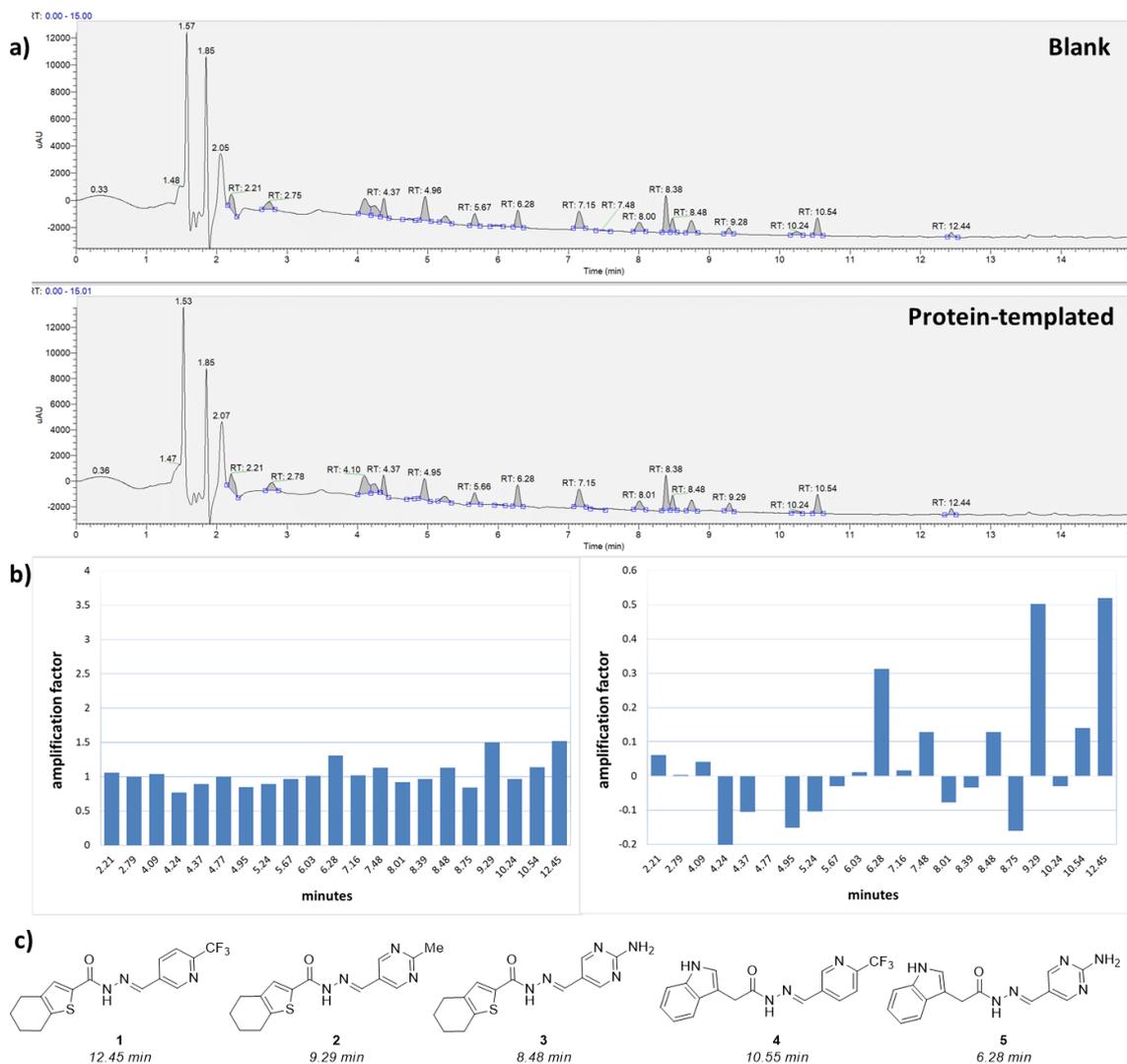
3.4. SUPPLEMENTARY INFORMATION FOR CHAPTER 2.4



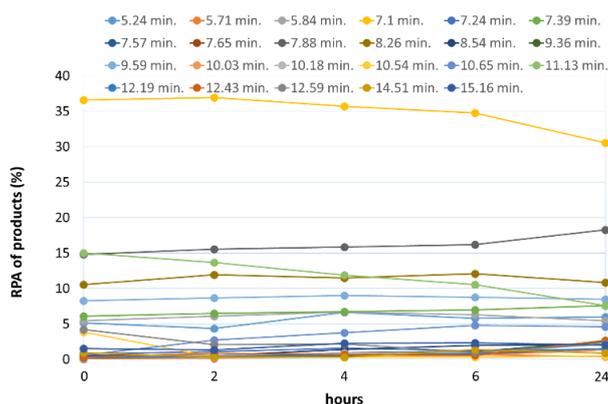
Supplementary Figure 8 e) Different acylhydrazones formed in DCC-1, their retention times, and observed mass.



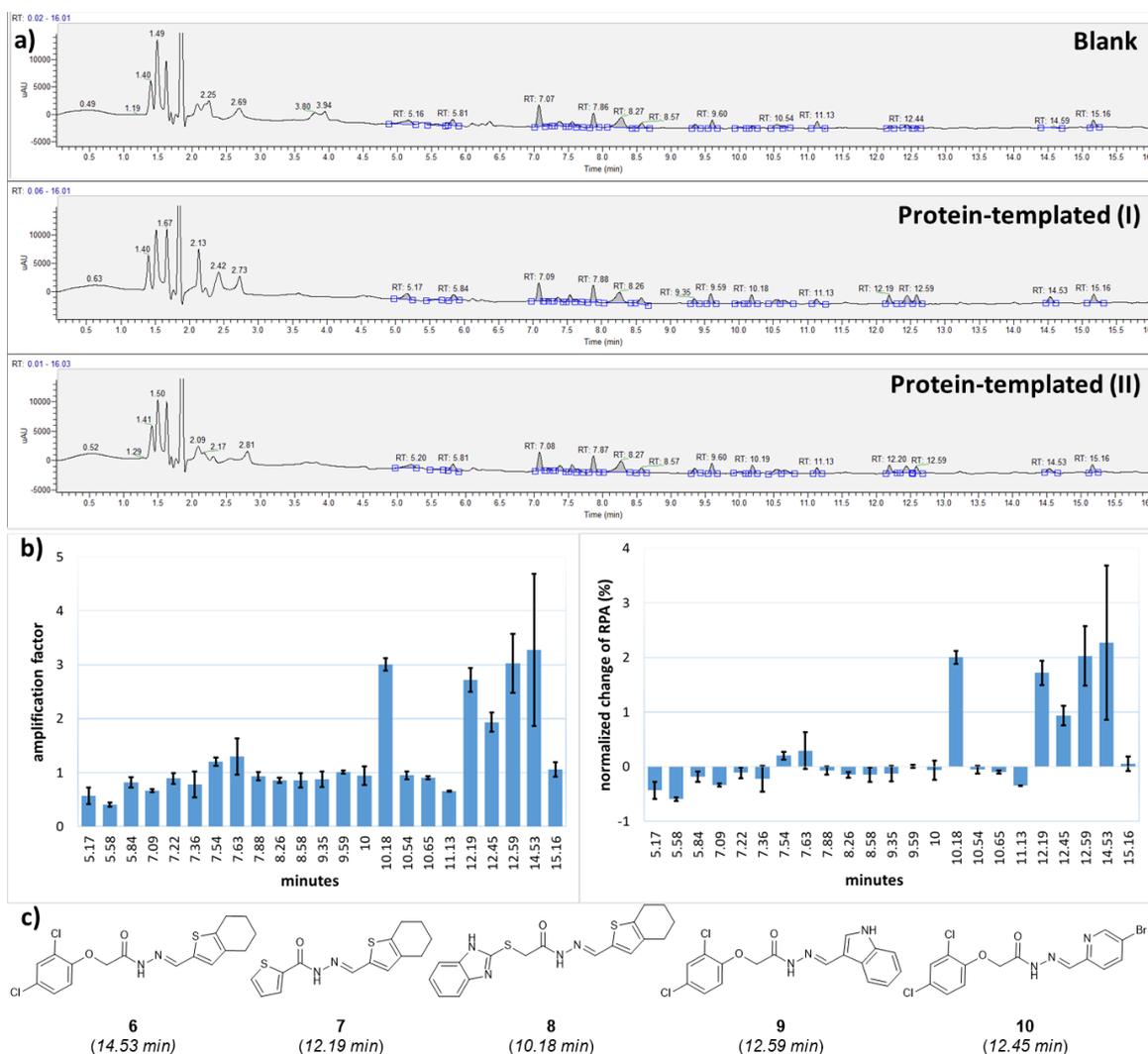
**Supplementary Figure 9** a) Comparison between blank and protein-templated **DCC-1 (20 mol% protein)** at 6 h, b) amplification factor and normalized change of RPA of products, c) amplified acylhydrazone products in protein-templated DCC.



**Supplementary Figure 10** a) Comparison between blank and protein-templated **DCC-1** (20 mol% protein pre-equilibrated) at 6 h, b) amplification factor and normalized change of RPA of products, c) amplified acylhydrazone products in protein-templated DCC.

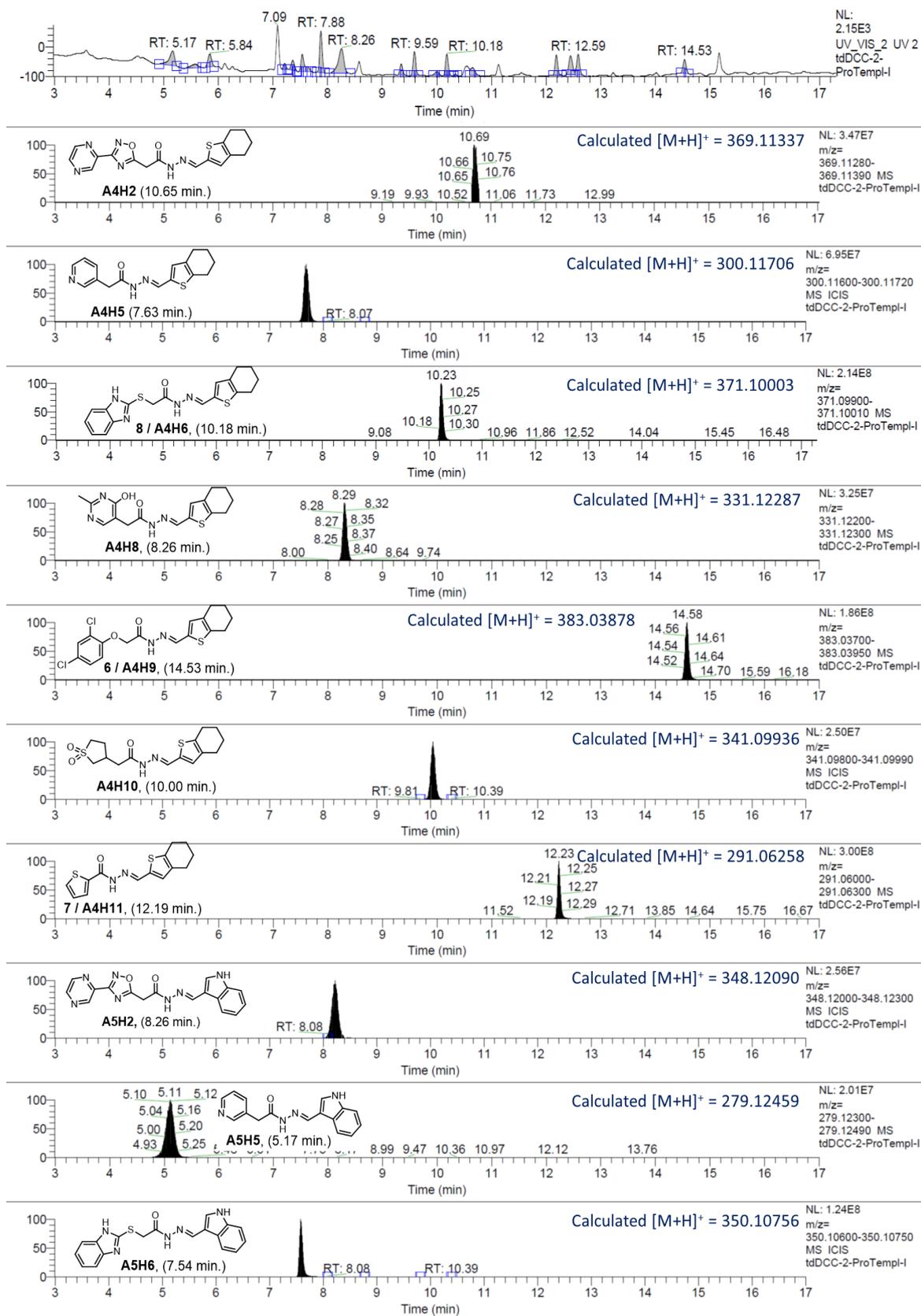


**Supplementary Figure 11** Evaluation of the equilibrium state of acylhydrazone formation in blank **DCL-2** by comparing relative peak areas of products formed over time. Lines represent the formation of acylhydrazone products over time.



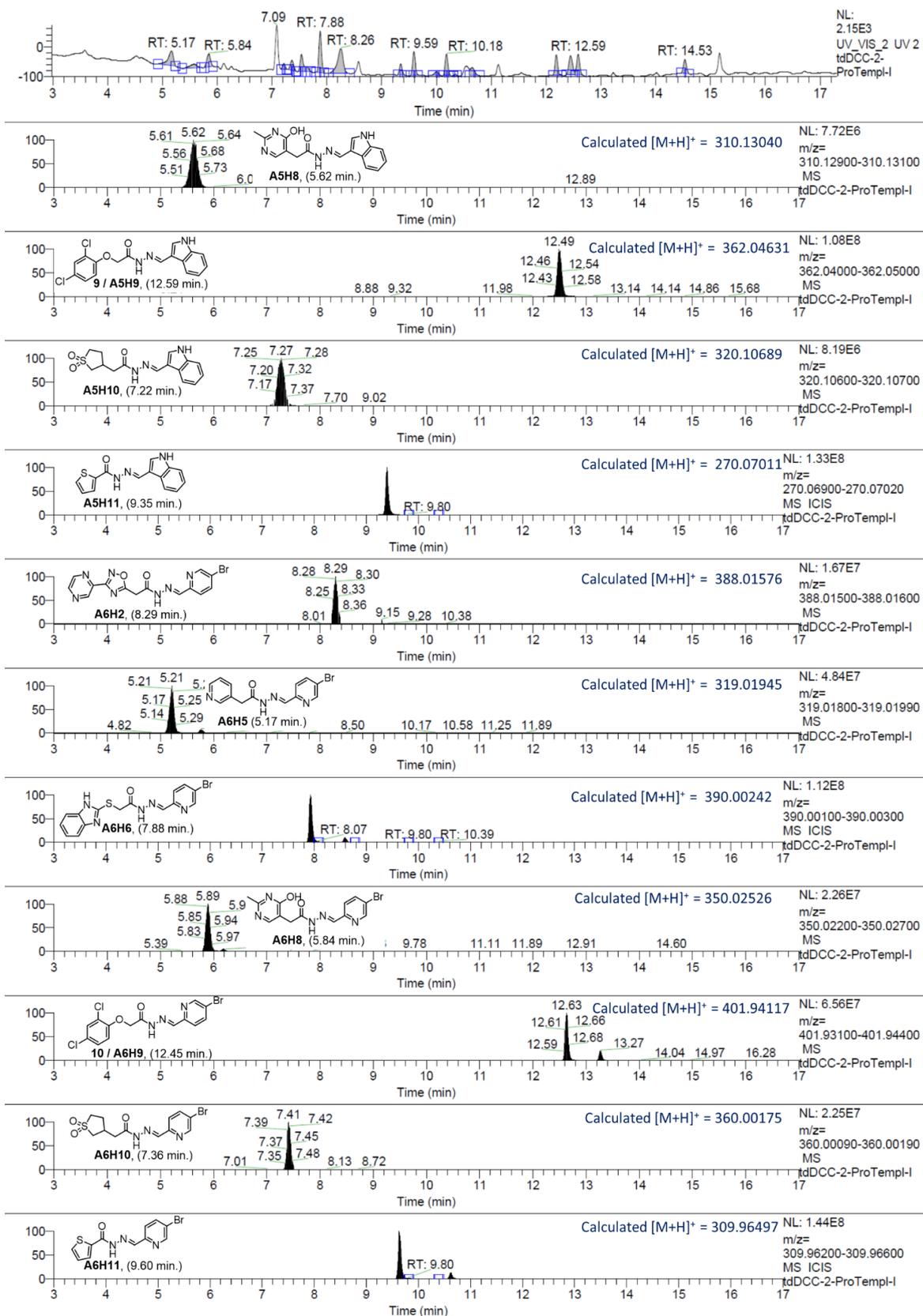
**Supplementary Figure 12** a) Comparison between blank and protein-templated **DCC-2** at 24 h, b) amplification factor and normalized change of RPA of products, c) amplified acylhydrazone products in protein-templated DCC.

## 3.4. SUPPLEMENTARY INFORMATION FOR CHAPTER 2.4

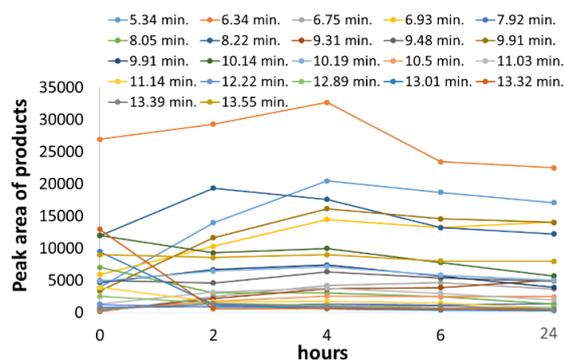


Supplementary Figure 12 d) Different acylhydrazones formed in DCC-2, their retention times, and observed mass.

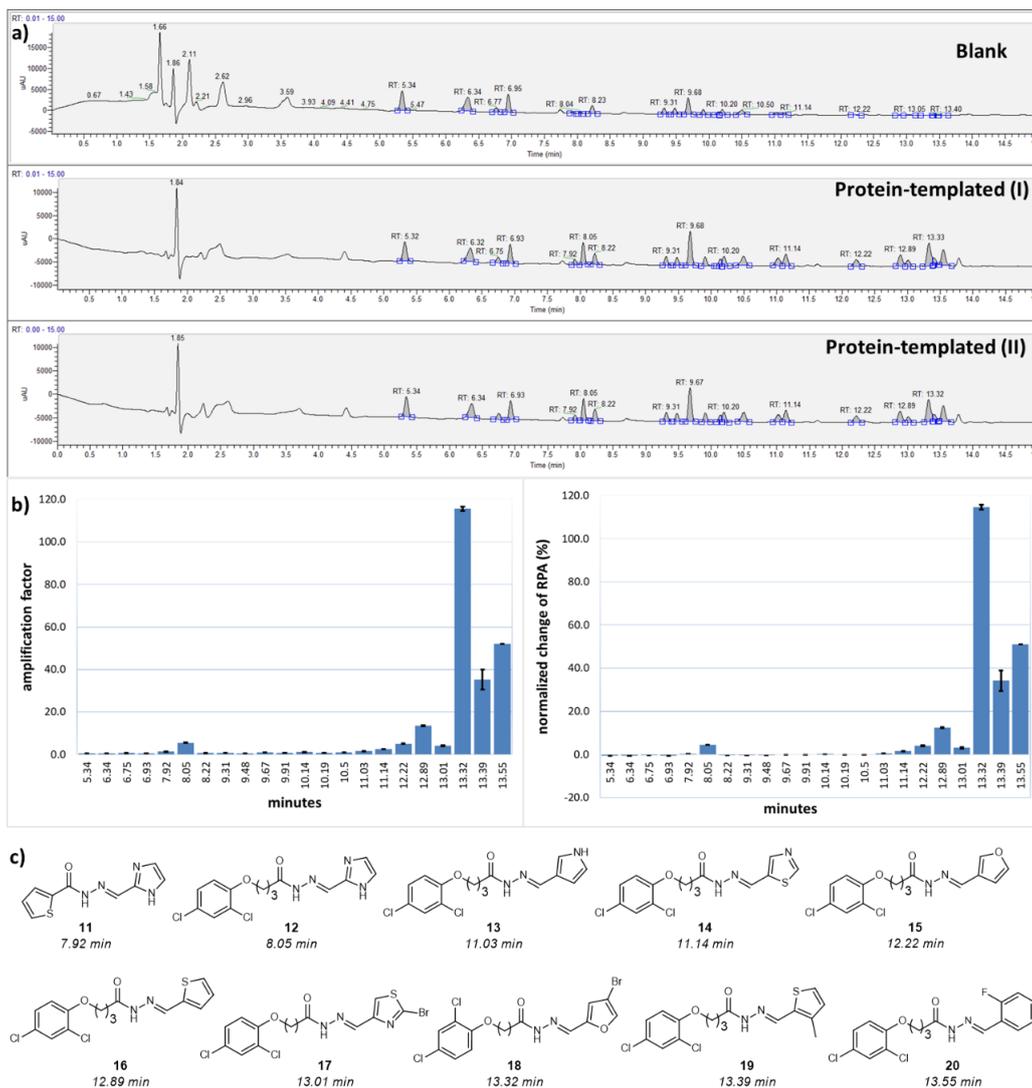
PART 3. SUPPLEMENTARY INFORMATION



Supplementary Figure 12 e) Different acylhydrazones formed in DCC-2, their retention times, and observed mass.

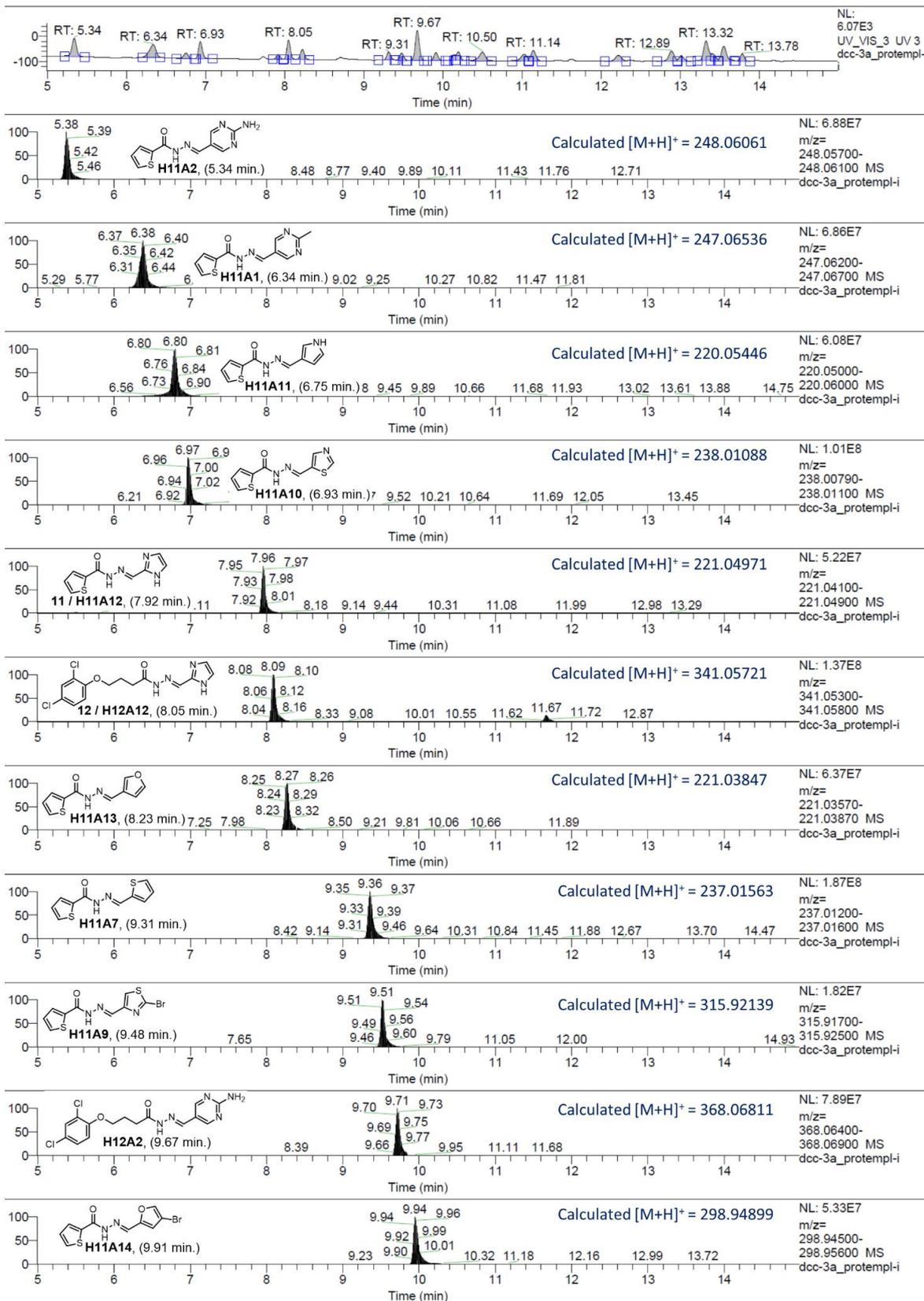


**Supplementary Figure 13** Evaluation of the equilibrium state in a blank **DCL-3a** by comparing peak areas of products formed over time. Lines represent the formation of acylhydrazone products over time.



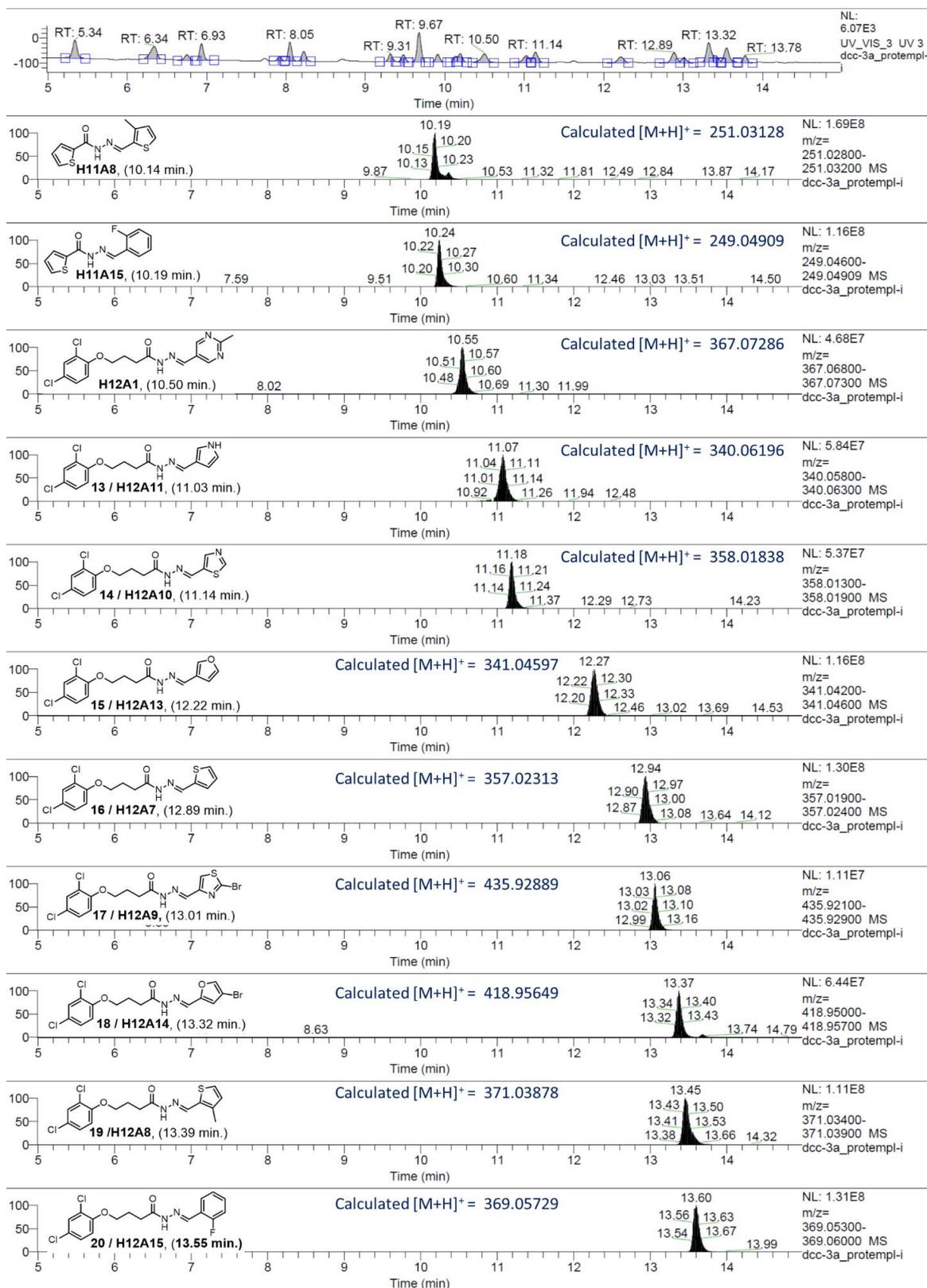
**Supplementary Figure 14** a) Comparison between blank and protein-templated **DCC-3a** at 24 h, b) amplification factor and normalized change of RPA of products, c) amplified acylhydrazone products in protein-templated DCC.

PART 3. SUPPLEMENTARY INFORMATION



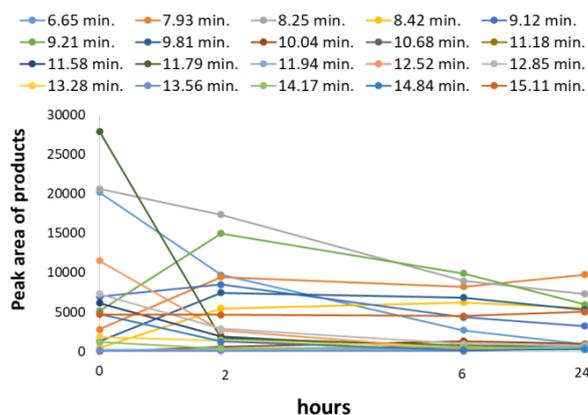
Supplementary Figure 14 d) Different acylhydrazones formed in DCC-3a, their retention times, and observed mass.

## 3.4. SUPPLEMENTARY INFORMATION FOR CHAPTER 2.4

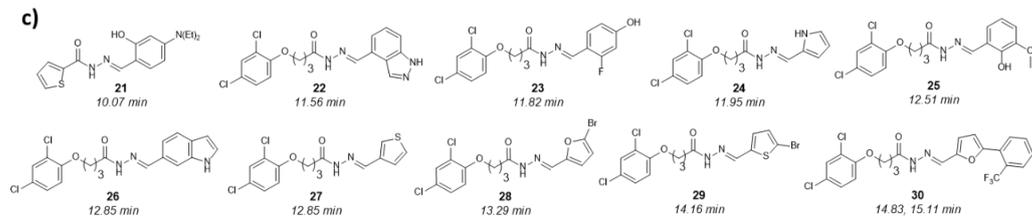
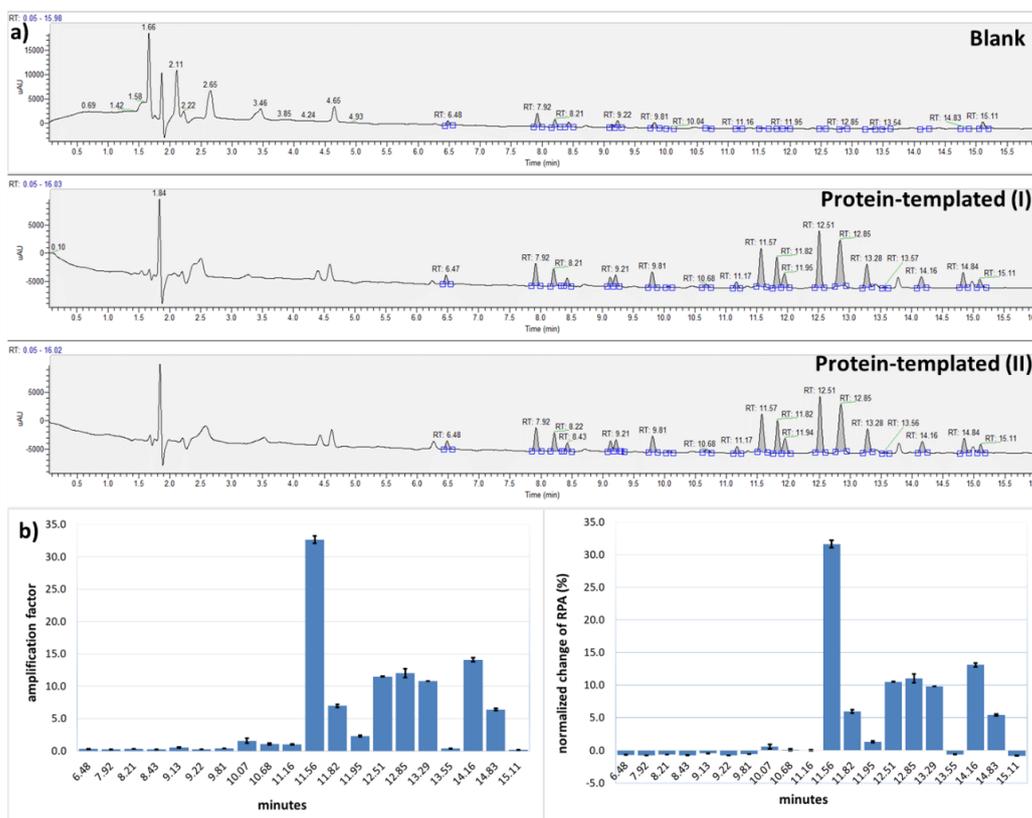


Supplementary Figure 14 e) Different acylhydrazones formed in DCC-3a, their retention times, and observed mass.

## Results from DCC-3b:

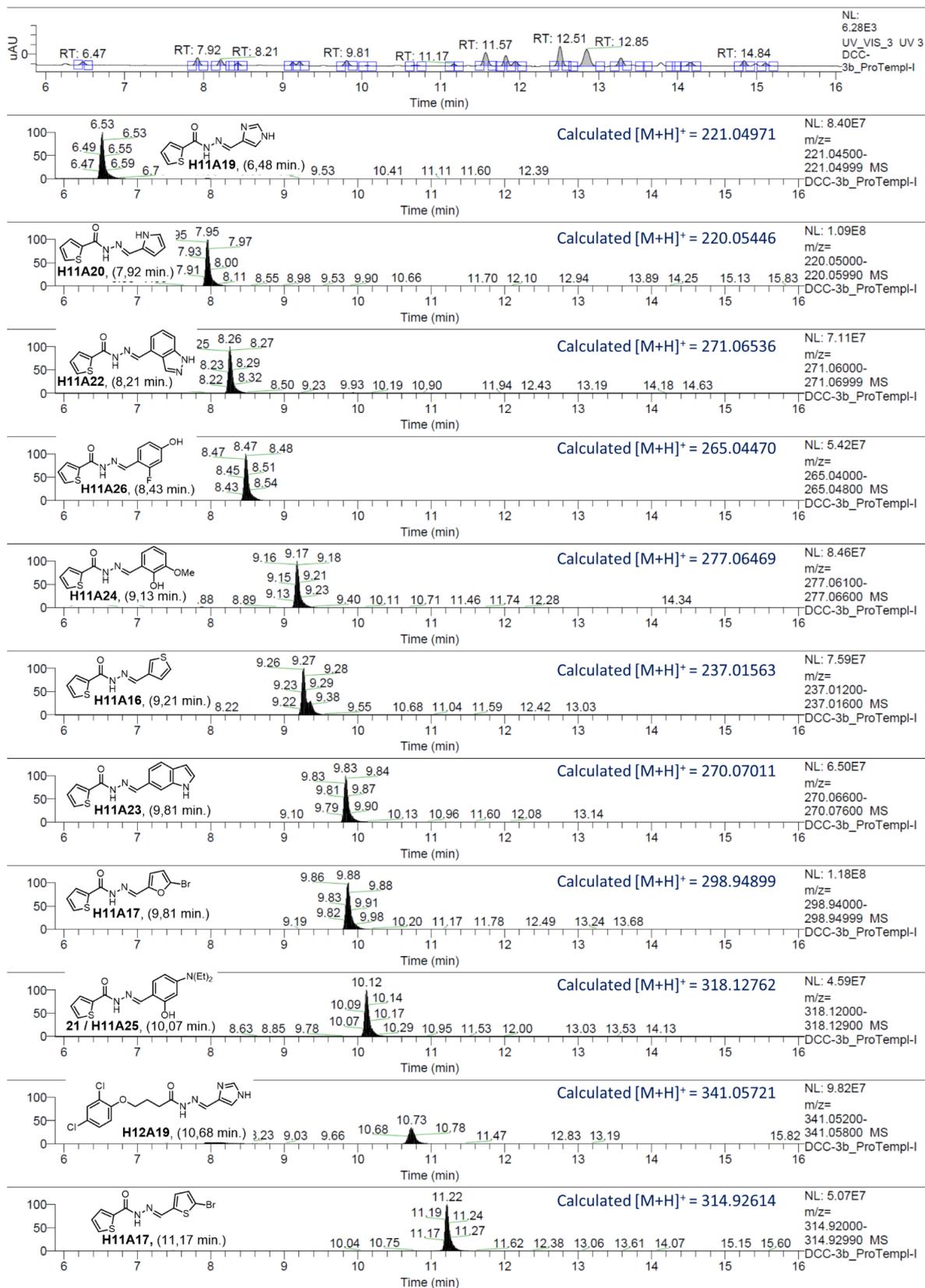


**Supplementary Figure 15** Evaluation of the equilibrium state in a blank **DCL-3b** by comparing peak areas of products formed over time. Lines represent the formation of acylhydrazone products over time.



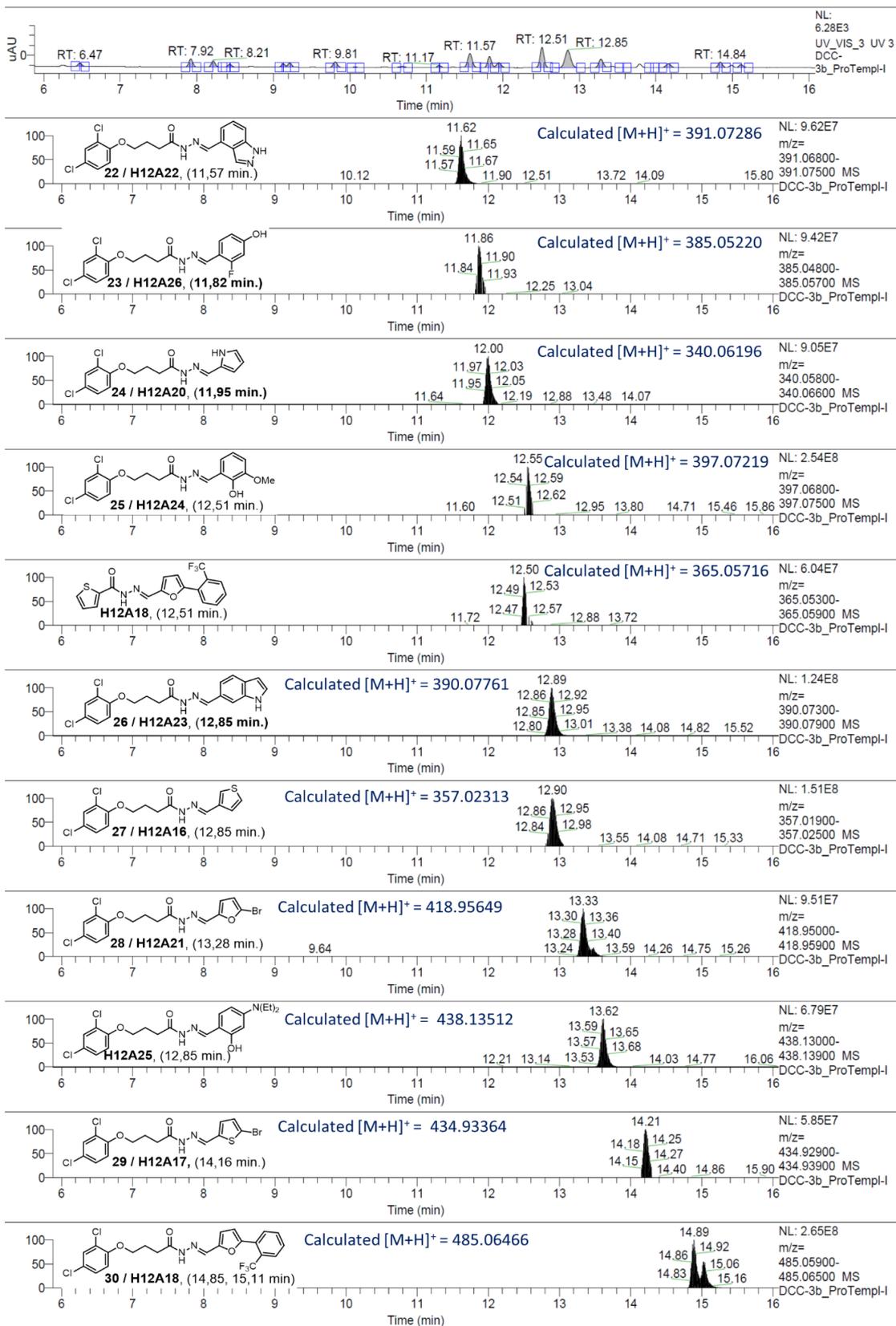
**Supplementary Figure 16** a) Comparison between blank and protein-templated **DCC-3b** at 24 h, b) amplification factor and normalized change of RPA of products, c) amplified acylhydrazone products in protein-templated DCC.

## 3.4. SUPPLEMENTARY INFORMATION FOR CHAPTER 2.4

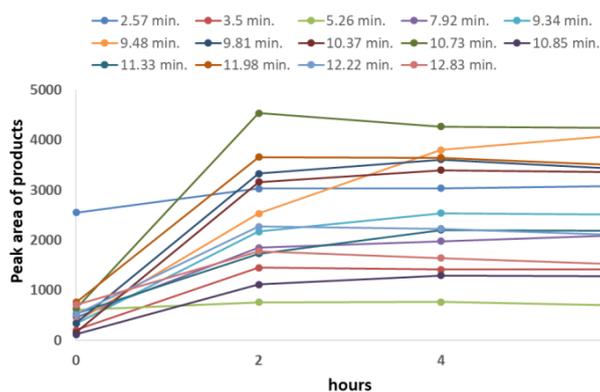


**Supplementary Figure 16 d)** Different acylhydrazones formed in DCC-3b, their retention times, and observed mass.

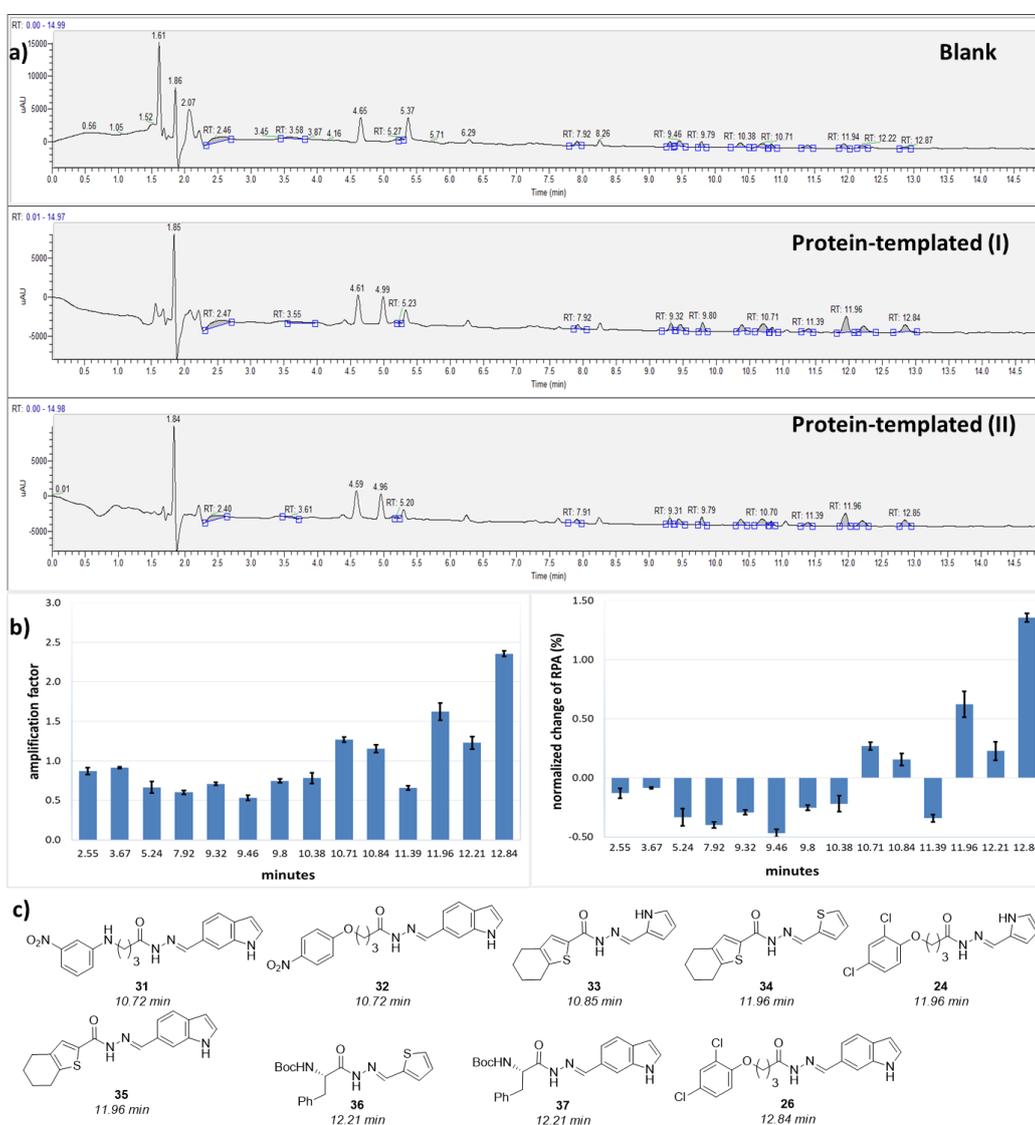
PART 3. SUPPLEMENTARY INFORMATION



Supplementary Figure 16 e) Different acylhydrazones formed in DCC-3b, their retention times, and observed mass.

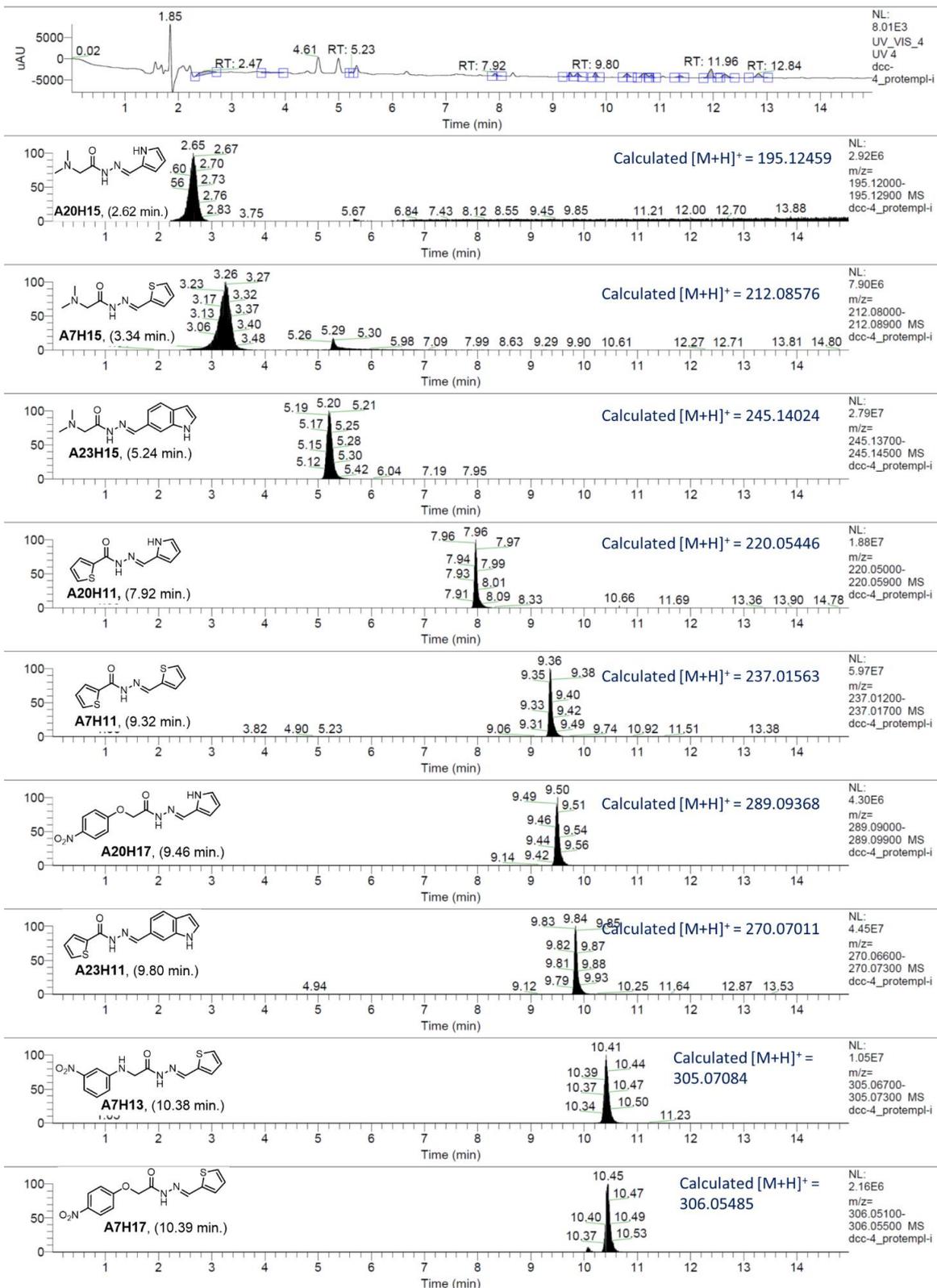


**Supplementary Figure 17** Evaluation of the equilibrium state in a blank **DCL-4** by comparing peak areas of products formed over time. Lines represent the formation of acylhydrazone products over time.



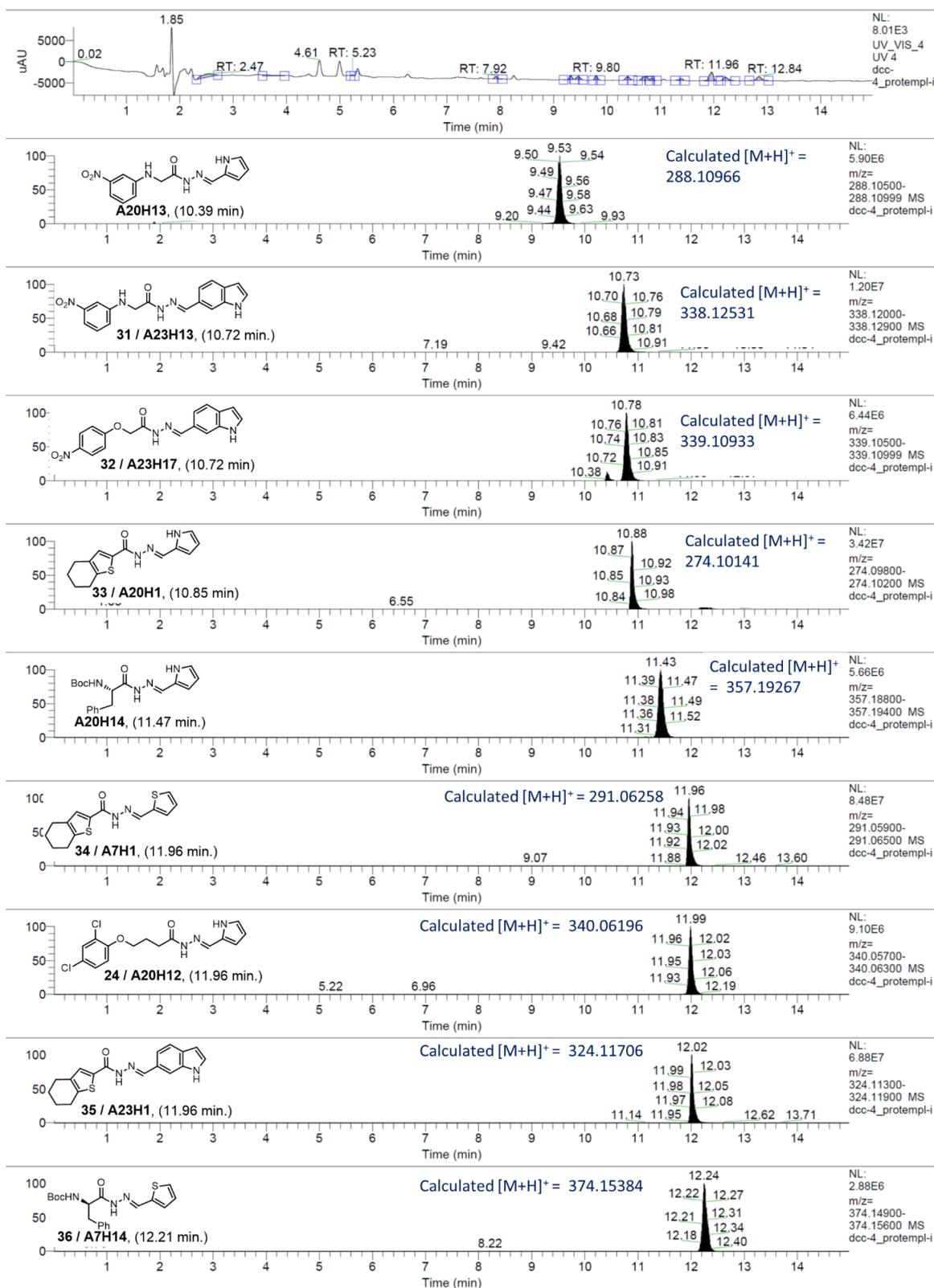
**Supplementary Figure 18** a) Comparison between blank and protein-templated **DCC-4** at 6 h, b) amplified acylhydrazone products in protein-templated **DCC**.

PART 3. SUPPLEMENTARY INFORMATION



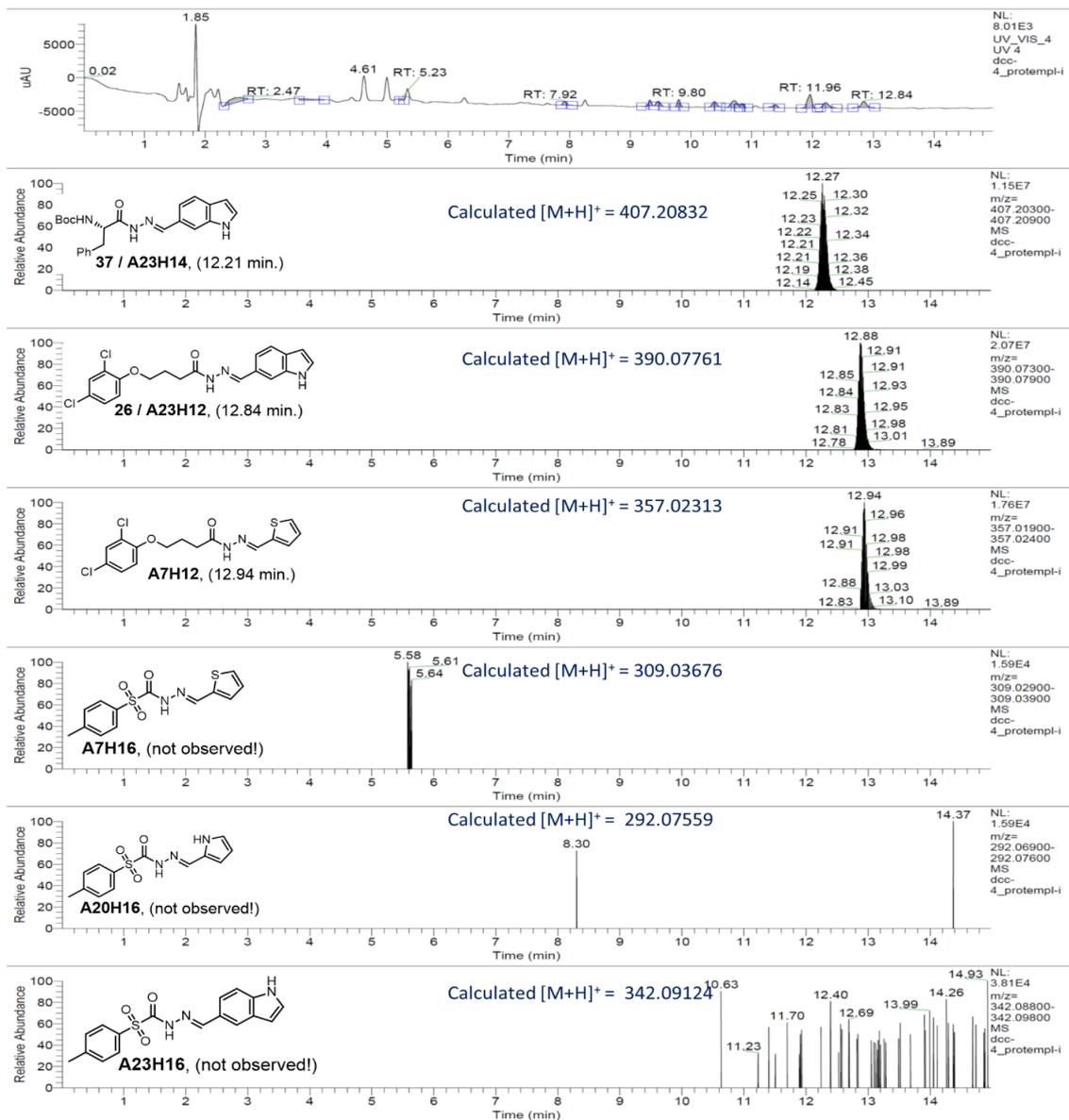
Supplementary Figure 18 e) Different acylhydrazones formed in DCC-3b, their retention times, and observed mass.

## 3.4. SUPPLEMENTARY INFORMATION FOR CHAPTER 2.4



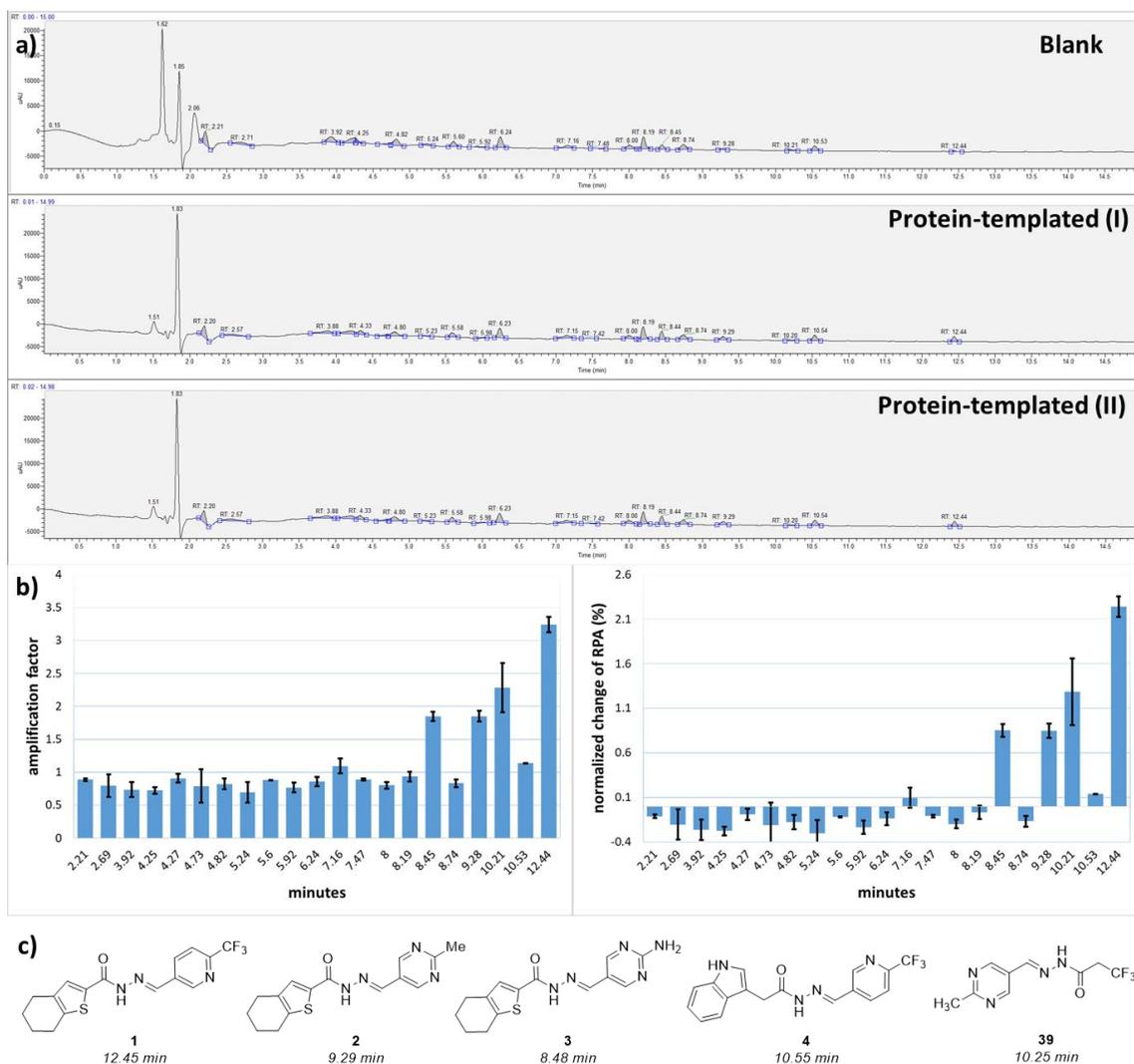
Supplementary Figure 18 f) Different acylhydrazones formed in DCC-3b, their retention times, and observed mass.

PART 3. SUPPLEMENTARY INFORMATION

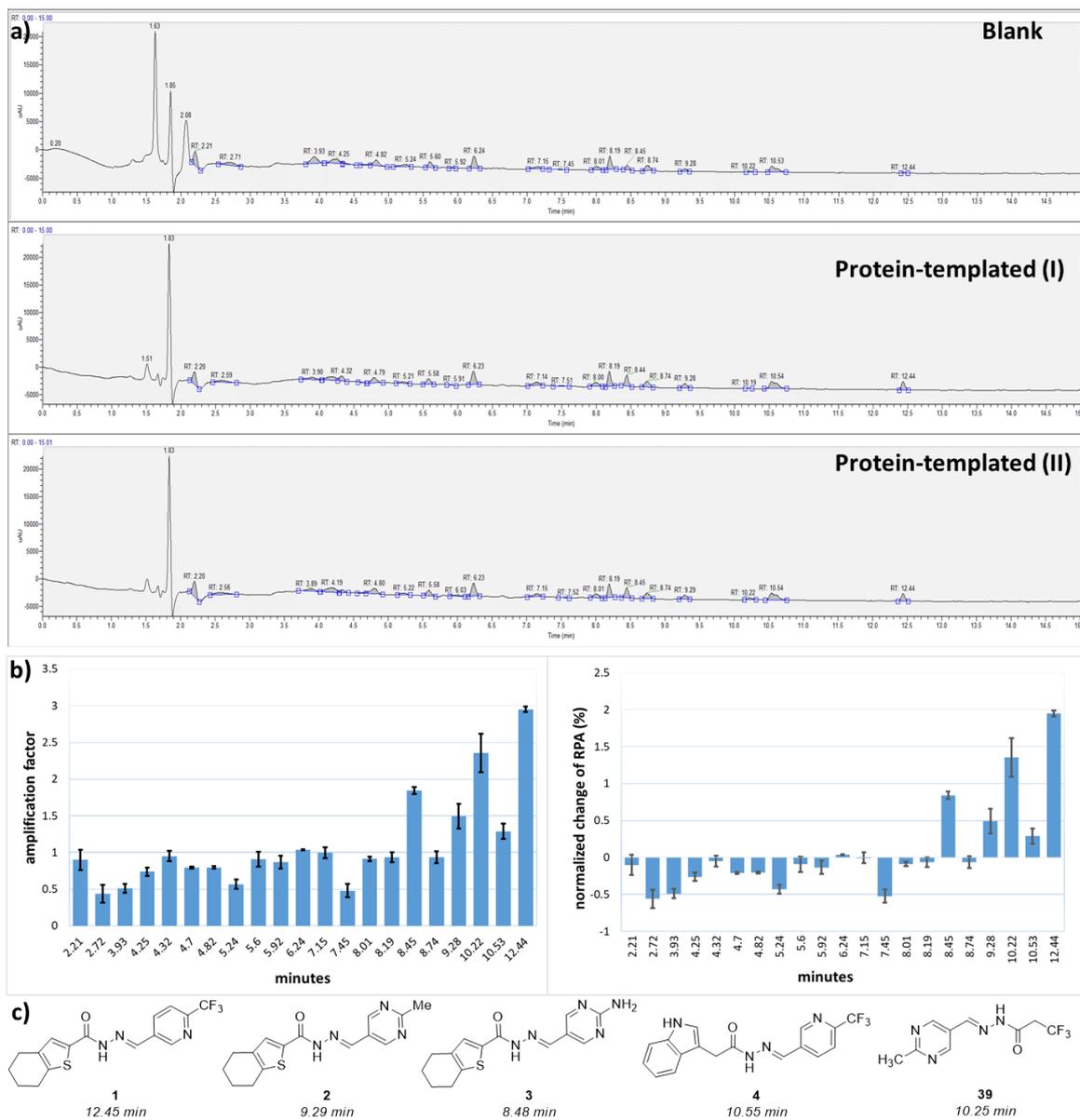


Supplementary Figure 18 g) Different acylhydrazones formed in DCC-3b, their retention times, and observed mass.

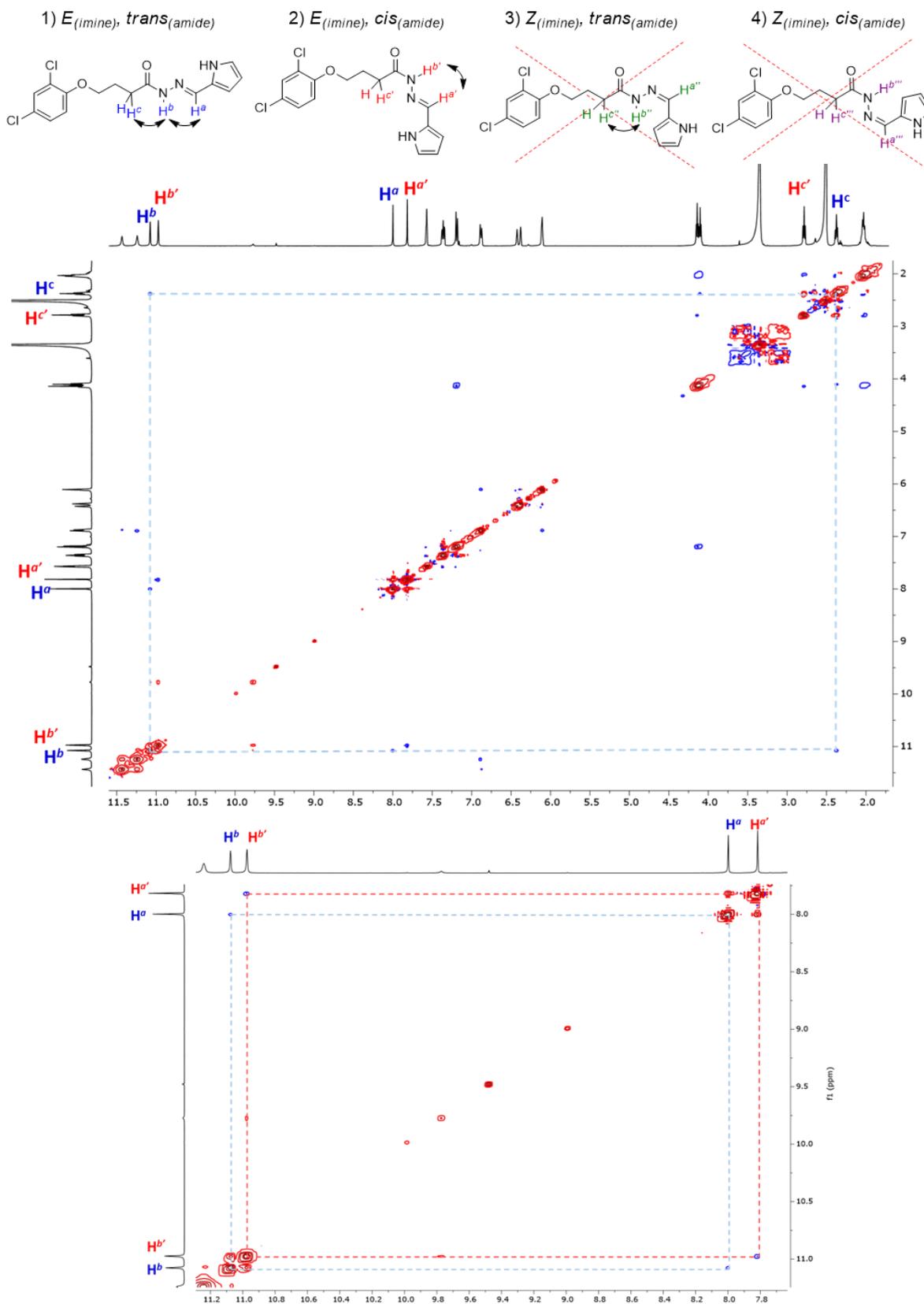
### 3.4. SUPPLEMENTARY INFORMATION FOR CHAPTER 2.4



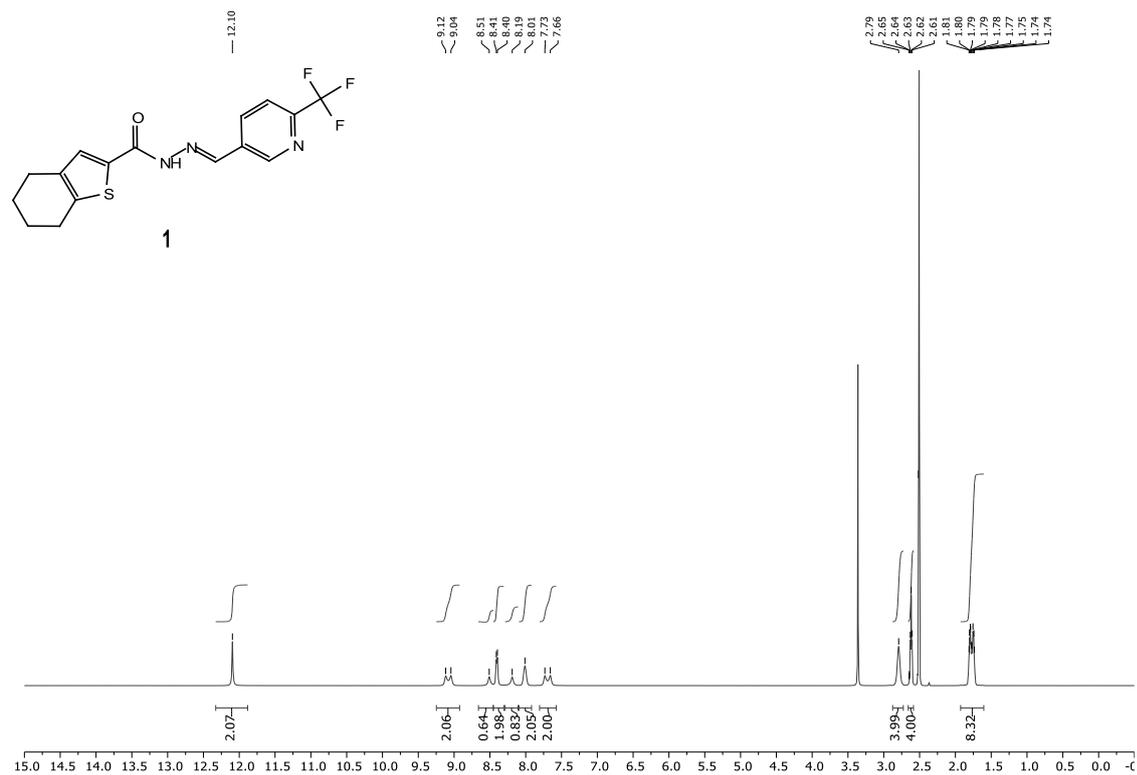
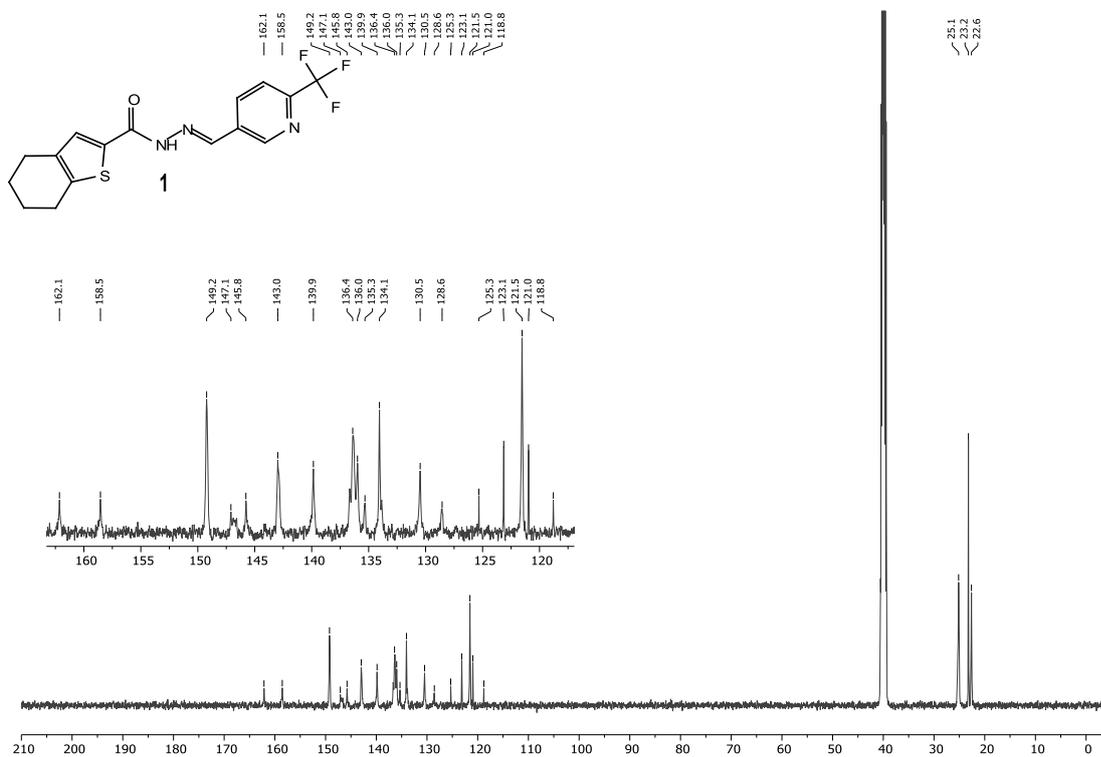
**Supplementary Figure 19** a) Comparison between blank and protein-templated **DCC-5 (ThDP)** at 6 h, b) amplification factor and normalized change of RPA of products, c) amplified acylhydrazone products in protein-templated DCC.

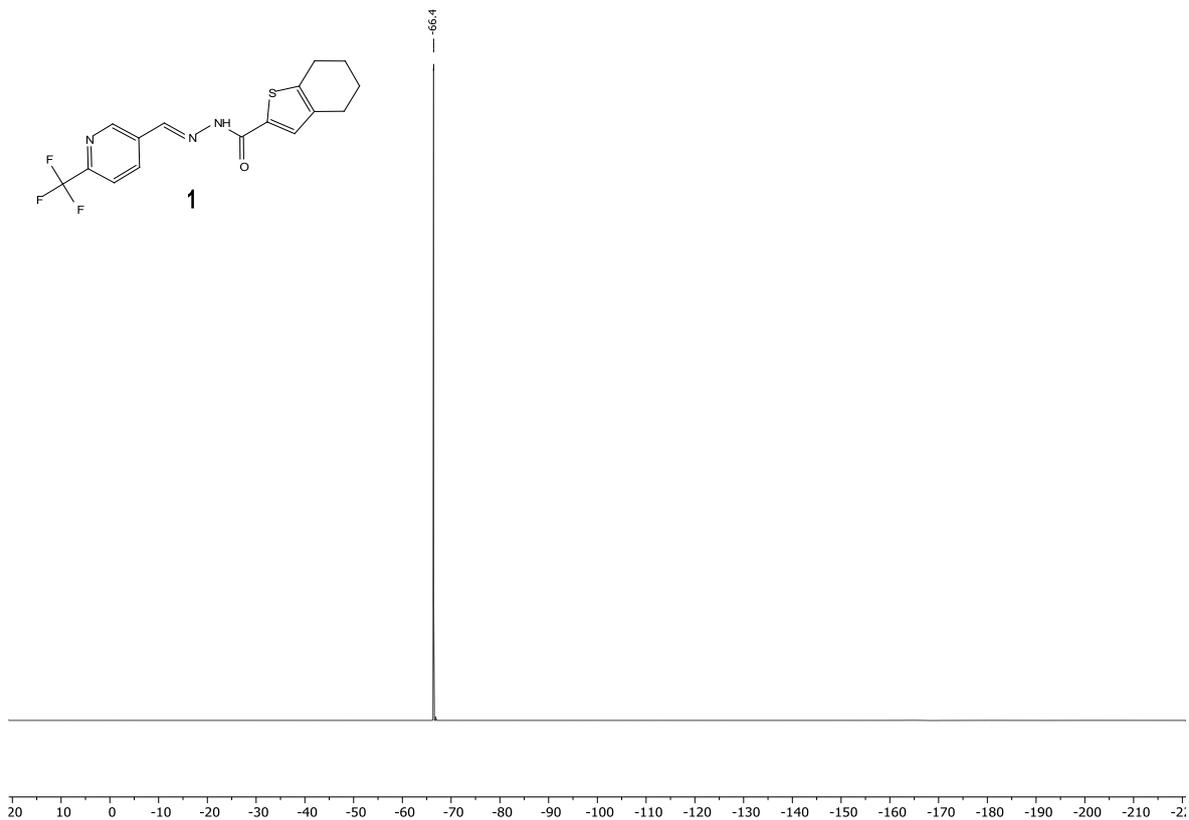
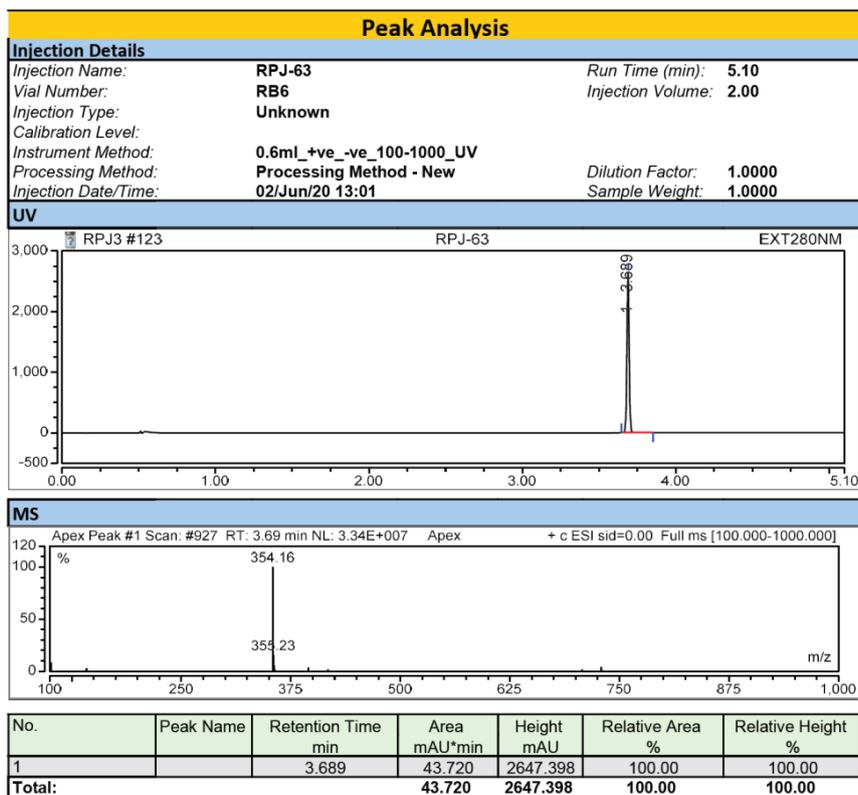


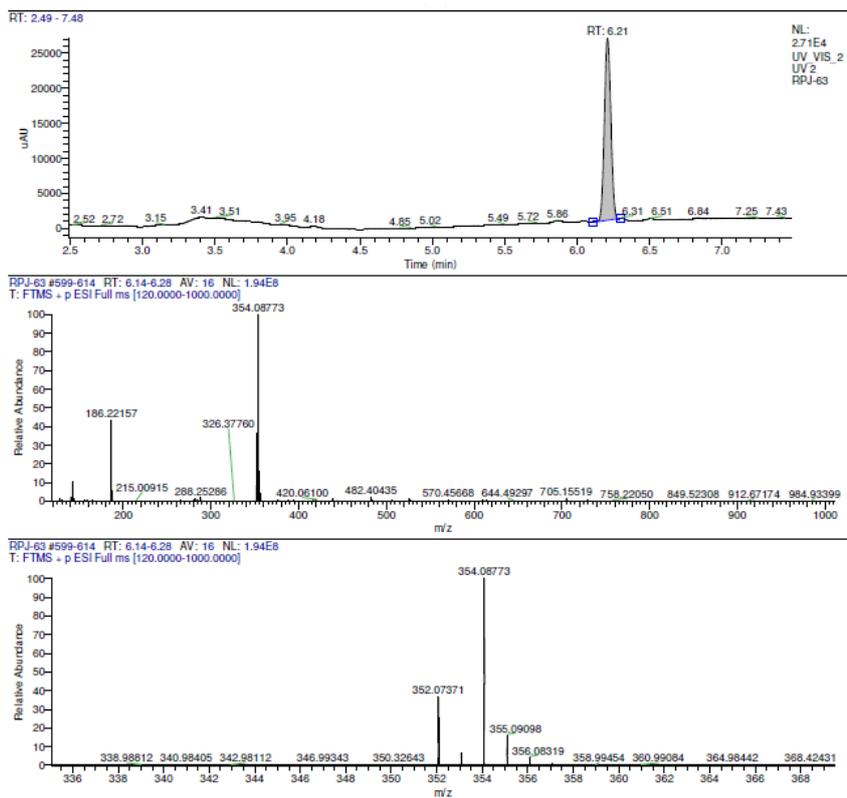
**Supplementary Figure 20** a) Comparison between blank and protein-templated DCC-6 at 6 h, b) amplification factor and normalized change of RPA of products, c) amplified acylhydrazone products in protein-templated DCC.



Supplementary Figure 21 NOESY NMR spectra of compound 24.

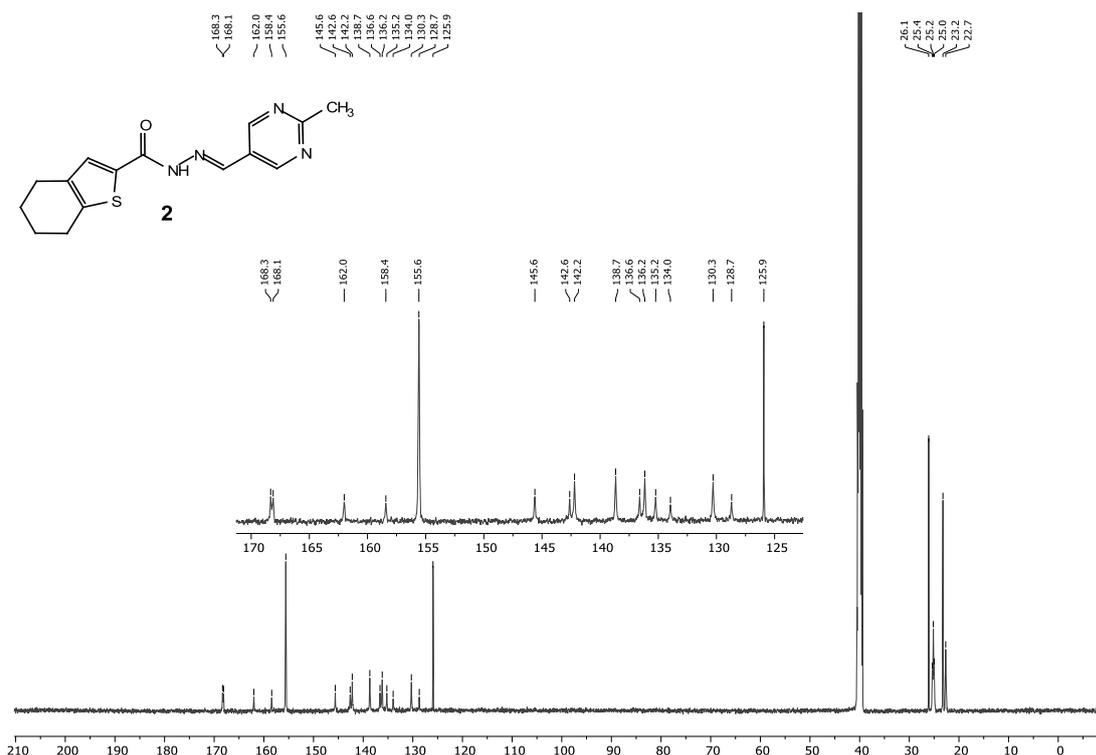
Supplementary Figure 22 <sup>1</sup>H NMR spectrum of **1**.Supplementary Figure 23 <sup>13</sup>C NMR spectrum of **1**.

Supplementary Figure 24 <sup>19</sup>F NMR spectrum of **1**.Supplementary Figure 25 HPLC purity analysis of **1**.

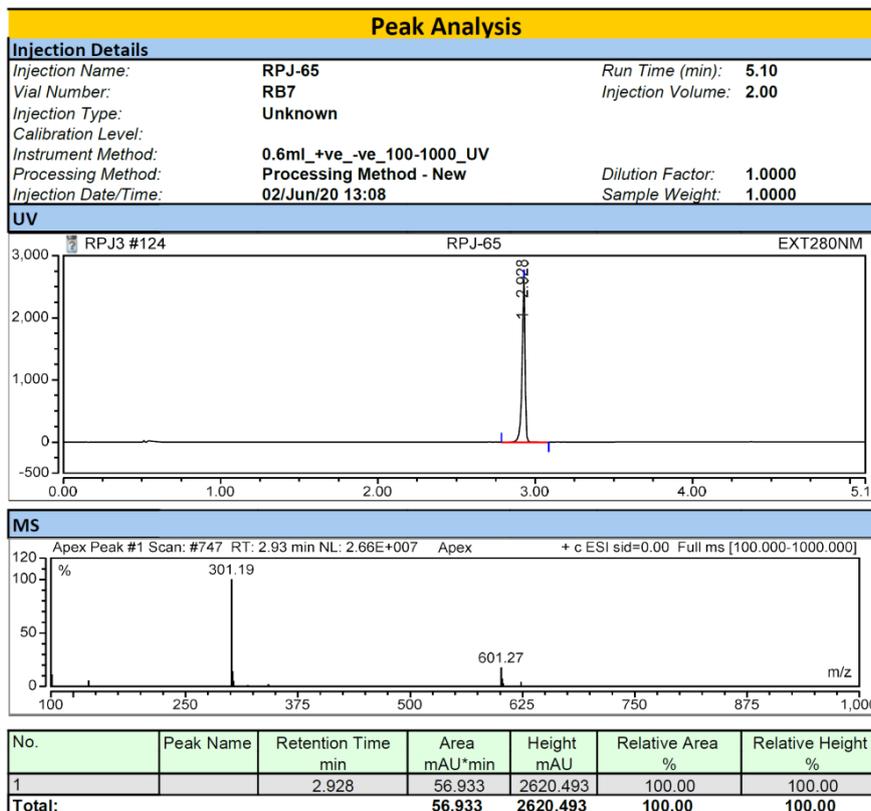


Supplementary Figure 26 HRMS of 1.

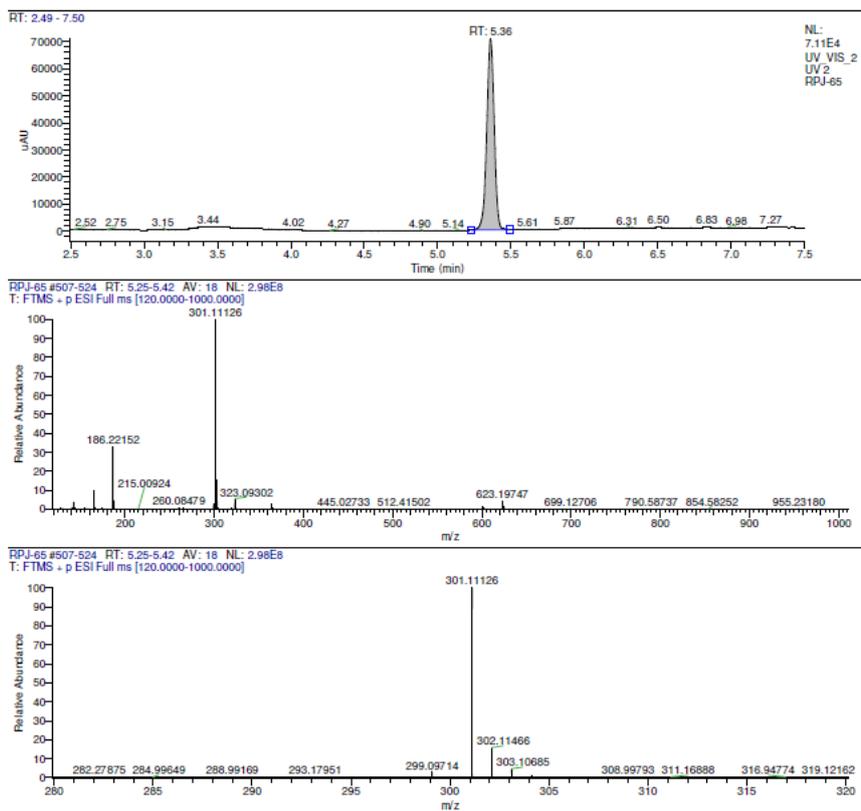
Supplementary Figure 27  $^1\text{H}$  NMR spectrum of 2.



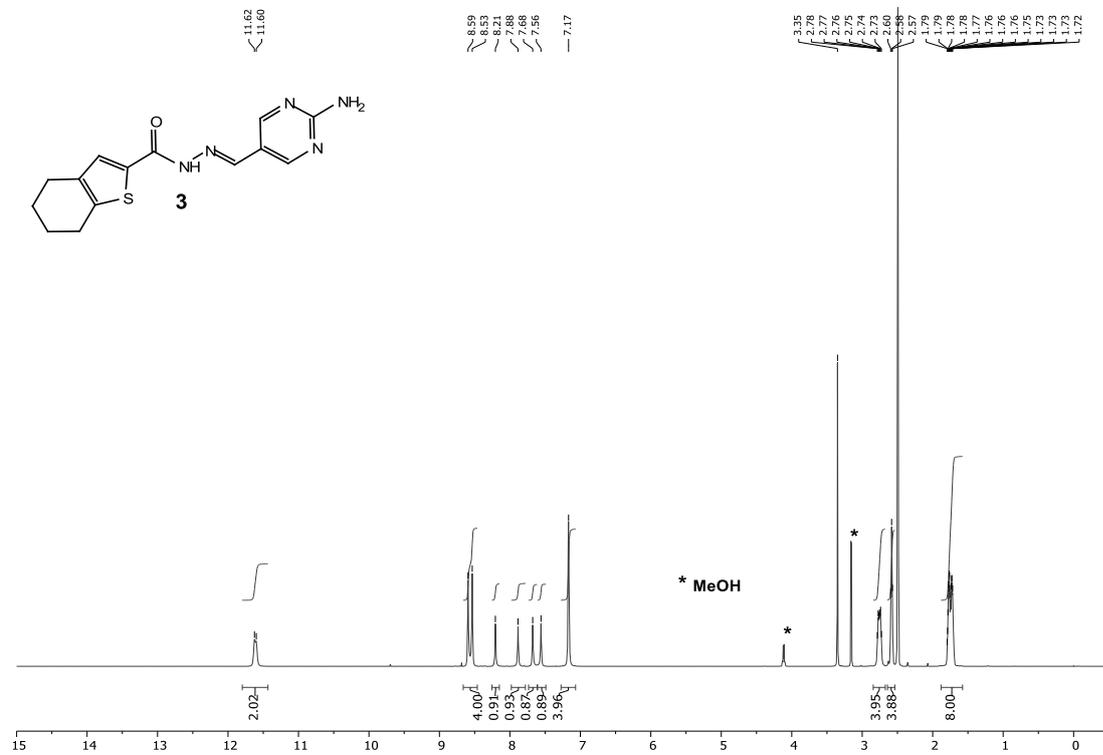
Supplementary Figure 28 <sup>13</sup>C NMR spectrum of 2.

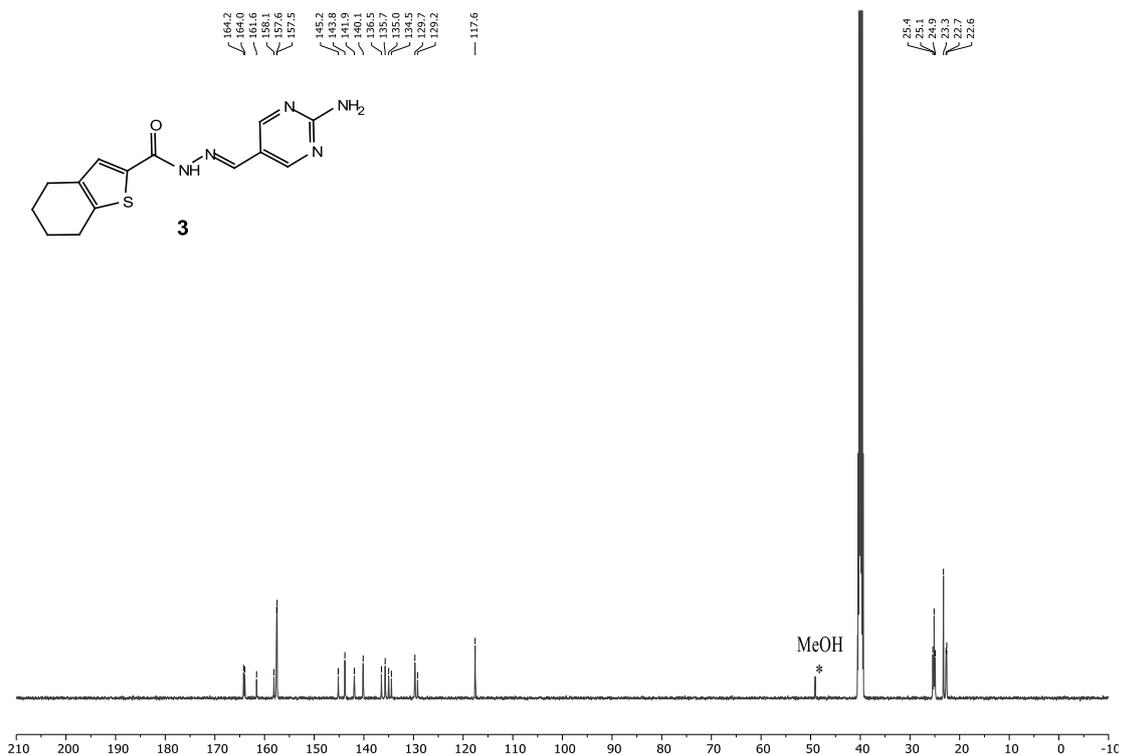


Supplementary Figure 29 HPLC purity analysis of 2.

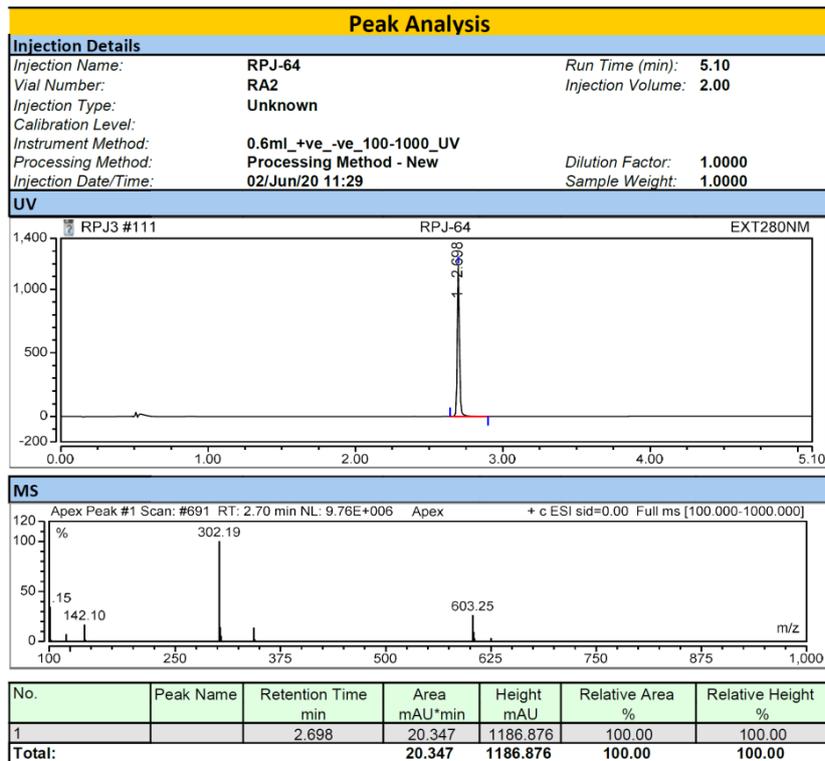


Supplementary Figure 30 HRMS of 2.

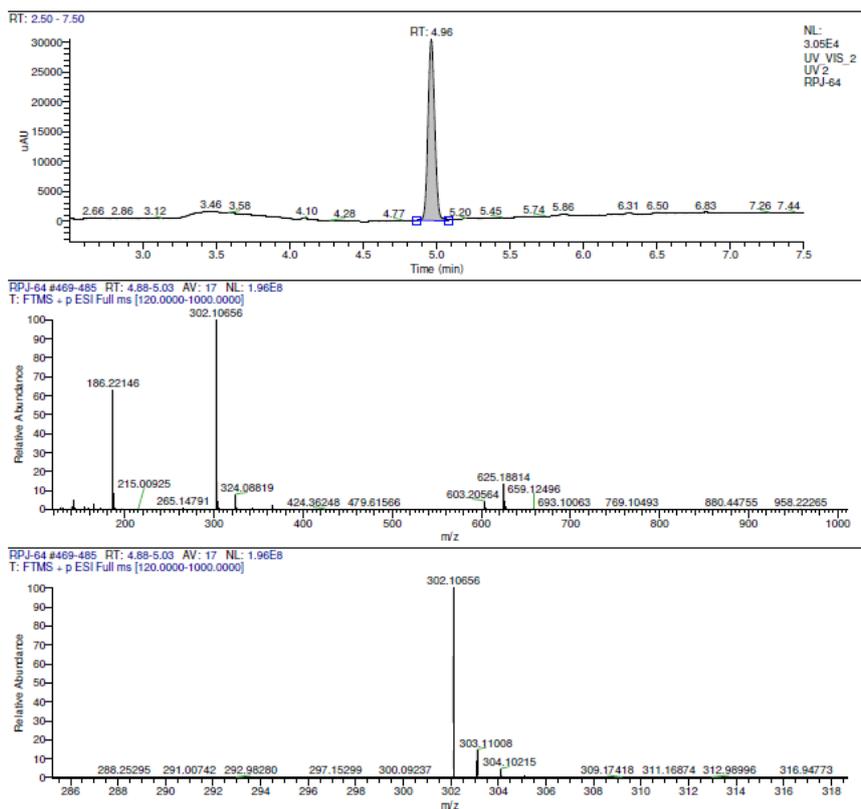
Supplementary Figure 31  $^1\text{H}$  NMR spectrum of 3.



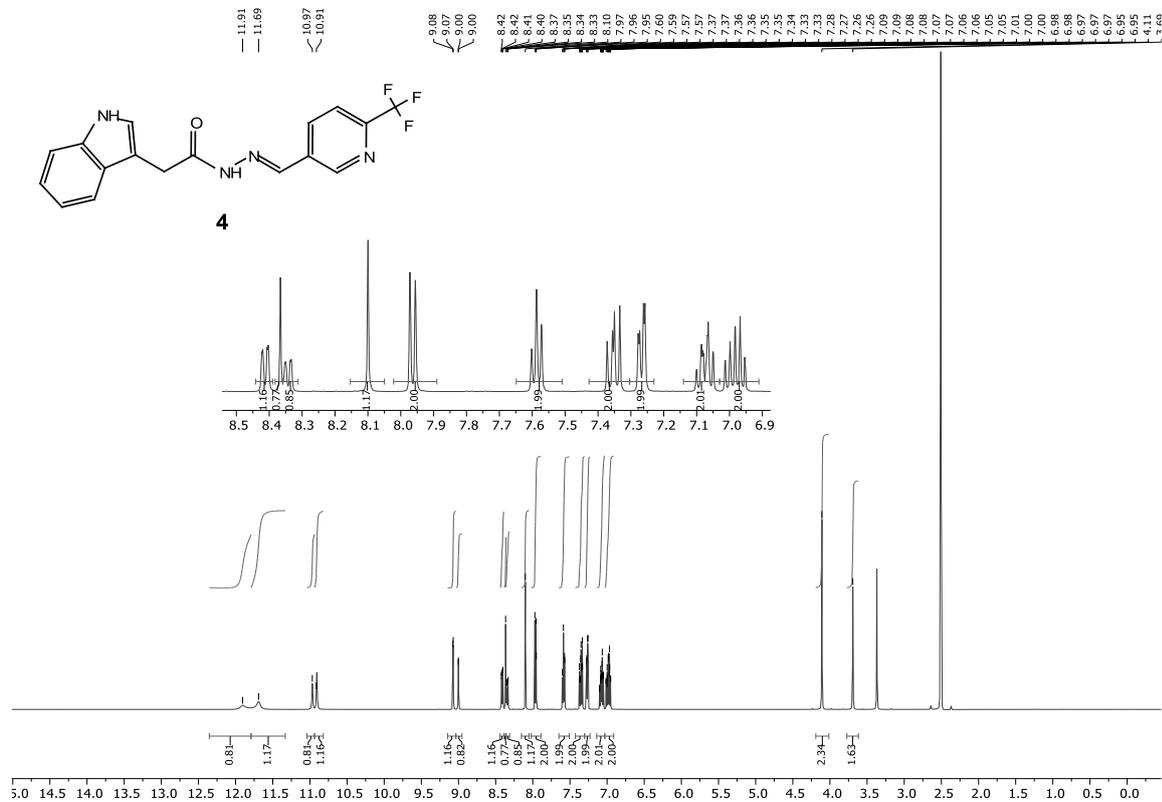
Supplementary Figure 32 <sup>13</sup>C NMR spectrum of 3.



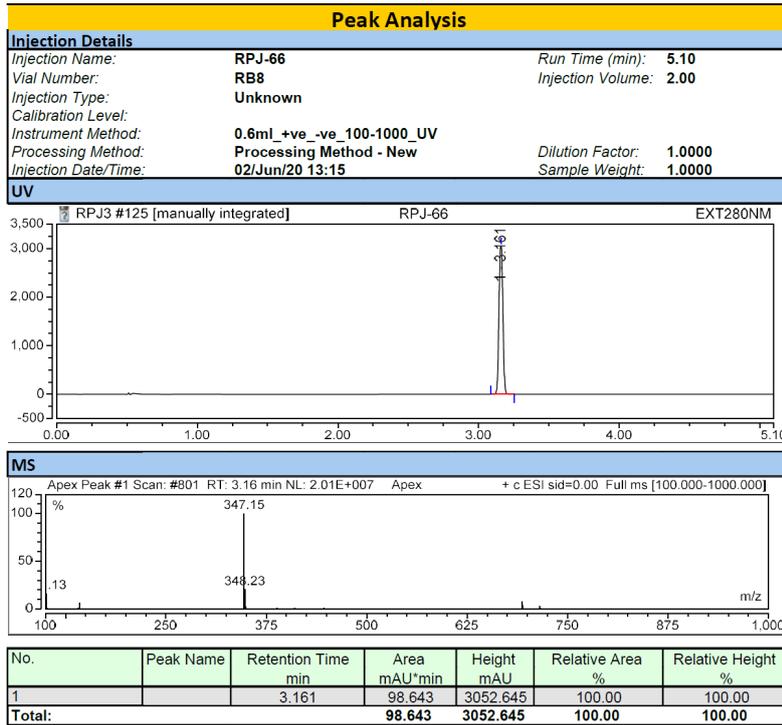
Supplementary Figure 33 HPLC purity analysis of 3.



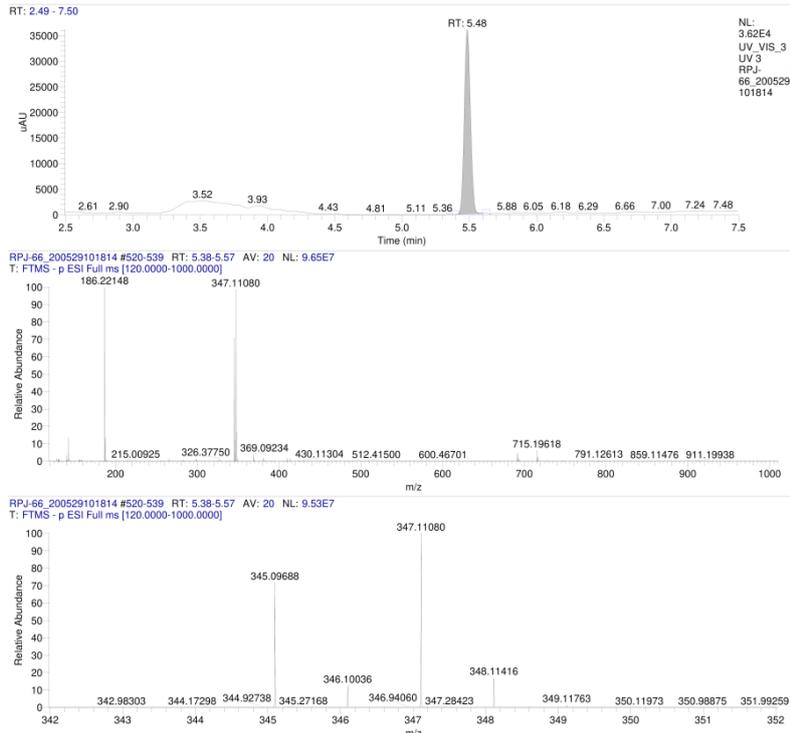
Supplementary Figure 34 HRMS of 3.

Supplementary Figure 35 <sup>1</sup>H NMR spectrum of 4.

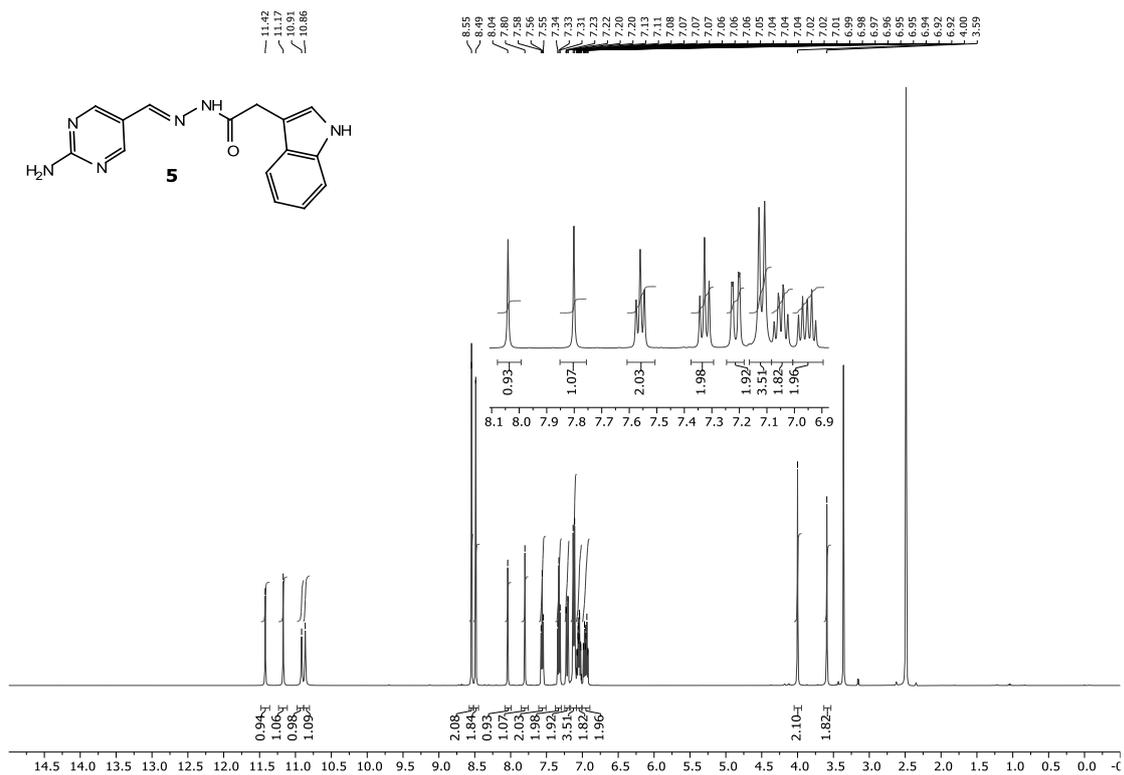
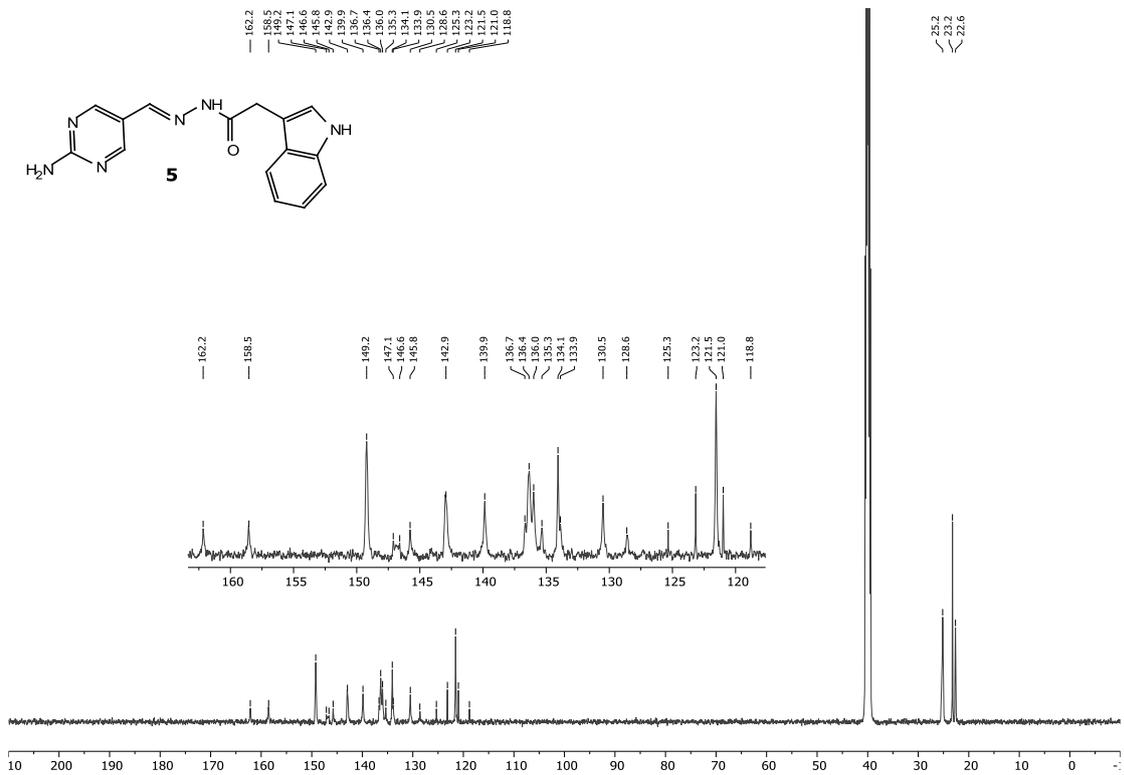


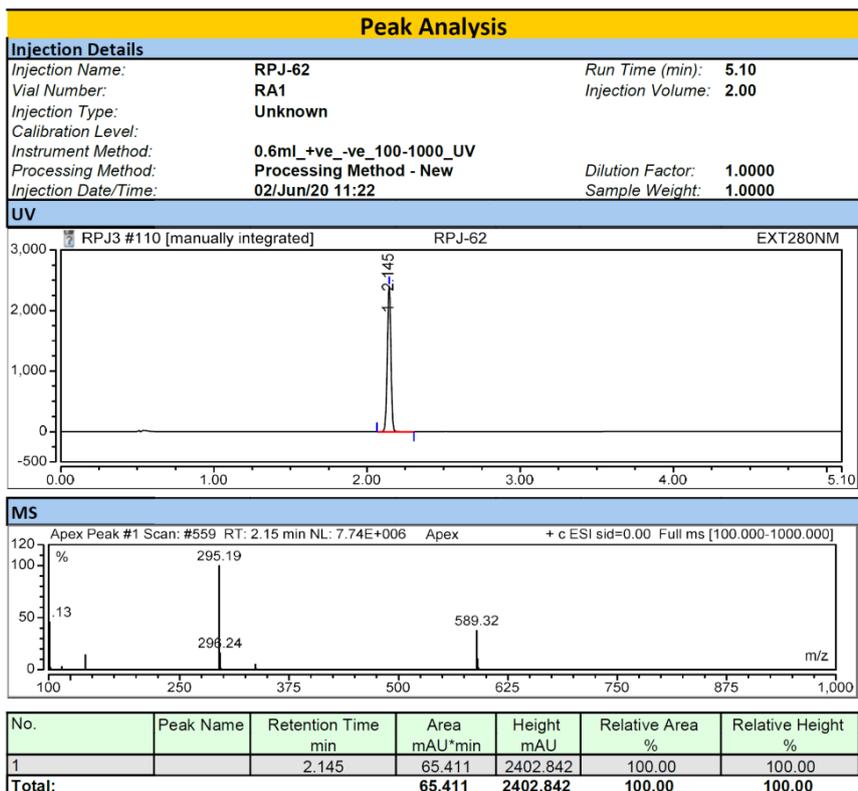


Supplementary Figure 38 HPLC purity analysis of 4.

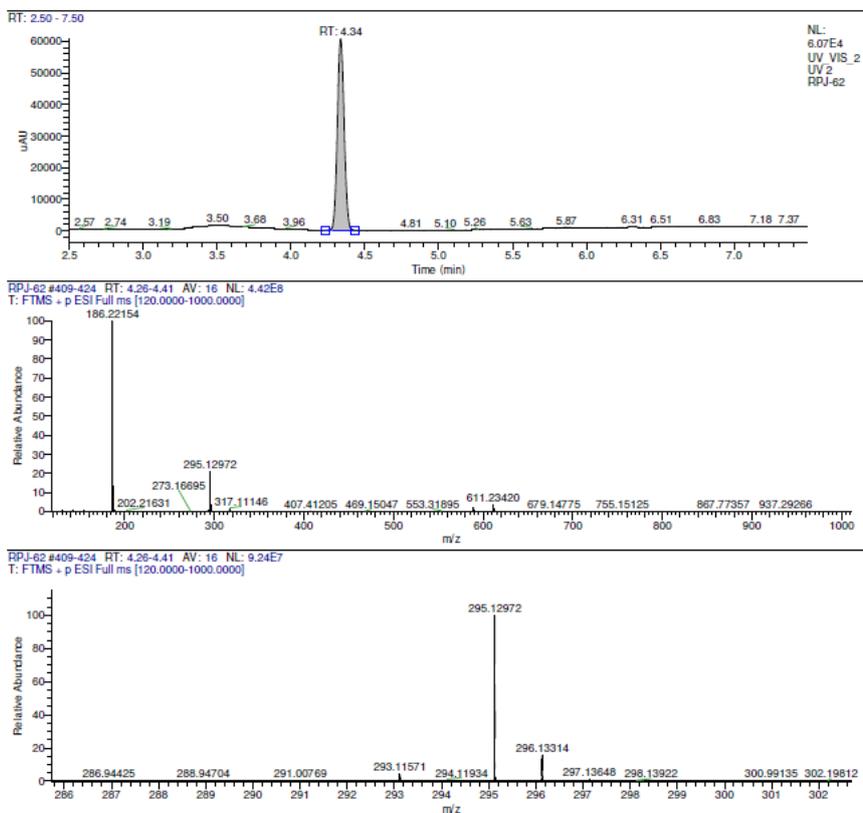


Supplementary Figure 39 HRMS of 4.

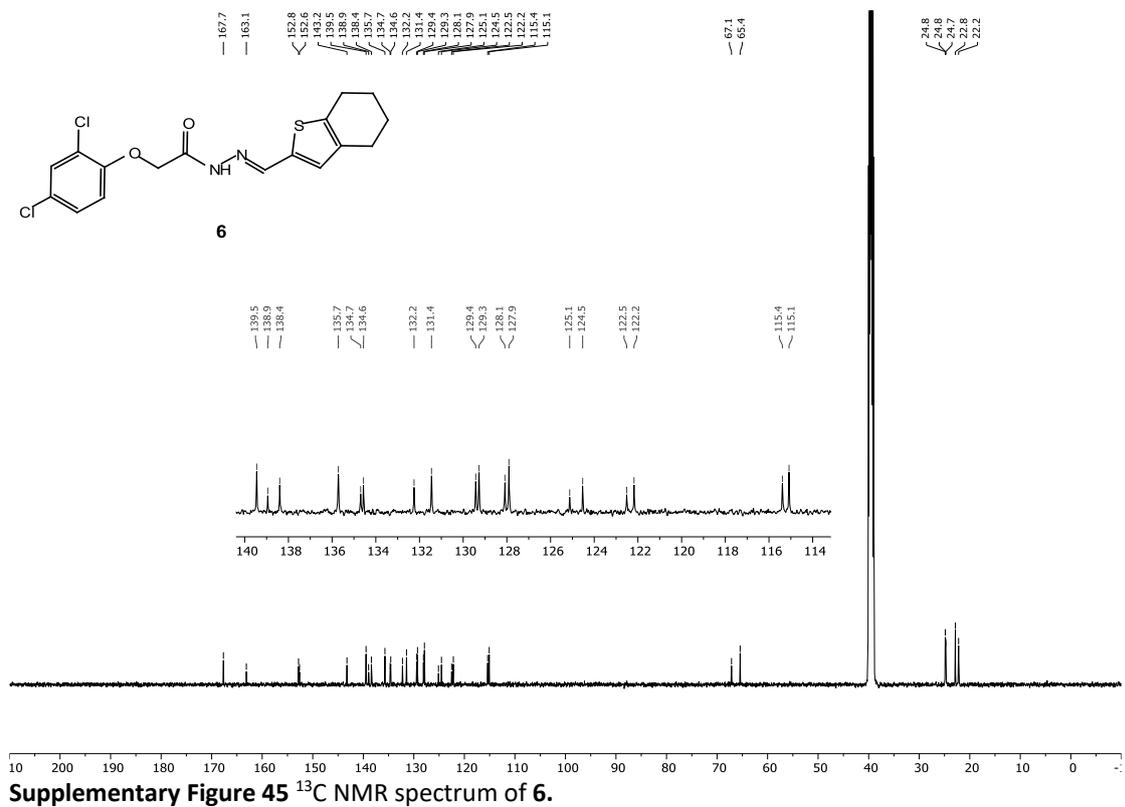
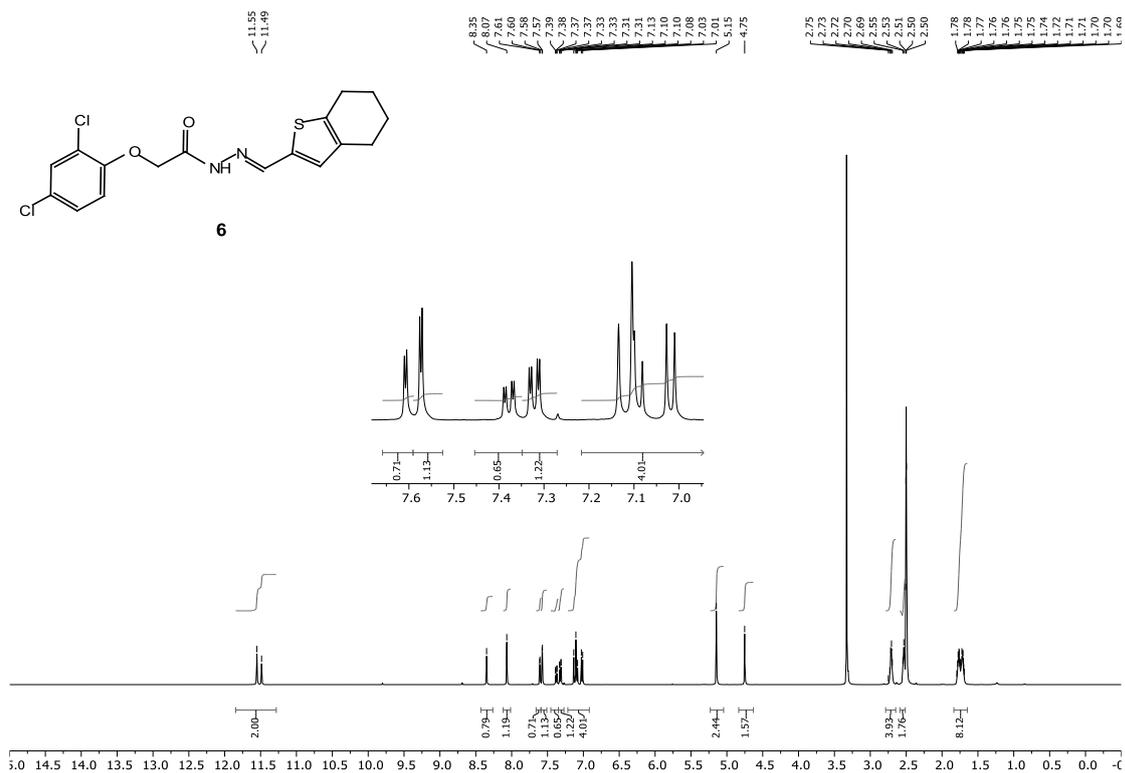
Supplementary Figure 40  $^1\text{H}$  NMR spectrum of **5**.Supplementary Figure 41  $^{13}\text{C}$  NMR spectrum of **5**.

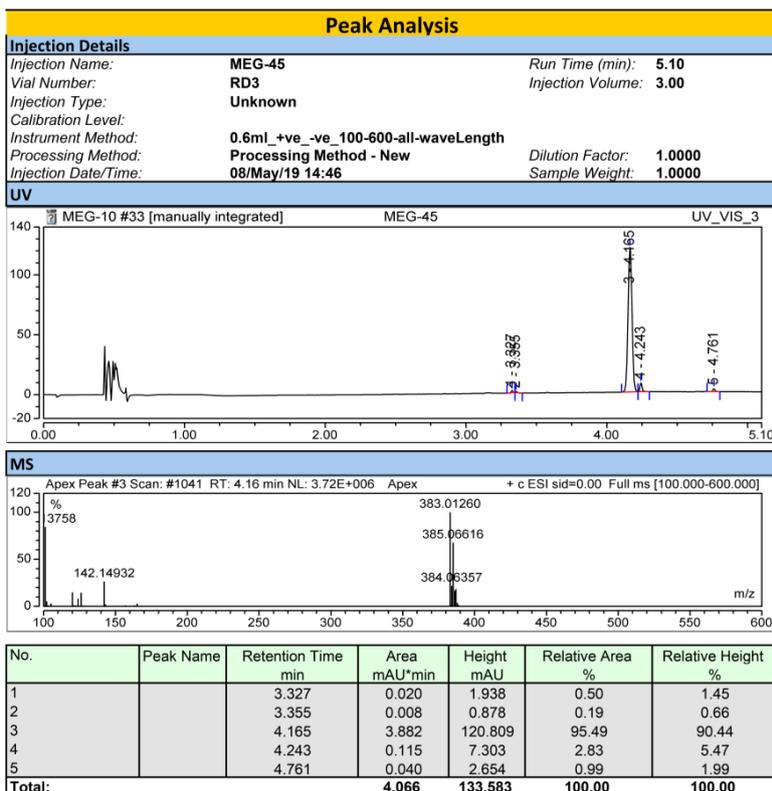


Supplementary Figure 42 HPLC purity analysis of 5.

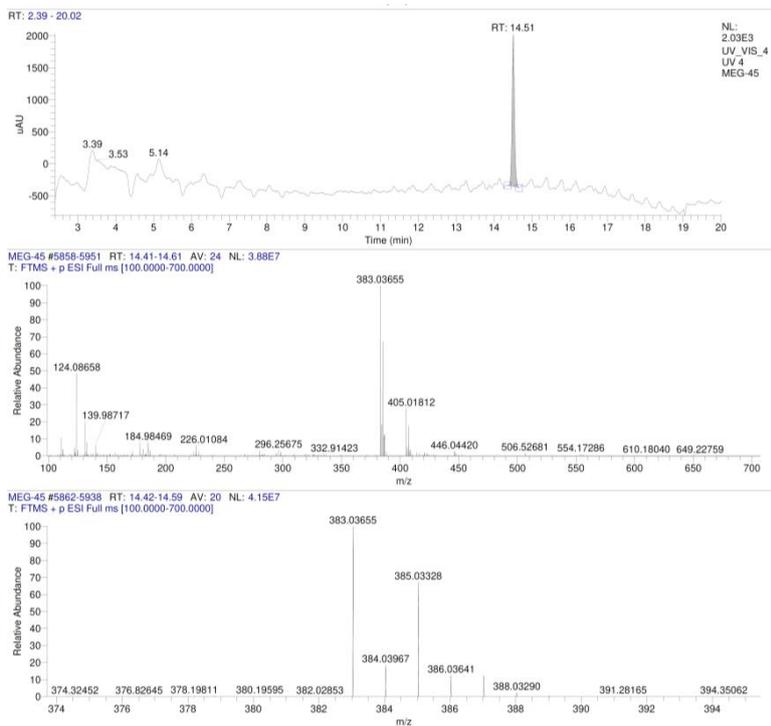


Supplementary Figure 43 HRMS of 5.

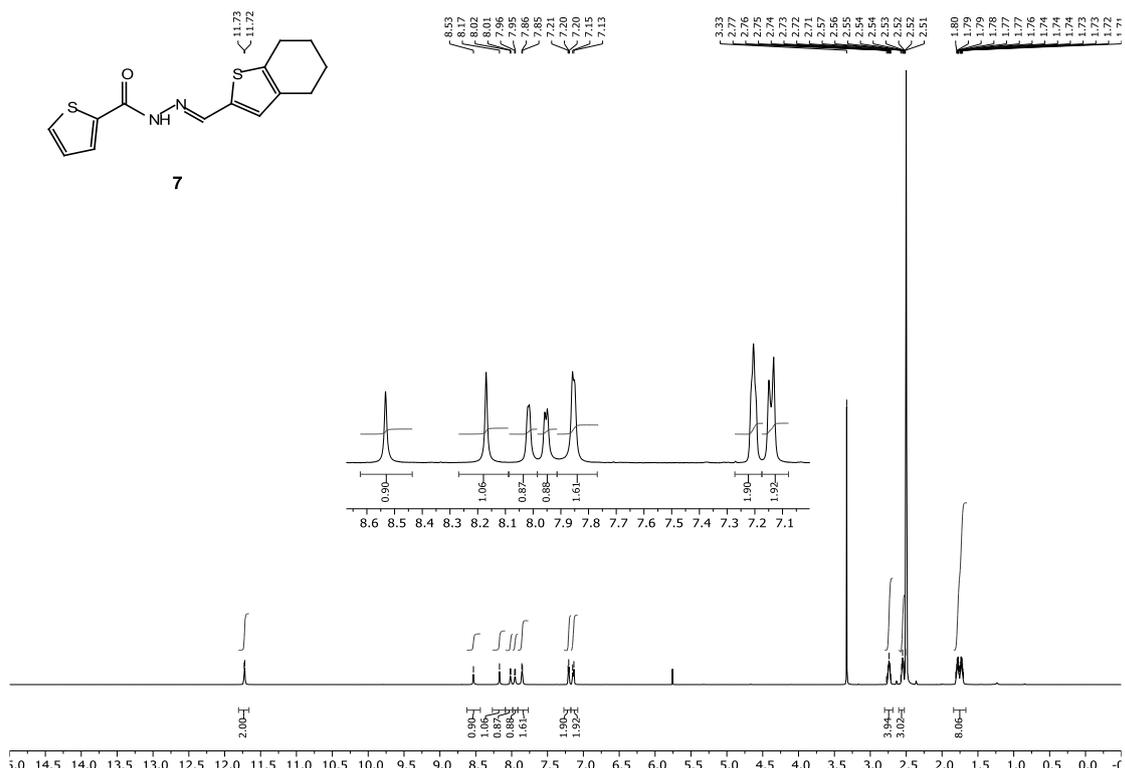




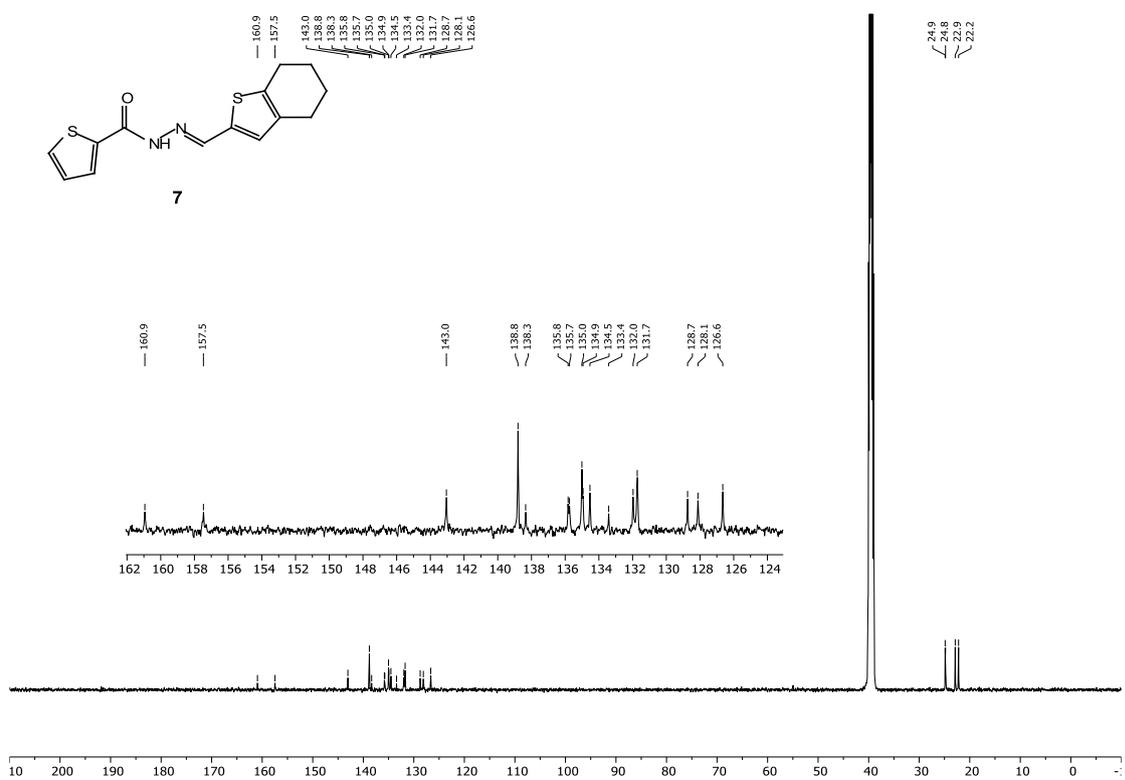
Supplementary Figure 46 HPLC purity analysis of 6.



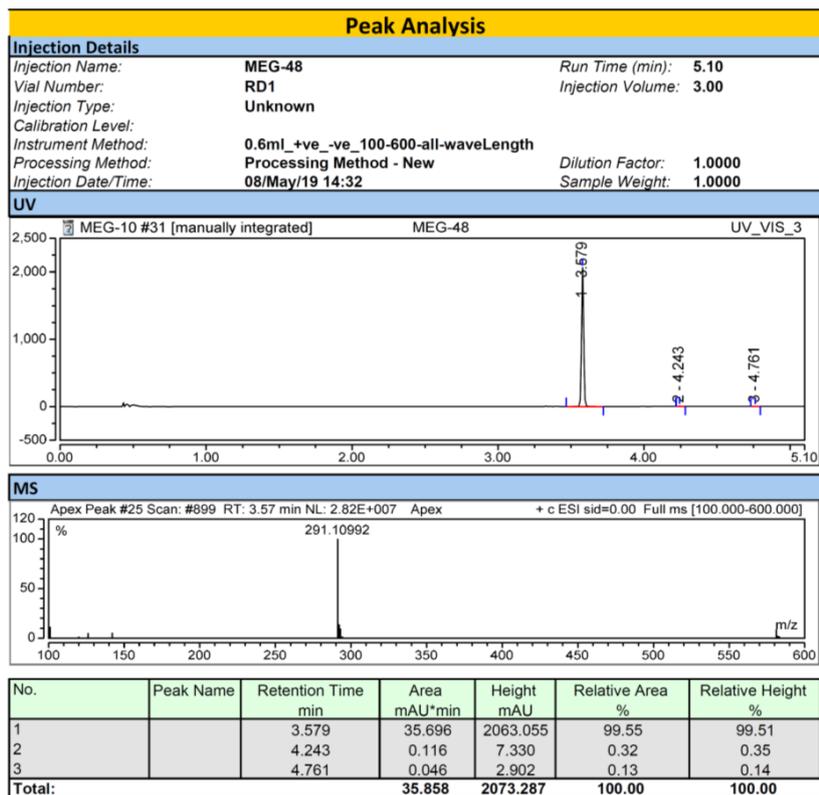
Supplementary Figure 47 HRMS of 6.



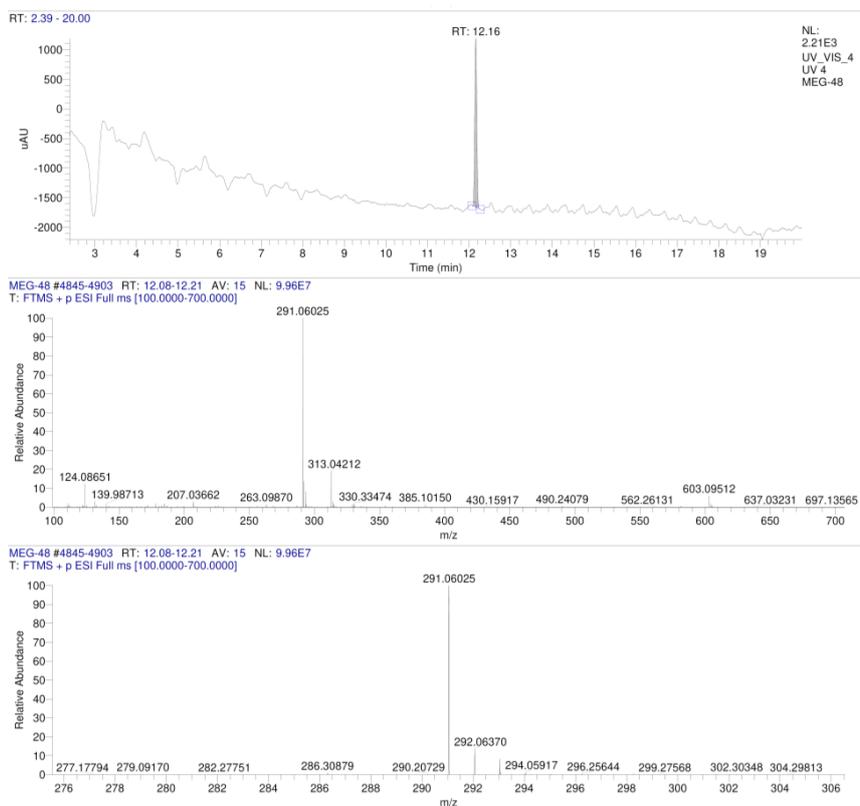
**Supplementary Figure 48**  $^1\text{H}$  NMR spectrum of **7**.



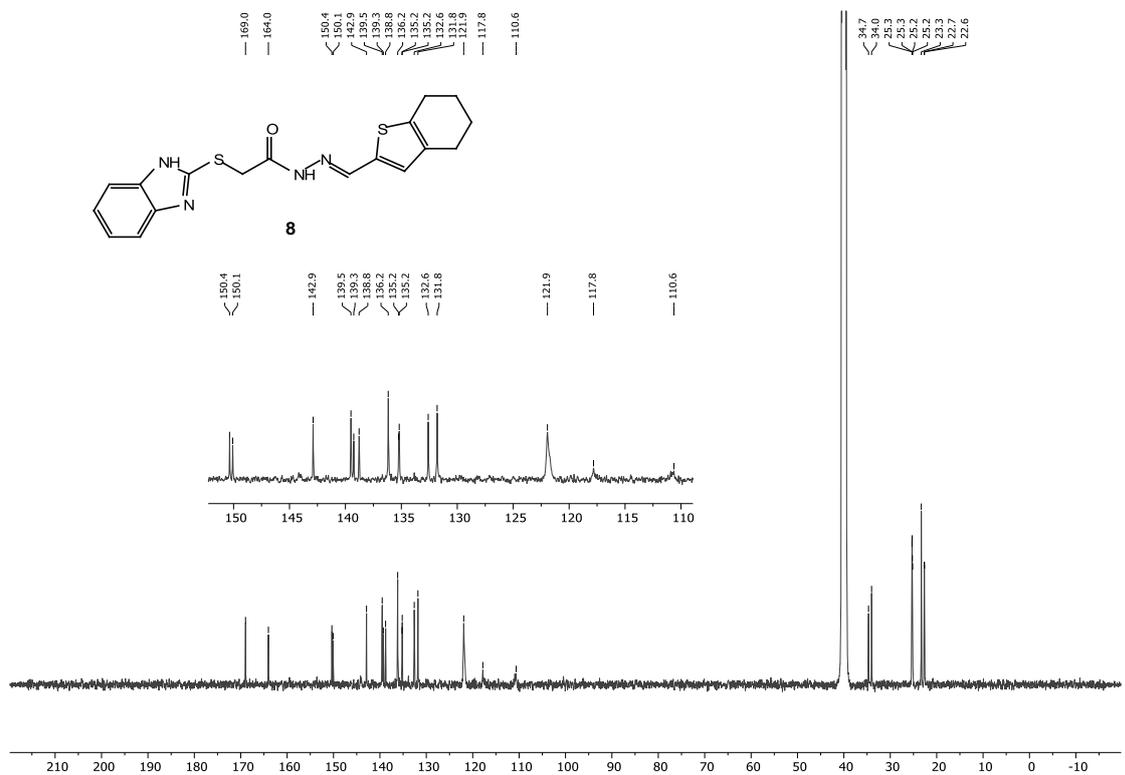
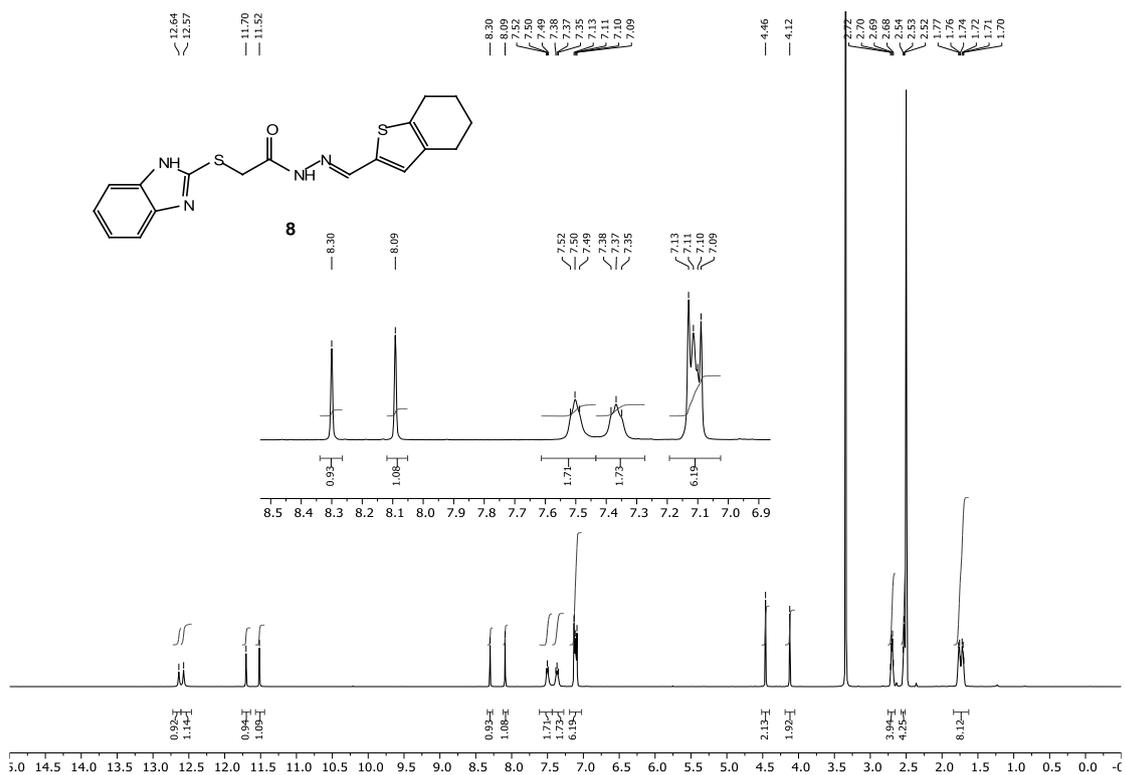
**Supplementary Figure 49**  $^{13}\text{C}$  NMR spectrum of **7**.

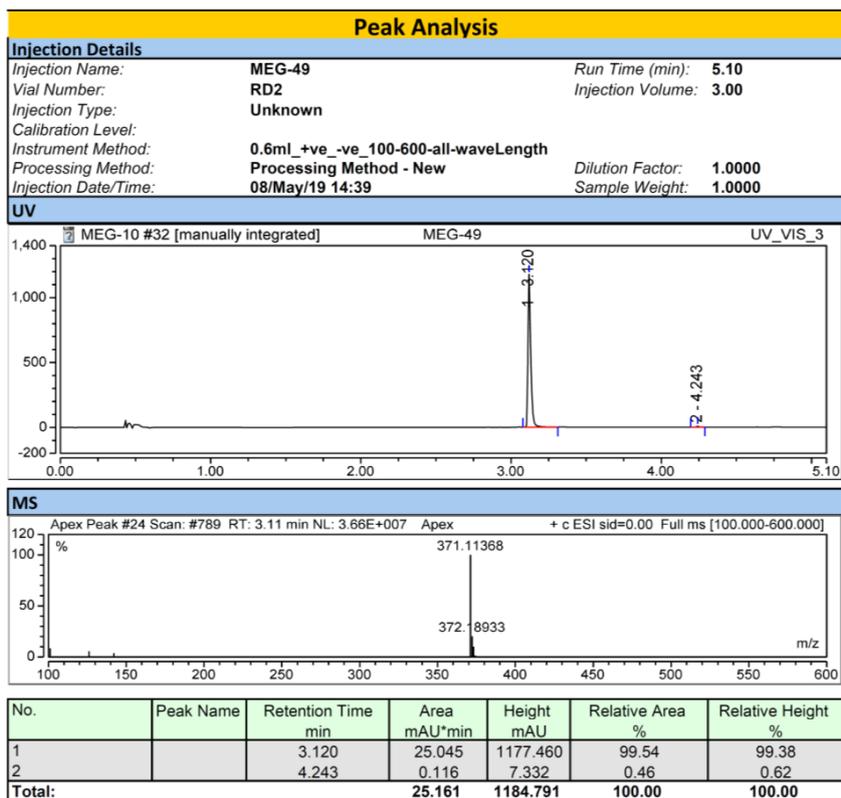


Supplementary Figure 50 HPLC purity analysis of 7.

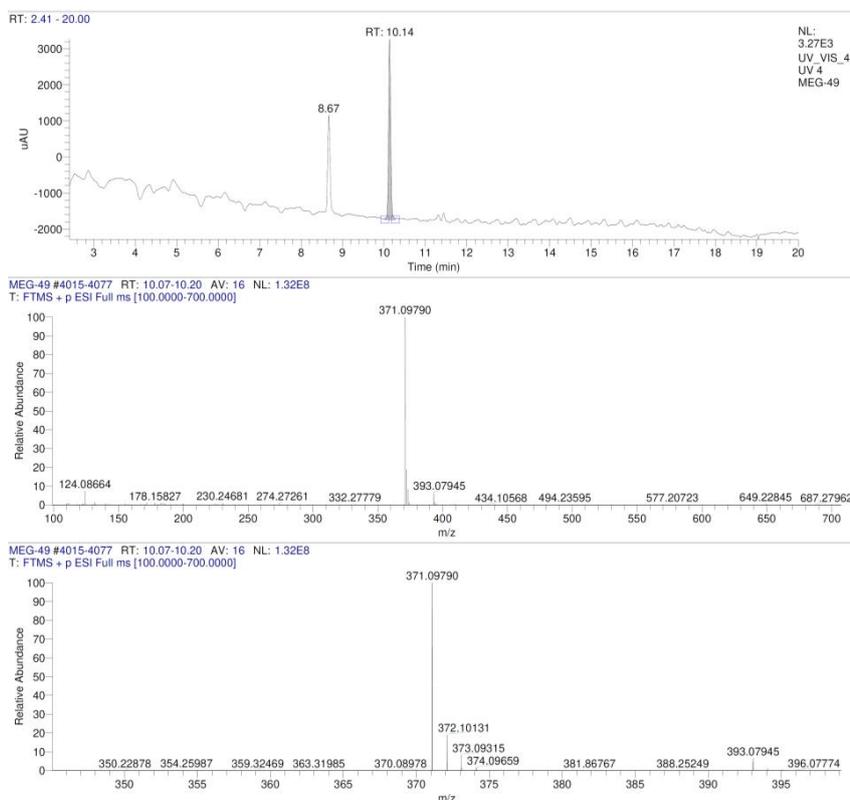


Supplementary Figure 51 HRMS of 7.

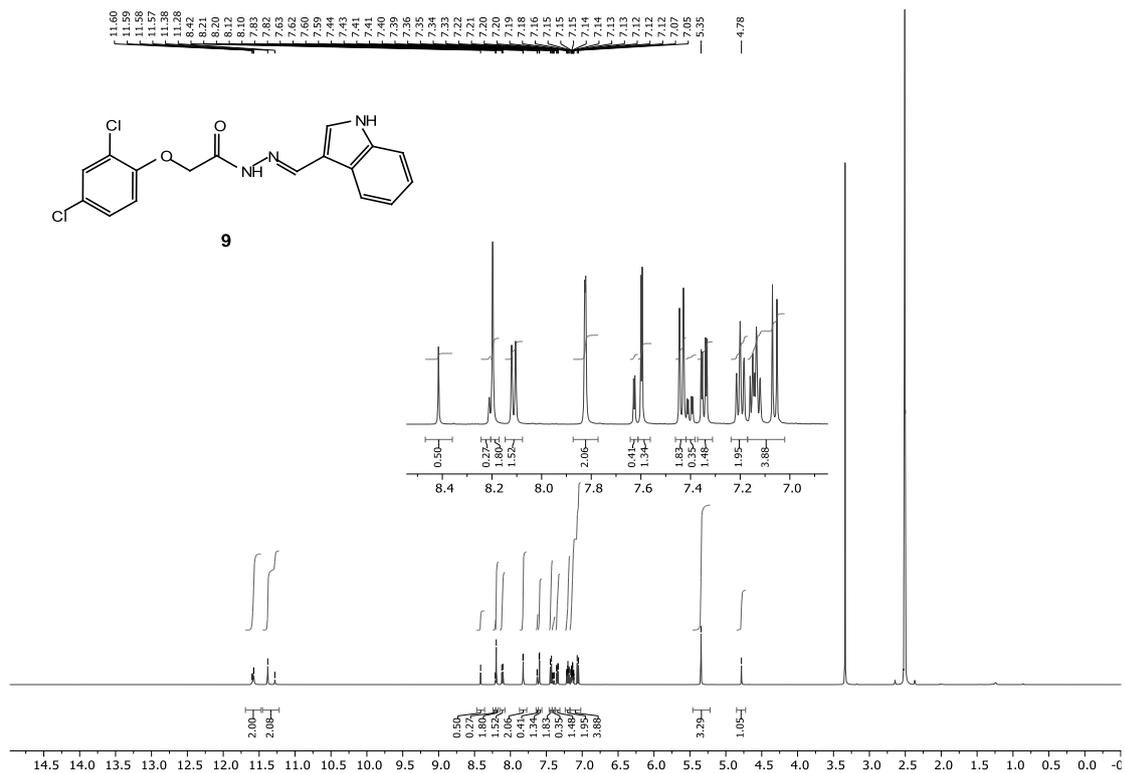




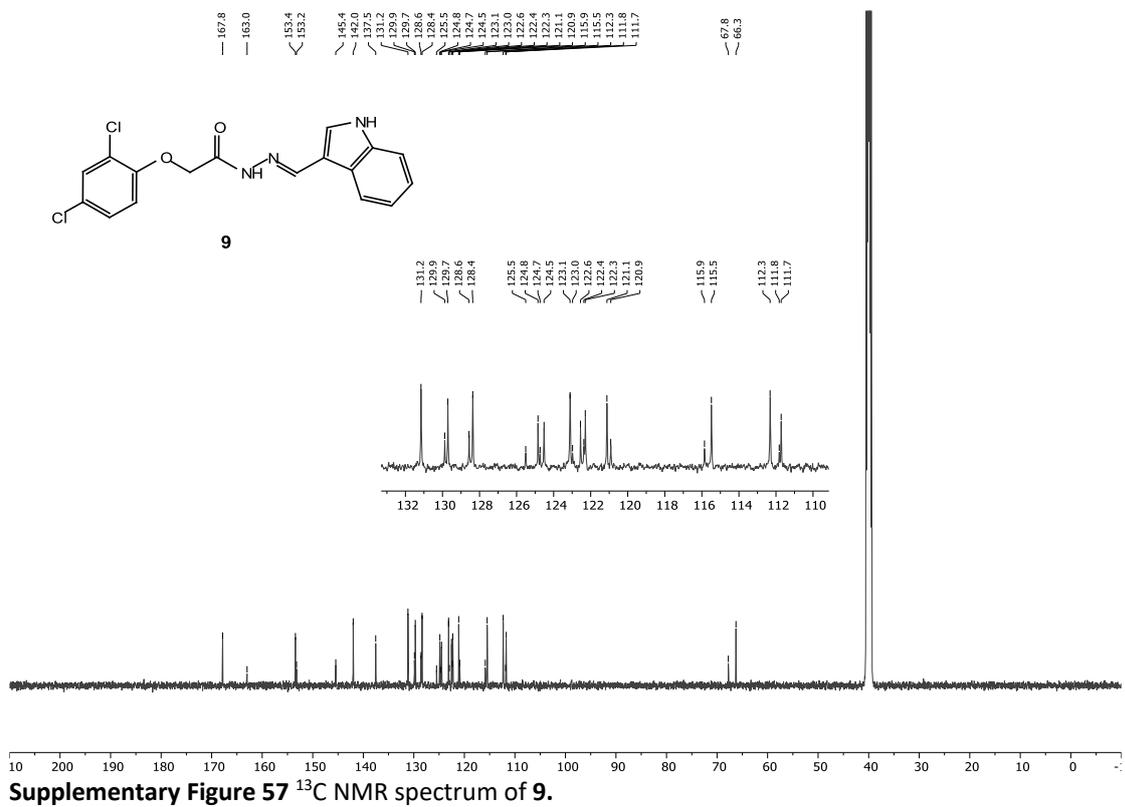
Supplementary Figure 54 HPLC purity analysis of 8.



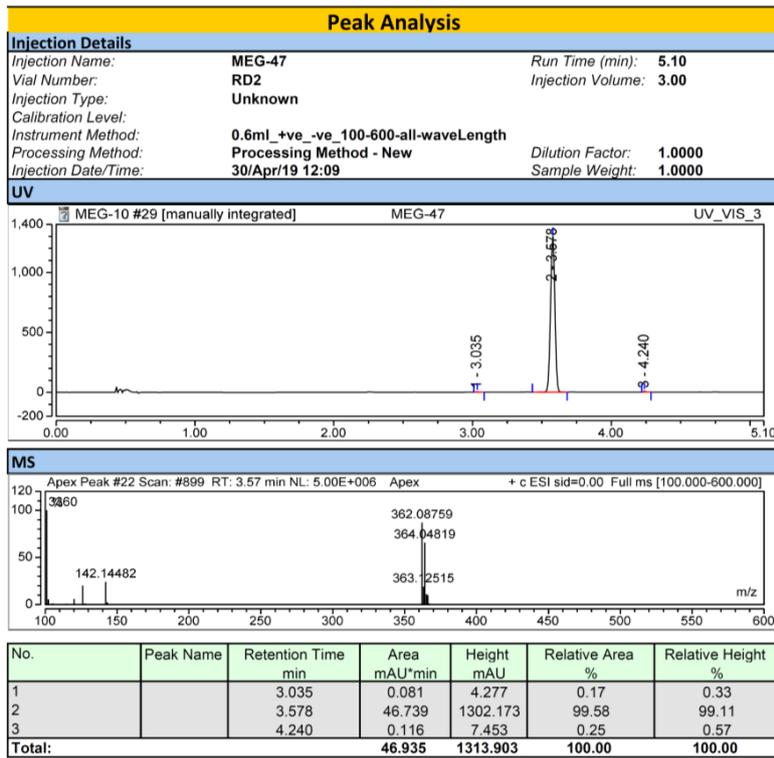
Supplementary Figure 55 HRMS of 8.



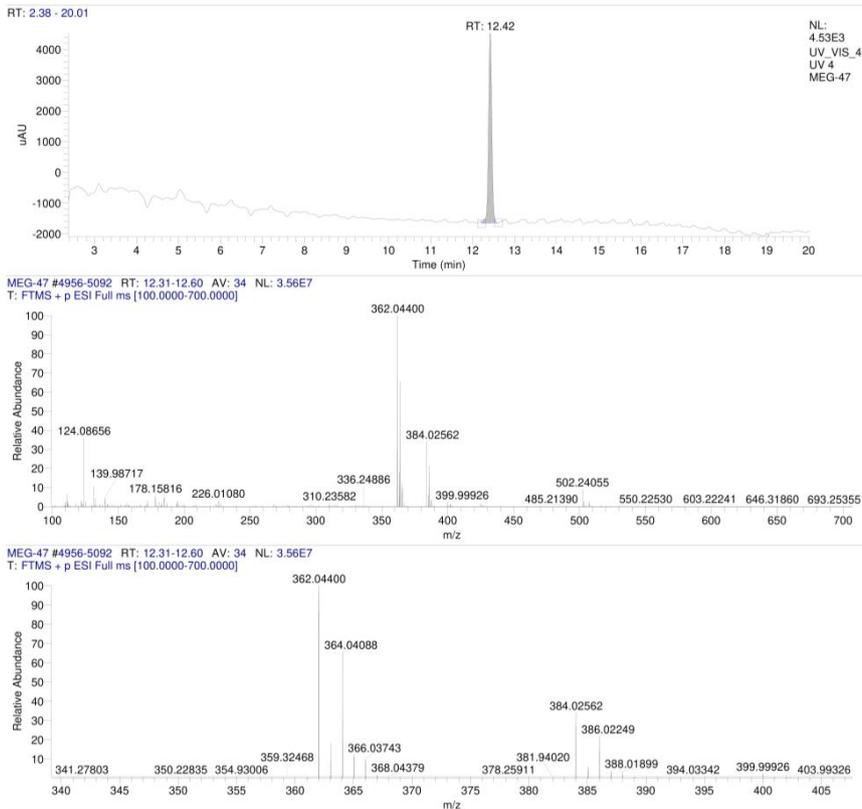
**Supplementary Figure 56**  $^1\text{H}$  NMR spectrum of **9**.



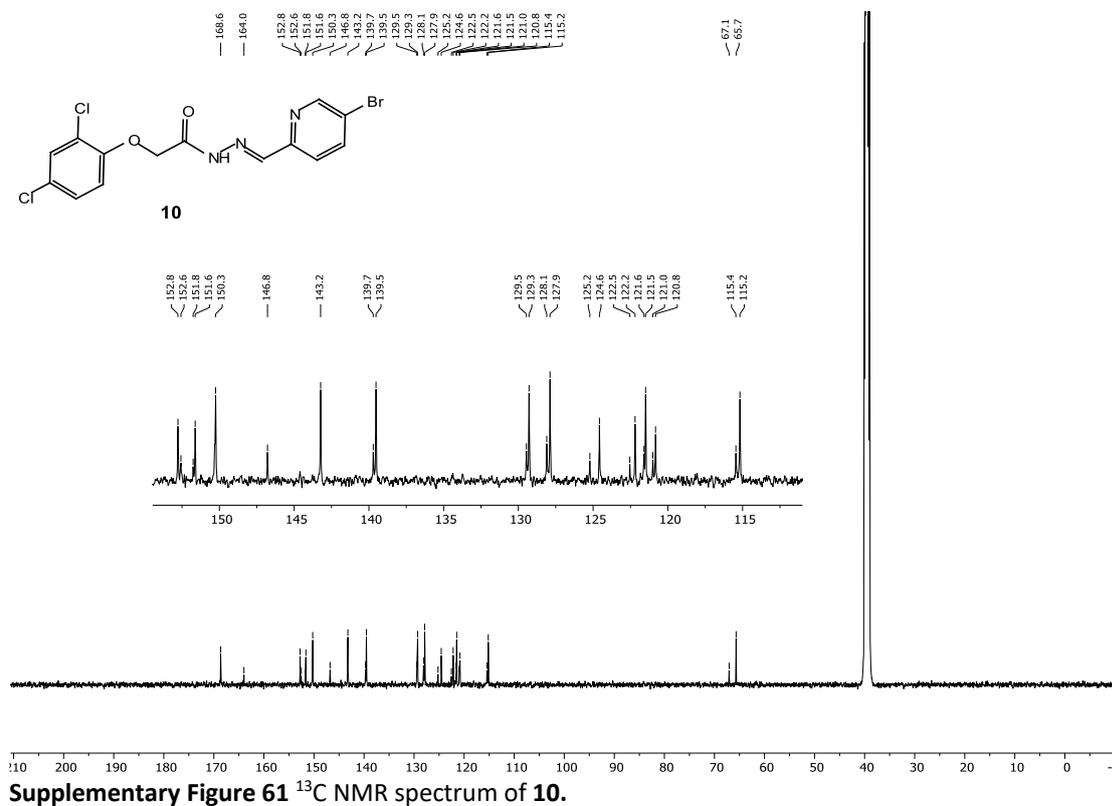
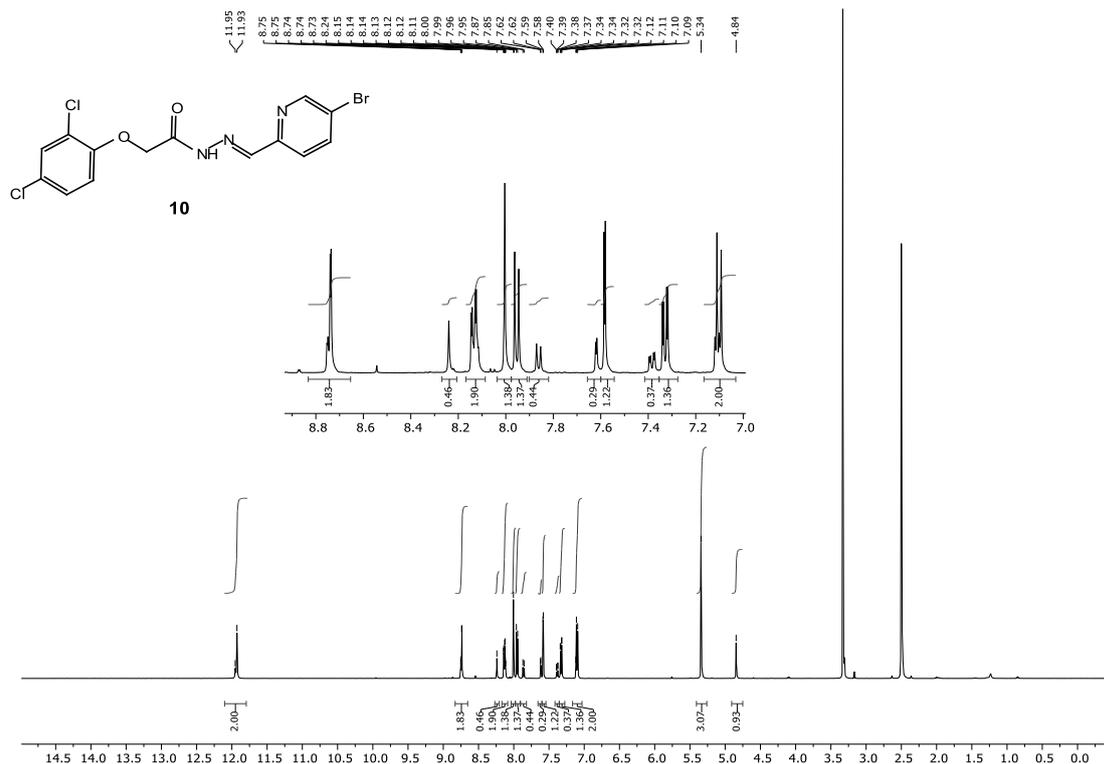
**Supplementary Figure 57**  $^{13}\text{C}$  NMR spectrum of **9**.

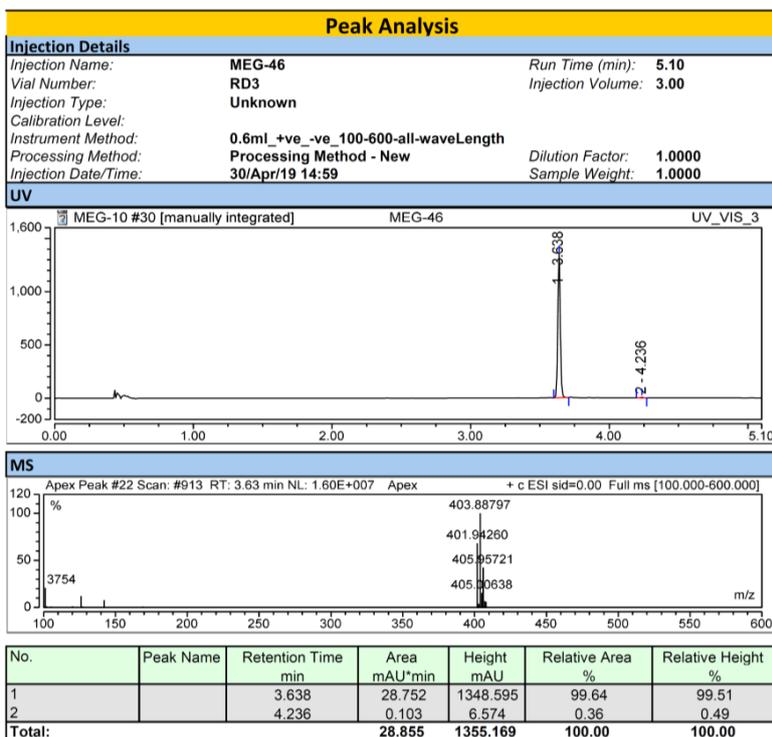


Supplementary Figure 58 HPLC purity analysis of 9.

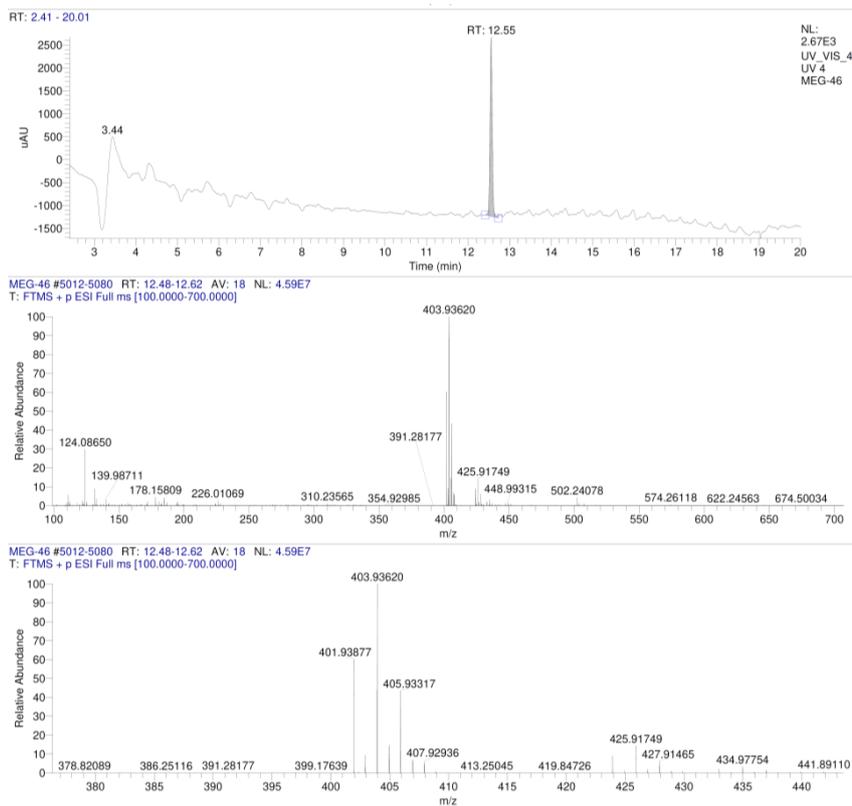


Supplementary Figure 59 HRMS of 9.

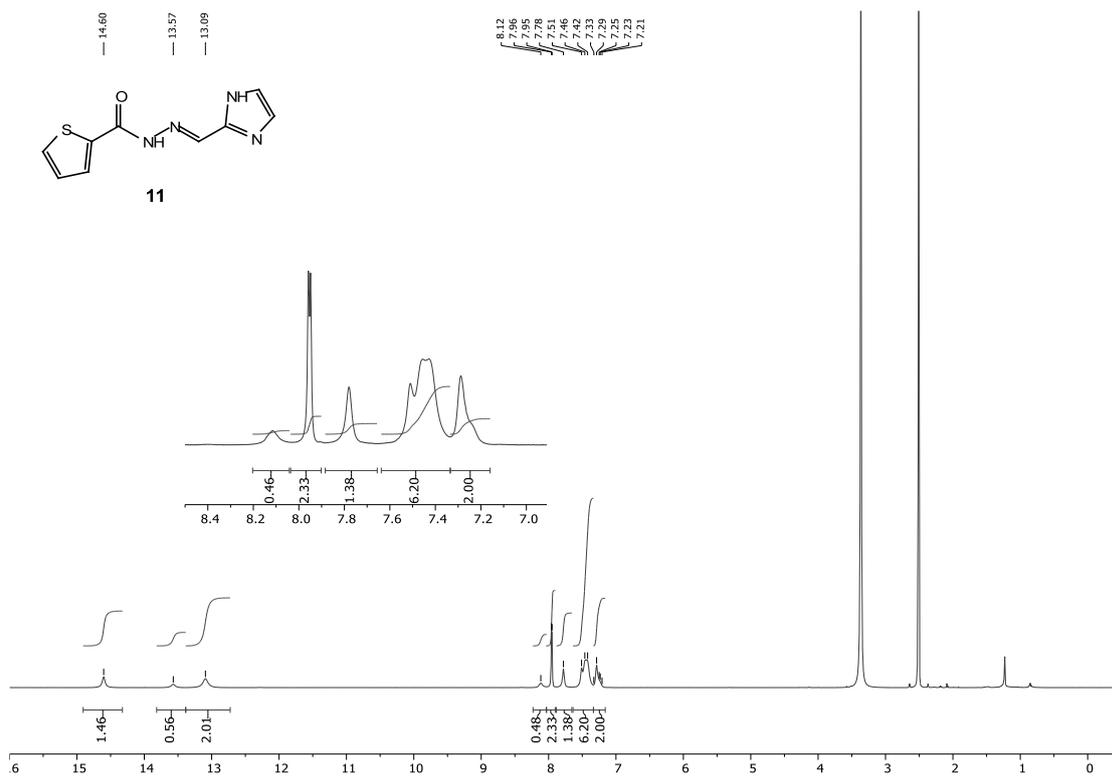
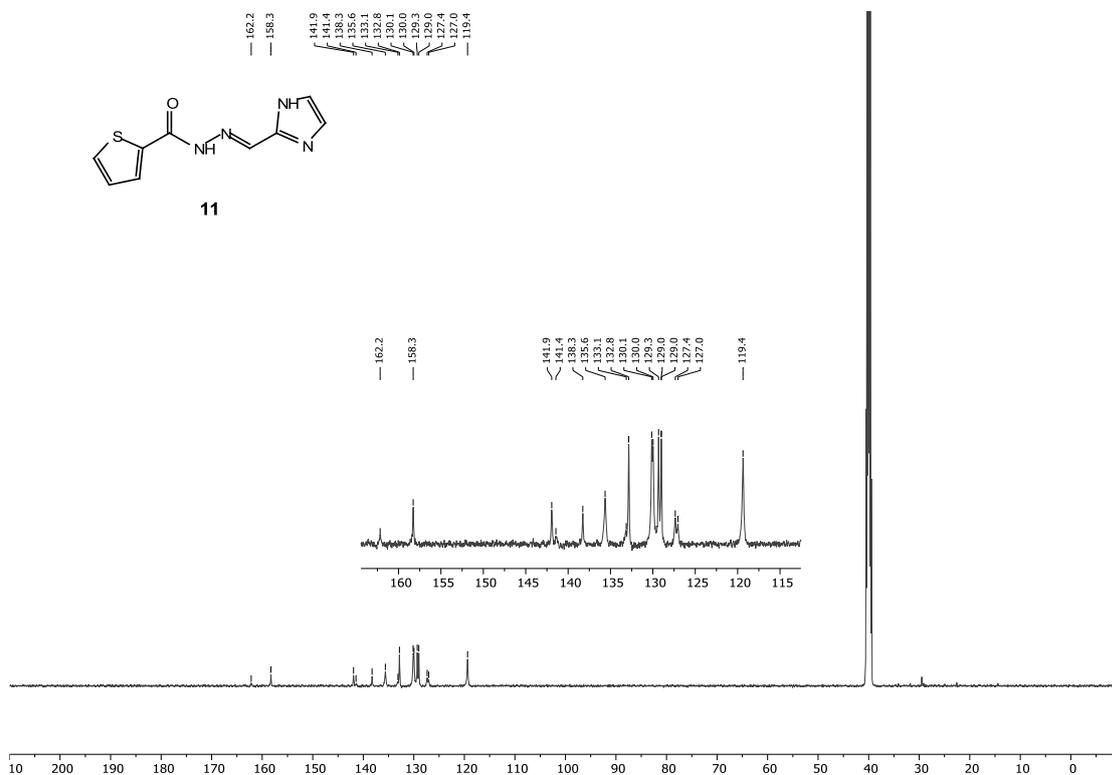


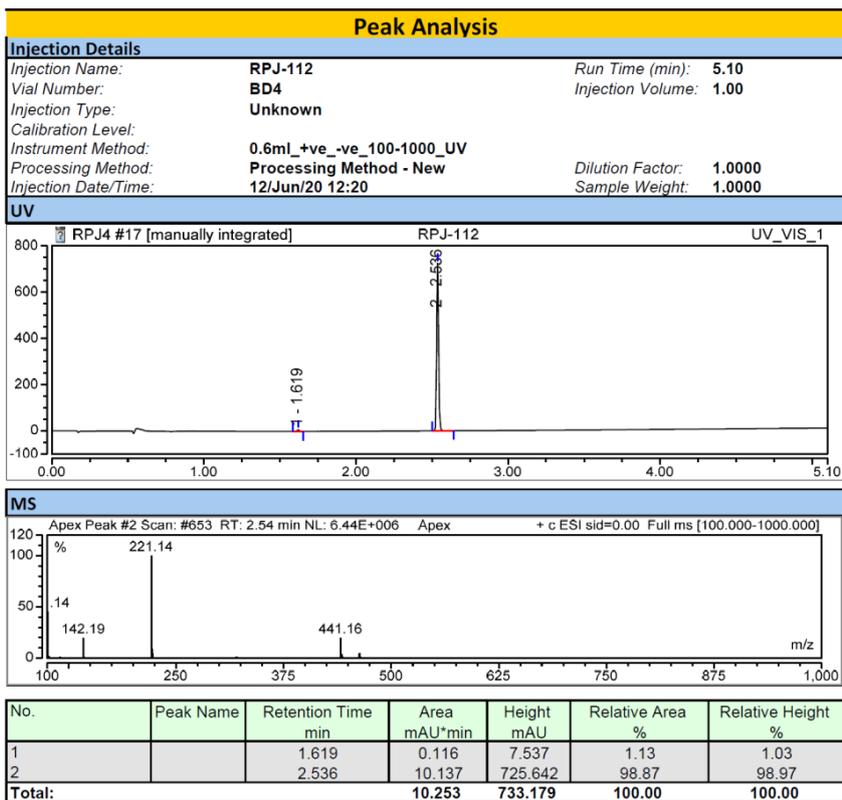


Supplementary Figure 62 HPLC purity analysis of 10.

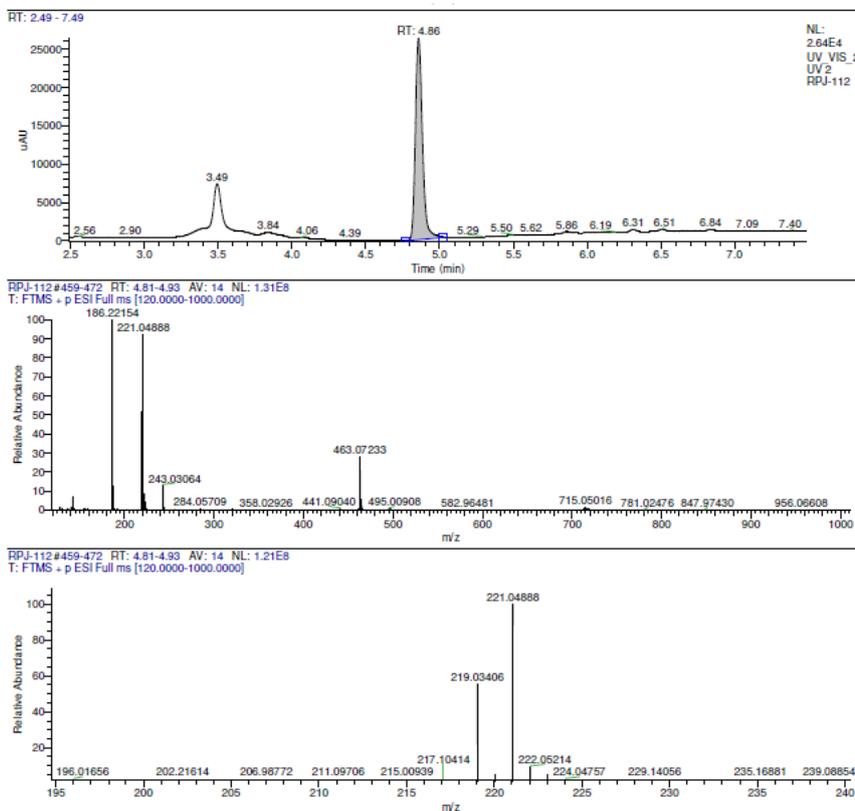


Supplementary Figure 63 HRMS of 10.

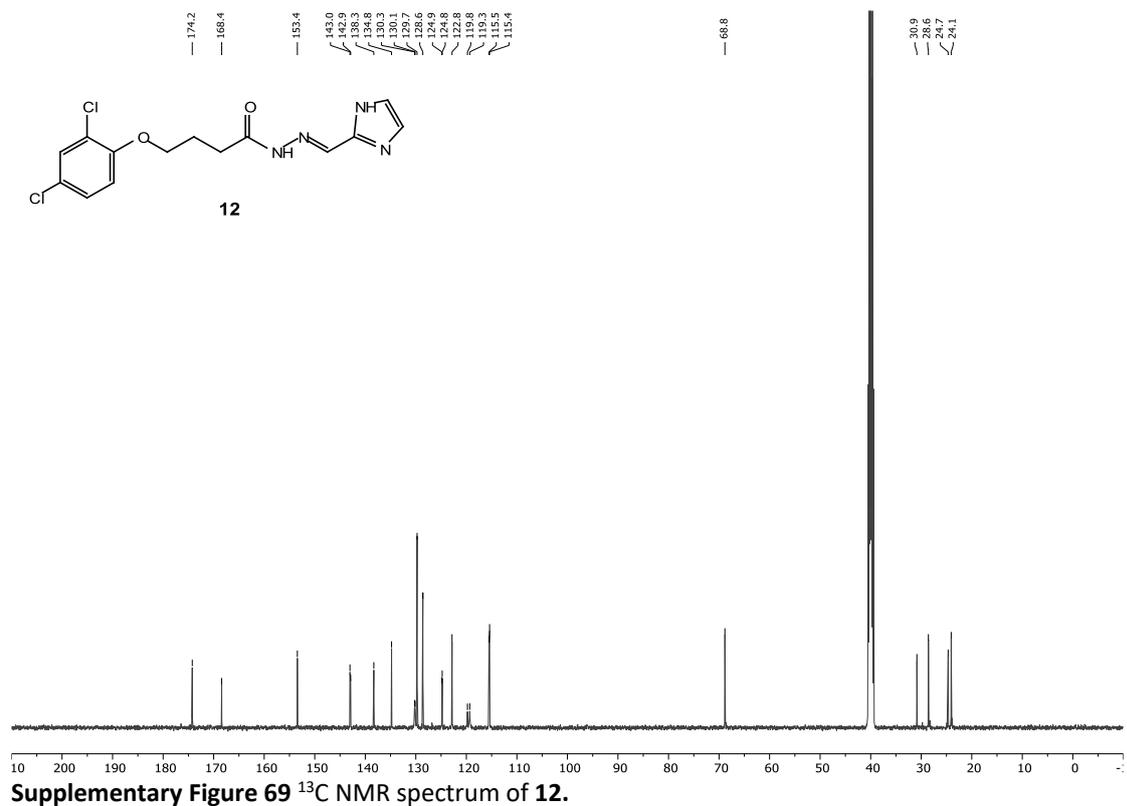
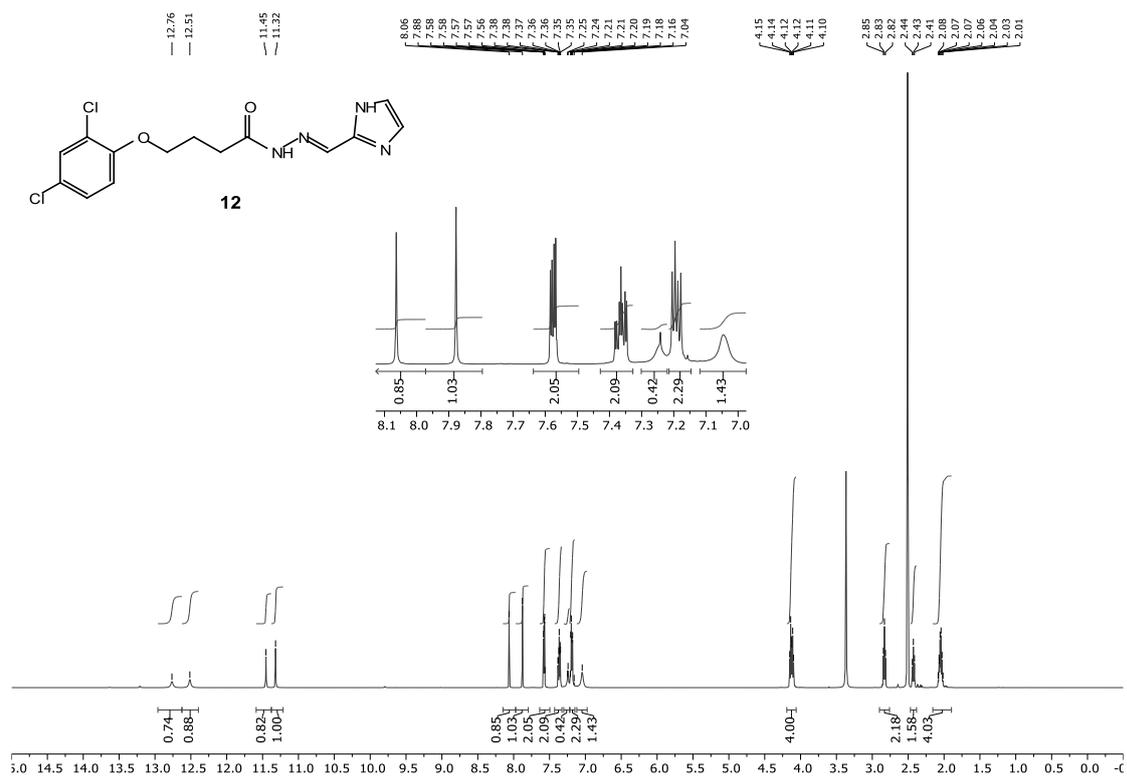
Supplementary Figure 64  $^1\text{H}$  NMR spectrum of **11**.Supplementary Figure 65  $^{13}\text{C}$  NMR spectrum of **11**.

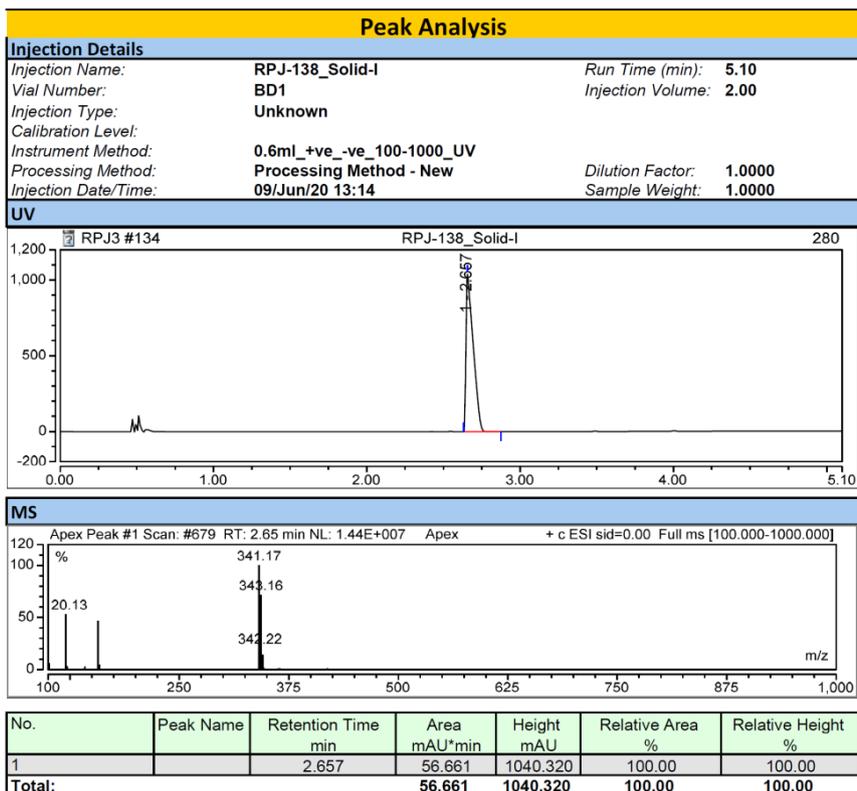


Supplementary Figure 66 HPLC purity analysis of 11.

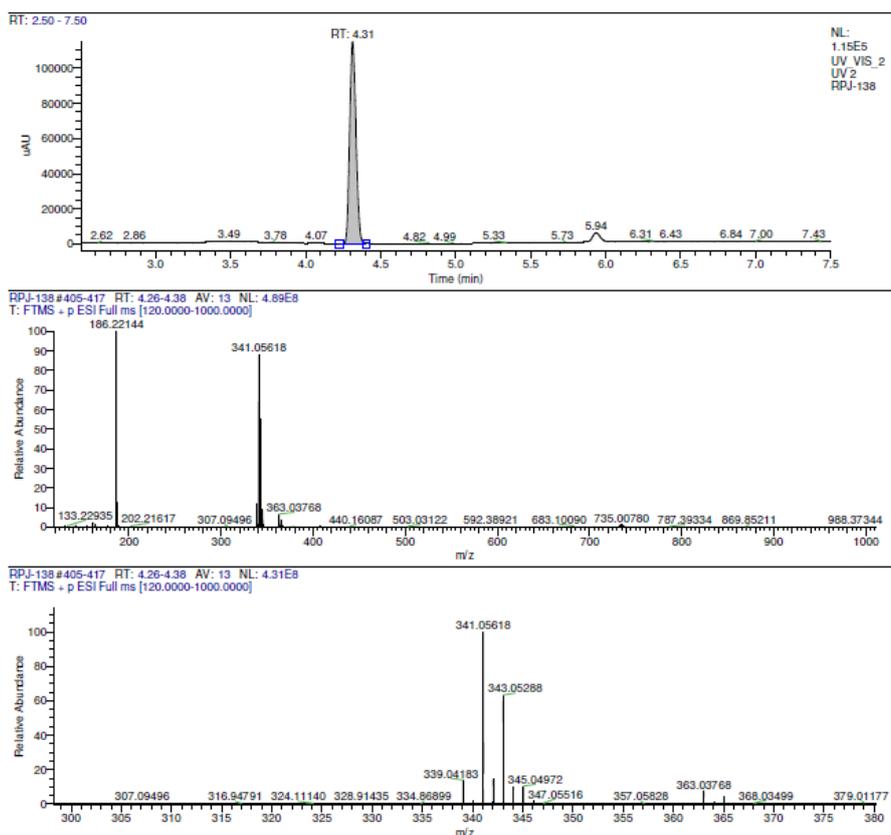


Supplementary Figure 67 HRMS of 11.

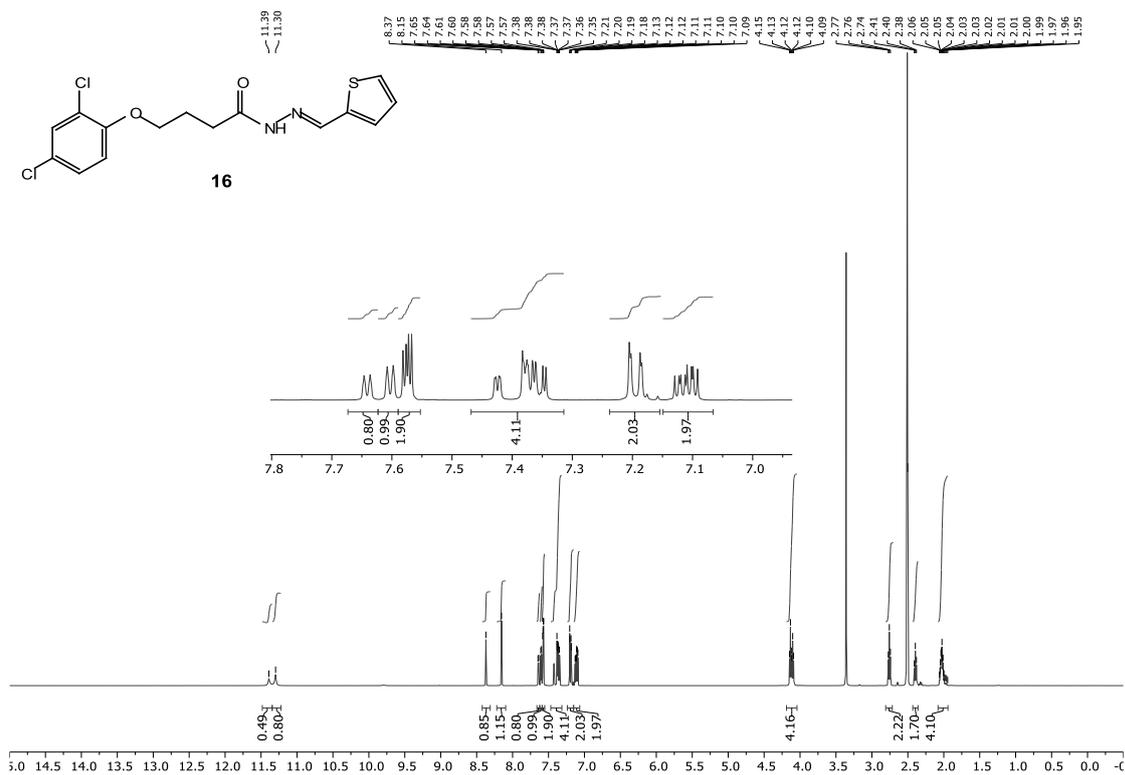
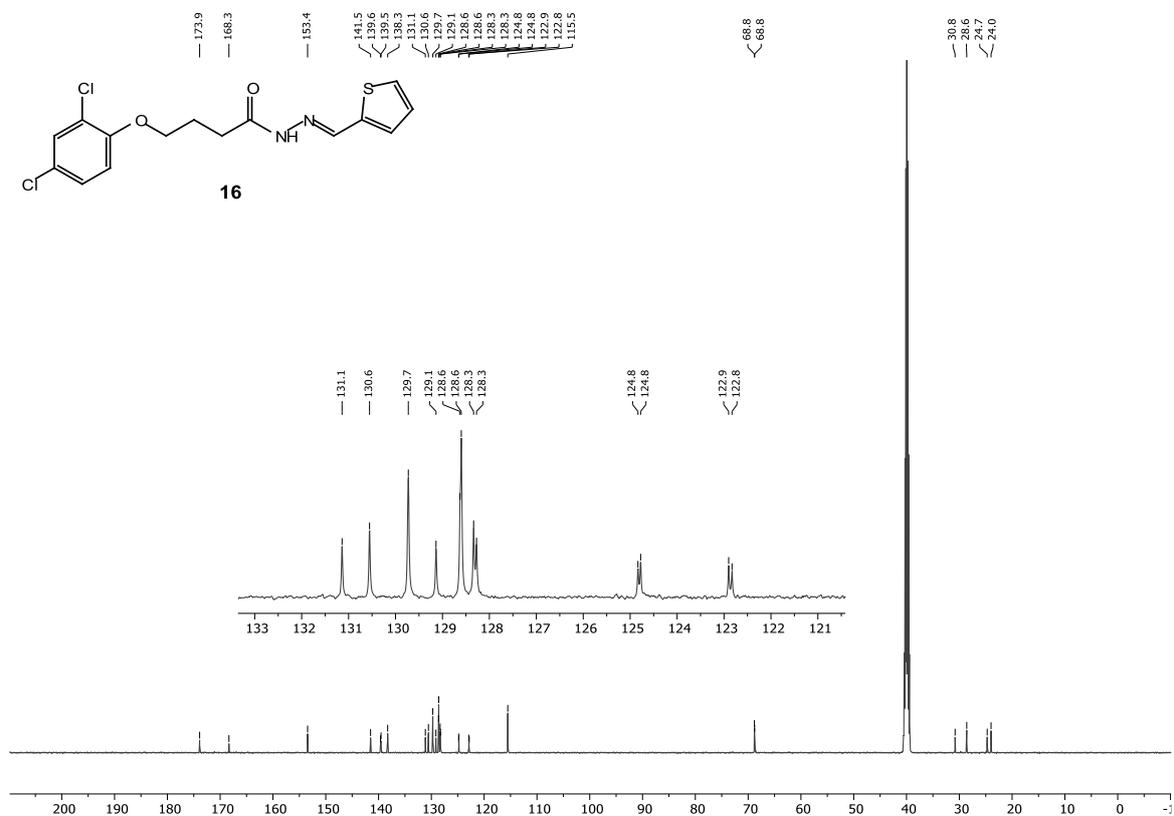


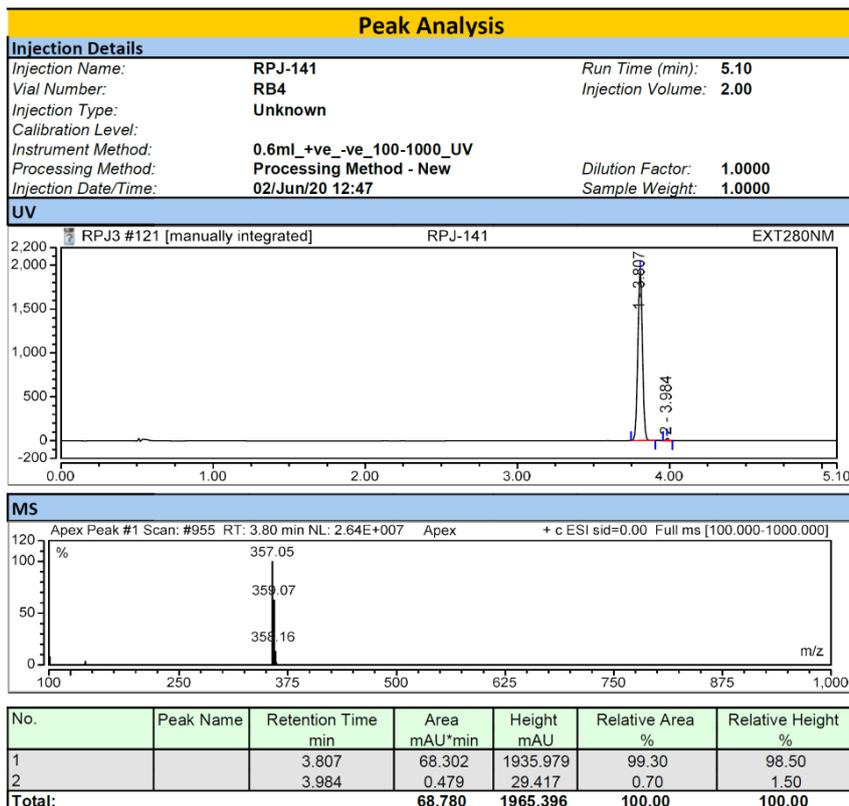


Supplementary Figure 70 HPLC purity analysis of 12.

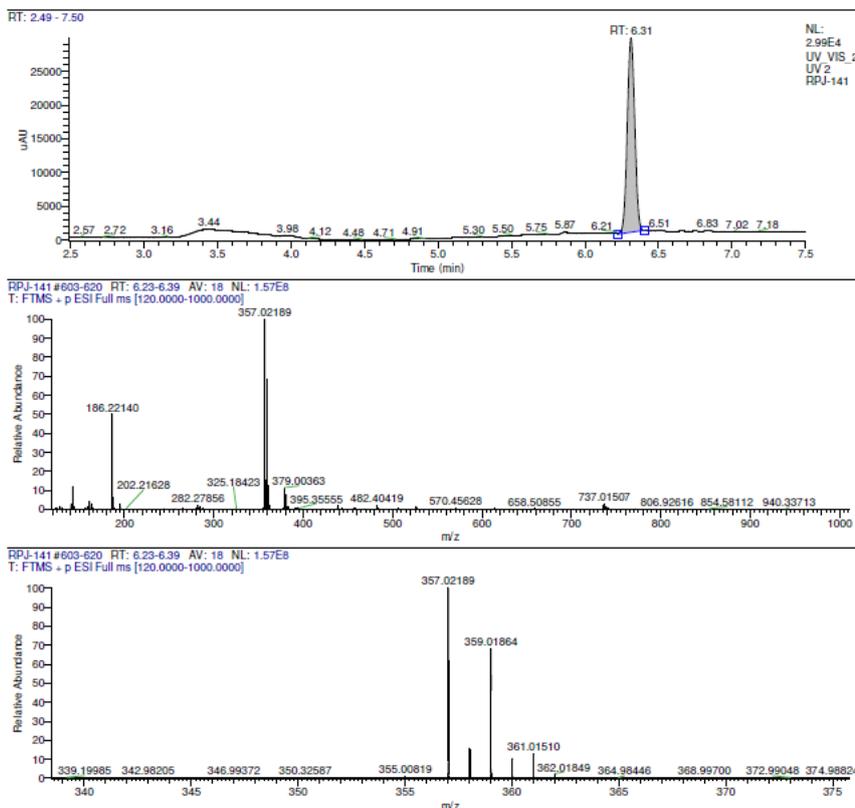


Supplementary Figure 71 HRMS of 12.

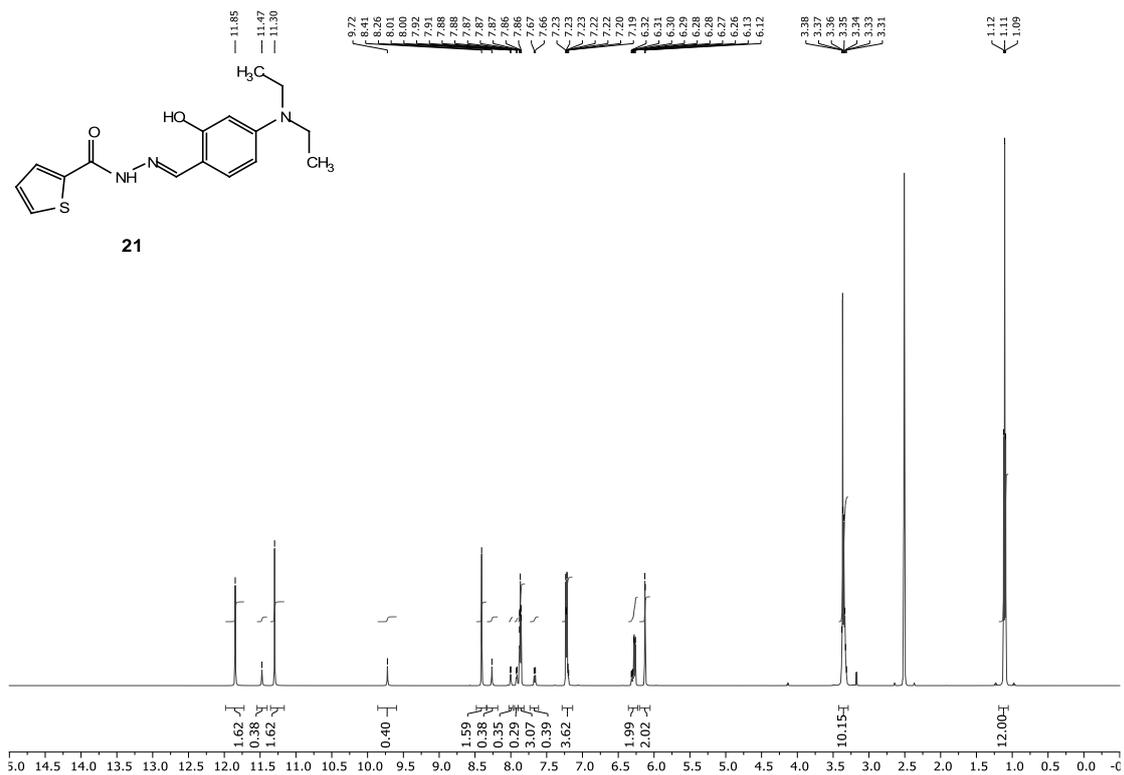
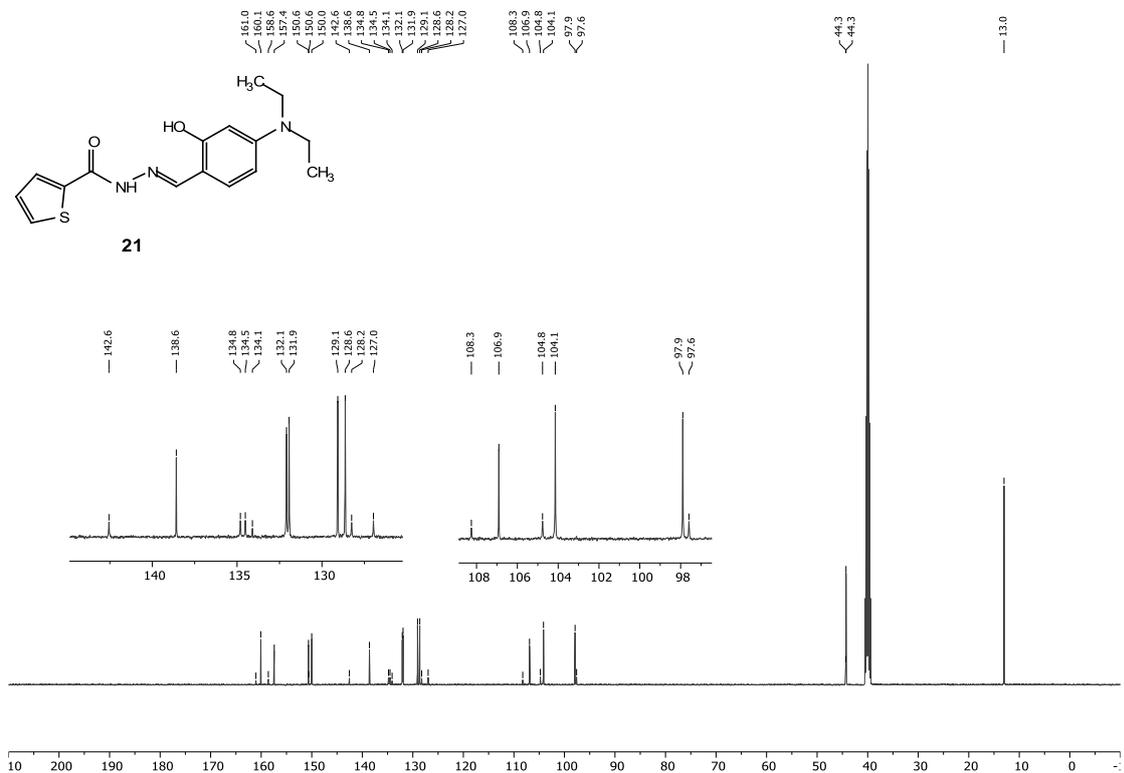
Supplementary Figure 72 <sup>1</sup>H NMR spectrum of **16**.Supplementary Figure 73 <sup>13</sup>C NMR spectrum of **16**.

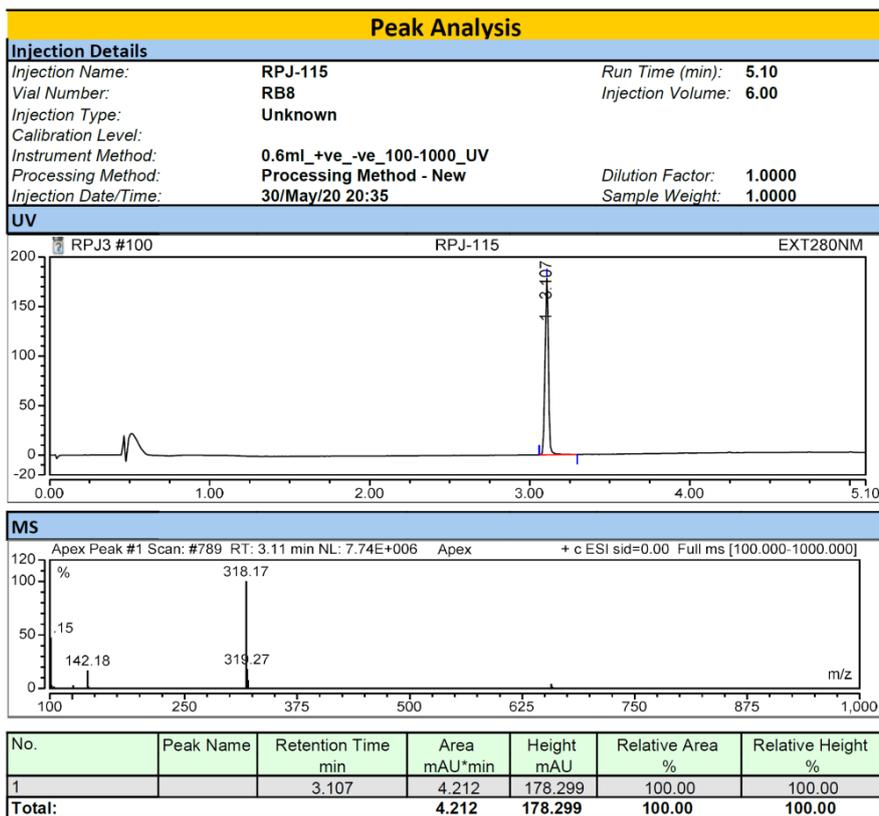


Supplementary Figure 74 HPLC purity analysis of 16.

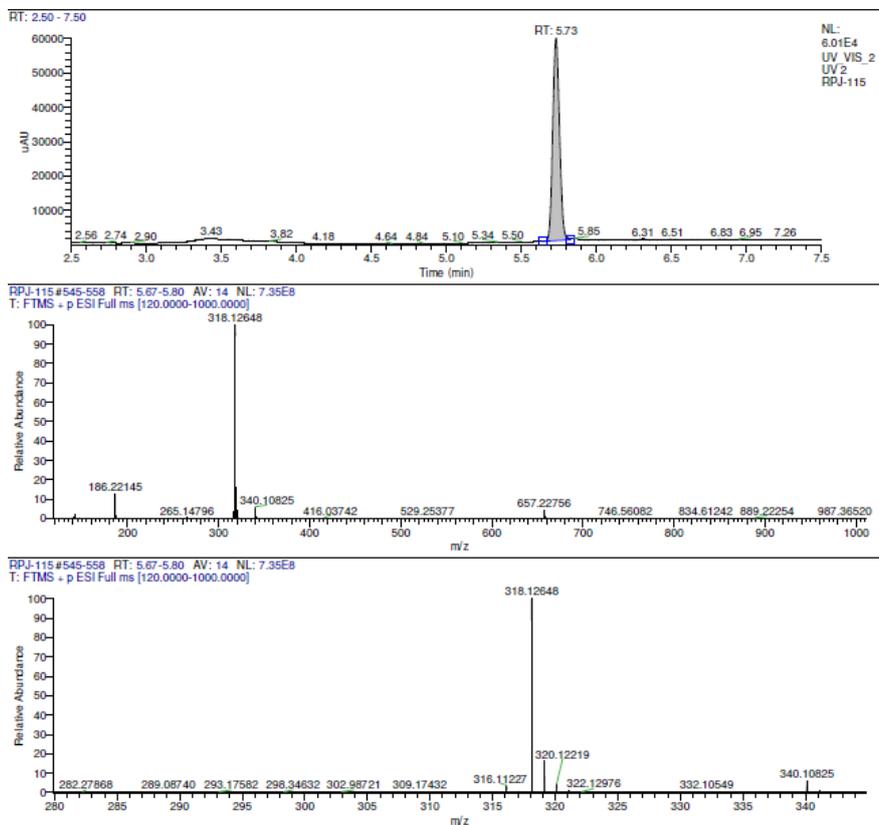


Supplementary Figure 75 HRMS of 16.

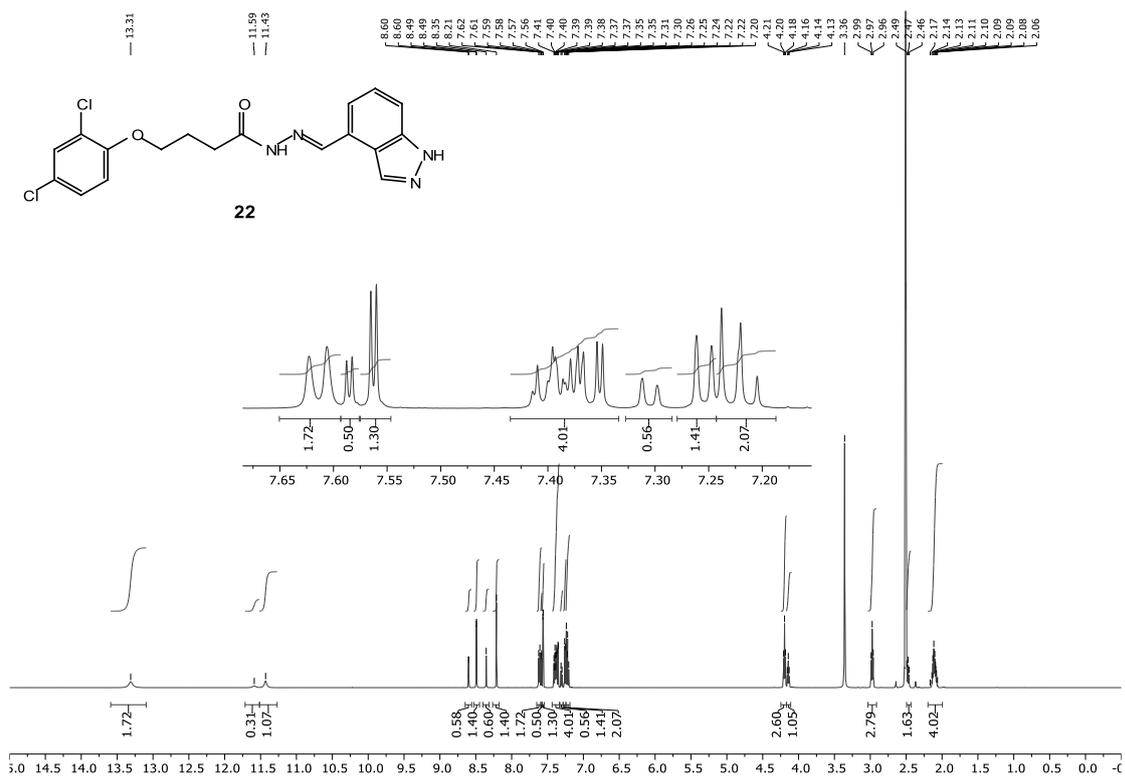
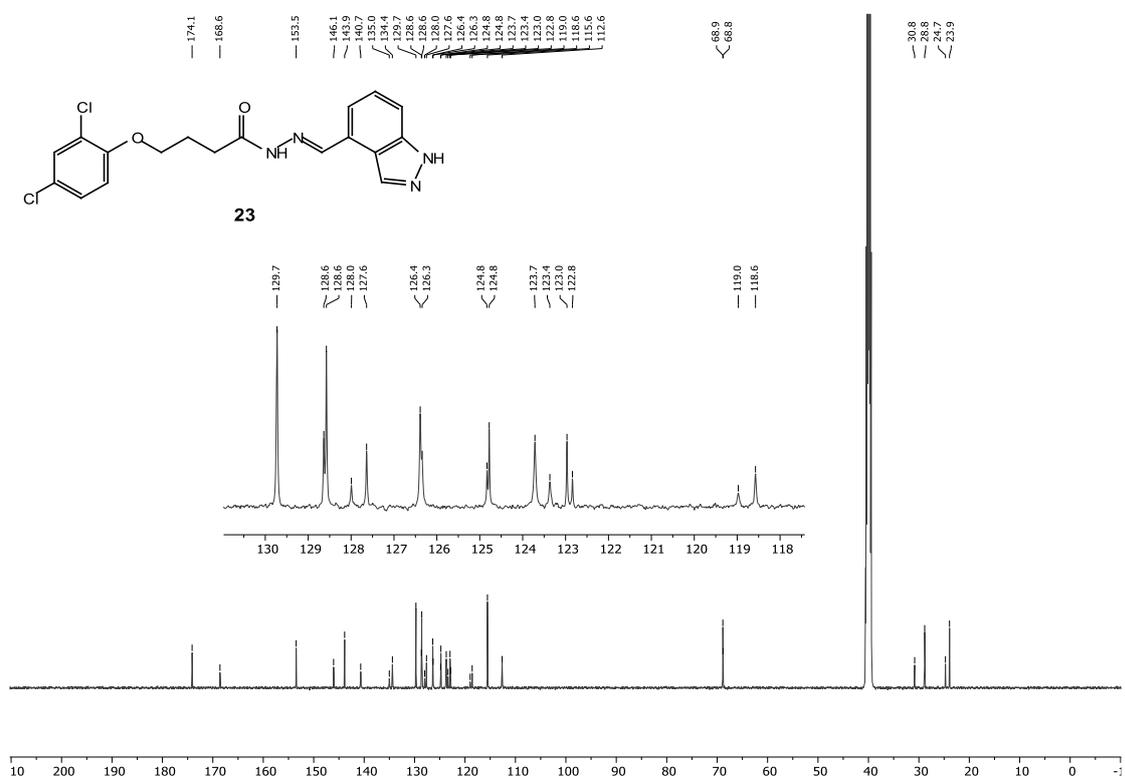
Supplementary Figure 76  $^1\text{H}$  NMR spectrum of **21**.Supplementary Figure 77  $^{13}\text{C}$  NMR spectrum of **21**.

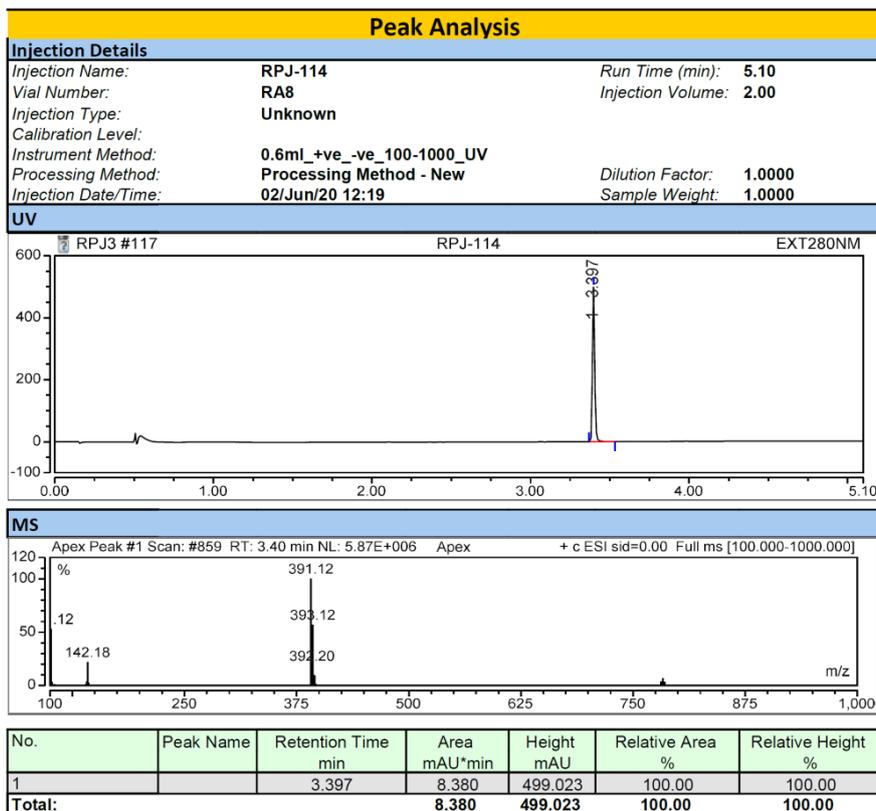


Supplementary Figure 78 HPLC purity analysis of 21.

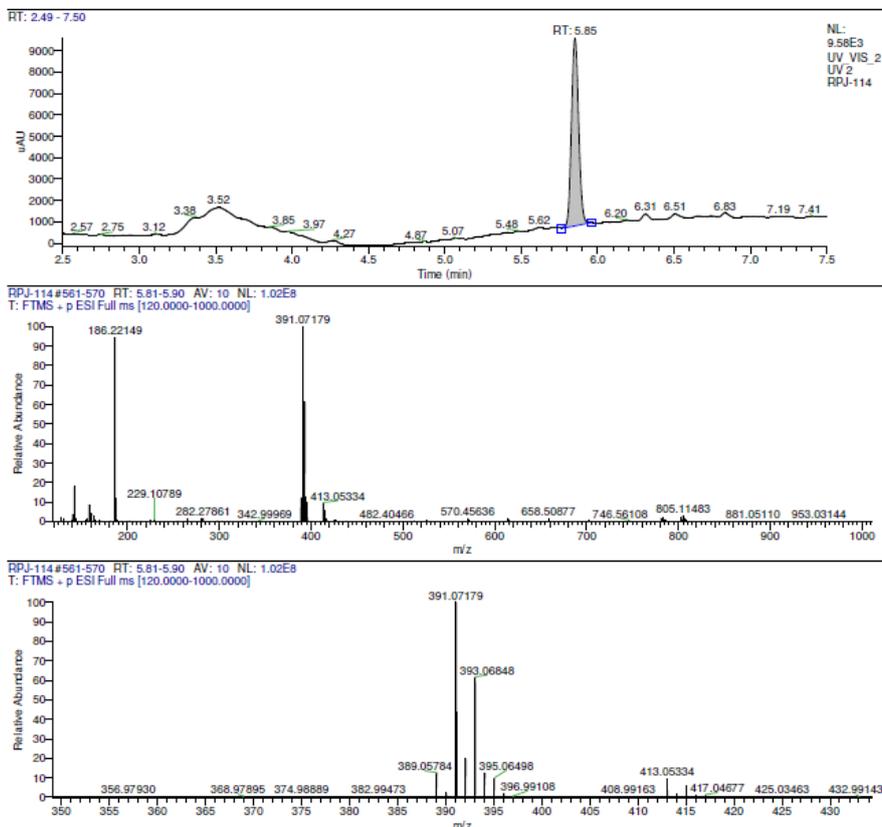


Supplementary Figure 79 HRMS of 21.

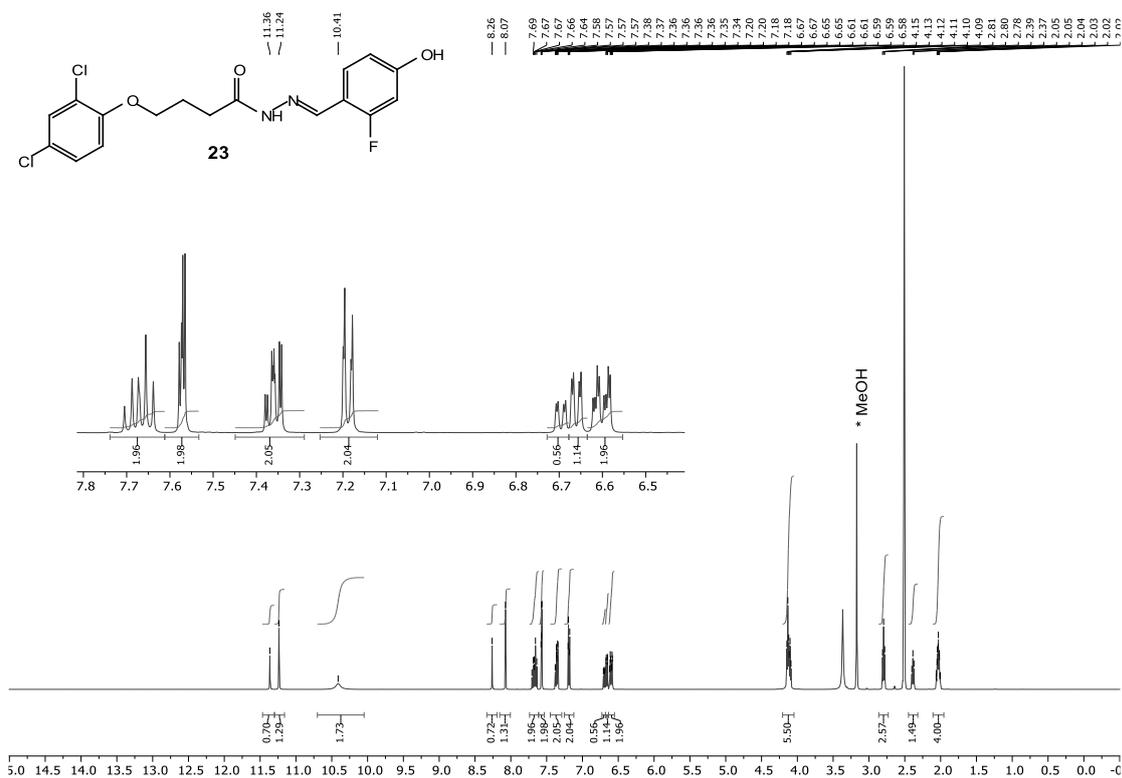
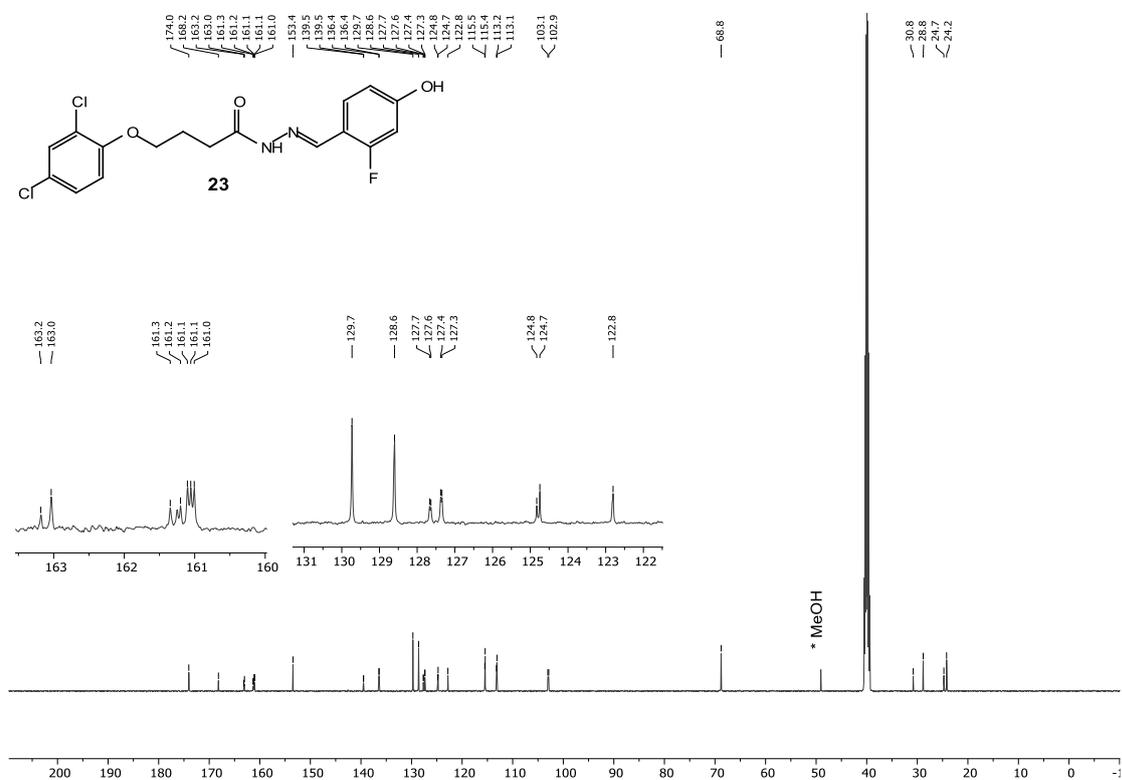
Supplementary Figure 80  $^1\text{H}$  NMR spectrum of **22**.Supplementary Figure 81  $^{13}\text{C}$  NMR spectrum of **22**.

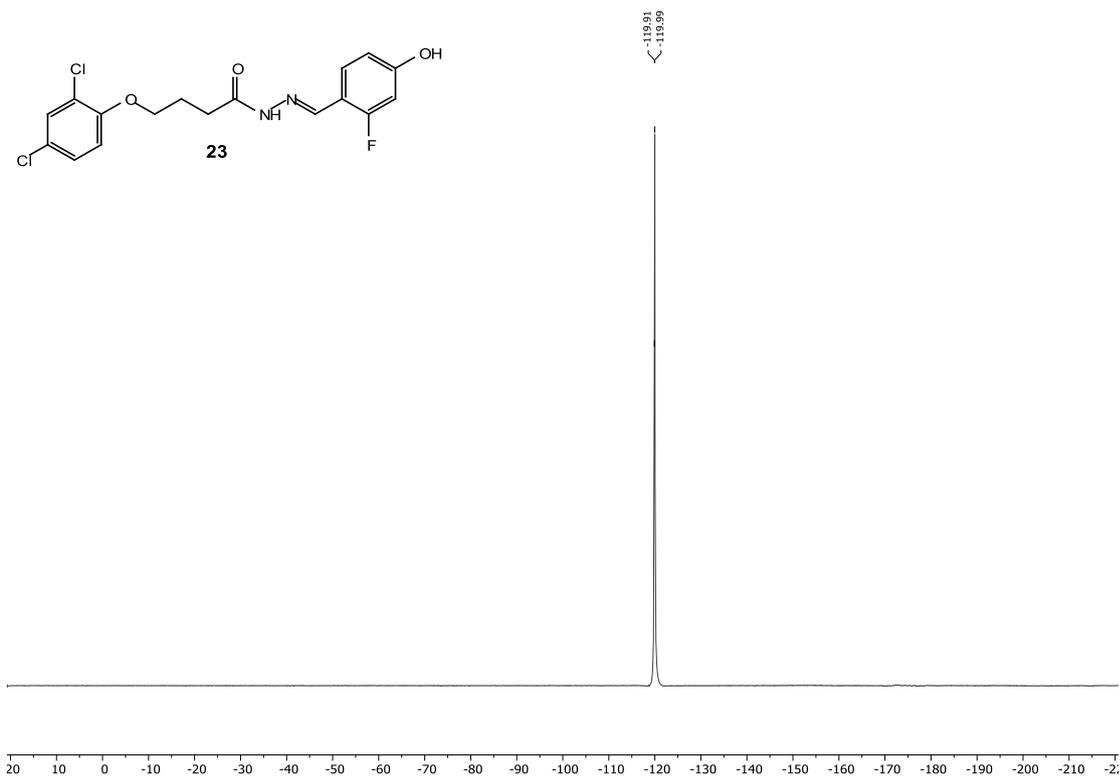


Supplementary Figure 82 HPLC purity analysis of 22.

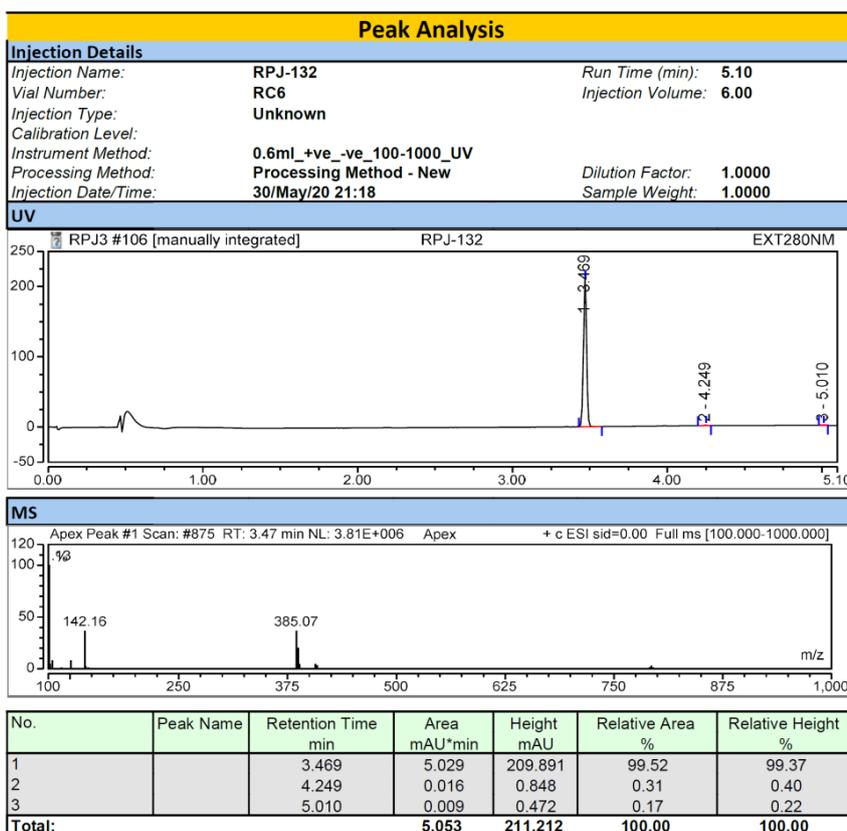


Supplementary Figure 83 HRMS of 22.

Supplementary Figure 84  $^1\text{H}$  NMR spectrum of **23**.Supplementary Figure 85  $^{13}\text{C}$  NMR spectrum of **23**.

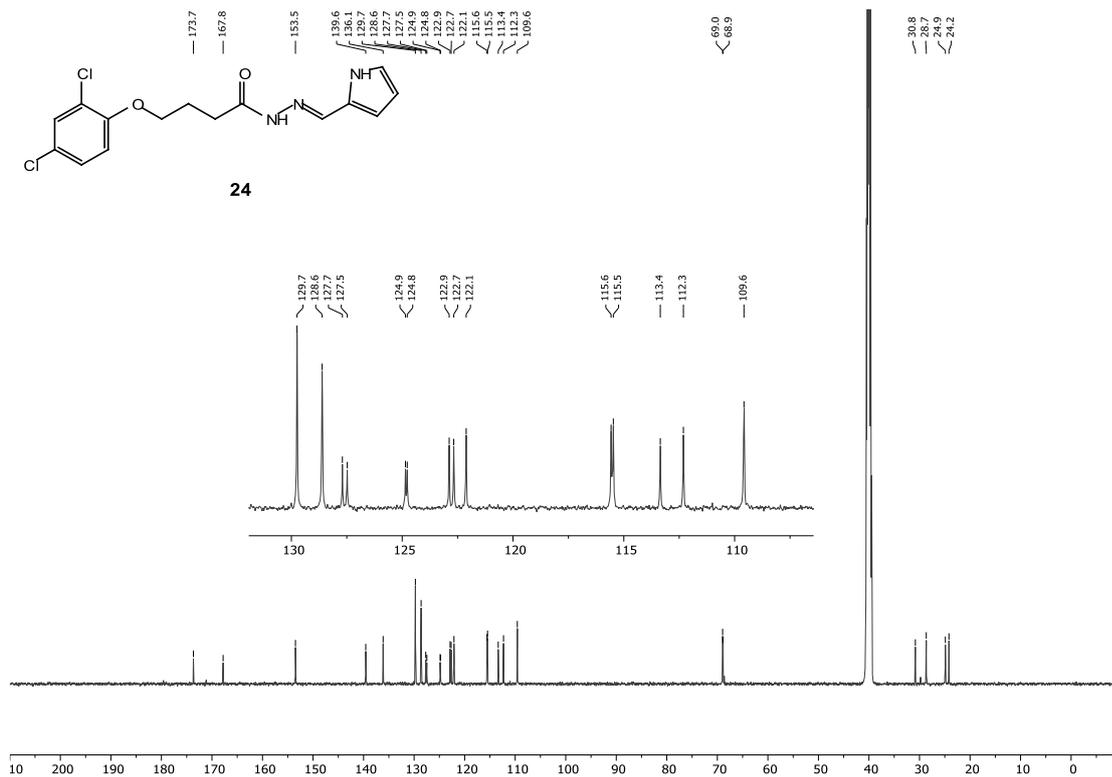


Supplementary Figure 86 <sup>19</sup>F NMR spectrum of **23**.

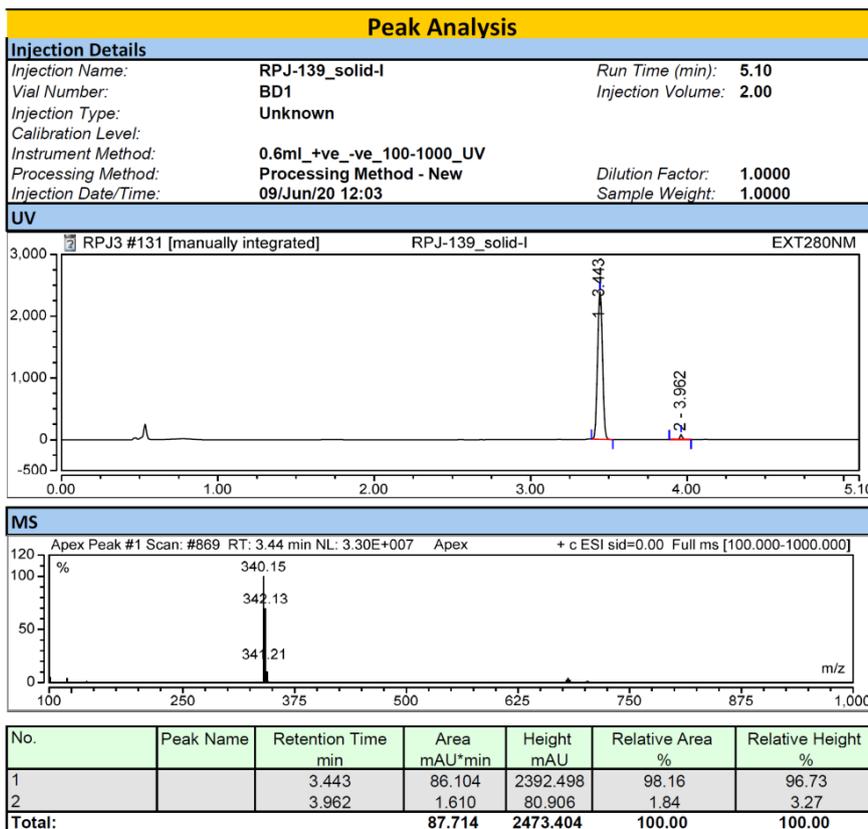


Supplementary Figure 87 HPLC purity analysis of **23**.

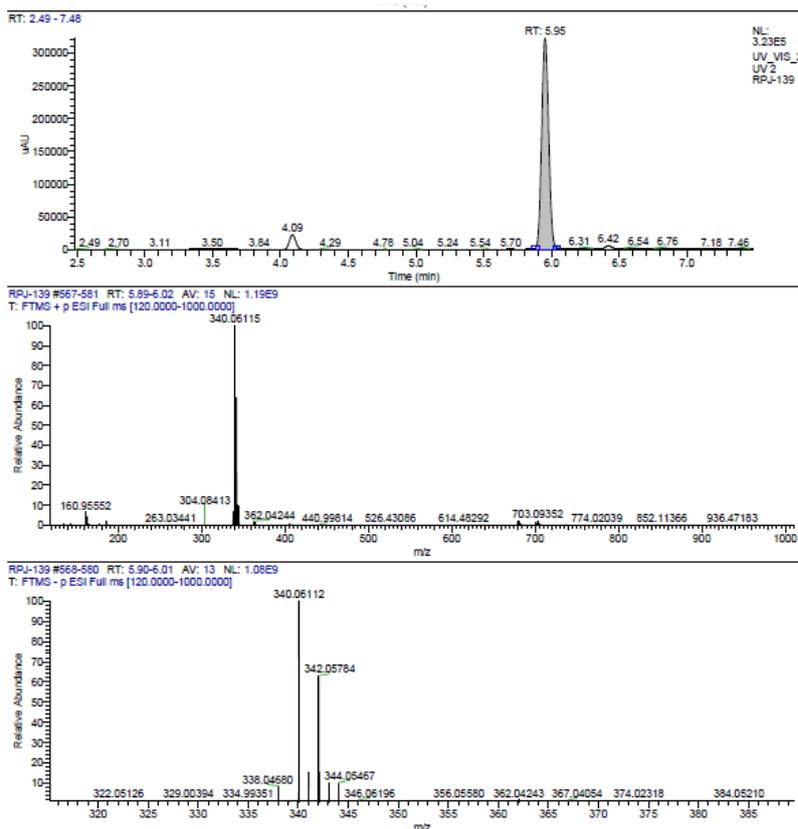




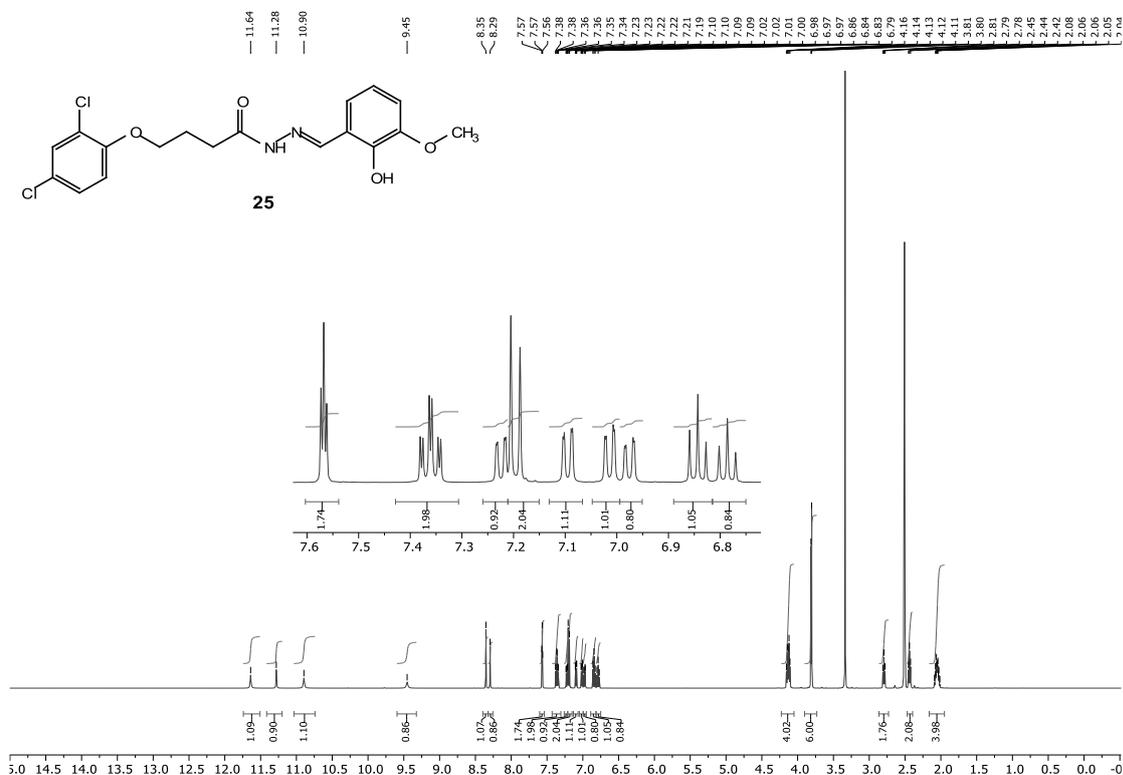
Supplementary Figure 90 <sup>13</sup>C NMR spectrum of 24.

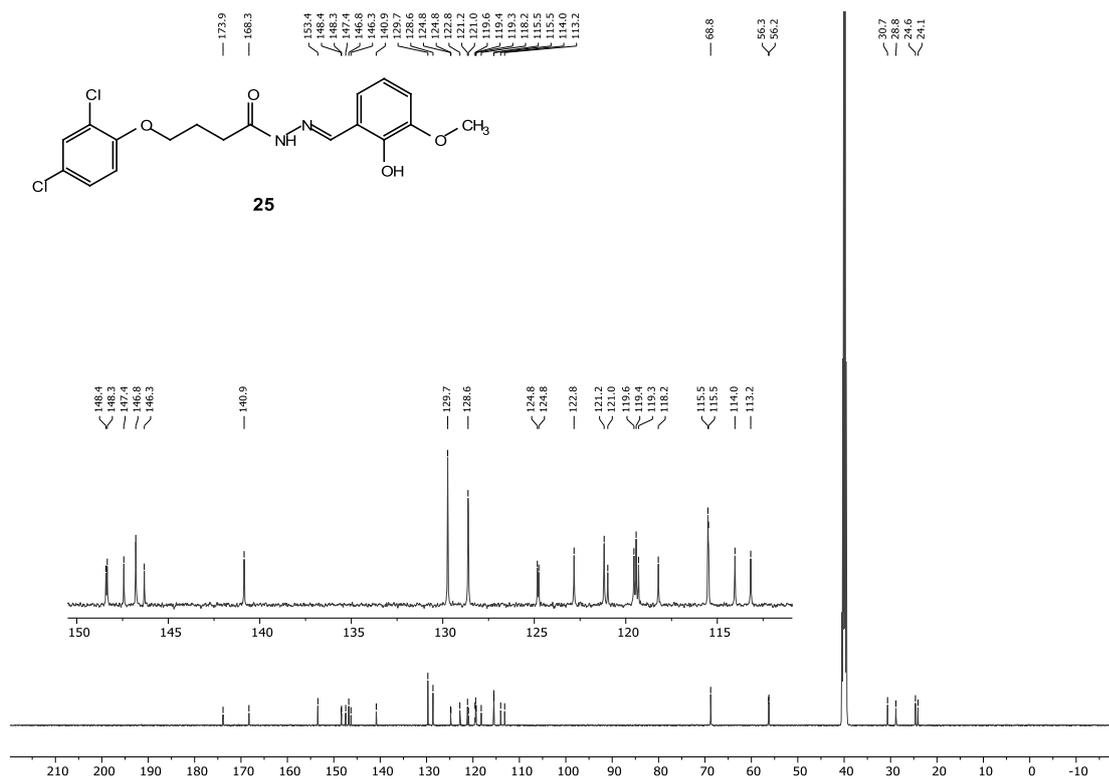


Supplementary Figure 91 HPLC purity analysis of 24.

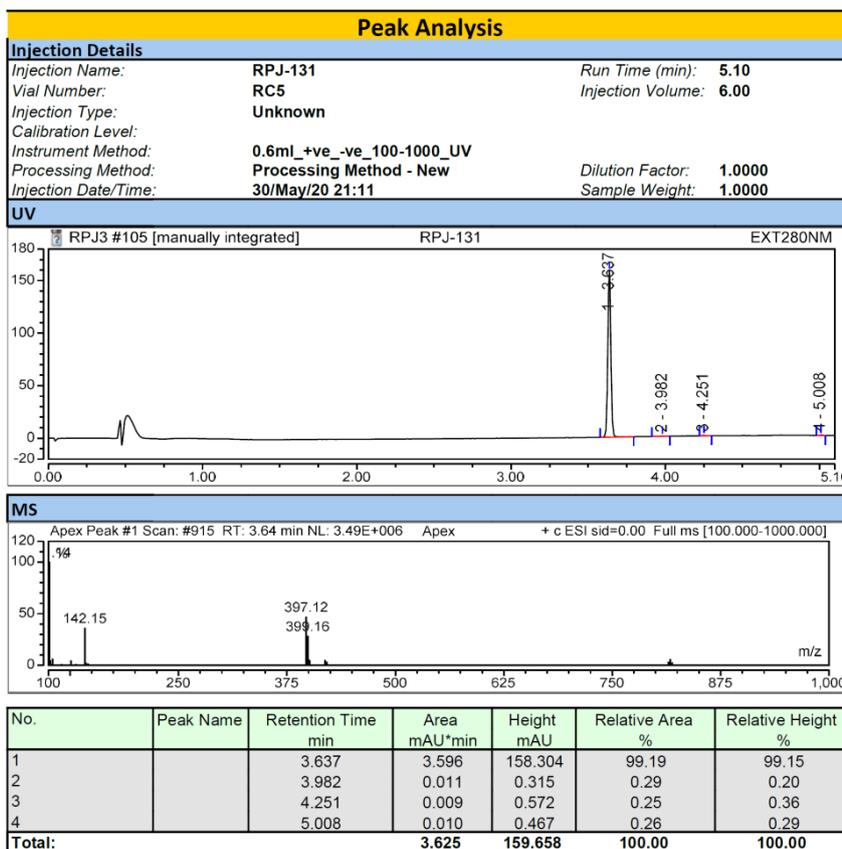


Supplementary Figure 92 HRMS of 24.

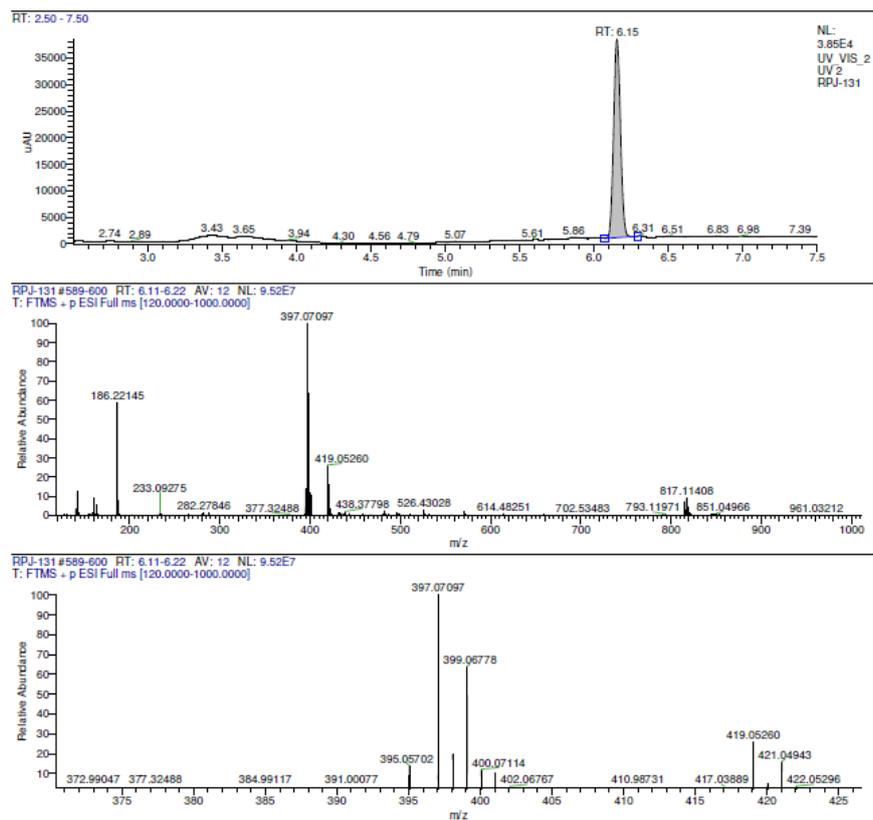
Supplementary Figure 93 <sup>1</sup>H NMR spectrum of 25.



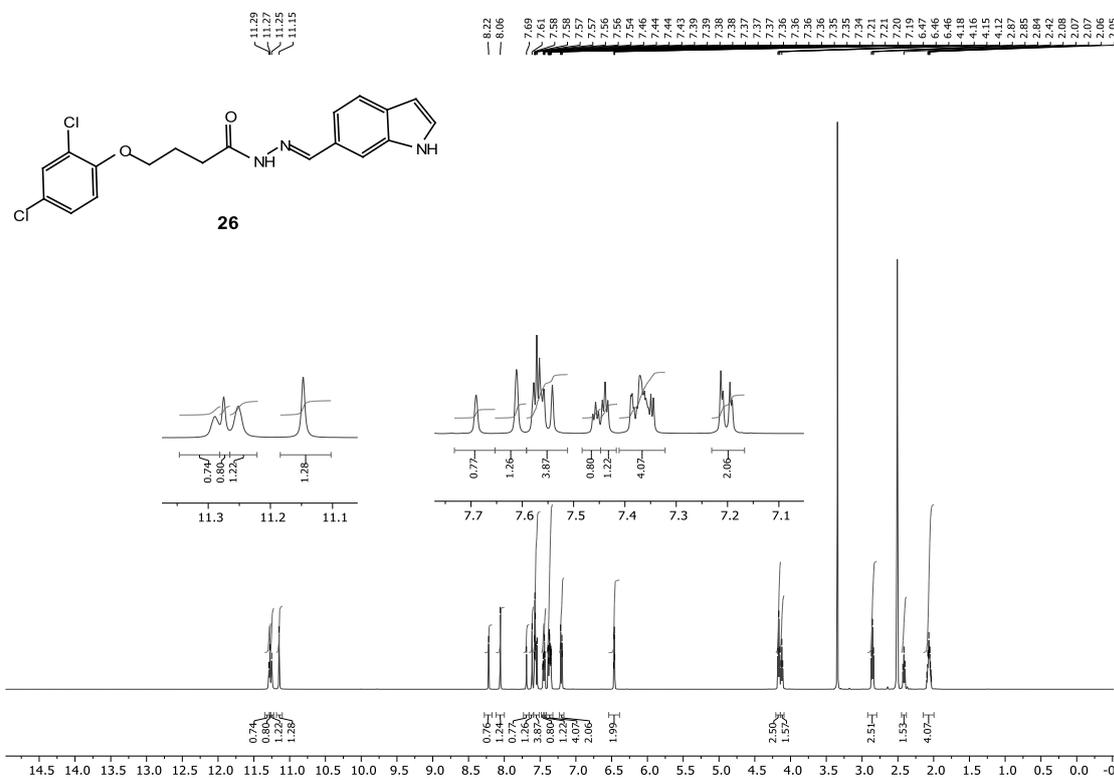
Supplementary Figure 94 <sup>13</sup>C NMR spectrum of 25.

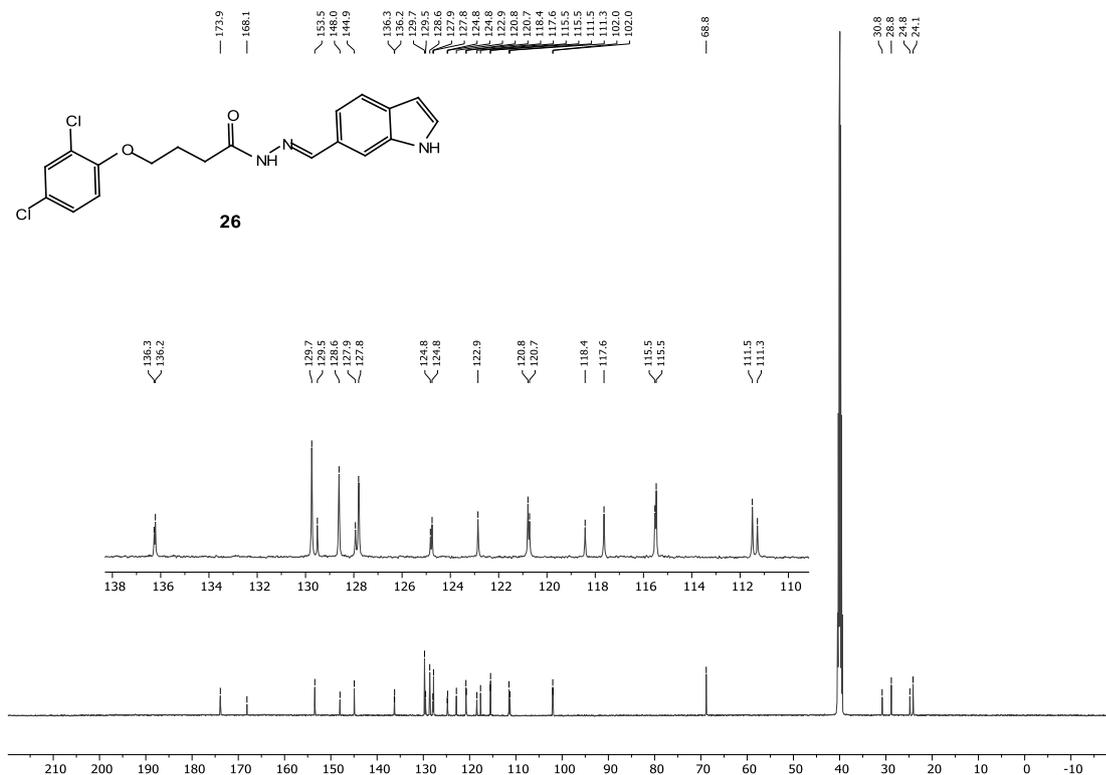
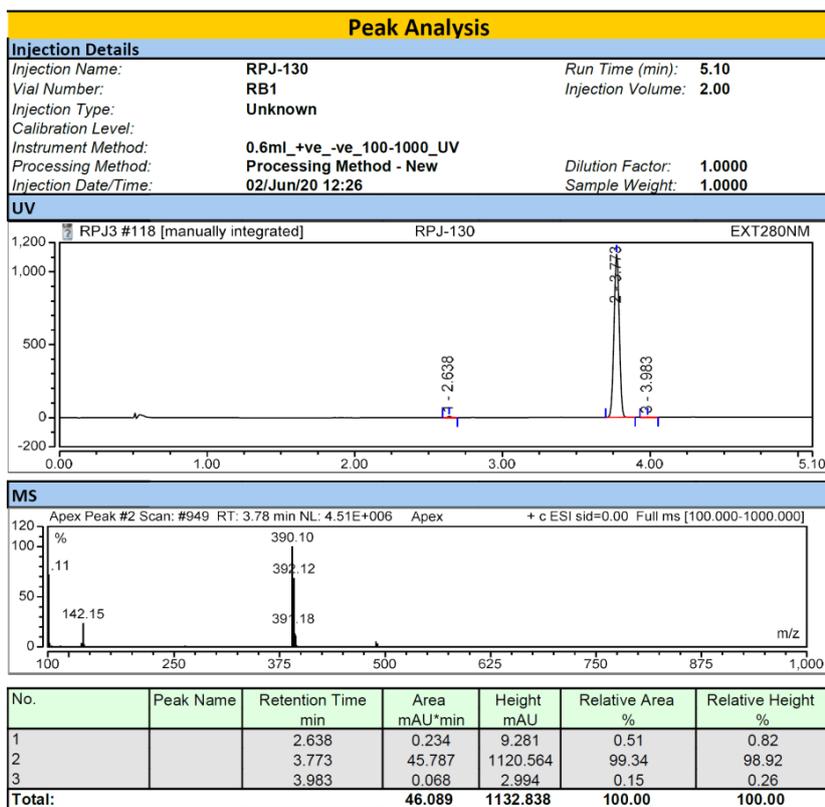


Supplementary Figure 95 HPLC purity analysis of 25.

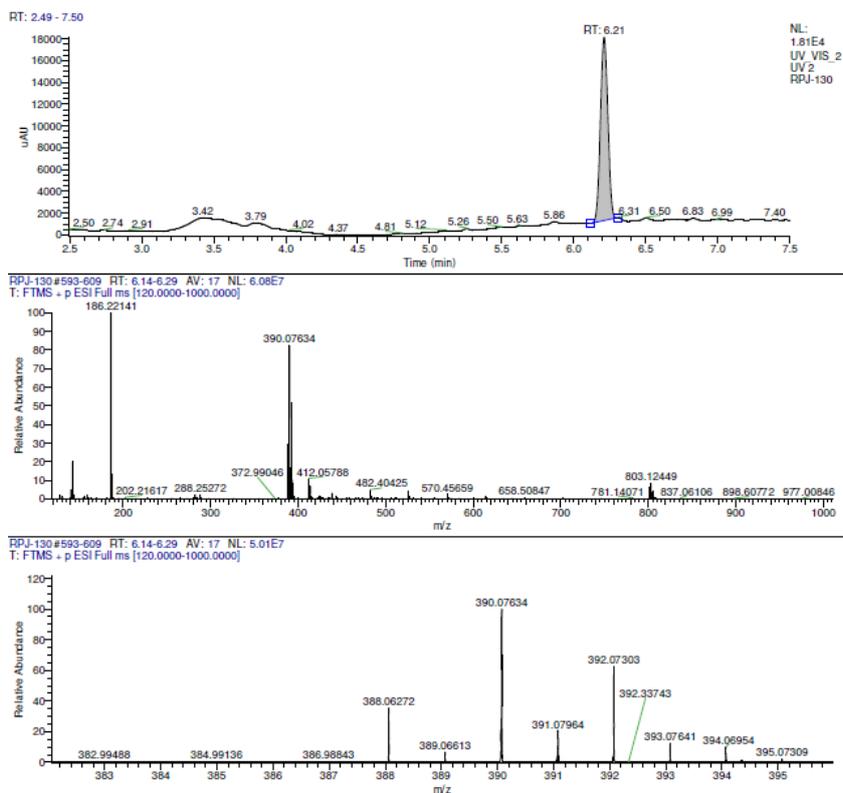


Supplementary Figure 96 HRMS of 25.

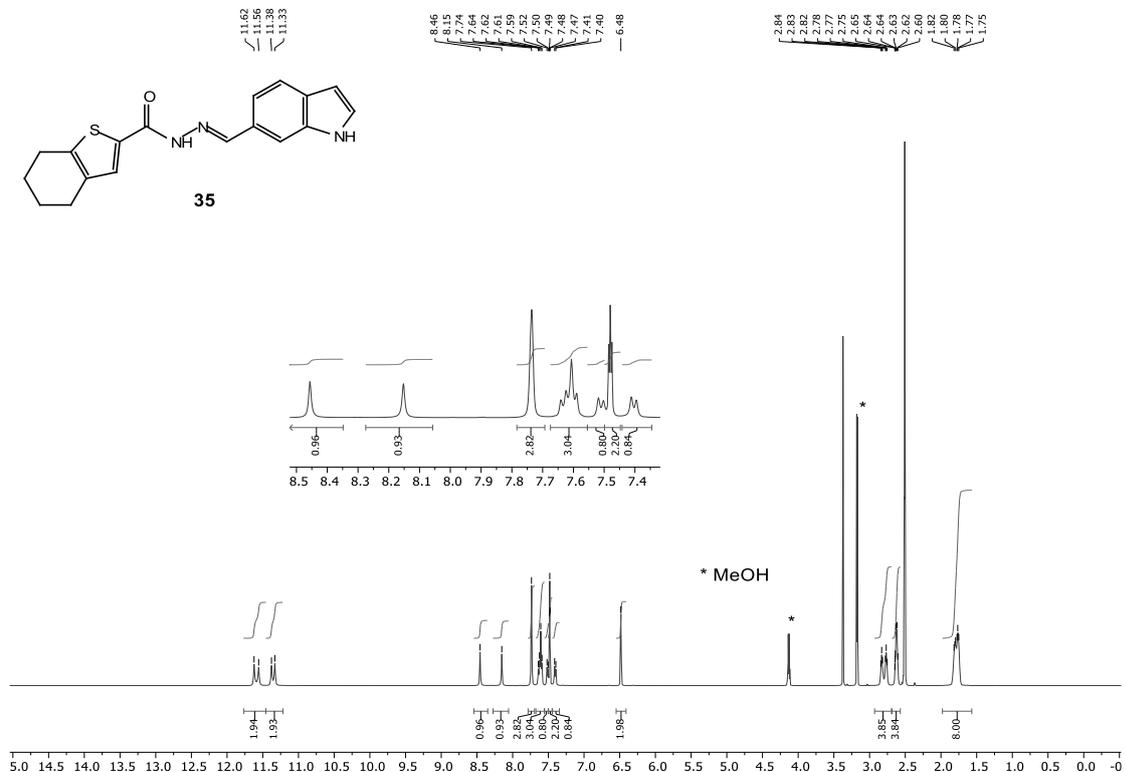


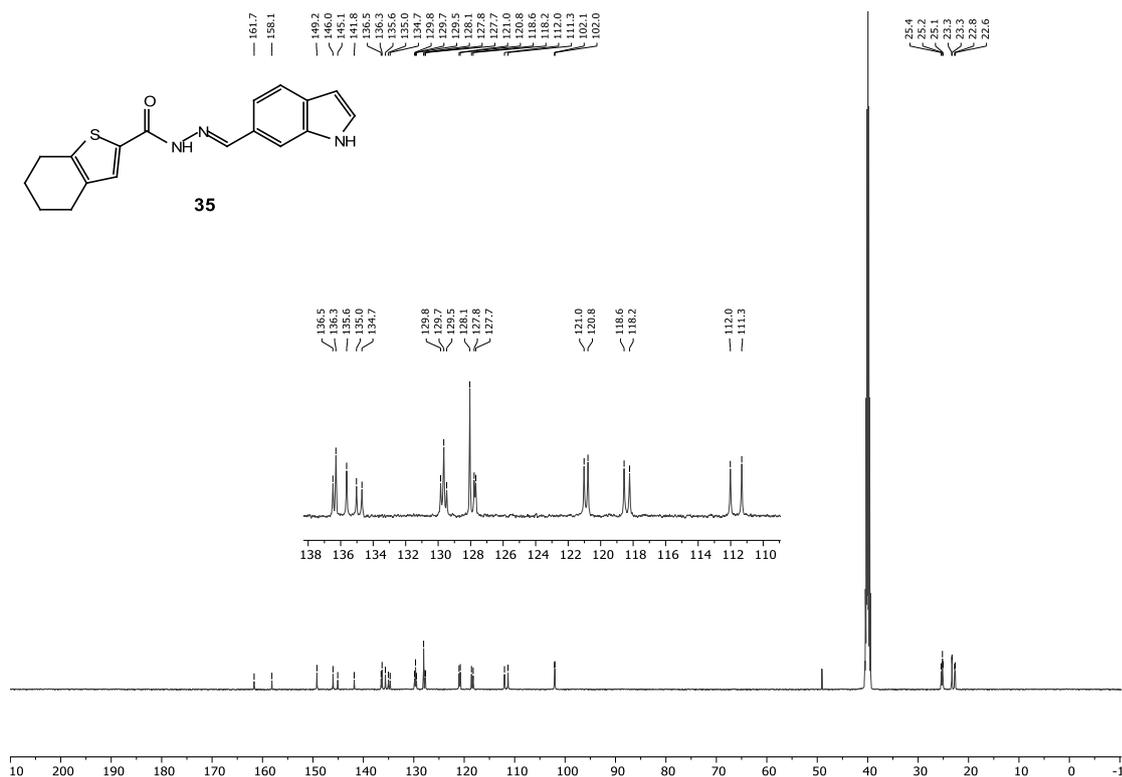
Supplementary Figure 97 <sup>1</sup>H NMR spectrum of 26.Supplementary Figure 98 <sup>13</sup>C NMR spectrum of 26.

Supplementary Figure 99 HPLC purity analysis of 26.

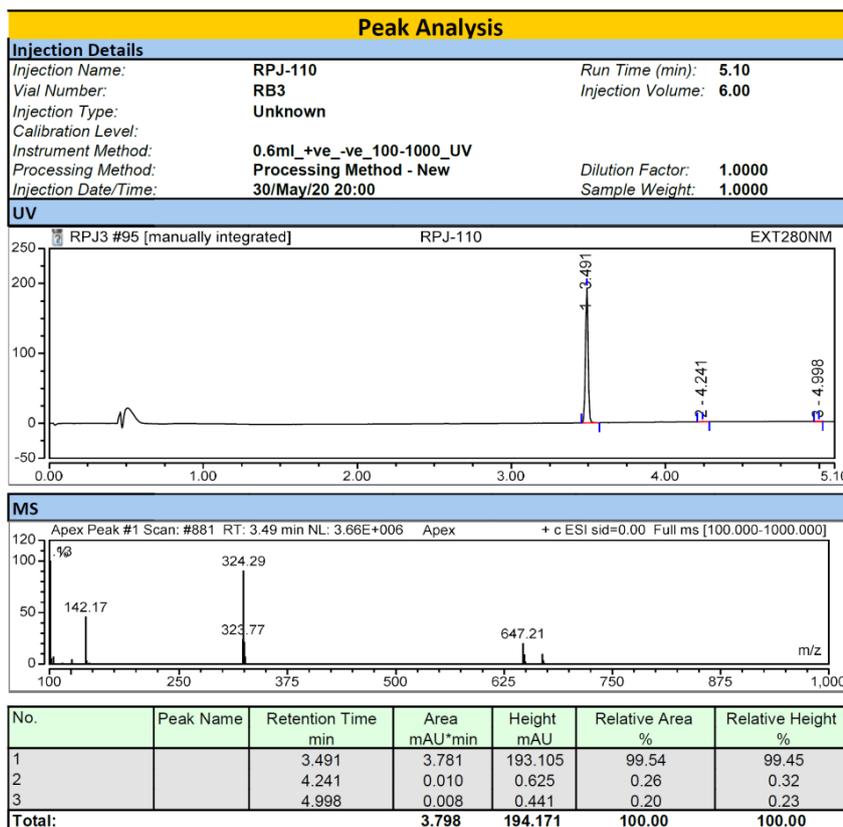


Supplementary Figure 100 HRMS of 26.

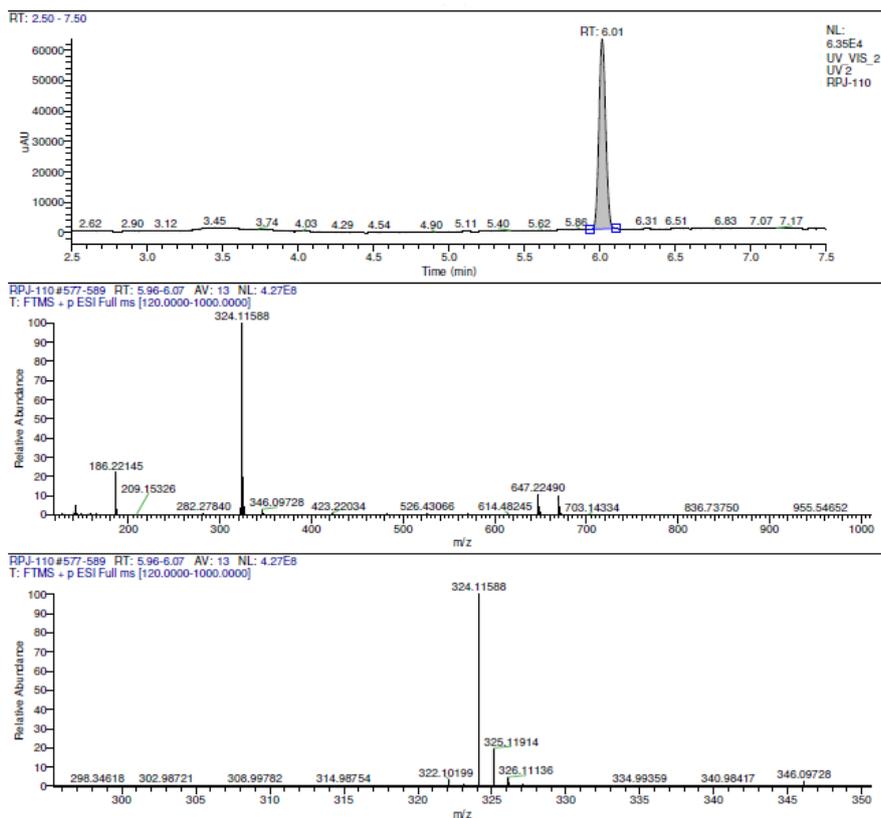
Supplementary Figure 101  $^1\text{H}$  NMR spectrum of 35.



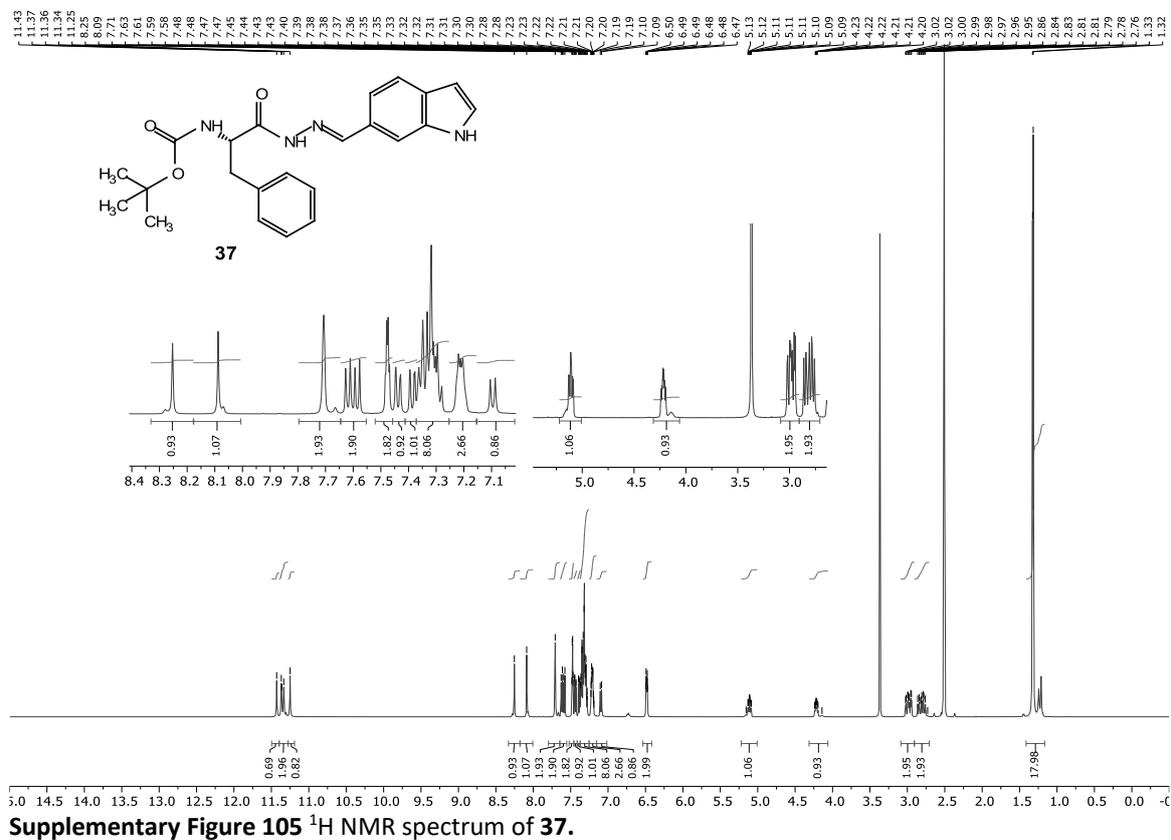
Supplementary Figure 102 <sup>13</sup>C NMR spectrum of 35.

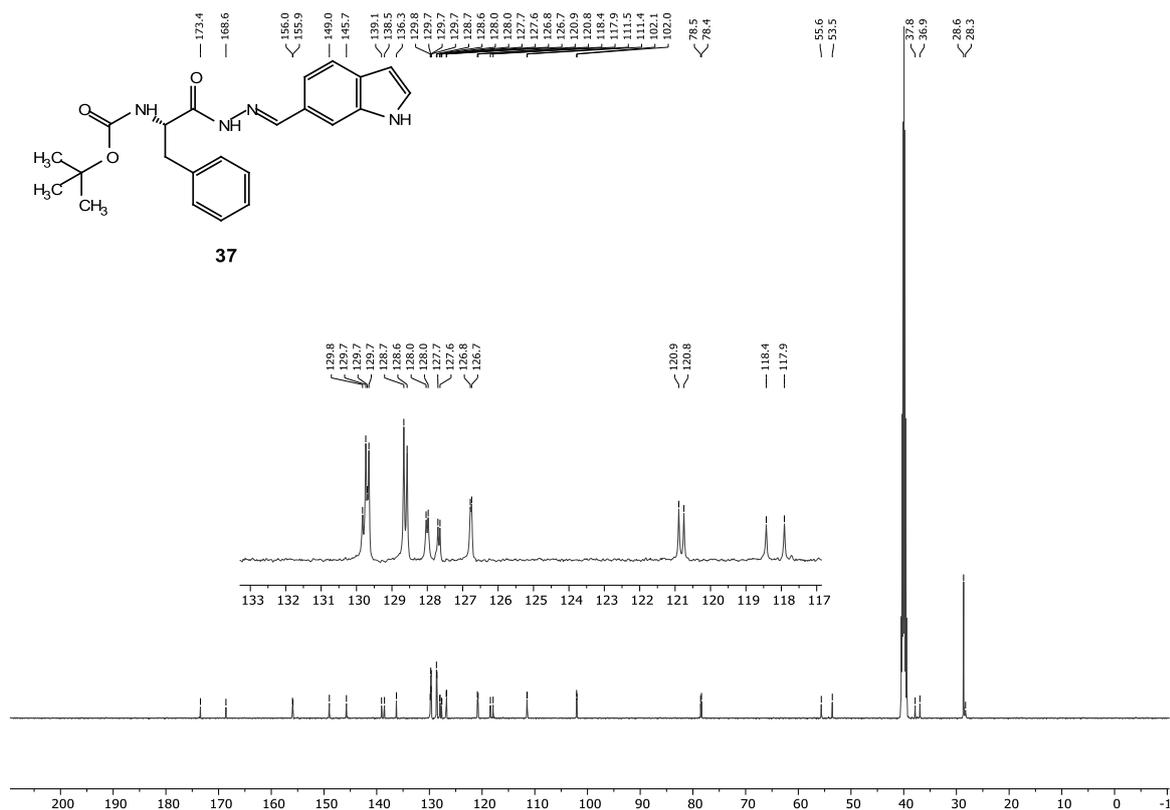
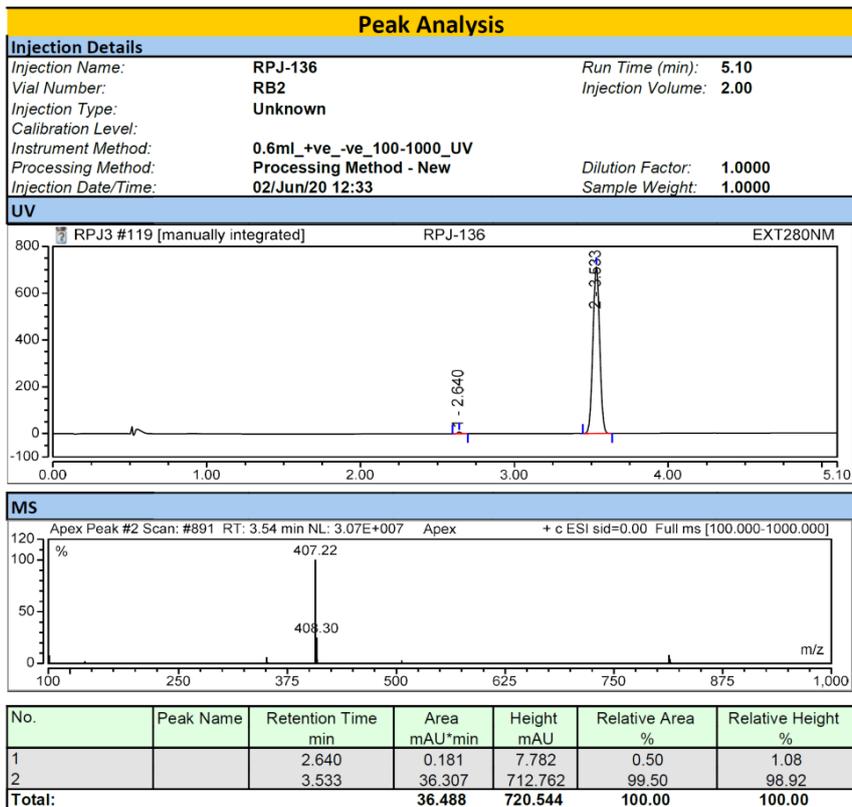


Supplementary Figure 103 HPLC purity analysis of 35.

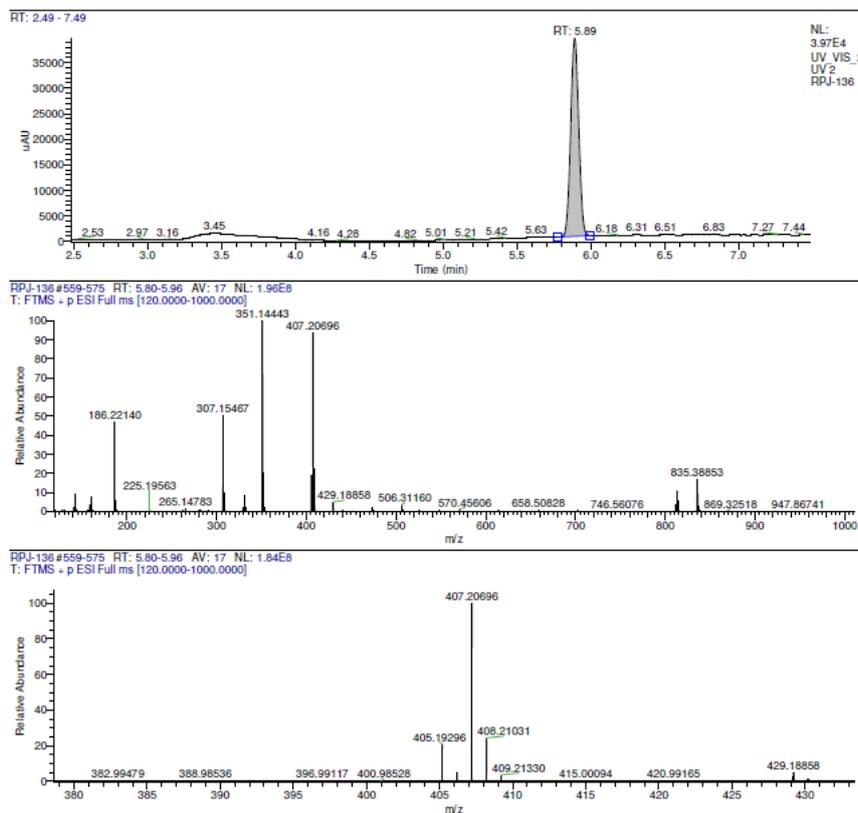


Supplementary Figure 104 HRMS of 35.

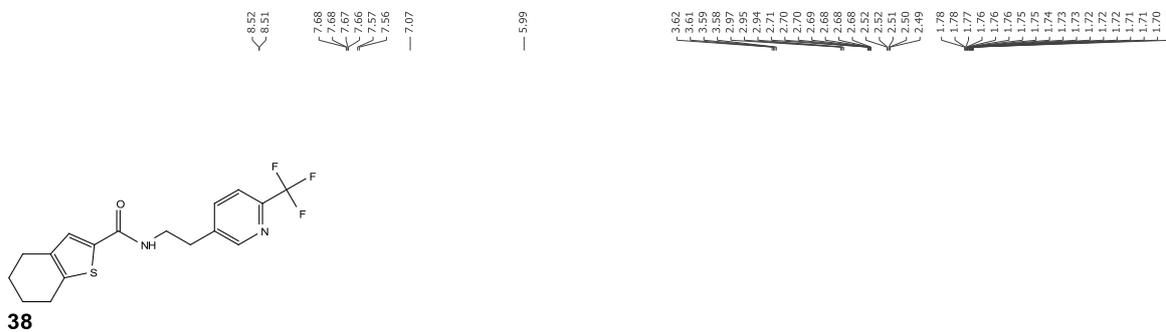


Supplementary Figure 106 <sup>13</sup>C NMR spectrum of 37.

Supplementary Figure 107 HPLC purity analysis of 37.

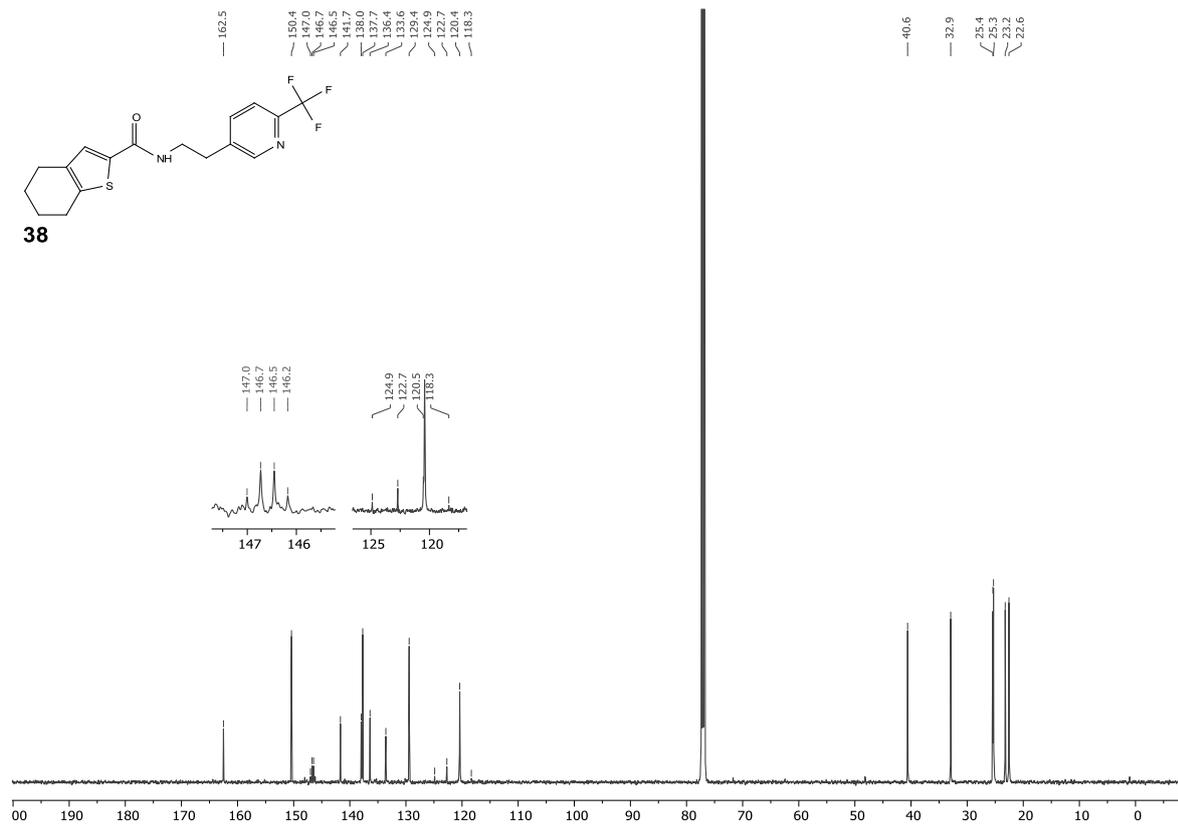
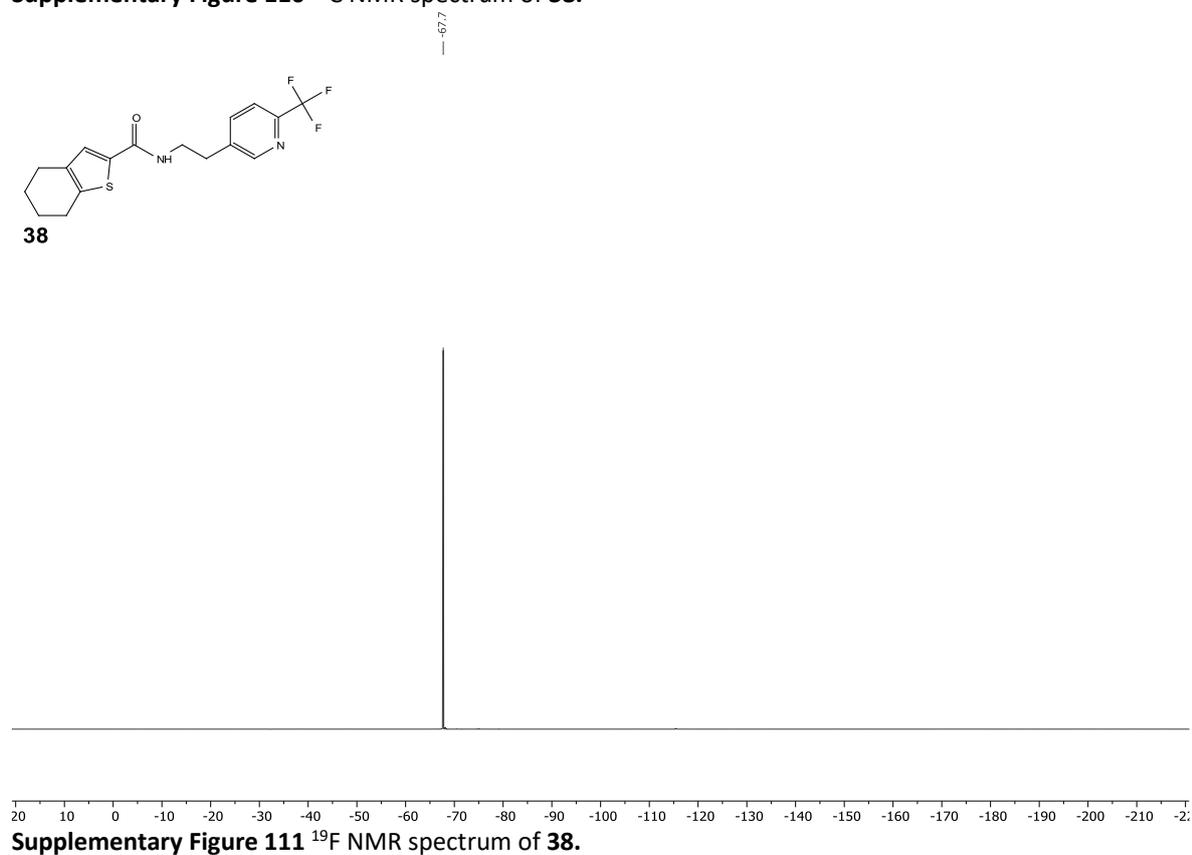


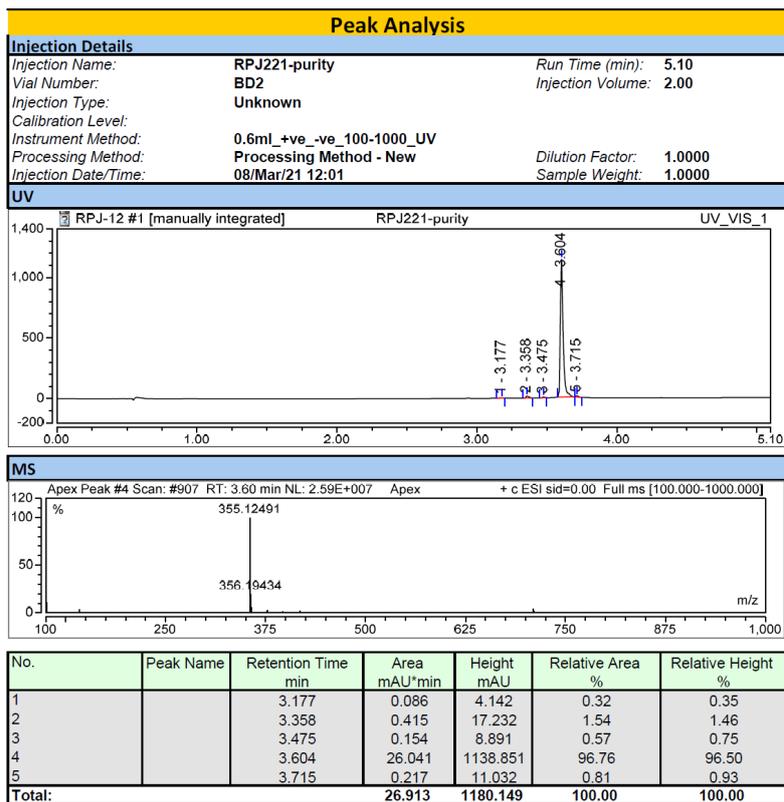
Supplementary Figure 108 HRMS of 37.



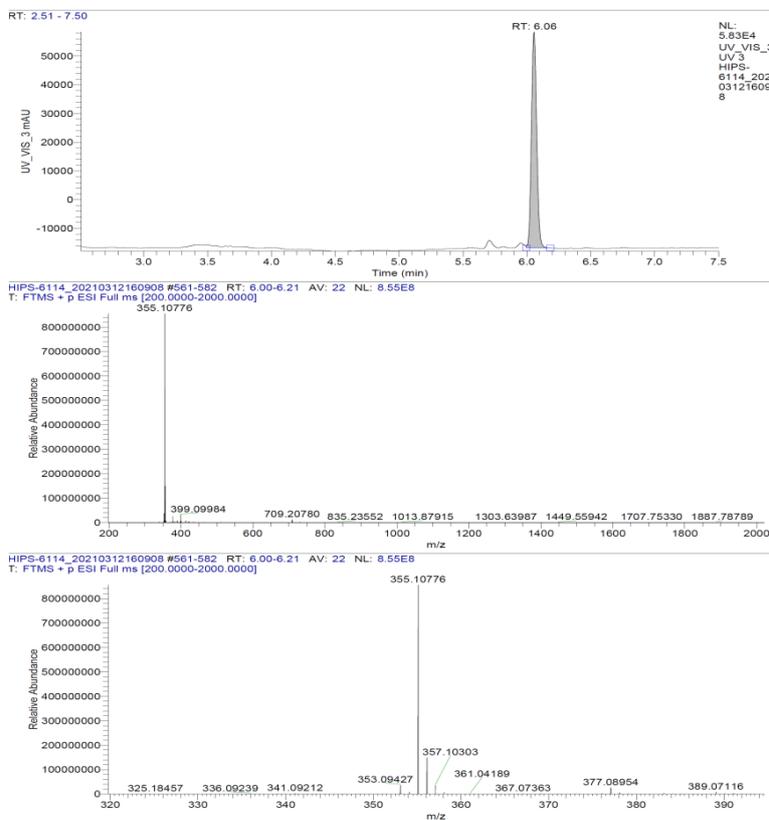
38

Supplementary Figure 109 <sup>1</sup>H NMR spectrum of 38.

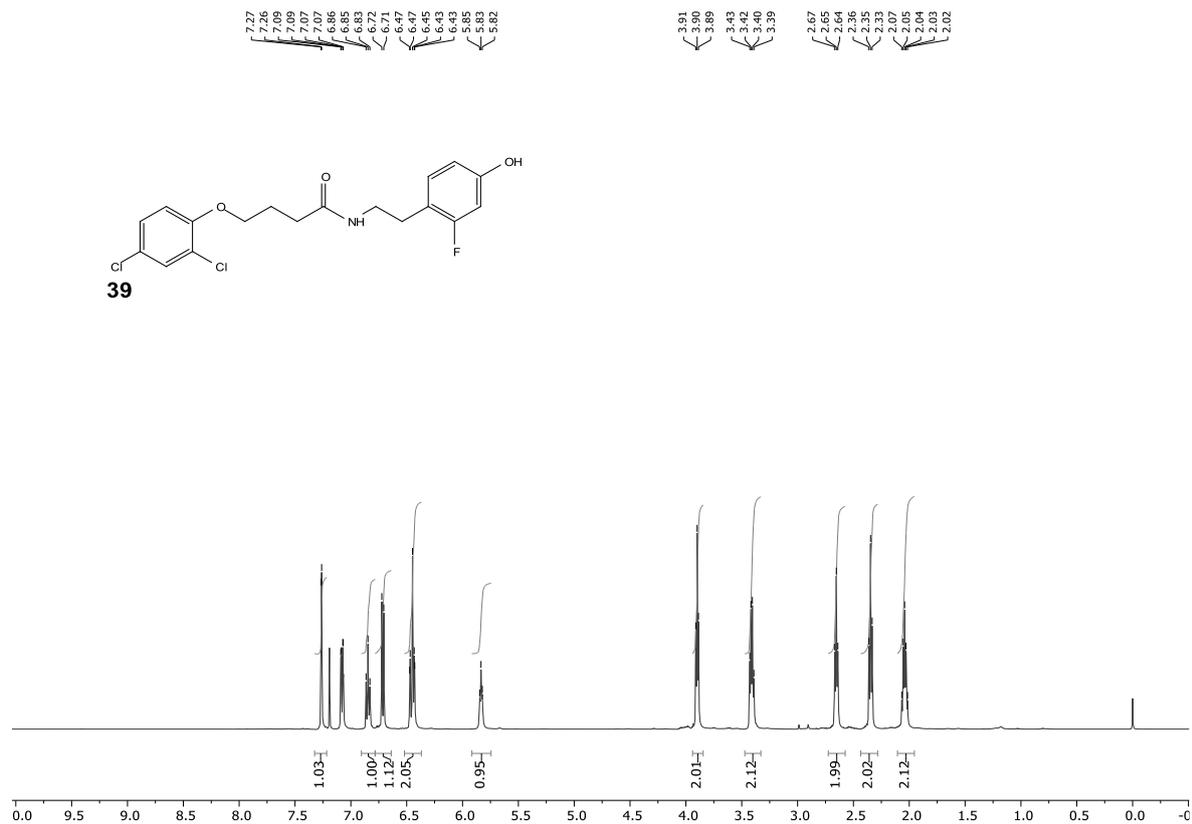
Supplementary Figure 110  $^{13}\text{C}$  NMR spectrum of **38**.Supplementary Figure 111  $^{19}\text{F}$  NMR spectrum of **38**.



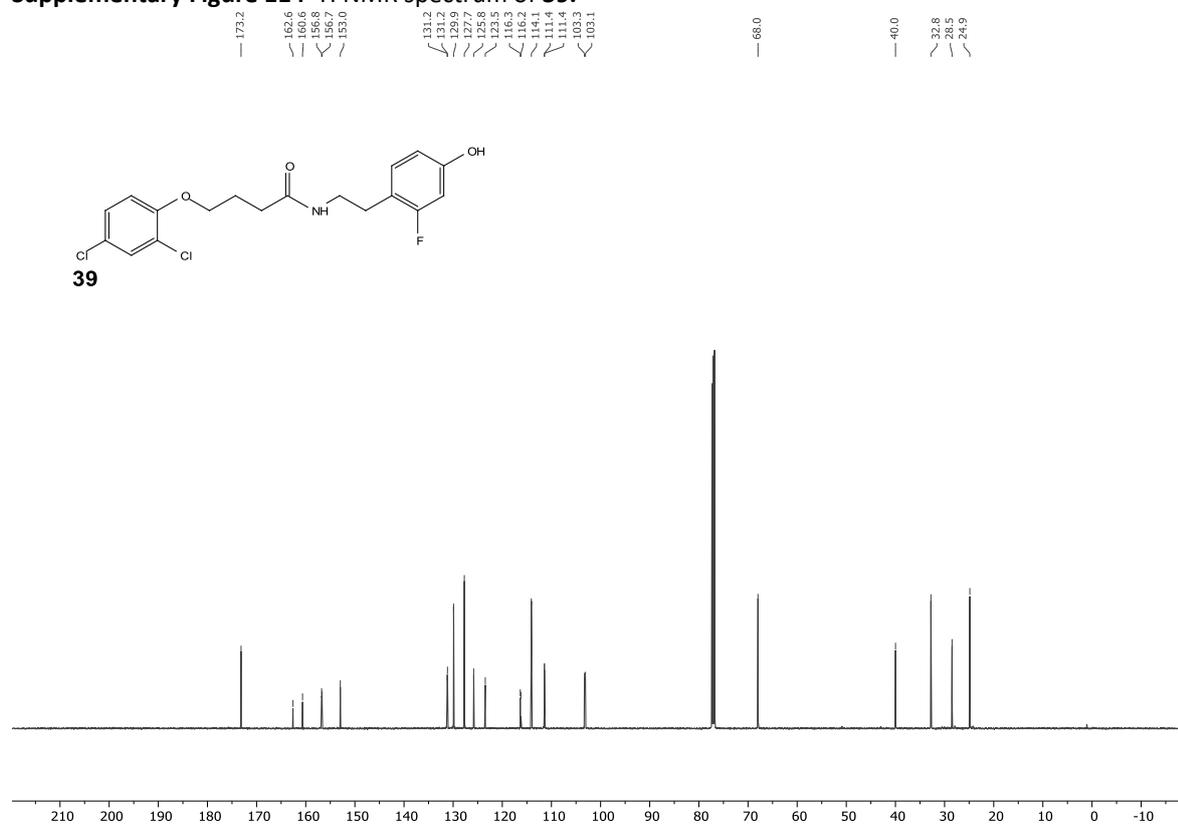
Supplementary Figure 112 HPLC purity analysis of 38.



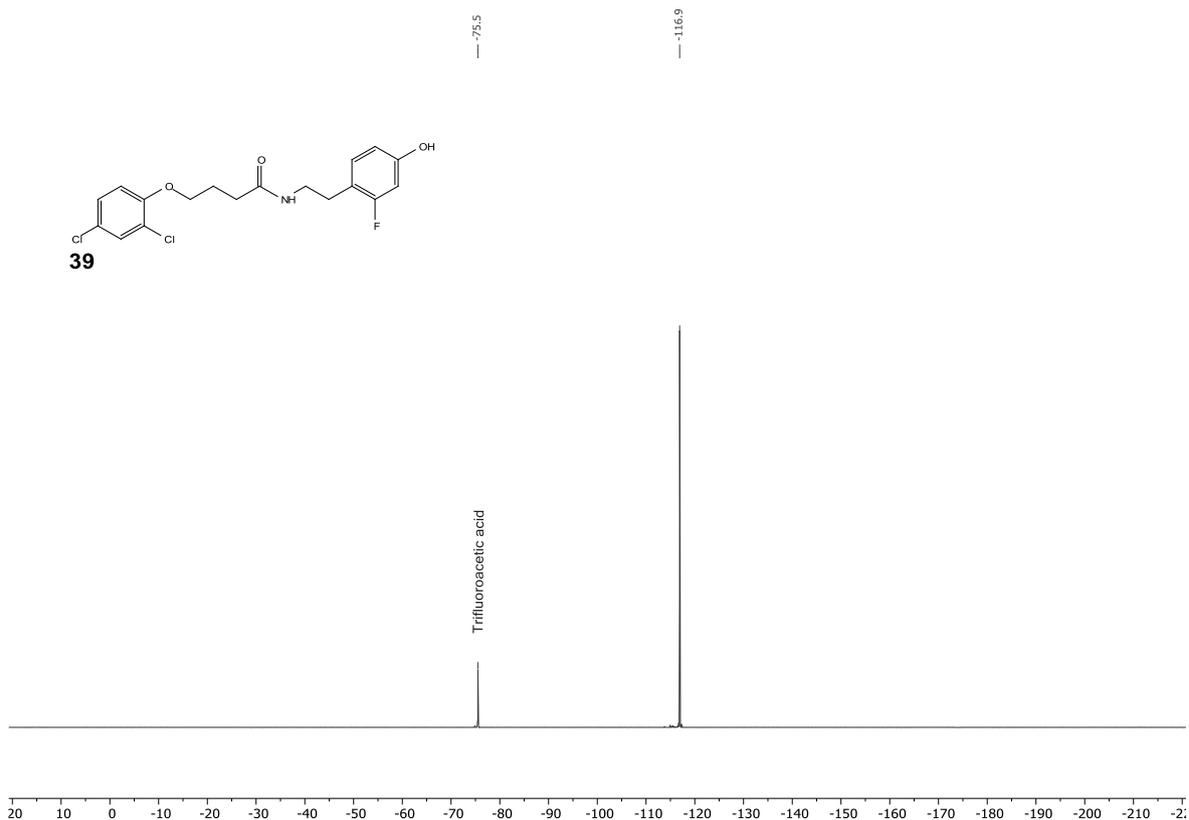
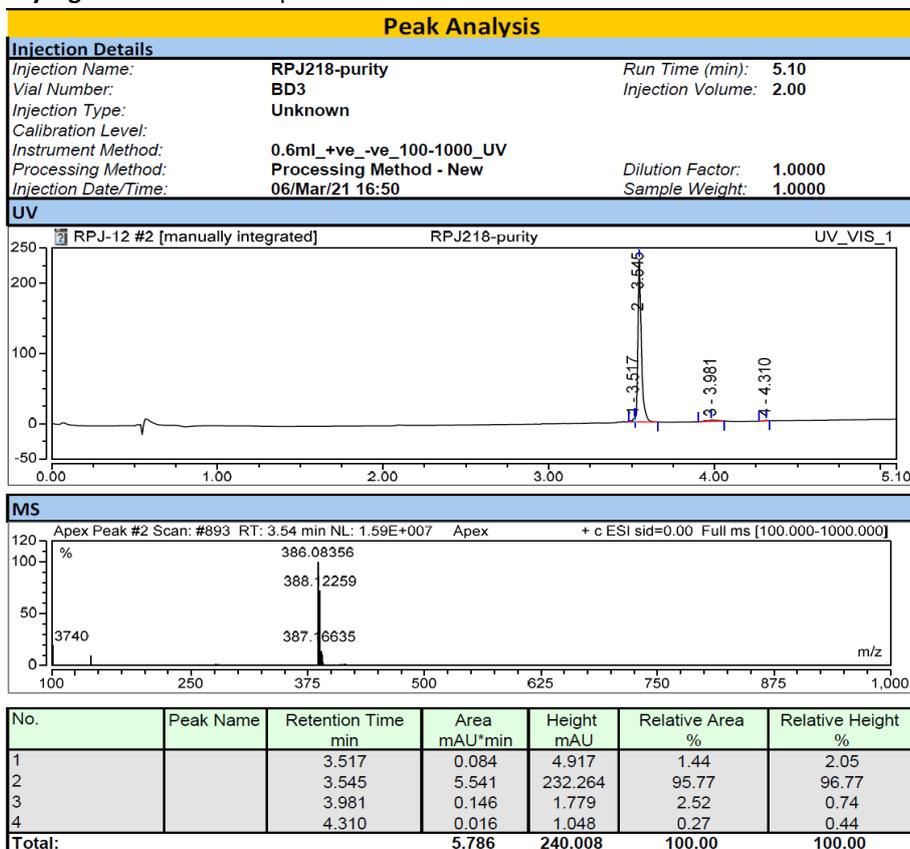
Supplementary Figure 113 HRMS of 38.



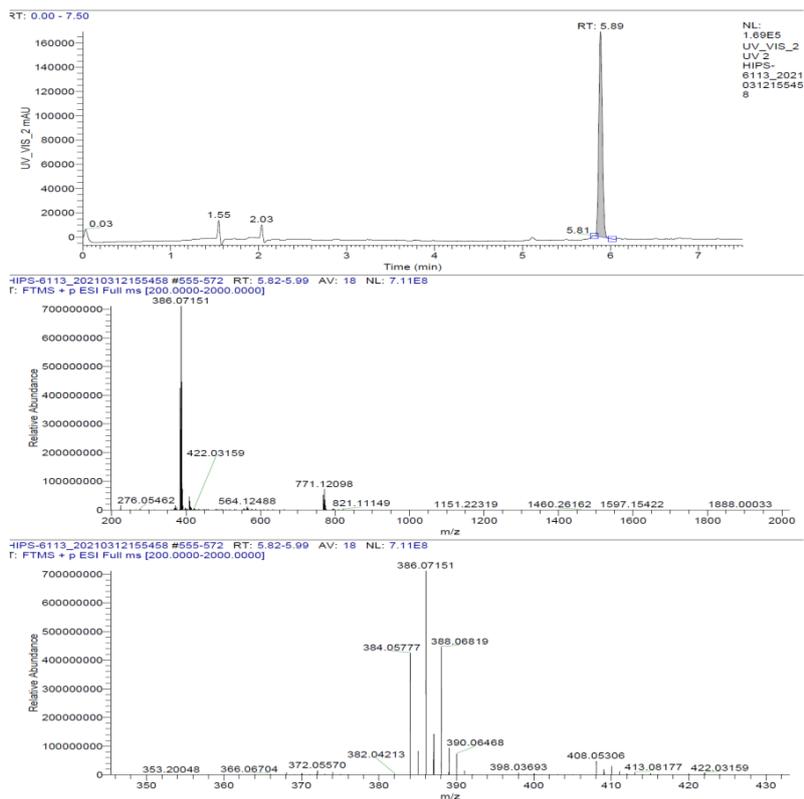
**Supplementary Figure 114**  $^1\text{H}$  NMR spectrum of **39**.



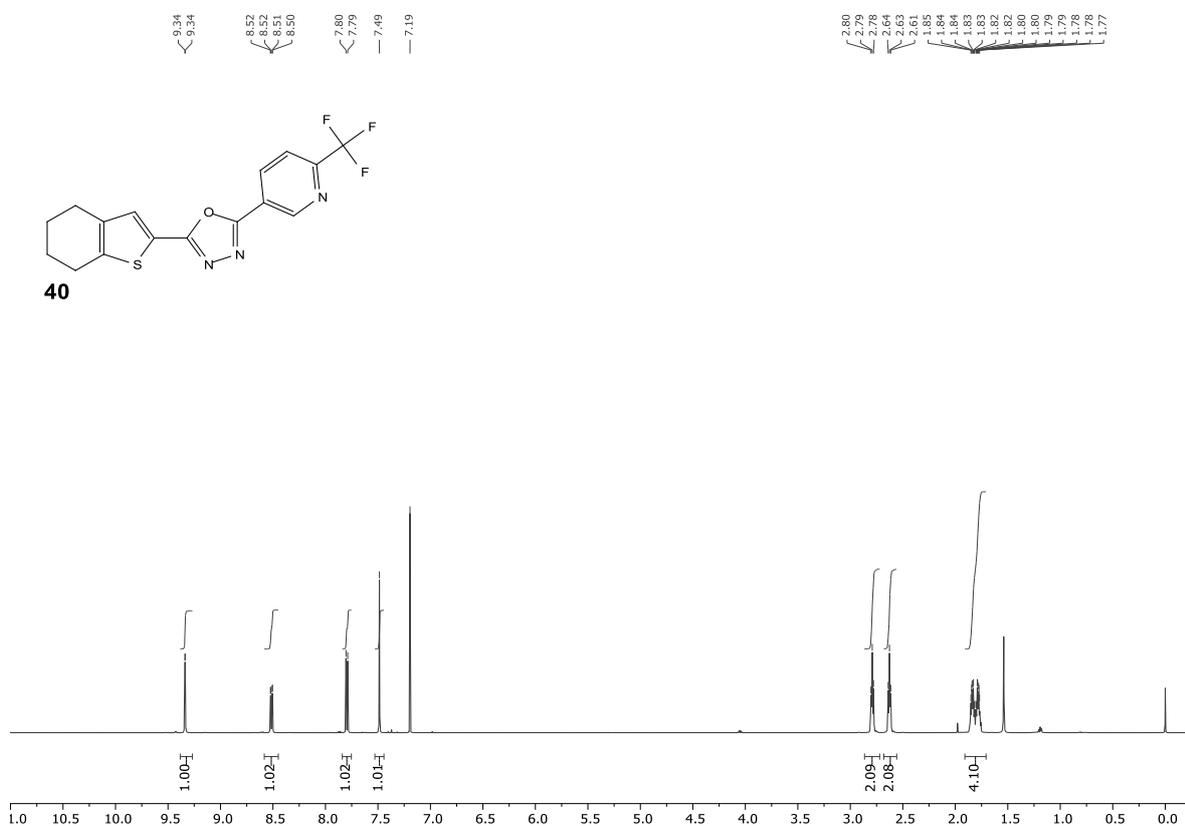
**Supplementary Figure 115**  $^{13}\text{C}$  NMR spectrum of **39**.

Supplementary Figure 116 <sup>19</sup>F NMR spectrum of 39.

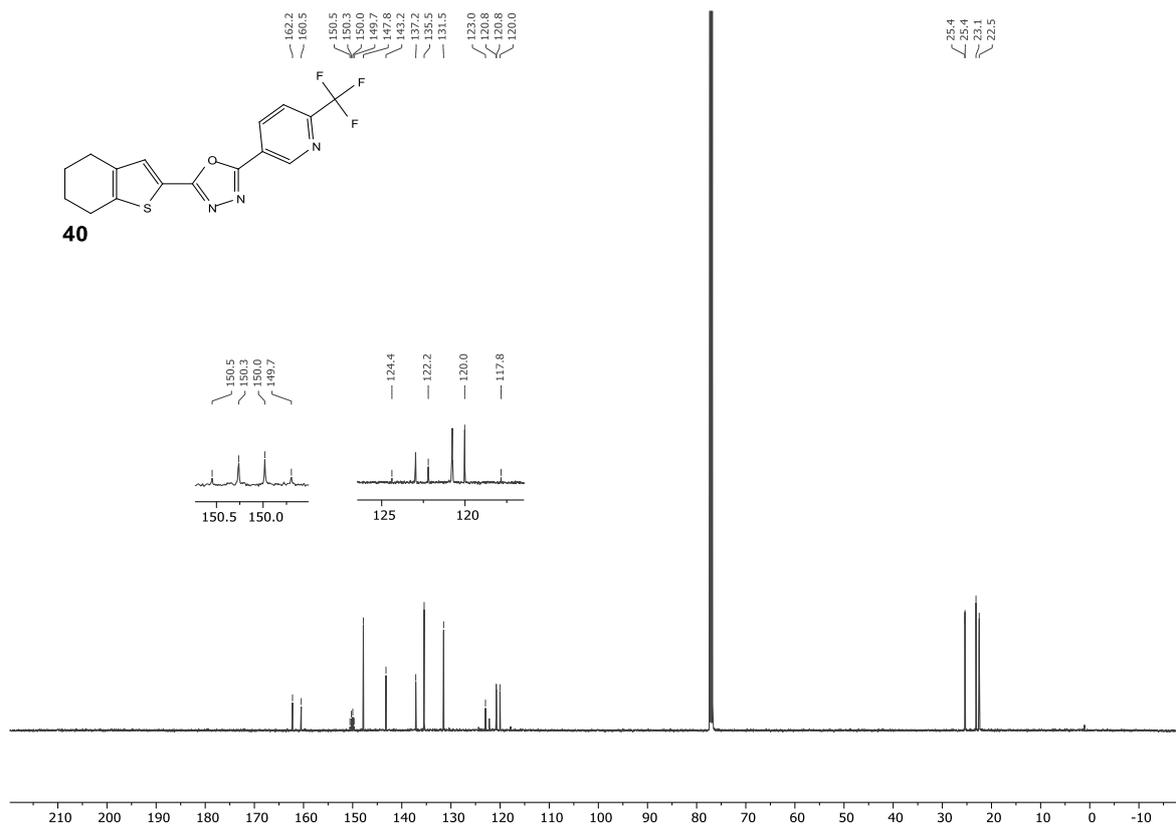
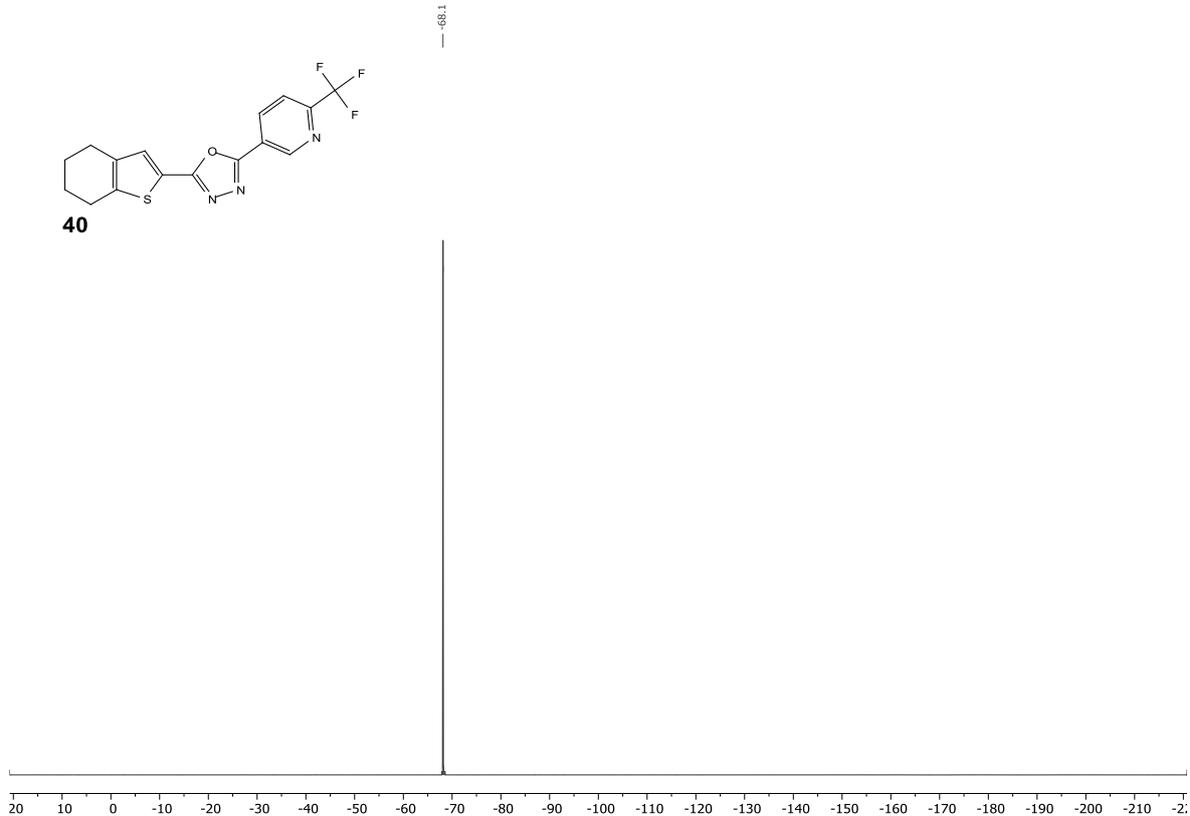
Supplementary Figure 117 HPLC purity analysis of 39.

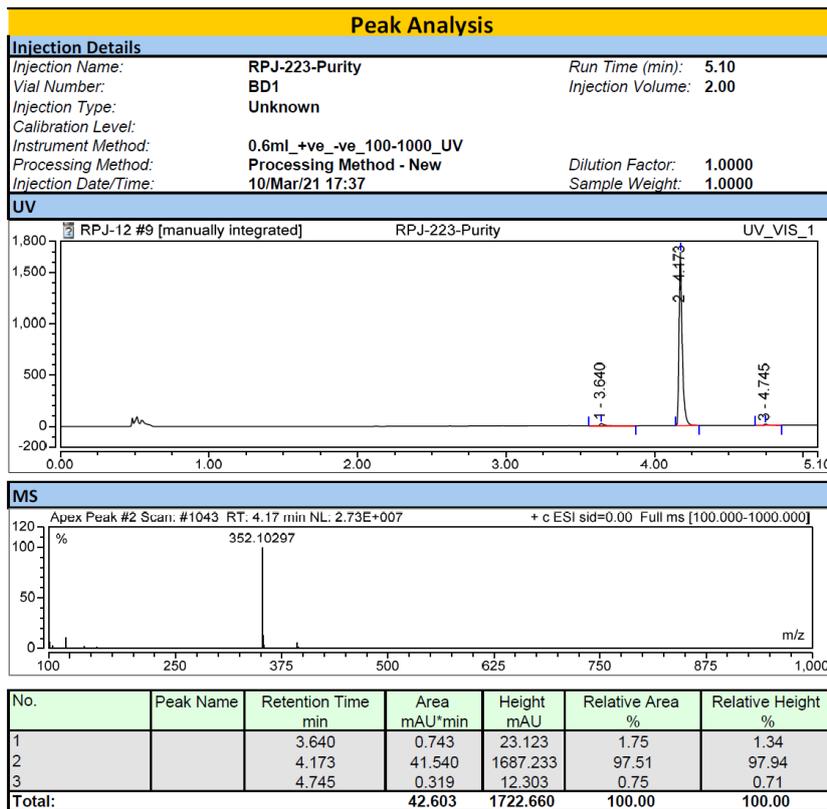


Supplementary Figure 118 HRMS of 39.

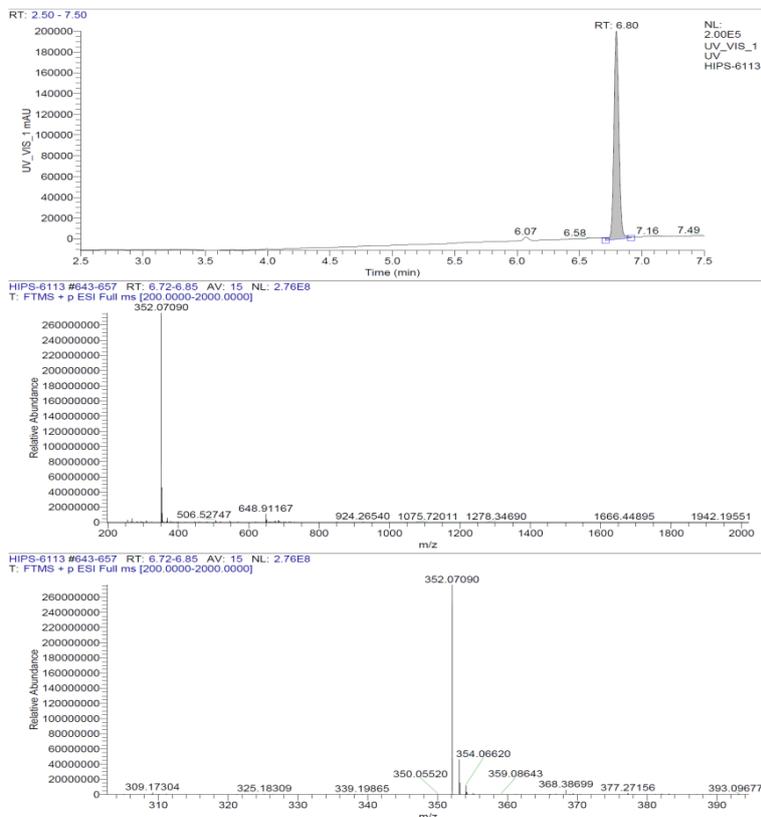


Supplementary Figure 119 <sup>1</sup>H NMR spectrum of 40.

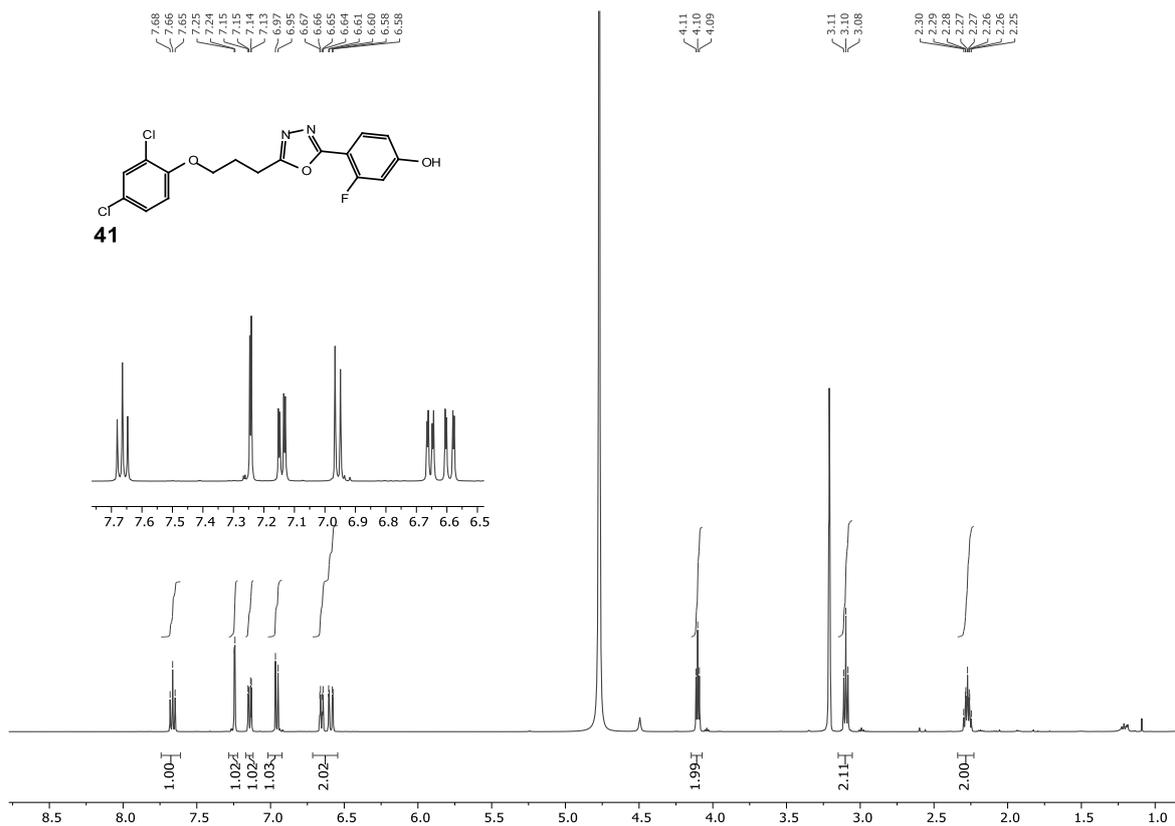
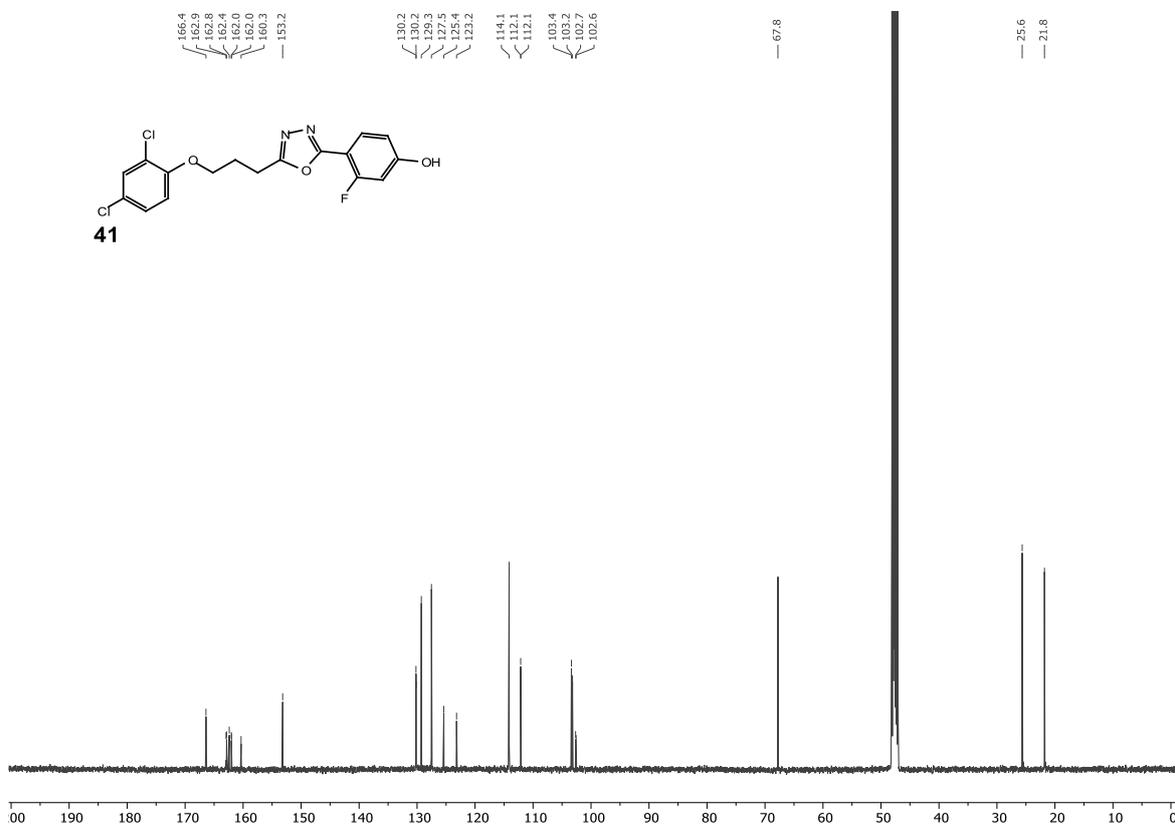
Supplementary Figure 120  $^{13}\text{C}$  NMR spectrum of **40**.Supplementary Figure 121  $^{19}\text{F}$  NMR spectrum of **40**.



Supplementary Figure 122 HPLC purity analysis of 40.



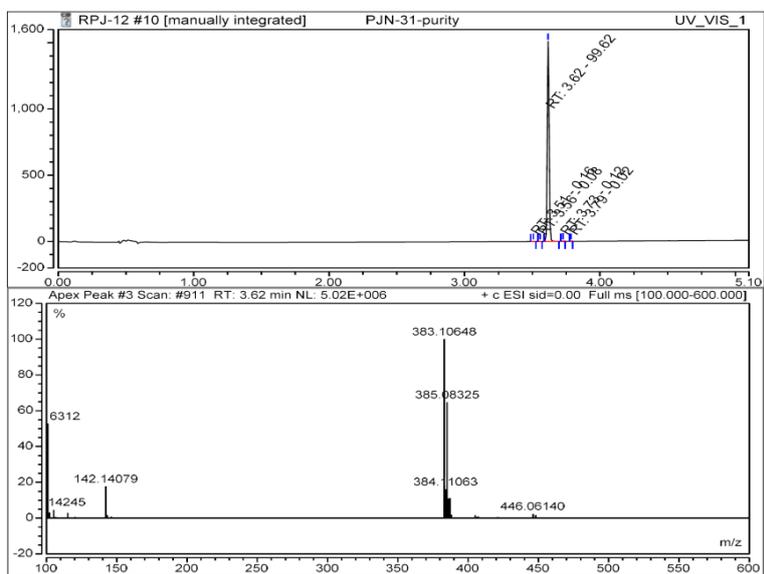
Supplementary Figure 123 HRMS of 40.

Supplementary Figure 124 <sup>1</sup>H NMR spectrum of **41**.Supplementary Figure 125 <sup>13</sup>C NMR spectrum of **41**.



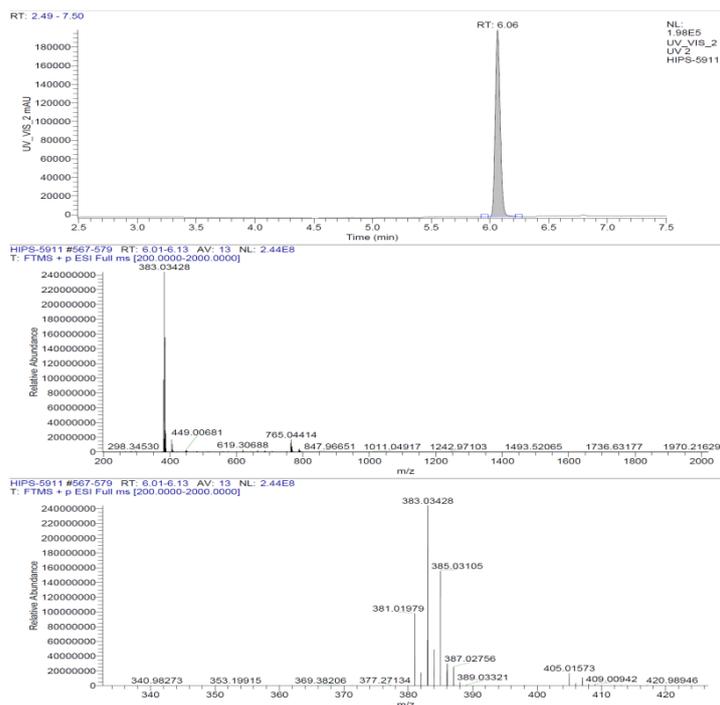
Supplementary Figure 126 <sup>19</sup>F NMR spectrum of 41.

Name PJN-31-purity  
 position BD1  
 inj. vol. 2.00  
 time 16/Sep/2020 07:44  
 instr. method 0.6mL+ve\_-ve\_100-600-all-waveLength  
 proc. method Processing Method - New

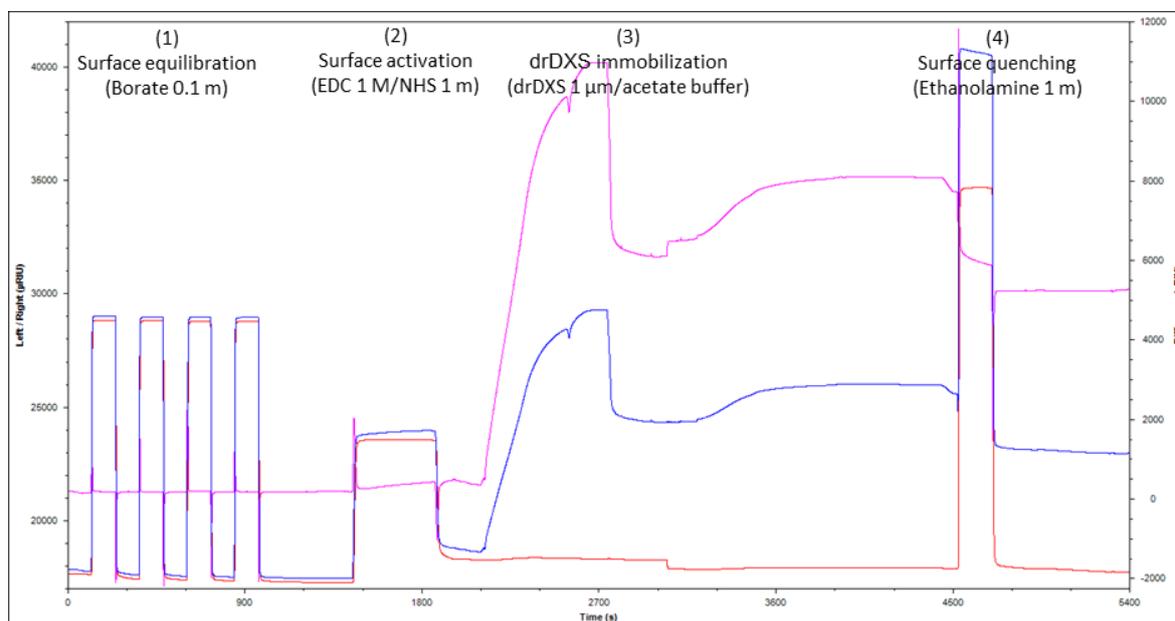


No.	Peak Name	Ret. Time min	Area mAU*min	Height mAU	Rel. Area %	Rel. Height %	Amount n.a.
1		3.508	0.036	1.770	0.16	0.12	n.a.
2		3.557	0.019	1.347	0.08	0.09	n.a.
3		3.617	22.933	1510.529	99.62	99.64	n.a.
4		3.726	0.028	1.931	0.12	0.13	n.a.
5		3.786	0.004	0.337	0.02	0.02	n.a.
Total:			23.020	1515.914	100.00	100.00	

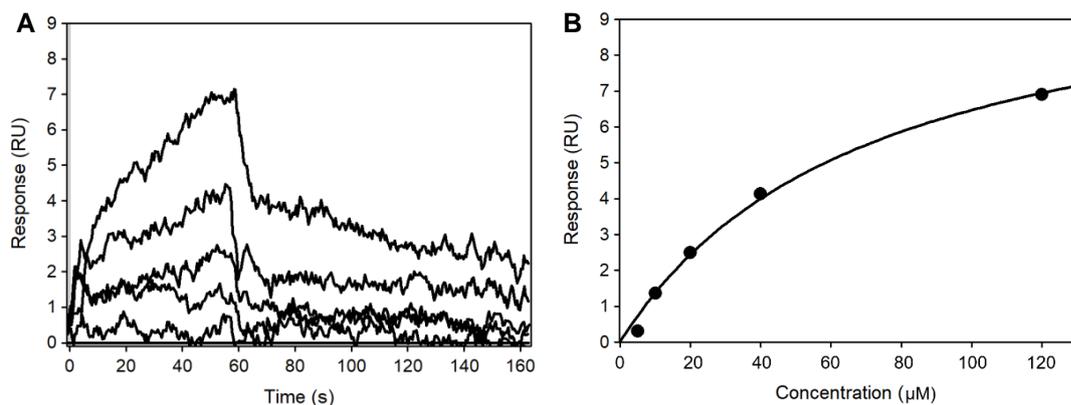
Supplementary Figure 127 HPLC purity analysis of 41.



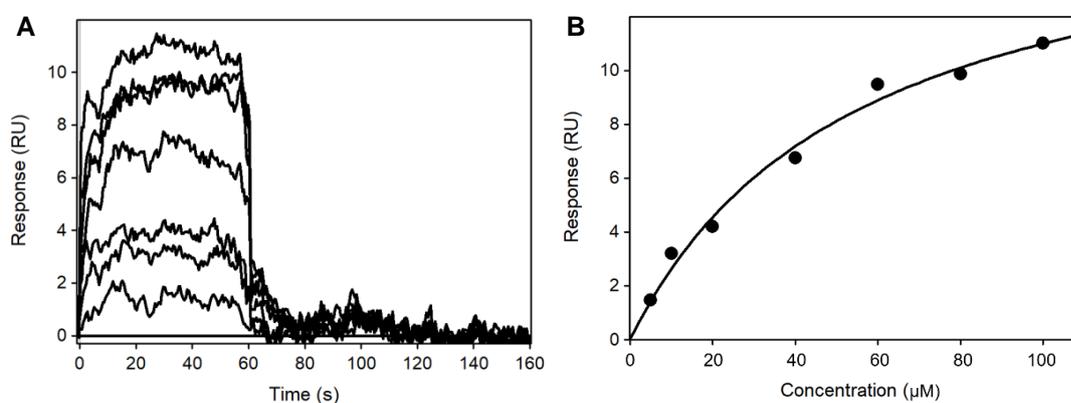
Supplementary Figure 128 HRMS of 41.



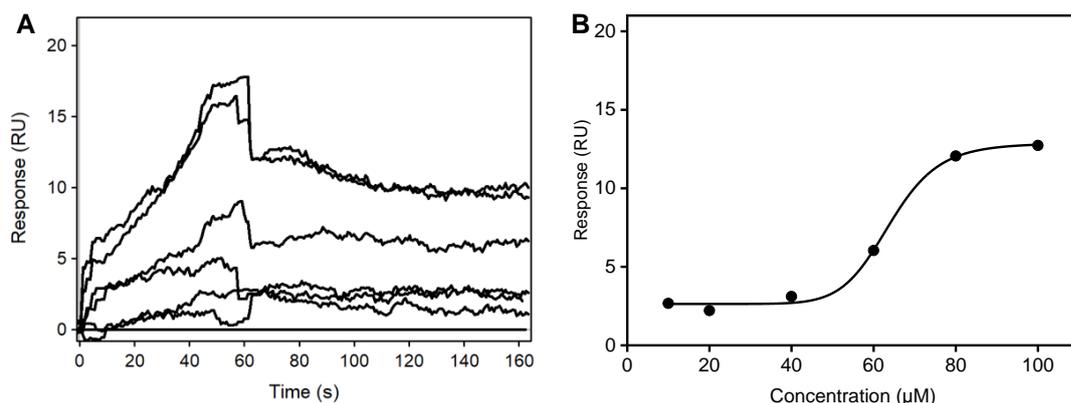
**Supplementary Figure 129** Sensorgram of the immobilization procedure for drDXPS on CMD500M sensor chip: (1) Four injections of cleaning solution, (2) activation solution, (3) drDXPS, and (4) quenching solution. The blue, red, and magenta curves represent the left (active) channel, right (reference) channel, and the difference, respectively.



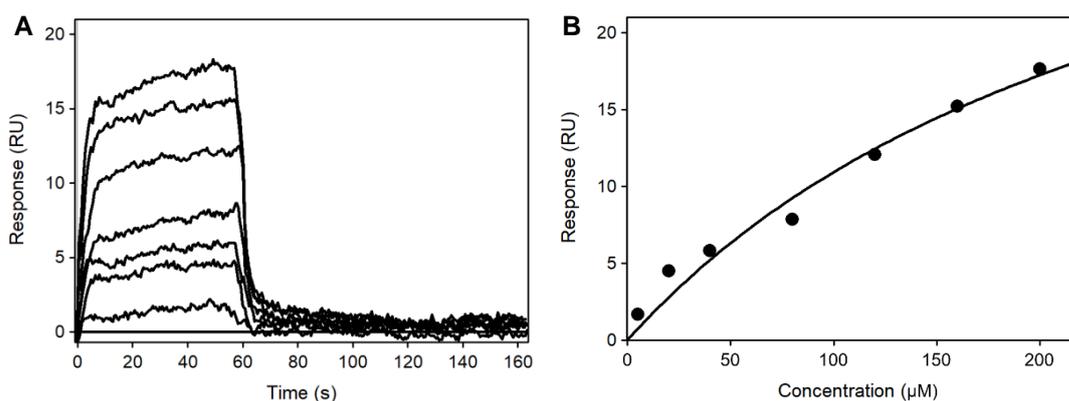
**Supplementary Figure 130** (A) Overlay of sensorgrams of **1** injected at concentrations 5–120  $\mu\text{M}$  over an immobilized drDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $70 \pm 5 \mu\text{M}$ ).



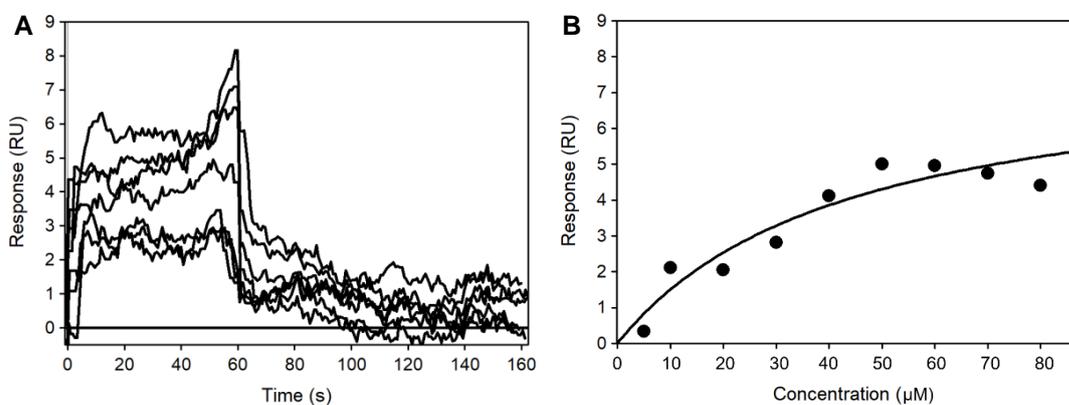
**Supplementary Figure 131** (A) Overlay of sensorgrams of **2** injected at concentrations 5–100  $\mu\text{M}$  over an immobilized drDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $55 \pm 5 \mu\text{M}$ ).



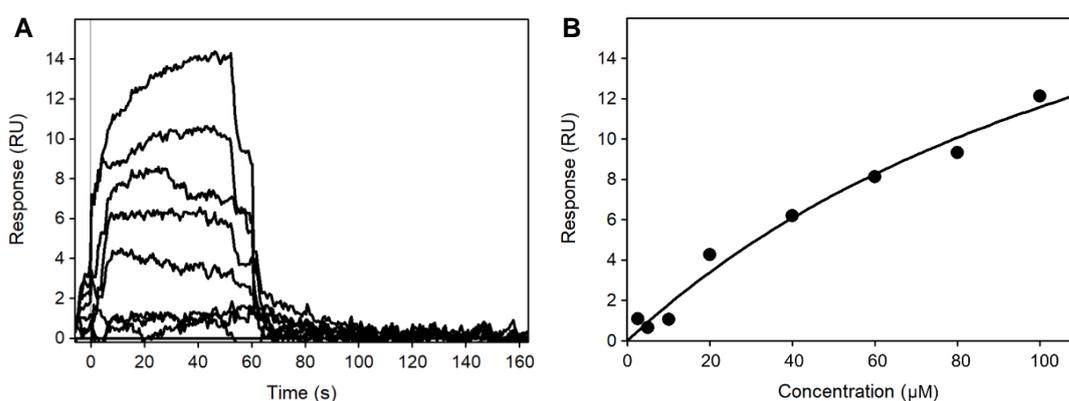
**Supplementary Figure 132** (A) Overlay of sensorgrams of **3** injected at concentrations 10–100  $\mu\text{M}$  over an immobilized drDXPS; (B) Fitting of responses at equilibrium according to Hill equation ( $K_D$ :  $64 \pm 2 \mu\text{M}$ ).



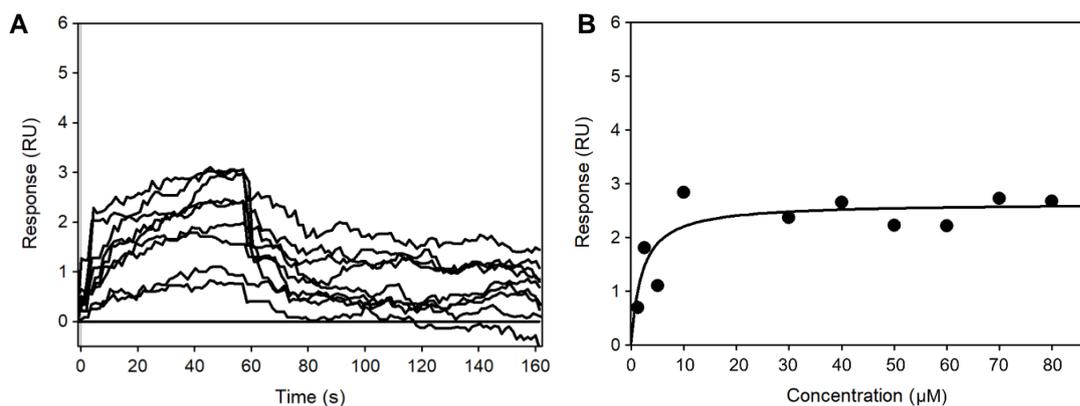
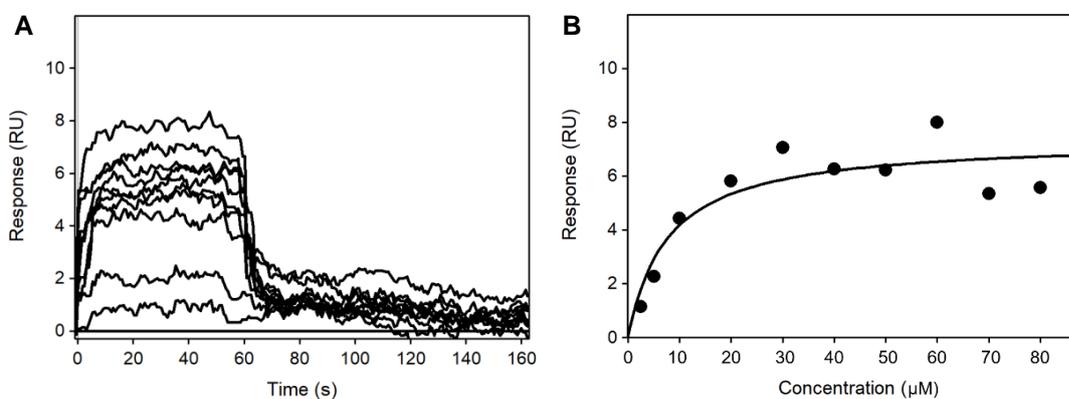
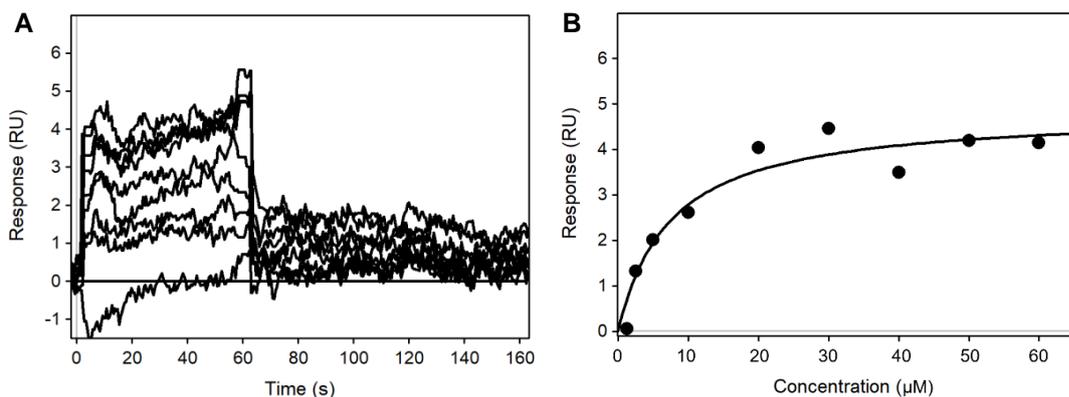
**Supplementary Figure 133.** (A) Overlay of sensorgrams of **5** injected at concentrations 5–200  $\mu\text{M}$  over an immobilized drDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $270 \pm 40 \mu\text{M}$ ).

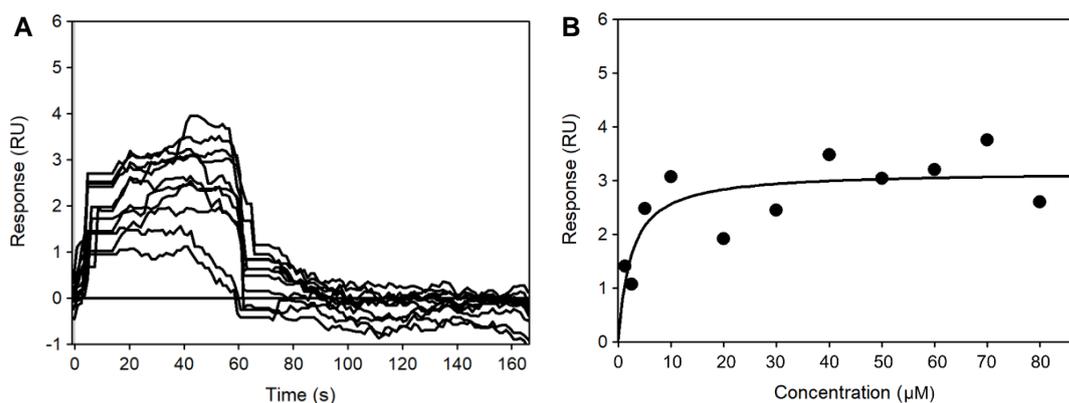


**Supplementary Figure 134** (A) Overlay of sensorgrams of **7** injected at concentrations 5–80  $\mu\text{M}$  over an immobilized drDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $40 \pm 10 \mu\text{M}$ ).

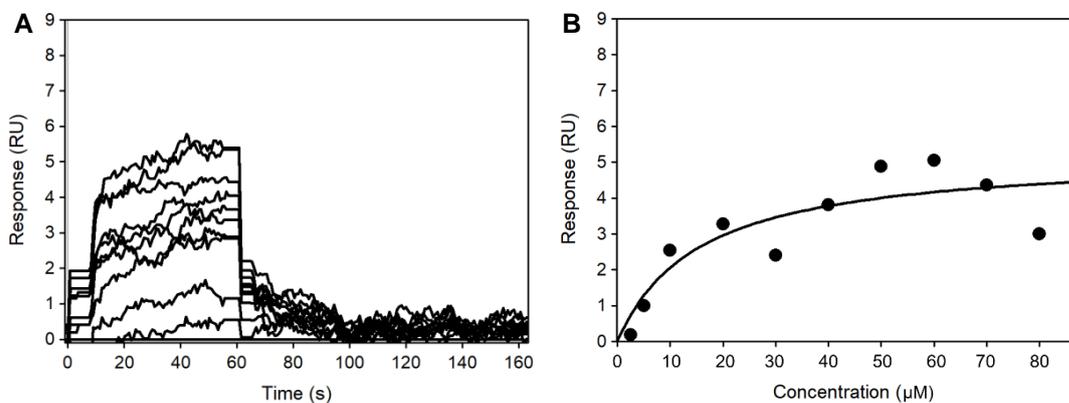


**Supplementary Figure 135** (A) Overlay of sensorgrams of **9** injected at concentrations 2.5–80  $\mu\text{M}$  over an immobilized drDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $150 \pm 30 \mu\text{M}$ ).

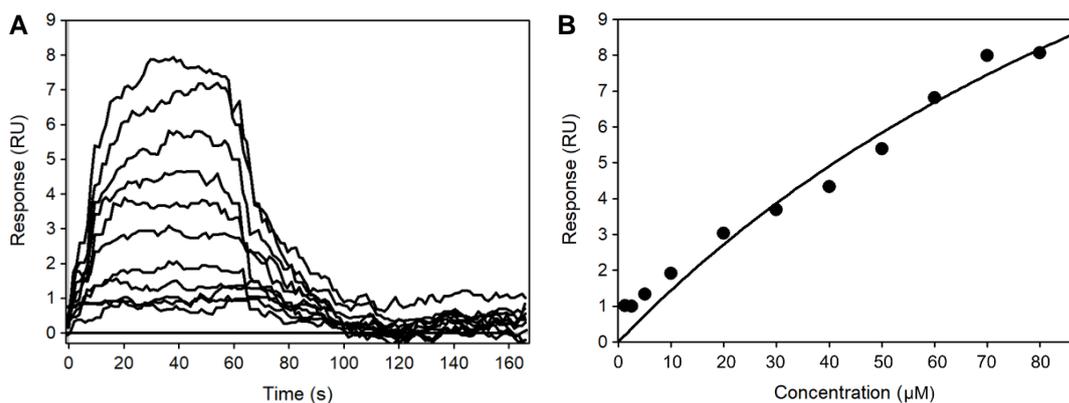




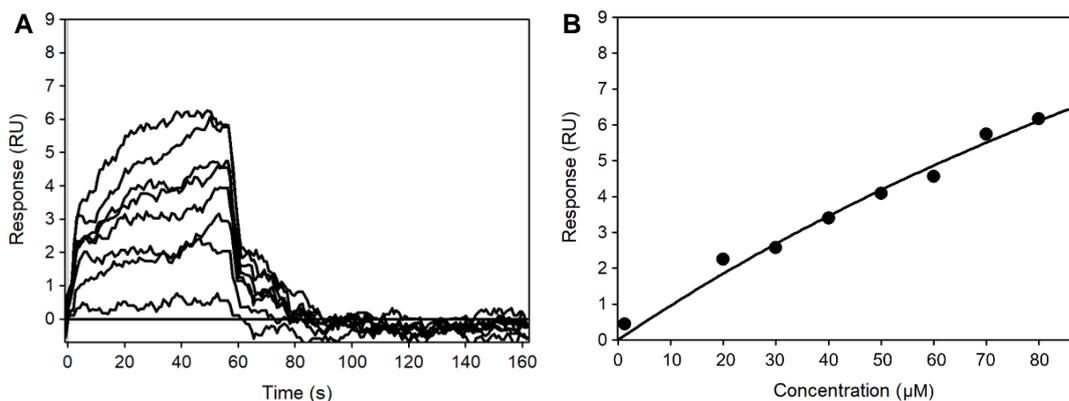
**Supplementary Figure 139** (A) Overlay of sensorgrams of **21** injected at concentrations 1.25–80  $\mu\text{M}$  over an immobilized drDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $3 \pm 1 \mu\text{M}$ ).



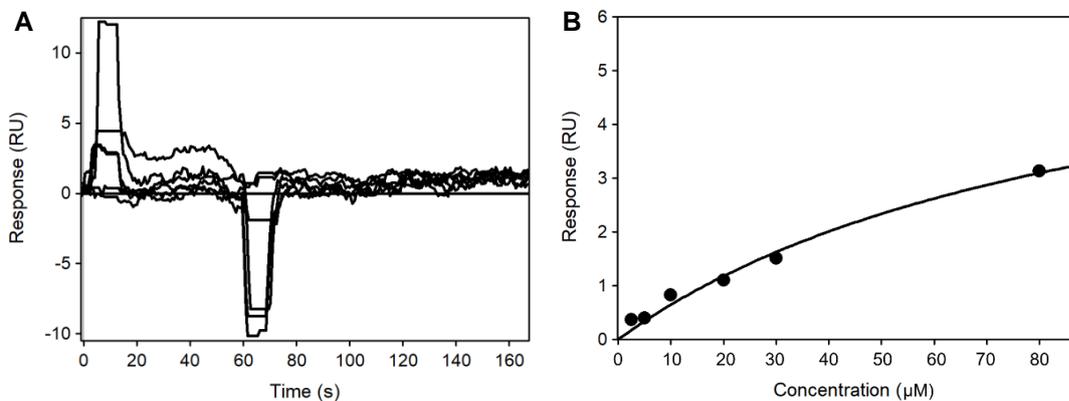
**Supplementary Figure 140** (A) Overlay of sensorgrams of **22** injected at concentrations 1.25–80  $\mu\text{M}$  over an immobilized drDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $15 \pm 3 \mu\text{M}$ ).



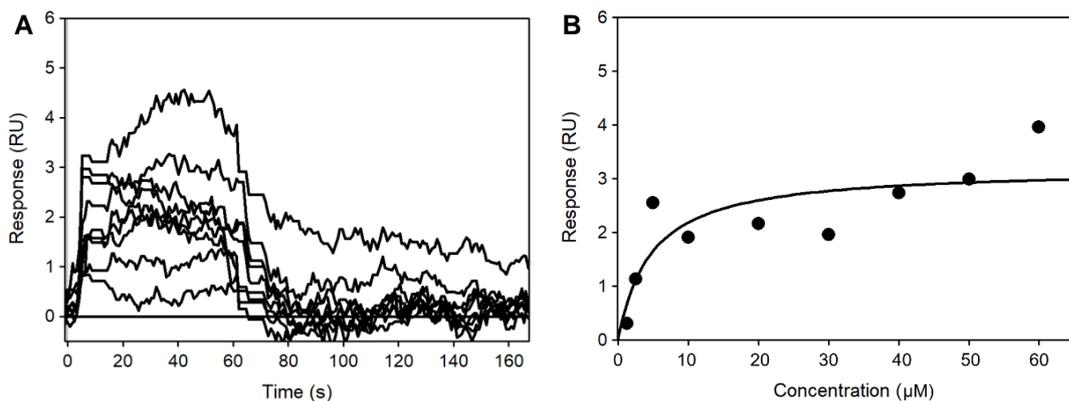
**Supplementary Figure 141** (A) Overlay of sensorgrams of **23** injected at concentrations 1.25–80  $\mu\text{M}$  over an immobilized drDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $160 \pm 50 \mu\text{M}$ ).



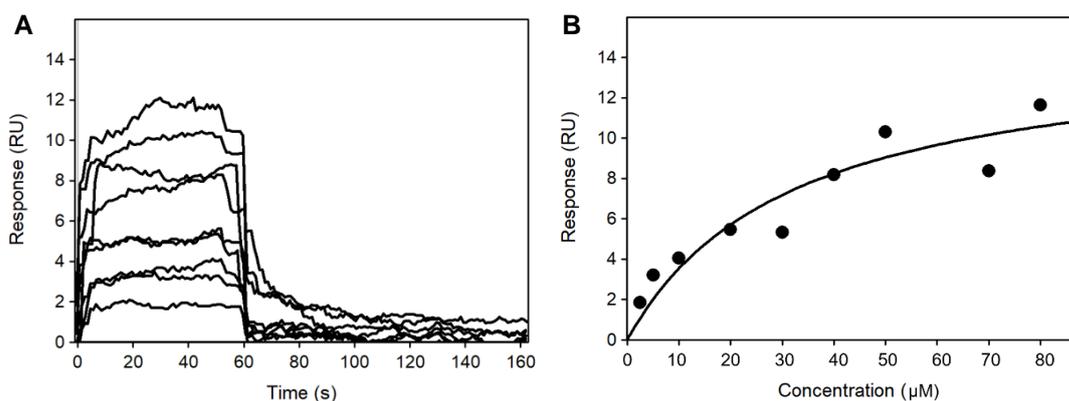
**Supplementary Figure 142** (A) Overlay of sensorgrams of **24** injected at concentrations 1.25–80  $\mu\text{M}$  over an immobilized drDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ : 260  $\pm$  80  $\mu\text{M}$ ).



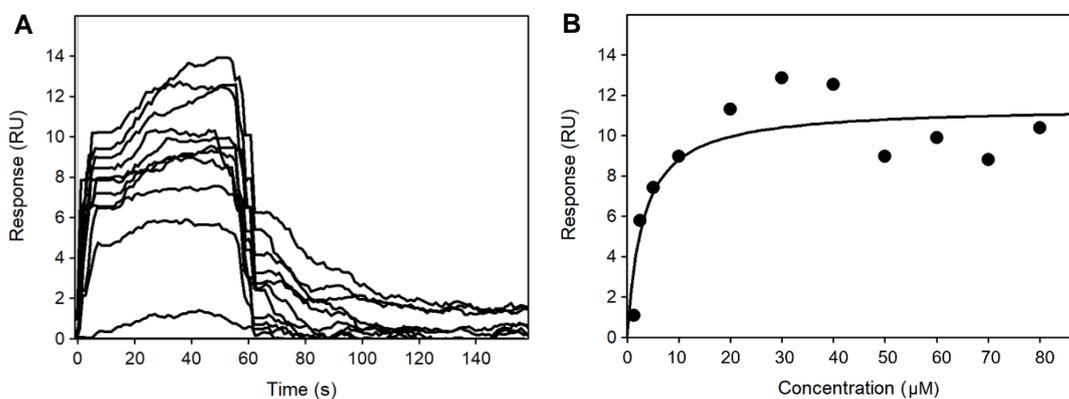
**Supplementary Figure 143** (A) Overlay of sensorgrams of **25** injected at concentrations 1.25–80  $\mu\text{M}$  over an immobilized drDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ : 90  $\pm$  20  $\mu\text{M}$ ).



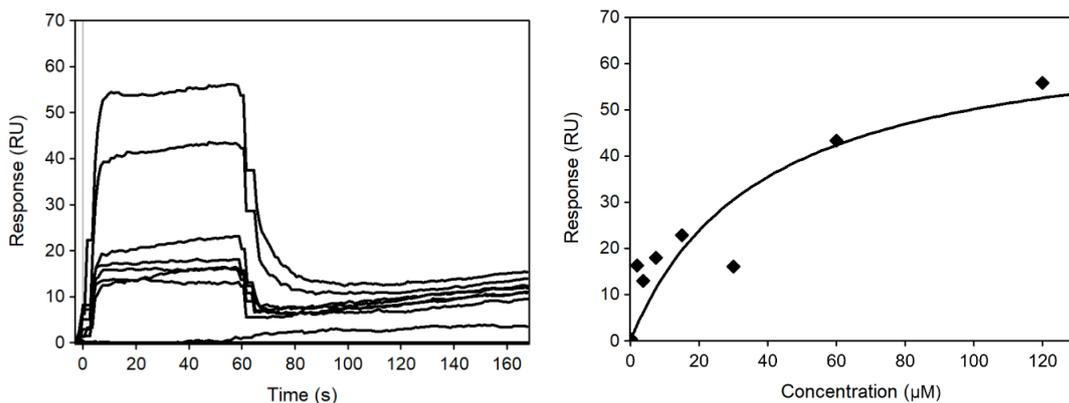
**Supplementary Figure 144** (A) Overlay of sensorgrams of **26** injected at concentrations 1.25–60  $\mu\text{M}$  over an immobilized drDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ : 5  $\pm$  2  $\mu\text{M}$ ).



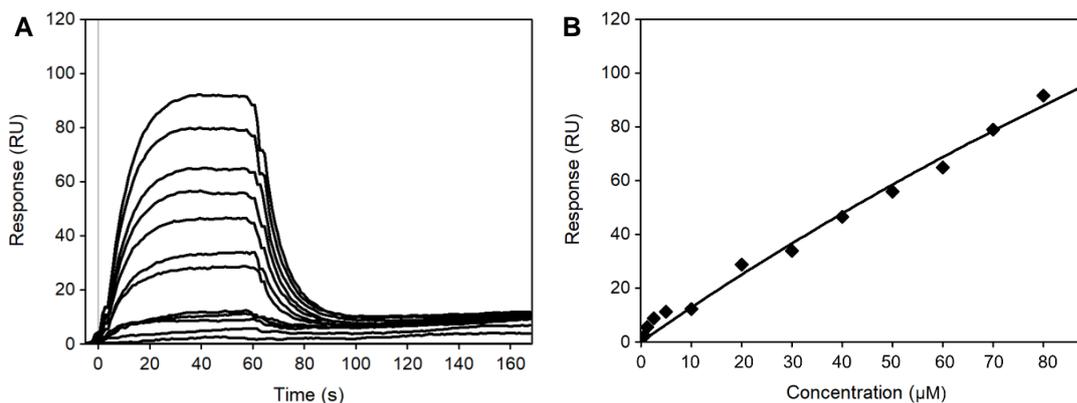
**Supplementary Figure 145** (A) Overlay of sensorgrams of **35** injected at concentrations 2.5–80  $\mu\text{M}$  over an immobilized drDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $32 \pm 8 \mu\text{M}$ ).



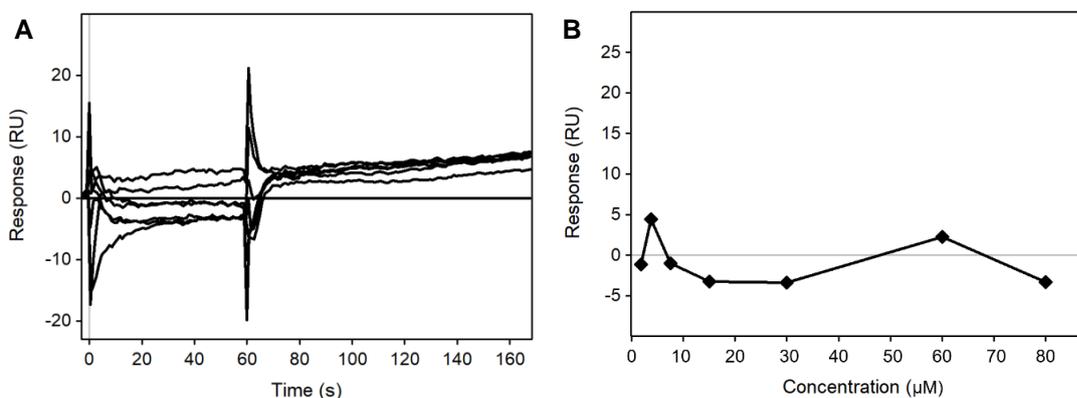
**Supplementary Figure 146** (A) Overlay of sensorgrams of **37** injected at concentrations 1.25–80  $\mu\text{M}$  over an immobilized drDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $3 \pm 1 \mu\text{M}$ ).



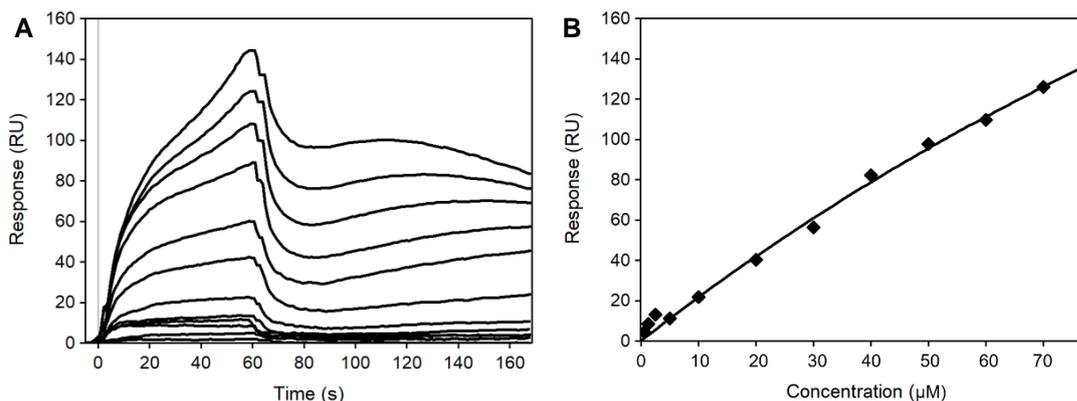
**Supplementary Figure 147** (A) Overlay of sensorgrams of **38** injected at concentrations 0.2–120  $\mu\text{M}$  over an immobilized drDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $40 \pm 10 \mu\text{M}$ ).



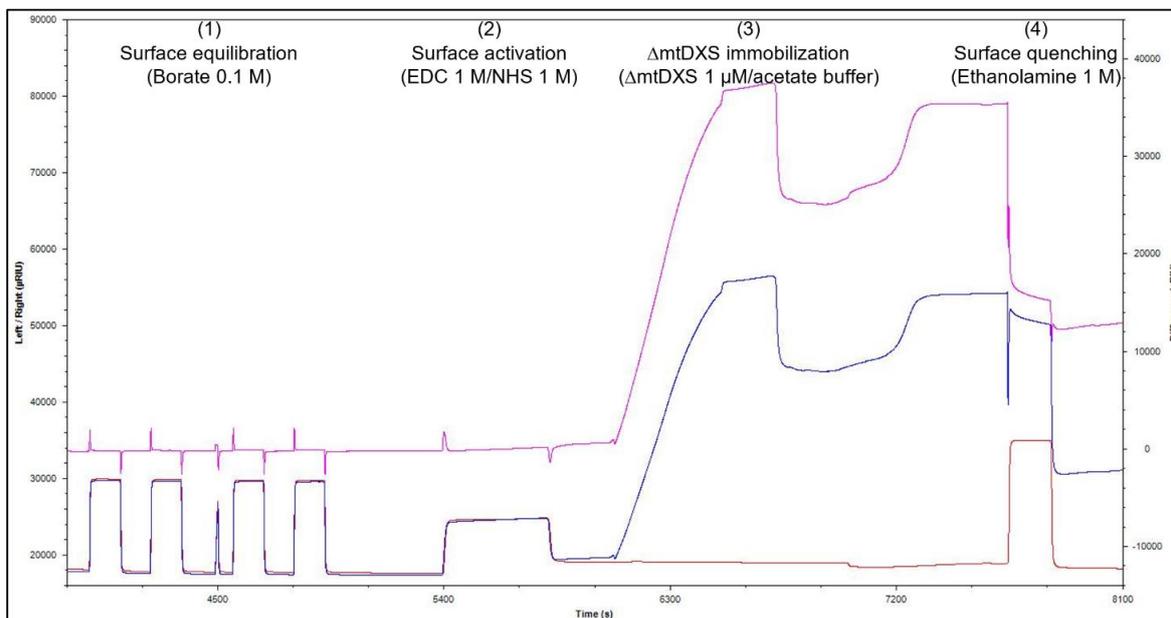
**Supplementary Figure 148** (A) Overlay of sensorgrams of **39** injected at concentrations 0.6–80 μM over an immobilized drDXS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ : 290 ± 30 μM).



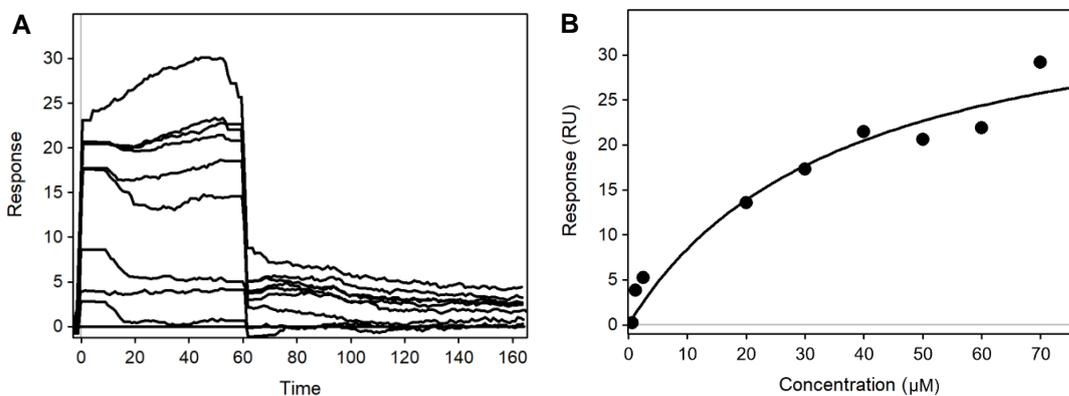
**Supplementary Figure 149** (A) Overlay of sensorgrams of **40** injected at concentrations 1.9–80 μM over an immobilized drDXS; (B) Responses at equilibrium are concentration-independent and could not be fitted to Langmuir nor Hill binding model, indicating no significant binding.



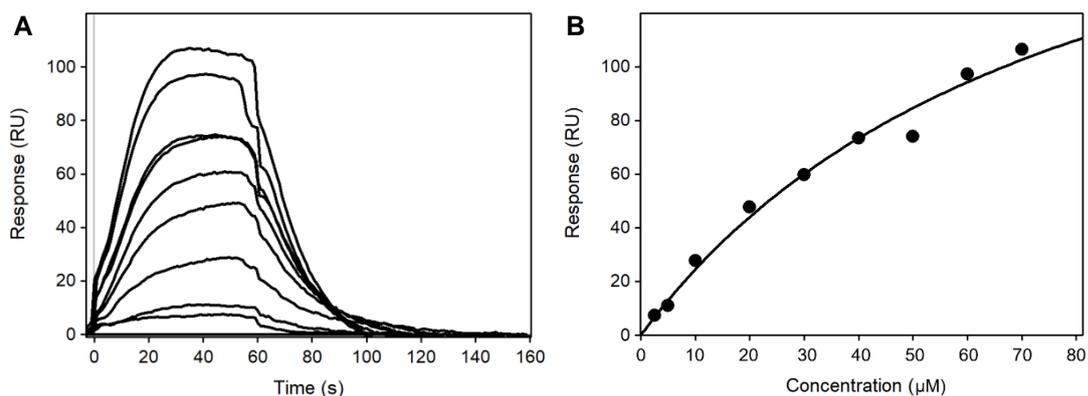
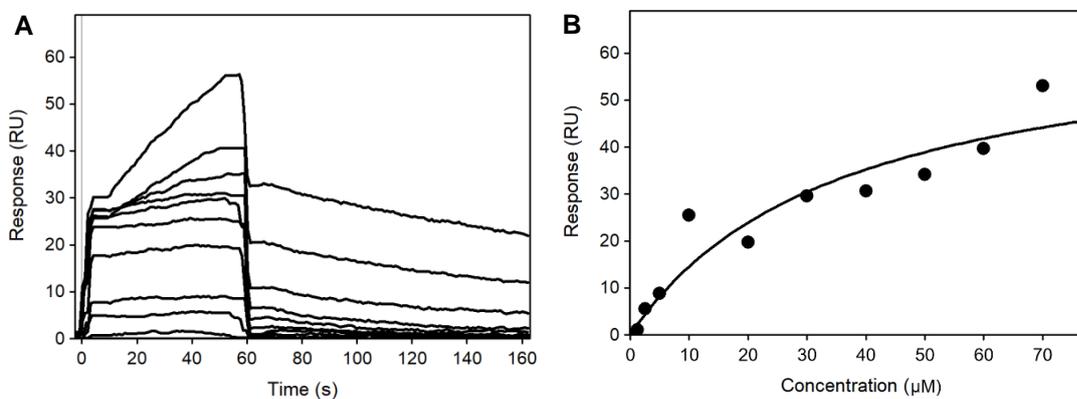
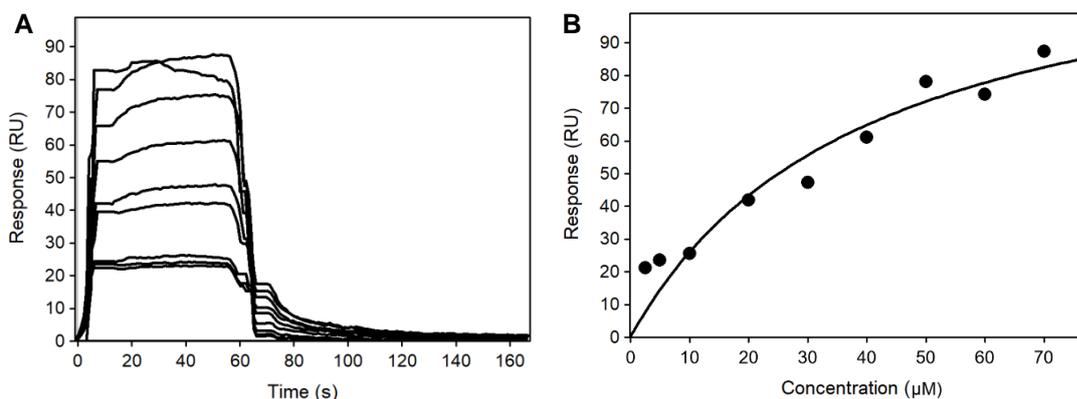
**Supplementary Figure 150** (A) Overlay of sensorgrams of **41** injected at concentrations 0.6–70 μM over an immobilized drDXS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ : 290 ± 40 μM).

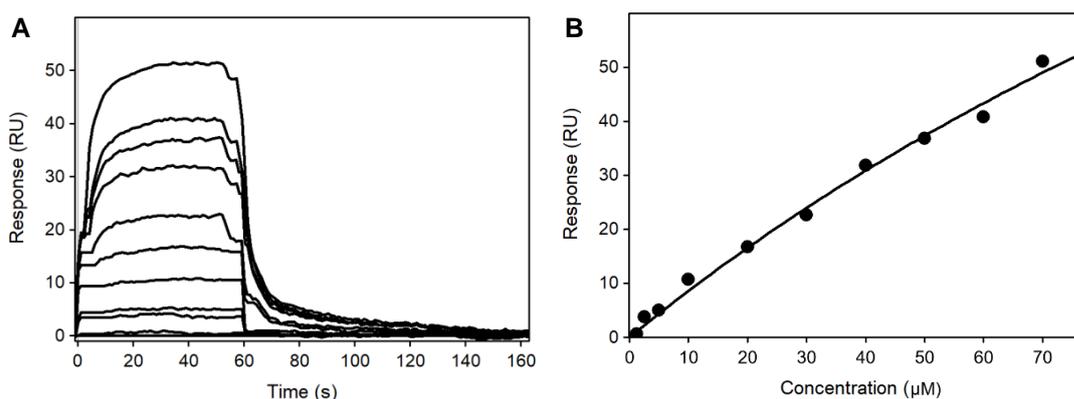


**Supplementary Figure 151** Sensorgram of the immobilization procedure for  $\Delta$ mtDXPS on CMD500M sensor chip: (1) Four injections of cleaning solution, (2) activation solution, (3)  $\Delta$ mtDXPS, and (4) quenching solution. The blue, red, and magenta curves represent the left (active) channel, right (reference) channel, and the difference, respectively.

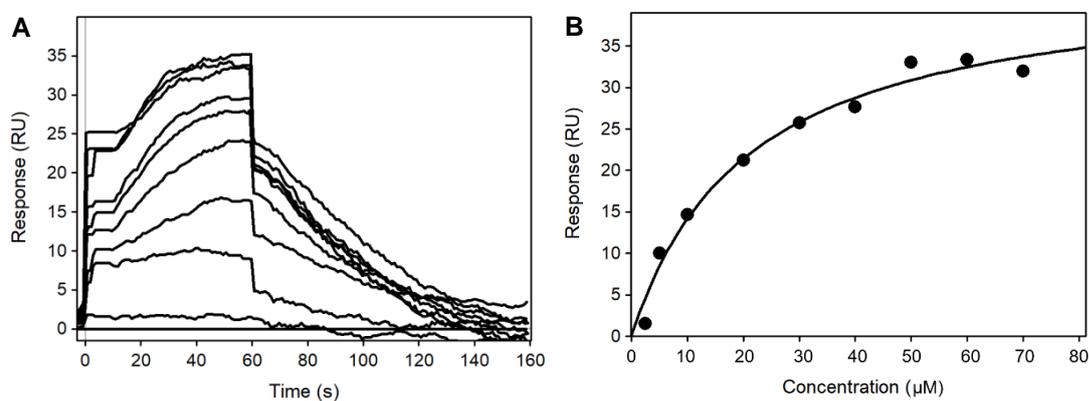


**Supplementary Figure 152** (A) Overlay of sensorgrams of **1** injected at concentrations 1.25–70  $\mu$ M over an immobilized  $\Delta$ mtDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $40 \pm 10 \mu$ M).

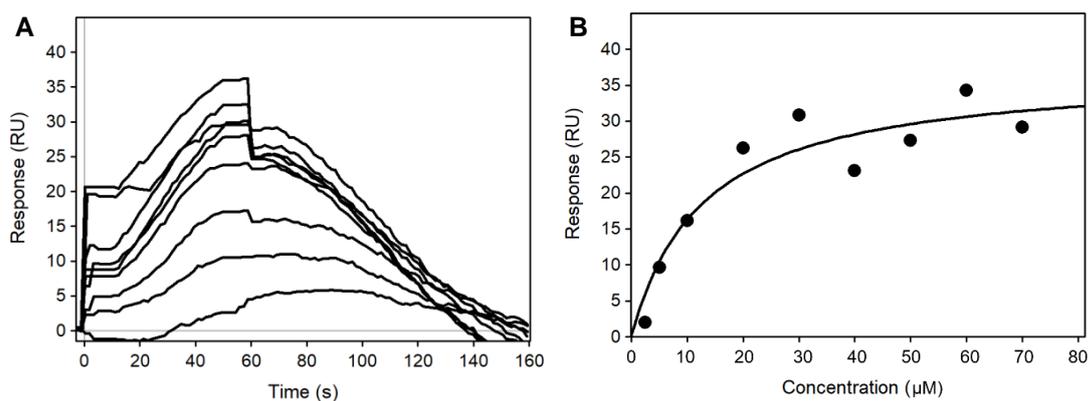




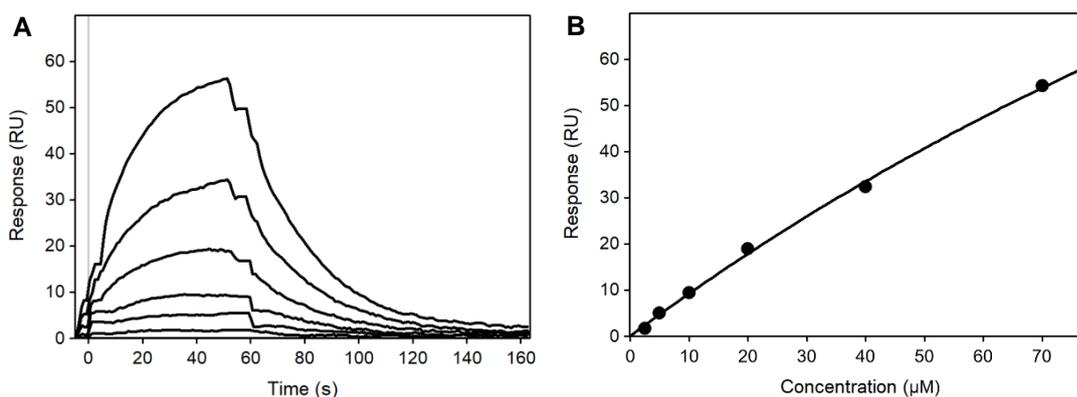
**Supplementary Figure 156** (A) Overlay of sensorgrams of **5** injected at concentrations 1.25–70  $\mu\text{M}$  over an immobilized  $\Delta\text{mtDXPS}$ ; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ : 250  $\pm$  20  $\mu\text{M}$ ).



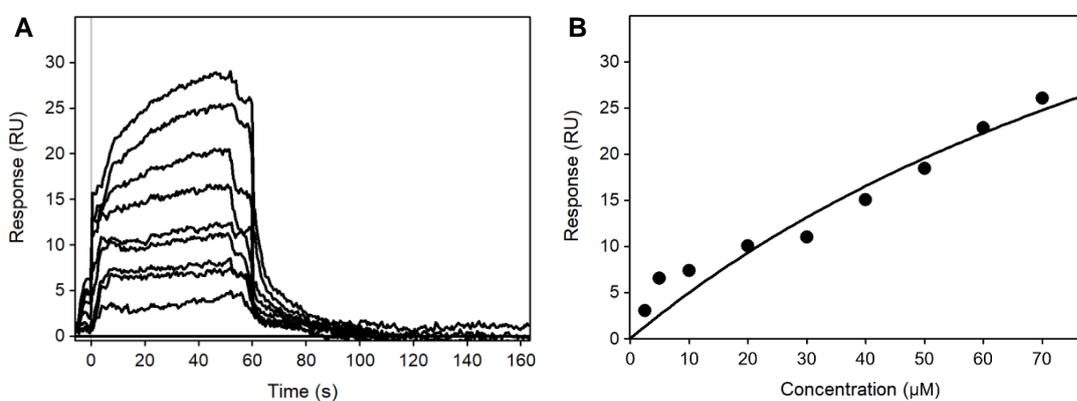
**Supplementary Figure 157** (A) Overlay of sensorgrams of **6** injected at concentrations 2.5–70  $\mu\text{M}$  over an immobilized  $\Delta\text{mtDXPS}$ ; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ : 21  $\pm$  3  $\mu\text{M}$ ).



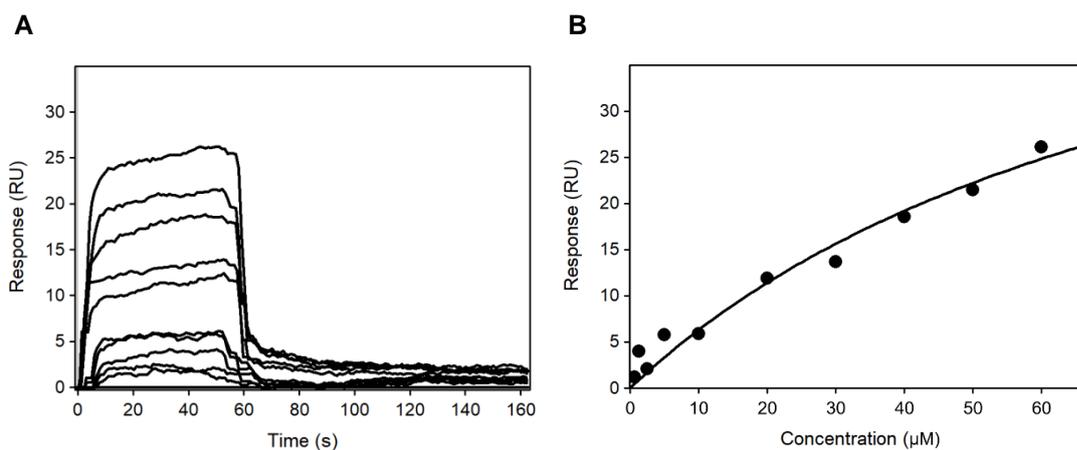
**Supplementary Figure 158** (A) Overlay of sensorgrams of **7** injected at concentrations 2.5–70  $\mu\text{M}$  over an immobilized  $\Delta\text{mtDXPS}$ ; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ : 13  $\pm$  3  $\mu\text{M}$ ).



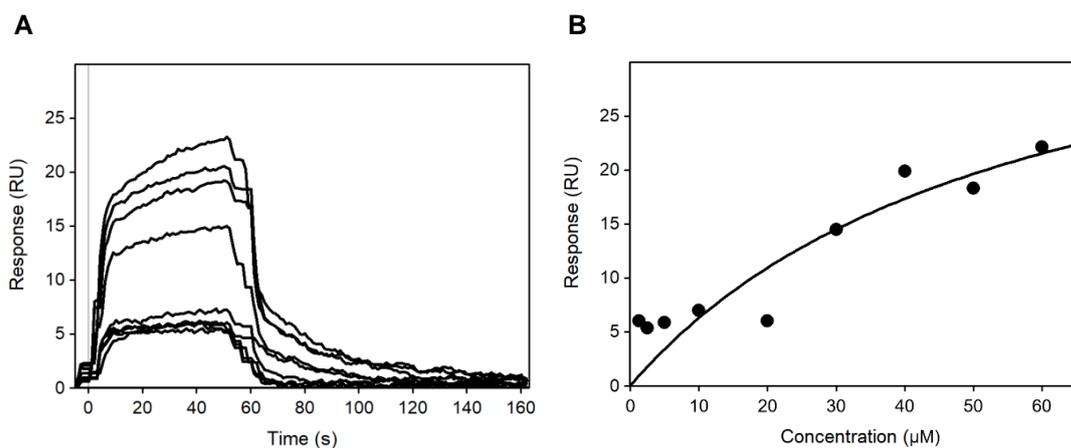
**Supplementary Figure 159** (A) Overlay of sensorgrams of **8** injected at concentrations 2.5–70  $\mu\text{M}$  over an immobilized  $\Delta\text{mtDXPS}$ ; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $230 \pm 20 \mu\text{M}$ ).



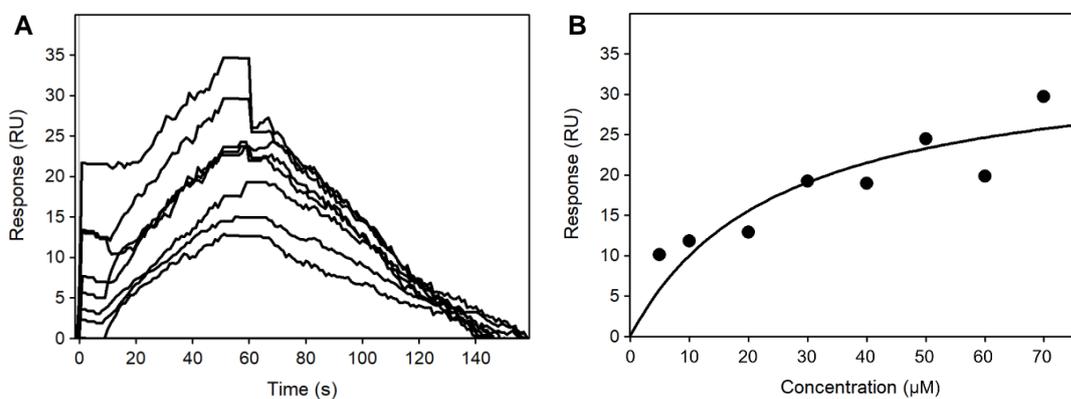
**Supplementary Figure 160** (A) Overlay of sensorgrams of **9** injected at concentrations 2.5–70  $\mu\text{M}$  over an immobilized  $\Delta\text{mtDXPS}$ ; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $140 \pm 50 \mu\text{M}$ ).



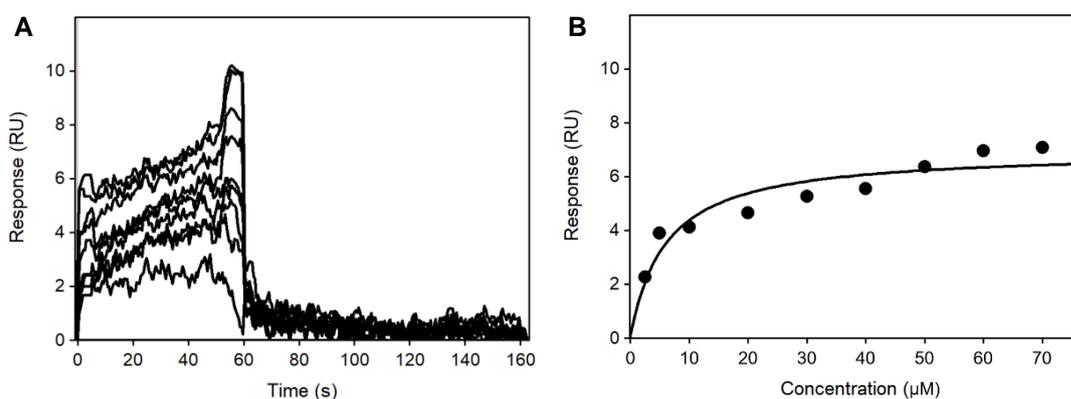
**Supplementary Figure 161** (A) Overlay of sensorgrams of **11** injected at concentrations 0.625–60  $\mu\text{M}$  over an immobilized  $\Delta\text{mtDXPS}$ ; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $90 \pm 30 \mu\text{M}$ ).



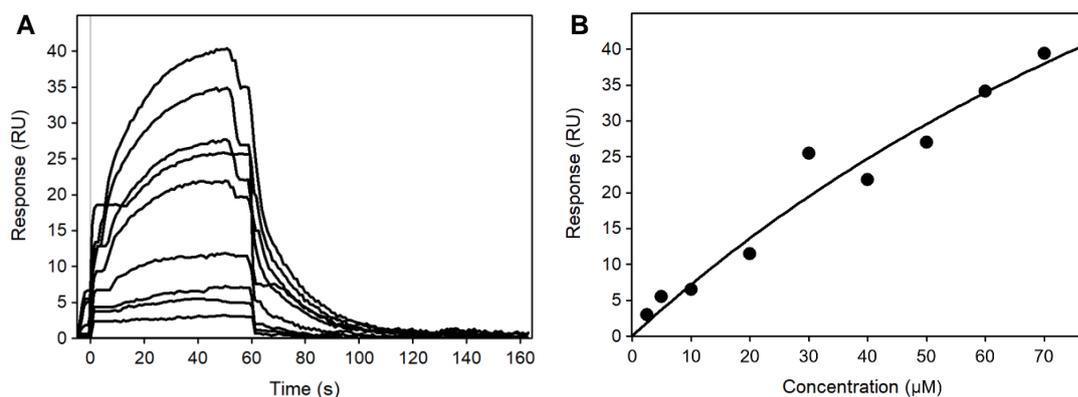
**Supplementary Figure 162** (A) Overlay of sensorgrams of **12** injected at concentrations 1.25–60  $\mu\text{M}$  over an immobilized  $\Delta\text{mtDXPS}$ ; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $60 \pm 20 \mu\text{M}$ ).



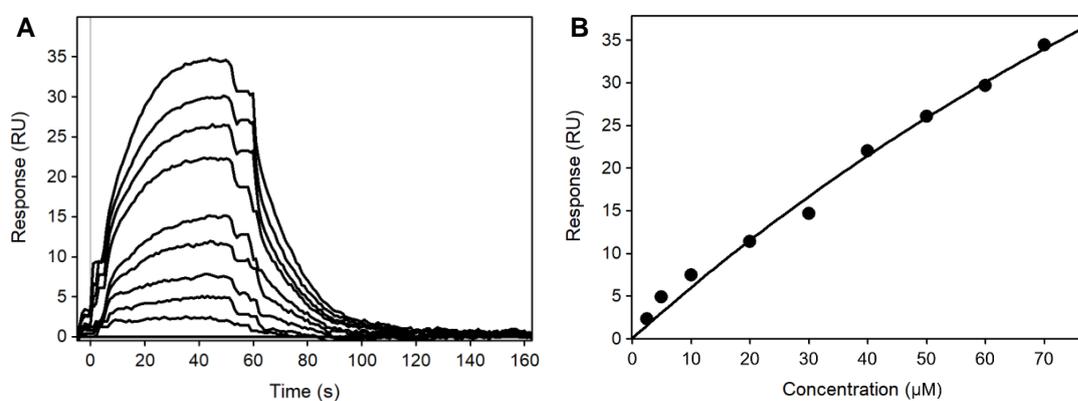
**Supplementary Figure 163** (A) Overlay of sensorgrams of **21** injected at concentrations 5–70  $\mu\text{M}$  over an immobilized  $\Delta\text{mtDXPS}$ ; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $25 \pm 8 \mu\text{M}$ ).



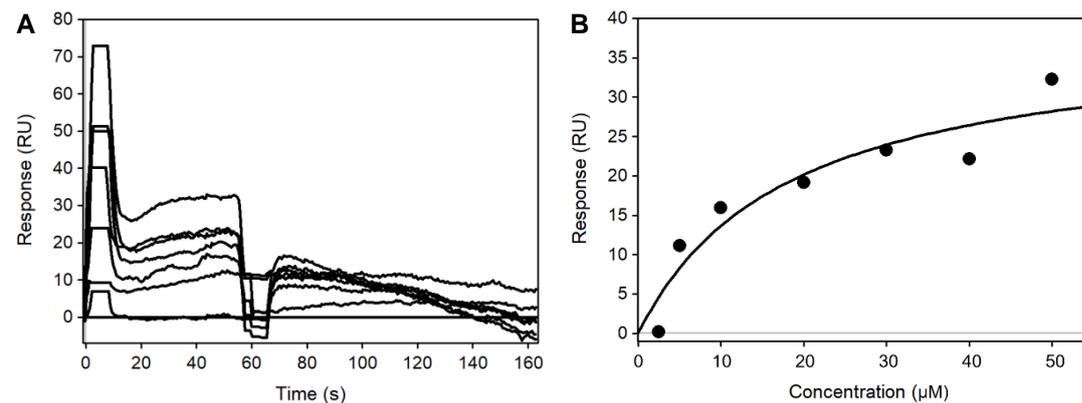
**Supplementary Figure 164** (A) Overlay of sensorgrams of **22** injected at concentrations 2.5–70  $\mu\text{M}$  over an immobilized  $\Delta\text{mtDXPS}$ ; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $6 \pm 1 \mu\text{M}$ ).



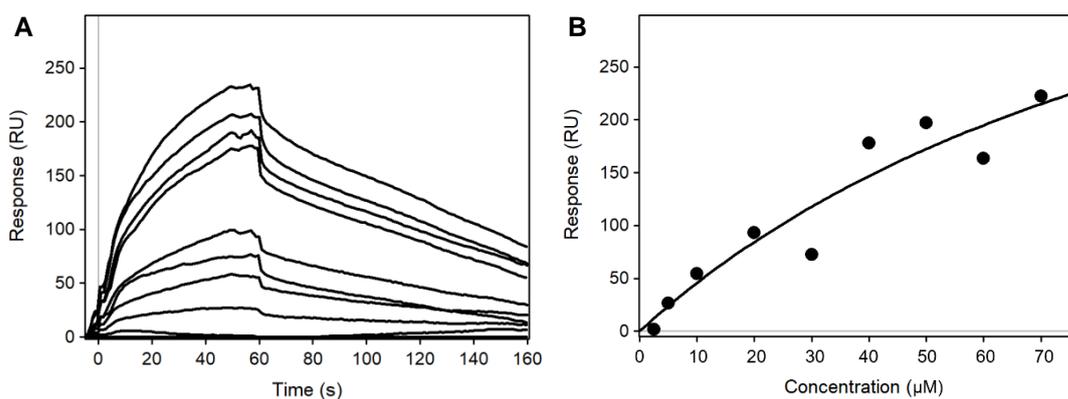
**Supplementary Figure 165** (A) Overlay of sensorgrams of **23** injected at concentrations 2.5–70  $\mu\text{M}$  over an immobilized  $\Delta\text{mtDXPS}$ . (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $160 \pm 40 \mu\text{M}$ )



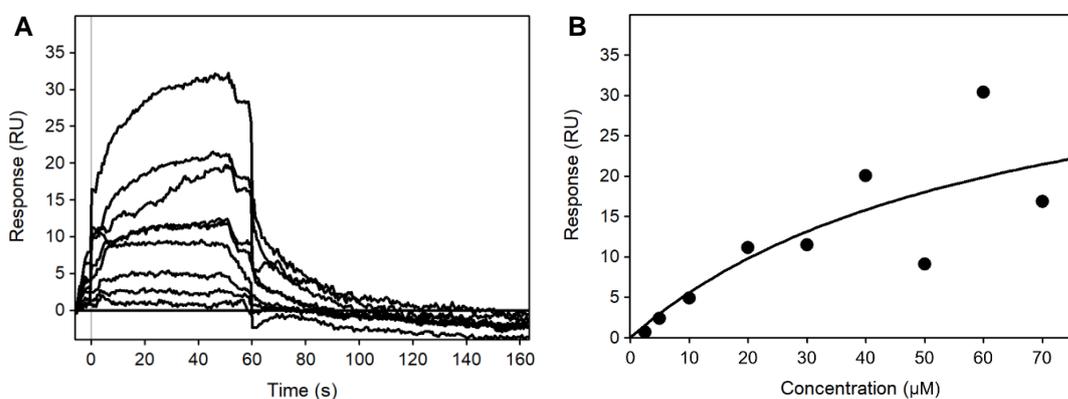
**Supplementary Figure 166** (A) Overlay of sensorgrams of **24** injected at concentrations 2.5–70  $\mu\text{M}$  over an immobilized  $\Delta\text{mtDXPS}$ ; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $250 \pm 60 \mu\text{M}$ ).



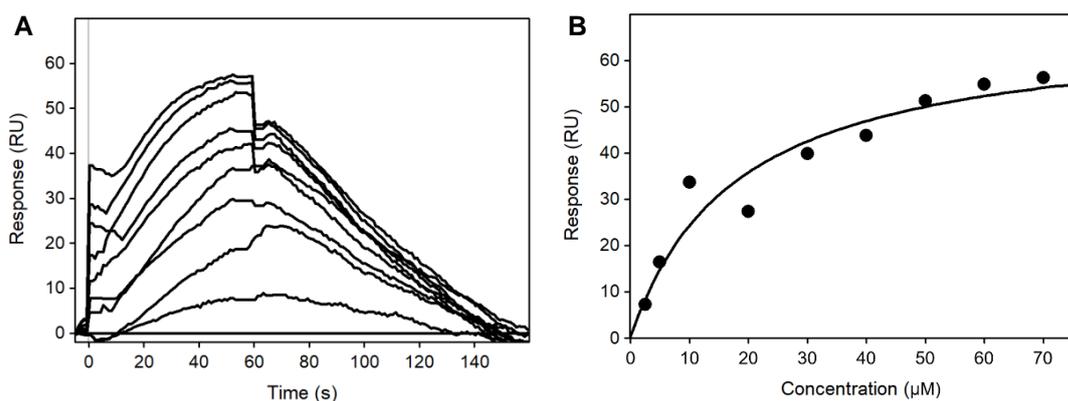
**Supplementary Figure 167** (A) Overlay of sensorgrams of **25** injected at concentrations 2.5–50  $\mu\text{M}$  over an immobilized  $\Delta\text{mtDXPS}$ ; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $20 \pm 4 \mu\text{M}$ ).



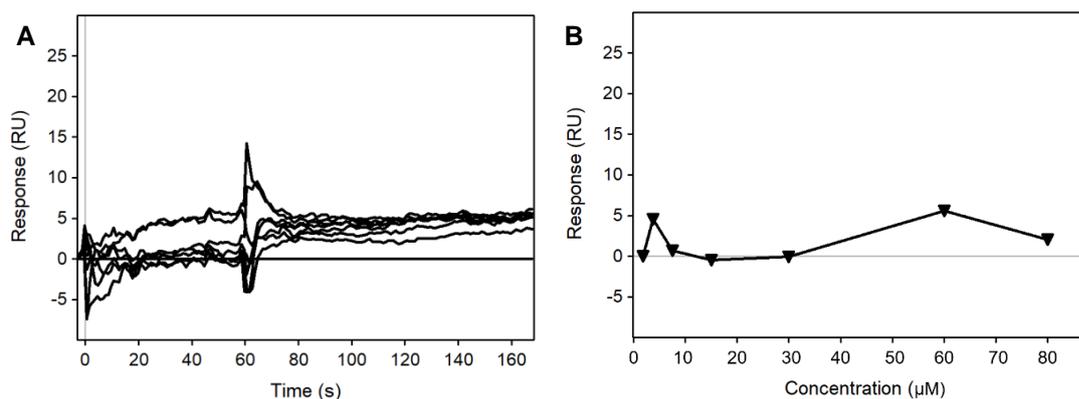
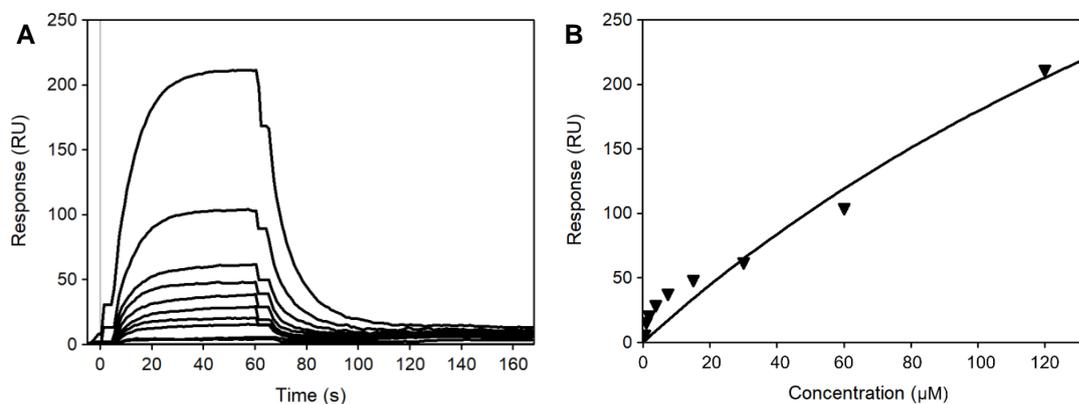
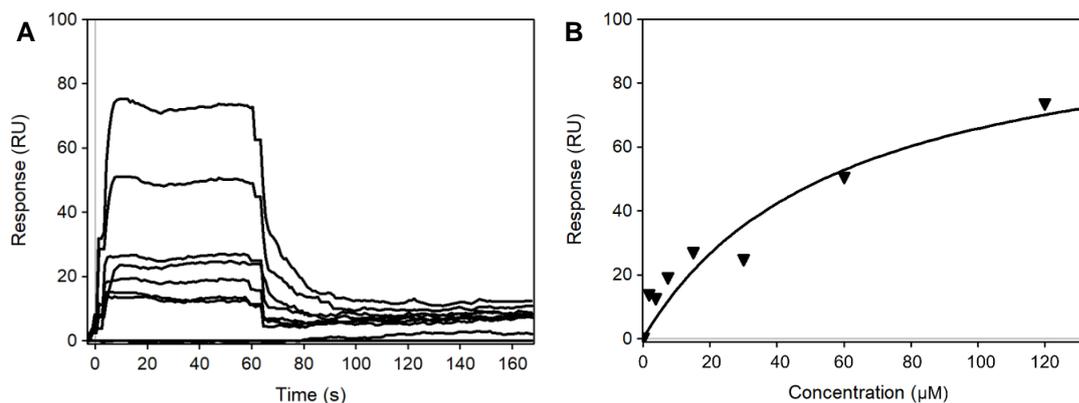
**Supplementary Figure 168** (A) Overlay of sensorgrams of **26** injected at concentrations 2.5–70 μM over an immobilized  $\Delta$ mtDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ : 110 ± 40 μM).

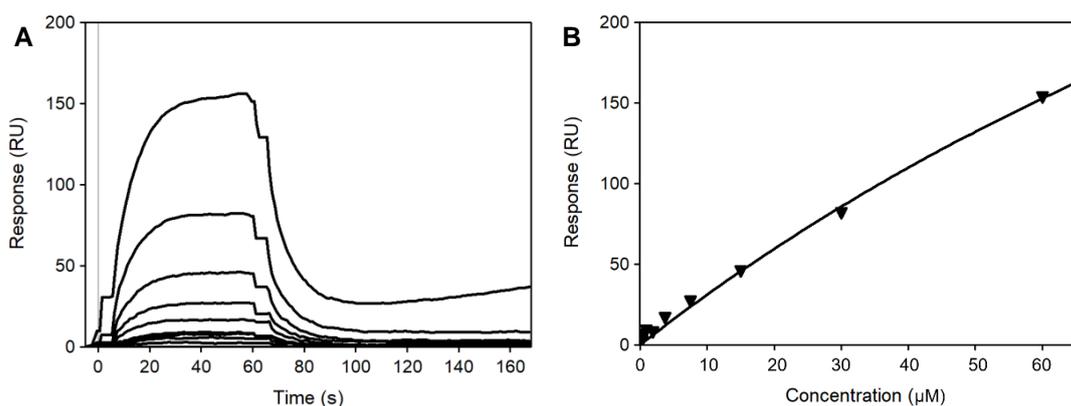


**Supplementary Figure 169** (A) Overlay of sensorgrams of **35** injected at concentrations 2.5–70 μM over an immobilized  $\Delta$ mtDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ : 60 ± 20 μM).

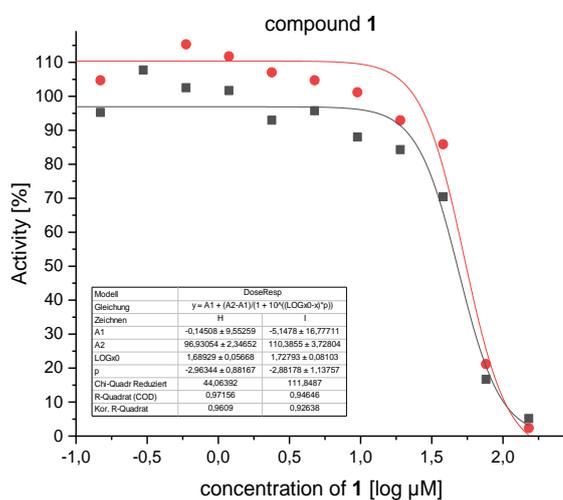


**Supplementary Figure 170** (A) Overlay of sensorgrams of **37** injected at concentrations 2.5–70 μM over an immobilized  $\Delta$ mtDXPS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ : 18 ± 3 μM).

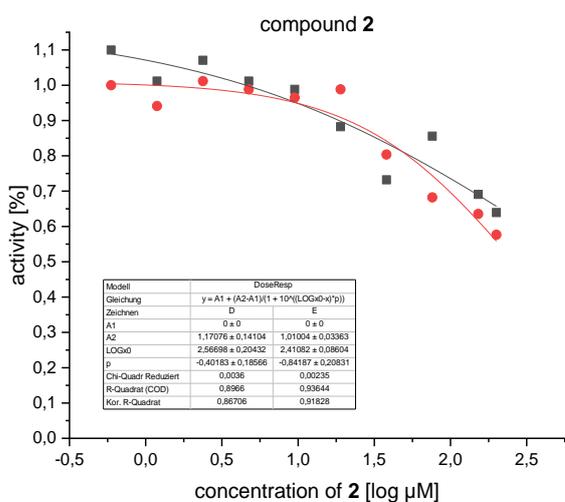




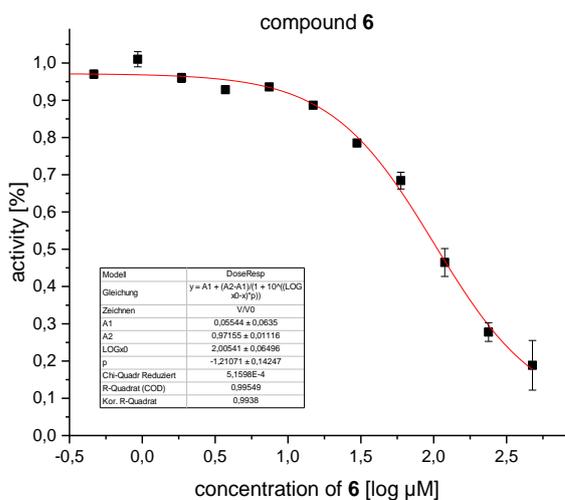
**Supplementary Figure 174** (A) Overlay of sensorgrams of **41** injected at concentrations 0.2–60  $\mu\text{M}$  over an immobilized dmtDXS; (B) Fitting of responses at equilibrium according to 1:1 Langmuir binding model ( $K_D$ :  $210 \pm 50 \mu\text{M}$ ).



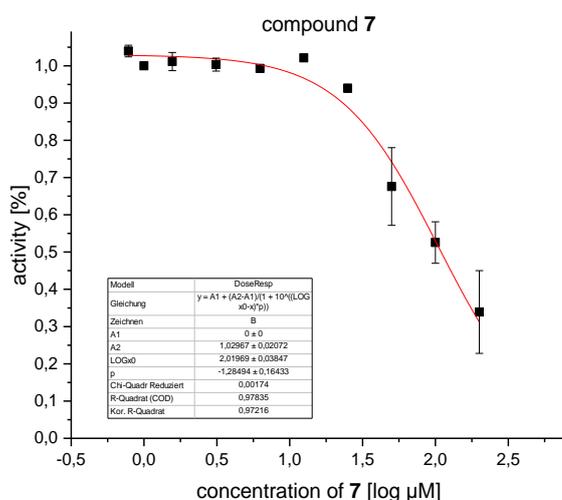
**Supplementary Figure 175** Enzymatic activity of the enzyme drDXPS at different concentrations of compound **1**. Individual  $IC_{50}$  values are  $53.4 \pm 9.9$  and  $48.8 \pm 6.3 \mu\text{M}$ ; mean  $IC_{50}$ :  $51.2 \pm 3.1 \mu\text{M}$ . Triton X-100 was used in the assay.



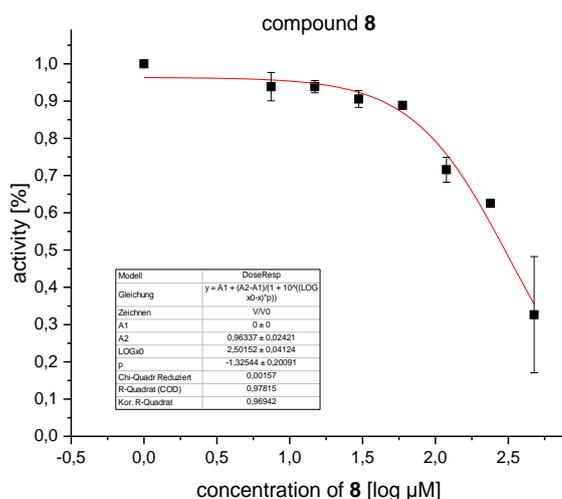
**Supplementary Figure 176** Enzymatic activity of the drDXPS at different concentrations of compound 2. Full inhibition could not be reached due to solubility limitations. Inhibition at 200  $\mu\text{M}$  is  $39.3 \pm 4.4\%$ , derived inhibition at 120  $\mu\text{M}$  is  $30.9 \pm 3.6\%$ . Triton X-100 was used in the assay.



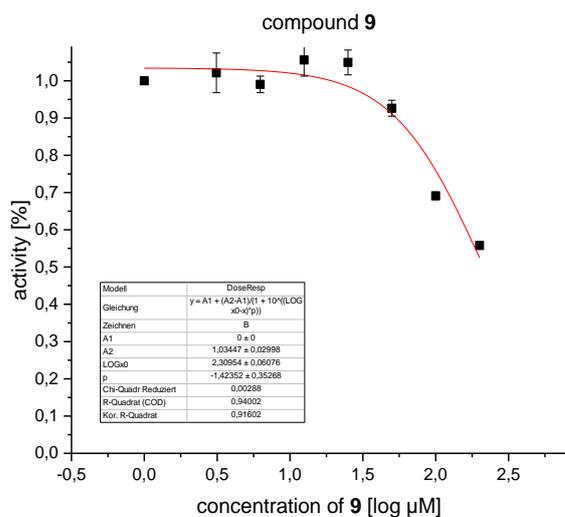
**Supplementary Figure 177** Enzymatic activity of the enzyme drDXPS at different concentrations of compound 6. Duplicates were averaged and fitted together, mean  $\text{IC}_{50}$ :  $101.3 \pm 15.1 \mu\text{M}$ . Triton X-100 was used in the assay.



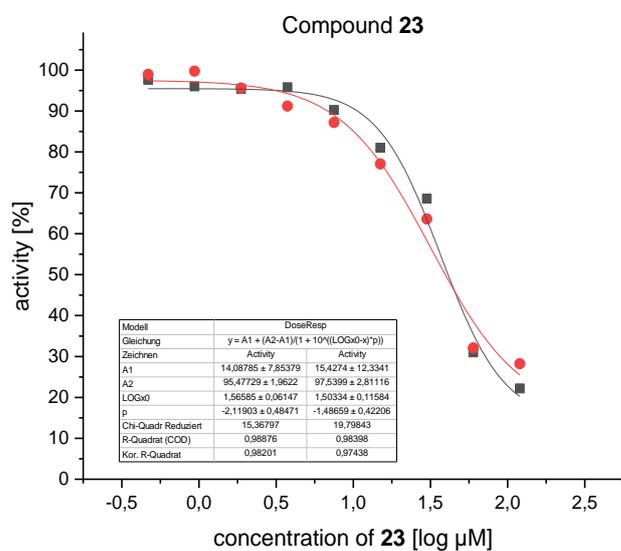
**Supplementary Figure 178** Enzymatic activity of the enzyme drDXPS at different concentrations of compound 7. Duplicates were averaged and fitted together, full inhibition could not be reached due to solubility limitations. Assuming full inhibition at high concentrations, an **IC<sub>50</sub> of 104.6 ± 9.3 μM** was estimated. Triton X-100 was used in the assay.



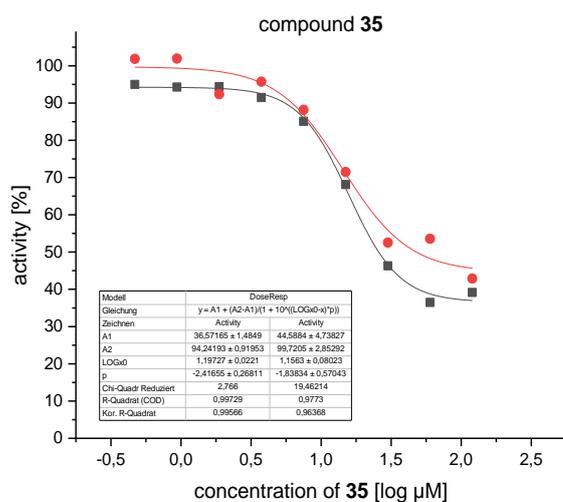
**Supplementary Figure 179** Enzymatic activity of the enzyme drDXPS at different concentrations of compound 8. Duplicates were averaged and fitted together, full inhibition could not be reached due to solubility limitations. Assuming full inhibition at high concentrations, an **IC<sub>50</sub> of 317.3 ± 30.1 μM** was estimated. Triton X-100 was used in the assay.



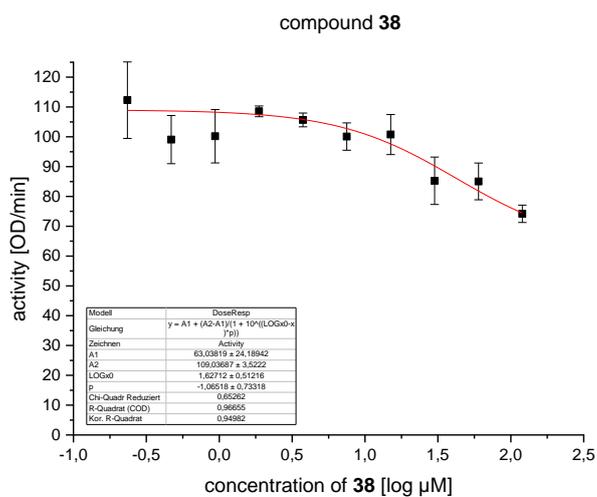
**Supplementary Figure 180** Enzymatic activity of the enzyme drDXPS at different concentrations of compound 9. Duplicates were averaged and fitted together, full inhibition could not be reached due to solubility limitations. Inhibition at 200 µM is 45.3%, derived inhibition at 120 µM is 30.1%. Triton X-100 was used in the assay.



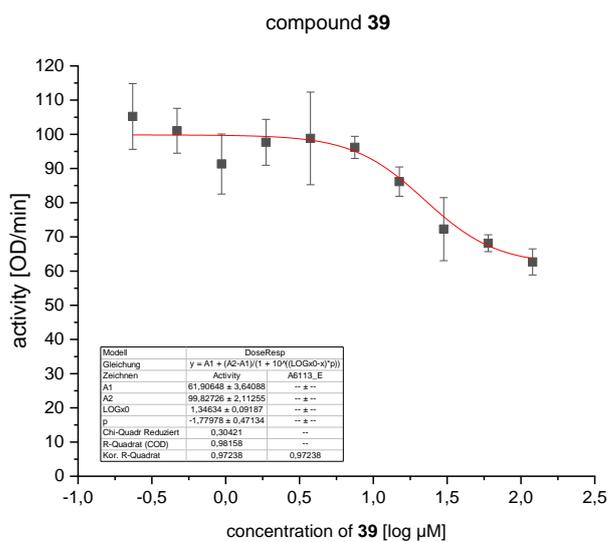
**Supplementary Figure 181** Enzymatic activity of the drDXPS at different concentrations of compound 23. Individual IC<sub>50</sub> values are 36.8 ± 5.2 and 31.8 ± 8.5 µM; mean IC<sub>50</sub>: **34.3 ± 3.5 µM**.



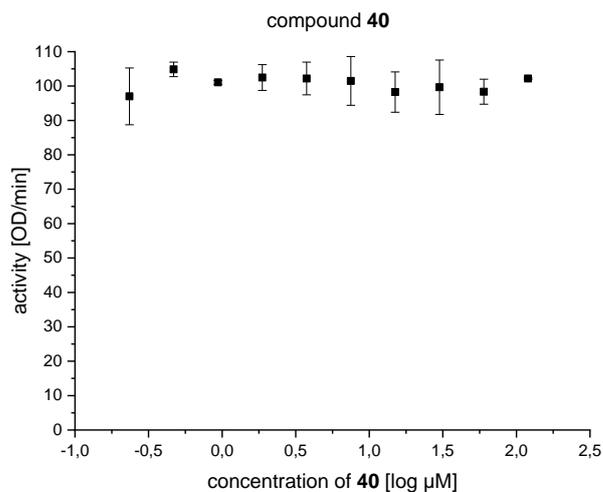
**Supplementary Figure 182** Enzymatic activity of the drDXPS at different concentrations of compound 35. Individual  $IC_{50}$  values are 15.75 and 14.33  $\mu\text{M}$ . The compound seems to be a partial inhibitor, as the residual activity of the enzyme is 40 %.



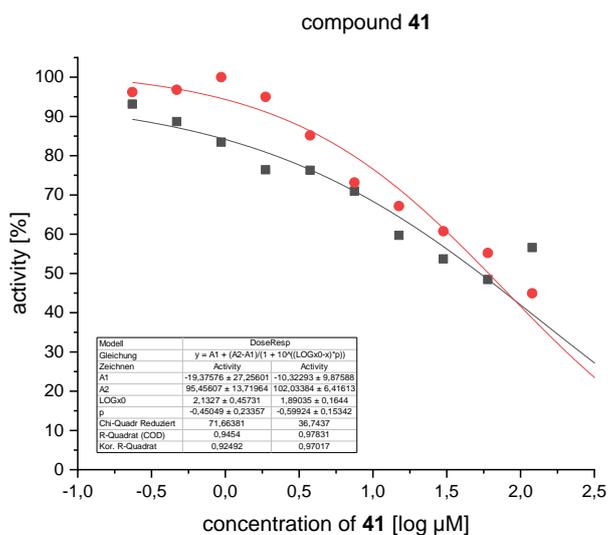
**Supplementary Figure 183** Enzymatic activity of the drDXPS at different concentrations of compound 38. Full inhibition could not be reached. Inhibition at 120  $\mu\text{M}$  is  $25.8 \pm 2.9\%$ .



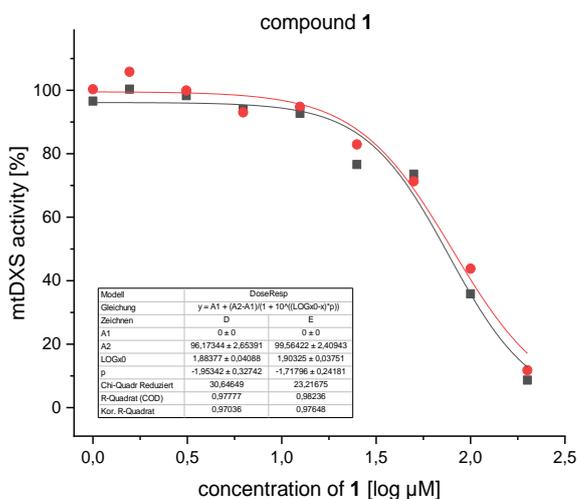
**Supplementary Figure 184** Enzymatic activity of the drDXPS at different concentrations of compound **39**. Full inhibition could not be reached. Inhibition at 120  $\mu\text{M}$  is  $37.3 \pm 3.8\%$



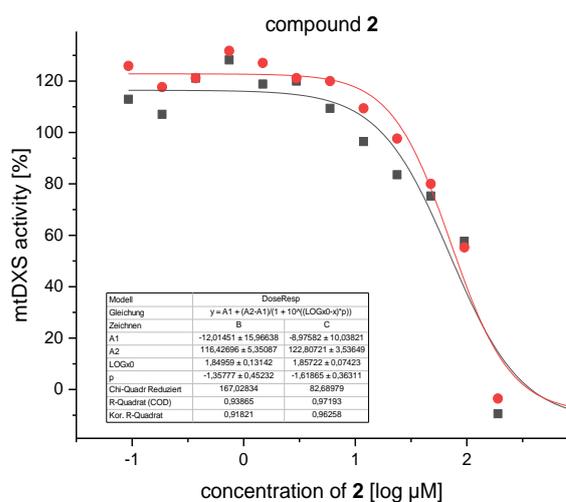
**Supplementary Figure 185** Enzymatic activity of the drDXPS at different concentrations of compound **40**. Full inhibition could not be reached. Inhibition at 120  $\mu\text{M}$  is 0 %



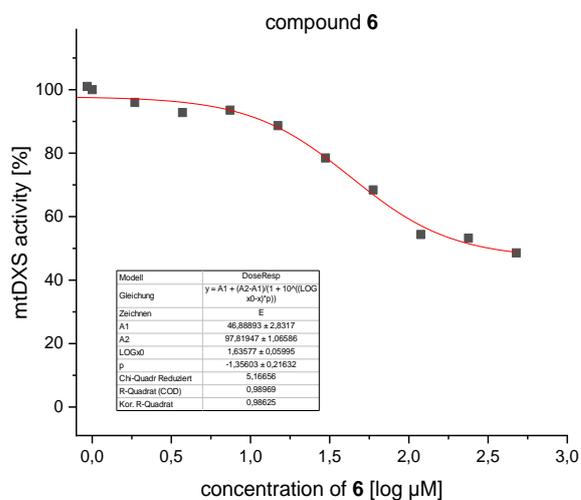
**Supplementary Figure 186** Enzymatic activity of the drDXPS at different concentrations of compound **41**. Full inhibition could not be reached. Inhibition at 60  $\mu\text{M}$  is  $48.2 \pm 3.4\%$  and at 120  $\mu\text{M}$  is  $49.2 \pm 5.8\%$ . Assuming full inhibition at high concentrations, an  $\text{IC}_{50}$  of  $99 \pm 2 \mu\text{M}$  was estimated.



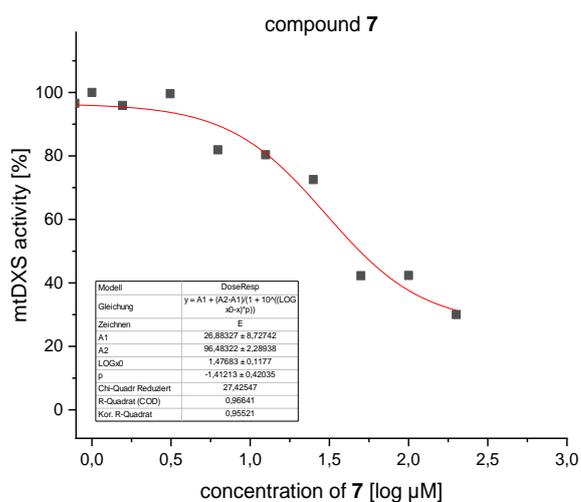
**Supplementary Figure 187** Enzymatic activity of the mtDXPS at different concentrations of compound **1**. Individual  $\text{IC}_{50}$  values are  $80.0 \pm 6.9 \mu\text{M}$  and  $76.5 \pm 7.2 \mu\text{M}$ ; mean  $\text{IC}_{50}$ :  $78.3 \pm 2.5 \mu\text{M}$ . Triton X-100 was used in the assay.



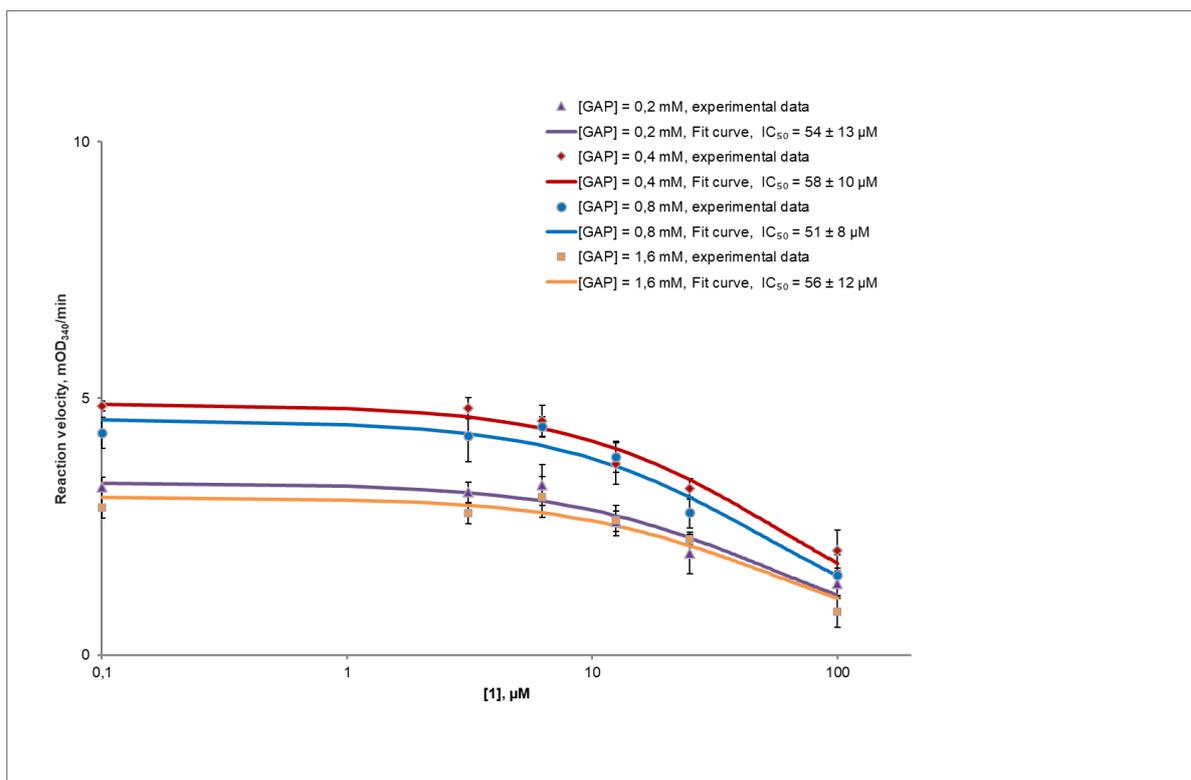
**Supplementary Figure 188** Enzymatic activity of the mtDXPS at different concentrations of compound 2. Individual  $IC_{50}$  values are  $70.7 \pm 21.4$  and  $72.0 \pm 12.3$   $\mu M$ ;  $IC_{50}$ :  $71.3 \pm 0.99$   $\mu M$ . Triton X-100 was used in the assay.



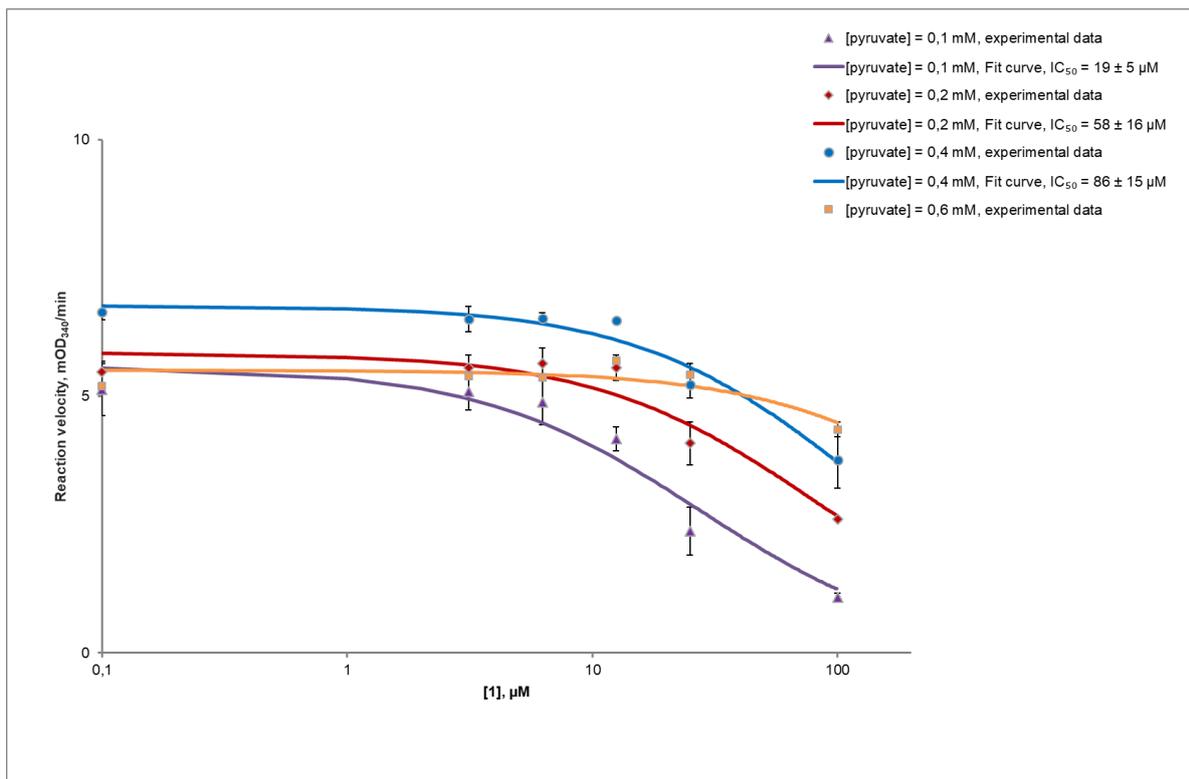
**Supplementary Figure 189** Enzymatic activity of the mtDXPS at different concentrations of compound 6.  $IC_{50}$ : 43.1  $\mu M$  at the residual activity of 50 %. Triton X-100 was used in the assay.



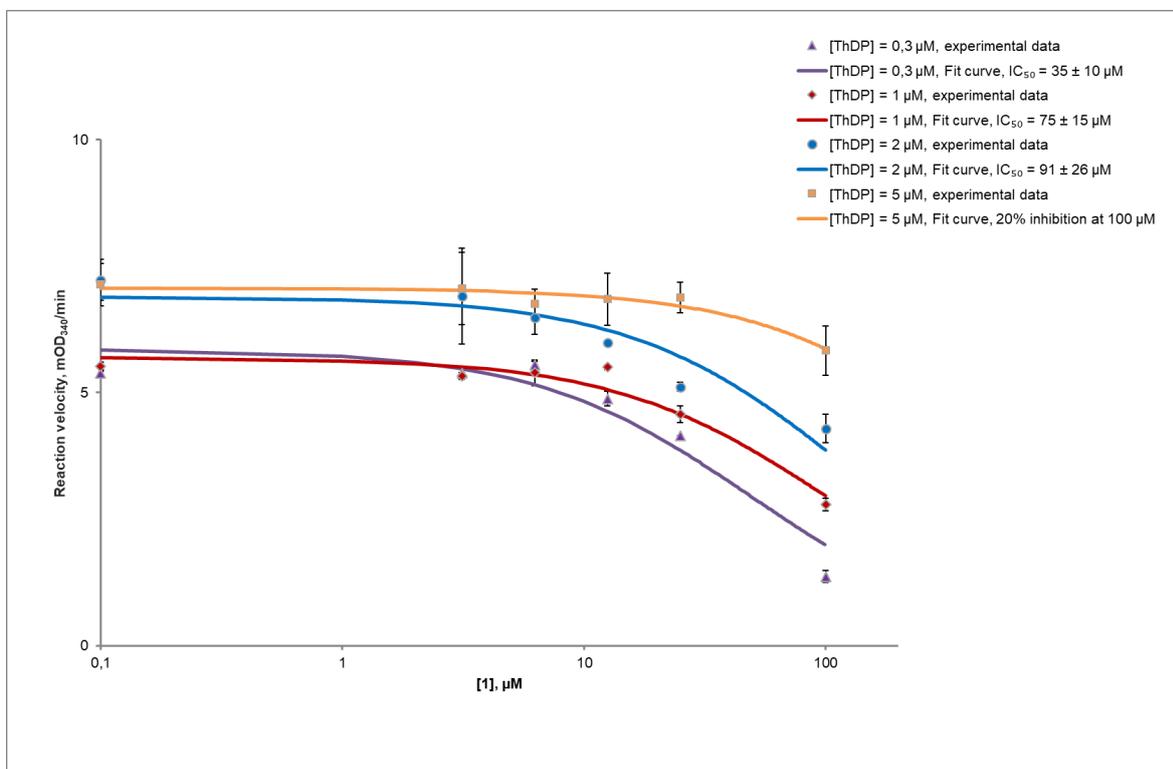
**Supplementary Figure 190** Enzymatic activity of the mtDXPS at different concentrations of compound 7.  $IC_{50}$ : 30  $\mu$ M at the residual activity of 30 %. Triton X-100 was used in the assay.



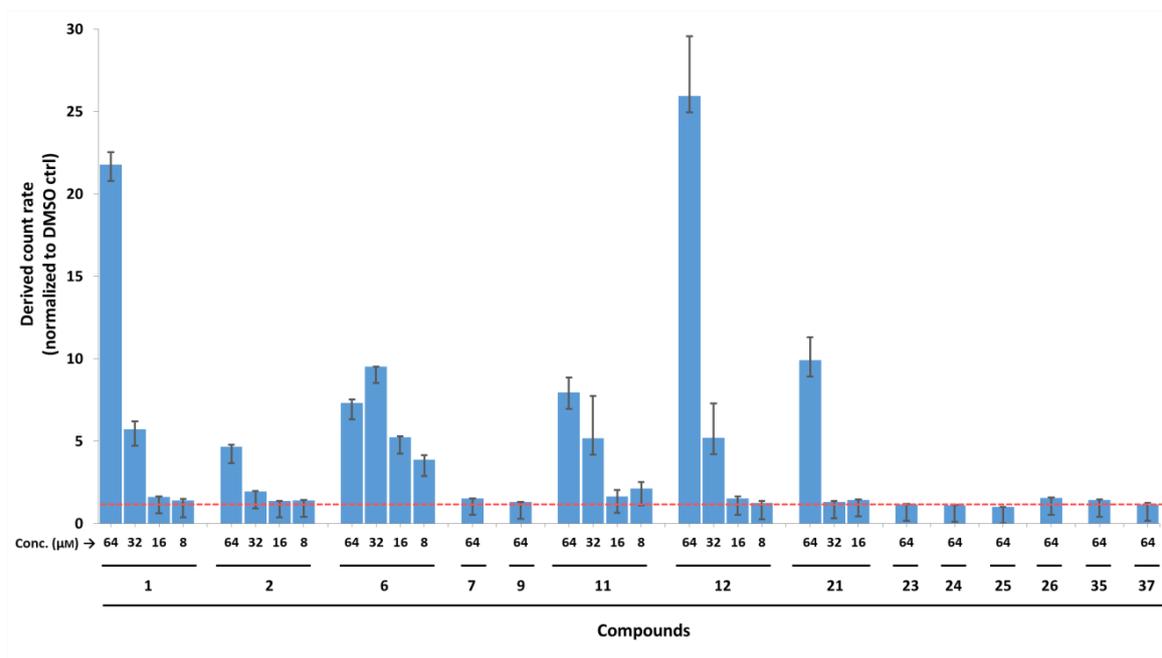
**Supplementary Figure 191** Inhibition of drDXPS by 1 at different concentrations of D/L-GAP.



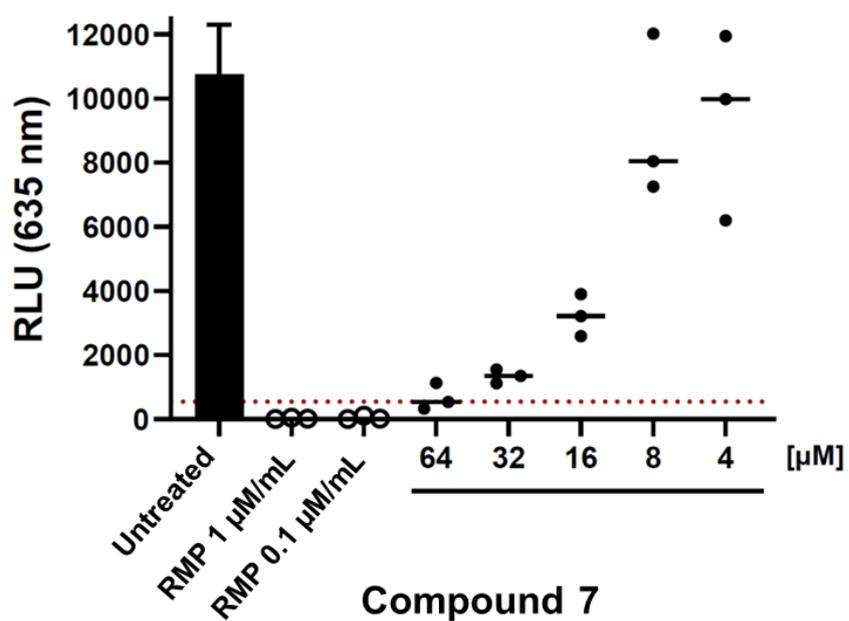
Supplementary Figure 192 Inhibition of drDXPS by **1** at different concentrations of pyruvate.



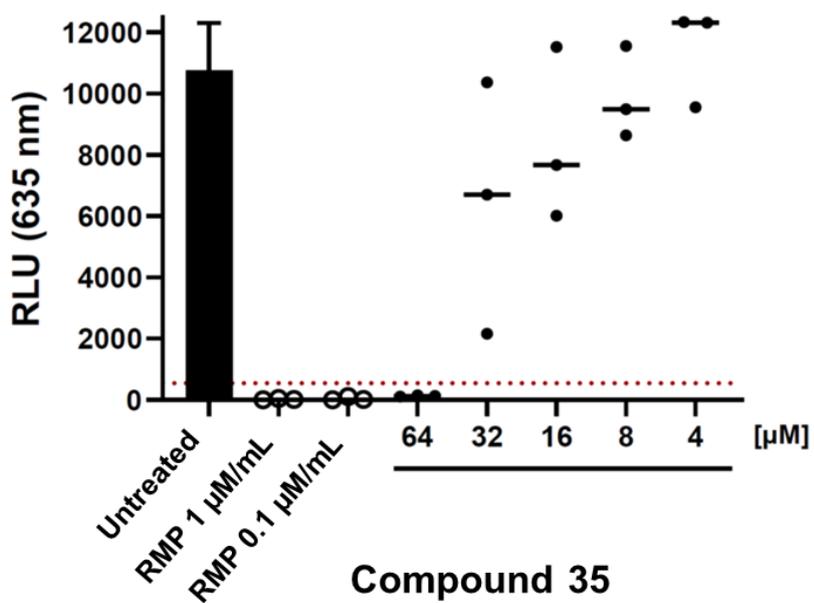
Supplementary Figure 193 Inhibition of drDXPS by **1** at different concentrations of ThDP.



**Supplementary Figure 194** Dynamic light scattering (DLS) measurements of DCC hits.



**Supplementary Figure 195** Growth inhibition of *M. tuberculosis* (H37Rv) bacteria at different concentrations of compound 7 for MIC analysis.



**Supplementary Figure 196** Growth inhibition of *M. tuberculosis* (H37Rv) bacteria at different concentrations of compound 35 for MIC analysis.



## SUPPLEMENTARY REFERENCES

- 1 P. Frei, L. Pang, M. Silbermann, D. Eriş, T. Mühlethaler, O. Schwardt and B. Ernst, *Chem. Eur. J.*, 2017, **23**, 11570.
- 2 P. Frei, R. Hevey, B. Ernst, *Chem. Eur. J.* 2019, **25**, 60.
- 3 S. Xiang, G. Usunow, G. Lange, M. Busch, L. Tong, *J. Biol. Chem.*, 2007, **282**, 2676.
- 4 LeadIT version 2.3.2; BioSolveIT GmbH, Sankt Augustin, Germany, [www.biosolveit.de/LeadIT](http://www.biosolveit.de/LeadIT), **2017**.
- 5 K. Stierand, M. Rarey, *ACS Med. Chem. Lett.*, 2010, **1**, 540.
- 6 SeeSAR version 10.1; BioSolveIT GmbH, Sankt Augustin, Germany, [www.biosolveit.de/SeeSAR](http://www.biosolveit.de/SeeSAR), **2020**.
- 7 N. N. Farshori, A. Rauf, M. A. Siddiqui, E. S. Al-Sheddi and M. M. Al-Oqail, *Arab. J. Chem*, 2017, **10**, S2853
- 8 T. Masini, J. Pilger, B. S. Kroezen, B. Illarionov, P. Lottmann, M. Fischer, C. Griesinger, A. K. H. Hirsch, *Chem. Sci.*, 2014, **5**, 3543
- 9 S. Hecht, J. Wungsintaweekul, F. Rohdich, K. Kis, T. Radykewicz, C. A. Schuhr, W. Eisenreich, G. Richter and A. Bacher, *J. Org. Chem.*, 2001, **66**, 7770.
- 10 W. A. M. Elgaher, M. Fruth, M. Groh, J. Hauptenthal, R. W. Hartmann, *RSC Adv.*, 2014, **4**, 2177
- 11 A. Zelmer, P. Carroll, N. Andreu, K. Hagens, J. Mahlo, N. Redinger, B. D. Robertson, S. Wiles, T. H. Ward, T. Parish, J. Ripoll, G. J. Bancroft and U. E. Schaible, *J Antimicrob. Chemother.*, 2012, **67**, 1948.
- 12 K. Kolbe, L. Möckl, V. Sohst, J. Brandenburg, R. Engel, S. Malm, C. Bräuchle, O. Holst, T. K. Lindhorst, N. Reiling, *ChemBioChem*, 2017, **18**, 1172.
- 13 A. Michelucci, T. Cordes, J. Ghelfi, A. Pailot, N. Reiling, O. Goldmann, T. Binz, A. Wegner, A. Tallam, A. Rausell, M. Buttini, C. L. Linster, E. Medina, R. Balling and K. Hiller, *Proc. Natl. Acad. Sci.*, 2013, **110**, 7820.
- 14 D. H. Schwarz, W. A. M. Elgaher, K. Hollemeyer, A. K. H. Hirsch and G. Wenz, *J. Mater. Chem. B*, 2019, **7**, 6148.
- 15 P. Kirsch, V. Jakob, W. A. M. Elgaher, C. Walt, K. Oberhausen, T. F. Schulz and M. Empting *ACS Chem. Biol.*, 2020, **15**, 388.
- 16 A. M. Hartman, V. R. Jumde, W. A. M. Elgaher, E. M. Te Poele, L. Dijkhuizen and A. K. H. Hirsch, *ChemMedChem*, 2021, **16**, 113.
- 17 A. Marcozzi, T. Masini, D. Zhu, D. Pesce, B. Illarionov, M. Fischer, A. Herrmann, A. K. H. Hirsch, *ChemBioChem*, 2018, **19**, 58.
- 18 P. Kuzmic, *Anal. Biochem.*, 1996, **237**, 260.



## 3.5 Supplementary Information for Chapter 2.5



### Supporting Information

#### **Design and Synthesis of Bioisosteres of Acylhydrazones as Stable Inhibitors of the Aspartic Protease Endothiapepsin**

Varsha R. Jumde<sup>+, [a, b]</sup> Milon Mondal<sup>+, [a]</sup> Robin M. Gierse,<sup>[a, b, c]</sup> M. Yagiz Unver,<sup>[a]</sup>  
Francesca Magari,<sup>[d]</sup> Roos C. W. van Lier,<sup>[a]</sup> Andreas Heine,<sup>[d]</sup> Gerhard Klebe,<sup>[d]</sup> and  
Anna K. H. Hirsch<sup>\*, [a, b, c]</sup>

cmdc\_201800446\_sm\_miscellaneous\_information.pdf

# Supporting Information

## Table of Contents

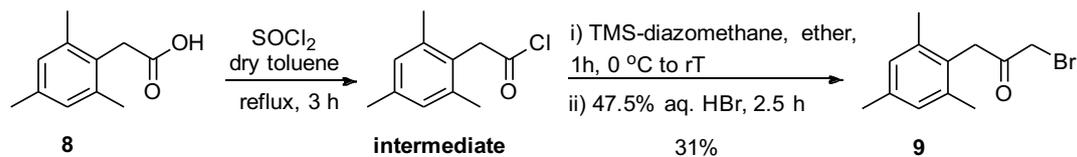
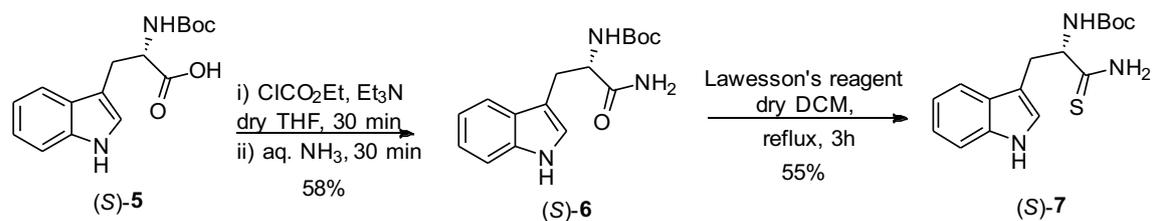
<b>General Experimental Details</b>	S3
<b>Schemes S1 – S4</b>	S4
<b>Experimental procedures</b>	S6
<b>Inhibition assay</b>	S10
<b>Modeling and docking</b>	S13
<b>Crystallization, data collection and processing</b>	S14
<b>Structural analysis of torsion angles</b>	S16
<b>Structural analysis of water hydrogen bond network</b>	S17
<b>Heterocyclic biosiosters</b>	S17
<b>NMR spectra</b>	S18
<b>References and notes</b>	S22

**General Experimental Details**

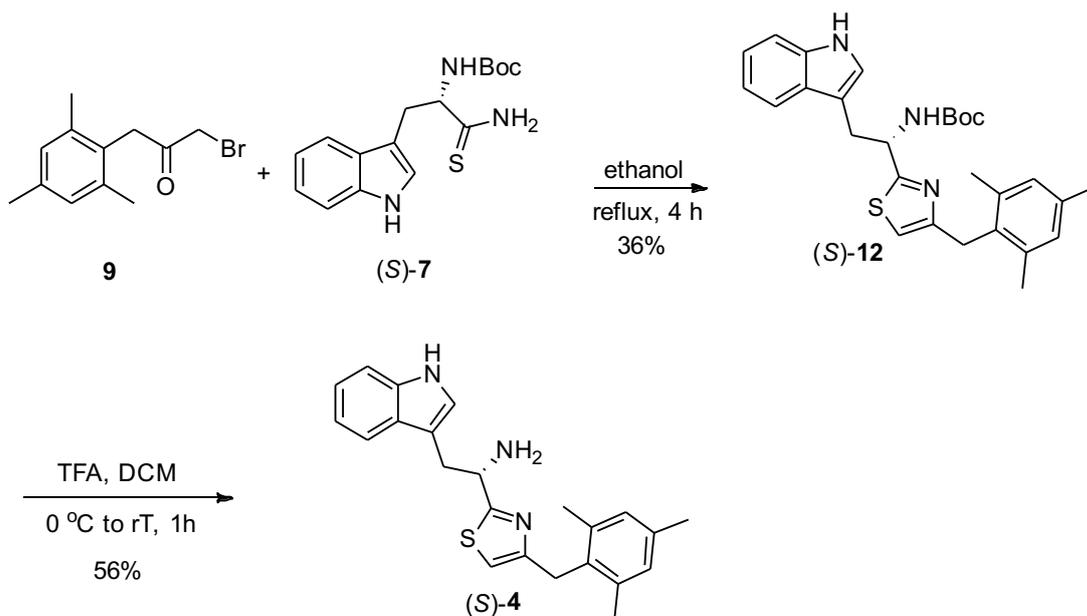
Starting materials and reagents were purchased from Aldrich, Acros or TCI. Yields refer to analytically pure compounds and have not been optimized. All solvents were reagent-grade and if necessary, dried and distilled prior to use. Reactions were carried out under nitrogen gas atmosphere. Column chromatography was performed on silica gel (Silicycle® SiliaSep™ 40-63  $\mu\text{M}$  60  $\text{\AA}$ ). TLC was performed with silica gel 60/Kieselguhr F254. Optical rotations were measured in MeOH on a Schmidt & Haensch polarimeter (Polartronic MH8) with a 10 cm cell (*c* given in g/100 mL).  $^1\text{H}$ -,  $^{13}\text{C}$ -NMR spectra were recorded at 400 MHz on a Varian AMX400 spectrometer (400 MHz for  $^1\text{H}$ , 101 MHz for  $^{13}\text{C}$ ) at 25 °C.

Chemical shifts ( $\delta$ ) are reported relative to the residual solvent peak. Splitting patterns are indicated as (s) singlet, (d) doublet, (t) triplet, (q) quartet, (m) multiplet, (br) broad. High-resolution mass spectra were recorded with a FTMS orbitrap (Thermo Fisher Scientific) mass spectrometer.

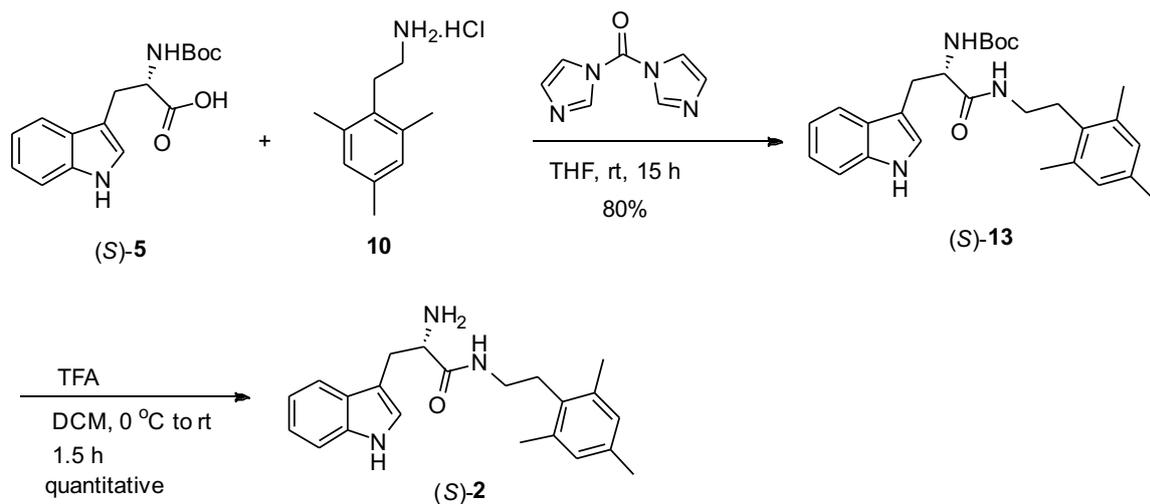
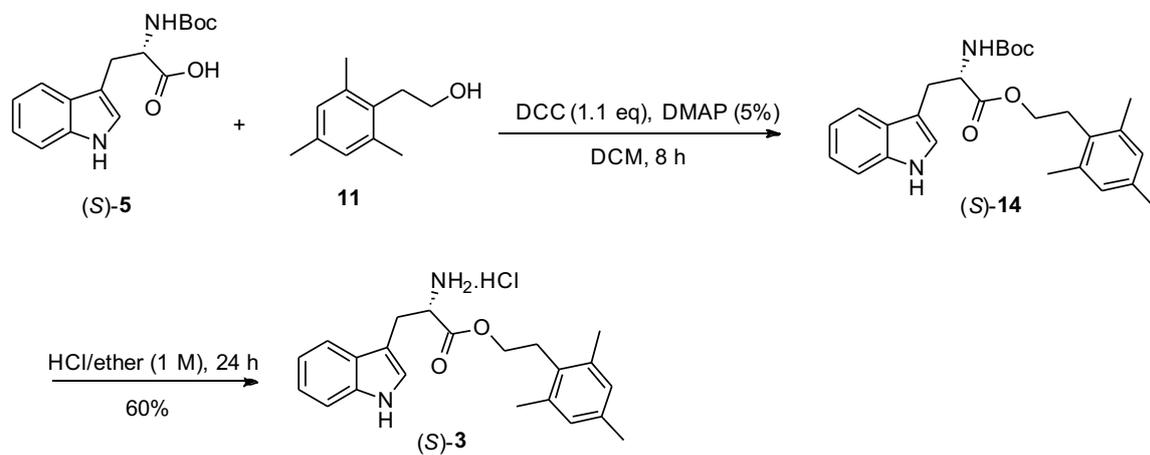
## Schemes



Scheme 1: Synthesis of building blocks.



Scheme 2: Synthesis of thiazole bioisostere (S)-4.

**Scheme 3:** Synthesis of amide bisostere (S)-2.**Scheme 4:** Synthesis of ester bisostere (S)-3.

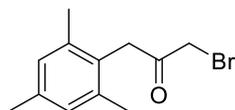
## Experimental procedures

### Synthesis of the building blocks:

Compounds (S)-**6** and (S)-**7** were synthesized according to procedures reported in the literature and their spectroscopic data are in agreement.<sup>1,2</sup>

#### 1-Bromo-3-mesitylpropan-2-one (**9**)

(adapted from<sup>1-3</sup>)



In a dry 100-mL three-necked flask, mesitylacetic acid (**8**) (1 eq., 1.5 g, 8.4 mmol) was dissolved in dry toluene (10 mL). Thionyl chloride (13.8 eq., 8.5 mL, 116 mmol) was added to this reaction mixture, and was refluxed for 3 h. The colorless solution was concentrated *in vacuo* to obtain a yellow crude oil, which was dissolved in ether (8 mL). This solution was added dropwise to trimethylsilyl-diazomethane in ether (2.6 eq., 10.8 mL, 2 M, 21.7 mmol) at 0–5 °C. The yellow solution was left to stir at room temperature for 1 h followed by the dropwise addition of aqueous HBr (10.8 mL, 47.5%), and the orange emulsion was refluxed for 2.5 h. It was cooled down to room temperature, and the ether layer was separated and washed with water (1 x 50 mL). The aqueous layer was extracted with ether (3 x 35 mL). The combined organic layers were washed with saturated aqueous NaHCO<sub>3</sub> solution and water sequentially, dried over MgSO<sub>4</sub>, filtered and concentrated *in vacuo*. The white solid was purified by column chromatography (SiO<sub>2</sub>, DCM/pentane, 2:3, stained in phosphomolybdic acid) to yield **9** (2.56 mmol, 655 mg, 31%) as a white solid. M.p. 52–55 °C

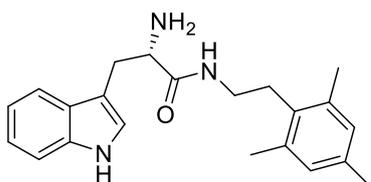
**<sup>1</sup>H-NMR (400 MHz, CDCl<sub>3</sub>-d)**: δ 6.88 (d, *J* = 6.6 Hz, 2H), 3.98 (d, *J* = 1.8 Hz, 1H), 3.91 (d, *J* = 1.9 Hz, 1H), 3.67 (dd, *J* = 5.7, 1.9 Hz, 2H), 2.49 – 1.89 (m, 9H).

**<sup>13</sup>C-NMR (101 MHz, CDCl<sub>3</sub>-d)**: (signals assigned corresponds to 2 products coexist due to the equilibrium of keto-enol form) δ 199.8, 172.1, 137.2, 137.1, 136.9, 136.7, 129.3, 129.0, 128.6, 127.9, 52.1, 41.2, 35.0, 33.7, 21.0, 21.0, 20.4, 20.3.

**GC-MS (250 °C)**: 20.08 min, *m/z* = 256 (30, M), 133 (100, M-COCH<sub>2</sub>Br), 117 (20, M-H<sub>2</sub>COCH<sub>2</sub>Br), 105 (20, Ph(CH<sub>3</sub>)<sub>2</sub>), 91 (25, PhCH<sub>3</sub>), 77 (10, Br), 41 (10, CH<sub>3</sub>CO).

### Synthesis of bioisosteres

#### (+)-**(S)**-2-Amino-3-(1*H*-indol-3-yl)-*N*-(2,4,6-trimethylphenethyl)propanamide (**2**)



To an ice cold solution of *N*-Boc-L-tryptophan (**5**) (1 eq., 610 mg, 2 mmol) in dry THF (5 mL), 1,1'-carbonyl bis-1*H*-imidazole (1 eq., 324 mg, 2 mmol) was added. After stirring the solution at room

temperature for 2 h, under ice cooling, mesitylethylamine<sup>4</sup> (**10**) (1 eq., 400 mg, 2 mmol) was added, and the reaction mixture was left to stir at room temperature overnight. This solution was concentrated *in vacuo*, the residue obtained was dissolved in EtOAc (50 mL), washed with water, 0.1 M HCl, saturated aqueous NaHCO<sub>3</sub> solution and with a saturated aqueous NaCl solution sequentially. The organic layer was dried over anhydrous MgSO<sub>4</sub>, filtered, and concentrated *in vacuo* to obtain (*S*)-**13**. It was monitored by TLC (DCM/ MeOH 98:2, stained with Ninhydrin), showing complete conversion. The crude was used in the next step without any further purification.

Trifluoroacetic acid (16mmol, 21.5 mmol, 1.65 mL) was added to a solution of protected amide compound (*S*)-**13** from the previous step (1 eq., 1.334 mmol, 600 mg) in dry DCM (6 mL) with ice cooling and left to stir for 5 minutes 1 h at room temperature. After concentrating the reaction mixture *in vacuo*, chloroform (25 mL) was added to the residue. The organic layer was washed with saturated aqueous NaHCO<sub>3</sub> solution (3 x 15 mL) and saturated aqueous NaCl solution (15 mL), dried over anhydrous MgSO<sub>4</sub>, filtered and concentrated *in vacuo* to obtain the final product **2** as white solid, (400 mg, quantitative). M.p. 56–60 °C

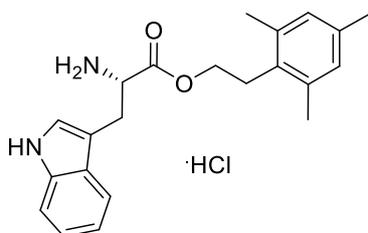
<sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD-*d*<sub>4</sub>) δ 7.65 (dt, *J* = 8, 1.0 Hz, 1H), 7.37 (dt, *J* = 8, 0.9 Hz, 1H), 7.15 – 6.93 (m, 3H), 6.78 (s, 2H), 3.61 (t, *J* = 6.7 Hz, 1H), 3.27 – 2.93 (m, 4H), 2.60 (t, *J* = 8.3 Hz, 2H), 2.25 (s, 6H), 2.21 (s, 3H).

<sup>13</sup>C NMR (101 MHz, CD<sub>3</sub>OD-*d*<sub>4</sub>) δ 178.6, 139.7, 138.9, 138.0, 134.9, 131.4, 130.4, 126.2, 124.0, 121.4, 121.1, 113.9, 112.8, 58.6, 40.9, 33.9, 31.7, 22.5, 21.5.

$$[\alpha]_D^{20} = +0.387 \text{ (} c = 0.39 \text{ in MeOH)}$$

HRMS (ESI) calcd for C<sub>22</sub>H<sub>28</sub>N<sub>3</sub>O [*M*+H]<sup>+</sup>: 350.2226, found 350.2232

### (–) 2,4,6-Trimethylphenethyl L-tryptophanate (**3**)



*N*-Boc- L-Tryptophan (**5**) (1eq., 365.4 mg, 1.2 mmol) was dissolved in anhydrous DCM. 2-Mesitylethanol (**11**) (1 eq., 197.5 mg, 1.2 mmol,) was added to the reaction followed by a catalytic amount of DMAP (5 mol%, 7.33 mg, 0.06 mmol). The reaction was cooled to 0 °C, and a solution of *N,N'*-dicyclohexylcarbodiimide (1.1 eq., 272.3 mg, 1.32 mmol) in anhydrous DCM (10 mL) was added dropwise over 10 min. The reaction was allowed to warm to room temperature and a precipitate was observed after 20 min. The reaction was stirred for further 6 h. The reaction was filtered to remove the *N,N'*-Dicyclohexylcarbodiimide urea by-product and the solvent removed *in-vacuo*. The residue was placed onto a column of flash silica gel where the product was eluted with (DCM / EtOAc 1:0 to 9:1). The solvent was removed under reduced pressure to afford the product (*S*)-**14**, which was used in the next step without further purification.

The crude product was dissolved in DCM (3 mL), and HCl/diethyl ether (1 M, 15 mL) was added. After stirring at room temperature for 24 h, the resulting white precipitate was collected and washed with ether, to afford the final product (**S**-**3**) as a white solid (279 mg, 60%). M.p. >170 °C (degradation).

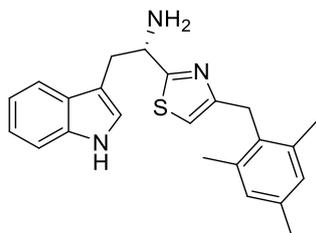
**<sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD-*d*<sub>4</sub>)** δ 7.47 (d, *J* = 8.0 Hz, 1H), 7.40 (d, *J* = 8.0 Hz, 1H), 7.17 – 7.11 (m, 2H), 7.05 (t, *J* = 7.5 Hz, 1H), 6.82 (s, 2H), 4.32 (t, *J* = 6.5 Hz, 1H), 4.20 (t, *J* = 7.7 Hz, 2H), 3.45 – 3.33 (m, 2H), 2.97 – 2.82 (m, 2H), 2.28 (s, 6H), 2.20 (s, 3H).

**<sup>13</sup>C NMR (101 MHz, CD<sub>3</sub>OD-*d*<sub>4</sub>)** δ 170.5, 138.4, 137.7, 137.1, 131.3, 130.0, 128.3, 125.8, 122.9, 120.3, 118.8, 112.7, 107.5, 65.9, 54.7, 28.9, 27.6, 20.9, 20.1.

$[\alpha]_D^{19} = -21$  (*c* = 0.02 in MeOH)

**HRMS (ESI)** calcd for C<sub>22</sub>H<sub>27</sub>N<sub>2</sub>O<sub>2</sub> [*M*+*H*]<sup>+</sup>: 351.2067, found: 351.2070.

**(+)** (**S**)-**2**-(**1H**-Indol-**3**-yl)-**1**-(**4**-mesitylthiazol-**2**-yl)ethan-**1**-amine (**4**)  
(adapted from<sup>1, 2, 5</sup>)



In a 25-mL round-bottomed flask was dissolved 1-bromo-3-mesitylpropan-2-one (**9**) (1 eq., 51.4 mg, 0.201 mmol) and *N*-Boc-L-tryptophan thioamide (**7**) (1.5 eq., 96.3 mg, 0.302 mmol) in absolute ethanol (6 mL). The yellow reaction mixture was heated to reflux for 4 h, and conversion was monitored by TLC (EtOAc/pentane, 3:7). The reaction mixture was concentrated *in vacuo*, and the crude material was dissolved in a mixture of water and DCM. The organic layer was separated, and the aqueous layer was extracted with DCM (3 x 15 mL). The combined organic layers were collected, dried over MgSO<sub>4</sub>, filtered and concentrated *in vacuo* to afford the crude, which was purified by column chromatography (SiO<sub>2</sub>, EtOAc/pentane, 3:7) to yield (**S**)-**12** (0.072 mmol, 34.1 mg, 36%) as a white solid, which was used in the next step without further purification.

In a 25-mL round-bottomed flask, was dissolved (**S**)-**12** (1 eq., 26.1 mg, 0.055 mmol) in dry DCM (1 mL). The reaction mixture was cooled to 0 °C, and trifluoroacetic acid (TFA) (16 eq., 0.067 mL, 0.878 mmol) was added. The solution was stirred in an ice bath for 5 min and at room temperature for 1 h. The reaction mixture was concentrated *in vacuo*, and the crude was dissolved in chloroform. The organic layer was washed with saturated aqueous NaHCO<sub>3</sub> solution and saturated aqueous NaCl solution, dried over MgSO<sub>4</sub>, filtered and concentrated *in vacuo*. The crude was purified by column chromatography (SiO<sub>2</sub>, EtOAc/pentane, 1:1 → 1:0) to yield (**S**)-**4** (0.031 mmol, 11.5 mg, 56%) as a yellow-orange semi-solid.

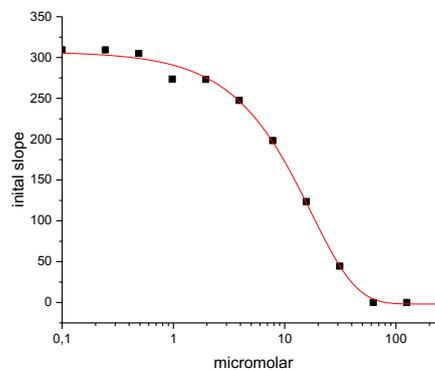
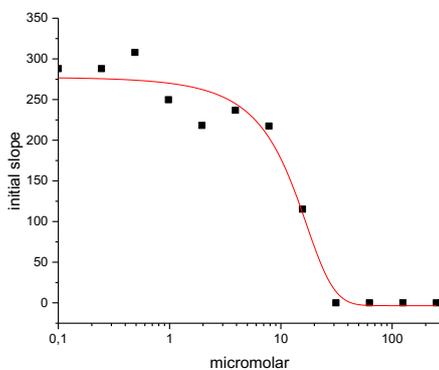
**<sup>1</sup>H-NMR (400 MHz, CDCl<sub>3</sub>-*d*)**: δ 8.12 (br s, -NH, 1H), 7.60 (d, *J* = 8.0 Hz, 1H), 7.35 (d, *J* = 8.0 Hz, 1H), 7.24 – 7.15 (m, 1H), 7.14 – 7.01 (m, 2H), 6.90 (s, 2H), 6.32 (s, 1H), 4.61 (dd, *J* = 8.8, 4.4 Hz, 1H), 4.13 (s, 2H),

3.50 (dd,  $J = 14.3, 4.4$  Hz, 1H), 3.13 (dd,  $J = 14.3, 8.8$  Hz, 1H), 2.29 (s, 3H), 2.27 (s, 6H), 2.07 (br s, -NH, 2H).

**$^{13}\text{C-NMR}$  (101 MHz,  $\text{CD}_3\text{OD-}d_4$ ):**  $\delta$  178.1, 156.4, 138.1, 137.7, 137.1, 133.4, 129.9, 128.8, 124.8, 122.4, 119.8, 119.2, 114.0, 112.3, 111.1, 55.6, 35.7, 32.2, 21.0, 20.1.

$$[\alpha]_D^{20} = +0.073 \text{ (} c = 0.15 \text{ in MeOH)}$$

**HRMS** (ESI): calcd for  $\text{C}_{23}\text{H}_{26}\text{N}_3\text{S}$   $[M+H]^+$ : 376.1842, found 376.1837; calcd for  $\text{C}_{23}\text{H}_{25}\text{N}_3\text{SNa}$   $[M+Na]^+$ : 398.1661, found 398.1656.

**Inhibition Assay:****Amide bioisostere (S)-2**

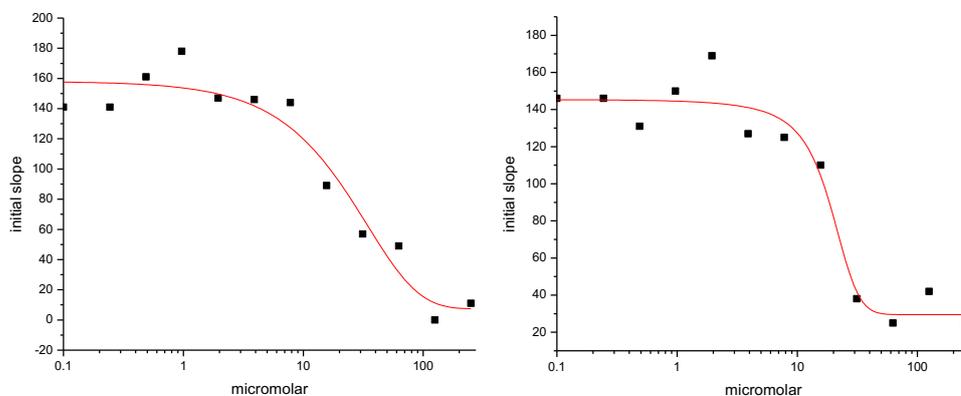
Model	DoseResp
Equation	$y = A1 + (A2-A1)/(1 + 10^{((LOGx0-x)*p)})$
Plot	Slope
A1	$-3,4032 \pm 13,06186$
A2	$336,63103 \pm 84,03253$
LOGx0	$10,85956 \pm 4,34377$
p	$-0,06219 \pm 0,02827$
Reduced Chi-Sqr	541,70386
R-Square(COD)	0,97589
Adj. R-Square	0,96685
IC50	13,47

Model	DoseResp
Equation	$y = A1 + (A2-A1)/(1 + 10^{((LOGx0-x)*p)})$
Plot	slope
A1	$-1,84765 \pm 4,70921$
A2	$1886,22955 \pm 2845,35958$
LOGx0	$-24,92074 \pm 31,97799$
p	$-0,02843 \pm 0,0055$
Reduced Chi-Sqr	53,59639
R-Square(COD)	0,99774
Adj. R-Square	0,99689
IC 50	12,34

**IC<sub>50</sub> = 12.9 ± 0.79 μM**

**Figure S1:** IC<sub>50</sub> inhibition curves of amide bioisostere (S)-2 (IC<sub>50</sub> = 12.90 ± 0.7 μM) measured in duplicate. The errors are given in standard deviations (SD), 11 different concentrations of inhibitor were used starting at 250 μM. In the duplicate measurement one well was excluded which contained air bubbles.

## Ester bioisostere (S)-3



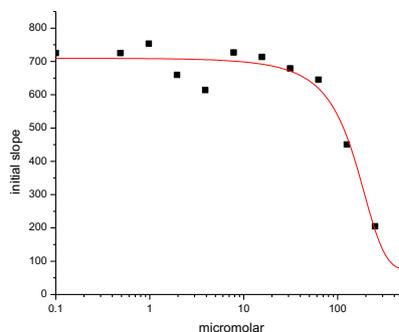
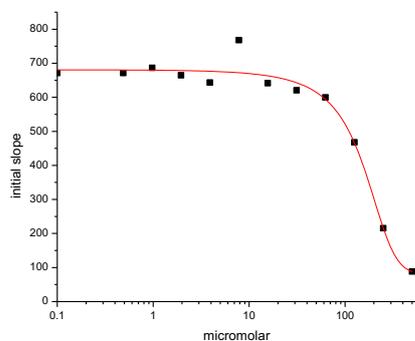
Model	DoseResp		
Equation	$y = A1 + (A2-A1)/(1 + 10^{((LOGx0-x)*p)})$		
Reduced Chi-Sqr	274.22045		
Adj. R-Square	0.92745		
		Value	Standard Error
	A1	7.33209	14.56058
	A2	39996.54923	1.53299E7
	LOGx0	-190.45171	13331.00848
	p	-0.01272	0.01337
	span	39989.21714	
	IC50	32.8	

Model	DoseResp		
Equation	$y = A1 + (A2-A1)/(1 + 10^{((LOGx0-x)*p)})$		
Reduced Chi-Sqr	174.7847		
Adj. R-Square	0.94071		
		Value	Standard Error
	A1	29.49153	7.60674
	A2	150.2924	12.27535
	LOGx0	18.51362	3.05574
	p	-0.07414	0.03707
	span	120.80087	
	IC50	24.6	

**IC<sub>50</sub> = 28.7 ± 4.1 μM**

**Figure S2:** IC<sub>50</sub> inhibition curves of ester bioisostere (S)-3 (IC<sub>50</sub> = 28.7 ± 4.1 μM) measured in duplicate. The errors are given in standard deviations (SD), 11 different concentrations of inhibitor were used starting at 250 μM.

## Thiazole bioisostere (S)-4



Model	DoseResp		
Equation	$y = A1 + (A2 - A1) / (1 + 10^{((LOGx0 - x) * p)})$		
Reduced Chi-Sqr	1455.65534		
Adj. R-Square	0.96546		
		Value	Standard Error
	A1	81.83853	43.56349
	A2	772.05744	91.09158
	LOGx0	142.4383	27.55815
	p	-0.00573	0.00191
	span	690.21891	
	IC50	205.1	

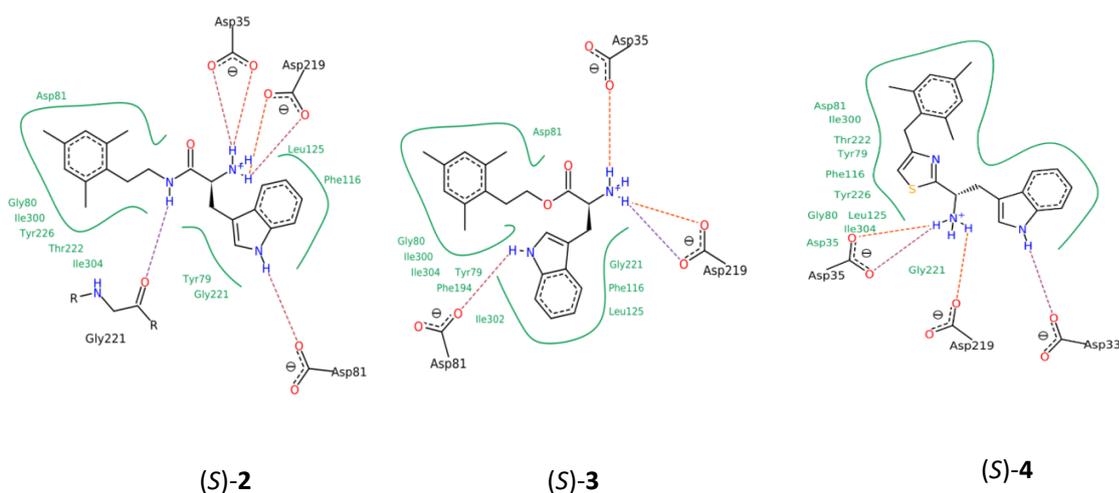
Model	DoseResp		
Equation	$y = A1 + (A2 - A1) / (1 + 10^{((LOGx0 - x) * p)})$		
Reduced Chi-Sqr	1977.14175		
Adj. R-Square	0.96051		
		Value	Standard Error
	A1	71.0996	48.06852
	A2	788.16021	85.00661
	LOGx0	142.27601	24.75058
	p	-0.00639	0.00218
	span	717.06061	
	IC50	182.3	

**IC<sub>50</sub> = 193.7 ± 11.4 μM**

**Figure S3:** IC<sub>50</sub> inhibition curves of thiazole bioisostere (S)-4 (IC<sub>50</sub> = 193.7 ± 11.4 μM) measured in duplicate. The errors are given in standard deviations (SD), 11 different concentrations of inhibitor were used starting at 500 μM.

**Modeling and Docking:**

The X-ray crystal structure of the complex of acylhydrazone (*S*)-**1** with endothiapepsin (PDB code: 4KUP) was used for our modeling.<sup>6</sup> Several ligands were designed in order to mimic the acylhydrazone interactions with the protein whilst finding replacements of the acylhydrazone linker, using the computer program MOLOC.<sup>7</sup> The energy of the system was minimized using the MAB force field as implemented in MOLOC, whilst keeping the protein coordinates fixed. Several acylhydrazone-mimetics were designed and subsequent energy minimization (MAB force field) was done using MOLOC. All types of interactions (hydrogen bonds and lipophilic interactions) between designed acylhydrazones and protein were measured in MOLOC. All the designed bioisosteres of acylhydrazones were subsequently docked into the active site of endothiapepsin by using the FlexX docking module in the LeadIT suite.<sup>8</sup> During the docking, the binding site in the protein was restricted to 9 Å around the bound ligand and the 30 top (FlexX)-scored solutions were retained, and subsequently post-scored with the HYDE<sup>9</sup> module in LeadIT v.2.1.3. After careful visualization to exclude poses with significant inter- or intramolecular clash terms or unfavorable conformations, the resulting solutions were subsequently ranked according to their binding energies. The top-ranked solutions identified in this way were included in the bioisostere library. In all cases, the designed acylhydrazone-mimetics address the catalytic dyad directly via hydrogen-bonding interactions.



**Figure S4:** Schematic representation of the predicted binding modes of bioisosteres of acylhydrazones. (S)-**2**, (S)-**3**, (S)-**4** in the active site of endothiapepsin. These binding modes are the result of a docking run using the FlexX docking module with 30 poses and represent the top-scoring pose after HYDE scoring with SEESAR<sup>10</sup> and careful visual inspection to exclude poses with significant inter- or intramolecular clash terms or unfavorable conformations. The Figure was generated with PoseView<sup>11</sup> as implemented in the LeadIT suite.

## Crystallization, data collection and processing

The protein endothiapepsin was purified from commercially available Suparen 600 by ultrafiltration using a cut-off of 10 kDa and repeated washing with 0.1 mol/l NaOAc buffer, pH= 4.6. After washing the protein concentration was adjusted to 5 mg/ml.

Crystals were obtained using sitting drop vapor-diffusion method with a drop volume of 4  $\mu$ L, using streak-seeding with microcrystals. The conditions were 0.1 mol/L NaOAc, 0.1 mol/L  $\text{NH}_4\text{OAc}$ , 50% PEG4000. Crystallization occurred within two days. Soaking of the ligand was achieved within 24 h in mother liquor containing additional 0.1 mol/L ligand and 25% glycerol.

Crystals were fished and flash-frozen in liquid nitrogen. Data was collected using an Incoatec 1  $\mu$ m X-ray source equipped with an Oxford Cryosystems cryosystem and a Mar 345dtb image plate detector.

The structure was determined using molecular replacement and refined using ccp4i2 and phenix.refine.

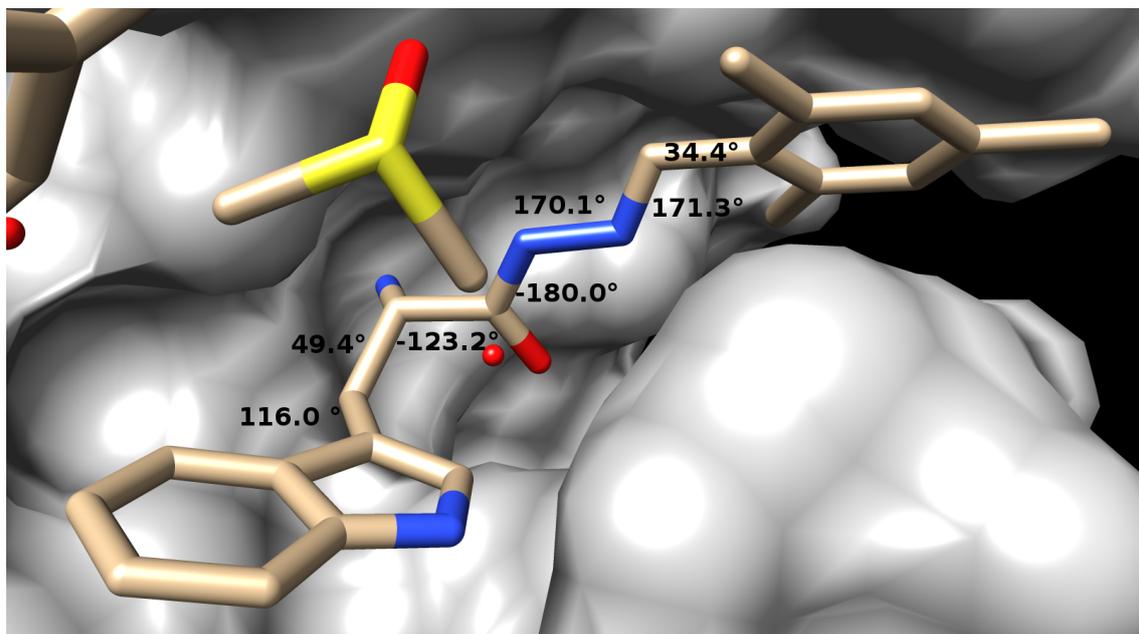
**Table 1:** Data collection and refinement statistics

PDB code	5OJE
wavelength (Å)	1.542
Resolution range (Å) <sup>a</sup>	28.01 - 1.58 (1.64-1.58)
Space group	P 1 21 1
Unit cell: <i>a</i> , <i>b</i> , <i>c</i> (Å)	45.4 73.0 53.1
$\alpha$ , $\beta$ , $\gamma$ (deg)	90 110.1 90
Total reflections	176205 (16432)
Unique reflections	42401 (4399)
multiplicity	4.0 (3.7)
Completeness (%)	99.6 (99.1)
Mean $I/\sigma(I)$	22.9 (6.8)
Wilson B-factor	10.7
R-merge	0.041 (0.185)
R-work	0.143 (0.127)
R-free	0.174 (0.176)
No. of non-hydrogen atoms	2912
Macromolecules	2422
Ligands	51
Water	439
Protein residues	330
RMS bonds (Å)	0.007
RMS angles (deg)	0.94
Ramachandran favored (%) <sup>b</sup>	98.5
Ramachandran outliers (%) <sup>b</sup>	1.5

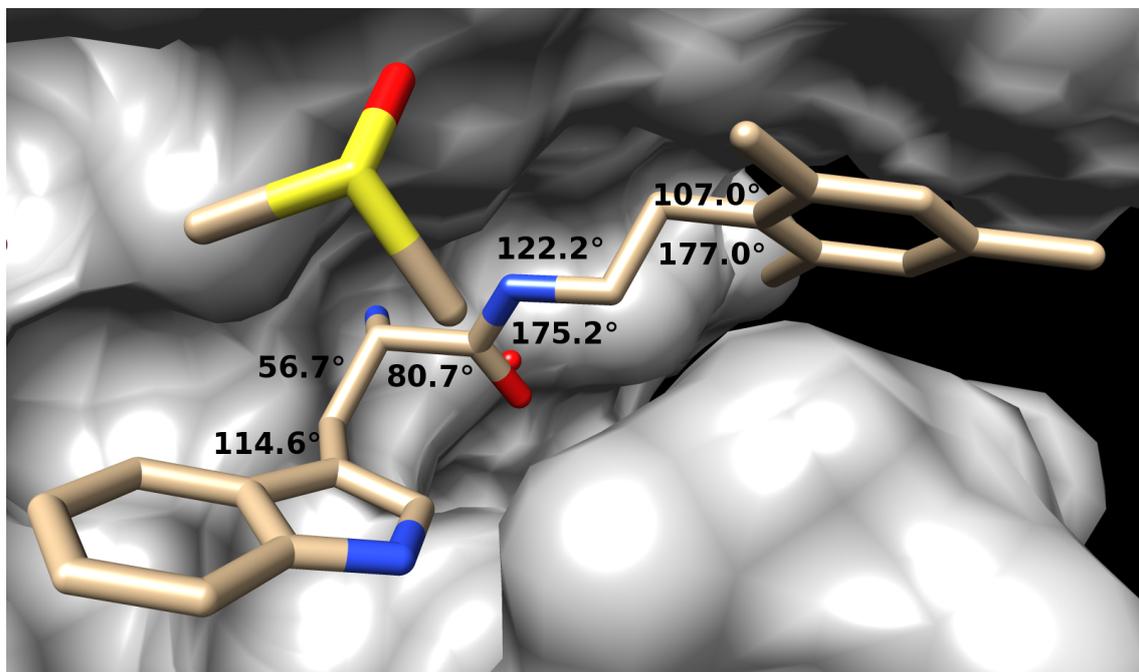
---

Ramachandran allowed (%) <sup>b</sup>	0
Average B-factor (Å <sup>3</sup> )	12.5
macromolecules	10.6
Ligands	17.7
Solvent	22.2
$B_{\text{ligand}}/B_{\text{protein}}$	1.41
Coordinate errors (Å) <sup>c</sup>	0.10

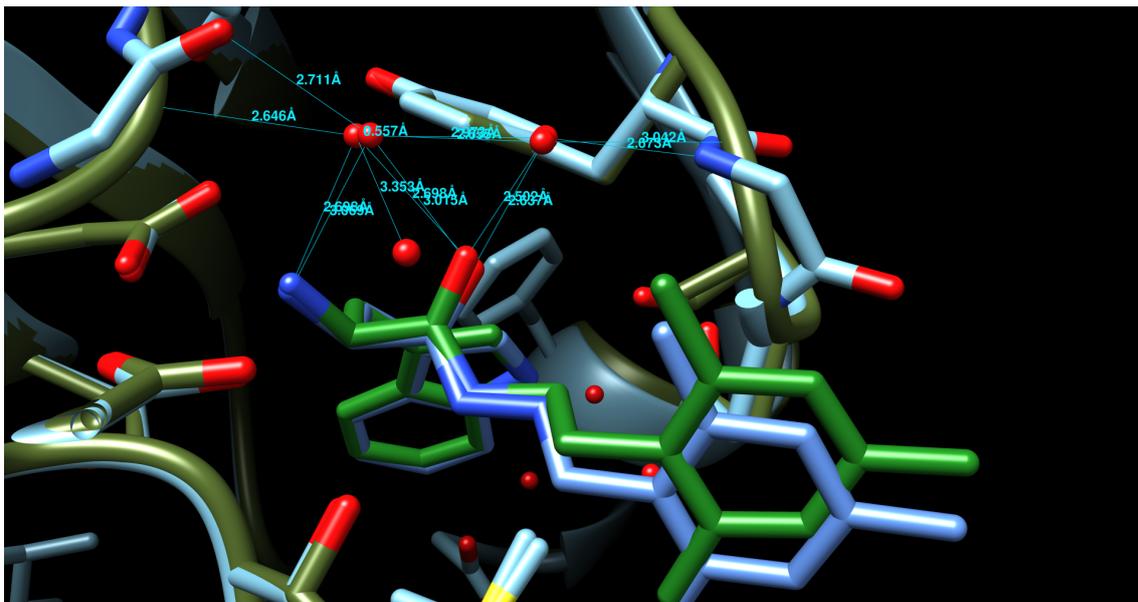
<sup>a</sup> Statistics for the highest-resolution shell are shown in parentheses. <sup>b</sup> Calculated by MOLPROBITY. <sup>c</sup> Maximum-likelihood based

**Structural analysis of torsional angles**

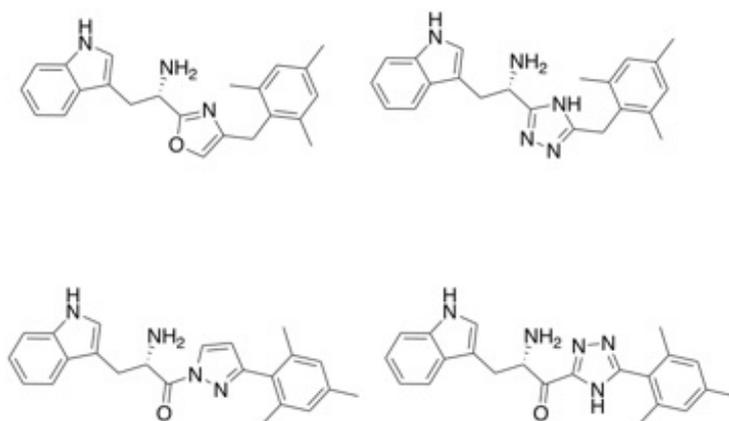
**Figure S5:** The observed orientation of the acylhydrazone (S)-1 in the crystal structure (PDB code: 4kup) with shown torsional angles. The figure was created using UCSF Chimera 1.11 .



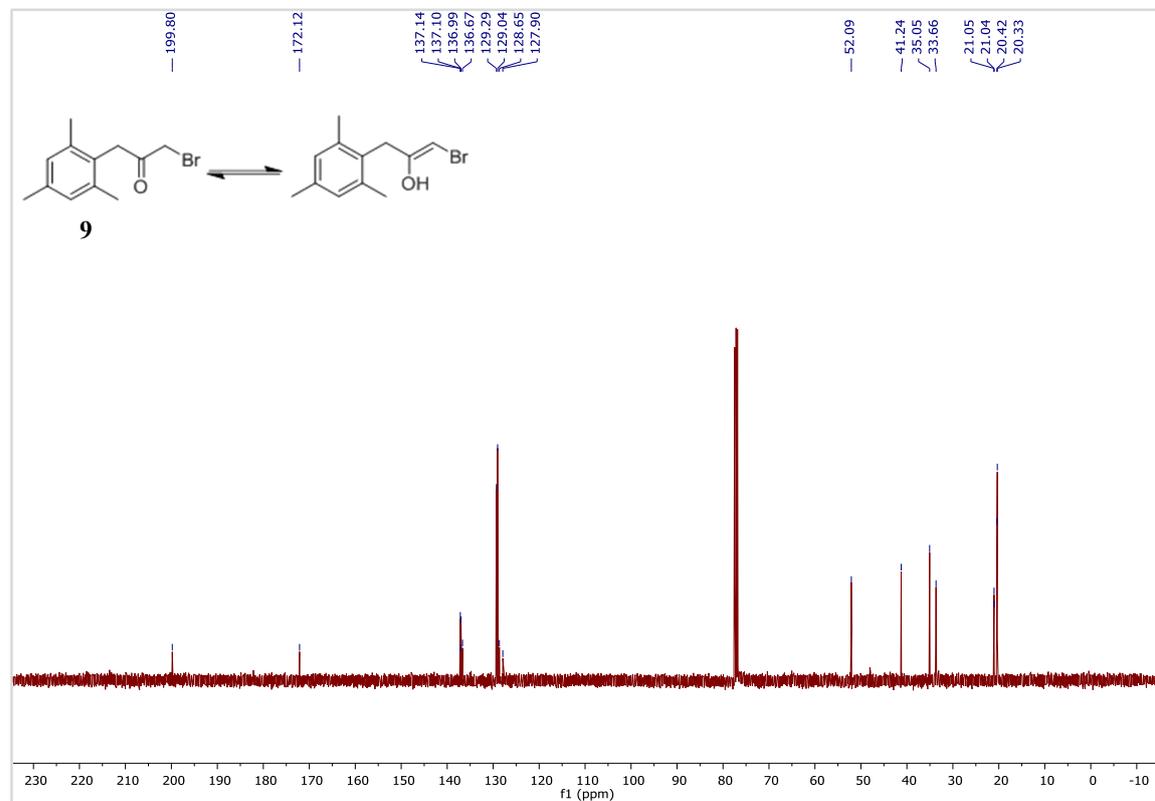
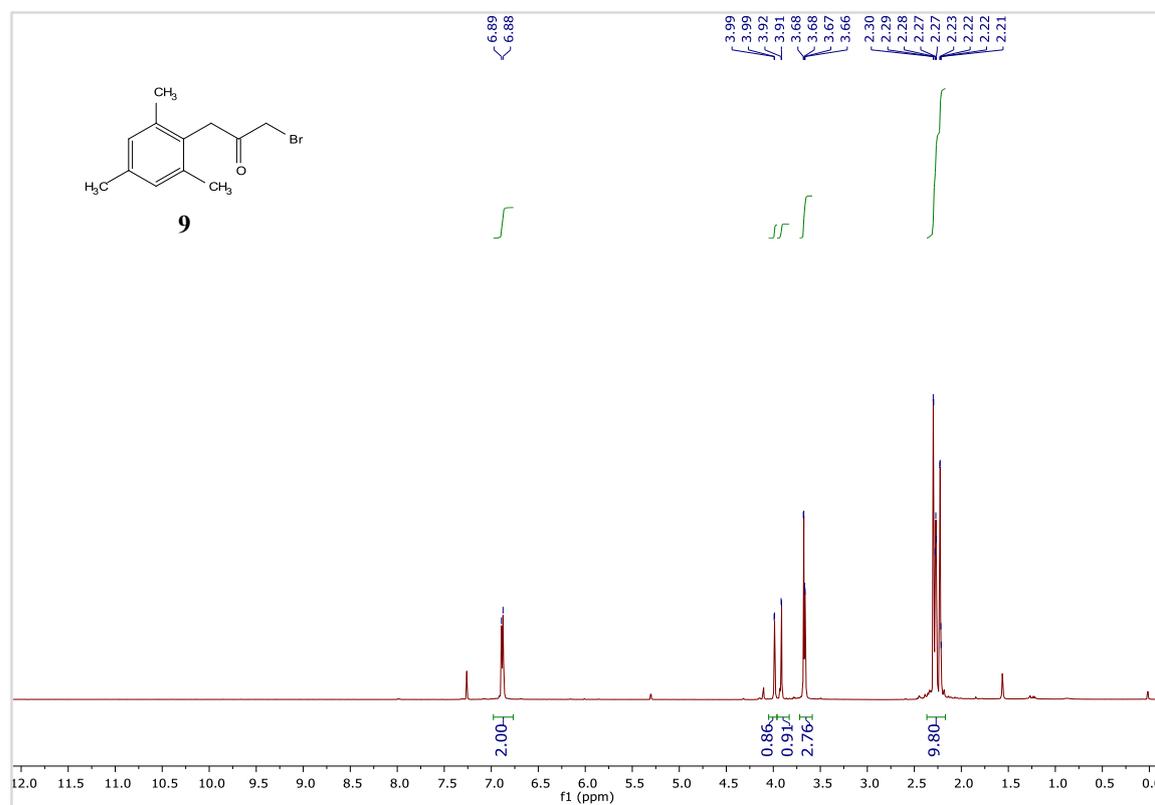
**Figure S6:** The observed orientation of the peptidic bioisostere (S)-2 in the crystal structure (PDB code: 5oje) with shown torsional angles. The figure was created using UCSF Chimera 1.11 .



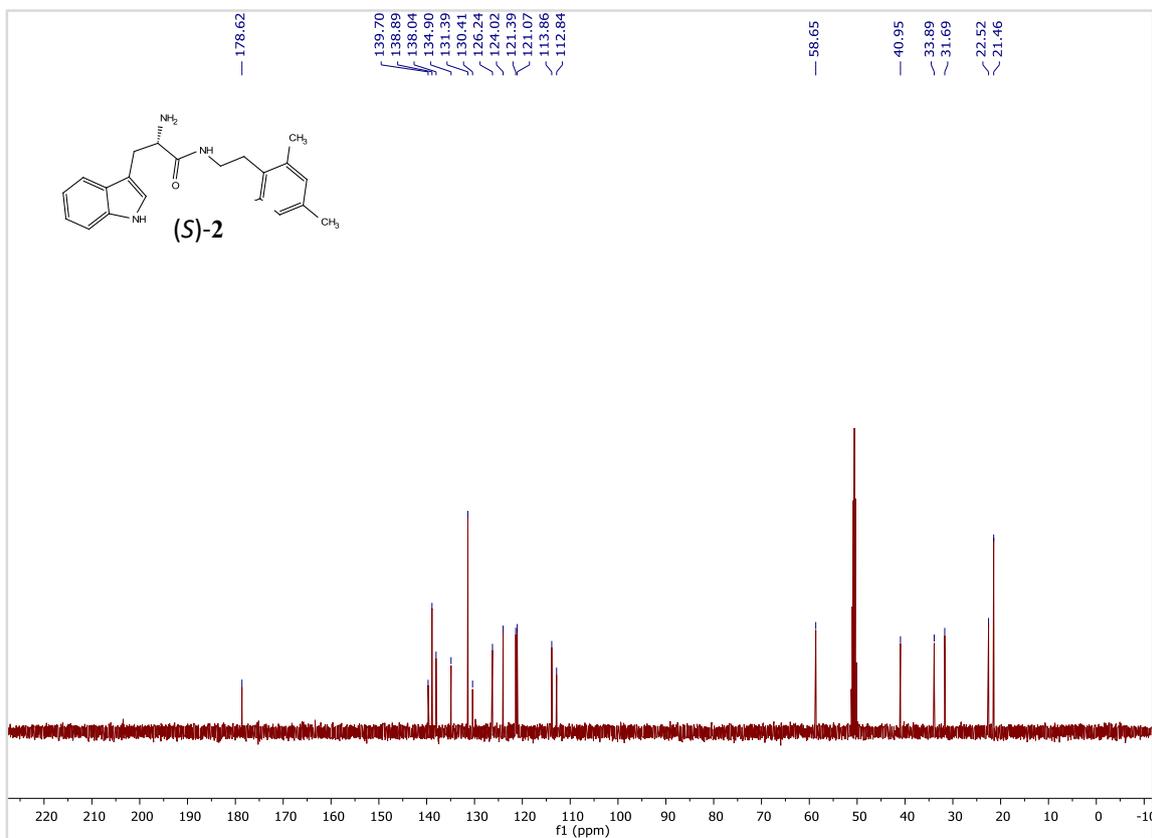
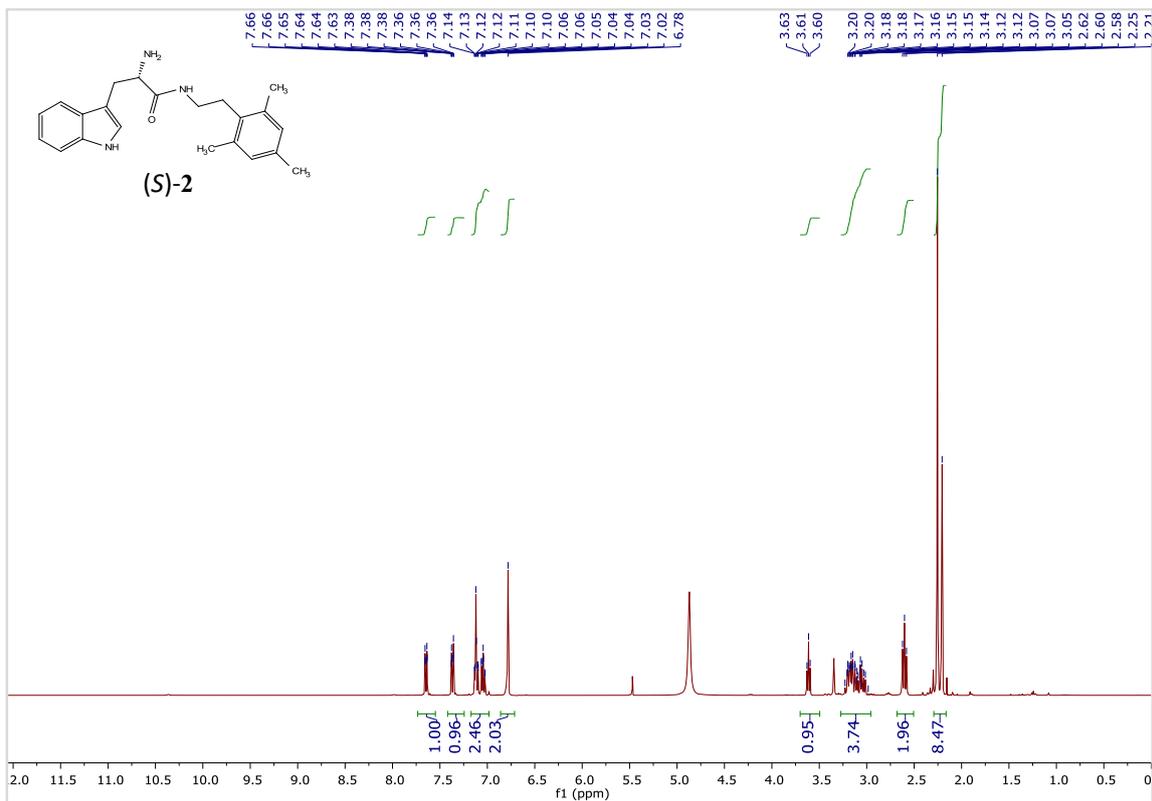
**Figure S7:** The observed network of conserved water molecules. Shown is the superposition of the 4kup structure (blue) with the new crystal structure. Two conserved water molecules can be identified which are interacting as H-bond donors with the oxygen atom of the peptide bond (pdb: 5oje) or acylhydrazone (pdb: 4kup). Bond lengths between the atoms are shown in Angstrom.

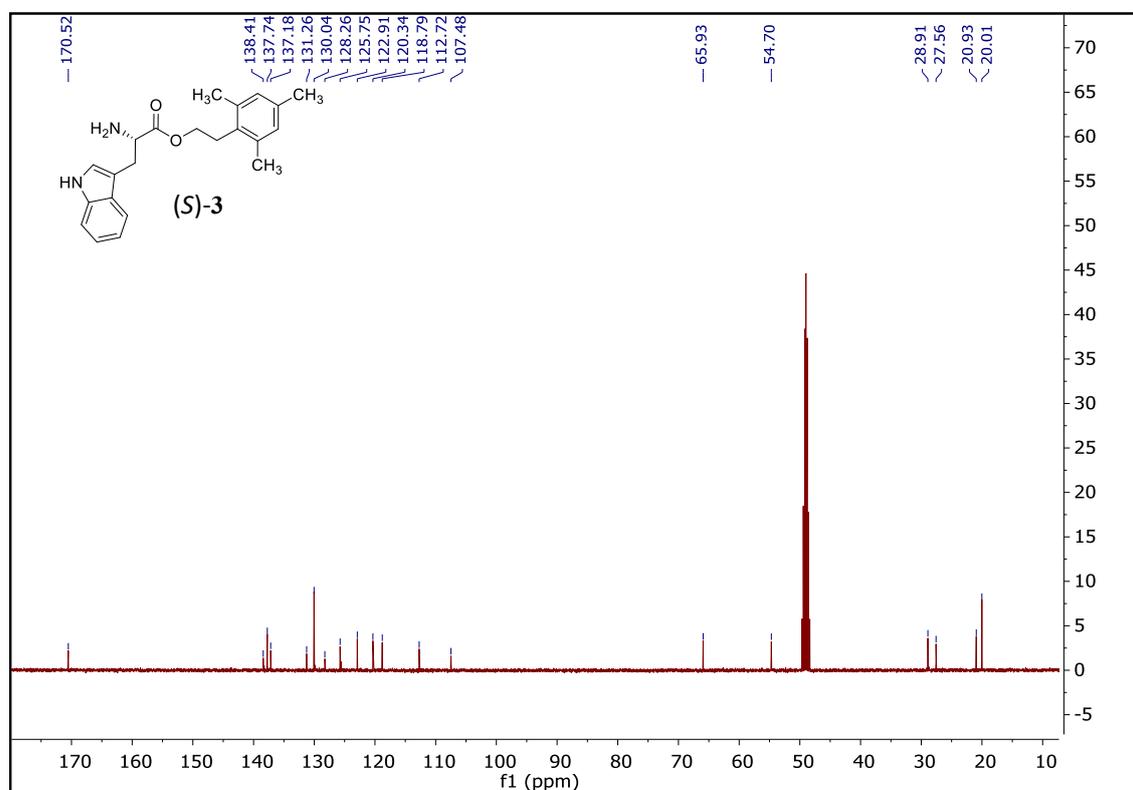
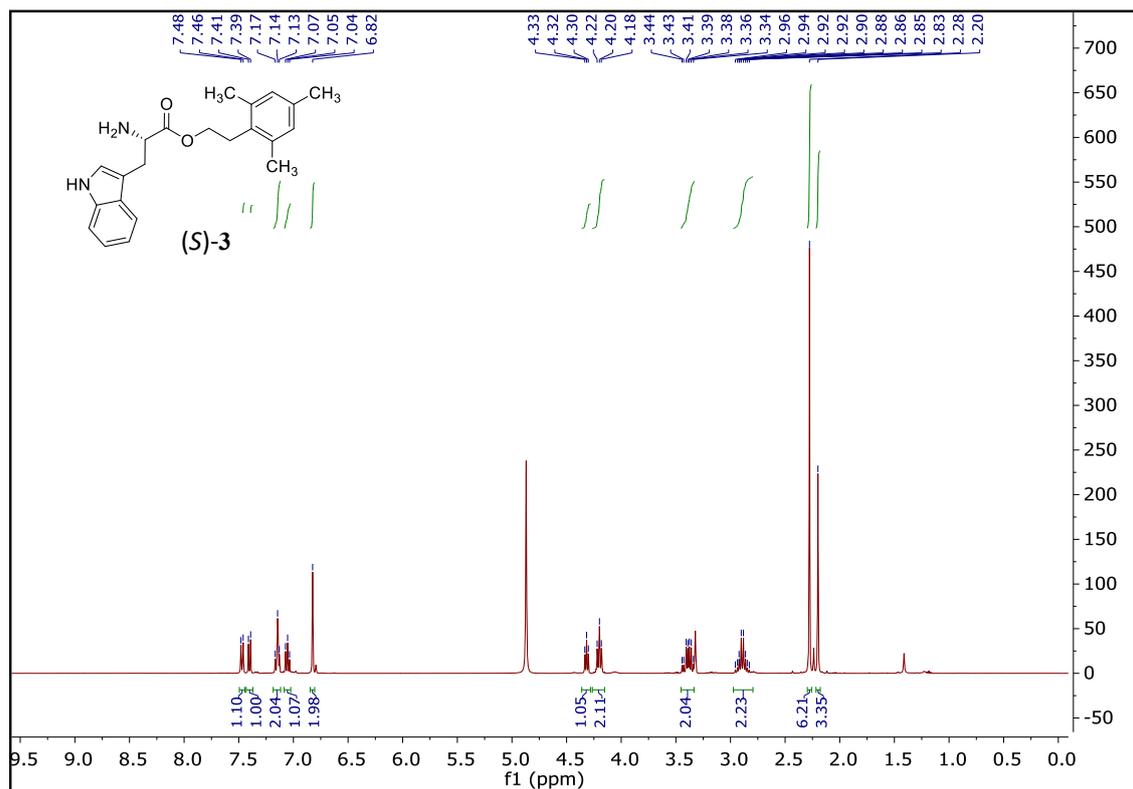


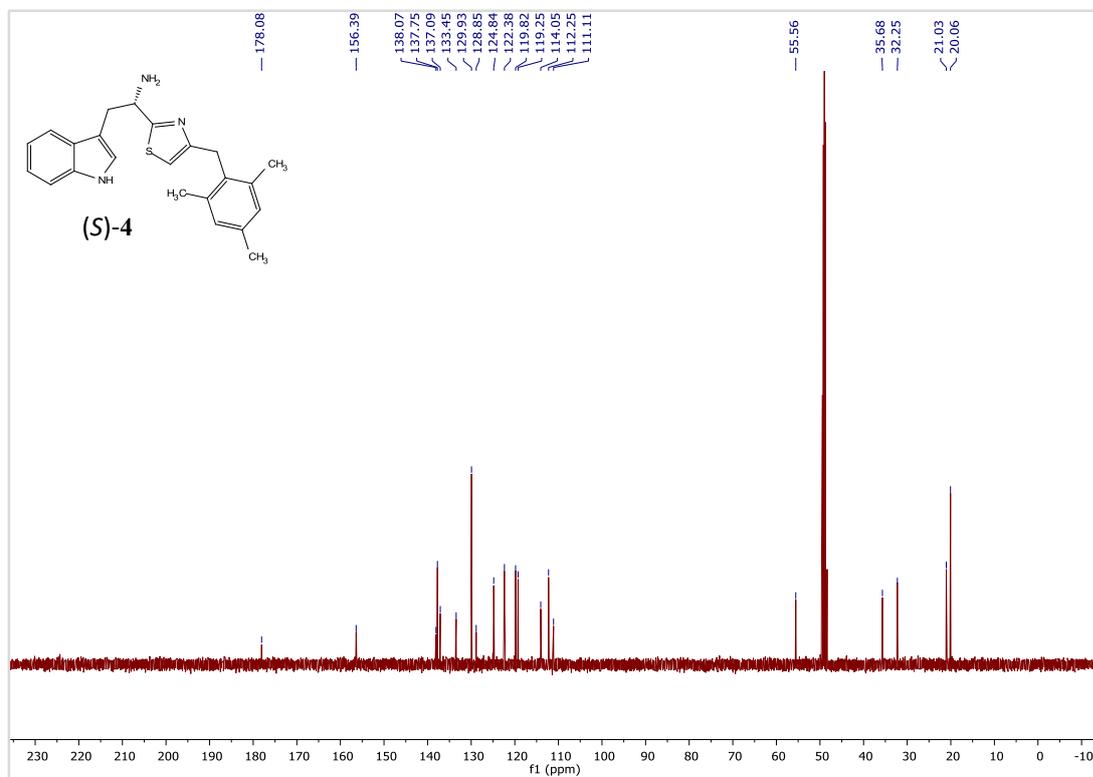
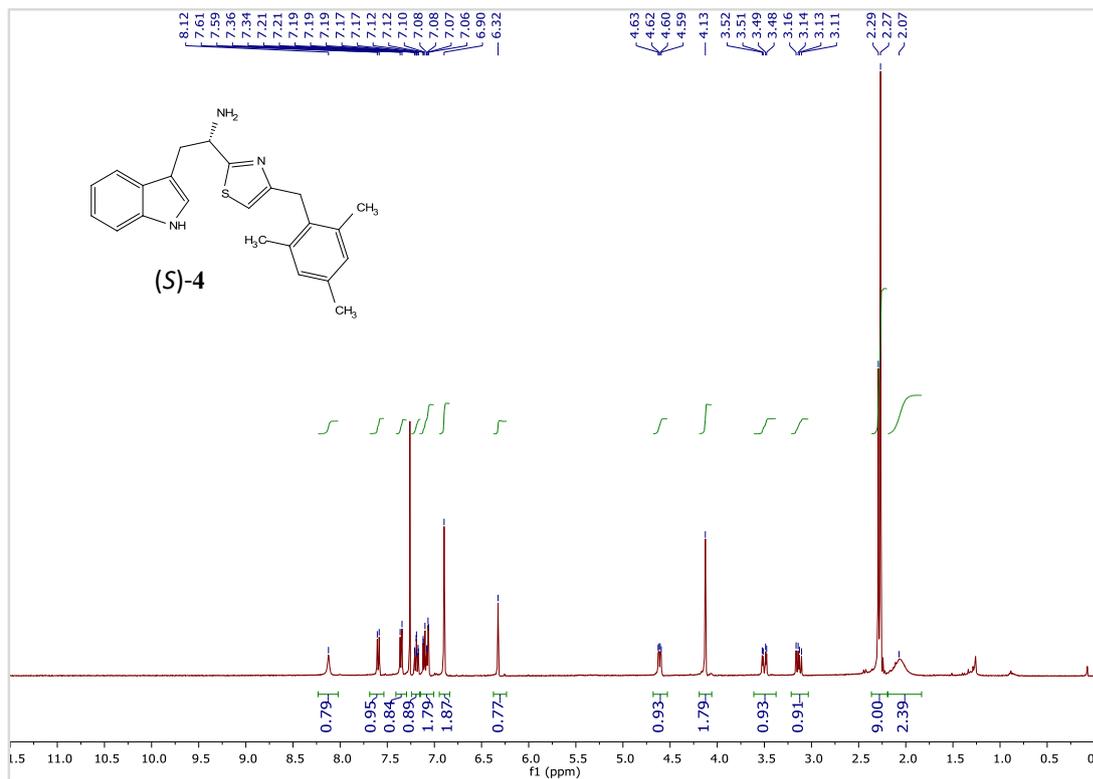
**Figure S8:** Heterocyclic replacements for the hydrazine linker that resulted from modelling and docking studies

**$^1\text{H}$  and  $^{13}\text{C}$  NMR Spectra**

PART 3. SUPPLEMENTARY INFORMATION







## References

- (1) Kavanagh, M. E.; Gray, J. L.; Gilbert, S. H.; Coyne, A. G.; McLean, K. J.; Davis, H. J.; Munro, A. W.; Abell, C. Substrate Fragmentation for the Design of *M. Tuberculosis* CYP121 Inhibitors. *ChemMedChem* **2016**, *11* (17), 1924–1935.
- (2) Kawamoto, I.; Endo, R.; Ishikawa, K.; Kojima, K.; Miyauchi, M.; Nakayama, E. A Convenient Synthesis of Versatile Side-Chain Intermediates for Carbapenem Antibiotics. *Synlett* **1995**, 575–577.
- (3) Gangjee, A.; Yang, J.; Ihnat, M. A.; Kamat, S. Antiangiogenic and Antitumor Agents: Design, Synthesis, and Evaluation of Novel 2-Amino-4-(3-Bromoanilino)-6-Benzylsubstituted pyrrolo[2,3-D]pyrimidines as Inhibitors of Receptor Tyrosine Kinases. *Bioorganic Med. Chem.* **2003**, *11* (23), 5155–5170.
- (4) Note: Triethylamine (2eq., 557  $\mu$ L, 4 mmol) was added to the mesitylethylamine.hydrochloride salt to neutralize and deprotonate the amine.
- (5) Lange, J. H. M.; Kruse, C. G.; van Stuivenberg, H. H.; Sliedregt, L. A. J. M.; *United States Patent Application Publication* **2005**, US 20050065189 A1, 4.
- (6) Mondal, M.; Radeva, N.; Köster, H.; Park, A.; Potamitis, C.; Zervou, M.; Klebe, G.; Hirsch, A. K. H. Structure-Based Design of Inhibitors of the Aspartic Protease Endothiapepsin by Exploiting Dynamic Combinatorial Chemistry. *Angew. Chemie - Int. Ed.* **2014**, *53* (12), 3259–3263.
- (7) Gerber, P. R.; Müller, K. MAB, a Generally Applicable Molecular Force Field for Structure Modelling in Medicinal Chemistry. *J. Comput. Aided. Mol. Des.* **1995**, *9* (3), 251–268.
- (8) BioSolveIT GmbH, Sankt Augustin. <http://www.biosolveit.de>, *LeadIT*, 2.1.2.
- (9) Schneider, N.; Hindle, S.; Lange, G.; Klein, R.; Albrecht, J.; Briem, H.; Beyer, K.; Claußen, H.; Gastreich, M.; Lemmen, C.; et al. Substantial Improvements in Large-Scale Redocking and Screening Using the Novel HYDE Scoring Function. *J. Comput. Aided. Mol. Des.* **2012**, *26* (6), 701–723.
- (10) SeeSAR (version 4.1), BioSolveIT GmbH, An Der Ziegelei 79, 53757 Sankt Augustin, Germany, **2015**;
- (11) Stierand, K.; Rarey, M. Drawing the PDB: Protein–Ligand Complexes in Two Dimensions. *ACS Med. Chem. Lett.* **2010**, *1* (9), 540–545.



# Summary and Outlook

## Summary

As an introduction, **Chapter 1.2** gives a general overview on the DXPS enzyme, with special attention to its suitability for SBDD. In **Chapter 1.3** a brief update on the recent developments on DXPS-related research since 2017 is given. **Chapter 1.4** introduces the hit-identification method protein-templated dynamic combinatorial chemistry (ptDCC).

In **Chapter 2.1**, we describe the optimization of the DXPS protein from *D. radiodurans* to obtain a truncated protein with improved crystallographic properties.

In the initial publication of the DXPS structure, partial proteolysis was reported as essential for protein crystal formation. While reproducing the literature conditions, we obtained protein crystals diffracting to a low resolution that also showed proteolysis. We identified the exact position of the labile loop using native LC-MS, which was in agreement with a missing part in the protein crystal structure. A multiple sequence alignment (MSA) of ~500 bacterial DXPS sequences further confirmed the identified loop, as the sequence of this part had a high evolutionary variability.

A protein construct replacing the loop with a polyglycine linker was designed and obtained commercially as a synthetic gene. The truncated protein expressed well, had similar biophysical properties and improved stability against degradation. It also formed protein crystals that diffracted well up to an improved resolution of 2.1 Å. Since the identified part of the protein is present in most species, we postulated that its removal should also work on other DXPS homologues, enabling their structural determination.

In **Chapter 2.2**, we describe the application of the truncation strategy of Chapter 2.1 to the DXPS homologue from *M. tuberculosis*, which until then was not forming protein crystals.

Transfer of the truncation method to *M. tuberculosis* DXPS (mtDXPS) resulted in a stable protein, which was expressing in higher yields. The improved protein formed protein crystals that diffracted well, and the protein structure of the DXPS homologue was solved with a resolution of 1.85 Å. A comparison of the active site of known DXPS structures with the new mtDXPS model is pointing out a small, but surprising difference. Close to the active site, a motif of the amino acids GSSH can be found, which might be the reason for the very slow catalysis rate of mtDXPS, compared with other DXPS enzymes. The mtDXPS structure is also the first one with a crystallographic hint on the 3-GAP binding site, allowing the docking of GAP competitive DXPS inhibitors. We showed this using the hydroxybenzaldoxime inhibitor class as an example.

In **Chapter 2.3**, we describe the results of a ligand-based approach to design inhibitors targeting the DXPS enzyme from *P. falciparum*.

Parallel to the work on the protein crystal structures, we performed a ligand-based approach to design new inhibitors similar to ThDP, the cofactor of DXPS. The initial hits from a study on *M. tuberculosis* were evaluated and optimized for their biological activity on *P. falciparum*. The best compound class showed an IC<sub>50</sub> of 0.6 μM. Although extensive experiments to confirm DXPS as the target of the three compound classes were performed, we could not obtain conclusive results. However, the compounds were showing clear anti-plasmodial activities in cell-based assays, measured independently in three laboratories. Because of this, a bioinformatic search for other possible targets was performed and the most promising targets discussed.

In **Chapter 2.4**, ptDCC was applied to the target DXPS from *D. radiodurans* as a hit-identification and -optimization method.

Starting with the DXPS protein structure from *D. radiodurans*, a focused ptDCC library of aldehydes and hydrazides was designed targeting the ThDP-binding pocket of DXPS. The initial hits were subsequently optimized in two additional rounds of ptDCC, in which the library compounds were selected based on the activity of compounds of the previous ptDCC library, improving the binding affinity into the single-digit micromolar range for the best compounds of the third ptDCC round. The mode of inhibition (MOI) of the representative hit **1** was determined to be ThDP- and pyruvate-competitive and non-competitive with respect to 3-GAP.

In addition, the best compounds from all three rounds of DCC were evaluated for their biological activity against *M. tuberculosis*, with compounds **7** and **35** showing an MIC of ~64  $\mu\text{M}$ . Finally, we applied the findings obtained in Chapter 2.5 and synthesized bioisosteres for the hit compounds **1** and **23**, which had similar binding affinities and only slightly reduced inhibitory activity.

**Chapter 2.5** describes a proof-of-concept study for the development of bioisosteres for acylhydrazone-based ptDCC-hits.

Based on a previously obtained Endothiapepsin co-crystal structure with an acylhydrazone hit-compound the two different softwares ReCore and LeadIT were used to model and dock bioisosteric linkers, replacing the acylhydrazone. The three best scoring compounds were synthesized and their activities evaluated. The amide linker had, with an  $\text{IC}_{50}$  of  $12.9 \pm 0.7 \mu\text{M}$ , a nearly identical activity compared to the acylhydrazone hit with an  $\text{IC}_{50}$  of  $12.8 \pm 0.4 \mu\text{M}$ . In a 1.58 Å co-crystal structure that was obtained by soaking, the binding mode of the bioisostere could be compared with the acylhydrazone. Both compounds showed a similar pose, with the amide retaining the atomic interactions of the acylhydrazone.

## Outlook

Based on the findings reported in Chapters 2.1 and 2.2, our group is now able to create DXPS proteins with improved properties. We recently designed constructs for the homologues from *Klebsiella pneumoniae* and *Pseudomonas aeruginosa*, which are both looking promising. The proteins already yielded first protein crystals and we are expecting to solve and publish their structures in the near future.

Besides obtaining the structures of new homologues, our next aim in the structural biology part of this project are co-crystal structures of DXPS with inhibitors. Experiments to obtain these complex structures with our hit-compounds are ongoing. Limiting factors have been the relatively low binding affinity and solubility of the compounds. We are tackling these by synthesizing more soluble derivatives, and expanding the chemical space of our compounds by using different hit-identification strategies. It might also be possible that the protein crystals exists in a non-soakable crystal form and co-crystallization needs to be used. However, as we are working on all these points, it should be possible to solve complex structures in the future.

If obtained, the co-crystal structures will give us valuable information on the binding mode of the inhibitors and enable us to perform structure-based optimization of our hits.

Regarding the development of the new anti-plasmodials of Chapter 2.3, additional experiments validating DXPS, or finding an alternative target, are ongoing. Although the expression and purification of *P. falciparum* DXPS proved to be challenging, we recently made good progress and should be able to test the compounds in an *in-vitro* activity assay soon. As second approach, we are aiming to generate resistant strains and analyze the mode of resistance by whole-genome sequencing. In parallel, we would like to experimentally test the proposed alternative targets, in particular the Valosin-containing protein (VCP) and started to reach out for collaboration partners who have the assay established in their lab.

Regarding drug design, the next big steps will be the use of the new structures from pathogens to start different structure-based projects. A structure-based virtual screening (SBVS) of our in-house HIPS-library will be one of our first steps, expanding the screen to commercial compound libraries, if needed.

During the time of this PhD project, an application on a AI-based SBVS was written and submitted and we were granted the AIMS-Award by the company Atomwise. The collaboration already started and the virtual screening was performed on the first DXPS crystal structure of *M. tuberculosis*, previous to its deposition in the pdb-archive. As part of the award, the resulting VS-hits were supplied by custom synthesis from the company Enamine and characterized in our lab. Expanding the initial grant, a second set of compounds based on our initial hits was ordered and the compounds are currently undergoing characterization.

## Samenvatting

Bij wijze van inleiding wordt in **Hoofdstuk 1.2** een algemeen overzicht gegeven van het DXPS-enzym, met specifieke aandacht voor zijn geschiktheid voor SBDD. In **Hoofdstuk 1.3** wordt een korte update gegeven van de recente ontwikkelingen op het gebied van DXPS-gerelateerd onderzoek sinds 2017. **Hoofdstuk 1.4** introduceert de hit-identificatie methode eiwit-gerichte dynamische combinatorische chemie (ptDCC).

In **Hoofdstuk 2.1** beschrijven we de optimalisatie van het DXPS-eiwit uit *D. radiodurans* om een getrunceerd eiwit met verbeterde kristallografische eigenschappen te verkrijgen.

In de eerste publicatie van de DXPS-structuur, werd gedeeltelijke proteolyse gerapporteerd als essentieel voor de kristalvorming van het eiwit. Terwijl we de omstandigheden uit de literatuur reproduceerden, verkregen we eiwitkristallen met een lage resolutie die ook proteolyse vertoonden. We identificeerden de exacte positie van de labiele lus met behulp van natieve LC-MS, die in overeenstemming was met een ontbrekend deel in de kristalstructuur van het eiwit. Een meervoudige sequentie uitlijning (MSA) van ~500 bacteriële DXPS-sequenties bevestigde de geïdentificeerde lus, aangezien de sequentie van dit deel een hoge evolutionaire variabiliteit had.

Een eiwitconstruct dat de lus vervangt door een polyglycine linker werd ontworpen en commercieel verkregen als een synthetisch gen. Het getrunceerde eiwit kwam goed tot expressie, had vergelijkbare biofysische eigenschappen en een verbeterde stabiliteit tegen degradatie. Het vormde ook eiwitkristallen die een goede diffractie lieten zien; tot een verbeterde resolutie van 2.1 Å. Aangezien het geïdentificeerde deel van het eiwit aanwezig is in de meeste soorten, veronderstelden we dat de verwijdering ervan ook zou moeten werken op andere homologen van DXPS, waardoor hun structuur kan worden bepaald.

In **Hoofdstuk 2.2** beschrijven we de toepassing van de truncatiestrategie van hoofdstuk 2.1 op de DXPS-homoloog van *M. tuberculosis*, die tot dan toe geen eiwitkristallen vormden.

Het toepassen van de truncatiemethode op *M. tuberculosis* DXPS (mtDXPS) resulteerde in een stabiel eiwit, dat in hogere opbrengsten tot expressie werd gebracht. Het verbeterde eiwit vormde eiwitkristallen die goede diffractiedata opleverde; de eiwitstructuur van de DXPS-homoloog werd opgelost met een resolutie van 1.85 Å. Een vergelijking van het actieve centrum van bekende DXPS-structuren met het nieuwe mtDXPS-model wijst op een klein, maar verrassend verschil. Dichtbij het actieve centrum kan een motief van de aminozuren GSSH worden gevonden, dat de reden zou kunnen zijn voor de zeer trage katalysesnelheid van mtDXPS, vergeleken met andere DXPS-enzymen. De mtDXPS-structuur is ook de eerste met een kristallografische hint wat betreft de 3-GAP bindingsplaats, waardoor het docken van GAP competitieve DXPS-remmers mogelijk wordt. Wij hebben dit aangetoond met de remmer klasse hydroxybenzaldoxime als voorbeeld.

In **Hoofdstuk 2.3** beschrijven we de resultaten van een ligand-gebaseerde benadering voor het ontwerpen van remmers gericht tegen het DXPS-enzym van *P. falciparum*.

Parallel aan het werk aan de kristalstructuren van de eiwitten, hebben we een ligand-gebaseerde benadering ondernomen om nieuwe remmers te ontwerpen die vergelijkbaar zijn met ThDP, de cofactor van DXPS. De eerste hits uit een studie met *M. tuberculosis* werden geëvalueerd en geoptimaliseerd voor hun biologische activiteit tegen *P. falciparum*. De beste klasse van verbindingen vertoonde een IC<sub>50</sub> van 0.6 µM. Hoewel uitgebreide experimenten om DXPS te bevestigen als het doelwit van de drie samengestelde klassen werden uitgevoerd, konden we geen eenduidige resultaten verkrijgen. De verbindingen vertoonden echter duidelijke anti-plasmodiale activiteiten in cel-gebaseerde testen, onafhankelijk gemeten in drie laboratoria. Daarom werd een bioinformatische zoektocht naar andere mogelijke doelwitten uitgevoerd en de meest veelbelovende doelwitten besproken.

In **Hoofdstuk 2.4** werd ptDCC toegepast op het doelwit DXPS van *D. radiodurans* als een hit-identificatie en -optimalisatie methode.

Uitgaande van de DXPS-eiwitstructuur van *D. radiodurans*, werd een doelgerichte ptDCC-bibliotheek van aldehyden en hydraziden ontworpen, gericht op de ThDP-binding pocket van

DXPS. De initiële treffers werden vervolgens geoptimaliseerd in twee extra rondes van ptDCC, waarbij de verbindingen uit de bibliotheek werden geselecteerd op basis van de activiteit van verbindingen uit de vorige ptDCC-bibliotheek, waardoor de bindingsaffiniteit verbeterde tot in het micromolaire voor de beste verbindingen van de derde ptDCC-ronde. De wijze van remming (MOI) van de representatieve hit **1** werd bepaald als ThDP- en pyruvaat-competitief en niet-competitief ten opzichte van 3-GAP. Bovendien werden de beste verbindingen uit alle drie de DCC-rondes geëvalueerd op hun biologische activiteit tegen *M. tuberculosis*, waarbij verbindingen **7** en **35** een MIC van  $\sim 64 \mu\text{M}$  lieten zien. Tenslotte pasten we de bevindingen uit hoofdstuk 2.5 toe en synthetiseerden bio-isosteren voor de hit-verbindingen **1** en **23**, die vergelijkbare bindingsaffiniteiten hadden en slechts een licht verminderde remmende activiteit.

**Hoofdstuk 2.5** beschrijft een proof-of-concept studie voor de ontwikkeling van bio-isosteren voor acylhydrazon-gebaseerde ptDCC-hits.

Gebaseerd op een eerder verkregen co-kristalstructuur van Endothiapepsine met een acylhydrazon-gebaseerde hit-verbinding werden twee verschillende programmas, ReCore en de LeadIT-suite, gebruikt om bio-isosterische linkers ter vervanging van de acylhydrazon te modelleren en te docken. De drie best scorende verbindingen werden gesynthetiseerd en hun activiteiten werden geëvalueerd. De amide-linker had, met een  $\text{IC}_{50}$  van  $12.9 \pm 0.7 \mu\text{M}$ , een bijna identieke activiteit vergeleken met de acylhydrazon-hit met een  $\text{IC}_{50}$  van  $12.8 \pm 0.4 \mu\text{M}$ . In een  $1.58 \text{ \AA}$  co-kristalstructuur die werd verkregen door inweken, kon de bindingswijze van de bio-isosteer worden vergeleken met die van de acylhydrazon. Beide verbindingen vertoonden een vergelijkbare bindingswijze, waarbij het amide de atomaire interacties van het acylhydrazon behield.

## Zusammenfassung

Einleitend wird in **Kapitel 1.2** ein allgemeiner Überblick über das DXPS-Enzym gegeben, mit besonderem Augenmerk auf die Eignung des Enzymes als SBDD-Ziel. Im **Kapitel 1.3** werden die aktuellen Entwicklungen im Bereich der DXPS-Forschung seit 2017 zusammengefasst. **Kapitel 1.4** gibt eine Einführung in die Hit-Identifikationsmethode protein-templated dynamic combinatorial chemistry (ptDCC).

In **Kapitel 2.1** wird die Optimierung des DXPS-Proteins von *D. radiodurans* beschrieben, mit dem Ziel ein Protein mit verbesserten kristallographischen Eigenschaften zu erzeugen.

In der initialen Strukturaufklärung wurde die partielle Proteolyse als wesentlich für die Kristallbildung des Proteins beschrieben. Wir konnten die Literaturbedingungen reproduzieren und erhielten Proteinkristalle mit niedriger Auflösung, die ebenfalls Proteolyse zeigten. Mittels nativer LC-MS konnte die genaue Position der labilen Stelle identifiziert werden, diese stimmt mit dem fehlenden Teil in der Kristallstruktur des Proteins überein. Ein multiples Sequenzalignment (MSA) von ~500 bakteriellen DXPS-Sequenzen bestätigte die identifizierte Stelle, die Sequenz dieses Teils weist eine hohe evolutionäre Variabilität auf.

Ein Proteinkonstrukt, in dem die identifizierte Schleife durch einen Polyglycin-Linker ersetzt ist, wurde entworfen und als synthetisches Gen kommerziell erworben. Das verkürzte Protein wurde gut exprimiert, hatte ähnliche biophysikalische Eigenschaften und zeigte eine verbesserte Stabilität gegenüber proteolytischen Abbau. Das verkürzte Protein hatte eine höhere Tendenz zur Proteinkristallbildung, die erhaltenen Kristalle beugten bis zu einer verbesserten Auflösung von 2,1 Å. Da der identifizierte, labile Teil des Proteins in den meisten Spezies vorhanden ist wurde postuliert, dass seine Entfernung auch bei anderen DXPS-Homologen funktionieren, und deren Strukturbestimmung ermöglichen sollte.

In **Kapitel 2.2** wird die Anwendung der Trunkierungsstrategie aus Kapitel 2.1 auf das DXPS-Homolog von *M. tuberculosis* beschrieben, das bis dahin keine Proteinkristallbildung zeigte.

Die Anwendung der Trunkierungsmethode auf *M. tuberculosis* DXPS führte zu einem stabilen Protein, das in höherer Ausbeute als der Wildtyp exprimiert werden konnte. Das verbesserte Protein bildete Proteinkristalle, die gut beugten, und die Proteinstruktur des DXPS-Homologs konnte mit einer Auflösung von 1,85 Å gelöst werden. Ein Vergleich des aktiven Zentrums bekannter DXPS-Strukturen mit dem neuen DXPS-Modell von *M. tuberculosis* deutet auf einen kleinen, aber überraschenden Unterschied hin. In der Nähe des aktiven Zentrums ist ein Motiv der Aminosäuren GSSH zu finden, das möglicherweise der Grund für die sehr langsame Katalysegeschwindigkeit von mtDXPS im Vergleich zu anderen DXPS-Enzymen sein könnte. Die mtDXPS-Struktur ist auch die erste mit einem kristallographischen Hinweis auf die 3-GAP Bindungsstelle, was Docking-Berechnungen von GAP-kompetitiven DXPS-Inhibitoren ermöglicht. Wir haben dies am Beispiel der Hydroxybenzaldoxim-Inhibitor-Klasse gezeigt.

In **Kapitel 2.3** werden die Ergebnisse eines ligandenbasierten Ansatzes zur Entwicklung von Inhibitoren gegen das DXPS-Enzym von *P. falciparum* beschrieben.

Parallel zu den Arbeiten an den Proteinkristallstrukturen wurde ein ligandenbasierter Ansatz verfolgt, um Inhibitoren mit Ähnlichkeit zum DXPS-cofaktor ThDP zu finden. Aktive Verbindungen aus einer Studie an *M. tuberculosis* wurden evaluiert und auf ihre biologische Aktivität gegen *P. falciparum* hin optimiert. Die beste Verbindungsklasse wies einen IC<sub>50</sub> von 0,6 µM auf. Obwohl umfangreiche Experimente zur Bestätigung von DXPS als Ziel der drei Verbindungsklassen durchgeführt wurden, konnten wir dieses nicht eindeutig validieren.

Die Verbindungen zeigten jedoch in zellbasierten Assays eine deutliche Anti-Plasmodien-Aktivität, die unabhängig voneinander in drei Laboratorien gemessen wurde. Aus diesem Grund wurde eine bioinformatische Suche nach anderen möglichen Zielenzymen durchgeführt und die vielversprechendsten Enzyme diskutiert.

In **Kapitel 2.4** wurde protein-templated dynamic combinatorial chemistry (ptDCC) sowohl als

Hit-Identifizierung, als auch als Hit-Optimierungsmethode genutzt, um Inhibitoren von DXPS aus *D. radiodurans* zu entwickeln.

Ausgehend von der *D. radiodurans* DXPS-Proteinstruktur wurde eine fokussierte ptDCC Bibliothek aus Aldehyden und Hydraziden entworfen, angepasst an die ThDP-Bindungstasche von DXPS. Die initialen Treffer wurden anschließend in zwei weiteren Runden ptDCC optimiert, in denen die Verbindungen der Bibliothek auf der Grundlage der Aktivität der Verbindungen der ptDCC-Bibliothek ausgewählt wurden. Die Bindungsaffinität der besten Verbindungen aus der dritten Runde erreichte hierbei den einstelligen mikromolaren Bereich. Der Wirkmechanismus (MOI) des repräsentativen Treffers **1** wurde bestimmt. Die Verbindung **1** zeigte ein ThDP- und Pyruvat-kompetitive und 3-GAP nicht-kompetitive Inhibition. Darüber hinaus wurden die besten Verbindungen aus allen drei Runden der DCC auf ihre biologische Aktivität gegen *M. tuberculosis* untersucht, wobei die Verbindungen **7** und **35** eine minimum inhibitory concentration (MIC) von 64  $\mu$ M zeigten. Schließlich wurden die in Kapitel 2.5 gewonnenen Erkenntnisse angewandt und Bioisostere für die Verbindungen **1** und **23** synthetisiert, die eine ähnliche Bindungsaffinität und nur geringfügig reduzierte hemmende Aktivität aufwiesen.

**Kapitel 2.5** beschreibt eine *proof-of-concept*-Studie um Bioisostere für Acylhydrazon-basierte ptDCC-Hitverbindungen zu entwickeln.

Auf der Grundlage einer zuvor erhaltenen Endothiapepsin Co-Kristallstruktur mit einem gebundenen Acylhydrazon wurden die beiden verschiedenen Softwareprogramme Recore und LeadIT verwendet, um Bioisostere zu dem Acylhydrazon-Linker zu modellieren und ihre Bindungsaffinität zu berechnen. Die drei am besten bewerteten Verbindungen wurden synthetisiert und getestet. Die Verbindung mit einem Amidlinker hatte mit einem  $IC_{50}$  von  $12,9 \pm 0,7 \mu$ M eine nahezu identische Aktivität wie das Acylhydrazon mit einem  $IC_{50}$  von  $12,8 \pm 0,4 \mu$ M. Mittels einer Co-Kristallstruktur mit einer Auflösung von 1.58 Å, die durch soaking erhalten wurde, konnte der Bindungsmodus des Bioisosters mit dem des Acylhydrazons verglichen werden. Beide Verbindungen zeigten eine sehr ähnliche Orientierung, wobei das Amid die atomaren Wechselwirkungen des Acylhydrazons nachbildete.



## Publications not Included in this Thesis

The following publications were written in the time period from 2016 to 2022 and not included in this thesis.

**Publication 1:** M. Yagiz Unver ‡, [Robin M. Gierse](#) ‡, Harry Ritchie and Anna K. H. Hirsch: Druggability Assessment of Targets Used in Kinetic Target-Guided Synthesis; **2018**, *Journal of Medicinal Chemistry*, Volume 61, Issue 21, Pages 9395-9409

**Publication 2:** Di Zhu ‡, Sandra Johannsen ‡, Tiziana Masini ‡, Céline Simonin ‡, Jörg Hauptenthal, Boris Illarionov, Anastasia Andreas, Mahendra Awale, Robin M. Gierse, Tridia van der Laan, Ramon van der Vlag, Rita Nasti, Mael Poizat, Eric Buhler, Norbert Reiling, Rolf Müller, Markus Fischer, Jean-Louis Reymond,\* Anna K. H. Hirsch\* Discovery of novel drug-like antitubercular hits targeting the MEP pathway enzyme DXPS by strategic application of ligand-based virtual screening; **2022**, *ChemRxiv*, DOI:10.26434/chemrxiv-2022-73fb5-v2

‡: Authors contributed equally

\*: Corresponding Authors



# Bibliography

- [1] Creativecommons license BY-NC-3.0, . URL <https://creativecommons.org/licenses/by-nc/3.0/>.
- [2] Creativecommons license BY-NC-3.0, . URL <https://creativecommons.org/licenses/by-nc/4.0/>.
- [3] drDXS UniProt entry with the accession code Q9RUB5, accessed 08/2021. URL <https://www.uniprot.org/uniprot/Q9RUB5>.
- [4] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell. 5th edition*. Garland Science, New York, 5th editio edition, 2007.
- [5] M. Albrecht, N. Misawa, and G. Sandmann. Metabolic engineering of the terpenoid biosynthetic pathway of *Escherichia coli* for production of the carotenoids  $\beta$ -carotene and zeaxanthin. *Biotechnol. Lett.*, pages 791–795, 1999.
- [6] B. Altincicek, M. Hintz, S. Sanderbrand, J. Wiesner, E. Beck, and H. Jomaa. Tools for discovery of inhibitors of the 1-deoxy-D-xylulose 5-phosphate (DXP) synthase and DXP reductoisomerase: An approach with enzymes from the pathogenic bacterium *Pseudomonas aeruginosa*. *FEMS Microbiol. Lett.*, 190(2):329–333, 2000. ISSN 03781097. doi: 10.1016/S0378-1097(00)00357-8.
- [7] D. Arigoni, W. Eisenreich, C. Latzel, S. Sagner, T. Radykewicz, M. H. Zenk, and A. Bacher. Dimethylallyl pyrophosphate is not the committed precursor of isopentenyl pyrophosphate during terpenoid biosynthesis from 1-deoxyxylulose in higher plants. 96(February):1309–1314, 1999. doi: 10.1073/pnas.96.4.1309.
- [8] T. J. Bach. Some New Aspects of Isoprenoid Biosynthesis in Plants - A Review. *Lipids*, 30(3):191–202, 1995.
- [9] D. Bartee and C. L. Freel Meyers. Targeting the unique mechanism of bacterial 1-deoxy-D-xylulose 5-phosphate (DXP) synthase. *Biochemistry*, 2018. ISSN 0006-2960. doi: 10.1021/acs.biochem.8b00548.
- [10] D. Bartee and C. L. Freel Meyers. Toward Understanding the Chemistry and Biology of 1-Deoxy- d -xylulose 5-Phosphate (DXP) Synthase: A Unique Antimicrobial Target at the Heart of Bacterial Metabolism. *Acc. Chem. Res.*, 51(Scheme 1):2546–2555, 2018. ISSN 15204898. doi: 10.1021/acs.accounts.8b00321.
- [11] D. Bartee, S. Sanders, P. D. Phillips, M. J. Harrison, A. T. Koppisch, and C. L. Freel Meyers. Enamide Prodrugs of Acetyl Phosphonate Deoxy- d -xylulose-5-phosphate Synthase Inhibitors as Potent Antibacterial Agents. *ACS Infect. Dis.*, 5(3):406–417, 2019. ISSN 23738227. doi: 10.1021/acsinfecdis.8b00307.
- [12] A. Bateman, M. J. Martin, S. Orchard, M. Magrane, R. Agivetova, S. Ahmad, E. Alpi, E. H. Bowler-Barnett, R. Britto, B. Bursteinas, H. Bye-A-Jee, R. Coetzee, A. Cukura, A. D. Silva, P. Denny, T. Dogan, T. G. Ebenezer, J. Fan, L. G. Castro, P. Garmiri, G. Georghiou,

- L. Gonzales, E. Hatton-Ellis, A. Hussein, A. Ignatchenko, G. Insana, R. Ishtiaq, P. Jokinen, V. Joshi, D. Jyothi, A. Lock, R. Lopez, A. Luciani, J. Luo, Y. Lussi, A. MacDougall, F. Madeira, M. Mahmoudy, M. Menchi, A. Mishra, K. Moulang, A. Nightingale, C. S. Oliveira, S. Pundir, G. Qi, S. Raj, D. Rice, M. R. Lopez, R. Saidi, J. Sampson, T. Sawford, E. Speretta, E. Turner, N. Tyagi, P. Vasudev, V. Volynkin, K. Warner, X. Watkins, R. Zaru, H. Zellner, A. Bridge, S. Poux, N. Redaschi, L. Aimo, G. Argoud-Puy, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, M. C. Blatter, J. Bolleman, E. Boutet, L. Breuza, C. Casals-Casas, E. de Castro, K. C. Echioukh, E. Coudert, B. Cucho, M. Doche, D. Dornevil, A. Estreicher, M. L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, N. Hyka-Nouspikel, F. Jungo, G. Keller, A. Kerhornou, V. Lara, P. Le Mercier, D. Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. B. Neto, S. Paesano, I. Pedruzzi, S. Pilbout, L. Pourcel, M. Pozzato, M. Pruess, C. Rivoire, C. Sigrist, K. Sonesson, A. Stutz, S. Sundaram, M. Tognolli, L. Verbregue, C. H. Wu, C. N. Arighi, L. Arminski, C. Chen, Y. Chen, J. S. Garavelli, H. Huang, K. Laiho, P. McGarvey, D. A. Natale, K. Ross, C. R. Vinayaka, Q. Wang, Y. Wang, L. S. Yeh, and J. Zhang. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.*, 49(D1):D480–D489, 2021. ISSN 13624962. doi: 10.1093/nar/gkaa1100.
- [13] K. Bloch. Sterol molecule : structure , biosynthesis , and function. *Steroids*, 57:378 – 383, 1992. doi: 10.1016/0039-128X(92)90081-J.
- [14] Y. Boucher and W. F. Doolittle. The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Mol. Microbiol.*, 37(4):703–716, 2000.
- [15] L. A. Brammer and C. F. Meyers. Revealing Substrate Promiscuity of 1-Deoxy-D-xylulose 5-Phosphate Synthase. 2009.
- [16] L. A. Brammer, J. M. Smith, H. Wades, and C. F. Meyers. 1-Deoxy-D-xylulose 5-phosphate synthase catalyzes a novel random sequential mechanism. *J. Biol. Chem.*, 286(42):36522–36531, 2011. ISSN 00219258. doi: 10.1074/jbc.M111.259747.
- [17] C. ChEMBL2285355, ChEMBL3559641. ChEMBL, 2021. URL [https://www.ebi.ac.uk/chembl/target\\_{\\_}report\\_{\\_}card/](https://www.ebi.ac.uk/chembl/target_{_}report_{_}card/).
- [18] P. Y. T. Chen, A. A. DeColli, C. L. Freel Meyers, and C. L. Drennan. X-ray crystallography-based structural elucidation of enzyme-bound intermediates along the 1-deoxy-D-xylulose 5-phosphate synthase reaction coordinate. *J. Biol. Chem.*, 294(33):12405–12414, 2019. ISSN 1083351X. doi: 10.1074/jbc.RA119.009321.
- [19] C. Courtens, M. Risseuw, G. Caljon, P. Cos, A. Martin, and S. Van Calenbergh. Phosphonodiamidate prodrugs of N-alkoxy analogs of a fosmidomycin surrogate as antimalarial and antitubercular agents. *Bioorganic Med. Chem. Lett.*, 29(9):1051–1053, 2019. ISSN 14643405. doi: 10.1016/j.bmcl.2019.03.008. URL <https://doi.org/10.1016/j.bmcl.2019.03.008>.
- [20] A. M. Davis, S. J. Teague, and G. J. Kleywegt. Application and limitations of x-ray crystallographic data in structure-based ligand and drug design. *Angew. Chemie - Int. Ed.*, 42(24):2718–2736, 2003. ISSN 14337851. doi: 10.1002/anie.200200539.
- [21] A. A. DeColli, N. S. Nemeria, A. Majumdar, G. J. Gerfen, F. Jordan, and C. L. Freel Meyers. Oxidative decarboxylation of pyruvate by 1-deoxy-D-xylulose 5-phosphate synthase, a central metabolic enzyme in bacteria. *J. Biol. Chem.*, 293(28):10857–10869, 2018. ISSN 1083351X. doi: 10.1074/jbc.RA118.001980.
- [22] A. Endo. The discovery and development of HMG-CoA reductase inhibitors. *J. Lipid Res.*, 33(11):1569–1582, 1992. ISSN 0022-2275. doi: 10.1016/S0022-2275(20)41379-3. URL [http://dx.doi.org/10.1016/S0022-2275\(20\)41379-3](http://dx.doi.org/10.1016/S0022-2275(20)41379-3).

- [23] A. ENDO. Review A historical perspective on the discovery of statins. *Proc. Japan Acad. Ser. B*, 86:484–493, 2010. doi: 10.2183/pjab.86.484.
- [24] J. M. Estévez, A. Cantero, A. Reindl, S. Reichler, and P. León. 1-Deoxy-D-xylulose-5-phosphate Synthase, a Limiting Enzyme for Plastidic Isoprenoid Biosynthesis in Plants. *J. Biol. Chem.*, 276(25):22901–22909, 2001. ISSN 00219258. doi: 10.1074/jbc.M100854200.
- [25] L. M. Eubanks and C. D. Poulter. Rhodobacter capsulatus 1-Deoxy-D-xylulose 5-Phosphate Synthase : Steady-State Kinetics and Substrate Binding. *Biochemistry*, (42):1140–1149, 2003.
- [26] R. A. W. Frank, F. J. Leeper, and B. F. Luisi. Review Structure, mechanism and catalytic duality of thiamine-dependent enzymes. *Cell. Mol. Life Sci*, 64:892–905, 2007. doi: 10.1007/s00018-007-6423-5.
- [27] L. M. Gomez. *Structure-based design of ligands for vitamin transporters in bacteria*. PhD thesis, University of Groningen, 2016.
- [28] A. M. Goswami. Computational analysis, structural modeling and ligand binding site prediction of Plasmodium falciparum 1-deoxy-D-xylulose-5-phosphate synthase. *Comput. Biol. Chem.*, 66:1–10, 2017. ISSN 14769271. doi: 10.1016/j.compbiolchem.2016.10.010. URL <http://dx.doi.org/10.1016/j.compbiolchem.2016.10.010>.
- [29] F. M. Hahn, L. M. Eubanks, C. a. Testa, S. J. Blagg, J. a. Baker, and C. D. Poulter. 1-Deoxy-D-Xylulose 5-Phosphate Synthase , the Gene Product of Open Reading Frame ( ORF) 2816 and ORF 2895 in Rhodobacter capsulatus. 183(1):1–11, 2001. doi: 10.1128/JB.183.1.1.
- [30] S. Heuston, M. Begley, C. G. Gahan, and C. Hill. Isoprenoid biosynthesis in bacterial pathogens. *Microbiology*, 158:1389–1401, 2012. doi: 10.1099/mic.0.051599-0.
- [31] A. K. Hirsch and F. Diederich. The non-mevalonate pathway to isoprenoid biosynthesis: A potential source of new drug targets. *Chimia (Aarau)*, 62(4):226–230, 2008. ISSN 00094293. doi: 10.2533/chimia.2008.226.
- [32] W. N. Hunter. The non-mevalonate pathway of isoprenoid precursor biosynthesis. *J. Biol. Chem.*, 282(30):21573–21577, 2007. ISSN 00219258. doi: 10.1074/jbc.R700005200.
- [33] M. ITN:. Exploiting the methylerythritol phosphate pathway as a source of drug targets for novel anti-infectives. <https://cordis.europa.eu/project/id/860816>, 2020. URL <https://cordis.europa.eu/project/id/860816>.
- [34] H. Jomaa, J. Wiesner, S. Sanderbrand, B. Altincicek, C. Weidemeyer, M. Hintz, I. Türbachova, M. Eberl, J. Zeidler, H. K. Lichtenthaler, D. Soldati, and E. Beck. Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science (80-. )*, 285(5433):1573–1576, 1999. ISSN 00368075. doi: 10.1126/science.285.5433.1573.
- [35] P. S. Karunanithi and P. Zerbe. Terpene Synthases as Metabolic Gatekeepers in the Evolution of Plant Terpenoid Chemical Diversity. *Front. Plant Sci.*, 10(October):1–23, 2019. doi: 10.3389/fpls.2019.01166.
- [36] Y. KURODA, M. OKUHARA, T. GOTO, M. OKAMOTO, H. TERANO, M. KOHSAKA, H. AOKI, and H. IMANAKA. STUDIES ON NEW PHOSPHONIC ACID ANTIBIOTICS. *J. Antibiot. (Tokyo)*, XXXIII(1):29–35, 1979.
- [37] T. Kuzuyama, T. Shimizu, S. Takahashi, and H. Seto. Fosmidomycin, a Specific Inhibitor of 1-Deoxy-D-Xylulose 5-Phosphate Reductoisomerase in the Nonmevalonate Pathway for Terpenoid Biosynthesis. *Tetrahedron Lett.*, 39:7913–7916, 1998.

- [38] T. Kuzuyama, M. Takagi, S. Takahashi, and H. Seto. Cloning and Characterization of 1-Deoxy- D -Xylulose 5- Phosphate Synthase from *Streptomyces* sp . Strain CL190 , Which Uses both the Mevalonate and Nonmevalonate Pathways for Isopentenyl Diphosphate Biosynthesis. *J. Bacteriol.*, 182(4):891–897, 2000.
- [39] R. Li, P. Chen, L. Zhu, F. Wu, Y. Chen, P. Zhu, and K. Ji. Characterization and Function of the 1-Deoxy-D-xylose-5-Phosphate Synthase (DXS) Gene Related to Terpenoid Synthesis in *Pinus massoniana*. *Int. J. Mol. Sci.*, 22(2):848, 2021. ISSN 14220067. doi: 10.3390/ijms22020848.
- [40] P.-h. Liang, T.-p. Ko, and A. H. Wang. Structure , mechanism and function of prenyltransferases. *Eur. J. Biochem.*, 3354(May):3339–3354, 2002. doi: 10.1046/j.1432-1033.2002.03014.x.
- [41] Y. F. Liang, H. Liu, H. Li, and W. Y. Gao. Determination of the Activity of 1-Deoxy-d-Xylulose 5-Phosphate Synthase by Pre-column Derivatization-HPLC Using 1,2-Diamino-4,5-Methylenedioxybenzene as a Derivatizing Reagent. *Protein J.*, 38(2):160–166, 2019. ISSN 15734943. doi: 10.1007/s10930-019-09816-9. URL <http://dx.doi.org/10.1007/s10930-019-09816-9>.
- [42] X. Lv, J. Gu, F. Wang, W. Xie, M. Liu, L. Ye, and H. Yu. Combinatorial pathway optimization in *Escherichia coli* by directed co-evolution of rate-limiting enzymes and modular pathway engineering. *Biotechnol. Bioeng.*, pages 1–29, 2016. doi: 10.1002/bit.26034.
- [43] B. Markus Lange, M. R. Wildung, D. Mccaskill, and R. Croteau. A family of transketolases that directs isoprenoid biosynthesis via a mevalonate-independent pathway. *Proc. Natl. Acad. Sci. U. S. A.*, 95(5):2100–2104, 1998. ISSN 00278424. doi: 10.1073/pnas.95.5.2100.
- [44] T. Masini. *The enzyme DXS as an anti-infective target*. PhD thesis, University of Groningen, 2015.
- [45] T. Masini, B. Lacy, L. Monjas, D. Hawksley, A. R. de Voogd, B. Illarionov, A. Iqbal, F. J. Leeper, M. Fischer, M. Kontoyianni, and A. K. H. Hirsch. Validation of a homology model of *Mycobacterium tuberculosis* DXS: rationalization of observed activities of thiamine derivatives as potent inhibitors of two orthologues of DXS. *Org. Biomol. Chem.*, 13:11263–11277, 2015. ISSN 1477-0520. doi: 10.1039/C5OB01666E. URL [www.rsc.org/obchttp://xlink.rsc.org/?DOI=C5OB01666E](http://www.rsc.org/obchttp://xlink.rsc.org/?DOI=C5OB01666E).
- [46] Y. Matsue, H. Mizuno, T. Tomita, T. Asami, M. Nishiyama, and T. Kuzuyama. The herbicide ketoclozomazone inhibits 1-deoxy-D-xylulose 5-phosphate synthase in the 2-C-methyl-D-erythritol 4-phosphate pathway and shows antibacterial activity against *Haemophilus influenzae*. *J. Antibiot. (Tokyo)*, 63(10):583–588, 2010. ISSN 00218820. doi: 10.1038/ja.2010.100.
- [47] J. McMurry. *Organic Chemistry with Biological Applications 2e*. Mary Finch, 2 edition, 2010.
- [48] E. Oldfield and F.-y. Lin. Terpene Biosynthesis : Modularity Rules. *Angew. Chem. Int. Ed.*, 51:1124–1137, 2012. doi: 10.1002/anie.201103110.
- [49] S. Pandian, S. Saengchjant, and T. S. Raman. An alternative pathway for the biosynthesis of isoprenoid compounds in bacteria. *Biochem. J.*, 196:675–681, 1981.
- [50] S. R. Putra, A. Disch, J.-m. Bravo, and M. Rohmer. Distribution of mevalonate and glyceraldehyde 3-phosphate / pyruvate routes for isoprenoid biosynthesis in some Gram-negative bacteria and mycobacteria. *FEMS Microbiol. Lett.*, 164:169–175, 1998.

- [51] J. Querol, O. Besumbes, L. M. Lois, A. Boronat, and S. Imperial. A fluorometric assay for the determination of 1-deoxy-D-xylulose 5-phosphate synthase activity. *Anal Biochem*, 296(1):101–105, 2001. ISSN 0003-2697. doi: 10.1006/abio.2001.5234. URL <http://www.ncbi.nlm.nih.gov/pubmed/11520037>.
- [52] F. Rohdich, A. Bacher, and W. Eisenreich. Perspectives in anti-infective drug design. The late steps in the biosynthesis of the universal terpenoid precursors, isopentenyl diphosphate and dimethylallyl diphosphate. *Bioorg. Chem.*, 32(5):292–308, 2004. ISSN 00452068. doi: 10.1016/j.bioorg.2004.05.012.
- [53] M. Rohmer. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep.*, 16(5):565–574, 1999. ISSN 02650568. doi: 10.1039/a709175c.
- [54] M. Rohmer, M. Knani, P. Simonin, B. Sutter, and H. Sahm. Isoprenoid biosynthesis in bacteria: A novel pathway for the early steps leading to isopentenyl diphosphate. *Biochem. J.*, 295(2):517–524, 1993. ISSN 02646021. doi: 10.1042/bj2950517.
- [55] M. Rohmer, C. Grosdemange, M. Seemann, and D. Tritsch. Isoprenoid biosynthesis as a novel target for antibacterial and antiparasitic drugs. *Curr. Opin. Investig. drugs*, pages 154–162, 2004.
- [56] M. D. Rosa, S. D. Rosa, A. Gambacorta, L. Minale, R. H. Thomson, and R. D. Worthington. Caldariellaquinone, a Unique Benzo[b]thiophen-4,7-quinone from *Caldariella acidophila*, an Extremely Thermophilic and Acidophilic Bacterium. *J. Chem. Soc.*, pages 653–657, 1977. doi: 10.1039/P19770000653.
- [57] R. Ruangweerayut, S. Looareesuwan, D. Hutchinson, A. Chauemung, V. Banmairuroi, and K. Na-bangchang. Assessment of the pharmacokinetics and dynamics of two combination regimens of fosmidomycin-clindamycin in patients with acute uncomplicated falciparum malaria. *Malar. J.*, 11:1–11, 2008. doi: 10.1186/1475-2875-7-225.
- [58] E. Rubio-Rodríguez, I. Vera-Reyes, E. B. Sepúlveda-García, A. C. Ramos-Valdivia, and G. Trejo-Tapia. Secondary metabolite production and related biosynthetic genes expression in response to methyl jasmonate in *Castilleja tenuiflora* Benth. *in vitro* plants. *Plant Cell. Tissue Organ Cult.*, (0123456789), 2021. ISSN 15735044. doi: 10.1007/s11240-020-01975-3. URL <https://doi.org/10.1007/s11240-020-01975-3>.
- [59] A. Runthala, T. H. Sai, V. Kamjula, S. C. Phulara, V. S. Rajput, and K. Sangapillai. Excavating the functionally crucial active-site residues of the DXS protein of *Bacillus subtilis* by exploring its closest homologues. *J. Genet. Eng. Biotechnol.*, 18(1), 2020. ISSN 20905920. doi: 10.1186/s43141-020-00087-x.
- [60] S. Sanders, R. J. Vierling, D. Bartee, A. A. Decolli, M. J. Harrison, J. L. Aklinski, A. T. Koppisch, and C. L. Freel Meyers. Challenges and Hallmarks of Establishing Alkylacetylphosphonates as Probes of Bacterial 1-Deoxy- d -xylulose 5-Phosphate Synthase. *ACS Infect. Dis.*, 3(7):467–478, 2017. ISSN 23738227. doi: 10.1021/acsinfecdis.6b00168.
- [61] M. K. Schwarz. *Terpen-Biosynthese in Ginkgo biloba: eine überraschende Geschichte*. PhD thesis, 1994.
- [62] S. Shaw and R. Ghosh. A modified kulka micromethod for the rapid and safe analysis of fructose and 1-deoxy-d-xylulose-5-phosphate. *Metabolites*, 8(4), 2018. ISSN 22181989. doi: 10.3390/metabo8040077.
- [63] Sigma-Aldrich. Pyruvate Assay Kit, 2021. URL <https://www.sigmaaldrich.com/catalog/product/sigma/mak071?lang=de&region=DE>.

- [64] J. M. Smith, R. J. Vierling, and C. F. Meyers. Selective inhibition of *E. coli* 1-deoxy-D-xylulose-5-phosphate synthase by acetylphosphonates. *Medchemcomm*, 3:65–67, 2012. doi: 10.1039/c1md00233c.
- [65] M. Srinath, A. Shailaja, B. B. V. Bindu, and C. C. Giri. Molecular Cloning and Differential Gene Expression Analysis of 1-Deoxy-D-xylulose 5-Phosphate Synthase (DXS) in *Andrographis paniculata* (Burm. f) Nees. *Mol. Biotechnol.*, 63(2):109–124, 2020. ISSN 15590305. doi: 10.1007/s12033-020-00287-3. URL <https://doi.org/10.1007/s12033-020-00287-3>.
- [66] P. A. Tavormina and M. H. Gibbs. THE METABOLISM OF  $\beta,\gamma$ -DIHYDROXY- $\beta$ -METHYLVALERIC ACID BY LIVER HOMOGENATES. *J. Am. Chem. Soc.*, 78(22): 6210, 1956.
- [67] J.-Y. van der Meer and A. K. H. Hirsch. The isoprenoid-precursor dependence of *Plasmodium* spp. *Nat. Prod. Rep.*, 29:721, 2012. ISSN 0265-0568. doi: 10.1039/c2np20013a.
- [68] R. Verpoorte, R. V. D. Heijden, H. J. G. Hoopen, and J. Memelink. Metabolic engineering of plant secondary metabolite pathways for the production of fine chemicals. *Biotechnol. Lett.*, (Verpoorte):467–479, 1999.
- [69] X. Wang and C. S. Dowd. The Methylerythritol Phosphate Pathway: Promising Drug Targets in the Fight against Tuberculosis. *ACS Infect. Dis.*, 4(3):278–290, 2018. ISSN 23738227. doi: 10.1021/acsinfecdis.7b00176.
- [70] G. L. Warren, T. D. Do, B. P. Kelley, A. Nicholls, and S. D. Warren. Essential considerations for using protein-ligand structures in drug discovery. *Drug Discov. Today*, 17(23-24):1270–1281, 2012. ISSN 13596446. doi: 10.1016/j.drudis.2012.06.011. URL <http://dx.doi.org/10.1016/j.drudis.2012.06.011>.
- [71] H. Wei, A. Movahedi, C. Xu, W. Sun, A. A. Z. Yaghuti, P. Wang, D. Li, and Q. Zhuge. Overexpression of ptDXS enhances stress resistance in poplars. *Int. J. Mol. Sci.*, 20(7):1–21, 2019. ISSN 14220067. doi: 10.3390/ijms20071669.
- [72] J. K. White, S. Handa, S. L. Vankayala, D. J. Merkler, and H. L. Woodcock. Thiamin Diphosphate Activation in 1-Deoxy-D-Xylulose 5-Phosphate Synthase: Insights into the Mechanism and Underlying Intermolecular Interactions. *J. Phys. Chem. B*, page acs.jpbc.6b07248, 2016. ISSN 1520-6106. doi: 10.1021/acs.jpbc.6b07248. URL <http://pubs.acs.org/doi/abs/10.1021/acs.jpbc.6b07248>.
- [73] D. E. Wolf, C. H. Hoffman, P. E. Aldrich, H. R. Skeggs, L. D. Wright, and K. Folkers.  $\beta$ -HYDROXY- $\beta$ -METHYL- $\delta$ -VALEROLACTONE (DIVALONIC ACID), A NEW BIOLOGICAL FACTOR. *J. Am. Chem. Soc.*, 78(17):4498, 1956.
- [74] L. P. Wright, J. M. Rohwer, A. Ghirardo, A. Hammerbacher, M. Ortiz-alcaide, B. Raguschke, J.-p. Schnitzler, J. Gershenson, M. A. Phillips, M. Plank, C. Ecology, and L. P. W. Germany. Deoxyxylulose 5-Phosphate Synthase Controls Flux through the Methylerythritol 4-Phosphate. *Plant Physiol.*, 165(August):1488–1504, 2014. doi: 10.1104/pp.114.245191.
- [75] S. Xiang, G. Usunow, G. Lange, M. Busch, and L. Tong. Crystal Structure of 1-Deoxy-D-xylulose 5-Phosphate Synthase, a Crucial Enzyme for Isoprenoids Biosynthesis. *J. Biol. Chem.*, 282(4):2676–2682, jan 2007. ISSN 0021-9258. doi: 10.1074/jbc.M610235200. URL <http://www.ncbi.nlm.nih.gov/pubmed/17135236>.
- [76] C. Xu, H. Wei, A. Movahedi, W. Sun, X. Ma, D. Li, T. Yin, and Q. Zhuge. Evaluation, characterization, expression profiling, and functional analysis of DXS and DXR genes of *Populus trichocarpa*. *Plant Physiol. Biochem.*, 142(June):94–105, 2019. ISSN 09819428. doi: 10.1016/j.plaphy.2019.05.034. URL <https://doi.org/10.1016/j.plaphy.2019.05.034>.

- [77] J. Zhou, L. Yang, A. DeColli, C. Freel Meyers, N. S. Nemeria, and F. Jordan. Conformational dynamics of 1-deoxy-d-xylulose 5-phosphate synthase on ligand binding revealed by H/D exchange MS. *Proc. Natl. Acad. Sci.*, 114(35):9355–9360, 2017. ISSN 0027-8424. doi: 10.1073/pnas.1619981114. URL <http://www.pnas.org/lookup/doi/10.1073/pnas.1619981114>.



# Acknowledgments

First I would like to thank you, Anna, for accepting me in your group and being my PhD supervisor. It was a pleasure to work with you and I am thankful I could still experience some time in Groningen and also share the exciting move to our new lab at the HIPS.

Many thanks to my co-supervisor Matthew Groves for his supervision, our two day intensive course in crystallography and the weekly discussions at the University Medical Center Groningen (UMCG).

A big part of the last five years was shaped by all the people I met in the lab. You all were present during the successful and fun moments in research, but also shared the failures. For this I would like to thank the whole DDOP and CBCH group at the HIPS institute, as well as the chemical biology group from Adriaan Minnaard, the structural biology group from Matthew Groves at the UMCG as well as the Dijkstra X-ray crystallography group at the RUG.

The time that I worked at the University of Groningen was short, but intense. I would like to thank the whole Hirsch-Group for their warm welcome, my office-mates Ramon, Steven and Ji for all the help and lively discussions as well as Eswar and Nils, who introduced me into the crystallography lab. I am particularly happy to have met Alwin, who not only is a great coworker (and helped me with the Dutch summary) but also always a pleasure to spend time with.

The group of people that shared a common enemy with me (DXS) and worked to find inhibitors for it grew over time to a respectable DXS-Team. Thank you Ravi, Sandra, Eswar, Alaa, Zhoor, Melissa, Tiziana and Judy for our good teamwork and the exciting discussions.

I also own a big thank you to my Master's Students Alaa, Harry and Dominik that joined me in working on my projects. It was nice to have someone to share the long days in the lab with and discuss our ideas. I hope you all will achieve your goals.

Although not every piece of our collaborations made it into this thesis, I would like to thank all my collaboration partners. I am very positive about the number of helpful collaborations that are possible and very uncomplicated in academia. Thanks to Dr. Boris Illarionov, University of Hamburg, for supplying us with a start stock of enzyme and helping with assay guidance. Thanks to Prof. Carsten Wrenger, Arne and Camila, University of São Paulo, for creating a transgenic *P. falciparum* cell line. Thanks to Rachel L. Edwards and Prof. Audrey R. Odom John, University of Pennsylvania, for performing target validation experiments. Thanks to PD Dr. Michael Hutter, University of Saarbrücken, for the bioinformatic studies. Thanks to Tessa and Christian, Atomwise, for performing an AI-based virtual screening. Thanks to Prof. Jana Held, University of Tübingen, for measuring our compounds on *P. falciparum* and thanks to Prof. Norbert Reiling, Leibniz Lungenzentrum Borstel, for testing our compounds on *M. tuberculosis*. A big thank you to Prof. Klebe for allowing me to work in his lab at the University of Marburg with Francesca. The two weeks at the beginning of my PhD time in this experienced and perfectly settled medicinal chemistry lab were very inspiring.

One of the best places I found in Saarbrücken was the whole house of my WG. A big thank you to all of my former flatmates that quickly became friends: Aza and Wolle, Johanna and Tyler, Chrissie, Frankie, Kira, Kevin, Lennart, Natalie, Jan, Marianne and Bernd. I will remember all of

our discussions, games and parties. May the WG continue to exist for another 24 years!

If I was not in the WG or lab, I was in the climbing gym with Eike, Lennart, Jan and Varvara. Thank you for many hours of excitement and for all your motivation, Eike!

All my friends from my hometown Schermbeck are a constant source of energy for me. Thanks to Stephan, Moritz, Christoph, Jonas, Jojo, Steffi, Nils, Lisa, Tine, Maria, Verena, Martin, Ruth, Diana, Laura and everyone I forgot for still sticking together, maybe more than ever before.

Nico, thanks for being my friend and mentor during my studies and for encouraging me to search for a PhD position abroad. Niklas, you always know the answer for practical questions regarding crystallography, in particular hot-keys. Thanks for the conferences we enjoyed together and for sharing some of your excitement for protein crystallography with me.

Although it comes at the end of this part, it might be the most important one. My family. A very big and warm Thank You to my parents Elfi and Dieter for unconditional support and a similar sized Thanks to Claudia for starting the adventure of founding a new family with me.

And thank you, Flora, for giving me some extra motivation to finish this thesis.



