Probing biological activities of RNA-binding proteins from bacterial and mammalian cells

Dissertation

Zur Erlangung des Grades des Doktors der Naturwissenschaften
der Naturwissenschaftlich-Technischen Fakultät
der Universität des Saarlandes

von

M.Sc. Yingwen Wu

Saarbrücken
2023

Die vorliegende Arbeit wurde von September 2019 bis August 2023 unter Anleitung von PD. Dr. Martin Empting in der Fachrichtung Pharmazeutische und Medizinische Chemie der Naturwissenschaftlich-Technischen Fakultät der Universität des Saarlandes sowie am Helmholtz-Institut für Pharmazeutische Forschung Saarland (HIPS) in der Abteilung Antiviral and Antivirulence Drugs (AVID) angefertigt.

„This book, therefore, ends not with answers but with a wish. I want to love

and be loved. I want to find a way where I don't hurt myself. I want to live a

life where I say things are good more than things are bad. I want to keep

failing and discovering new and better directions.[...] Some day, I will."

"I Want to Die but I Want to Eat Tteokbokki" - Baek Sehee

Acknowledgements

Summary

The growing danger of antimicrobial resistance and the ongoing fight against cancer present an immense challenge for public health. The need for new and innovative treatments is evident in both infection research and cancer therapy.

This doctoral thesis focuses on two RNA-binding proteins as potential drug targets: The Carbon Storage Regulator System A (CsrA) in bacterial systems and the insulin-like growth factor 2 (IGF2) mRNA binding proteins (IGF2BPs/IMPs) in mammalian systems.

Within this work, the reasons behind selecting both proteins as promising targets for anti-virulence/cancer therapy are explained. Furthermore, the process of identifying of novel inhibitors against these targets through biophysical screening methods, followed by hit prioritization, medicinal chemistry optimization and validation via cell-based assays is described. The aim is to discover cell-active compounds suitable for initial proof-of-concept studies.

Zusammenfassung

Die zunehmende Bedrohung durch antimikrobielle Resistenz und der fortlaufende Kampf gegen Krebs stellen immense Herausforderungen für das globale Gesundheitswesen dar. Die Dringlichkeit für neue und innovative Behandlungsmöglichkeiten zeigt sich sowohl in der Infektionsforschung als auch in der Krebstherapie.

Diese Doktorarbeit konzentriert sich auf zwei RNA-bindende Proteine als Wirkstoffziele: Das Carbon Storage Regulator System A (CsrA) in bakteriellen Systemen und das Insulin-like growth factor 2 (IGF2) mRNA-bindende Proteine(IGF2BPs/IMPs) in Säugetiersystemen.

Im Rahmen dieser Arbeit wird die Auswahl beider Proteine als vielversprechende Targets für Anti-Virulenz-/Krebstherapien erläutert. Des Weiteren wird der Prozess der Identifizierung neuartiger Inhibitoren gegen diese Proteine durch biophysikalische Screening-Methoden beschrieben, gefolgt von der Priorisierung der Hitverbindungen, der Optimierung mittels medizinischen Chemie und der Validierung über zellbasierte Assays. Das Ziel besteht darin, zellaktive Verbindungen zu finden, die sich für anfängliche Proof-of-Concept-Studien eignen.

Publication included in this Thesis

Publication 1: V. Jakob[*], B. G. E. Zoller*, J. Rinkes, Y. Wu, A. F. Kiefer, M. Hust, S. Polten, A. M. White, P. J. Harvey, T. Durek, D. J. Craik, A. Siebert, U. Kazmaier, M. Empting, *Phage display-based discovery of cyclic peptides against the broad spectrum bacterial anti-virulence target CsrA, Eur. J. Med. Chem.* 2022 https://doi.org/10.1016/j.ejmech.2022.114148

* These authors contributed equally

Publication 2: Y. Wu, B. G.E. Zoller, M. A. M. Kamal, S.-K. Hotop, C.-M.Lehr, M. Brönstrup, P. Dersch, M. Empting, *Establishment of an In Bacterio assay for the Assessment of Carbon Storage Regulator A (CsrA) inhibitors, ChembioChem* 2023 https://doi.org/10.1002/cbic.202300369

Publication 3: C. Dahlem*, A. Abuhaliema*, S. M. Kessler*, T. Kröhler, B. G. E. Zoller, S. Chanda, Y. Wu, S. Both, F. Müller, K. Lepikhov, S. H. Kirsch, S. Laggai, R. Müller, M. Empting, A. K. Kiemer, *First Small-Molecule Inhibitors Targeting the RNA-Binding Protein IGF2BP2/IMP2 for Cancer Therapy, ACS Chem. Biol.* 2022 https://doi.org/10.1021/acschembio.1c00833

* These authors contributed equally.

Publication 4: C. E. Cunningham*, F. S. Vizeacoumar*, Y. Zhang*, L. Kyrylenko*, P. Gao, V. Maranda, A. Ganapathysamy, H. Dong, S. Both, K. Wagner, Y. Wu, J. Price, T. Katrii, B. G. E. Zoller, K. Rajamanickam, P. Walke, L. Hui, H. Patel, A. Morejon Morales, K. K. Bhanumathy, H. Elhasasna, R. Dahiya, O. Abuhussein, T. Freywald, E. Prando Munhoz, A. Krishnan, E. Ruppin, , J. Sang Lee, K. Rox, B. Toosi, R. Datla, S. Yadav, Y. Wu, M. Empting, A, K. Kiemer, A. Freywald, F. J. Vizeacoumar, *Integration of in vivo pooled CRISPR and direct capture Perturb-seq screens identifies synthetic dosage lethal interactions of Polo-Like Kinase 1 in cancer cells* Manuscript in preparation

* These authors contributed equally.

Abbreviations

ADME	absorption, distribution, metabolism, excretion
AMR	antimicrobial resistance
Arg	Arginine
cAMP	Cyclic adenosine monophosphate
CF	Cystic Fibrosis
CFTR	cystic fibrosis transmembrane conductance regulator channel
CIN	chromosomal instability
CRD	coding region determinant
CsrA	Carbon Storage Regulator A
E. coli	Escherichia coli
EIIAGlu	glucose-specific enzyme IIA
EPEC	Enteropathogenic *E. coli*
FP	fluorescence polarization
HGT	horizontal gene transfer
hnRNP	Heterogeneous nuclear ribonucleoproteins
HTS	high-throughput screening
IGF2BPs/IMPs	insulin-like growth factor 2 mRNA binding proteins
IM	inner membrane
KO	knockouts
Leu	Leucine
LPS	lipopolysaccharides
MDR	multi-drug-resistant
mRNPs	target mRNAs distinct ribonucleoproteins
MST	microscale thermophoresis
mTORC1	mammalian target of rapamycin complex 1
NMR	nuclear magnetic resonance
nt	nucleotides
OM	outer membrane
PA	Pseudomonas aeruginosa
PK	pharmacokinetic
PLK1	polo-like kinase 1
PNPase	polynucleotide phosphorylase

PTS	phosphotransferase system
RBPs	RNA-binding proteins
RRMs	RNA recognition motif
RsmA	regulator of secondary metabolites
RT- qPCR	Reverse Transcription-quantitative Polymerase Chain Reaction
SAR	structure-activity relationship
SCFAs	short-chain fatty acids
STD	saturation transfer difference
T2D	type 2 diabetes
T3SS	the type three secretion system
TCS	two-component signal transduction system
TEV	tobacco etch virus
TNBC	triple-negative breast cancer
TSA	thermal shift assays
UTR	untranslated region
WHO	World Health Organization

Table of contents

I.   Chapter 1: Introduction

"The time may come when penicillin can be bought by anyone in the shops. Then there is the danger that the ignorant man may easily underdose himself and by exposing his microbes to non-lethal quantities of the drug make them resistant."[1]

78 years ago, Sir Alexander Fleming, who discovered the first antibiotic, warned the world in his Nobel lecture about the threat of antibiotic resistance. As he predicted, the misuse and overuse of antibiotics have contributed to the development of antibiotic resistance, which arose to be a global concern.[2] To acquire antibiotic resistance, bacteria mostly go through various genetic mutations that occurs naturally over time.[3,4] Mutations typically arises in several categories of genes. Firstly, they can occur in genes encoding the targets of the antibiotics themselves, leading to modifications such as overproduction of the targets.[5,6] Secondly, mutations can affect genes responsible for encoding transporters that facilitate the entry or efflux of antibiotics.[5] Moreover, modulation of the membrane permeability can be achieved through mutations.[6] Finally, mutations can affect genes encoding regulators that normally suppress the expression of antibiotic-modifying enzymes, including chromosomally-encoded antibiotic-modifying enzymes and multidrug efflux pumps.[5]

Another way to acquire antibiotic resistance is through the horizontal gene transfer (HGT), where resistance genes are transferred between different organisms.[3] The acquisition of the genetic material occurs normally using three main mechanisms: Transformation, transduction and conjugation.[3,7]

In case of transformation, bacteria incorporate naked DNA from the environment in order to repair the DNA or using genetic diversification to increase the adaptability.[7] The antibiotic-resistant strains of *Streptococcus spp.* for example used transformation for their evolution.[7] Transduction is based on phage-mediated gene transfer and might be involved in resistance development of *Staphylococcus aureus*. However, transduction and transformation mechanisms are difficult to detect outside of laboratory environments.[7]

The most widespread mechanism for prevalence of resistant genes is plasmid-mediated conjugation.[7] Gene transfer using conjugation occurs in high-density settings such as biofilms, co-infections conditions and the human or animal gut.[7] The

plasmid is able to autonomously replicate and transfer genes to the host.[7] Therefore plasmid-mediated conjugation leads to the global spread of resistant determinants.[7] Thus, it allows bacteria to acquire resistance against multiple antibiotics.

In addition to the mechanisms mentioned above, there is a so-called intrinsic antibiotic resistance, which is independent of any prior exposure to antibiotics. This resistance refers to genetic traits inherent within the genome of certain bacterial species. These traits make them naturally less susceptible to certain antibiotics.[4,7] The mechanisms of intrinsic resistance are usually chromosome-encoded, for example in response to environmental toxins, non-specific active efflux pumps such as AcrAB/TolC pumps in *Escherichia coli* (*E. coli*) have actively evolved.[7] Additionally, limited outer membrane permeability, such as the case of vancomycin resistance in *E. coli*, is also a common mechanism of intrinsic antibiotic resistance. While intrinsic resistance is not typically considered a mutation, it can still be subject to evolutionary changes over time.[4,5,7] Genetic variations and selection pressures may influence the level of intrinsic resistance in bacterial populations.[5,7]

Nowadays, antimicrobial resistance (AMR) becomes more and more a threat for the public health because the resistance is against all types of antimicrobial drugs including antibiotics, antivirals, antifungals and antiparasitics.[2,8] According to the World Health Organization (WHO) AMR is one of the top 10 global public health and development threat.[2] On the one hand, the risks of AMR include the limitation or complete lack of treatment options. On the other hand, there is the accelerated emergence and spread of AMR due to factors such as the lack of access to appropriate hygiene facilities, including clean water, as well as the absence of strong infection and disease prevention measures. This so-called "silent pandemic" is not as visible as the COVID-19 pandemic we had between 2019-2022 but the consequences might be as dramatic or even worse. Based on estimation statistics, without any counteractions AMR would cause more than 10 million deaths per year globally by 2050.[7–9]

1. Polymicrobial infections

In general, a healthy human microbiome comprises a rich and diverse community of organisms, exhibiting stability and resilience against pathogenic species. To suppress the colonization by pathogenic species, commensal bacteria produce their own

antimicrobial peptides as a defense mechanism.[10] Conversely, dysbiotic communities, that are commonly associated with disease progression, exhibit a reduced diversity and are more dominated by a limited number of pathogenic species. Co-infection involving pathogenic species leads to enhanced virulence, alteration of infected niche through nutrient availability and other factors that affect the microbial growth; or modulation of the host immune response such as enhanced tolerance to immune radicals, antibiotics or other toxins.[10] These outcomes contribute to the emergence of polymicrobial infections.[10,11]

In polymicrobial infections, a prevailing pathogen often enhances the virulence of co-infecting microbes through various mechanisms, such as inhibiting competing microbes. This is known as microbial interference, facilitating the provision, and sharing of nutrients, such as carbon sources or subverting immunity. In other words, the specific interactions between pathogenic species, the environment and the host organism have a significant influence on the outcome of the disease.[11,12]

Polymicrobial infections can occur in various parts of the body, including the respiratory tract, urinary tract, skin, and wounds. In the following section, the focus will be given to the respiratory and gastrointestinal associated infections.

1.1 Respiratory and gastrointestinal infections

Polymicrobial infections mediate acute and chronic diseases, which are associated with many respiratory and gastrointestinal infections. In the case of respiratory tract viruses, they can destroy the epithelium, which increases the bacterial adhesion and induces immunosuppression that leads to bacterial superinfections.[11] An example of such a condition is Cystic Fibrosis (CF), an autosomal recessive disorder primarily affecting the respiratory system (Figure 1).[13] A mutation of cystic fibrosis transmembrane conductance regulator channel (CFTR), which regulates chloride and bicarbonate transport, leads to the production of thick and sticky mucus.[14,15] This mucus buildup creates an environment in which bacteria and other microorganisms can survive. Thus, CF patients are susceptible to polymicrobial infections due to compromised immune responses, which makes them suffer from repeated infections. These CF-associated infections are facilitated by formation of biofilms altering the pH of the airway surface liquid. Some of the pathogens responsible for these infections are

Staphylocooccus aureus, *Haemophilus influenza* and *Pseudomonas aeruginosa (PA)*.[10,14] PA is responsible for the transition of acute respiratory infections into chronic infections depending on the age of the patients. This pathogen along with another gram-negative bacterium *Burkholderia cepacia* strongly shape and dominate the microbial communities in a CF lung.[14]
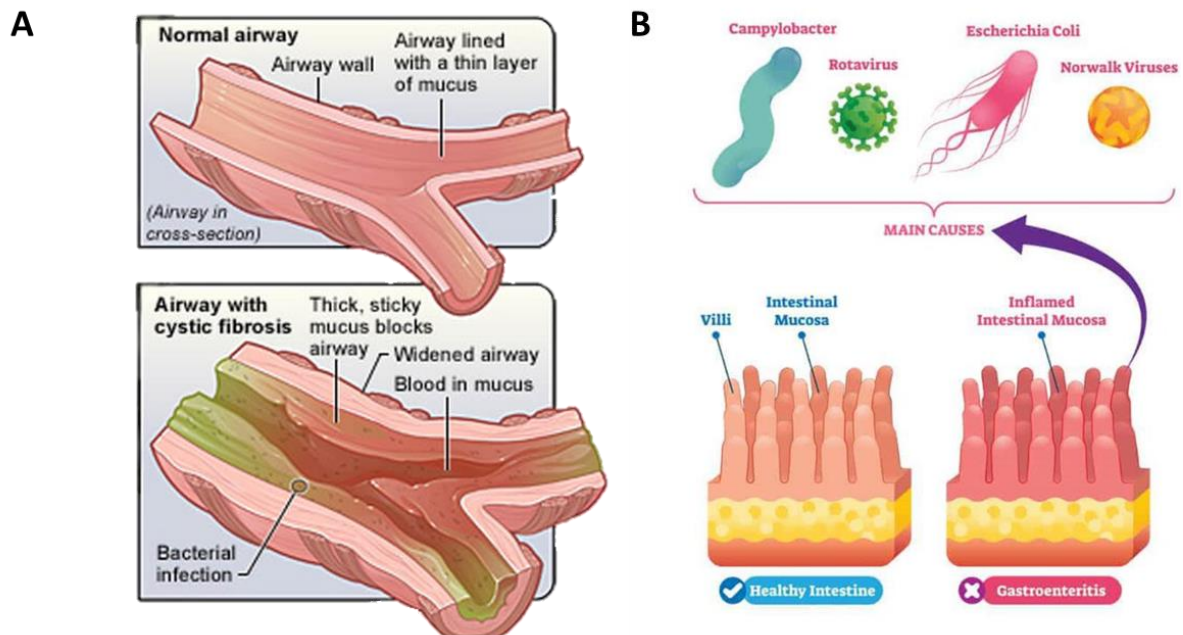
Figure 1: Illustrations of polymicrobial respiratory and gastrointestinal infections, originated from [14,15]. **A**: Lung segment of a CF infected patient. Airways are dilated and blocked by thick and sticky mucus.[15] **B**: Gastroenteritis causes inflamed intestinal mucosa. Several bacteria that are responsible for this polymicrobial infection are shown in the illustration as well.[16]

Based on the same principle, gastrointestinal infections arise from the loss of diversity in the intestinal microbiome. The human intestine usually harbors a rich community of more than thousand bacterial species.[17] However, it has been suggested that the actual number of bacterial species residing in the intestine could reach up to 36,000.[17] According to literature, antibiotics alter gut microbiota for up to 4 years post-exposure.[13,18,17,19] The effect of antibiotic mistreatment increases the abundance of pathogens of *Escherichia*, *Salmonella*, *Clostridia* and *Klebsiella genera* among others (Figure 1).[13,18,17,19]

The complexity of polymicrobial infections increases challenges for diagnosis and treatment. Identifying and targeting all the involved microorganisms is difficult, especially since different pathogens may require specific therapies. Furthermore, interactions among microorganisms can influence the effectiveness of treatments because the presence of one microorganism can protect others from the effects of antimicrobial drugs.[10,11]

2. Targeting Gram-negative bacteria

Gram-negative carbapenem-resistant pathogens, for example *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacteriaceae* are the among top three difficult-to-treat pathogens on the WHO global priority pathogens list reported in 2017.[20] Antibiotics with alternative mode-of-actions acting on these Gram-negative bacteria are therefore urgently needed.[21]

One of the major difficulties in targeting these pathogens is their unique cell envelope. The cell wall of a Gram-negative bacteria consist of an outer membrane (OM) as well as an inner membrane (IM) and the peptidoglycan layer.[22,23] OM consists of a sophisticated asymmetric glycerol phospholipid bilayer as an inner leaflet, while glycolipid lipopolysaccharides (LPS) form the outer leaflet (Figure 2).[22,23] LPS is important to protect the extracellular environment from harmful compounds. IM is also formed by phospholipid bilayer.[22,23] The periplasm separates both membranes, which consists of a peptidoglycan layer. While it is not considered as a physical barrier for drugs, this compartment is important as the site of action for β-lactam antibiotics and their potential inactivation by β- lactamases.[23,24] Besides these physical barriers, several efflux pumps are also located in the cell envelope that decrease the permeability of compounds into the cytoplasm (Figure 2).[23,24]
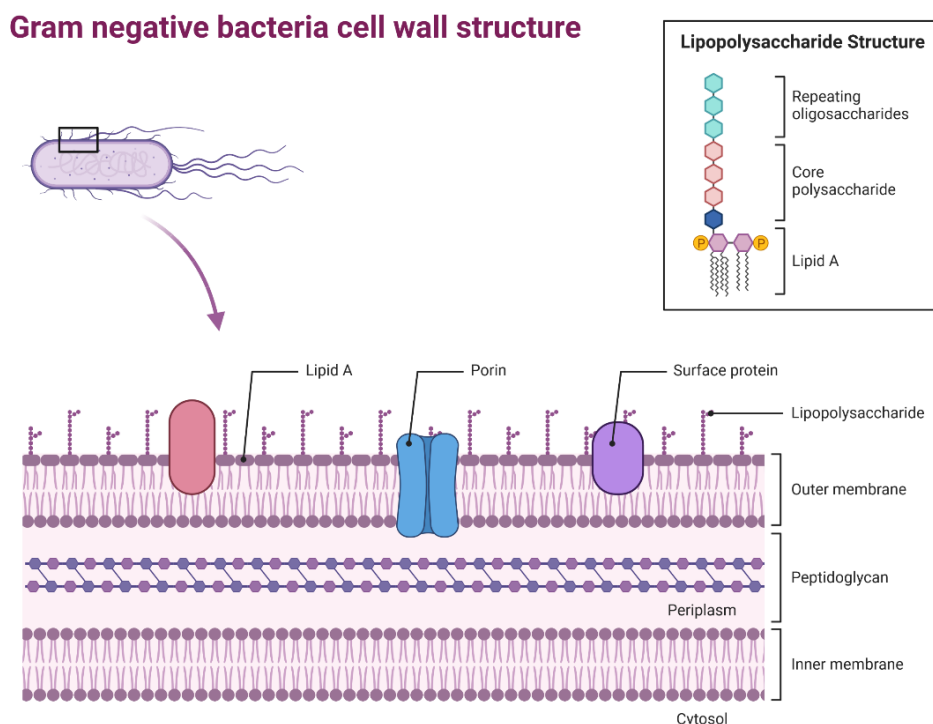
Figure 2: Cell wall structure of gram-negative bacteria: It is a bilayer cell envelope consists of an OM and an IM separated by peptidoglycan layer. Inner leaflet of both membranes is composed of a glycerol phospholipid bilayer while LPS from the outer leaflet of the OM. The structure of the LPS is described in the illustration as well (scheme from biorender)

The OM plays a crucial role in enhancing AMR as it evolves to protect against the damaging effects of antibiotics produced within microbial communities, which arise from microbial communication and competition.[24] Furthermore, an effective antibacterial drug is not only required to overcome all these obstacles but also have to be stable and active enough to reach and address the targets that are beyond these surface layers.[24]

2.1 Anti-virulence therapy

The AMR crisis and the resulting consequences discussed in chapter one emphasize the difficulty and intrinsic flaws of using conventional approaches such as bacteriocidal and bacteriostatic antibiotics to combat the resistant strains. In the recent years, a shift away from the traditional antibiotics has been developed and the "pathoblocker approach" came more and more into the focus.[21,25–27]

This alternative mode of action also referred as anti-virulence therapy, aims to disarm the pathogen by targeting the bacterial virulence factors. The principle of virulence is described as the ability of a pathogen to use virulence factors such as proteases or toxins or host tissue-damaging mechanisms that cause diseases.[26] Instead of affecting the viability of the pathogens, anti-virulence agents disrupt the interaction between the host system and the pathogen by disabling those virulence factors.[26,28,25,27] This leads to a lowered selection pressure for resistant mutants resulting in a decreased sensitivity for resistance development.[26,28,25]

Additionally, the spectrum of novel (anti-virulence) targets is broader than the limited number of traditional antimicrobial targets. This also helps to prevent cross resistance with already existing active agents.[26,27] Since virulence factors are specific to the pathogens, anti-virulence therapeutics will typically not affect the natural microbiome of the host, thereby colonization of pathogens can be reduced by host immune response.[26] In general, there are two types of pathoblockers: Those that directly target specific virulence factors and those who target the regulators of the virulence factors.[27] Targeting the regulatory systems offers the advantage of simultaneously eliminating multiple virulence factors.[27] One of the potential targeted regulatory systems, which will be introduced in the following chapters, is the Csr/Rsm system.

3. Csr/Rsm system as anti-virulence target

The main focus of this doctoral thesis is the investigation of the Carbon Storage Regulator A (CsrA, alternatively also called regulator of secondary metabolites RsmA in some species), a post-transcriptional regulator affecting mRNA translation and/or stability.[29–31] The *csrA* gene was first discovered by a transposon mutagenesis screen for identification of regulators in the stationary phase of growth.[32,33] The reason behind it is to understand global gene expression patterns of bacteria confronting fluctuant nutrient availability. To adapt and cope with various environmental changes bacteria need global regulators for gene expression reprogramming. These investigations demonstrated that *csrA* mutation had pleiotropic effects on various genes related to glycogenesis, gluconeogenesis and phenotype such as adherence and cell morphology.[32–34] It showed that CsrA allows bacteria to control coordinately stationary-phase gene expression and cell growth.[29] Early evidence showed that CsrA is not only essential for fundamental physiological properties and metabolism, but also for regulation of virulence systems required for host infection, which will be described later in the chapter.[29]

An additional reason why CsrA is considered a promising target for pathoblocker compounds is the high degree of homology observed among different CsrA proteins from various bacteria. CsrA is widespread among Gram-negative pathogens and highly conserved in its sequence and function.[29–31,35] For example, the homology between CsrA from *Y. pseudotuberculosis* and CsrA from *E. coli* is 95%.[29] Furthermore, CsrA homologs can also be found in a diverse range of animal and plant pathogens.[33] The phylogenetic tree and the amino acid sequences of the homologs are illustrated in Figure 3.[29]

This high homology suggests that inhibiting CsrA could have a broad-spectrum effect across multiple bacterial species, making it an attractive target for developing antimicrobial strategies that can combat a range of pathogenic bacteria.

A

- CsrA *Y. pseudotuberculosis*
- RsmA *E. carovotora* 98%
- CsrA *V. cholerae* 88%
- CsrA *L. pneumophila* Paris 73%
- CsrA *B. burgdorferi* 31%
- RsmE *P. fluorescens* 67%
- RsmA *P. fluorescens* 73%
- RsmA *P. aeruginosa* 85%
- CsrA *E. coli* 95%
- CsrA *S. flexneri* 95%
- CsrA *S. typhimurium* 95%

B

| β1 | β2 | β3 | β4 | β5 | α |

```
Y.pseudotuberculosis:  MLILTRRVGETLMIGDEVTVTVLGVKGNQVRIGVNAPKEVSVHREEIYQRIQAEKSQPTT
E.coli:                MLILTRRVGETLMIGDEVTVTVLGVKGNQVRIGVNAPKEVSVHREEIYQRIQAEKSQQSS
P.aeruginosa:          MLILTRRVGETLMVGDDVTVTVLGVKGNQVRIGVNAPKEVAVHREEIYQRIQKEKDQEPN
```

Figure 3: Phylogenetic tree of the CsrA homologs. The most important homologs and subjects of the thesis are encircled in red (**A**).[29] Amino acid sequences of the three homologs with the assignment of the β-sheets and α- helix are shown in **B**. The sequences are highly conserved and the important binding residue Arg44 is highlighted in green.

3.1   CsrA structure

The *csrA* gene encodes the CsrA protein that consists of 61 amino acids. Based on nuclear magnetic resonance (NMR) solution and crystallographic studies of CsrA from *E. coli* and homologs, the structure is composed of a homodimer.[32,35–37] Each monomer (about 7 kDa) contains five tandem β-strands (β$_1$ to β$_5$), one α-helix and a flexible C terminus (Figure 4). The β-strands are intertwined so that it forms a hydrophobic core while the α-helices with the C terminus extend away from the protein.[32,35–37]
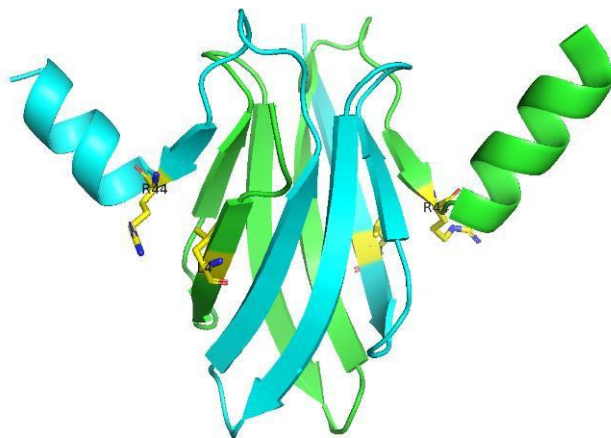
Figure 4: CsrA dimer is represented in ribbon diagram with chain A in cyan and chain B in green. Leu4 and Arg44 in the β$_1$ and β$_5$ strands are important RNA binding residues highlighted in yellow. (PDB: 1VPZ)

The two identical RNA-binding surfaces are positively charged and formed by the β$_1$-

8

and β5-strands of opposite polypeptides. The highly conserved sequence-specific mRNA recognition site contains a core-binding motif GGA. This motif is preferentially located within a short hairpin hexaloop (AR**GGA**U), which is usually overlapping or adjacent to the Shine-Dalgarno (SD) sequence (Figure 5).[30,32,35,36,36,38] During RNA-protein interaction, each monomer interacts with all six bases in the hexaloop. Furthermore, through alanine scanning mutagenesis of *E.coli* CsrA, amino acid residues (Leu4 and Arg44) in the β1 and β5 strands were identified to be essential for the *in vitro* RNA binding.[30] Previous studies have elucidated that dual site binding of a single CsrA dimer to one mRNA target occurs when the two target sites are separated by a spacer of 18 nucleotides (nt).[30,32,35] An example for dual site binding is the regulation of *glgC* gene encoding a glycogen biosynthetic enzyme. The subsequent section provides a detailed description and explanation of this regulatory mechanism.[32,33,35,38]
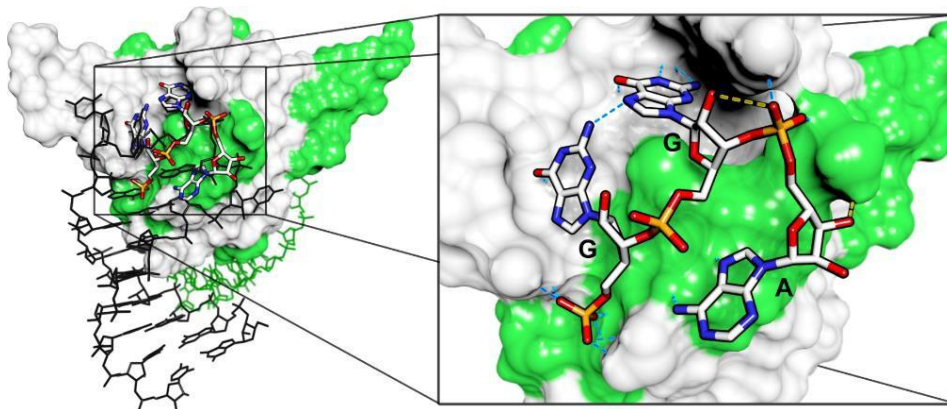
Figure 5: RNA shown in black and green binds to the surface of CsrA. The binding motif GGA is pointed out in the zoom-in view. (PDB: 1VPZ)

3.2   CsrA-mediated regulation

Both negative and positive regulation mediated by CsrA have been elucidated and are differentiated in acute infection (positive regulation) and in chronic infection (negative regulation). In most of the cases, CsrA acts negatively resulting in the decay of the mRNA targets.[31–33,39] The first identified regulatory mechanism of CsrA was the translational repression of the glycogen biosynthetic gene *glgC*. [33,38] One homodimer of the protein first binds to the high-affinity site of the *glgC* mRNA hairpin, which leads to the binding of the tethered CsrA homodimer to a low-affinity site overlapping the SD sequence. [33,38] All four sites are located in the untranslated leader of the *glgCAP* operon transcript. Bound CsrA blocks the ribosome binding, thus resulting in rapid degradation of the polycistronic transcript (Figure 6A).[33,38]

Another regulatory strategy of CsrA apart from translational repression is translational termination.[32,33] In. *E. coli*, the operon *pgaABCD* is responsible for the biosynthesis and secretion of biofilm polysaccharide adhesin. CsrA binds to the untranslated leader in *pgaABCD* mRNA, which contains six CsrA binding sites.[32,33] This is the largest number of binding sites in an mRNA among all CsrA targets identified so far, which reflects in complexity of the regulation.[32,33] Bound CsrA mediates Rho-dependent termination by remodeling the transcript and results in exposing the *rut* binding sites for Rho (Figure 6B).[32,33] This regulatory model is the first example of CsrA for directly controlling transcription.[33]
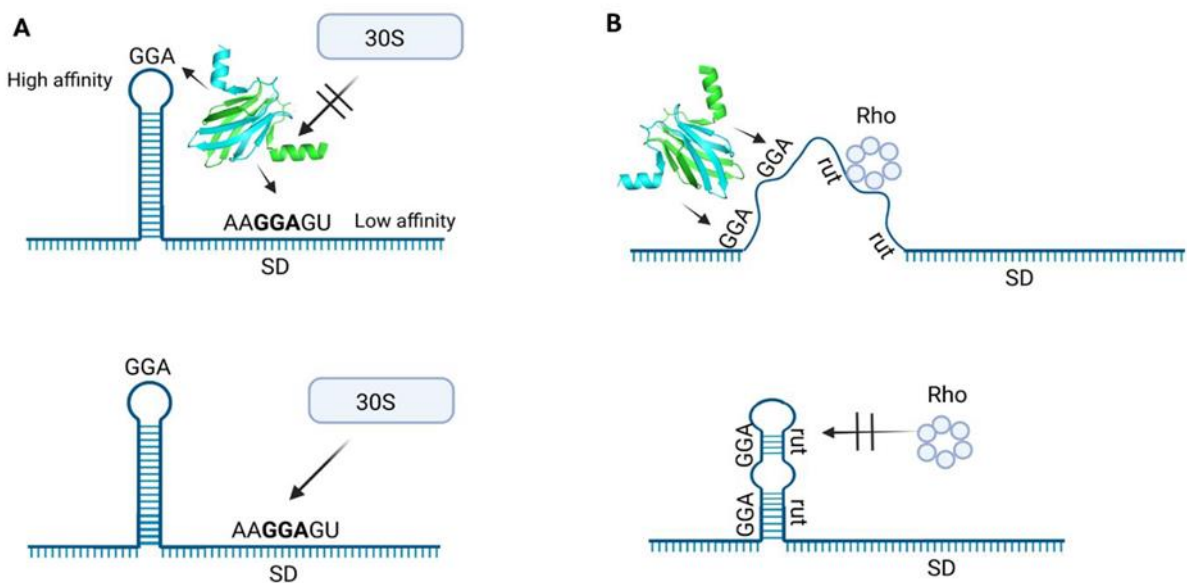
Figure 6: CsrA-mediated negative regulation with two different mechanisms: Translational repression (A) and Translational termination (B). CsrA regulates the *glgC* gene by binding to the dual sites in the untranslated region (UTR) of *glgC* operon and inhibits the ribosome binding (A). CsrA binds to the UTR of *pgaABCD* genes and mediates the Rho-dependent termination by reshaping the transcript. This leads to exposure of the *rut* binding sites for Rho.(B).[33]

Most of the CsrA-mediated activation/stabilization mechanisms have not been well-studied.[32,33] For instance, in *E. coli*, CsrA-binding has a positive influence on *flhDC* mRNA, which is a master operon for flagellum biosynthesis. CsrA stabilizes *flhDC* by binding to two sites at the far upstream (>150 nt) 5' end of the *flhDC* transcript. [32,33] It leads to prevention of the 5' end-dependent cleavage by RNase E. This transcript stabilization results in activation of *flhDC* expression, which enhances the motility of bacteria (Figure 7AB).[33]
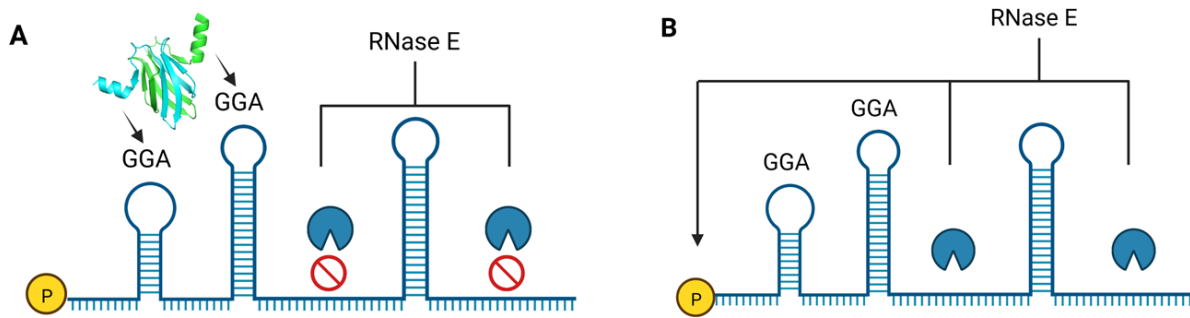
Figure 7: CsrA-mediated stabilization. CsrA binds to two binding sites of *flhDC* mRNAs and prevents 5' end- dependent cleavage by Rnase E (A). The cleavage of the master operon for flagellum biosynthesis occurs without the stabilization of CsrA and leads to limited motility of bacteria (B).[31]

3.3   Regulation of the Csr/Rsm System
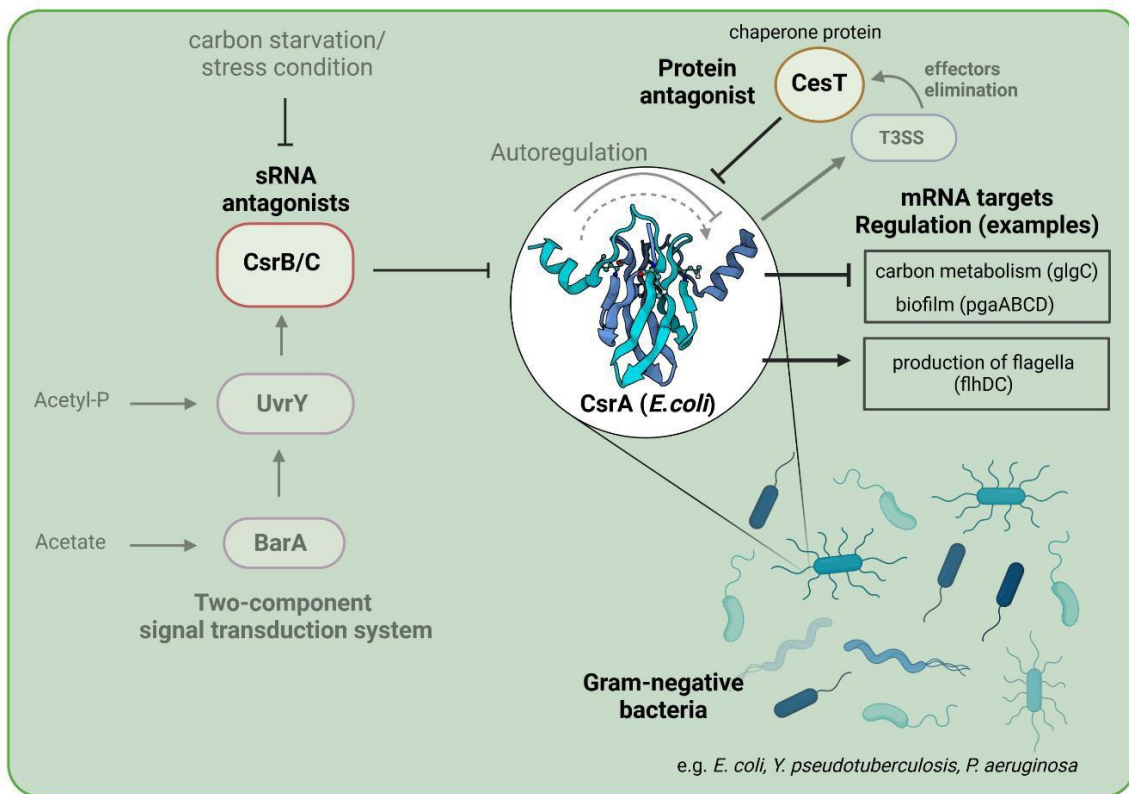
Figure 8. Simplified regulation circuit of CsrA in *E. coli*: CsrA is antagonized by sRNAs CsrB/C and chaperone protein CesT. While most Gammaproteobacteria use the BarA-UvrY two-component signal transduction system to activate CsrB/C transcription, *Yersinia pseudotuberculosis* relies on the cAMP/Crp system instead.[32,38,40] Furthermore, sRNA antagonists are controlled by carbon starvation/stress condition as well. [32,38,40] CsrA can autoregulate its own expression by simultaneously activating and repressing it.[32] However, these feedback loops and regulatory circuits are not relevant to the current study and are depicted more transparently.[32,33] CsrA mediated regulations have been already discussed in **chapter 3.2**. The regulation of CesT will be described in **chapter 3.3.2**.

3.3.1 sRNA antagonists CsrB/C

The activity of *E. coli* CsrA is controlled by the sequestration of inhibitory sRNAs CsrB/CsrC. According to literature, CsrB is the principle antagonist in *E. coli* during the growth conditions that were tested. CsrC shares both structural and functional

similarities with CsrB.[34] The affinity of CsrA for these sRNAs (~350 nt long) is higher than for its own target mRNAs. Furthermore, the amount of CsrB/C determines the level of CsrA available for binding targets. The reason for the high affinity is due to those 13-22 potential binding sites, which are able to sequester ~9 CsrA dimers. Suggestion for the CsrA binding element is the repetitive hairpin loop motif 5'-CAGGAUG-3' which can also be found in single-stranded segments between the loops.[34,38,41]
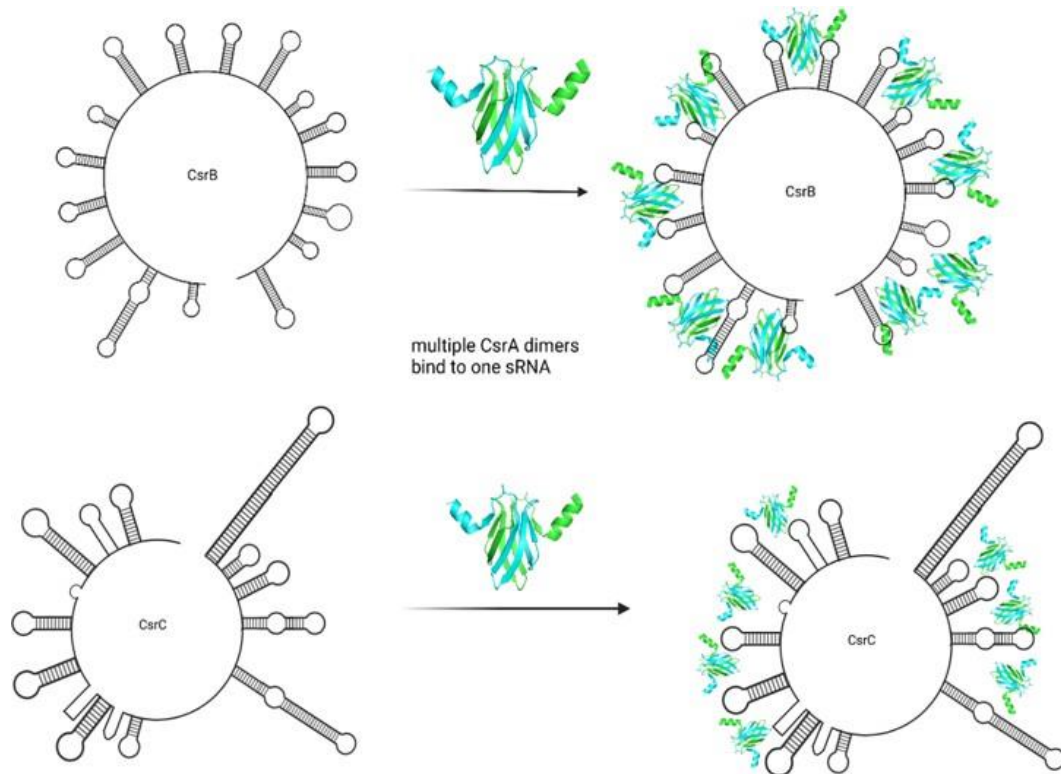
Figure 9. Structure of both sRNAs CsrB/CsrC. Multiple binding sites allow the sRNAs to sequester more than two CsrA dimers.[34,41]

Experiments with ∆*csrB* strain showed that its absence caused pleiotropic effects on *E. coli* physiology and expression of downstream targets regulated by CsrA are similar affected.[34] Overall, these sRNAs are produced to enable bacteria to fine-tune CsrA activity and thus increase the robustness of Csr regulatory circuit.[32,34,38,40,41]

3.3.2 Regulation of csrA via CesT

Besides applying sRNA antagonism for the regulation of CsrA/RsmA, bacteria also use protein to inhibit the activity of CsrA. For example in Enteropathogenic *E. coli* (EPEC) a chaperon protein called CesT (Figure 10) binds to CsrA leading to alterations in virulence and metabolic gene expression.[32,42,43]
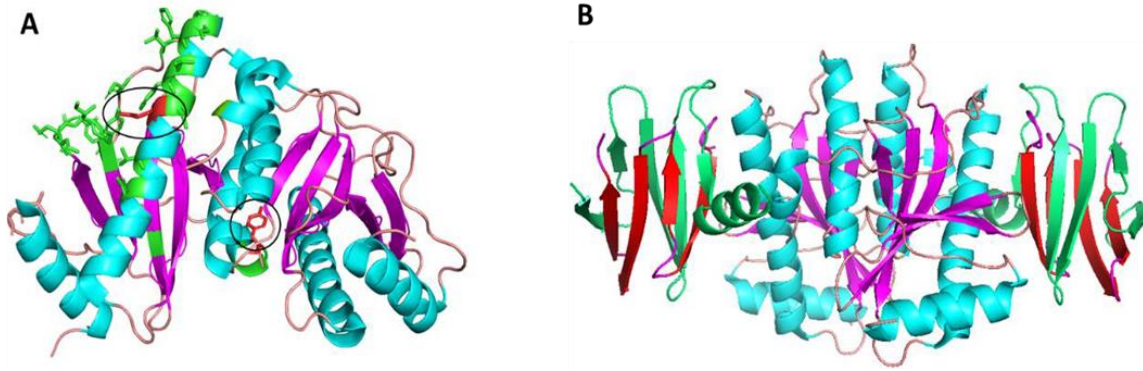
Figure 10. Structure of CesT with CsrA binding sites (PDB: 5Z38). CesT is a dimeric protein and one monomer consists of 5 β-sheets (magenta); 3 α-helices (cyan) and loops are shown in salmon. CsrA binding regions are highlighted in green and located mainly at the C-terminal region. Tyr152 and Glu121 are the important binding residues highlighted in red and encircled (A).[42,43] One CesT dimer is able to bind to two CsrA dimers (red and green).(B). [42,43]

CesT is responsible for stabilizing and translocating the virulence factors (effectors) needed by the type three secretion system (T3SS). In host environment, this strategy of T3SS-mediated bacterial infection is adapted for pathogenic survival.[32,42,43] Furthermore, previous studies showed that liberated CesT binds to the CsrA regulator after injecting the effectors into the host cells. Bound protein decreases the T3SS activity and this leads to accumulation of the effectors, which in turn sequester CesT. These findings suggested that CsrA and T3SS activities regulate each other indirectly in a negative-feedback loop (Figure 8).[32,42,43]

4. IGF2BP2/IMP2 as anti-cancer target

RNA-binding proteins are important in numerous physiological processes, as previously elucidated alongside the post-transcriptional regulator CsrA. A side project in this doctoral thesis focused on the human insulin-like growth factor 2 (*IGF2*) mRNA binding proteins (IGF2BPs/IMPs). There are three known members of the IGF2BP family: IGF2BP1, IGF2BP2, and IGF2BP3. The discovery of these proteins dates back to 1999 and among them, IGF2BP2 or commonly known as IMP2 with a molecular mass of 66 kDA plays a critical role in the maintenance of RNA stability, translation and localization.[44–46]

4.1 IMP2 structure

All three mammalian IMPs have highly conserved amino acid sequences and shared share an overall sequence identity of 56%. Structurally, IMP proteins comprise two RNA recognition motif domains (RRMs) N-terminally followed by four consecutive

hnRNP (KH) domains. The arrangement of these domains is in three pairs, namely RRM12, KH12, and KH34. (Figure 11). Despite the high sequence similarity, each IMP protein regulates different RNA targets and they are distinguishable in their phenotypes according to knockouts (KO) experiments.[47–49]
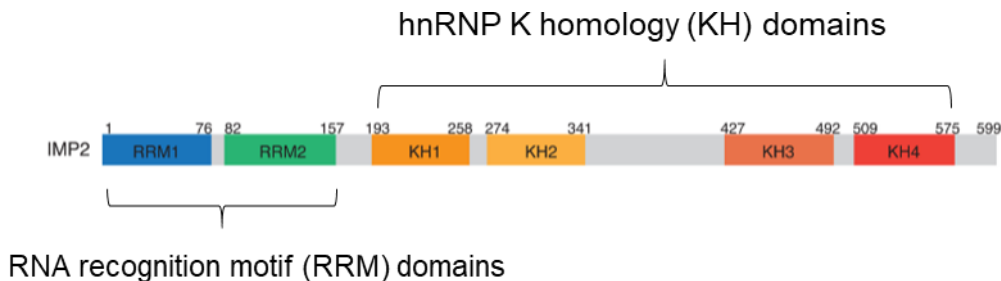
Figure 11. Schematic diagram of the conserved IMP2 domain arrangement.[50]

IMP2 shares a lower sequence homology with the two other IMP family members, while IMP1 and IMP2 share a sequence homology of 73%. Additionally, IMP2's expression lasts throughout life compared to the expression pattern of IMP1 and IMP3, which is from the mid-to-late period of embryonic development with little to no expression in adults. However, reactivation of their expression has been observed in various tumors through clinical studies, suggesting a potential involvement of IMP1 and IMP3 in the development of cancer.[45] Nevertheless, the structural sequence identity of all three IMPs' RNA binding domains is high with more than 70%.[48,49]The N-terminal RRM domains and all four KH domains have been proposed as contributors to stabilizing the association between IMPs and RNA.[48] Previous *in vitro* and *in vivo* investigations have emphasized the importance of the C-terminal KH34 domains in RNA recognition. Although the experimental evidence primarily based on IMP1, the extensive structural and sequence similarity suggests that the KH34 RNA binding properties could extend to the entire IMP protein family.[47–49]

The architecture of a single KH domain refers to the type 1 KH fold (βααββα) and together both domains arrange in an anti-parallel pseudo-dimer conformation (Figure 12). This orientation facilitates the interaction with target RNA by positioning the putative RNA binding surfaces at opposite ends of the molecule.[49]Based on such structural arrangement, IMPs can select for RNA targets containing two distinct protein binding regions interspacing by a number of nucleotides.[49]

According to previous studies, sequence-specific recognition is primarily facilitated by

the variable loop regions of KH34 domains.[45] For example, mutations in the loop regions, particularly specific variable loop amino acid, confirmed the differences observed in the sequence specificity between IMP1 and IMP2.[45] Mutated residues in the linker between KH domains or the KH domain variable loop also reduce the affinity of RNA-binding proteins (RBPs) for their targets.[45] These findings emphasized the importance of the various loop for sequence-specific recognition. In case of IMP2, the recognition of specific RNA is through the interactions between the GXXG motifs and the variable loops shown in Figure 12. IMP2-specific sequences are dominated by AGGU and UGGA.[45] In general, the GXXG motifs of IMP family members contain multiple positively charged residues, suggesting a preference for interaction with the negatively charged phosphate backbone of RNA.[45]
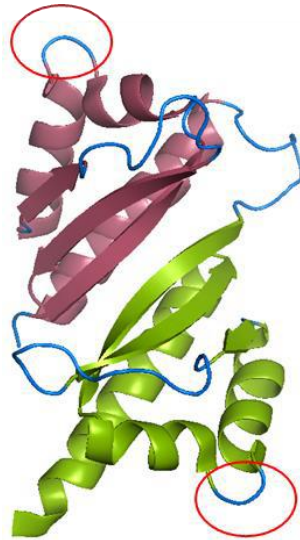
Figure 12. Secondary structure of IMP2 KH34 domains. KH3 is shown in red and KH4 is shown in green. Due to the anti-parallel arrangement of this pseudo dimer, potential RNA binding surfaces position at opposite ends of the molecule. The variable loops for sequence-specific recognition are encircled in red. (PDB: 6ROL)

4.2 IMP2 regulation

In general, IMPs are found in the cytoplasm and form with their target mRNAs distinct ribonucleoproteins (mRNPs)(Figure 13).[47] These stable protein-RNA complexes together with various RNA-binding proteins (RBPs) allows the control of mRNA translation, transport, and degradation as illustrated in Figure 13. As an example, the stabilization of c-myc mRNA through the coding region determinant (CRD) is ensured by four of these RBPs (HNRNPU, SYNCRIP, YBX1, and DHX9).[51] IMP1 is associated with these factors in a CRD-dependent manner and has similar distribution pattern as them in non-polysomal cellular fractions containing c-myc mRNA.[51] Moreover,

colocalization of IMP1 with the four RBPs in the cytoplasm has been observed.[51] The regulation is based on "caging" or releasing the mRNAs for specific target processes, which are mostly triggered by phosphorylation events.[47] For example, IMP1 activates the translation of β-actin (ACTB) mRNA by Src-directed tyrosine phosphorylation within the linker region that bridges KH2 and KH3 domains.[47,52] Another example is the IMP2-mediated regulation of increased IGF2 protein synthesis.[46] IMP2's N-terminal linker region connecting RRM2 and KH1 is phosphorylated by the mammalian target of rapamycin complex 1 (mTORC1).[47,53] The phosphorylated IMP2 indirectly regulate and influence the upregulation of IGF2 translation by associating with the IGF2 leader 3 of 5'- untranslated region (UTR). [47,53]
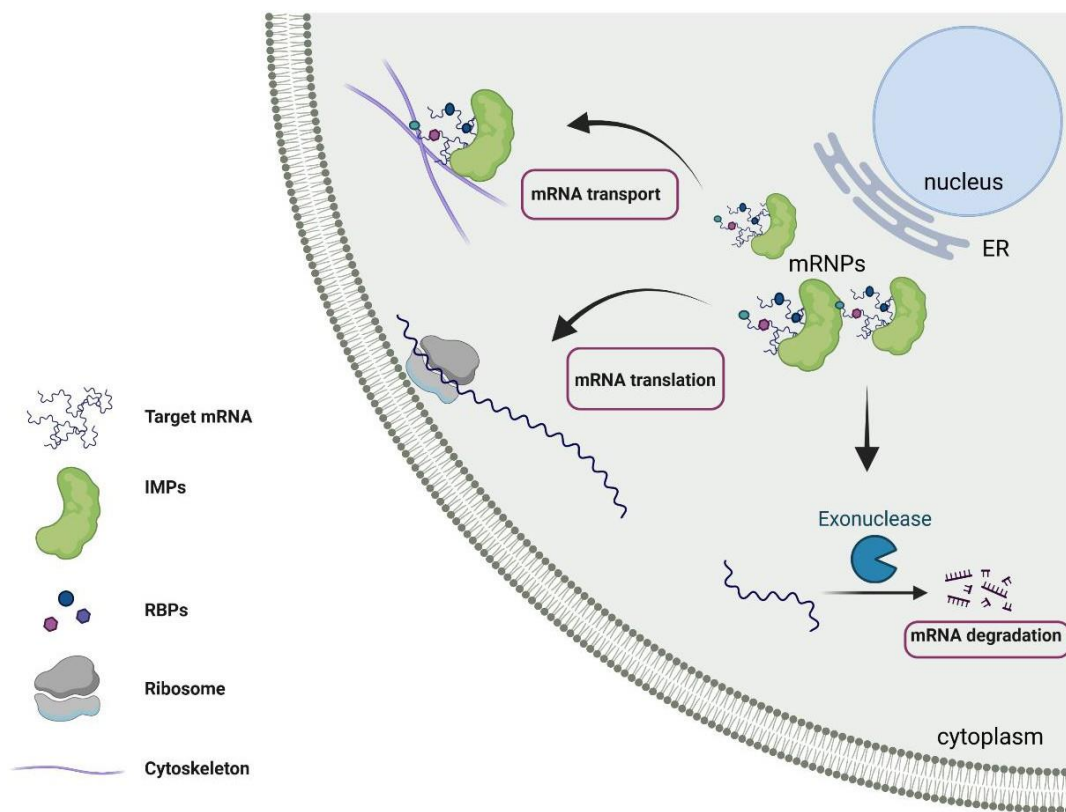
Figure 13. Simplified scheme of IMPs-mediated regulations. IMPs associated with specific target mRNAs along with other RBPs within cytoplasmic mRNPs. The process of dissociating these associated mRNAs from mRNPs can lead to either mRNA degradation or mRNA translation for protein synthesis. The reason for forming stable mRNPs is to facilitate the directed mRNA transport along the microtubule and/or actin cytoskeleton. To prevent uncontrolled translation of sorted mRNAs, translational silencing mechanisms of localized transcripts are used while undergoing transport.[47] (Scheme generated with Biorender)

Prior investigations have demonstrated a significant correlation between the IMP2 gene and the development of type 2 diabetes (T2D) by disrupting insulin secretion.[44–46] However, of particular importance is the involvement of IMP2 in the regulation of cancer development and progression. Previous studies have shown that IMP2 exhibits a higher amplification and expression across various cancer types compared to IMP1

and IMP3.[44–46] The ability of IMP2 to regulate directly and indirectly the expression and translation of diverse oncogenes, for example breast, colorectal and lung cancer[46] leads to a multitude of biological processes, such as an increase in growth rate, cell proliferation, migration, and invasion. Moreover, IMP2's impact on oncogenes influences cell metabolism and promotes epithelial-mesenchymal transition.[44–46]

Based on *in vitro* and *in vivo* data demonstrating the promotion of tumorigenesis and tumor progression in liver and colon cancer for example, IMP2 emerges as a promising target for gastrointestinal tumor treatment. Notably, the expression of IMP2 is higher in cancer cells compared to normal cells, making it a specific target. Consequently, this project aims to conduct fluorescence polarization (FP)-based screening to identify potential small-molecule inhibitors that disrupt the protein-RNA binding associated with IMP2.

5. Biophysical methods and *in bacterio* assays

5.1   Fluorescence Polarization-based Competition Assay

The discovery of novel drugs and therapeutics relies on important and well-established methods such as screening of compound libraries to identify novel inhibitor scaffolds against a certain biological target. Reliable, homogenous and robust assay technologies are required for high-throughput screening (HTS).[54,55] One of such technologies is the fluorescence polarization (FP)-based assay, which has been exploited for interrogating a broad range of molecular interactions (like protein-protein, protein- peptide, protein-nucleic acid and protein-small molecule).[54,55] The advantages of this methodology are for example evaluation of molecular processes in solution and the possibility to measure multiple times at once. Furthermore, it is less pricey and insensitive towards some assay interferences. Hence, the FP assay is one of the favorited methods nowadays for biophysical screening and determination of inhibitory activity.[54,55]
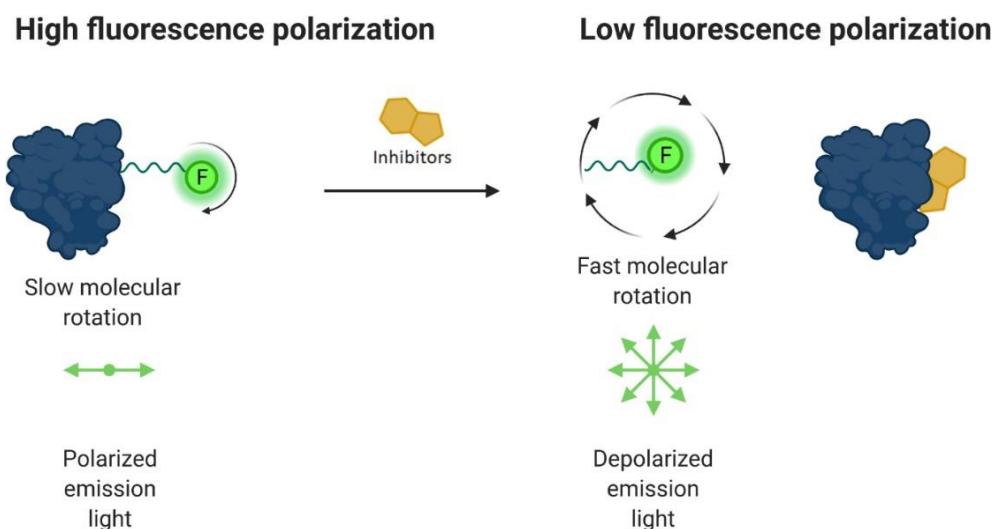
Figure 14. Principle of the fluorescence polarization assay. Due to the slow molecular rotation, bound fluorophore causes polarization retention of the emitted light (high polarization value). Free fluorophore, which is replaced by potential inhibitors and therefore rapid rotation occurs, emits highly depolarized light (low polarization value).(Scheme generated with Biorender)

The principle of FP assay is based on the irradiation of fluorophore probe with linear polarized light and detecting its emitted fluorescence with a degree of polarization, which is dependent on the molecular rotation. If the fluorescent molecule is bound to a larger molecule or a surface, its rotation and mobility are restricted, resulting in a higher degree of polarization.[54,55]
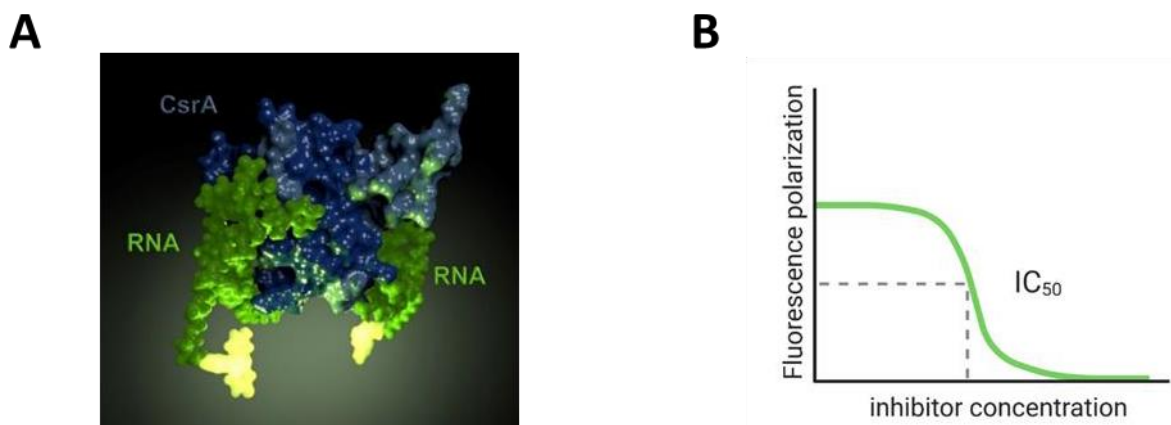
Figure 15. Two specific RNA sequences labeled with fluorophore bind to CsrA. Each CsrA homolog has a different RNA sequence (A). Sigmoidal inhibition curve of a typical competition experiment in which an inhibitor competes with labeled RNA sequence. Fluorescence polarization (millipolarization mP) is plotted against the logarithmic inhibitor concentration .Determination of the binding affinity is based on the IC_{50} values (B).(Scheme generated by Biorender)

In the frame of this work, the specific RNA sequence is labeled with a fluorophore, which is designed to bind to CsrA. (Figure 15A) To perform the assay the fluorescent RNA sequence and CsrA are mixed in a solution. (Details of the procedure are shown in the following chapters). The degree of polarization is measured using fluorescence polarization reader (Clariostar). This instrument emits polarized light and detects the

emitted light's polarization intensity. By measuring the change in polarization, the binding affinity of potential inhibitors in competition to the labeled RNA sequence can be quantitatively assessed. The determination of the IC$_{50}$ values is possible as well as single point measurements in inhibitors screening, which have been used for the CsrA and IMP2 project (Figure 15B).[30,31,44]

5.2   Luciferase reporter gene assay

The general term for emission of photons occurs without being caused by heat is Luminescence. It means any substance that can emit the light from their electronically excited states is luminescent. Fluorescence, as described in the previous paragraph, is only a type of Luminescence, which occurs in chemical or biological systems through spontaneous relaxation of electrons usually from a singlet excited state to a singlet ground state after absorption of electromagnetic radiation.[56,57]

The emission of light by living organisms mediated by biochemical processes is described as bioluminescence. The bioluminescent properties of bacterial luciferases for example can be utilized to investigate the real-time inhibitory effects on target protein levels. These luciferases are capable of emitting light when exposed to luciferin (reduced riboflavin phosphate, FMNH$_2$), a substrate that undergoes oxidation in association with a long-chain aldehyde and an oxygen molecule.[56,57] By introducing the luxCDABE operon derived from bacteria, cells can produce detectable light at a wavelength of 490 nm. This operon consists of the genes encoding the luciferase enzyme (LuxAB) and the enzymes responsible for synthesizing the substrate (LuxCDE).[56,57]

In this study an *in bacterio* assay based on luciferase reporter gene assay is established in order to assess potential hits targeting CsrA (Figure 16). The whole procedure will be described in detail in **chapter III.2**.
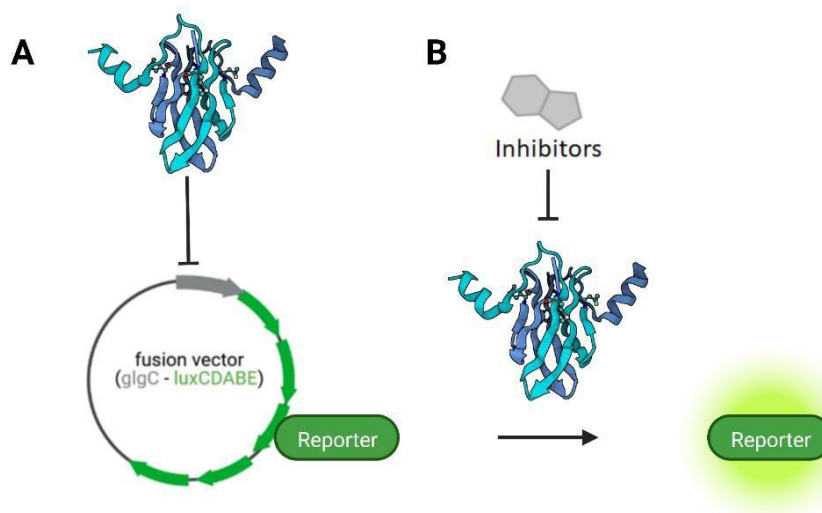
Figure 16. *In bacterio* assay based on luminescence reporter gene assay for assessment potential CsrA inhibitors. A fusion vector containing *glgC* promoter region and *luxCDABE* genes is illustrated in (A). Inhibition of CsrA leads to upregulation of *glgC* expression resulting in enhancement of the bioluminescence (B). (Scheme generated with Biorender)

5.3   RT -qPCR

In order to understand the impacts of novel inhibitory compounds on CsrA, the analysis of the gene expression of either CsrA's target or CsrA-modulated genes is crucial and important for this research. Among the techniques available for quantifying gene expression, Reverse Transcription-quantitative Polymerase Chain Reaction (RT-qPCR) has emerged as a reliable, powerful and commonly used tool. This technique combines the principles of reverse transcription (RT) and quantitative polymerase chain reaction (qPCR), enabling measurement and quantification of RNA molecules of interest (Figure 17).[58,59]

The principle of RT-qPCR involves the conversion of RNA molecules into complementary DNA (cDNA) through the process of reverse transcription. A limiting factor of this step is the contamination of the RNA template by (genomic) DNA. That is why it is important to add a DNAse incubation step during the procedure of RT-PCR. The resulting cDNA from RT-PCR serves as the template for subsequent amplification and quantification of the target RNA using qPCR. The determination of the initial RNA concentration can be monitored in real-time during each PCR cycle. RT-qPCR enables measurement of gene expression levels, even when RNA quantities are limited. Using gene-specific primers the sensitivity and specificity of target cDNA can be increased and reduce the unspecific background.[58,59]
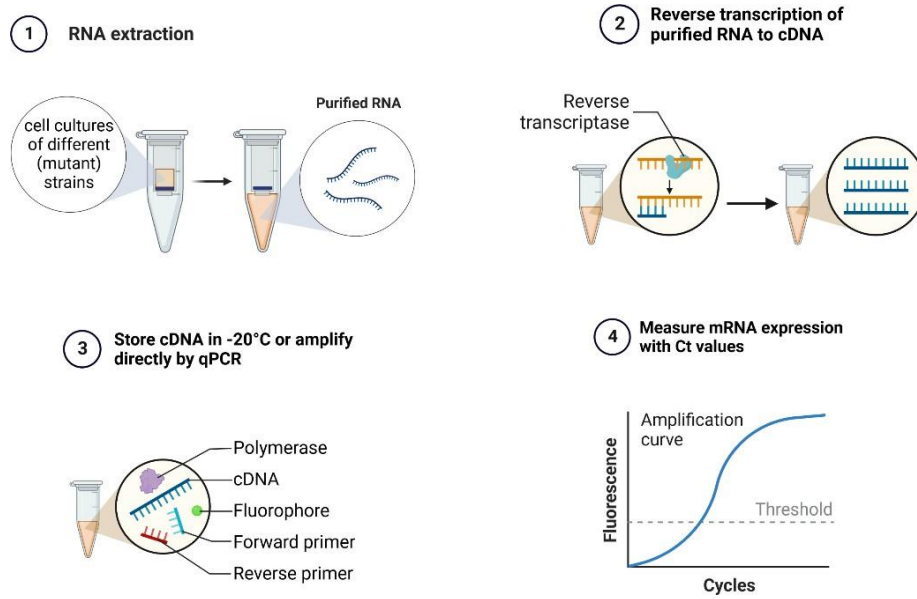
Figure 17. Simplified workflow of a typical RT-qPCR. This scheme describes the steps from RNA extraction of different strains to RT-PCR of the purified RNAs and ends at qPCR procedure. (Scheme generated with Biorender)

II.   Chapter 2: Aim and Scope

Polymicrobial infections pose a serious threat for the public health in the present days. Due to the increasing prevalence of antimicrobial resistance, the development of new drugs that bypass resistance mechanisms is a critical step in modern infection research. Concurrently, the battle against cancer diseases remains an ongoing challenge, with no definitive end in sight. Both areas of research require continuous efforts and innovative approaches to improve patient outcomes and address these complex and evolving health threats.

In both indications, the aim is to address unexplored promising target proteins through medicinal chemistry. In the present work, these target proteins undergo macromolecule-macromolecule interactions. To be specific, these targets are the Carbon Storage Regulator System A (CsrA) in bacterial systems and the human insulin-like growth factor 2 (IGF2) mRNA binding proteins (IGF2BPs/IMPs) in mammalian systems. The commonality of both target proteins is their RNA-binding ability.

In the context of this doctoral thesis, new starting points (hits) against these targets will initially be identified using biophysical methods. Recombinant production of the target proteins will be carried out, followed by screening of compound libraries using Fluorescence Polarization. These hits will be prioritized based on their properties (synthetic accessibility, solubility, metabolic stability). Subsequently, these compounds will be selected and verified through cell-based assays and undergo medicinal chemistry optimization to improve pharmacodynamic and pharmacokinetic properties. The goal is to generate initially cell-active substances suitable for general proof-of-concept studies.

The first part of this work (**chapter III**) focuses on targeting CsrA. Phage display-based screening led to the discovery of a novel peptide-based inhibitor. (**chapter III.1**) Furthermore, **chapter III.2** describes the development of the first *in bacterio* assay for evaluating potential CsrA inhibitors. A follow up manuscript is currently being prepared, highlighting the identification of promising inhibitor scaffolds from a commercial library through virtual and FP-based screening.
Moving to **chapter IV**, this section explores the potential of IGF2BP2/IMP2 as a

promising target for cancer therapy. Utilizing FP-based screening (**chapter IV.1**), small-molecule inhibitors were identified, which were further validated through additional assays like thermal shift assays (TSA) and saturation transfer difference (STD-) NMR. Subsequent *in vivo* tests, including xenograft models, confirmed the therapeutic potential of these inhibitor scaffolds against IMP2. **Chapter IV.2** delves into the crucial role of IMP2 in anti-cancer therapy. It discusses the association between IMP2 and the regulation of polo-like kinase 1 (PLK1), a critical player in cancer. Dysregulation of PLK1 can lead to chromosomal instability (CIN) in cancer cells. By targeting IMP2, a reduction in the proliferation of PLK1-overexpressing tumor cells was observed among others, indicating the potential of IMP2 as a key therapeutic target in cancer treatment.

III. Chapter 3: Disputing the Csr/Rsm system in Gram-negative pathogens

1. Publication Title:

Phage display-based discovery of cyclic peptides against the broad spectrum bacterial anti-virulence target CsrA

Authors:

Valentin Jakob*, Ben G. E. Zoller*, Julia Rinkes, Yingwen Wu, Alexander F. Kiefer, Michael Hust, Saskia Polten, Andrew M. White, Peta J. Harvey, Thomas Durek, David J. Craik, Andreas Siebert, Uli Kazmaier and Martin Empting

* These authors contributed equally.

Phage display-based discovery of cyclic peptides against the broad spectrum bacterial anti-virulence target CsrA

Valentin Jakob [a, b, 1], Ben G.E. Zoller [a, b, 1], Julia Rinkes [a, b], Yingwen Wu [a, b], Alexander F. Kiefer [a, b], Michael Hust [c], Saskia Polten [c], Andrew M. White [d], Peta J. Harvey [d], Thomas Durek [d], David J. Craik [d], Andreas Siebert [e], Uli Kazmaier [e], Martin Empting [a, b, *]

[a] Department of Drug Design and Optimization (DDOP), Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS) - Helmholtz Centre for Infection Research (HZI), Campus E8.1, 66123, Saarbrücken, Germany
[b] Department of Pharmacy, Saarland University, Campus E8.1, 66123, Saarbrücken, Germany
[c] Technische Universität Braunschweig, Institut für Biochemie, Biotechnologie und Bioinformatik, Spielmannstr. 7, 38106, Braunschweig, Germany
[d] ARC Centre of Excellence for Innovations in Peptide and Protein Science, Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, 4072, Australia
[e] Institut für Organische Chemie Saarland University Campus C4.2, 66123, Saarbrücken, Germany

ABSTRACT

Small macrocyclic peptides are promising candidates for new anti-infective drugs. To date, such peptides have been poorly studied in the context of anti-virulence targets. Using phage display and a self-designed peptide library, we identified a cyclic heptapeptide that can bind the carbon storage regulator A (CsrA) from Yersinia pseudotuberculosis and displace bound RNA. This disulfide-bridged peptide, showed an IC50 value in the low micromolar range. Upon further characterization, cyclisation was found to be essential for its activity. To increase metabolic stability, a series of disulfide mimetics were designed and a redox-stable 1,4-disubstituted 1,2,3-triazole analogue displayed activity in the double-digit micromolar range. Further experiments revealed that this triazole peptidomimetic is also active against CsrA from *Escherichia coli* and RsmA from *Pseudomonas aeruginosa*. This study provides an ideal starting point for medicinal chemistry optimization of this macrocyclic peptide and might pave the way towards broad-acting virulence modulators.

© 2022 The Authors. Published by Elsevier Masson SAS. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

For many years, researchers have warned about the antimicrobial resistance (AMR) crisis [1–3]. The rampant spread of multidrug resistant bacterial pathogens combined with the lack of novel treatment options, especially against Gram-negative species, poses a great threat for our modern healthcare systems [4]. For this reason, the discovery of new anti-infective candidates with novel and innovative mechanisms-of-action are needed. We consider the carbon storage regulator A (CsrA; in some species also called the regulator of secondary metabolites, RsmA) [5] as an attractive, yet underexplored, virulence-modulating target [6,7]. It is widespread in Gram-negative pathogens [8] where its sequence and function is highly conserved [9]. Knock-out studies in *Pseudomonas aeruginosa*, Yersinia pseudotuberculosis and *Helicobacter pylori* [10] have demonstrated its critical role in bacterial virulence and highlighted its potential as a therapeutic target [11]. The CsrA/RsmA protein is a post-transcriptional regulator [12], that binds and regulates translation of mRNA and, thus exerts pleiotropic effects on the bacterial transcriptome (Fig. 1) [13,14].

Through its mRNA binding activity it is involved in the regulation of quorum sensing [15], motility [16], carbon metabolism [17], peptide uptake *via* cstA [18–20] and biofilm development [21]. To disrupt the function of CsrA/RsmA at the molecular level, protein-RNA interaction inhibitors need to be devised. CsrA usually occurs as a homodimer, with two identical RNA-binding sites [22].
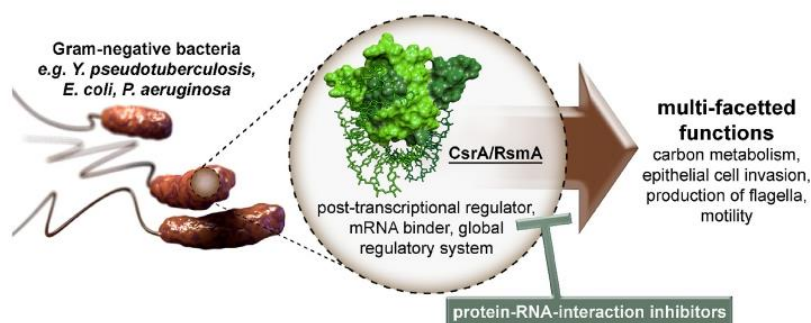
Fig. 1. CsrA/RsmA as a promising drug target for multi-pathogen virulence modulation by disruption of an essential protein-RNA interaction.

In a previous study using *Yersinia* CsrA and a short piece of RNA that contained the important core binding motif GGA, it was shown that this RNA can be displaced by small molecules [10,23]. In the present work, we sought to find novel lead molecules within the extended Lipinski space (MW between 500 and 1000 Da) [24–26]. These molecules should provide a suitable basis for disrupting macromolecule-macromolecule interactions while still retaining the potential for membrane permeability and oral bioavailability [26,27].

2. Materials and methods

2.1. Phage display

A detailed description of the experimental procedures used for oligomer design, cloning, library packaging and phage display was previously published as a protocol [28].

In brief, an oligomer was designed, which encodes for a very small peptide library, which has the structure XCXXXCX. There are two fixed cysteine positions in it and X encodes for any amino acid except cysteine. The oligomer was synthesized by Ella Biotech GmbH. This library was cloned into the phagemid pHAL30 [29] where it displays $2.48 \cdot 10^6$ different peptides. The library was packed into M13K07 phages, which are able to present the peptides on their surface to a protein target.

Under oxidative conditions, the cysteines are forming a macrocycle over the disulfide bond. For selection of potential CsrA binders, phage display was established. For this process, CsrA_-biot_His$_6$ was bound to a streptavidin-coated ELISA plate well. After blocking with BSA/milk powder, the pre-selected library, where BSA and streptavidin binders were excluded, was added. Unspecific binders were eliminated with a plate washer (Tecan Hydroflex), where PBS pH 7.4 containing Tween-20 was used. After the third panning round the clones were separated on agar plates. 32 clones were sequenced and checked for plausibility. Criteria for selection were an intact sequence with low tryptophane content and one glutamic acid, while carboxyl groups are beneficial for binding to positively charged surfaces on CsrA. We identified two interesting sequences containing a glutamate residue within the macrocycle. Peptide **1** (*Ac*-V-[CSELC]$_{cyclic}$-W-*NH$_2$*) could be successfully synthesized. Synthesis and macrocyclization of the alternative sequence *Ac*-H-[CQEVC]$_{cyclic}$-P-*NH$_2$* yielded only dimerized product (data not shown). Nevertheless, this finding underscores the potential ionic interaction between the glutamate side chain and basic residues on the protein surface.

CsrA coating amount of the wells, detergent amount in PBS buffer as well as washing stringency with the plate washer were optimized. Before each panning round, the wells were coated with

either 4 μg or 40 μg CsrA_biot_His$_6$. The better results were obtained with 4 μg 0 μg CsrA was used as negative control. Before phage elution with trypsin, PBS containing different amounts of Tween-20 (0.05%, 0.1% or 0.2%) was tried to rinse the wells with the plate washer. Since the variation effect was not significant, 0.1% Tween-20 was chosen as the detergent amount. The number of washing cycles, however, had a significant impact on the results. Tested were 2/4/6, 4/8/10 and 10/10/10 washing cycles. The first number corresponds to the number of washing cycles after the first panning round, the second after the second and the third after the third round. The more washing rounds were used, the more sequences containing frame shifts or predominantly hydrophobic amino acids (i.e., more than one tryptophan residues) were found. This was the case for the 4/8/10 and 10/10/10 variant. If, on the other hand, fewer washing cycles were used (2/4/6), a large proportion of pHAL30 origin empty vector containing no peptide encoding sequence was found, but also a few desired peptide sequences as potential binders. Peptide 1 was one of those useful sequences.

2.2. Fluorescence polarization assay

The fluorescence polarization assay has been established by Maurer et al. [10] Fluorescence polarization was recorded using a CLARIOstar microplate reader (BMG LABTECH, Ortenberg, Germany) with an extinction filter at 485 nm and emission filter at 520 nm. Gain adjustment was performed before starting each measurement to achieve maximum sensitivity. The FP values were measured in millipolarization units (mP). The assay was performed two times in duplicates and the IC$_{50}$ value was calculated using sigmoidal logistic fit in Origin. Fluorescein-labeled RNA (for *Yersinia* CsrA: 5'-UUCACGGAGAA[flc]; for *E. coli* CsrA: 5'-AGACAAGGAUGU [flc]) was obtained from Sigma Aldrich in HPLC purity. The results of the dose-dependent measurement are shown in Figure S3 and S4.

A 20 mM peptide in DMSO stock solution was diluted with assay buffer (10 mM HEPES, 150 mM NaCl, 0.005% (v/v) Tween-20, ad DEPC-treated H$_2$O (RNase free water), pH 7.4) in a way that 3 mM peptide in 15% DMSO was achieved (21 μL 20 mM peptide in DMSO + 119 μL assay buffer). Afterwards, a 1:2 dilution series containing 12 steps was utilized by diluting 70 μL of assay buffer containing 15% DMSO with 70 μL of the peptide in assay buffer with 15% DMSO from this solution (figure S2), starting from 3 mM ended in 1,46 μM. Using a 12-channel pipette, 10 μL of each concentration were transferred to a 384 well microtiter plate (black, flat bottom, Greiner Bio-One) in two replicates and another 10 μL of 1.2 μM (2.4 μM for *E. coli* CsrA) of the corresponding CsrA-biot-His$_6$ protein (in assay buffer) were added to each well and quickly centrifuged to be preincubated for 1 h on a Duomax 1030 shaker under light

exclusion. 10 μM fluorescein-labeled RNA (RNAflc) was diluted with assay buffer to a concentration of 45 nM obtaining an end concentration of 15 nM in the assay. After short centrifugation the plate was incubated for 1.5 h on the shaker under light exclusion. The final concentrations in the assay were 400 nM (800 nM for *E. coli* CsrA) CsrA-biot-His$_6$ (monomer concentration), 5% DMSO, 15 nM RNAflc and 1000 μM−0.49 μM peptide.

Furthermore, a high control was prepared to check for the homogeneity of fluorescence for the complex between protein and RNAflc, a low control to verify the homogeneity of fluorescence for the free RNAflc as well as a blank to exclude any deviation due to the matrix of the assay. For the high control components were 10 μL of 15% DMSO in assay buffer, 10 μL of protein and 10 μL of RNAflc, for the low control corresponding 10 μL of 15% DMSO in assay buffer, 10 μL of assay buffer and 10 μL of RNAflc and the blank consisted of 10 μL of 15% DMSO in assay buffer and two times 10 μL of assay buffer. These three controls were measured in 24-lets.

Moreover, a fluorescence control was performed for the peptides measured to check for the possibility of fluorescence quenching. Therefore, the first component was 10 μL of the dilution series of the corresponding peptide, second component was 10 μL of assay buffer and third component was 10 μL of RNAflc.

Thereby, fluorescence intensity was calculated by determination of the sum of blank corrected based on raw data parallel and perpendicular for the highest concentration on the one hand and for the lowest concentration on the other hand. Afterwards, the average of these two values was determined and the deviation from fluorescence intensity to the average value should be under 20% for no fluorescence quenching. This was the reason why the 1000 μM and 500 μM value was not included in the assay for **3d** and **5a**.

2.3. Microscale thermophoresis assay (MST)

The MST assay was performed according to the protocol of the Monolith NT™ His-Tag labelling Kit RED-tris-NTA and was used for Peptide **1** only. The *Yersinia* CsrA-biot-His$_6$ monomer concentration was adjusted with assay buffer (10 mM HEPES, 150 mM NaCl, 0.005% (v/v) Tween-20, ad DEPC-treated H$_2$O (RNase free water), pH 7.4) to 200 nM in a volume of 100 μL, mixed with 100 μL 100 nM dye (Nano RED) and incubated for 30 min in the dark at room temperature. The sample was centrifuged for 10 min at 4 °C and 15000 g. This was the ready-labeled protein. A 20 mM peptide DMSO stock solution was diluted with assay buffer to 2 mM, that the highest end concentration in the assay was 1000 μM with 5% DMSO. 20 μL of the 2 mM peptide was transferred into a first PCR tube and 10 μL of assay buffer containing 10% DMSO was transferred into each next PCR tube 2−16. For the serial dilution series of the peptide, 10 μL of the ligand from tube 1 were transferred to tube 2 with a pipette and mixed by pipetting up-and-down several times. The procedure was repeated for tube 3−16 and 10 μL from tube 16 were discarded. Finally, 10 μL of the labeled protein were added to each PCR tube, mixed with a pipette and incubated in the dark for 45 min. All 16 dilutions were loaded into Monolith NT™ Standard Capillaries and measured in the Monolith NT.115™ device with 60% excitation power and 40% MST power. The protein concentration in the assay was 50 nM. The assay was performed three times in duplicates and the K$_d$ value of 10.5 ± 1.4 μM was calculated using sigmoidal logistic fit in Origin. The results from the MST assay for peptide **1** can found in figure S5.

2.4. Peptide synthesis and macrocyclization

2.4.1. General information

All resins were purchased from Rapp Polymere. The azide/alkyne building blocks Fmoc-ʟ-azidoalanine (Fmoc-Aza-OH), Fmoc-

ʟ-propargylglycine (Fmoc-Pra-OH) and Fmoc-ʟ-homoazidoalanine (Fmoc-Aha-OH) were purchased from Carl Roth VG. Fmoc-Val-OH, Fmoc-Ala-OH, Fmoc-Glu(O*t*Bu)-OH, Fmoc-Leu-OH and Fmoc-Cys(Trt)-OH were purchased from Novabiochem. Fmoc-Ser(*t*Bu)-OH was purchased from TCI.

2.4.2. General Fmoc-SPPS procedure

Most peptides were synthesized manually *via* solid phase peptide synthesis (SPPS) using Fmoc chemistry. The resin was swollen for 30 min in DMF. For Fmoc deprotection piperidine/DMF (1:4, v:v) was added and shaken for 5 min, twice. It was then washed five times with DMF followed by the second round of adding piperidine/DMF (1:4) with incubating 5 min on a shaker. It was washed five times with DMF, five times with DCM and again one time with DMF. We used double coupling for each amino acid. The amino acid (4.0 eq.) was solved in DMF together with 3.9 eq 3-[Bis(dimethylamino)methyliumyl]-3*H*-benzotriazol-1-oxide hexafluorophosphate (HBTU) followed by adding 8.0 eq. *N*-Ethyl-*N*-(propan-2-yl)propan-2-amine (DIPEA). This solution was pre-activated for 5 min on a shaker. The activated solution was added to the resin and incubated for 1 h on a shaker. After washing five times with DMF, it was added an activated amino acid/HBTU/DIPEA/DMF solution again and incubated 1 h on a shaker. The resin was washed five times with DMF and five times with DCM. This was followed by two deprotection cycles and two coupling cycles of the next amino acid.

2.4.3. General acetylation procedure

For Acetylation, DMF/DIPEA/Ac$_2$O (12:8:5, v:v:v) was added to the resin and shaken for 0.5 h. Then it was washed five times with DMF, five times with DCM and again one time with DMF.

2.4.4. General cleavage procedure

For protein cleavage from the solid support and removal of the side chain protecting groups a cleavage cocktail containing trifluoroacetic acid (TFA)/triisopropylsilane (TIS)/H$_2$O/anisole (95:2:2:1, v:v:v:v) with a spatula tip of dithiothreitol (DTT) was added to the resin and incubated 2.5−3.0 h on a shaker. The liquid was collected and TFA was removed under reduced pressure, followed by precipitation with cold (−20 °C) methyl *tert*-butyl ether (MTBE). The crude peptide was gained by centrifugation (4600 rpm, 4 °C, 10 min) followed by MTBE washing (3x) and repeated centrifugation.

2.4.5. General cyclisation procedure

For disulfide cyclisation the crude lyophilized peptide was dissolved in H$_2$O/ACN (1:1, v:v) with a concentration of 1 mg peptide per 1 mL solvent and 1−3% DMSO was added. The pH was adjusted to 7.7 using 1 M aq. ammonium carbonate solution. The solution was stirred for 1−4 days. The reaction was monitored by LC-MS.

2.4.6. General preparative HPLC procedure

The purification was done with a DIONEX UltiMate 3000 UHPLC$^+$ focused (Thermo Scientific), containing pump, diode array detector, and automated fraction collector. We used a VP 250/10 NUCLEODUR C18 Gravity, 5 μm (Macherey-Nagel) column with a gradient from 10 to 50% solvent B over 25 min (solvent A: H$_2$O (0.05% formic acid), solvent B: ACN (0.05% formic acid)) and a 5 mL/min flowrate. Pure fractions were checked by LC-MS, combined and lyophilized.

3. Results and discussion

3.1. Phage display

We devised a strategy to screen a library of disulfide-constrained heptapeptides covering a mass range between 548

3

and 1193 Da *via* phage display (Scheme 1) [28]. The use of phage-encoded libraries displaying millions of compound variants [30] has proven to be an excellent method for finding novel binders for several targets [31]. An important example is the search for small antibody fragments, so called single-chain variable fragments (scFvs) for any desired target [32,33]. Phage display can also be transferred to libraries encoding for short peptides [31,34]. This method allows screening of whole peptide libraries to find potential binders for a given target [35].

Our self-designed phage library encodes for a peptide library with the general structure XCXXXCX (2.48 · 10^6 variants) [28]. It contains two cysteine residues at fixed positions, which form a disulfide bond under oxidative conditions; X encodes for any amino acid except cysteine. This design provides the means to identify very small peptides with a mass range of around 550−1200 Da that are rigidified by a well-defined macrocyclization motif. We screened this library against immobilized *Yersinia* CsrA (biotinylated and His-tagged CsrA construct CsrA_biot_His$_6$; more details on phage display and CsrA expression can be seen in Supporting Information and a published protocol) bound to a streptavidin-coated ELISA well. After three rounds of panning, phage binding with high affinity were separated on agar plates. After sequencing of 32 clones, we identified one sequence as a

potential CsrA binder. The criteria for selection were intact sequences and avoidance of a high tryptophan content (more than two Trp), which usually leads to unspecific binding [36]. Notably, the selected sequence contained a glutamic acid residue - a feature we deemed plausible as anionic carboxyl groups should be of benefit for binding the positively charged surface of CsrA possessing a high content of basic amino acids.

3.2. First evaluation of peptidic hit

This peptidic hit (**1**) was synthesized by solid phase peptide synthesis (SPPS) in disulfide-cyclized, *N*-terminally acetylated and *C*-terminally amidated form. These modifications were chosen because the sequence is presented within a peptide backbone extending beyond its *N*- and *C*-termini on the phage during the panning experiment. The peptide was characterized by LC-MS, HRMS and NMR (Supporting Information). Using an established fluorescence polarization assay [10] (2.4), peptide **1** was tested for its ability to displace mRNA from CsrA. In this assay a fluorescein (flc)-labeled RNA (5'-UUCACGGAGAA[flc]) and CsrA_biot_His$_6$ were used to probe the protein-RNA interaction. The labeled RNA was successfully displaced by peptide **1** with an IC$_{50}$ value in the micromolar range (6.9 ± 1.3 μM, Fig. 2). This peptide amongst the
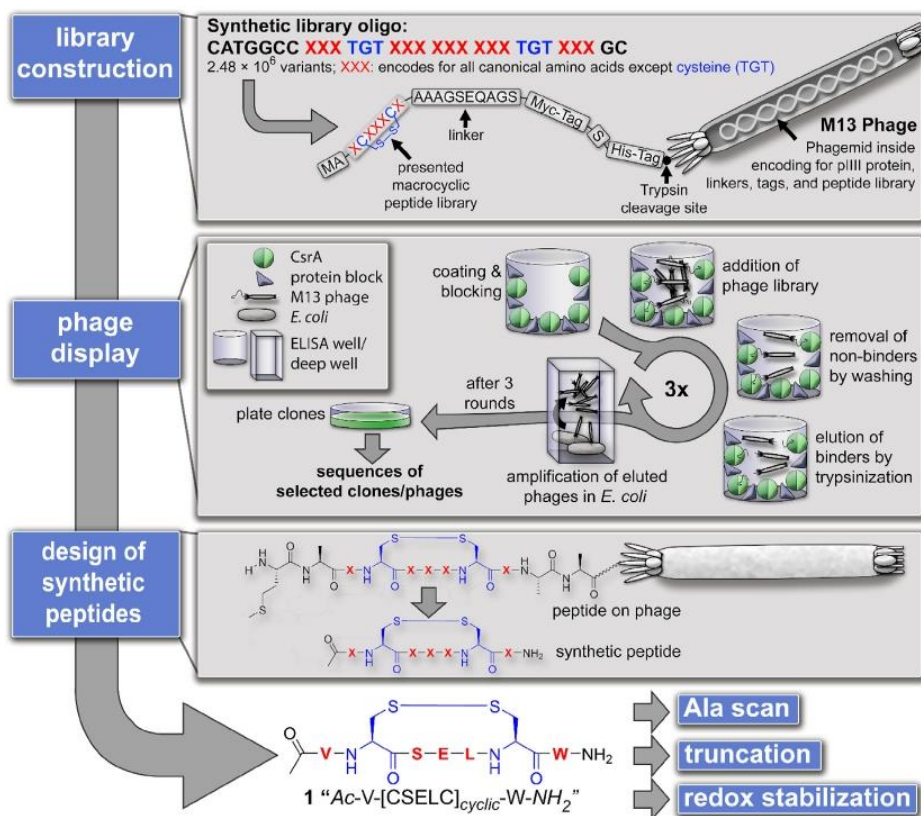
Scheme 1. Phage display-based selection process from library design to peptide hit identification. Library construction: An oligomer was constructed to code for the subsequent peptide library. After cloning and packaging in M13 phage, genotype and phenotype are coupled by presenting the encoded peptide including tags and linkers simultaneously. Phage display: The peptide phage library was used in the panning process, in which the enrichment of potential CsrA binders is achieved. Sequences of the bound peptides were identified by sequencing the coding phage gene. Design of synthetic peptides: Representation of a phage from panning to which a general library peptide including peptide backbone is linked, compared to the synthesized peptides. Here, *N*-terminal acetylation and *C*-terminal amidation simulate the peptide backbone. Furthermore: A representation of the selected peptide **1** is given, which was characterized in more detail as a "hit" in the context of this paper.

4

most potent compounds discovered against CsrA to date and is readily synthetically accessible. Previously identified natural products such as MM14 and tubulysin Ar-672, have shown similar potency (4 ± 1 μM and 11 ± 1 μM, respectively) [10], but are much more challenging to synthesize.

When the assay was conducted in the presence of 5 mM DTT, peptide **1** lost its activity almost completely. Under these conditions the disulfide bond is reduced and the macrocycle linearized. Thus, we concluded that the conformational constraint induced by the disulfide bond is essential for activity. This observation also supports a conformation-specific (structure-dependent) interaction between the peptide and CsrA. Additionally, a microscale thermophoresis (MST) assay was performed with peptide **1** and CsrA_-biot_His$_6$ yielding a K_d of 10.5 ± 1.4 μM (2.5). This assay further supports a direct specific interaction between the peptide and CsrA.

3.3. Alanine scan and truncation

To gain further insights into the underlying structure-activity relationship (SAR) of peptide **1** we synthesized an array of derivatives by Fmoc-SPPS, oxidized them with DMSO and tested for inhibitory activity in the fluorescence polarization assay. The resulting IC$_{50}$ values are listed in Table 1.

To investigate the importance of the *N*-terminal acetylation as well as the *C*-terminal amide, peptides 2a and 2b were synthesized, respectively. We observed slightly increased IC$_{50}$ values (27.6 ± 4.0 μM and 17.4 ± 2.0 μM, respectively) indicating that both modifications contribute to the overall affinity of peptide 1. To identify possible interaction hotspots, an alanine scan of peptide 1 was performed.

As expected, activity was abolished when both cysteine residues were replaced by alanine (3a) corroborating our earlier findings when using DTT to linearize peptide 1 (Fig. 1). Similarly, when Ser3 (3c) or Trp7 (3f) were changed to alanine, dramatic losses of activity
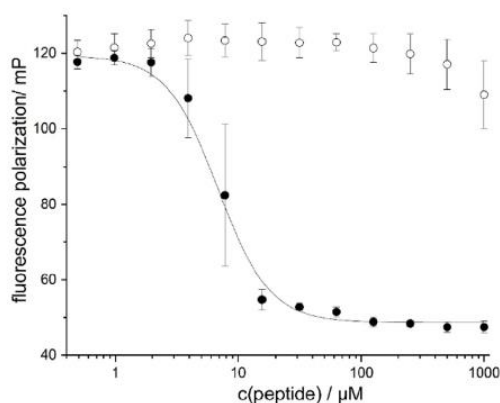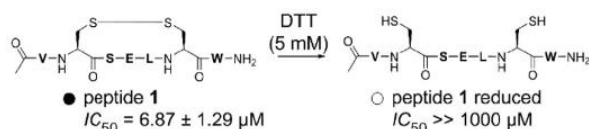
Fig. 2. Displacement of RNAflc from CsrA_biot_His$_6$ with peptide **1** and its reduced derivative measured via fluorescence polarization. Data shown are from two independent experiments measured in duplicate and were fitted to a sigmoidal logistic, Levenberg Marquardt inhibition model (solid line). The results of peptide **1** (filled circles) as well as peptide **1** in the presence of 5 mM DTT in the assay are shown (open circles).

Table 1
Peptides **1**–**4b** with corresponding IC$_{50}$ values obtained from the fluorescence polarization assay and their activity relative to peptide **1**.

Entry	Sequence[a]	IC$_{50}$/μM[b] *Y. pseudotuberculosis*	Relative activity[c]
1	*Ac*-V-[CSELC]$_{cyclic}$-W-*NH$_2$*	6.9 ± 1.3	1
2a	**H**-V-[CSELC]$_{cyclic}$-W-*NH$_2$*	27.6 ± 4.0	0.25
2b	*Ac*-V-[CSELC]$_{cyclic}$-W-**OH**	17.4 ± 2.0	0.4
3a	*Ac*-V**ASELAW**-*NH$_2$*	≫1000	—
3b	*Ac*-**A**-[CSELC]$_{cyclic}$-W-*NH$_2$*	114 ± 8	0.06
3c	*Ac*-V-[C**A**ELC]$_{cyclic}$-W-*NH$_2$*	>1000	—
3d	*Ac*-V-[CS**A**LC]$_{cyclic}$-W-*NH$_2$*	22.8 ± 0.7	0.3
3e	*Ac*-V-[CSE**A**C]$_{cyclic}$-W-*NH$_2$*	57.9 ± 2.1	0.12
3f	*Ac*-V-[CSELC]$_{cyclic}$-**A**-*NH$_2$*	>500	—
4a	*Ac*- − -[CSELC]$_{cyclic}$-W-*NH$_2$*	128 ± 4	0.05
4b	*Ac*-V-[CSELC]$_{cyclic}$- − -*NH$_2$*	≫1000	—

[a] Differences relative to peptide **1** are shown in bold or as "−" for deletions. Each peptide is disulfide-cyclized (except **3a**) over the cysteines.
[b] Standard error of the sigmoidal curve fit is given (two independent experiments, measured in duplicates).
[c] Relative activity for a peptide **x** is given as the ratio IC$_{50}$(peptide **1**)/IC$_{50}$(peptide **x**).

were observed. Therefore, the interactions mediated by the serine and tryptophan sidechains are essential hotspots for high affinity. Furthermore, the Ala-scan allowed us to conclude that substitution of residues Val1 (**3b**), Glu4 (**3d**) or Leu5 (**3e**) has a less pronounced effect on activity, showing IC$_{50}$ values of 114 ± 8 μM, 22.8 ± 0.7 μM, and 57.9 ± 2.1 μM, respectively. In the case of the Glu4Ala mutation, this result was surprising. CsrA is an RNA-binding protein possessing a positively charged surface area due to an abundance of lysine and arginine residues. Hence, the presence of the carboxylic acid function in peptide **1** hinted at a potential salt bridge as an important contribution to affinity. If the proposed ionic interaction between Glu4 and the basic amino acid sidechains of CsrA was optimally positioned, a dramatic loss of affinity would have been expected for compound **3d**. As this was not the case, this position should be investigated in more detail in future optimization efforts.

Two truncated versions were tested for inhibitory activity to check whether further reduction in size is possible. A version without Val1 (**4a**) had an IC$_{50}$ of 128 ± 4 μM, which is comparable to the value obtained for the Val1Ala mutant, **3b** (114 ± 8 μM). If tryptophan is omitted (**4b**), the activity in the measured concentration range is completely lost and in line with our findings with the Trp7Ala mutant (**3f**). Hence, we conclude that the complete seven amino acid sequence is required for high activity.

3.4. Disulfide replacement by triazole bridge

In a final step, we sought to protect peptide **1** from reductive linearization, which we consider essential for achieving intracellular activity. To this end, we made use of the "triazole bridge" approach [37] and replaced the cysteine residues with non-natural amino acids bearing alkyne and azide functions in their sidechain for facile click chemistry-based macrocyclization [38−40]. Notably, this strategy provides selective access to either a 1,4-disubstituted or 1,5-disubstituted bridging motif depending on whether copper(I)- or ruthenium(II)-catalyzed azide-alkyne cycloaddition is applied (abbreviated CuAAC or RuAAC, respectively). This method had recently been used, with great success, by the groups of Tomassi et al. [41], Tala et al. [42] and Pacifico et al. [43] to generate redox stable derivates of disulfide bridge containing peptides. By this means, several different triazole-bridged peptides were generated (Table 2). The linear precursor peptides were synthesized using commercially available building blocks Fmoc-protected propargylglycine (Fmoc-Pra-OH) and Fmoc-protected azidoalanine

(Fmoc-Aza-OH) or Fmoc-protected azidohomoalanine (Fmoc-Aha-OH). In-solution CuAAC macrocyclization of the unprotected peptides in separate reactions delivered three 1,4-disubstituted 1,2,3-triazole variants (**5a-5c**), which were characterized by LC-MS, HRMS and NMR (Supporting Information) and tested in the fluorescence polarization assay (Table 2). **5a**, originating from an azidoalanine-bearing precursor, showed an IC_{50} of 35.3 ± 0.6 μM, which correlates to a moderate 5-fold reduction in potency compared to the disulfide counterpart **1**. Installing an elongated macrocyclization motif by using azidohomoalanine instead (**5c**) leads to a further reduction of activity (76.0 ± 3.3 μM). Changing the orientation of the triazole ring by switching positions of the propargylglycine and azidoalanine residues (**5b**) resulted in an IC_{50} of 92.8 ± 4.0 μM.

Previous work on a 14-amino acid, backbone cyclic protease inhibitor peptide SFTI-1, demonstrated the utility of 1,5-disubstituted bridging motifs, which are installed via RuAAC in solution or on resin [37,44]. In the case of our current CsrA-RNA-interaction inhibitor **1**, this strategy was surprisingly not beneficial. Macrocyclic peptide **6a** achieved only an IC_{50} of 178 ± 12 μM. If the azidoalanine in position 2 was replaced by azidohomoalanine (**6b**), the IC_{50} value increased even further to 337 ± 34 μM. Finally, exchanging the positions of Aha and Pra (**6c**) did not show any significant difference in comparison to **6b** (IC_{50} = 309 ± 15 μM).

To demonstrate the potential for a broader anti-Gram-negative activity we tested peptide **1** and our triazole-stabilized derivatives against the *E. coli* and the *Pseudomonas. aeruginosa* homologs of CsrA (RsmA, Table 2). Surprisingly, disulfide-cyclized inhibitor **1** showed a reduced activity (IC_{50} (*E. coli*) = 182 ± 67 μM, IC_{50} (*P. aeruginosa*) = 272 ± 68 μM), while the 1,4-disubstituted triazoles **5a**, **5b** and **5c** now outperformed the parent peptide (IC_{50} (*E. coli*) = 4.9 ± 0.9 μM, 6.8 ± 1.5 μM, and 3.4 ± 0.6 μM, IC_{50} (*P. aeruginosa*) = 20 ± 5.4 μM, 22.8 ± 5.0 μM, and 30.2 ± 3.2 μM). The 1,5-disubstituted congeners again showed reduced activity compared to their 1,4-counterparts, albeit still being more active than the disulfide compound **1**. Considering the high sequence identity between CsrA from *Yersinia* and *E.coli* (95%) [45], it is fair to assume that this finding provides evidence for the potential site of interaction of the macrocyclization motif for our inhibitor scaffold. The only differences in amino acid sequence between the *Y. pseudotuberculosis* and *E. coli* proteins are at distinct residues of the *C*-terminus, including Pro58Gln, Thr59Ser, and Thr60Ser, respectively (see sections regarding protein expression in the Supporting Information). This region is also close to the protein-RNA-interaction interface (Fig. 3). Hence, we hypothesize that the inhibitor scaffold covers an area encompassing interactions to both sites (*C*-terminus and RNA-binding site). Unfortunately, attempts to co-crystallize the peptide with CsrA have not been successful to date. To gain access to structural information, we solved the structure of peptide **1** by NMR (PDB ID 7M7X, BMRB ID 30895, Fig. 3a and Supporting Information). With this ligand structure ensemble in hand we performed a docking experiment based on a *Y. pseudotuberculosis* CsrA homology model derived from a protein-RNA complex determined by NMR [46]. The result of the docking experiment is shown in Fig. 3b (see also Supporting Information). The binding pose of peptide **1** is in line with the SAR derived via the Ala-scan and truncation experiments. For example, the side chains of "hot spot" residues Ser3 and Trp7 form key contacts with Lys38 and Val40, while the other residues are primarily involved in backbone-based interactions (Fig. 3c). Although this pose will need further validation in future studies, it provides a basis for explaining the observed differences between *Y. pseudotuberculosis* and *E. coli* CsrA. We hypothesize that the differences in activities seen for compounds in Table 2 are potentially resulting from the Pro58Gln mutation changing the *C*-terminal interaction site from a rather hydrophobic environment to a more polar one, which might favor the hydrogen acceptor functions of the triazole. Implementation of a 1,5-motif (6a − 6c), however, could result in steric clashes between the ligand and the protein target rendering them less effective in this scenario.

Along similar lines, the decreased sequence identity between CsrA (*Y. pseudotuberculosis*, *E. coli*) and RsmA (*P. aeruginosa*) of 85% (*E. coli − P. aeruginosa*) or 86% (*Y. pseudotuberculosis − P. aeruginosa*) arising again mainly at the *C*-terminal end might explain the different SAR observation made (Table 2, Supporting Information).

4. Conclusion

In summary, we have shown that the 1,4-disubstituted triazole bridging motif established in 5a is a suitable disulfide replacement that is active against *Y. pseudotuberculosis*, *E. coli* and *P. aeruginosa*

Table 2

Peptide **1** and triazole-bridged variants **5a-6c** with corresponding IC_{50} values obtained for *Yersinia* and *E. coli* CsrA and RsmA from *P. aeruginosa* from a fluorescence polarization assay.

Entry	Sequence	IC_{50}/μM[a] *Yersinia*	IC_{50}/μM[a] E. Coli	IC_{50}/μM[a] *P. aeruginosa*
1	*Ac*-V-[CSELC]*cyclic*-W-*NH₂*	6.9 ± 1.3	182 ± 67	272 ± 68
5a		35.3 ± 0.6	4.9 ± 0.9	20 ± 5.4
5b		92.8 ± 4.0	6.8 ± 1.5	22.8 ± 5.0
5c		76.0 ± 3.3	3.4 ± 0.6	30.2 ± 3.2
6a		178 ± 12	48.1 ± 16	50 ± 15
6b		337 ± 34	51.6 ± 27.1	44.4 ± 9.7
6c		309 ± 15	83.4 ± 47.1	37 ± 17

[a] Standard error of the sigmoidal curve fit is given (two independent experiments, measured in duplicate).
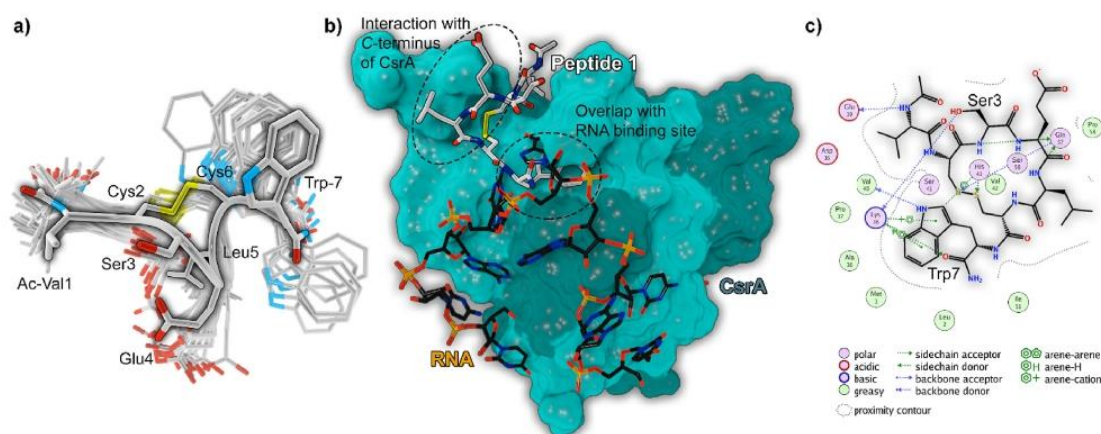
Fig. 3. *In silico* analysis of the peptide-CsrA interaction. a) Overlay of 20 NMR-derived solution structures of peptide 1 (PDB ID 7M7X, BMRB ID 30895) showing the peptide backbone as a tube and highlighting conformer 1 (entry 1 in pdb) for clarity. b) Depiction of docking-derived interaction hypothesis highlighting key interaction sites. Carbons of peptide 1 are shown in white and RNA carbons in black. Surface of the two CsrA chains shown in light cyan and dark cyan. c) 2D interaction profile of binding hypothesis for peptide 1. "Hot spot" residues identified via Ala-scan (Ser3 and Trp7) are indicated.

CsrA. In combination with our phage display-based screening methodology, we have provided a generic approach towards the identification, initial qualification, and subsequent redox-protection of short macrocyclic peptides as protein-RNA-interaction inhibitors. The phage display methodology proved to be a rapid approach towards identification of the first macrocyclic peptide able to disrupt the CsrA-RNA interaction. The starting scaffold peptide **1** was thoroughly characterized by fluorescence polarization-based functional activity tests as well as MST-based protein binding assay. Exchanging the disulfide bond with a redox stable 1,2,3-triazole bridge gave us active non-natural derivatives suitable for future cell-based assays. Contrary to previous studies, we have observed that in the current system the synthetically easier accessible 1,4-disubstituted 1,2,3-triazole was the superior disulfide mimic showing an IC_{50} value in the 2-digit micromolar range. Based on NMR-based solution structure determination of the native peptide sequence and docking experiments structure-guided optimization can now be attempted. This novel scaffold serves as a suitable starting point for the generation of high potency CsrA inhibitors, also because it is applicable against CsrA from further bacterial species with high medical need (*P. aeruginosa*).

Author contribution

The author contributions can be found in the supporting information.

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejmech.2022.114148.

References

[1] M. Jemal, T. Deress, T. Belachew, Y. Adem, Antimicrobial resistance patterns of bacterial isolates from blood culture among HIV/AIDS patients at felege hiwot referral hospital, northwest Ethiopia, Internet J. Microbiol. 2020 (2020) 8893266, https://doi.org/10.1155/2020/8893266.
[2] H. Inoue, R. Minghui, Antimicrobial resistance: translating political commitment into national action, Bull. World Health Organ. 95 (2017) 242, https://doi.org/10.2471/BLT.17.191890.
[3] Z. Kmietowicz, Few novel antibiotics in the pipeline, WHO warns, BMJ 358 (2017) j4339, https://doi.org/10.1136/bmj.j4339.
[4] S. Vasoo, J.N. Barreto, P.K. Tosh, Emerging issues in gram-negative bacterial resistance: an update for the practicing clinician, Mayo Clin. Proc. 90 (2015) 395–403, https://doi.org/10.1016/j.mayocp.2014.12.002.
[5] Y. Irie, M. Starkey, A.N. Edwards, D.J. Wozniak, T. Romeo, M.R. Parsek, Pseudomonas aeruginosa biofilm matrix polysaccharide Psl is regulated transcriptionally by RpoS and post-transcriptionally by RsmA, Mol. Microbiol. 78 (2010) 158–172, https://doi.org/10.1111/j.1365-2958.2010.07320.x.
[6] A.M. Nuss, A.K. Heroven, P. Dersch, RNA regulators: formidable modulators of Yersinia virulence, Trends Microbiol. 25 (2017) 19–34, https://doi.org/10.1016/j.tim.2016.08.006.
[7] F.M. Barnard, M.F. Loughlin, H.P. Fainberg, M.P. Messenger, D.W. Ussery, P. Williams, P.J. Jenks, Global regulation of virulence and the stress response by CsrA in the highly adapted human gastric pathogen Helicobacter pylori, Mol. Microbiol. 51 (2004) 15–32, https://doi.org/10.1046/j.1365-2958.2003.03788.x.
[8] A.H. Potts, C.A. Vakulskas, A. Pannuri, H. Yakhnin, P. Babitzke, T. Romeo, Global role of the bacterial post-transcriptional regulator CsrA revealed by integrated transcriptomics, Nat. Commun. 8 (2017) 1596, https://doi.org/10.1038/s41467-017-01613-1.
[9] E.R. Morris, G. Hall, C. Li, S. Heeb, R.V. Kulkarni, L. Lovelock, H. Silistre, M. Messina, M. Cámara, J. Emsley, P. Williams, M.S. Searle, Structural rearrangement in an RsmA/CsrA ortholog of Pseudomonas aeruginosa creates a dimeric RNA-binding protein, RsmN, Structure 21 (2013) 1659–1671, https://doi.org/10.1016/j.str.2013.07.007.
[10] C.K. Maurer, M. Fruth, M. Empting, O. Avrutina, J. Hoßmann, S. Nadmid, J. Gorges, J. Herrmann, U. Kazmaier, P. Dersch, R. Müller, R.W. Hartmann, Discovery of the first small-molecule CsrA-RNA interaction inhibitors using biophysical screening technologies, Future Med. Chem. 8 (2016) 931–947, https://doi.org/10.4155/fmc-2016-0033.
[11] G. Sharma, S. Sharma, P. Sharma, D. Chandola, S. Dang, S. Gupta, R. Gabrani, Escherichia coli biofilm: development and therapeutic strategies, J. Appl. Microbiol. 121 (2016) 309–319, https://doi.org/10.1111/jam.13078.
[12] J. Timmermans, L. van Melderen, Post-transcriptional global regulation by CsrA in bacteria, Cell, Mol. Life Sci. 67 (2010) 2897–2908, https://doi.org/10.1007/s00018-010-0381-z.
[13] E. van Assche, S. van Puyvelde, J. Vanderleyden, H.P. Steenackers, RNA-binding

proteins involved in post-transcriptional regulation in bacteria, Front. Microbiol. 6 (2015) 141, https://doi.org/10.3389/fmicb.2015.00141.

[14] N.A. Sabnis, H. Yang, T. Romeo, Pleiotropic regulation of central carbohydrate metabolism in Escherichia coli via the gene csrA, J. Biol. Chem. 270 (1995) 29096–29104, https://doi.org/10.1074/jbc.270.49.29096.

[15] H. Yakhnin, C.S. Baker, I. Berezin, M.A. Evangelista, A. Rassin, T. Romeo, P. Babitzke, CsrA represses translation of sdiA, which encodes the N-acylho-moserine-L-lactone receptor of Escherichia coli, by binding exclusively within the coding region of sdiA mRNA, J. Bacteriol. 193 (2011) 6162–6170, https://doi.org/10.1128/JB.05975-11.

[16] B.L Wei, A.M. Brun-Zinkernagel, J.W. Simecka, B.M. Prüss, P. Babitzke, T. Romeo, Positive regulation of motility and flhDC expression by the RNA-binding protein CsrA of Escherichia coli, Mol. Microbiol. 40 (2001) 245–256, https://doi.org/10.1046/j.1365-2958.2001.02380.x.

[17] M.Y. Liu, G. Gui, B. Wei, J.F. Preston, L. Oakford, U. Yüksel, D.P. Giedroc, T. Romeo, The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in Escherichia coli, J. Biol. Chem. 272 (1997) 17502–17510, https://doi.org/10.1074/jbc.272.28.17502.

[18] Y. Tan, Z.-Y. Liu, Z. Liu, H.-J. Zheng, F.-L. Li, Comparative transcriptome analysis between csrA-disruption Clostridium acetobutylicum and its parent strain, Mol. Biosyst. 11 (2015) 1434–1442, https://doi.org/10.1039/c4mb00600c.

[19] J.E. Schultz, A. Matin, Molecular and functional characterization of a carbon starvation gene of Escherichia coli, J. Mol. Biol. 218 (1991) 129–140, https://doi.org/10.1016/0022-2836(91)90879-B.

[20] A.K. Dubey, C.S. Baker, K. Suzuki, A.D. Jones, P. Pandit, T. Romeo, P. Babitzke, CsrA regulates translation of the Escherichia coli carbon starvation gene, cstA, by blocking ribosome access to the cstA transcript, J. Bacteriol. 185 (2003) 4450–4460, https://doi.org/10.1128/jb.185.15.4450-4460.2003.

[21] D.W. Jackson, K. Suzuki, L. Oakford, J.W. Simecka, M.E. Hart, T. Romeo, Biofilm formation and dispersal under the influence of the global regulator CsrA of Escherichia coli, J. Bacteriol. 184 (2002) 290–301, https://doi.org/10.1128/jb.184.1.290-301.2002.

[22] A.K. Dubey, C.S. Baker, T. Romeo, P. Babitzke, RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction, RNA 11 (2005) 1579–1587, https://doi.org/10.1261/rna.2990205.

[23] X. Ren, R. Zeng, M. Tortorella, J. Wang, C. Wang, Structural insight into inhibition of CsrA-RNA interaction revealed by docking, molecular dynamics and free energy calculations, Sci. Rep. 7 (2017) 14934, https://doi.org/10.1038/s41598-017-14916-6.

[24] G.B. Santos, A. Ganesan, F.S. Emery, Oral administration of peptide-based drugs: beyond lipinski's rule, ChemMedChem 11 (2016) 2245–2251, https://doi.org/10.1002/cmdc.201600288.

[25] C.A. Lipinski, Rule of five in 2015 and beyond: target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions, Adv. Drug Deliv. Rev. 101 (2016) 34–41, https://doi.org/10.1016/j.addr.2016.04.029.

[26] B.C. Doak, B. Over, F. Giordanetto, J. Kihlberg, Oral druggable space beyond the rule of 5: insights from drugs and clinical candidates, Chem. Biol. 21 (2014) 1115–1142, https://doi.org/10.1016/j.chembiol.2014.08.013.

[27] A. Capecchi, M. Awale, D. Probst, J.-L. Reymond, PubChem and ChEMBL beyond Lipinski, Mol. Inform. 38 (2019), e1900016, https://doi.org/10.1002/minf.201900016.

[28] V. Jakob, S. Helmsing, M. Hust, M. Empting, Restriction-free construction of a phage-presented very short macrocyclic peptide library, Methods Mol. Biol. 2070 (2020) 95–113, https://doi.org/10.1007/978-1-4939-9853-1_6.

[29] J. Kügler, S. Wilke, D. Meier, F. Tomszak, A. Frenzel, T. Schirrmann, S. Dübel, H. Garritsen, B. Hock, L. Toleikis, M. Schütte, M. Hust, Generation and analysis of the improved human HAL9/10 antibody phage display libraries, BMC Biotechnol. 15 (2015) 10, https://doi.org/10.1186/s12896-015-0125-0.

[30] S. Dübel, Handbook of Therapeutic Antibodies, secondnd ed., John Wiley & Sons Incorporated, Weinheim, 2014.

[31] C.G. Ullman, L. Frigotto, R.N. Cooley, In vitro methods for peptide display and their applications, Brief. Funct, Genomics 10 (2011) 125–134, https://doi.org/10.1093/bfgp/elr010.

[32] A. Frenzel, T. Schirrmann, M. Hust, Phage display-derived human antibodies in clinical development and therapy, mAbs 8 (2016) 1177–1194, https://doi.org/10.1080/19420862.2016.1212149.

[33] S. Dübel, O. Stoevesandt, M.J. Taussig, M. Hust, Generating recombinant antibodies to the complete human proteome, Trends Biotechnol. 28 (2010) 333–339, https://doi.org/10.1016/j.tibtech.2010.05.001.

[34] J. Kügler, J. Zantow, T. Meyer, M. Hust, Oligopeptide m13 phage display in pathogen research, Viruses 5 (2013) 2531–2545, https://doi.org/10.3390/v5102531.

[35] V. Baeriswyl, C. Heinis, Phage selection of cyclic peptide antagonists with increased stability toward intestinal proteases, Protein Eng. Des. Sel. 26 (2013) 81–89, https://doi.org/10.1093/protein/gzs085.

[36] N.B. Adey, A.H. Mataragnon, J.E. Rider, J.M. Carter, B.K. Kay, Characterization of phage that bind plastic from phage-displayed random peptide libraries, Gene 156 (1995) 27–31, https://doi.org/10.1016/0378-1119(95)00058-E.

[37] M. Empting, O. Avrutina, R. Meusinger, S. Fabritz, M. Reinwarth, M. Biesalski, S. Voigt, G. Buntkowsky, H. Kolmar, Triazole bridge": disulfide-bond replacement by ruthenium-catalyzed formation of 1,5-disubstituted 1,2,3-triazoles, Angew Chem. Int. Ed. Engl. 50 (2011) 5207–5211, https://doi.org/10.1002/anie.201008142.

[38] X. Jiang, X. Hao, L. Jing, G. Wu, D. Kang, X. Liu, P. Zhan, Recent applications of click chemistry in drug discovery, Expet Opin. Drug Discov. 14 (2019) 779–789, https://doi.org/10.1080/17460441.2019.1614910.

[39] M. Roice, I. Johannsen, M. Meldal, High capacity poly(ethylene glycol) based amino polymers for peptide and organic synthesis, QSAR Comb. Sci. 23 (2004) 662–673, https://doi.org/10.1002/qsar.200420021.

[40] G. Appendino, S. Bacchiega, A. Minassi, M.G. Cascio, L. De Petrocellis, V. Di Marzo, The 1,2,3-triazole ring as a peptide- and olefinomimetic element: discovery of click vanilloids and cannabinoids, Angew. Chem. 119 (2007) 9472–9475, https://doi.org/10.1002/ange.200703590.

[41] S. Tomassi, A.M. Trotta, C. Ieranò, F. Merlino, A. Messere, G. Rea, F. Santoro, D. Brancaccio, A. Carotenuto, V.M. D'Amore, F.S. Di Leva, E. Novellino, S. Cosconati, L. Marinelli, S. Scala, S. Di Maro, Disulfide bond replacement with 1,4- and 1,5-disubstituted 1,2,3-triazole on C-X-C chemokine receptor type 4 (CXCR4) peptide ligands: small changes that make big differences, Chemistry 26 (2020) 10113–10125, https://doi.org/10.1002/chem.202002468.

[42] S.R. Tala, A. Singh, C.J. Lensing, S.M. Schnell, K.T. Freeman, J.R. Rocca, C. Haskell-Luevano, 1,2,3-Triazole rings as a disulfide bond mimetic in chimeric AGRP-melanocortin peptides: design, synthesis, and functional characterization, ACS Chem. Neurosci. 9 (2018) 1001–1013, https://doi.org/10.1021/acschemneuro.7b00422.

[43] S. Pacifico, A. Kerckhoffs, A.J. Fallow, R.E. Foreman, R. Guerrini, J. McDonald, D.G. Lambert, A.G. Jamieson, Urotensin-II peptidomimetic incorporating a non-reducible 1,5-triazole disulfide bond reveals a pseudo-irreversible covalent binding mechanism to the urotensin G-protein coupled receptor, Org. Biomol. Chem. 15 (2017) 4704–4710, https://doi.org/10.1039/C7OB00959C.

[44] A.M. White, S.J. de Veer, G. Wu, P.J. Harvey, K. Yap, G.J. King, J.E. Swedberg, C.K. Wang, R.H.P. Law, T. Durek, D.J. Craik, Application and structural analysis of triazole-bridged disulfide mimetics in cyclic peptides, Angew Chem. Int. Ed. Engl. 59 (2020) 11273–11277, https://doi.org/10.1002/anie.202003435.

[45] A.K. Heroven, K. Böhme, P. Dersch, The Csr/Rsm system of Yersinia and related pathogens: a post-transcriptional strategy for managing virulence, RNA Biol. 9 (2012) 379–391, https://doi.org/10.4161/rna.19333.

[46] O. Duss, E. Michel, N. Diarra dit Konté, M. Schubert, F.H.-T. Allain, Molecular basis for the wide range of affinity found in Csr/Rsm protein-RNA recognition, Nucleic Acids Res. 42 (2014) 5332–5346, https://doi.org/10.1093/nar/gku141.

32

Phage Display-based Discovery of Cyclic Peptides against the Broad Spectrum Bacterial Anti-Virulence Target CsrA.

Valentin Jakob,[1,2,†] Ben G. E. Zoller,[1,2,†] Julia Rinkes,[1,2] Yingwen Wu,[1,2] Alexander F. Kiefer,[1,2] Michael Hust,[3] Saskia Polten,[3] Andrew M. White,[4] Peta J. Harvey,[4] Thomas Durek,[4] David J. Craik,[4] Andreas Siebert,[5] Uli Kazmaier,[5] and Martin Empting[1,2,*]

[1]Department of Drug Design and Optimization (DDOP), Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS) - Helmholtz Centre for Infection Research (HZI), Campus E8.1, 66123 Saarbrücken, Germany.

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

[3]Technische Universität Braunschweig, Institut für Biochemie, Biotechnologie und Bioinformatik, Spielmannstr. 7, 38106 Braunschweig, Germany

[4]ARC Centre of Excellence for Innovations in Peptide and Protein Science, Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, 4072, Australia

[5]Institut für Organische Chemie Saarland University Campus C4.2, 66123 Saarbrücken, Germany

[†] These authors contributed equally to this work.

1

SUPPORTING INFORMATION

Table of Contents

34

Peptide Synthesis and Macrocyclization

***Ac*-V-[CSELC]**$_{cyclic}$**-W-*NH*$_2$ (1); 3-((4*R*,7*S*,10*S*,13*S*,16*R*)-16-((*S*)-2-acetamido-3-methylbutanamido)-4-(((*S*)-1-amino-3-(1*H*-indol-3-yl)-1-oxopropan-2-yl)carbamoyl)-13-(hydroxymethyl)-7-isobutyl-6,9,12,15-tetraoxo-1,2-dithia-5,8,11,14-tetraazacycloheptadecan-10-yl)propanoic acid.**

The linear precursor peptide was synthesized on a Fmoc Trp(Boc) TentaGel® S RAM resin (load: 0.2 mmol/g) in a 0.2 mmol scale manually using Fmoc strategy with two coupling cycles and two deprotection cycles as described in the "general Fmoc-SPPS procedure". The cleavage of the peptide from the resin was performed by the "general cleavage procedure" protocol, where 7 mL cleavage cocktail was used with an incubation of 3 h on a shaker. After lyophilization, 45.7 mg of crude linear product was achieved, which was used for the "general cyclisation procedure" for disulfide cyclisation of the cysteines for 2 days. For purification, the "general preparative HPLC procedure" has been performed. A yield of 2.68 mg pure (≥ 98 %) cyclized peptide (3.05 µmol, 1.53 % according to initial load of the resin) was obtained and characterized by LC-MS, ^{1}H-NMR, ^{13}C-NMR and HRMS. LC-MS: Column: Phenomenex Luna C18(2), gradient: MeCN/H$_2$O + 0.05 % HCOOH, 5 % MeCN increase to 50% MeCN in 5.1 min, flow rate: 0.6 mL/min, t_R= 3.75 min, m/z = 878.3 ([M+H]$^+$). HRMS (ESI+) m/z calculated for C$_{38}$H$_{55}$N$_9$O$_{11}$S$_2$ [M+H]$^+$ 878.3530; found: 878.3528.
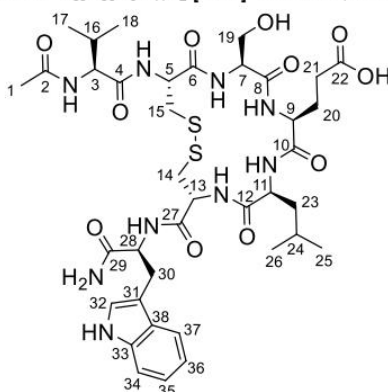
^{1}**H-NMR** (500 MHz, dmso-d6, δ in ppm): 10.82 (bs, 1 H, Indole NH), 8.41 (bs, 1 H, NH), 8.30 (d, ^3J = 7.80 Hz, 1 H, NH), 8.15 (d, ^3J = 8.2 Hz, 1 H, NH), 8.06 (d, ^3J = 6.8 Hz, 1 H, NH), 7.99 (d, ^3J = 6.6 Hz, 1 H, NH), 7.97–7.92 (m, 2 H, NH), 7.56 (d, $^3J_{37,36}$ = 7.9 Hz 1 H, H37), 7.30 (d, $^3J_{34,45}$ = 7.9 Hz, 2 H, H34, NH), 7.12 (m, 2 H, NH, 32H), 7.03 (t, $^3J_{35,36/34}$ = 7.3 Hz, 1 H, H35), 6.95 (t, $^3J_{36,35/37}$ = 7.4 Hz, 1 H, H36), 4.62-4.56 (m, 1 H, H5), 4.55-4.49 (m, 1H, 13-H), 4.43-4.37 (m, 1H, 28H), 4.28-4.24 (m, 1H, 7H), 4.19 (dd, ^3J = 8.8, 6.7 Hz, 1H, 3-H), 4-17-4.09 (m, 2H, 9H, 11H), 3.76-3.62 (m, 2H, 19H), 3.25 (dd, $^2J_{15a,15b}$ = 13.4, $^3J_{15a,5}$ = 4.2 Hz, 1H, H15$_a$), 3.13 (dd, $^2J_{30a,30b}$ = 13.4, $^3J_{30a,28}$ = 5.1 Hz, 1H, H30$_a$), 3-06-2.93 (m, 4H, H14, H15$_b$, H30$_b$), 2.23 (t, $^3J_{21,20}$ = 7.9 Hz, 2H, 21H), 2.00-1.90 (m, 3H, 20H, 16H), 1.88 (s, 3H, 1H), 1.60-1.53 (m, 1H, 24H), 1.52-1.47 (m, 2H, 23H), 0.87 (d, $^3J_{25,24}$, 3H, 25H), 0.85-0.80 (m, 9H, 17H, 18H, 26H).

^{13}**C-NMR** (126 MHz, dmso-d6, δ in ppm): 173.1 (C22), 172.2 (C29), 171.6 (C4), 171.5 (C10), 170.2 (C6, C8 or C12), 169.8 (C6, C8 or C12), 169.5 (C-2), 169.1 (C6, C8 or C12), 136.1 (C33), 127.4 (C38), 123.6 (C32), 120.9 (C35), 118.5 (C37), 118.3 (C36), 111.3 (C34), 110.1 (C31), 61.2 (C19), 57.6 (C3), 55.9 (C7), 54.0 (C99), 53.6 (C28), 52.8 (C13), 52.5 (C5), 52.0 (C-11), 41.9 (C15), 40.7 (C14), 40.1 (C23), 30.6 (C21), 27.7 (C30), 26.9(C20), 24.3 (C24), 23.1 (C25), 22.6 (C1), 21.6 (C26), 19.3 (C17 or C18), 18.2 (C17 or C18).

***H*-V-[CSELC]**$_{cyclic}$**-W-*NH*$_2$ (2a); 3-((4*R*,7*S*,10*S*,13*S*,16*R*)-4-(((*S*)-1-amino-3-(1*H*-indol-3-yl)-1-oxopropan-2-yl)carbamoyl)-16-((*S*)-2-amino-3-methylbutanamido)-13-(hydroxymethyl)-7-isobutyl-6,9,12,15-tetraoxo-1,2-dithia-5,8,11,14-tetraazacycloheptadecan-10-yl)propanoic acid.**

The linear precursor peptide was synthesized on a Fmoc Trp(Boc) TentaGel® S RAM resin (load: 0.2 mmol/g) in a 0.1 mmol scale manually using Fmoc strategy with two coupling cycles and two deprotection cycles as described in the "general Fmoc-SPPS procedure". The cleavage of the peptide from the resin was done following the "general cleavage procedure" protocol, where 7 mL cleavage cocktail was used with an incubation of 2.5 h on a shaker. After lyophilization, 87.4 mg of crude linear product was achieved, which was used for the "general cyclisation procedure" for disulfide cyclisation of the cysteines for 2 days. For purification, the "general preparative HPLC procedure" has been performed. A yield of 1.03 mg pure (≥ 93 %) cyclized peptide (1.23 µmol, 1.23 % according to initial load of the resin) was obtained. LC-MS: Column: Phenomenex Luna C18(2), gradient: MeCN/H$_2$O + 0.05 % HCOOH, 5 % MeCN increase to 50 % MeCN in 5.1 min, flow rate: 0.6 mL/min, t_R= 3.15 min, m/z = 834.4 ([M-H]$^-$). HRMS (ESI+) m/z calculated for C$_{36}$H$_{53}$N$_9$O$_{10}$S$_2$ [M-H]$^-$ 834.3279; found: 834.3311.

***Ac*-V-[CSELC]**$_{cyclic}$**-W-*OH* (2b); ((4*R*,7*S*,10*S*,13*S*,16*R*)-16-((*S*)-2-acetamido-3-methylbutanamido)-10-(2-carboxyethyl)-13-(hydroxymethyl)-7-isobutyl-6,9,12,15-tetraoxo-1,2-dithia-5,8,11,14-tetraazacycloheptadecane-4-carbonyl)-L-tryptophan.**

3

The linear precursor peptide was synthesized on a Fmoc Trp(Boc) TentaGel® S AC resin (load: 0.2 mmol/g) in a 0.05 mmol scale with a microwave-assisted peptide synthesizer (CEM Liberty Lite) using Fmoc strategy with two coupling cycles and two deprotection cycles, including acetylation of the *N*-terminus. The used coupling reagents were Oxyma (0.5 M) and DIC (0.25 M) in DMF. The cleavage of the peptide from the resin was done following the "general cleavage procedure" protocol, where 7 mL cleavage cocktail was used with an incubation of 2.5 h on a shaker. After lyophilization, 14.4 mg of crude uncyclized product was achieved, which was used for the "general cyclisation procedure" for disulfide cyclisation of the cysteines for 3 days. For purification, the "general preparative HPLC procedure" has been performed. A yield of 1.09 mg pure (≥ 98 %) cyclized peptide (1.24 µmol, 2.47 % according to initial load of the resin) was obtained. LC-MS: Column: Phenomenex Luna C18(2), gradient: MeCN/H_2O + 0.05 % HCOOH, 5 % MeCN increase to 50 % MeCN in 5.1 min, flow rate: 0.6 mL/min, t_R= 3.88 min, m/z = 879.3 ([M+H]$^+$). HRMS (ESI+) m/z calculated for $C_{38}H_{54}N_8O_{12}S_2$ [M-H]$^-$ 877.3224; found: 877.3251.

Ac*-VASELAW-*NH_2 (3a); (4*S*,7*S*,10*S*,13*S*)-13-(((*S*)-1-(((*S*)-1-amino-3-(1*H*-indol-3-yl)-1-oxopropan-2-yl)amino)-1-oxopropan-2-yl)amino)-4-methyl-1-oxopentan-2-yl)carbamoyl)-10-(hydroxymethyl)-4-isopropyl-7-methyl-2,5,8,11-tetraoxo-3,6,9,12-tetraazahexadecan-16-oic acid.

This linear peptide was synthesized on a Fmoc Trp(Boc) TentaGel® S RAM resin (load: 0.2 mmol/g) in a 0.1 mmol scale manually using Fmoc strategy with two coupling cycles and two deprotection cycles as described in the "General Fmoc-SPPS procedure" followed by the "general acetylation procedure". The cleavage of the peptide from the resin was done following the "general cleavage procedure" protocol, where 7 mL cleavage cocktail was used with an incubation of 2.5 h on a shaker. For purification, the "general preparative HPLC procedure" has been performed. A yield of 3.24 mg pure (≥ 98 %) peptide (3.97 µmol, 3.97 % according to initial load of the resin) was obtained. LC-MS: Column: Phenomenex Luna C18(2), gradient: MeCN/H_2O + 0.05 % HCOOH, 5 % MeCN increase to 50 % MeCN in 5.1 min, flow rate: 0.6 mL/min, t_R= 3.57 min, m/z = 816.5 ([M+H]$^+$). HRMS (ESI+) m/z calculated for $C_{38}H_{57}N_9O_{11}$ [M-H]$^-$ 814.4099; found: 814.4125.

Ac*-A-[CSELC]$_{cyclic}$-W-*NH_2 (3b); 3-((4*R*,7*S*,10*S*,13*S*,16*R*)-16-((*S*)-2-acetamidopropanamido)-4-(((*S*)-1-amino-3-(1*H*-indol-3-yl)-1-oxopropan-2-yl)carbamoyl)-13-(hydroxymethyl)-7-isobutyl-6,9,12,15-tetraoxo-1,2-dithia-5,8,11,14-tetraazacycloheptadecan-10-yl)propanoic acid. The linear precursor peptide was synthesized on a Fmoc Trp(Boc) TentaGel® S RAM resin (load: 0.2 mmol/g) in a 0.2 mmol scale manually using Fmoc strategy with two coupling cycles and two deprotection cycles as described in the "general Fmoc-SPPS procedure" followed by the "general acetylation procedure". The cleavage of the peptide from the resin was done following the "general cleavage procedure" protocol (in this case TFA/TIS/H_2O/Anisole 95:2:2:1 as cleavage cocktail), where 3 mL cleavage cocktail was used with an incubation of 3 h on a shaker. After lyophilization, the crude uncyclized product was used for the "general cyclisation procedure" for disulfide cyclisation of the cysteines for 3 days. For purification, the "general preparative HPLC procedure" has been performed. A yield of 1.07 mg pure (≥ 92 %) cyclized peptide (1.26 µmol, 0.63 % according to initial load of the resin) was obtained. LC-MS: Column: Phenomenex Luna C18(2), gradient: MeCN/H_2O + 0.05 % HCOOH, 5 % MeCN increase to 50 % MeCN in 5.1 min, flow rate: 0.6 mL/min, t_R= 3.38 min, m/z = 850.4 ([M+H]$^+$). HRMS (ESI+) m/z calculated for $C_{36}H_{51}N_9O_{11}S_2$ [M-H]$^-$ 848.3071; found: 848.3097.

Ac*-V-[CAELC]$_{cyclic}$-W-*NH_2 (3c); 3-((4*R*,7*S*,10*S*,13*S*,16*R*)-16-((*S*)-2-acetamido-3-methylbutanamido)-4-(((*S*)-1-amino-3-(1*H*-indol-3-yl)-1-oxopropan-2-yl)carbamoyl)-7-isobutyl-13-methyl-6,9,12,15-tetraoxo-1,2-dithia-5,8,11,14-tetraazacycloheptadecan-10-yl)propanoic acid. The linear precursor peptide was synthesized on a Fmoc Trp(Boc) TentaGel® S RAM resin (load: 0.2 mmol/g) in a 0.05 mmol scale with a microwave-assisted peptide synthesizer (CEM Liberty Lite) using Fmoc strategy with two coupling cycles and two deprotection cycles as described in the "general Fmoc-SPPS procedure", including acetylation of the *N*-terminus. The used coupling reagents were Oxyma (0.5 M) and DIC (0.25 M) in DMF. The cleavage of the peptide from the resin was done following the "general cleavage procedure" protocol, where 7 mL cleavage cocktail was used with an incubation of 2.5 h on a shaker. After lyophilization, the crude uncyclized product was used for the "general cyclisation procedure" for disulfide cyclisation of the cysteines for 7 days. For purification, the "general preparative HPLC procedure" has been performed. A yield of 3.44 mg pure (≥ 98 %) cyclized peptide (3.99 µmol, 7.98 % according to initial load of the resin) was obtained. LC-MS: Column: Phenomenex Luna C18(2), gradient: MeCN/H_2O + 0.05 % HCOOH, 5 % MeCN increase to 50 % MeCN in 5.1 min, flow rate: 0.6 mL/min, t_R= 3.80 min, m/z = 862.4 ([M+H]$^+$). HRMS (ESI+) m/z calculated for $C_{38}H_{55}N_9O_{10}S_2$ [M-H]$^-$ 860.3435; found: 860.3463.

Ac*-V-[CSALC]$_{cyclic}$-W-*NH_2 (3d); (4*R*,7*S*,10*S*,13*S*,16*R*)-16-((*S*)-2-acetamido-3-methylbutanamido)-*N*-((*S*)-1-amino-3-(1*H*-indol-3-yl)-1-oxopropan-2-yl)-13-(hydroxymethyl)-7-isobutyl-10-methyl-6,9,12,15-tetraoxo-1,2-dithia-5,8,11,14-tetraazacycloheptadecane-4-carboxamide. The linear precursor peptide was synthesized on a Fmoc Trp(Boc) TentaGel® S RAM resin (load: 0.2 mmol/g) in a 0.2 mmol scale manually using Fmoc strategy with two coupling cycles and two deprotection cycles as described in the "general Fmoc-SPPS procedure" followed by the "general acetylation procedure". The cleavage of the peptide from the resin was done following the "general cleavage procedure" protocol (in this case TFA/TIS/H_2O/Anisole 95:2:2:1 as cleavage cocktail), where 7 mL cleavage cocktail was used with an incubation of 4.5 h on a shaker. After lyophilization, 45 mg of crude uncyclized product was achieved, which was used for the "general cyclisation procedure" for disulfide cyclisation of the cysteines for 10 days. For purification, the "general preparative HPLC procedure" has been performed. A yield of 0.8 mg cyclized peptide (Purity: ≥ 71 %; 0.98 µmol, 0.49 % according to initial load of the resin) was obtained. LC-MS: Column: Phenomenex Luna C18(2), gradient: MeCN/H_2O + 0.05 % HCOOH, 5 % MeCN increase to 50 % MeCN in 5.1 min, flow rate: 0.6 mL/min, t_R= 3.82 min, m/z = 820.3 ([M+H]$^+$). HRMS (ESI+) m/z calculated for $C_{36}H_{53}N_9O_9S_2$ [M-H]$^-$ 818.3329; found: 818.3356.

4

***Ac*-V-[CSEAC]**$_{cyclic}$**-W-*NH$_2$** (3e); 3-((4*R*,7*S*,10*S*,13*S*,16*R*)-16-((*S*)-2-acetamido-3-methylbutanamido)-4-(((*S*)-1-amino-3-(1*H*-indol-3-yl)-1-oxopropan-2-yl)carbamoyl)-13-(hydroxymethyl)-7-methyl-6,9,12,15-tetraoxo-1,2-dithia-5,8,11,14-tetraazacycloheptadecan-10-yl)propanoic acid.** The linear precursor peptide was synthesized on a Fmoc Trp(Boc) TentaGel® S RAM resin (load: 0.2 mmol/g) in a 0.1 mmol scale manually using Fmoc strategy with two coupling cycles and two deprotection cycles as described in the "general Fmoc-SPPS procedure" followed by the "general acetylation procedure". The cleavage of the peptide from the resin was done following the "general cleavage procedure" protocol, where 5 mL cleavage cocktail was used with an incubation of 2 h on a shaker. After lyophilization, 44 mg of crude uncyclized product was achieved, which was used for the "general cyclisation procedure" for disulfide cyclisation of the cysteines for 3 days. For purification, the "general preparative HPLC procedure" has been performed. A yield of 1.21 mg pure (≥ 98 %) cyclized peptide (1.45 µmol, 1.45 % according to initial load of the resin) was obtained. LC-MS: Column: Phenomenex Luna C18(2), gradient: MeCN/H$_2$O + 0.05 % HCOOH, 5 % MeCN increase to 50 % MeCN in 5.1 min, flow rate: 0.6 mL/min, t_R= 3.10 min, m/z = 834.3 ([M-H]$^-$). HRMS (ESI+) m/z calculated for C$_{35}$H$_{49}$N$_9$O$_{11}$S$_2$ [M-H]$^-$ 834.2915; found: 834.2944.

***Ac*-V-[CSELC]**$_{cyclic}$**-A-*NH$_2$** (3f); 3-((4*R*,7*S*,10*S*,13*S*,16*R*)-16-((*S*)-2-acetamido-3-methylbutanamido)-4-(((*S*)-1-amino-1-oxopropan-2-yl)carbamoyl)-13-(hydroxymethyl)-7-isobutyl-6,9,12,15-tetraoxo-1,2-dithia-5,8,11,14-tetraazacycloheptadecan-10-yl)propanoic acid.** The linear precursor peptide was synthesized on a Fmoc Ala TentaGel® S RAM resin (load: 0.2 mmol/g) in a 0.1 mmol scale manually using Fmoc strategy with two coupling cycles and two deprotection cycles as described in the "general Fmoc-SPPS procedure" followed by the "general acetylation procedure". The cleavage of the peptide from the resin was done following the "general cleavage procedure" protocol, where 5 mL cleavage cocktail was used with an incubation of 2.5 h on a shaker. After lyophilization, 63 mg of crude uncyclized product was achieved, which was used for the "general cyclisation procedure" for disulfide cyclisation of the cysteines for 3 days. For purification, the "general preparative HPLC procedure" has been performed. A yield of 6 mg pure (≥ 98 %) cyclized peptide (7.86 µmol, 7.86 % according to initial load of the resin) was obtained. LC-MS: Column: Phenomenex Luna C18(2), gradient: MeCN/H$_2$O + 0.05 % HCOOH, 5 % MeCN increase to 50 % MeCN in 5.1 min, flow rate: 0.6 mL/min, t_R= 2.99 min, m/z = 761.3 ([M-H]$^-$). HRMS (ESI+) m/z calculated for C$_{30}$H$_{50}$N$_8$O$_{11}$S$_2$ [M-H]$^-$ 761.2985; found: 761.2985.

***Ac*- − -[CSELC]**$_{cyclic}$**-W-*NH$_2$** (4a); ((4*R*,7*S*,10*S*,13*S*,16*R*)-16-acetamido-10-(2-carboxyethyl)-13-(hydroxymethyl)-7-isobutyl-6,9,12,15-tetraoxo-1,2-dithia-5,8,11,14-tetraazacycloheptadecane-4-carbonyl)-L-tryptophan.** The linear precursor peptide was synthesized on a Fmoc Trp(Boc) TentaGel® S RAM resin (load: 0.2 mmol/g) in a 0.1 mmol scale manually using Fmoc strategy with two coupling cycles and two deprotection cycles as described in the "general Fmoc-SPPS procedure" followed by the "general acetylation procedure". The cleavage of the peptide from the resin was done following the "general cleavage procedure" protocol, where 5 mL cleavage cocktail was used with an incubation of 2.75 h on a shaker. The solution was suspended in 1.5 mL DCM. At -20 °C TFA (300 µL, 40 eq.) was added and incubated overnight. After lyophilization, 24 mg of crude uncyclized product was achieved, which was used for the "general cyclisation procedure" for disulfide cyclisation of the cysteines for 2 days. For purification, the "general preparative HPLC procedure" has been performed. A yield of 0.77 mg pure (≥ 98 %) cyclized peptide (0.99 µmol, 0.99 % according to initial load of the resin) was obtained. LC-MS: Column: Phenomenex Luna C18(2), gradient: MeCN/H$_2$O + 0.05 % HCOOH, 5 % MeCN increase to 50 % MeCN in 5.1 min, flow rate: 0.6 mL/min, t_R= 3.39 min, m/z = 777.3 ([M-H]$^-$). HRMS (ESI+) m/z calculated for C$_{33}$H$_{46}$N$_8$O$_{10}$S$_2$ [M-H]$^-$ 877.2700; found: 877.2720.

***Ac*-V-[CSELC]**$_{cyclic}$ **-*NH$_2$** (4b); (4*R*,7*S*,10*S*,13*S*,16*R*)-16-((*S*)-2-acetamido-3-methylbutanamido)-10-(2-carboxyethyl)-13-(hydroxymethyl)-7-isobutyl-6,9,12,15-tetraoxo-1,2-dithia-5,8,11,14-tetraazacycloheptadecane-4-carboxylic acid.** The linear precursor peptide was synthesized on a Fmoc Cys(Trt) TentaGel® S RAM resin (load: 0.2 mmol/g) in a 0.1 mmol scale manually using Fmoc strategy with two coupling cycles and two deprotection cycles as described in the "general Fmoc-SPPS procedure" followed by the "general acetylation procedure". The cleavage of the peptide from the resin was done following the "general cleavage procedure" protocol, where 7 mL cleavage cocktail was used with an incubation of 2.5 h on a shaker. After lyophilization, 55.9 mg of crude uncyclized product was achieved, which was used for the "general cyclisation procedure" for disulfide cyclisation of the cysteines for 3 days. For purification, the "general preparative HPLC procedure" has been performed. A yield of 0.27 mg pure (≥ 89 %) cyclized peptide (0.39 µmol, 0.39 % according to initial load of the resin) was obtained. LC-MS: Column: Phenomenex Luna C18(2), gradient: MeCN/H$_2$O + 0.05 % HCOOH, 5 % MeCN increase to 50 % MeCN in 5.1 min, flow rate: 0.6 mL/min, t_R= 2.99 min, m/z = 690.3 ([M-H]$^-$). HRMS (ESI+) m/z calculated for C$_{27}$H$_{45}$N$_7$O$_{10}$S$_2$ [M-H]$^-$ 690.2591; found: 690.2608.

General Fmoc-SPPS procedure for triazole-bridged peptides 5a, 5b, 5c. The amino acids were coupled *via* two coupling cycles. For proteinogenic amino acids a solution of the Fmoc protected amino acid (4.0 eq), HBTU (3.9 eq) and DIPEA (8.0 eq) was used. For non-proteinogenic amino acids a solution of the amino acid (3.5 eq), HATU (3.9 eq) and DIPEA (8.0 eq) was used. The resin was shaken 1 h at room temperature before every coupling step was followed by washing steps with DMF (5 x 6 mL) and DCM (5 x 6 mL). Fmoc deprotection was achieved by a reaction with 20 % piperidine in DMF for 5 min at room temperature, followed by a second deprotection step under same conditions. A solution of DMF/DIPEA/Ac$_2$O (5:3:2) was given on the resin and shook 30 min at room temperature for the Acetylation of the peptide. The cleavage of the peptide from the resin and removal of the side chain protecting groups was done with a cleavage cocktail of TFA/H$_2$O/anisole/TES (47:1:1:1). The mixture was shaken at room temperature for 3 h. After precipitation, the obtained solid was washed with MTBE (4 x 2 mL) and dried by freeze-drying.

Copper-catalyzed azide-alkyne cycloaddition (CuAAC). The linear peptide (1.0 eq) was dissolved in argon-flushed H$_2$O (1 mL/mg). CuSO$_4$·5H$_2$O (2.0 eq), Na-ascorbate (4.0 eq) and DIPEA (8.0 eq) were added and the reaction mixture stirred under argon at room temperature overnight. The solvent was removed by freeze-drying and the macrocyclic peptide was purified by preparative HPLC.

5

Ac*-V-[Pra-SEL-Aza]**cyclic**-W-*NH₂ (1,4-triazole) (5a); 3-((3*S*,6*S*,9*S*,12*S*,15*S*,*Z*)-15-((*S*)-2-acetamido-3-methylbutanamido)-3-(((*S*)-1-amino-3-(1*H*-indol-3-yl)-1-oxopropan-2-yl)carbamoyl)-12-(hydroxymethyl)-6-isobutyl-5,8,11,14-tetraoxo-11*H*-4,7,10,13-tetraaza-1(1,4)-triazolacyclohexadecaphane-9-yl)propanoic acid.** The linear precursor peptide *Ac*-V-Pra-SEL-Aza-W-*NH₂* was synthesized manually on a Fmoc Trp(Boc) Tenta Gel S RAM resin (0.20 mmol/g) at 0.10 mmol scale. According to the general Fmoc-SPPS procedure for triazole-bridged peptides, the following amino acids and building blocks were used: Fmoc-L-Aza-OH (0.35 mmol, 3.5 eq), Fmoc-Leu-OH (0.4 mmol, 4.0 eq), Fmoc-Glu(O*t*Bu)-OH (0.4 mmol, 4.0 eq), Fmoc-Ser(*t*Bu)-OH (0.4 mmol, 4.0 eq), Fmoc-Pra-OH (0.35 mmol, 3.5 eq) and Fmoc-Val-OH (0.4 mmol, 4.0 eq). Fmoc-L-Aza-OH and Fmoc-Pra-OH were used together with HATU (0.39 mmol, 3.9 eq) and DIPEA (0.8 mmol, 8.0 eq), while all other amino acids were used together with HBTU (0.39 mmol, 3.9 eq) and DIPEA (0.8 mmol, 8.0 eq). The product was received as a white solid (32.6 mg, 37.0 µmol, 37 % according to the initial load of the resin). According to "Copper-catalyzed azide-alkyne cycloaddition" the macrocyclic peptide was prepared by a reaction of 27.0 mg (30.6 µmol, 1.0 eq) linear precursor peptide, 15.4 mg CuSO₄·5H₂O (61.2 µmol, 2.0 eq), 24.4 mg Na-ascorbate (122 µmol, 4.0 eq) and 41.5 µL DIPEA (245 µmol, 8.0 eq). The solvent was removed by freeze-drying. The macrocyclic peptide was purified by preparative HPLC (H₂O:ACN 9:1→1:1) and was received as a pure (≥ 98 %) white solid (5.10 mg, 5.79 µmol, 19 %). The characterization was done by LC-MS, IR, ¹H-NMR, ¹³C-NMR, 2 D NMR and HRMS (m/z). LC-MS: Column: Phenomenex Luna C18(2), gradient: MeCN/H₂O + 0.05 % HCOOH, 5 % MeCN increase to 50 % MeCN in 5.1 min, flow rate: 0.6 mL/min, t$_R$= 3.49 min, m/z = 879.5 ([M-H]⁻). HRMS (ESI+) m/z calculated for C₄₀H₅₆N₁₂O₁₁ [M+H]⁺ 881.4270; found: 881.4236.
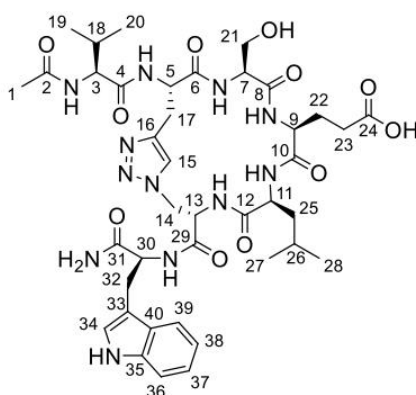
¹H-NMR (500 MHz, CD₃OD, δ in ppm): 10.0 (bs, 1 H, Indole NH), 8.46-8.55 (m, 1 H, NH), 8.27-8.35 (bs, 1 H, NH), 8.08 (d, ³$J_{NH,5}$ = 8.24 Hz, 1 H, NH), 7.76 – 7.81 (m, 1 H, NH), 7.65 (s, 1 H, H15), 7.60 (d, ³$J_{39,38}$ = 7.78 Hz 1 H, H39), 7.54 (d, ³$J_{NH,7}$ = 8.24 Hz, 1 H, NH), 7.34 (d, ³$J_{36,37}$ = 8.09 Hz, 1 H, H36), 7.09 (dd, ³$J_{37,36/38}$ = 7.68 Hz, 1 H, H37), 7.01-7.04 (m, 1 H, H38), 7.01 (s, 1 H, H34), 4.94-5.01 (m, 1 H, H13), 4.73-4.82 (m, 3 H, H5/H14/H30), 4.59 (bs, 1 H, NH), 4.49 (d, ³$J_{14,13}$ = 3.74 Hz, 1 H, H14), 4.30-4.34 (m, 2 H, H7/H9), 4.26-4.30 (m, 1 H, H3), 3.97-4.08 (m, 2 H, H11/H21), 3.76 (dd, ²$J_{21a,21b}$ = 3.74 Hz, ³$J_{21,7}$ = 11.4 Hz, 1 H, H21), 3.60-3.65 (m, 2 H, H23), 3.38 (dd, ²$J_{17a,17b}$ = 3.74 Hz, ³$J_{17,5}$ = 11.4 Hz,, 1 H, H17), 3.19-3.26 (m, 1 H, H17), 3.16 (dd, ³$J_{32a,30}$ = 3.66 Hz, ³$J_{32b,30}$ = 4.88 Hz, 2 H, H32), 2.02-2.19 (m, 2 H, H18/OH), 2.00 (s, 3 H, H1), 1.77-1.94 (m, 2 H, H22), 1.52-1.68 (m, 3 H, H25/H26), 0.97 (dd, ³$J_{19/20,18}$ = 6.48 Hz, 6 H, H19/H20), 0.92 (d, ³$J_{27,26}$ = 6.10 Hz 3 H, H27), 0.87 (d, ³$J_{28,26}$ = 6.10 Hz, 3 H, H28).

¹³C-NMR (126 MHz, CD₃OD, δ in ppm): 176.9 (C24), 175.3 (C31), 174.6 (C29), 173.8 (C2), 173.7 (C4), 173.1 (C12), 172.7 (C10), 171.1 (C6/C8), 138.2 (C16/C35), 128.9 (C40), 126.2 (C15), 125.0 (C34), 122.7 (C37), 120.1 (C39), 119.6 (C38), 112.4 (C36), 111.3 (C33), 63.5 (C21), 60.4 (C3), 57.0 (C7/C9), 55.9 (C30), 55.3 (C5), 54.3 (C11), 53.7 (C13), 51.3 (C14), 39.9 (C25), 32.1 (C18), 29.2 (C23), 28.6 (C32), 26.7 (C17/C22), 26.2 (C26), 23.5 (C27), 22.6 (C1), 21.6 (C28), 19.9 (C20), 18.8 (C19).

Ac*-V-[Aza-SEL-Pra]**cyclic**-W-*NH₂ (1,4-triazole) (5b); 3-((3*S*,6*S*,9*S*,12*S*,15*S*,*Z*)-15-((*S*)-2-acetamido-3-methylbutanamido)-3-(((*S*)-1-amino-3-(1*H*-indol-3-yl)-1-oxopropan-2-yl)carbamoyl)-12-(hydroxymethyl)-6-isobutyl-5,8,11,14-tetraoxo-11*H*-4,7,10,13-tetraaza-1(4,1)-triazolacyclohexadecaphane-9-yl)propanoic acid.** The linear precursor peptide *Ac*-V-Aza-SEL-Pra-W-*NH₂* was synthesized on a microwave-assisted peptide synthesizer (Liberty Lite) using Fmoc Trp(Boc) Tenta Gel S RAM resin (0.20 mmol/g) at 0.05 mmol scale. The engaged amino acids were used in concentrations of 0.2 M in DMF. The used coupling reagents were Oxyma (0.5 M) and DIC (0.25 M) in DMF. The product was received as a white solid (32.0 mg, 36.3 µmol, 73 % according to the initial load of the resin). According to "Copper-catalyzed azide-alkyne cycloaddition" the macrocyclic peptide was prepared by a reaction of 32.0 mg (36.3 µmol, 1.0 eq) linear precursor peptide, 18.6 mg CuSO₄·5H₂O (74.5 µmol, 2.05 eq), 28.2 mg Na-ascorbate (142 µmol, 3.92 eq) and 55 µL DIPEA (317 µmol, 8.7 eq). The solvent was removed by freeze-drying. The macrocyclic peptide was purified by preparative HPLC (H₂O:ACN 9:1→1:1) and was received as a white solid (6.42 mg, 7.29 µmol, 15 %). The characterization was done by LC-MS, ¹H-NMR, ¹³C-NMR and HRMS (m/z). LC-MS: Column: Phenomenex Luna C18(2), gradient: MeCN/H₂O + 0.05 % HCOOH, 5 % MeCN increase to 50 % MeCN in 5.1 min, flow rate: 0.6 mL/min, t$_R$= 3.52 min, m/z = 879.4 ([M-H]⁻). HRMS (ESI+) m/z calculated for C₄₀H₅₆N₁₂O₁₁ [M+H]⁺ 881.4270; found: 881.4253.
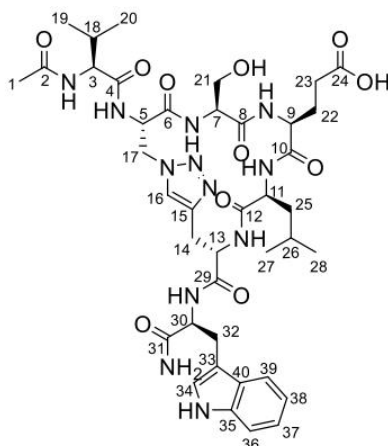
6

¹H-NMR (500 MHz, CD₃OD, δ in ppm): 7.64 (d, $^3J_{39,38}$ = 7.9 Hz, 1 H, H39), 7.58 (s, 1 H, H16), 7.34 (d, $^3J_{36,37}$ = 8.1 Hz, 1 H, H36), 7.15 (s, 1 H, H34), 7.09 (t, $^3J_{37,36/38}$ = 7.3 Hz, 1 H, H37), 7.02 (t, $^3J_{38,37/39}$ = 7.4 Hz, 1 H, H38), 5.13-5.09 (m, 1 H, H5), 4.75-4.66 (m, 3 H, H13/H17ₐ/H30), 4.62-4.55 (m, 2 H, H9, H11), 4.36 (t, $^3J_{17,5}$ = 3.5 Hz, 1 H, H17ᵦ), 4.21 (d, ^3J = 7.3 Hz, 1 H, H3), 4.11 (dd, $^2J_{21a,21b}$ = 11.5 Hz, $^3J_{21a,7}$ = 4.2 Hz, 1H, H21ₐ), 4.03 (dd, $^2J_{32a,32b}$ = 9.5 Hz, $^3J_{32a,30}$ = 5.6 Hz, 1H, H32ₐ), 3.81-3.77 (m, 1H, H32ᵦ), 3.74 (dd, $^2J_{21a,21b}$ = 11.5, $^3J_{21b,7}$ = 3.0, 1 H, H21ᵦ), 3.43 (dd, $^2J_{14a,14b}$ = 15.2 Hz, $^3J_{14a,13}$ = 3.9 Hz, 1 H, H14ₐ), 3.25-3.18 (m, 1 H, H14ᵦ), 2.97-2.90 (m, 1H, H23ₐ), 2.85-2.78 (m, 1H, H23ᵦ), 2.31 (br.s, 2 H, H18/OH), 2.12-2.03 (m, 2H, H22), 2.01 (s, 3 H, H1), 1.69-1.61 (m, 1H, H26), 1.58-1.54 (m, 2 H, 25H), 0.97 (dd, $^3J_{19/20,18}$ = 6.8 Hz, 6 H, H19/H20), 0.92 (d, $^3J_{27,26}$ = 6.5 Hz. 3 H, H27), 0.87 (d, $^3J_{28,26}$ = 6.4 Hz, 3 H, H28).

¹³C-NMR (126 MHz, CD₃OD, δ in ppm): 177.1 (C24), 175.6 (C31), 174.6 (C29), 174.0 (C2), 173.9 (C4), 173.9 (C12), 173.6 (C10), 172.3 (C8), 171.1 (C6), 138.1 (C35), 138.1 (C15), 129.0 (C40), 125.0 (C34), 124.8 (C16), 122.6 (C37), 120.1 (C39), 119.7 (C38), 112.4 (C36), 111.7 (C33), 63.5 (C21), 60.7 (C3), 57.1 (C7/C9), 56.1 (C30), 55.9 (C13), 55.4 (C5), 54.8 (C11), 51.5 (C17), 41.1 (C25), 31.9 (C18), 28.8 (C23), 28.6 (C32), 28.1 (C14/C22), 26.2 (C26), 23.5 (C27), 22.5 (C1), 21.5 (C28), 19.8 (C20), 18.9 (C19).

Ac*-V-[Pra-SEL-Aha]cyclic-W-*NH₂* (1,4-triazole) (5c); 3-((3*S*,6*S*,9*S*,12*S*,15*S*,*Z*)-3-((*S*)-2-acetamido-3-methylbutanamido)-15-(((*S*)-1-amino-3-(1*H*-indol-3-yl)-1-oxopropan-2-yl)carbamoyl)-6-(hydroxymethyl)-12-isobutyl-4,7,10,13-tetraoxo-11*H*-5,8,11,14-tetraaza-1(1,4)-triazolacycloheptadecaphane-9-yl)propanoic acid.** The linear precursor peptide *Ac*-**V-Pra-SEL-Aha-W-*NH₂ was synthesized on a microwave-assisted peptide synthesizer (Liberty Lite) using Fmoc Trp(Boc) Tenta Gel S RAM resin (0.20 mmol/g) at 0.05 mmol scale. The engaged amino acids were used in concentrations of 0.2 M in DMF. The used coupling reagents were Oxyma (0.5 M) and DIC (0.25 M) in DMF. The product was received as a white solid (35.5 mg, 39.7 μmol, 79 % according to the initial load of the resin). According to "Copper-catalyzed azide-alkyne cycloaddition" the macrocyclic peptide was prepared by a reaction of 35.5 mg (39.7 μmol, 1.0 eq) linear precursor peptide, 20.4 mg CuSO₄·5H₂O (81.7 μmol, 2.06 eq), 33.2 mg Na-ascorbate (168 μmol, 4.2 eq) and 55 μL DIPEA (317 μmol, 8.0 eq). The solvent was removed by freeze-drying. The macrocyclic peptide was purified by preparative HPLC (H₂O:ACN 9:1→1:1) and was received as a pure (≥ 98 %) white solid (8.45 mg, 9.4 μmol, 19 %). The characterization was done by LC-MS, ¹H-NMR, ¹³C-NMR and HRMS (m/z). LC-MS: Column: Phenomenex Luna C18(2), gradient: MeCN/H₂O + 0.05 % HCOOH, 5 % MeCN increase to 50 % MeCN in 5.1 min, flow rate: 0.6 mL/min, t$_R$= 3.48 min, m/z = 895.4 ([M+H]⁺). HRMS (ESI+) m/z calculated for C₄₀H₅₆N₁₂O₁₁ [M+H]⁺ 895.4426; found: 895.4406.

7
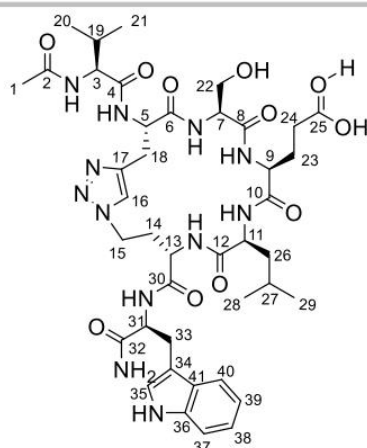
¹H-NMR (500 MHz, CD₃OD, δ in ppm): 7.66 (d, $^3J_{40,39}$ = 7.8 Hz, 1 H, H40), 7.39 (s, 1 H, H16), 7.32 (s, 1 H, H35), 7.05 (d, $^3J_{37,38}$ = 8.0 Hz, 1 H, H37), 7.00 (t, $^3J_{39,38/40}$ = 7.3 Hz, 1 H, H39), 6.94 (t, $^3J_{38,37/39}$ = 7.4 Hz, 1 H, H38), 4.85-4.81 (m, 1H, H5), 4.44-4.37 (m, 2H, H9, H31), 4.25-4.18 (m, 2H, H9, H31), 4.13 (d, $^3J_{13,14}$ = 6.1 Hz, 1H, H13), 4.08 (dd, $^3J_{11,26}$ = 4.1 Hz, 1H, H11), 3.89 (dd, $^2J_{22a,22b}$ = 10.3 Hz, $^3J_{22a,7}$ = 5.2 Hz, 1H, H22a), 3.54-3.33 (m, 3H, H22b, H15), 3.19-3.06 (m, 4H, H18, H33), 2.58 (br.s, 1H, H24), 2.49-2.40 (m, 1H, H14a), 2.23 (br.s, 1H, H23), 2.06-2.00 (m, 1H, H19), 1.99 (s, 3H, H1), 1.89-1.81 (m, 1H, H14b), 1.66-1.52 (m, 3H, H26, H27), 0.94 (dd, $^3J_{20,19 or 21,19}$ = 6.6 Hz, 4.6 Hz, 6H, H20, H21), 0.91 (d, $^3J_{29,27}$ = 5.8 Hz, 3H, H29), 0.87 (d, $^3J_{28,27}$ = 6.0 Hz, 3H, H28).

¹³C-NMR (126 MHz, CD₃OD, δ in ppm): 176.9 (C25), 176.0 (C32), 174.8 (C4 or C8 or C30), 174.7 (C4 or C8 or C30), 173.5 (C2 or C6 or C10), 173.4 (C2 or C6 or C10), 173.3 (C2 or C6 or C10), 172.2 (C12, C4 or C8 or C30), 138.0 (C36), 128.5 (C17, C41), 125.3 (C16, C35), 122.5 (C39), 120.0 (C38), 119.6 (C40), 112.3 (C37), 111.4 (C34), 62.4 (C22), 60.0 (C3), 57.2 (C7), 56.6 (C31), 54.9 (C5 or C9 or C11), 54.8 (C5 or C9 or C11), 53.6 (C13), 45.7 (C15), 39.7 (C26), 32.0 (C19), 31.2 (C14), 28.9 (C23, C24), 28.4 (C33), 26.1 (C27), 23.4 (C29), 22.4 (C1), 19.7 (C20 or C21), 18.7 (C20 or C21).

Ac-V-[Pra-SEL-Aza]*cyclic*-W-*NH₂* (6a); (4*S*,7*S*,10*S*,13*S*)-13-(((S)-1-(((S)-1-(((S)-1-amino-3-(1*H*-indol-3-yl)-1-oxopropan-2-yl)amino)-3-(5-methyl-1*H*-1,2,3-triazol-1-yl)-1-oxopropan-2-yl)amino)-4-methyl-1-oxopentan-2-yl)carbamoyl)-10-(hydroxymethyl)-4-isopropyl-7-methyl-2,5,8,11-tetraoxo-3,6,9,12-tetraazahexadecan-16-oic acid, *Ac*-V-[Pra-SEL-Aha]*cyclic*-W-*NH₂* (6b); (4*S*,7*S*,10*S*,13*S*)-13-(((S)-1-(((S)-1-(((S)-1-amino-3-(1*H*-indol-3-yl)-1-oxopropan-2-yl)amino)-4-(5-methyl-1*H*-1,2,3-triazol-1-yl)-1-oxobutan-2-yl)amino)-4-methyl-1-oxopentan-2-yl)carbamoyl)-10-(hydroxymethyl)-4-isopropyl-7-methyl-2,5,8,11-tetraoxo-3,6,9,12-tetraazahexadecan-16-oic acid and *Ac*-V[Aha-SEL-Pra]*cyclic* W-*NH₂* (6c); 3-((3*S*,6*S*,9*S*,12*S*,15*S*,*Z*)-15-((*S*)-2-acetamido-3-methylbutanamido)-3-(((S)-1-amino-3-(1*H*-indol-3-yl)-1-oxopropan-2-yl)carbamoyl)-12-(hydroxymethyl)-6-isobutyl-5,8,11,14-tetraoxo-11H-4,7,10,13-tetraaza-1(4,1)-triazolacycloheptadecaphane-9-yl)propanoic acid.

Each peptide was manually synthesized by Fmoc-SPPS using Rink amide-MBHA resin (0.8 mmol loading, 100-200 mesh, Chempep Inc). Fmoc-protected amino acids (Fmoc-Trp(Boc)-OH, Fmoc-Leu-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Ser(tBu)-OH, Fmoc-Val-OH) were purchased from either Mimotopes or CSBio, and the azide and alkyne precursors (Fmoc-Aza-OH, Fmoc-Aha-OH and Fmoc-Pra-OH) were purchased from Chem-Impex International Inc. The peptides were synthesized on a 0.25 mmol scale and the resins were first swelled in DMF for 30 min prior to Fmoc deprotection (standard condition used throughout assembly: 20 % piperidine, 5 mL, 15 min). Each amino acid was coupled using 4.0 eq. of amino acid, 4.0 eq. of benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) and 8.0 eq. of *N*,*N*-diisopropylethylamine (DIPEA) in 3 mL of DMF, shaken for 45 min. The resin was washed with DMF (3 x 5 mL) and DCM (3 x 5 mL) between each deprotection and coupling step. Following assembly of the sequences, the *N*-terminus was acetylated using acetic anhydride/DMF (1:4, 6 mL) and 2 eq. of DIPEA, shaken for 45 min. The resin was then washed thoroughly in DMF (5 x 5 mL) and DCM (5 x 5 mL), dried under a stream of N₂ and stored in a desiccator overnight.

The crude peptide-bound resin was next subject to ruthenium-catalyzed azide-alkyne cycloaddition (RuAAC) to install the 1,5-disubstituted 1,2,3-triazole bridge. For each analogue, 75 µmol of resin-bound peptide (6a: 160 mg; 6b: 170 mg; and 6c 170 mg; based on initial resin loading) was loaded into a glass vessel and suspended in 2 mL of anhydrous DMF. The mixture was sparged with argon for 30 min prior to the addition of 50 mol% of chloro(pentamethylcyclopentadienyl)(cyclooctadiene)ruthenium(II) (14.2 mg, 37.5 µmol). The reaction was heated to 80 ˚C for 18 h under an atmosphere of argon. The resin was thoroughly washed with DMF (3 x 5 mL), 0.5 % sodium diethyldithiocarbamate trihydrate in DMF (w/v, 3 x 5 mL), MeOH (3 x 5 mL), and DCM (3 x 5 mL) and dried under N₂. The peptides were next cleaved from the resin by suspending the resin in 5 mL of TFA/TIS/H₂O (95:2.5:2.5, v/v/v) for 2 h, followed by precipitation with cold Et₂O before being redissolved in 50 % MeCN and lyophilized. The peptide was next purified to >95 % purity by preparative RP-HPLC on a Shimadzu Prominence system with a Phenomenex Gemini C-18 column (5 µm, 250 x 10 mm) using a gradient of 20-50 % Solvent B (Solvent A: H₂O with 0.05 % trifluoroacetic acid; Solvent B: 95 % acetonitrile with 0.05 % trifluoroacetic acid) at 3 mL/min. The lyophilized peptides were obtained as a white solid (6a: purity: ≥ 95 %*, 0.67 mg, 0.76 µmol, 1.0 % overall yield

8

based on the initial resin loading; **6b**: purity: ≥ 98 %, 1.72 mg, 1.92 µmol, 2.6 %; and **6c**: purity: ≥ 98 %, 1.40 mg, 1.56 µmol, 2.1 %) and further characterized by LC-MS, ^{1}H-NMR and HRMS.

6a: HRMS (ESI+) m/z calculated for $C_{40}H_{56}N_{12}O_{11}$ [M+H]$^+$ 881.4270; found: 881.4280.

6b: LC-MS: Column: Phenomenex Luna C18(2), gradient: MeCN/H_2O + 0.05 % HCOOH, 5 % MeCN increase to 50 % MeCN in 5.1 min, flow rate: 0.6 mL/min, t_R= 3.52 min, m/z = 895.4 ([M+H]$^+$). HRMS (ESI+) m/z calculated for $C_{41}H_{58}N_{12}O_{11}$ [M+H]$^+$ 895.4426; found: 895.4428.

6c: LC-MS: Column: Phenomenex Luna C18(2), gradient: MeCN/H_2O + 0.05 % HCOOH, 5 % MeCN increase to 50 % MeCN in 5.1 min, flow rate: 0.6 mL/min, t_R= 3.52 min, m/z = 895.4 ([M+H]$^+$). HRMS (ESI+) m/z calculated for $C_{41}H_{58}N_{12}O_{11}$ [M+H]$^+$ 895.4426; found: 895.4402.

6a
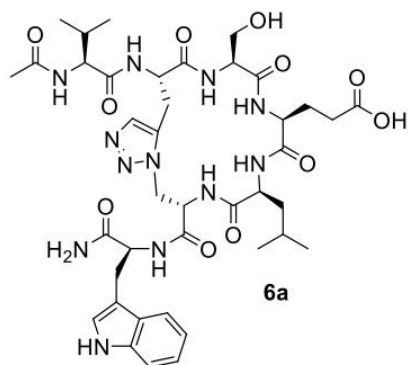
Table S1. NMR chemical shifts of **6a**.

Residue	H	N	Hα	Hβ	Others
1 Val	8.08	114.1	4.04	1.97	0.88 Hγ1; 0.87 Hγ2; 59.8 Cα; 30.0 Cβ; 17.7 Cγ1; 18.3 Cγ2; Ac: 1.97 3H, 21.7 CH₃
2 Pra	8.73	123.5	4.73	3.02, 3.16	7.63 Hδ; 25.5 Cβ; 133.3 Cδ
3 Ser	7.87	118.1	4.20	3.75, 3.85	56.1 Cα; 60.3 Cβ
4 Glu	8.40	120.9	3.86	1.91 (2H)	2.16 Hγ (2H); 56.1 Cα; 26.4 Cβ; 33.3 Cγ
5 Leu	7.73	118.6	4.09	1.25 (2H)	1.39 Hγ; 0.74 Hδ1; 0.77 Hδ2; 52.5 Cα; 39.5 Cβ; 24.3 Cγ; 20.6 Cδ1; 22.2 Cδ2
6 Aza	8.08	125.5	5.04	4.39 (2H)	52.2 Cα; 47.3 Cβ
7 Trp	7.93	123.5	4.74	3.14, 3.32	128.9 Nε1; 7.13 Hδ1; 9.98 Hε1; 7.58 Hε3; 7.41 Hζ2; 7.06 Hζ3; 7.16 Hη2; 27.0 Cβ; 124.6 Cδ1; 118.4 Cε3; 112.0 Cζ2; 119.4 Cζ3; 122.03 Cη2; NH₂: 7.09 H1, 7.43 H2, 108.0 N

9
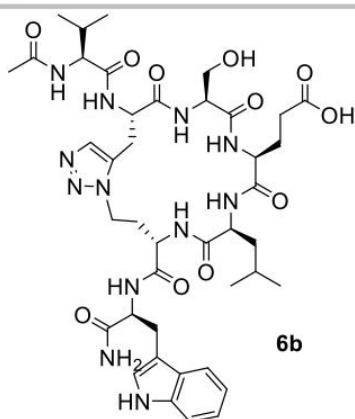
Table S2. NMR chemical shifts of **6b**.

Residue	H	N	Hα	Hβ	Others
1 Val	8.07	125.6	3.94	1.91	0.81 Hγ1; 0.83 Hγ2; 60.0 Cα; 29.8 Cβ; 18.3 Cγ1; 17.9 Cγ2; Ac: 1.95 3H, 21.6 CH₃
2 Pra	8.78	123.5	4.77	3.38, 3.11	7.54 Hδ; 25.6 Cβ; 131.7 Cδ
3 Ser	8.59	118.2	4.26	3.89, 3.81	57.1 Cα; 60.7 Cβ
4 Glu	7.87	119.3	4.37	2.07, 1.85	2.29 Hγ (2H); 52.2 Cα; 27.2 Cβ; 30.5 Cγ
5 Leu	7.95	119.8	3.98	1.45, 1.35	1.41 Hγ; 0.73 Hδ1; 0.79 Hδ2; 53.9 Cα; 39.2 Cβ; 24.3 Cγ; 21.4 Cδ1; 21.7 Cδ2
6 Aha	8.33	121.8	4.40	2.24, 1.82	4.16 Hγ1; 4.25 Hγ2; 50.7 Cα; 31.7 Cβ; 44.1 Cγ
7 Trp	7.99	121.4	4.62	3.27, 3.13	129.1 Nε1; 7.14 Hδ1; 10.01 Hε1; 7.55 Hε3; 7.34 Hζ2; 7.02 Hζ3; 7.09 Hη2; 27.2 Cβ; 124.6 Cδ1; 118.3 Cε3; 112.0 Cζ2; 119.3 Cζ3; 122.0 Cη2; NH₂: 7.02 H1, 7.46 H2, 108.2 N
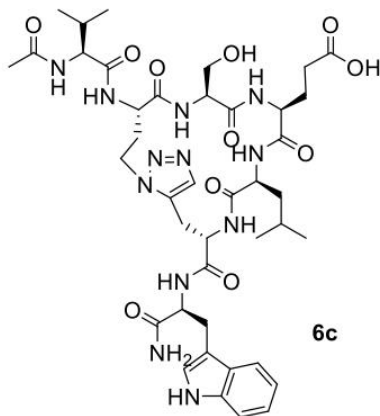
Table S3. NMR chemical shifts of **6c**.

Residue	H	N	Hα	Hβ	Others
1 Val	8.02	124.5	4.07	2.02	0.87 Hγ1; 0.86 Hγ2; 62.2 Cα; 32.6 Cβ; 21.2 Cγ1; 20.1 Cγ2; Ac 1.96 3H, 24.4 CH₃
2 Aha	8.70	124.4	4.08	2.27, 2.61	3.71 Hγ (2H); 46.1 Cα; 32.3 Cβ; 54.8 Cγ
3 Ser	8.18	117.7	4.23	3.96, 3.79	59.5 Cα; 63.2 Cβ
4 Glu	7.88	118.5	4.15	2.00 (2H)	2.36 Hγ2 (2H); 57.6 Cα; 28.7 Cβ; 33.3 Cγ
5 Leu	7.58	118.1	4.11	1.02, 1.24	1.43 Hγ; 0.69 Hδ1; 0.72 Hδ2; 56.2 Cα; 42.9 Cβ; 26.8 Cγ; 24.8 Cδ1; 23.0 Cδ2
6 Pra	7.98	117.0	4.62	2.89 (2H)	7.43 Hδ; 26.84 Cβ; 134.7 Cδ
7 Trp	7.68	121.1	4.62	3.16, 3.31	128.8 Nε1; 7.14 Hδ1; 10.0 Hε1; 7.5 Hε3; 7.41 Hζ2; 7.06 Hζ3; 7.16 Hη; 29.8 Cβ; 127.3 Cδ1; 121.1 Cε3; 114.6 Cζ2; 122.1 Cζ3; 124.6 Cη2; NH₂: 7.03 H1, 7.33 H2, 107.7 N

10

Table S4. Key facts about the synthesis of the individual peptides. Starting from the used resin over cleavage cocktail and cyclization and yield.

Peptide	Resin	Approach	Manually/ synthesizer	Cleavage cocktail	Crude linear peptide yield	Cyclisation	Yield after prep HPLC	Purity[a]
1	Fmoc Trp(Boc) TentaGel® S RAM	0.2 mmol	Manually	TFA/TIS/H$_2$O (37:1:1) + DTT, 7 mL, 3 h	45.7 mg	ACN/H$_2$O (1:1) 1 mg/mL, 2 % DMSO, 2 days	2.68 mg	≥98%
2a	Fmoc Trp(Boc) TentaGel® S RAM	0.1 mmol	Manually	TFA/TIS/H$_2$O (37:1:1) + DTT, 7 mL, 2.5 h	87.4 mg	ACN/H$_2$O (1:1) 1 mg/mL, 2 % DMSO, 2 days	1.03 mg	≥98%
2b	Fmoc Trp(Boc) TentaGel® S AC	0.05 mmol	Synthesizer	TFA/TIS/H$_2$O (37:1:1) + DTT, 7 mL, 2.5 h	14.4 mg	ACN/H$_2$O (1:1) 1 mg/mL, 2 % DMSO, 3 days	1.09 mg	≥98%
3a	Fmoc Trp(Boc) TentaGel® S RAM	0.1 mmol	Manually	TFA/TIS/H$_2$O (37:1:1) + DTT, 7 mL, 2.5 h	n.d.	–	3.24 mg	≥98%
3b	Fmoc Trp(Boc) TentaGel® S RAM	0.2 mmol	Manually	TFA/TIS/H$_2$O/Anisole (95:2:2:1), 3 mL, 3 h	n.d.	ACN/H$_2$O (1:1) 1 mg/mL, 2 % DMSO, 3 days	1.07 mg	≥95%
3c	Fmoc Trp(Boc) TentaGel® S RAM	0.05 mmol	Synthesizer	TFA/TIS/H$_2$O (37:1:1) + DTT, 7 mL, 2.5 h	n.d.	ACN/H$_2$O (1:1) 1 mg/mL, 3 % DMSO, 7 days	3.44 mg	≥98%
3d	Fmoc Trp(Boc) TentaGel® S RAM	0.2 mmol	Manually	TFA/TIS/H$_2$O/Anisole (95:2:2:1) + DTT, 7 mL, 4.5 h	45 mg	ACN/H$_2$O (2:1) 0.17 mg/mL, 1 % DMSO, 10 days	0.8 mg	≥91%[b]
3e	Fmoc Trp(Boc) TentaGel® S RAM	0.1 mmol	Manually	TFA/TIS/H$_2$O (37:1:1) + DTT, 5 mL, 2 h	44 mg	ACN/H$_2$O (1:1) 1 mg/mL, 2 % DMSO, 3 days	1.21 mg	≥98%
3f	Fmoc Ala TentaGel® S RAM resin	0.1 mmol	Manually	TFA/TIS/H$_2$O (37:1:1) + DTT, 5 mL, 2.5 h	63 mg	ACN/H$_2$O (1:1) 1 mg/mL, 2 % DMSO, 3 days	6 mg	≥98%
4a	Fmoc Trp(Boc) TentaGel® S RAM	0.1 mmol	Manually	TFA/TIS/H$_2$O (37:1:1) + DTT, 5 mL, 2.75 h	24 mg	ACN/H$_2$O (1:1) 1 mg/mL, 2 % DMSO, 2 days	0.77 mg	≥98%
4b	Fmoc Cys(Trt) TentaGel® S RAM resin	0.1 mmol	Manually	TFA/TIS/H$_2$O (37:1:1) + DTT, 7 mL, 2.5 h	55.9 mg	ACN/H$_2$O (1:1) 1 mg/mL, 2 % DMSO, 3 days	0.27 mg	≥92%
5a	Fmoc Trp(Boc) TentaGel® S RAM	0.1 mmol	Manually	TFA/H$_2$O/anisole/TES (47:1:1:1), 3 h	32.6 mg	CuAAC	5.1 mg	≥98%
5b	Fmoc Trp(Boc) TentaGel® S RAM	0.05 mmol	Synthesizer	TFA/H$_2$O/anisole/TES (47:1:1:1), 3 h	32 mg	CuAAC	6.42 mg	≥98%
5c	Fmoc Trp(Boc) TentaGel® S RAM	0.05 mmol	Synthesizer	TFA/H$_2$O/anisole/TES (47:1:1:1), 3 h	35.5 mg	CuAAC	8.45 mg	≥98%
6a/6b/6c	Rink amide-MBHA resin	0.25 mmol	Manually	TFA/TIS/H$_2$O (95:2.5:2.5, v/v/v), 2 h	160 mg / 170 mg / 170 mg	RuAAC	0.67 mg / 1.72 mg / 1.4 mg	≥95% / ≥98% / ≥98%

[a] Determined by LC-MS. [b] This peptide was very poorly soluble, therefore the cyclization reaction had to be diluted and carried out with a higher ACN amount. This also prolonged the cyclization time significantly. In addition, many side products were formed, which could not be separated easily, which explains the lower purity of ≥ 91 %.

11

Expression of *Y. pseudotuberculosis* CsrA-biot-His$_6$

The expression protocol of *Yersinia pseudotuberculosis* CsrA-biot-His$_6$ has already been published by Christine Maurer *et al.*[1] Two plasmids were transformed into electro competent *E. coli* BL21 by performing a double transformation. On the one hand pET28a with pAKH172_biotag insert for overexpression of His- and biotin-tagged CsrA and on the other hand pBirAcm for overexpression of biotin ligase for *in vivo* biotinylation at the lysine residue of the biotintag. pET28a has a kanamycin resistance, pBirAcm a chloramphenicol resistance. The amino acid sequence for the CsrA-biot-His$_6$ construct is MLILTRRVGE TLMIGDEVTV TVLGVKGNQV RIGVNAPKEV SVHREEIYQR IQAEKSQPTT Y<u>LEGLNDIFE AQKIEWHEL</u>E <u>HHHHHH</u>. Biotin tag and His tag are underlined. The molecular weight of the CsrA-biot-His$_6$ monomer is 10.2 kDa.

4 L of LB medium, containing 50 µg/mL kanamycin and 17 µg/mL chloramphenicol, were inoculated with an overnight preculture. This main culture was grown at 37 °C and 180 rpm until an O.D. 600 of 0.6 was reached. Then, 10 mL 5 mM biotin (50 µM end concentration), 3 mL 3 M MgCl$_2$ (10 mM end concentration) and 1.19 mL 0.84 M IPTG (1 mM end concentration) per liter of culture was added. The culture was grown again at 37 °C, 180 rpm overnight. Cells were harvested by centrifugation (4 °C, 6200 rpm, 20 min). The pellets were resuspended in 4.5 mL/g wet cells lysis buffer (50 mM dipotassium hydrogen phosphate trihydrate, 300 mM sodium chloride, 10 mM imidazole, pH 8.0) containing cOmplete™ (EDTA-free protease inhibitor cocktail, Roche). Afterwards the cells were disrupted by one passage through a microfluidizer. After centrifugation of the homogenisate (4 °C, 19000 rpm, 1 h), the supernatant was sterile-filtered through 0.22 µm membrane filter. For purification an ÄKTAxpress™ device with a 1 mL HisTrap™ HP column was used, which was equilibrated with 20 mL lysis buffer (4 mL/min flowrate). The clear lysate was loaded on the column with 1 mL/min. This was followed by two washing steps, first 15 mL of high salt buffer (50 mM dipotassium hydrogen phosphate trihydrate, 1 M sodium chloride, 10 mM imidazole, pH 8.0), second 20 mL of binding buffer (50 mM dipotassium hydrogen phosphate trihydrate, 300 mM sodium chloride, 30 mM imidazole, pH 8.0). Next, a linear gradient from 0 to 70 % elution buffer (25 mM dipotassium hydrogen phosphate trihydrate, 150 mM sodium chloride, 125 mM imidazole, pH 8.0) within 56 min was chosen. For the final elution step, it was switched to 100 % elution buffer and the fractions were collected. The CsrA-containing fractions were concentrated via Vivaspin® 20 spin filters (3 kDa MWCO, Sartorius™), before the buffer was exchanged to storage buffer (50 mM dipotassium hydrogen phosphate trihydrate, 300 mM sodium chloride, ad DEPC-treated water (RNase-free), pH 8.0) with a PD10 desalting column. The concentration was determined by UV spectroscopy with NanoDrop™ (ε_{280} = 8480 M^{-1}cm^{-1}, monomer). If required, the united fractions were concentrated again with Vivaspin® 20 spin filters (3 kDa MWCO, Sartorius™) to adjust a monomer concentration of about 200 µM. Glycerol (10 % end concentration) was added to the protein and divided into aliquots, flash frozen in liquid nitrogen and stored at - 80 °C. About 2 mg protein per liter of main culture were yielded.

12

Expression of *E. coli* CsrA-His$_6$

The amino acid sequence for the *Escherichia coli* CsrA-His$_6$ construct is MLILTRRVGE TLMIGDEVTV TVLGVKGNQV RIGVNAPKEV SVHREEIYQR IQAEKSQQSSY HHHHH. The molecular weight of the CsrA-His$_6$ monomer is 7.68 kDa. The construct is present in pET21a+ with an ampicillin resistance. The expression protocol of *E. coli* CsrA-His$_6$ is based on Dubey *et al*.[2]

TB medium, containing 100 µg/mL ampicillin, were inoculated with an overnight preculture. This main culture was grown at 37 °C and 180 rpm until an O.D. 600 of 0.6 was reached. Then, 1.19 mL 0.84 M IPTG (1 mM end concentration) per liter of culture was added. The culture was grown again at 37 °C, 180 rpm overnight. Cells were harvested by centrifugation (4 °C, 6200 rpm, 20 min). The pellets were resuspended in 4.5 mL/g wet cells lysis buffer (50 mM potassium dihydrogen phosphate, 300 mM sodium chloride, 10 mM imidazole, 10 % glycerol, pH 8.0) containing cOmplete™ (EDTA-free protease inhibitor cocktail, Roche). Afterwards the cells were disrupted by ultra-sonification (current = 50 %, every 30 sec and 5 cycles, break between every 5 cycles). After centrifugation of the homogenisate (4 °C, 19000 rpm, 1 h), the supernatant was sterile-filtered through 0.22 µm membrane filter. For purification an ÄKTAxpress™ device with a 1 mL HisTrap™ HP column was used, which was equilibrated with 20 mL lysis buffer (4 mL/min flowrate). The clear lysate was loaded on the column with 1 mL/min. This was followed by three washing steps, first washing buffer 1 (50 mM potassium dihydrogen phosphate, 300 mM sodium chloride, 20 mM imidazole, 10 % glycerol, pH 8.0), second two times washing buffer 2 (50 mM potassium dihydrogen phosphate, 300 mM sodium chloride, 50 mM imidazole, 10 % glycerol, pH 8.0). The elution was done with elution buffer (50 mM potassium dihydrogen phosphate, 300 mM sodium chloride, 250 mM imidazole, 10 % glycerol, pH 8.0) and the fractions were collected. The CsrA-containing fractions were concentrated via Vivaspin® 20 spin filters (3 kDa MWCO, Sartorius™), before the buffer was exchanged to storage buffer (50 mM dipotassium hydrogen phosphate trihydrate, 300 mM sodium chloride, ad DEPC-treated water (RNase-free), pH 8.0) with a PD10 desalting column. The concentration was determined by UV spectroscopy with NanoDrop™ (ε_{280} = 2980 M^{-1}cm^{-1}, monomer). If required, the united fractions were concentrated again with Vivaspin® 20 spin filters (3 kDa MWCO, Sartorius™) to adjust a monomer concentration of about 200 µM. Glycerol (10 % end concentration) was added to the protein and divided into aliquots, flash frozen in liquid nitrogen and stored at -80 °C.

Expression of *P. Aeruginosa* RsmA-His$_6$

The amino acid sequence for the *Pseudomonas aeruginosa* RsmA-His$_6$ construct is MLILTRRVGE TLMVGDDVTV TVLGVKGNQV RIGVNAPKEV AVHREEIYQRI QKEKDQEPNHKLE HHHHHH. The molecular weight of the RsmA-His$_6$ monomer is 8.1 kDa. The construct is present in pET28a with an kanamycin resistance. The expression protocol of *P. Aeruginosa* RsmA-His$_6$ is based on Jean-Pierre *et al*. but with variations. [3]

TSB medium, containing 50 µg/mL kanamycin, were inoculated with an overnight preculture. This main culture was grown at 37 °C and 180 rpm until an O.D. 600 of 0.7 was reached. Then, 2 mL 0.5 M IPTG (1 mM end concentration) per liter of culture was added. The culture was grown again at 37 °C, 180 rpm for 4 h. Cells were harvested by centrifugation (4 °C, 6200 rpm, 30 min). The pellets were resuspended in 4.5 mL/g wet cells lysis buffer (20 mM potassium dihydrogen phosphate, 500 mM sodium chloride and 20 mM Tris/HCL, pH 7.65). Afterwards the cells were disrupted by ultra-sonification (current = 50 %, every 30 sec and 5 cycles, break between every 5 cycles). After centrifugation of the homogenisate (4 °C, 15 0000 g, 45 min), the supernatant was sterile-filtered through 0.22 µm membrane filter. For purification an ÄKTAxpress™ device with a 1 mL HisTrap™ HP column was used, which was equilibrated with 20 mL lysis buffer (1 mL/min flowrate). The clear lysate was loaded on the column with 1 mL/min. This was followed by three washing steps, first washing buffer 1 (50 mM potassium dihydrogen phosphate, 300 mM sodium chloride, 20 mM imidazole, 10 % glycerol, pH 8.0), second two times washing buffer 2 (50 mM potassium dihydrogen phosphate, 300 mM sodium chloride, 50 mM imidazole, 10 % glycerol, pH 8.0). The elution was done with elution buffer (50 mM potassium dihydrogen phosphate, 300 mM sodium chloride, 250 mM imidazole, 10 % glycerol, pH 8.0) and the fractions were collected. The RsmA-containing fractions were concentrated via Vivaspin® 20 spin filters (5 kDa MWCO, Sartorius™), before the buffer was exchanged to storage buffer (10 mM Tris/HCl, 33% glycerol, pH 7.65) with a PD10 desalting column. The concentration was determined by UV spectroscopy with NanoDrop™ (ε_{280} = 2980 M^{-1}cm^{-1}, monomer). The protein was divided into aliquots, flash frozen in liquid nitrogen and stored at -80 °C.

13

SUPPORTING INFORMATION

Sequence Identities (BLAST).

Sequence identity RsmA(P.aeruginosa) – CsrA(E.coli) = 85%
Method:Compositional matrix adjust.,
Identities:52/61(85%), Positives:58/61(95%), Gaps:0/61(0%)

```
Query  1   MLILTRRVGETLMVGDDVTVTVLGVKGNQVRIGVNAPKEVAVHREEIYQRIQKEKDQEPN  60
           MLILTRRVGETLM+GD+VTVTVLGVKGNQVRIGVNAPKEV+VHREEIYQRIQ EK Q+ +
Sbjct  1   MLILTRRVGETLMIGDEVTVTVLGVKGNQVRIGVNAPKEVSVHREEIYQRIQAEKSQQSS  60

Query  61  H  61
           +
Sbjct  61  Y  61
```

Sequence identity CsrA(Y. pseudotuberculosis) – RsmA(P.aeruginosa) = 86%
Method:Compositional matrix adjust.,
Identities:55/64(86%), Positives:58/64(90%), Gaps:1/64(1%)

```
Query  1   MLILTRRVGETLMIGDEVTVTVLGVKGNQVRIGVNAPKEVSVHREEIYQRIQAEKSQ-PT  59
           MLILTRRVGETLM+GD+VTVTVLGVKGNQVRIGVNAPKEV+VHREEIYQRIQ EK Q P
Sbjct  1   MLILTRRVGETLMVGDDVTVTVLGVKGNQVRIGVNAPKEVAVHREEIYQRIQKEKDQEPN  60

Query  60  TYLE  63
             LE
Sbjct  61  HKLE  64
```

Sequence identity CsrA(Y. pseudotuberculosis) – CsrA(E.coli) = 95%
Method:Compositional matrix adjust.,
Identities:58/61(95%), Positives:60/61(98%), Gaps:0/61(0%)

```
Query  1   MLILTRRVGETLMIGDEVTVTVLGVKGNQVRIGVNAPKEVSVHREEIYQRIQAEKSQPTT  60
           MLILTRRVGETLMIGDEVTVTVLGVKGNQVRIGVNAPKEVSVHREEIYQRIQAEKSQ ++
Sbjct  1   MLILTRRVGETLMIGDEVTVTVLGVKGNQVRIGVNAPKEVSVHREEIYQRIQAEKSQQSS  60

Query  61  Y  61
           Y
Sbjct  61  Y  61
```

14

Phage Display
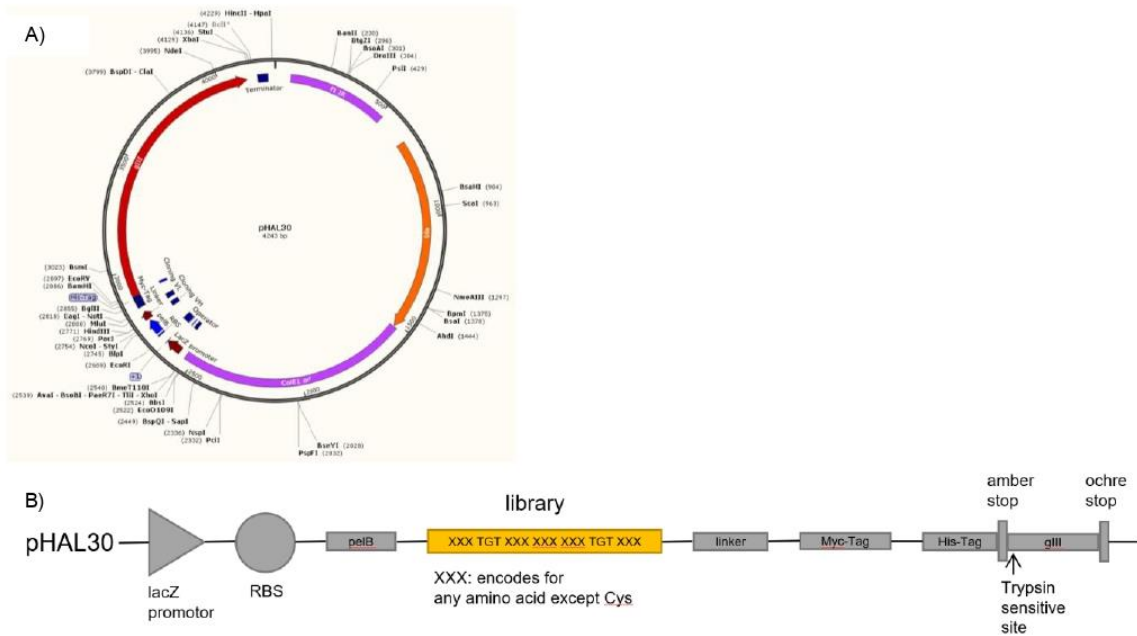
Figure S1. A) Vector map of pHAL30, B) Schematic drawing of pHAL30 with cloned peptide library. lacZ promoter: promoter of the bacterial lac operon; RBS: ribosome binding site; pelB: signal peptide sequence of bacterial pectate lyase *Erwinia caratovora*, mediating secretion into the periplasmic space; gIII: gene encoding for the phage protein III.[4]

Fluorescence Polarization Assay
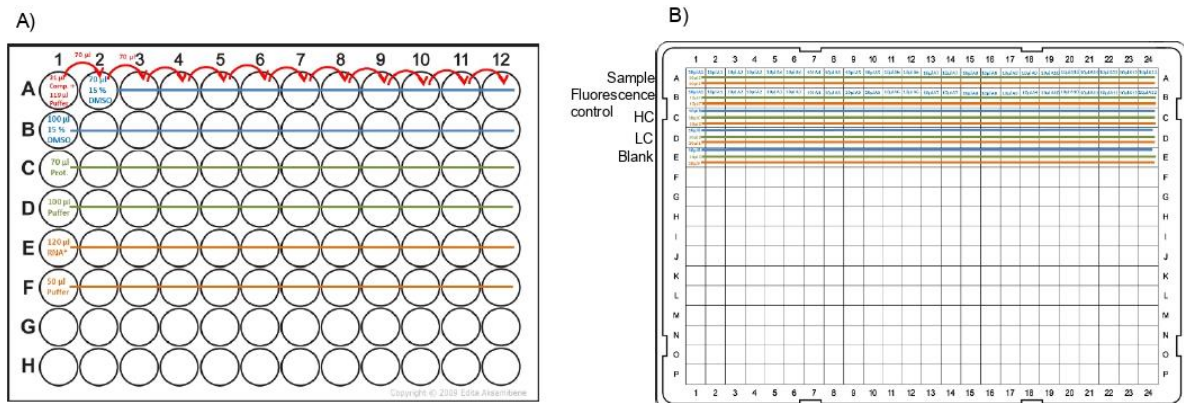
Figure S2. A) Preparation plate (96-well plate, clear, v-bottom, non-binding) scheme and B) measuring plate (Greiner 386 well, black, flat bottom). Transferring in duplicates with 12-channel pipette from preparation plate to measuring plate

The results from the fluorescence polarization assay for the peptides 2a-6c can be found in figure S3 for CsrA from Y. pseudotuberculosis and in figure S4 for CsrA from E. Coli.

47
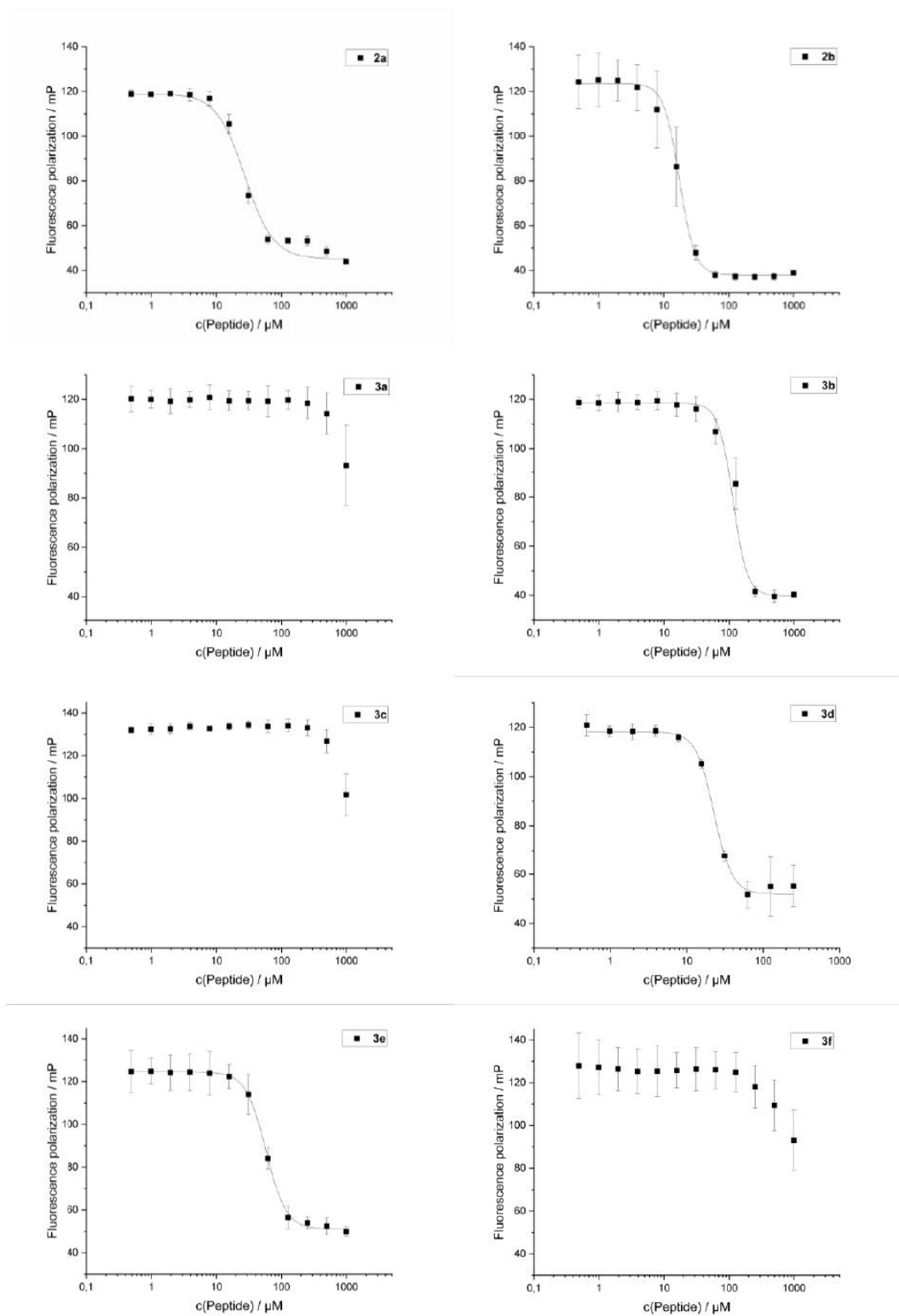
Figure S3 (part 1). Results of the dose-dependent fluorescence polarization assay with competition inhibition curves used to determine the half maximal inhibitory concentration (IC_{50}) for **peptides 2a-3f** with *Yersinia* CsrA_biot_His6. The assay has been performed two times in duplicates

17
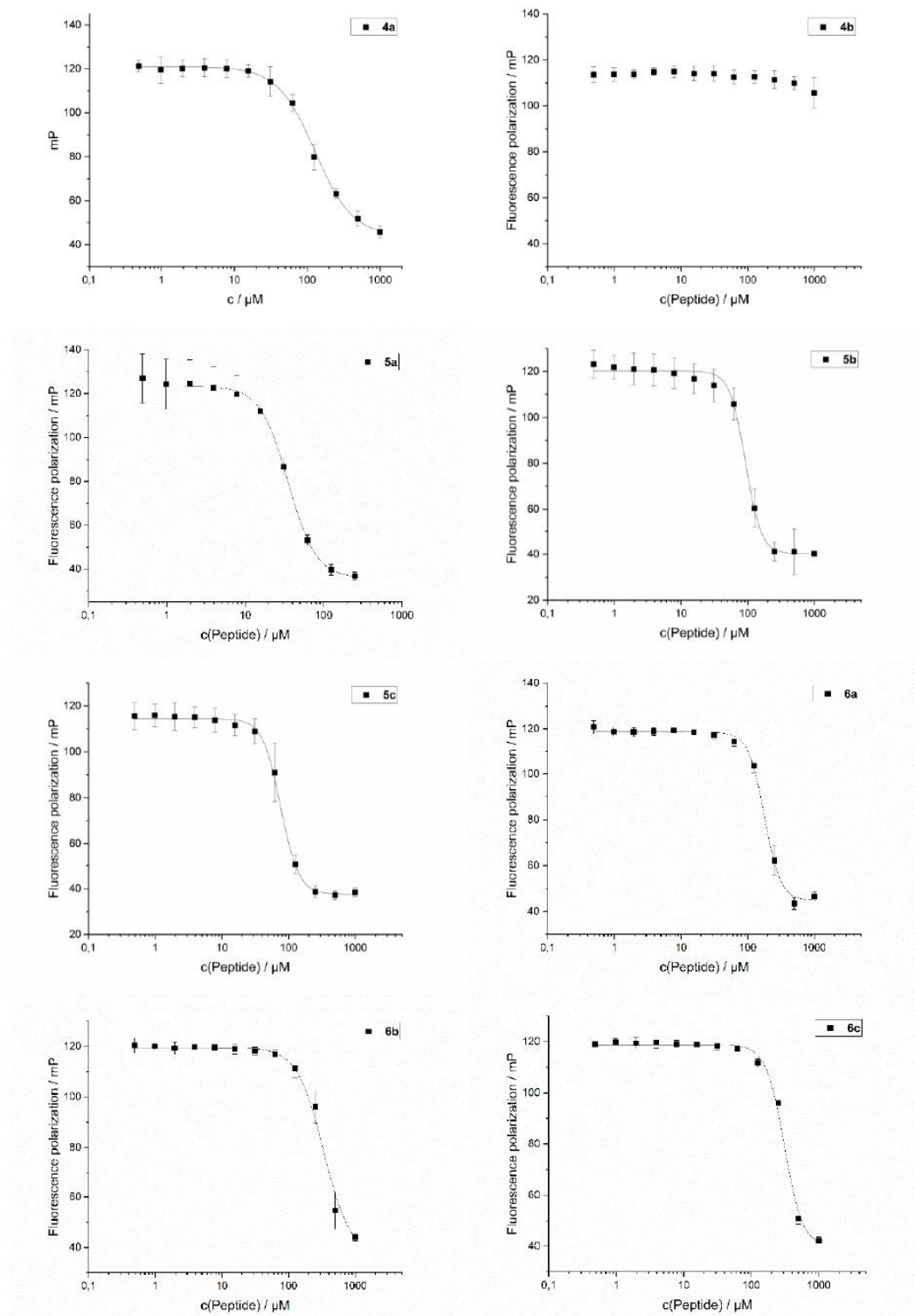
Figure S3 (part 2). Results of the dose-dependent fluorescence polarization assay with competition inhibition curves used to determine the half maximal inhibitory concentration (IC_{50}) for **peptides 4a-6c** with *Yersinia* CsrA_biot_His$_6$. The assay has been performed two times in duplicates.

18
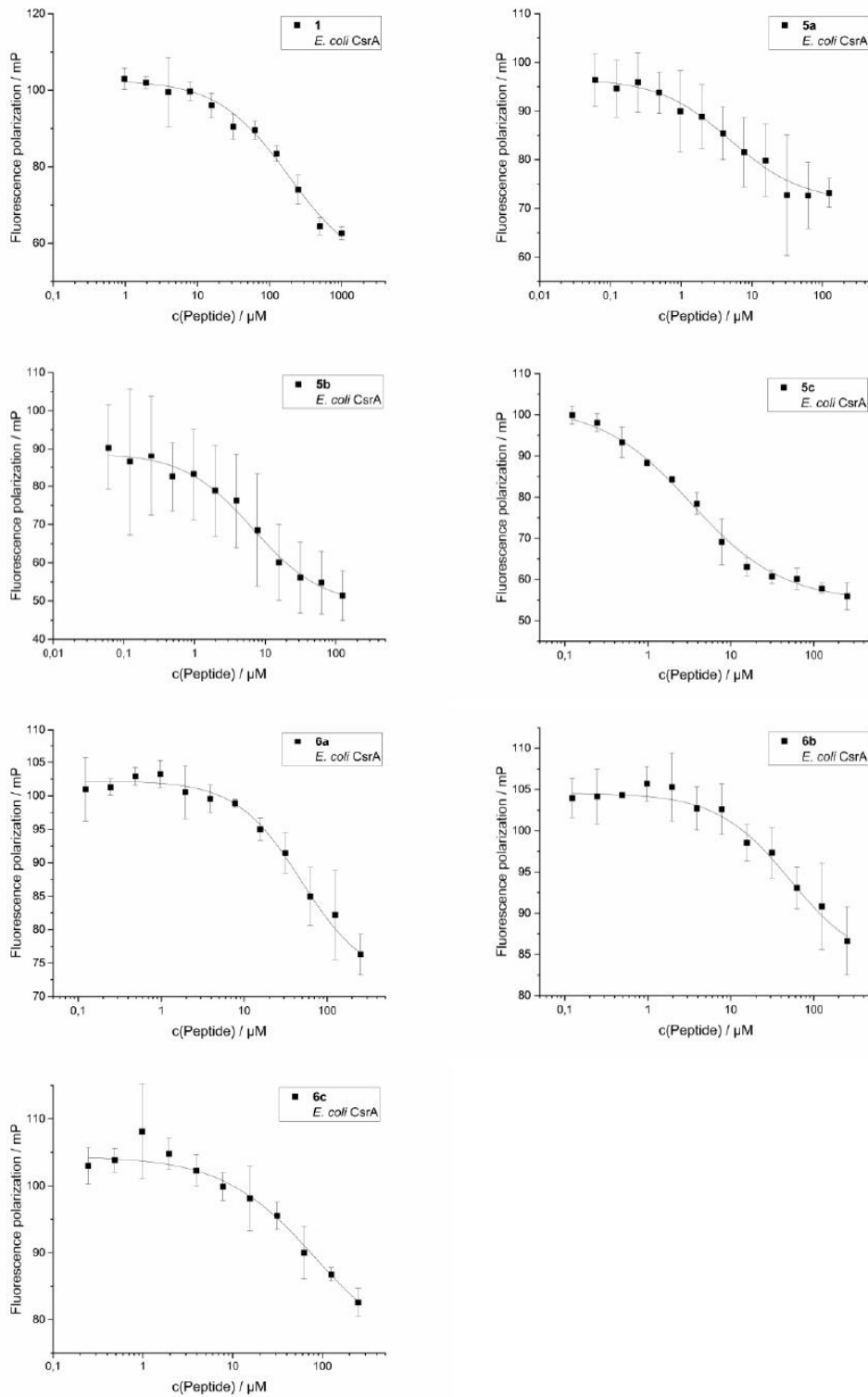
Figure S4. Result of the dose-dependent fluorescence polarization assay with competition inhibition curve used to determine the half maximal inhibitory concentration for peptide **5a** and **5b** with *E. coli* CsrA_His$_6$ as target. The assay has been performed two times in duplicates.

19

Microscale Thermophoresis Assay (MST)

Results from the MST-assay can be found in figure S5
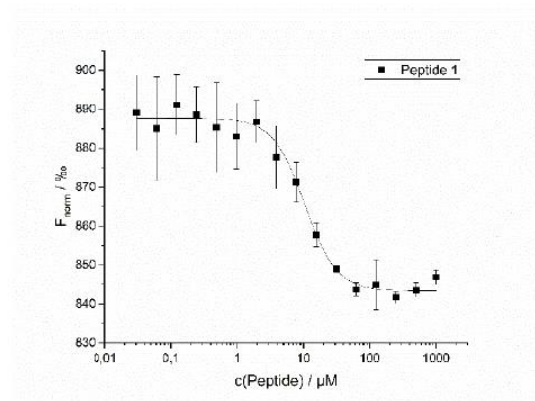
Figure S5. Dose-dependent MST interaction curve of peptide **1** with *Yersinia* CsrA_biot_His$_6$ used to determine the dissociation constant (K$_d$).

Calculation of the Error Bars in Fluorescence Polarization and MST Assay

Error bars are calculated with TINV function of Microsoft Excel, which returns the two-tailed inverse of the Student's t-distribution multiplied with the standard deviation of the mean of the measurements:

TINV(probability, deg_freedom) · (STDEV of the mean)

The argument probability is set to 95 % and the degree of freedom to 4.

20

NMR analysis and structure calculations

NMR analysis of peptide **1** was performed with a 2 mM solution in 50 % H_2O/50 % d3-acetonitrile (298 K, pH 3.4) on a Bruker Avance III 600 MHz spectrometer. 2D Spectra included TOCSY, ROESY, and natural abundance heteronuclear correlation spectroscopy (^{15}N- and ^{13}C-HSQC). TOCSY spectra were also recorded at variable temperatures (283 – 303 K) to determine amide proton temperature coefficients. Spectra were referenced to residual acetonitrile at 1.94 ppm. All spectra were processed using TopSpin v3.6 and assigned using CcpNMR Analysis.

Preliminary structures were calculated in CYANA based upon ROESY-derived distance restraints. Several dihedral angle restraints were also added as predicted by TALOS-N[5] along with a single hydrogen bond pair after consideration of preliminary structures and amide proton temperature coefficients. A final ensemble of 20 structures were generated within CNS[6] using torsion angle dynamics and refinement and energy minimization in explicit water solvent. MolProbity[7] was used to assess stereochemical quality (summarised in Table S5).

Table S5. Statistical analysis of peptide **1** structures[a]

Experimental restraints	
total no. distance restraints	45
intraresidue	22
sequential	23
medium range, i-j<5	2
hydrogen bond restraints	2
dihedral angle restraints	
phi	5
psi	1
Deviations from idealized geometry	
bond lengths (Å)	0.012 ± 0.001
bond angles (deg)	1.161 ± 0.107
impropers (deg)	1.14 ± 0.15
NOE (Å)	0.033 ± 0.005
cDih (deg)	0.032 ± 0.104
Mean energies (kcal/mol)	
overall	-194 ± 13
bonds	3.9 ± 0.4
angles	9.0 ± 1.8
improper	3.3 ± 0.7
van Der Waals	-14.5 ± 2.2
NOE	0.05 ± 0.02
cDih	0.01 ±0.03
electrostatic	-977 ± 33
Violations	
NOE violations exceeding 0.2 Å	0
Dihedral violations exceeding 2.0 Å	0
Rms deviation from mean structure, Å	
backbone atoms	0.79 ± 0.37
all heavy atoms	1.81 ± 0.70
Stereochemical quality[b]	
Residues in most favoured Ramachandran region, %	80.0 ± 11.0
Ramachandran outliers, %	0 ± 0
Unfavourable sidechain rotamers, %	0.0 ± 0.0
Clashscore, all atoms	0.0 ± 0.0
Overall MolProbity score	1.2 ± 0.3

[a] All statistics are given as mean ± SD.
[b] According to MolProbity[7]

In silico Investigations

General. All *in silico* experiments were performed with Molecular Operating Environment (MOE) by Chemical Computing Group (CCG) release 2020.09 employing the Amber10:EHT force field.[8]

Homology Model Building. Homology model of *Yersinia pseudotuberculosis* CsrA in complex with RNA was built using the first entry of NMR-solution structure 2MFH from *Pseudomonas fluorescens* in complex with RNA-oligo ucaggacau.[9] The template structure 2MFH was chosen from the available structures in the PDB based on the following requirements: resolved *C*-terminal residues and complex with short RNA oligo.

The sequence of the template structure (*P. fluorescens*, Sec1) and the target sequence (*Y. pseudotuberculosis*, Sec2) share an identity of 71% and homology of 89% as shown by following blast result:

Score	Expect	Method	Identities	Positives	Gaps
91.7 bits(226)	3e-32	Compositional matrix adjust.	41/58(71%)	52/58(89%)	0/58(0%)

```
Sec2      MLILTRRVGETLMIGDEVTVTVLGVKGNQVRIGVNAPKEVSVHREEIYQRIQAEKSQP   58
          MLILTR+VGE++ IGD++T+T+LGV G QVRIG+NAPK+V+VHREEIYQRIQA  + P
Sec1      MLILTRKVGESINIGDDITITILGVSGQQVRIGINAPKDVAVHREEIYQRIQAGLTAP   58
```

The built-in "homology model" function of MOE was used with standard parameters, while RNA atoms were used as environment to successfully yield a model in complex with RNA.

Docking. Docking was performed using the built-in "Docking" function of MOE. NMR structures of peptide **1** (all 20 entries of PDB ID 7M7X, BMRB ID 30895) were used as "ligand" structures and the above-mentioned homology model as "receptor". The docking site was defined by involving protein residues in 4.5 Å proximity to the RNA atoms as well as the *C*-terminal residues of one of the two identical RNA-interaction sites. The resulting selection was as follows:

```
>CsrAYP_1|Chain A|Translational repressor|Y.pseudotuberculosis HomologyModel
MLILTRR--E-----------------------------------------------
>CsrAYP_1|Chain B|Translational repressor|Y.pseudotuberculosis HomologyModel
-------------------T-L--K--Q-R----APK-VSVHR-EIYQRIQAEKSQPT
```

Placement algorithm was "Triangle Matcher" with "London dG" as Scoring function generating 10 initial poses for every peptide **1** conformer (entry). Refinement method was "Induced Fit" with "GBVI/WSA dG" as Scoring function and 5 keeper poses.

The resulting 100 docking poses (5 poses × 20 entries) were sorted according to the refinement/binding score. The 10 best-scoring poses were sorted according to the "rmsd_refine" parameter indicating binding hypothesis with minimal deviation from the initial (experimental) solution geometry. By this means, we selected the optimal pose scoring in number 8 of 100 regarding refinement/binding score and 3 of 100 regarding the "rmsd_refine" parameter.

Analysis and Visualisation. The pose derived by the docking procedure described above was analysed using the "Ligand Interactions" function of MOE for generating a 2D depiction of the interaction profile (see Figure 3c from the main text). Graphic processing for manuscript figures was done using YASARA structure (YASARA Biosciences GmbH)[10] and POV-Ray 3.7.0.

22

Ac-V**A**SELA**W**-*NH₂* (**3a**)

Ac-**A**-[CSELC]*cyclic*-W-*NH₂* (**3b**)

Ac-V-[CAELC]$_{cyclic}$-W-NH$_2$ (**3c**)

Ac-V-[CSALC]*cyclic*-W-*NH₂* (**3d**)

Ac-V-[CSEAC]*cyclic*-W-*NH₂* (**3e**)

Ac-V-[CSELC]*cyclic*-**A**-*NH₂* (**3f**)

Ac- – – *-[CSELC]*cyclic*-W-NH₂* (**4a**)

32

Ac-V-[CSELC]*cyclic* – -NH₂ (**4b**)

Ac-V-[**Pra**-SEL-**Aza**]*cyclic*-W-NH₂ (1,4-triazole) (**5a**)

34

Ac-V-[**Aza**-SEL-**Pra**]*cyclic*-W-*NH₂* (1,4-triazole) (**5b**)

Ac-V-[**Pra**-SEL-**Aha**]*cyclic*-W-*NH*₂ (1,4-triazole) (**5c**)

Ac-V-[**Pra**-SEL-**Aza**]*cyclic*-W-*NH₂* (1,5-triazole) (**6a**)

Ac-V-[**Pra**-SEL-**Aha**]*cyclic*-W-*NH₂* (1,5-triazole) (**6b**)

37

Ac-V-[**Aha**-SEL-**Pra**]*cyclic*-W-*NH₂* (1,5-triazole) (**6c**)

Purity

No.	Purity / %
1	100
2a	100
2b	100
3a	100
3b	95
3c	100
3d	91
3e	98
3f	100
4a	100
4b	92
5a	100
5b	100
5c	100
6a	100
6b	100
6c	100

38

^{1}H-NMR and ^{13}C-NMR Spectra

Compounds **1**, **5a-5c** were dissolved in 0.5 mL DMSO-d6 and measured with Bruker Fourier spectrometer model Ultrashield Plus 500 (500 MHz for ^{1}H-NMR and 126 MHz for ^{13}C-NMR). Chemical shifts are given in parts per million (ppm) and referenced against the residual proton or carbon resonances of the >99% deuterated solvents as 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, dd = doublet of doublets, dt =doublet of triplets, br = broad and combinations of these) coupling constants, and integration. NMR spectra were evaluated using MestReNova.

The 1,5-triazole analogues **6a**, **6b**, and **6c** were analyzed on a Bruker AVANCE III 600 MHz spectrometer equipped with a cryogenically cooled probe. The samples were prepared in 500 μL of H_2O/D_2O (9:1 v/v, ~2 mM, pH 3.4) and ^{1}H and ^{13}C experiments were acquired at 298 K (referencing to H_2O at 4.70 ppm). The spectra were manually assigned using CCPNMR analysis 2.4.2.
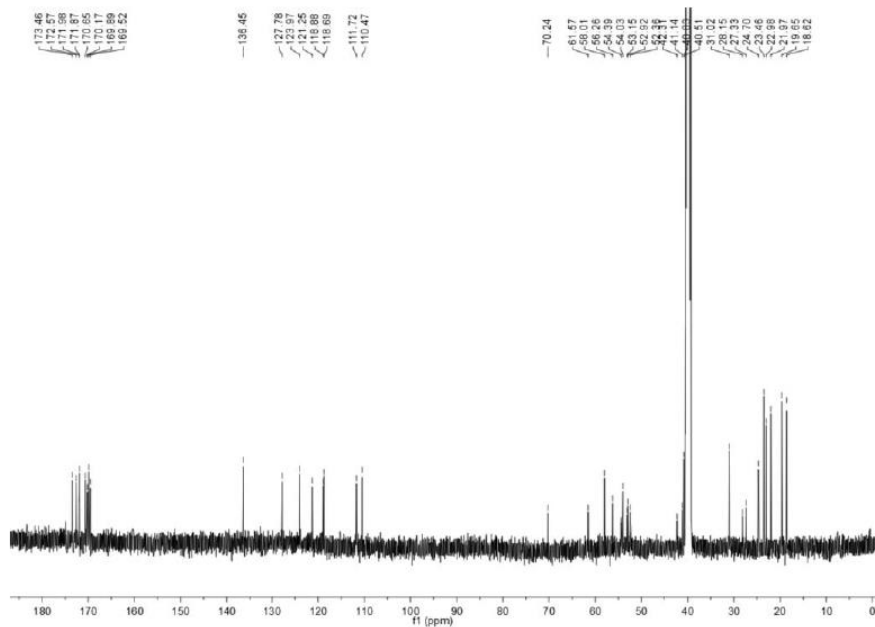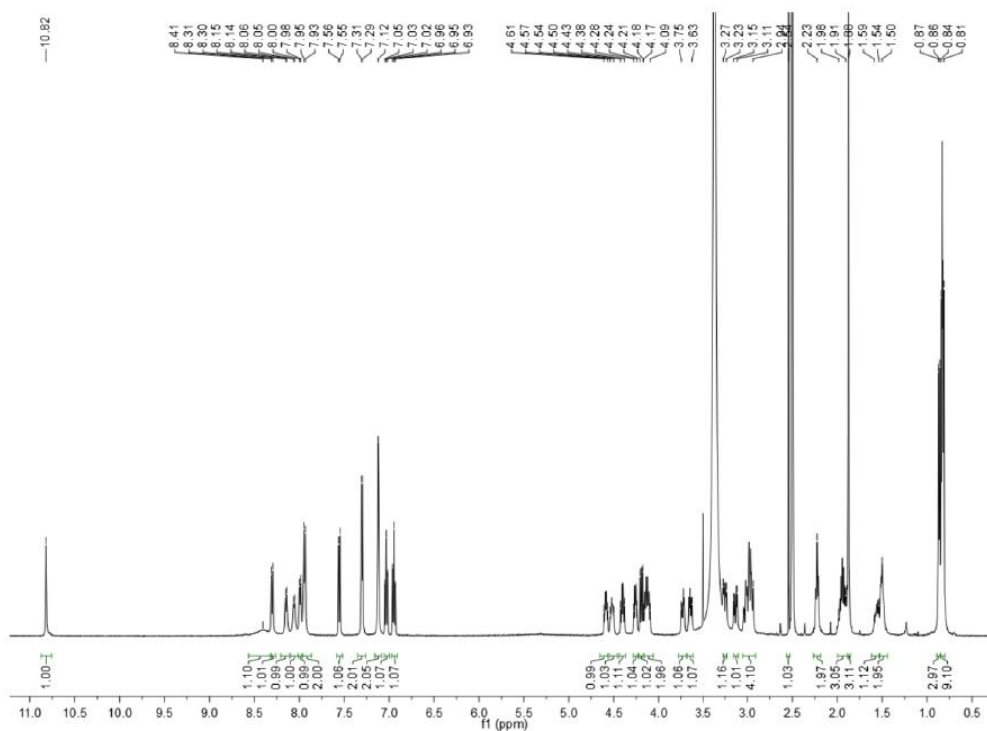
39

Ac-V-[CSELC]*cyclic*-W-*NH₂* (**1**)

40

Ac-V-[**Pra**-SEL-**Aza**]*cyclic*-W-*NH₂* (1,4-triazole) (**5a**)

Ac-V-[**Aza**-SEL-**Pra**]*cyclic*-W-*NH₂* (1,4-triazole) (**5b**)

41

Ac-V-[**Pra**-SEL-**Aha**]*cydic*-W-*NH₂* (1,4-triazole) (**5c**)

43
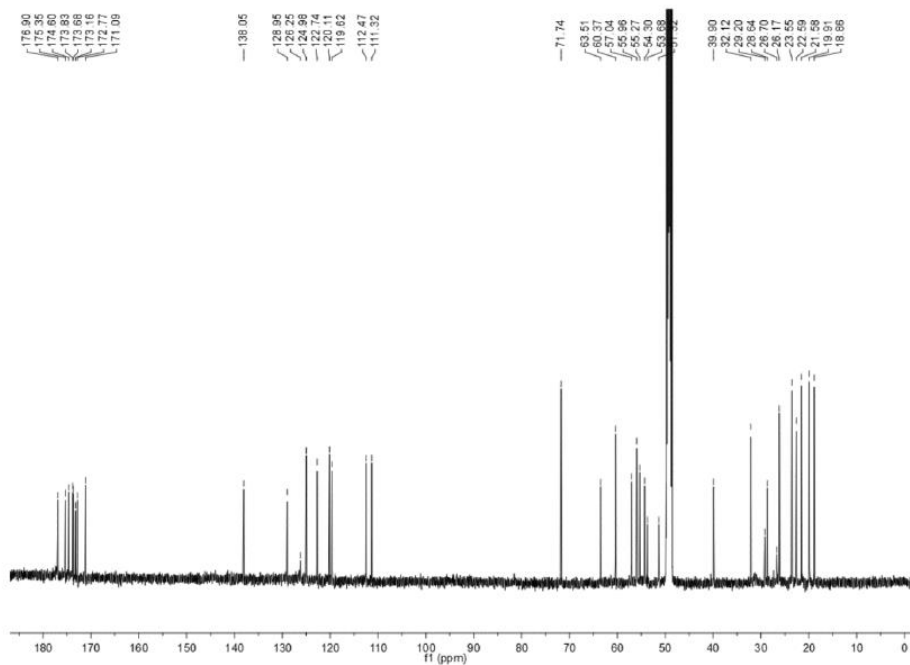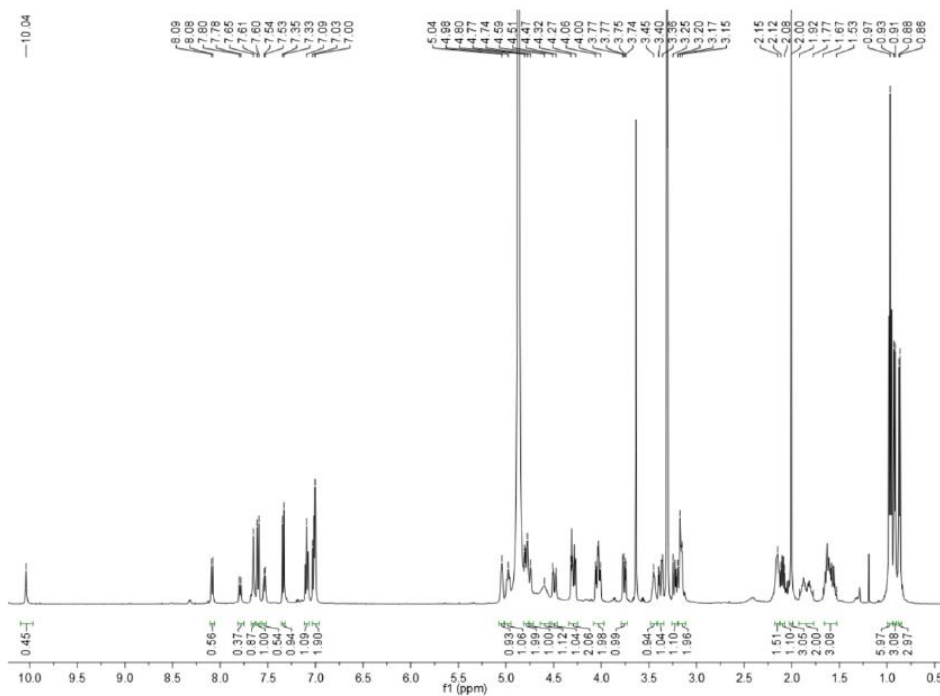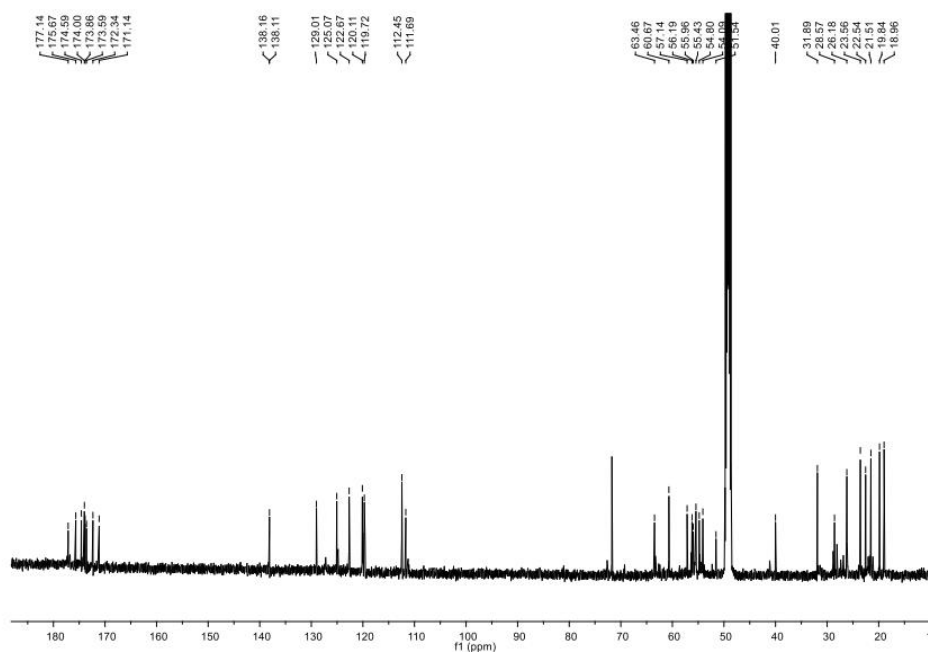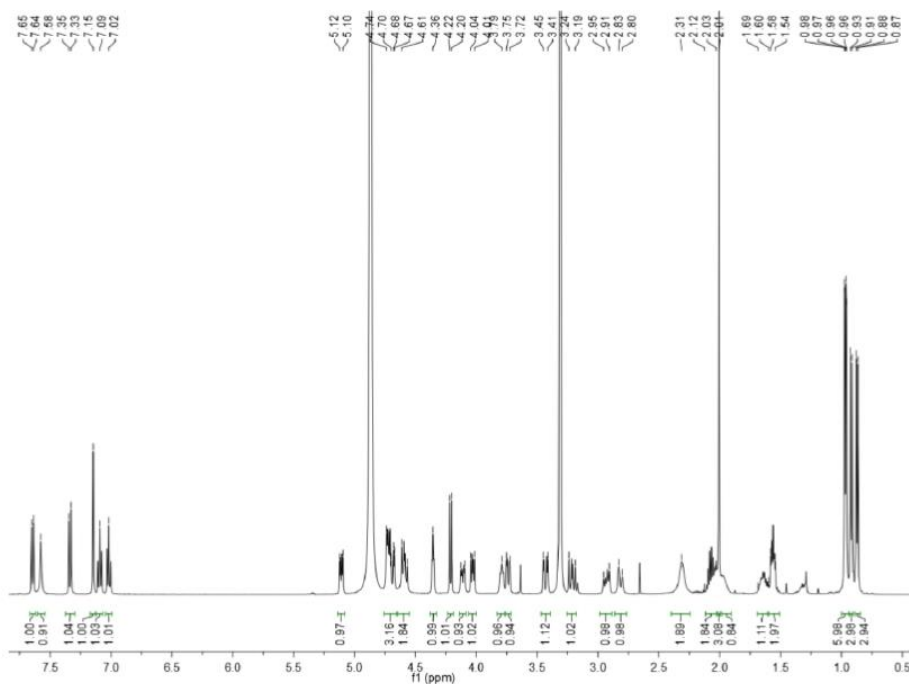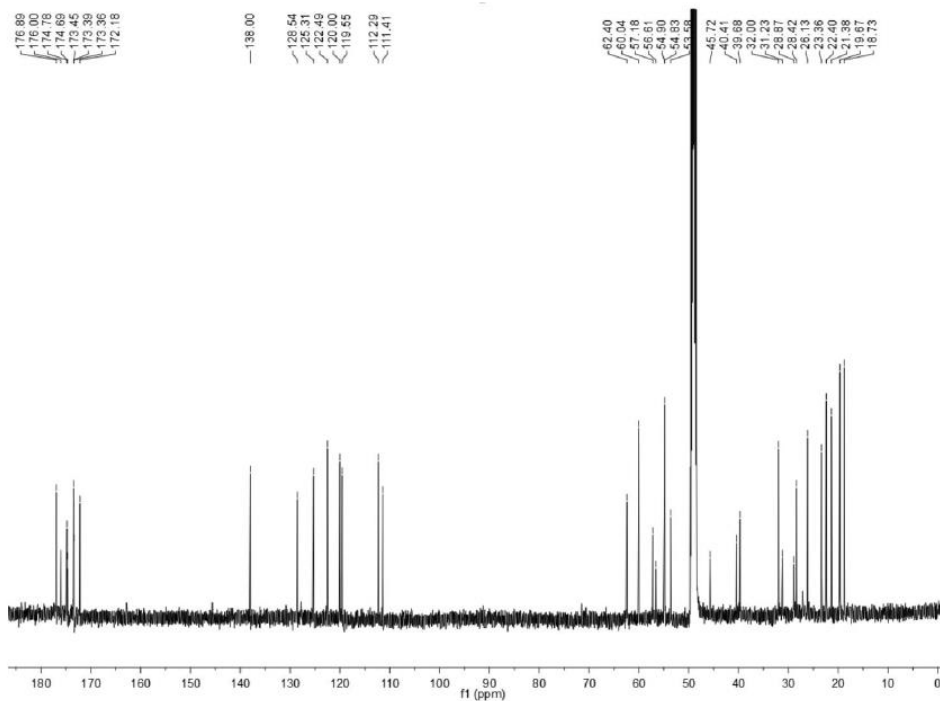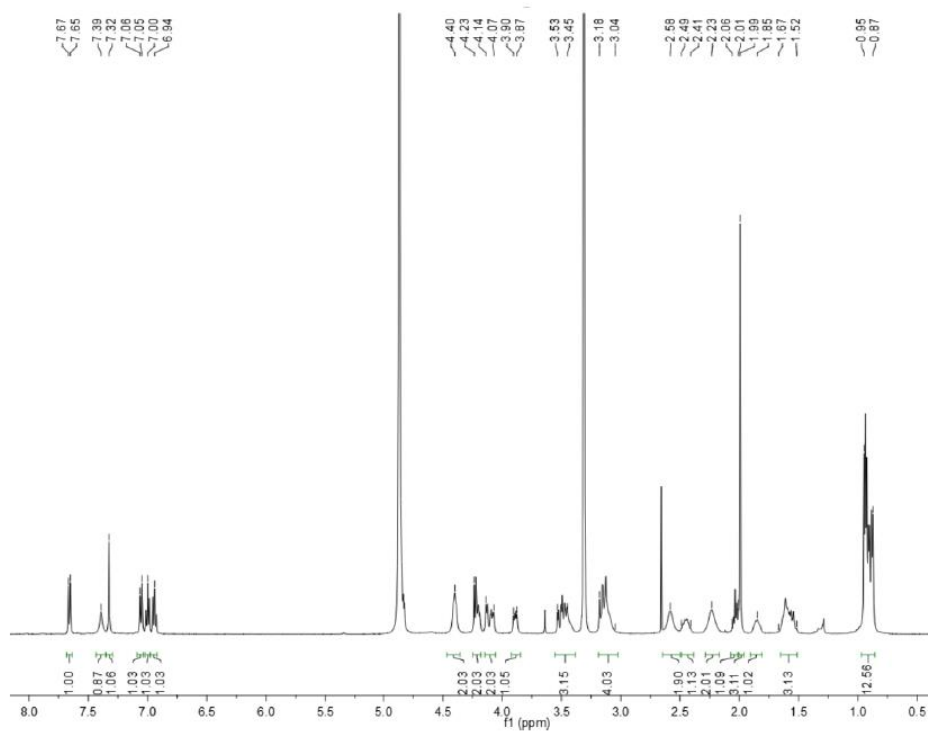
Ac-V-[**Pra**-SEL-**Aza**]*cyclic*-W-*NH₂* (1,5-triazole) (**6a**)

^{1}H NMR spectra of **6a** (600 MHz, H$_2$O/D$_2$O (9:1), pH 3.4)
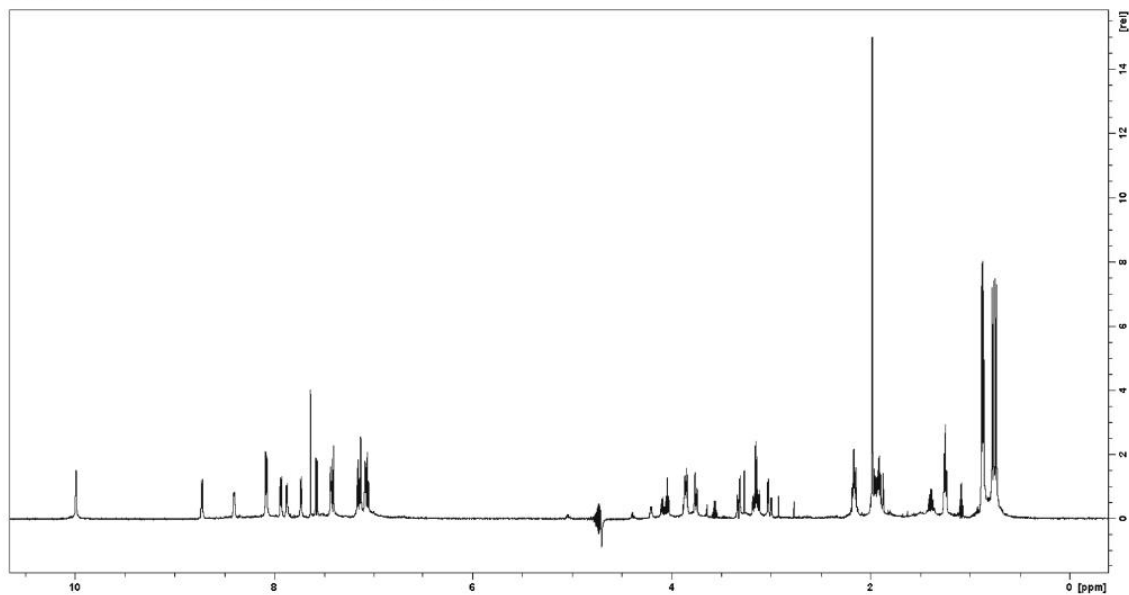
Ac-V-[**Pra**-SEL-]*cyclic*-W-*NH₂* (1,5-triazole) (**6b**)

^{1}H NMR spectra of **6b** (600 MHz, H$_2$O/D$_2$O (9:1), pH 3.4)

44
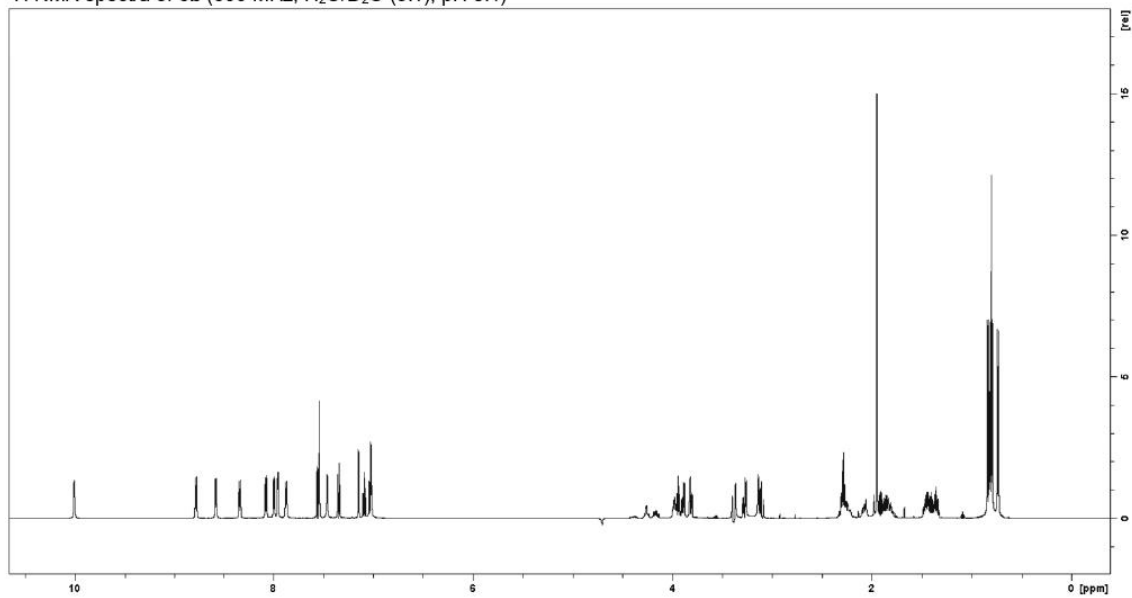
Ac-V-[**Aha**-SEL-**Pra**]*cyclic*-W-*NH₂* (1,5-triazole) (**6c**)

¹H NMR spectra of **6c** (600 MHz, H_2O/D_2O (9:1), pH 3.4)
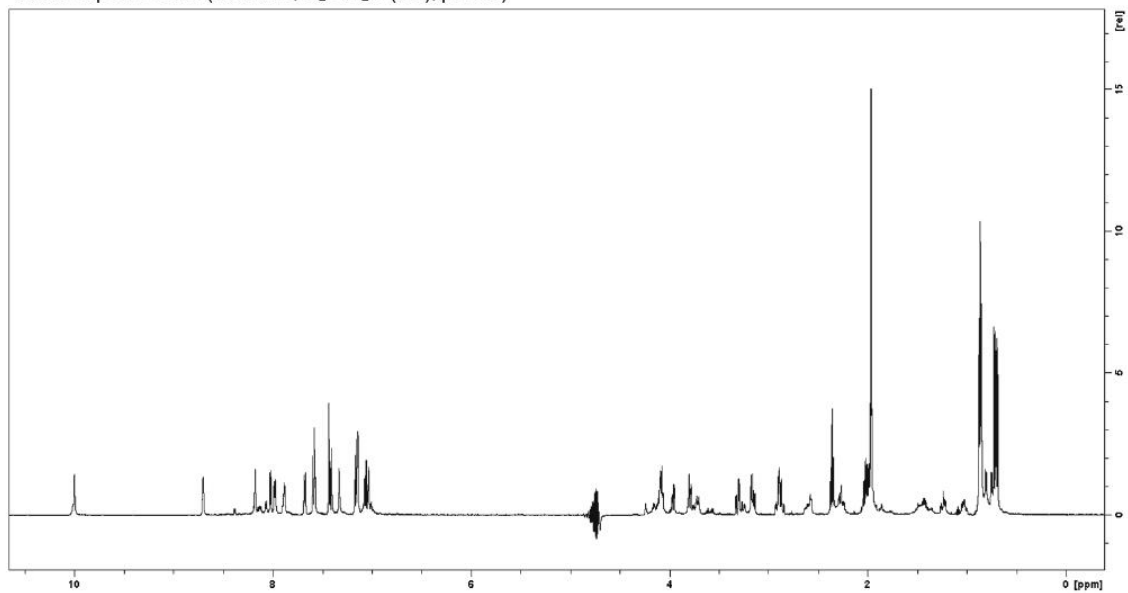
2D NMR Spectra

Ac-V-[**Pra**-SEL-**Aza**]*cyclic*-W-*NH2* (1,4-triazole) (**5a**)

[^{1}H, ^{13}C]-HSQC (full)
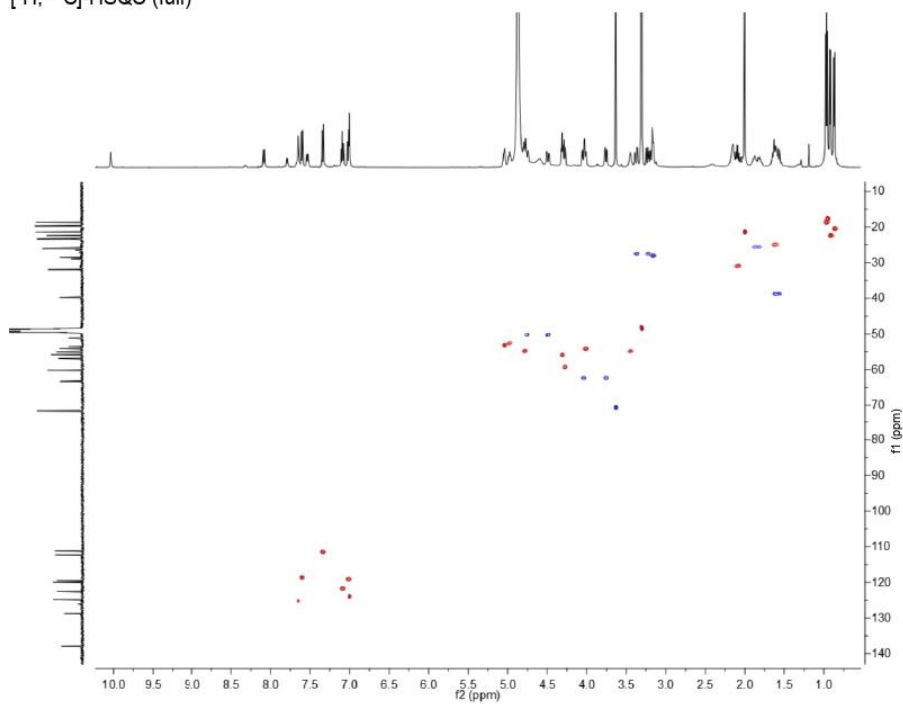
[^{1}H, ^{13}C]-HSQC (7.65, 126.2 C/H 15) Triazole

46
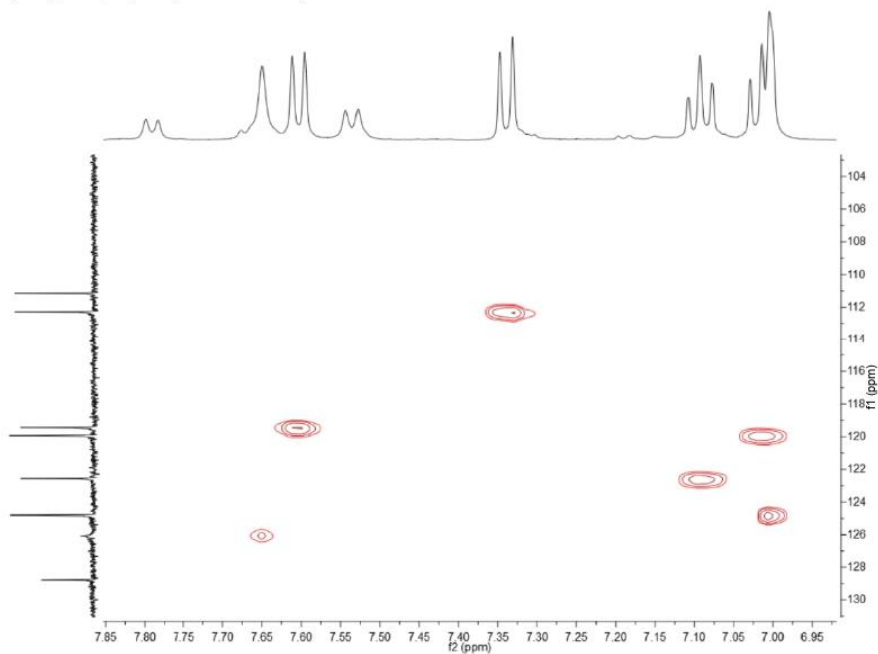
[^{1}H, ^{13}C]-HMBC (full)

[^{1}H, ^{13}C]-HMBC (C15/H17)

47
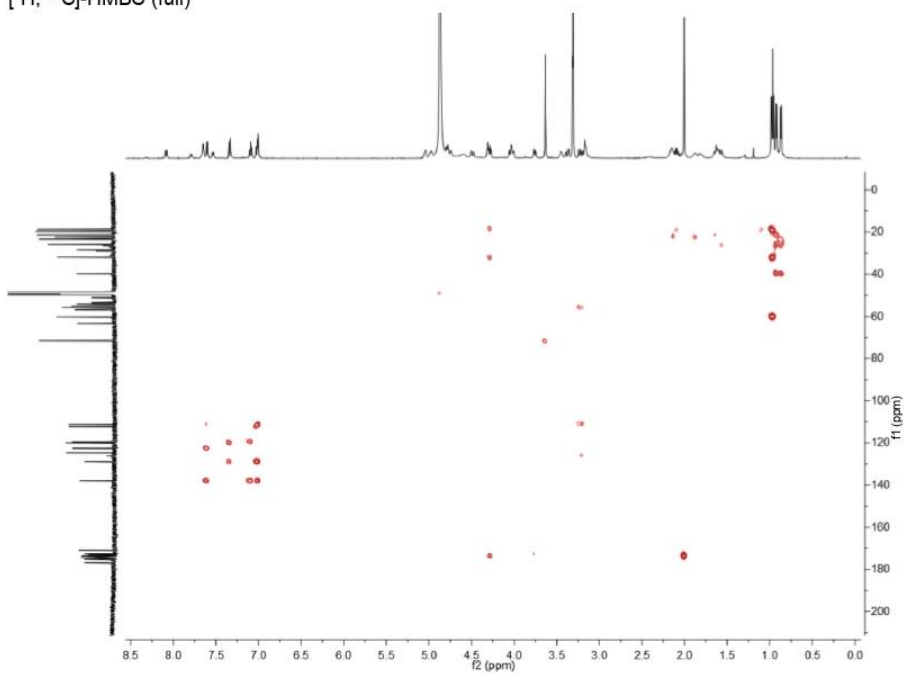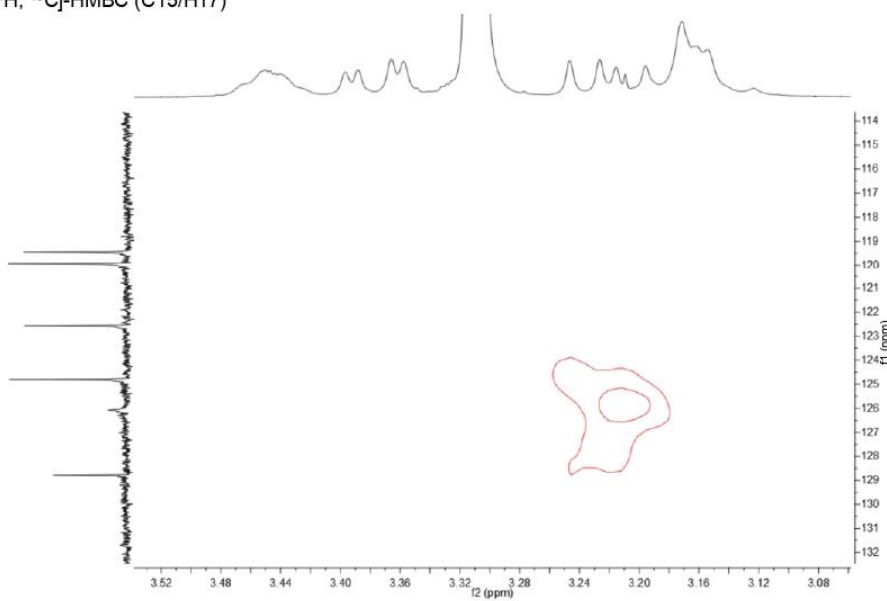
FTIR Spectra

IR spectra were measured on a PerkinElmer Spectrum 100 FT-IR Spectrometer. The samples were measured as solid.

Ac-V-[**Pra**-SEL-**Aza**]*cyclic*-W-*NH₂* (1,4-triazole) (**5a**)

Spectrum
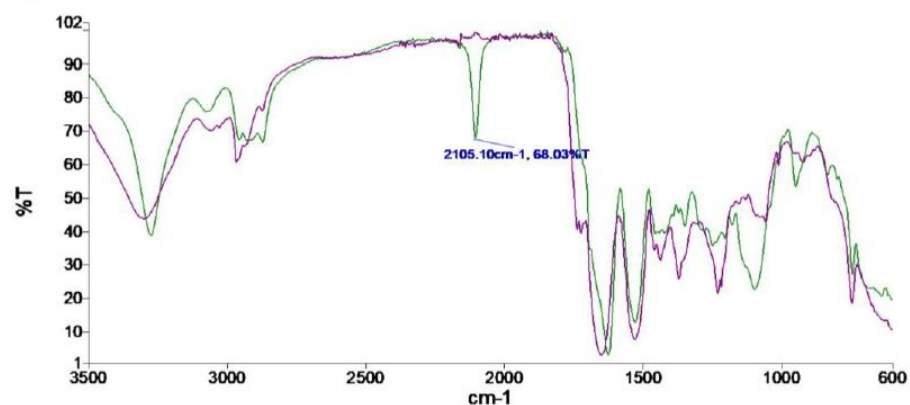
Green curve: before CuAAC
Purple curve: after CuAAC

HRMS

Compounds **1** and **5a-5c** were solved in H_2O/ACN (1:1) (v/v). HRMS was done with a DIONEX UltiMate 3000 UHPLC⁺ focused (Thermo Scientific), containing pump, autosampler, column compartment heated to 30 °C, diode array detector, and Q exactive focus. We used an EC 150/2 NUCLEODUR C18 Pyramid, 3 µm (Macherey-Nagel) column with a gradient from 90-5 % solvent B over 9 min and 1.5 min constant 5 % solvent B (solvent A: H_2O (0.05 % formic acid), solvent B: ACN (0.05 % formic acid)) and a 0.5 mL/min flowrate.

Compounds **6a**, **6b** and **6c** were analyzed by HRMS on a Shimadzu interfaced UPLC coupled to an AB Sciex 5600 TripleTOF MS using time-of-flight-MS (TOF-MS) scanning. The samples were run over a linear gradient of 20-40% acetonitrile in H_2O (v/v) on an Agilent Zorbax C18 column (100 x 2.1 mm, 1.8 µm) at 40 °C and a flow rate of 0.2 mL/min. The electrospray voltage was 5500 V with a source temperature of 500 °C. The data was processed using Analyst v1.6.3 software by AB Sciex.

Ac-V-[CSELC]*cyclic*-W-*NH2* (**1**)

H-V-[CSELC]*cyclic*-W-*NH2* (**2a**)

Ac-V-[CSELC]*cyclic*-W-*OH* (**2b**)

Ac-V**A**SELA**W**-*NH*₂ (**3a**)

Ac-**A**-[CSELC]*cyclic*-W-*NH*$_2$ (**3b**)

54

Ac-V-[C**A**ELC]*cyclic*-W-*NH₂* (**3c**)
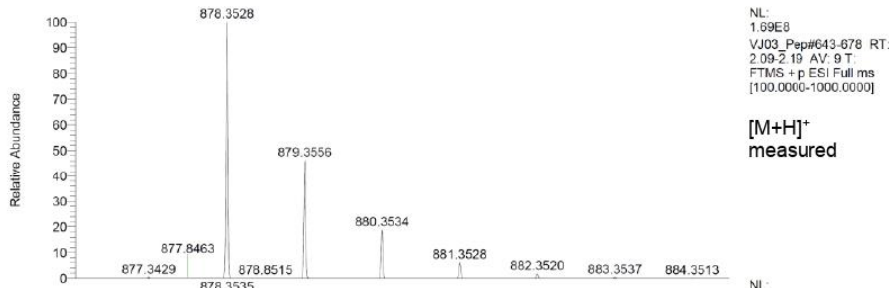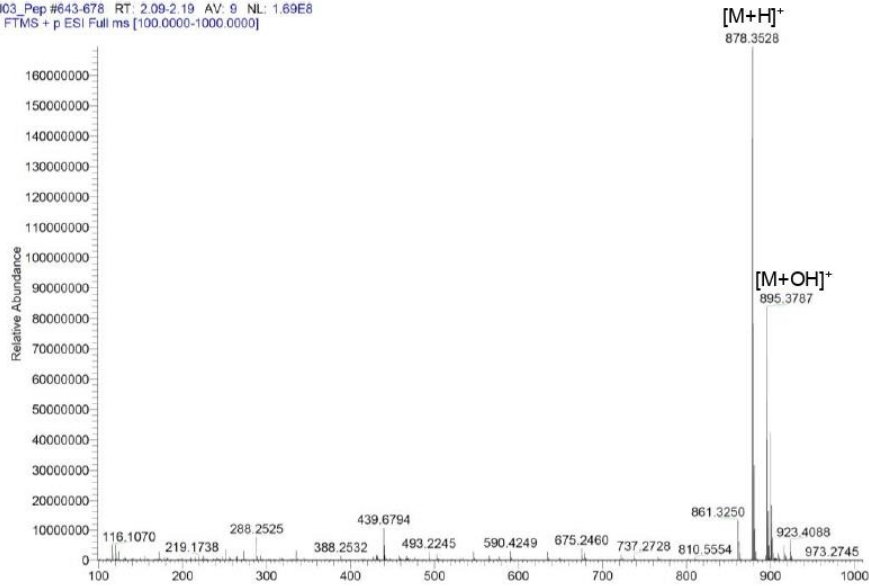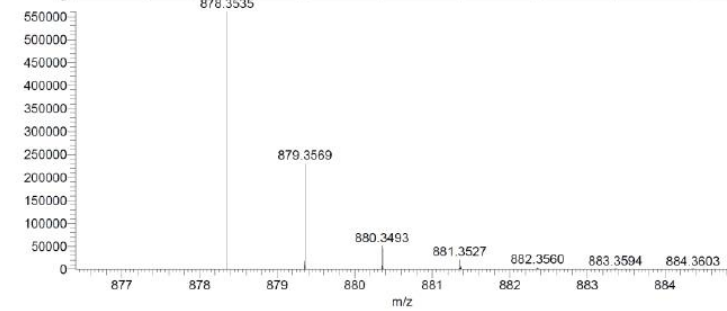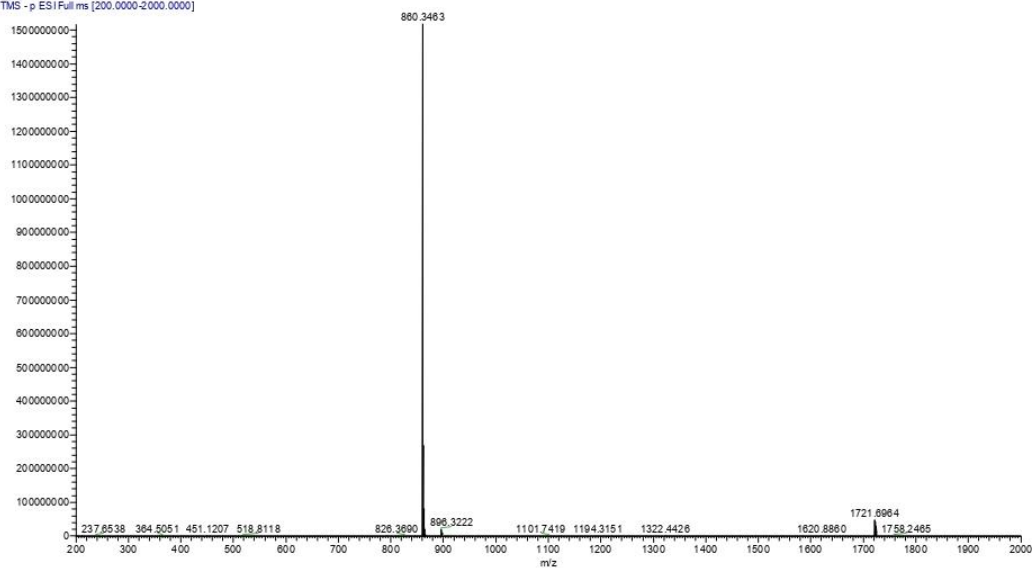
BZO_peptide3c #432 RT: 4.69 AV: 1 NL: 1.52E9
T: FTMS - p ESI Full ms [200.0000-2000.0000]

BZO_peptide3c #432 RT: 4.69 AV: 1 NL: 1.52E9
T: FTMS - p ESI Full ms [200.0000-2000.0000]

55

Ac-V-[CS**A**LC]*cyclic*-W-*NH₂* (**3d**)

Ac-V-[CSE**A**C]*cyclic*-W-*NH*2 (**3e**)

peptide_3e_2 #416   RT: 4.34   AV: 1   NL: 2.10E9
T: FTMS - p ESI Full ms [200.0000-2000.0000]

peptide_3e_2 #416   RT: 4.34   AV: 1   NL: 2.10E9
T: FTMS - p ESI Full ms [200.0000-2000.0000]

57

Ac-V-[CSELC]*cyclic*-**A**-*NH₂* (**3f**)

Ac- – – -[CSELC]*cyclic*-W-*NH₂* (**4a**)

59

Ac-V-[CSELC]*cyclic* – -*NH₂* (**4b**)

60

Ac-V-[**Pra**-SEL-**Aza**]*cyclic*-W-*NH2* (1,4-triazole) (**5a**)

Ac-V-[**Aza**-SEL-**Pra**]*cydic*-W-*NH₂* (1,4-triazole) (**5b**)

Ac-V-[**Pra**-SEL-**Aha**]*cyclic*-W-*NH2* (1,4-triazole) (**5c**)

Ac-V-[**Pra**-SEL-**Aza**]*cyclic*-W-*NH2* (1,5-triazole) (**6a**)

Ac-V-[**Pra**-SEL-**Aha**]*cyclic*-W-*NH₂* (1,5-triazole) (**6b**)

65

Ac-V-[**Aha**-SEL-**Pra**]*cyclic*-W-*NH*₂ (1,5-triazole) (**6c**)

References

[1] C.K. Maurer, M. Fruth, M. Empting, O. Avrutina, J. Hoßmann, S. Nadmid, J. Gorges, J. Herrmann, U. Kazmaier, P. Dersch, R. Müller, R.W. Hartmann, Discovery of the first small-molecule CsrA-RNA interaction inhibitors using biophysical screening technologies, Future Med. Chem. 8 (2016) 931–947. https://doi.org/10.4155/fmc-2016-0033.

[2] A.K. Dubey, C.S. Baker, T. Romeo, P. Babitzke, RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction, RNA 11 (2005) 1579–1587. https://doi.org/10.1261/rna.2990205.

[3] F. Jean-Pierre, J. Perreault, E. Déziel, Complex autoregulation of the post-transcriptional regulator RsmA in Pseudomonas aeruginosa, Microbiology (Reading) 161 (2015) 1889–1896. https://doi.org/10.1099/mic.0.000140.

[4] J. Kügler, S. Wilke, D. Meier, F. Tomszak, A. Frenzel, T. Schirrmann, S. Dübel, H. Garritsen, B. Hock, L. Toleikis, M. Schütte, M. Hust, Generation and analysis of the improved human HAL9/10 antibody phage display libraries, BMC Biotechnol. 15 (2015) 10. https://doi.org/10.1186/s12896-015-0125-0.

[5] Y. Shen, A. Bax, Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks, J. Biomol. NMR 56 (2013) 227–241. https://doi.org/10.1007/s10858-013-9741-y.

[6] A.T. Brünger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Crystallography & NMR system: A new software suite for macromolecular structure determination, Acta Cryst. D54 (1998) 905–921. https://doi.org/10.1107/s0907444998003254.

[7] V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, Acta Cryst. D66 (2010) 12–21. https://doi.org/10.1107/S0907444909042073.

[8] Molecular Operating Environment (MOE), 2020.09 Chemical computing group ULC 2020.

[9] O. Duss, E. Michel, N. Diarra dit Konté, M. Schubert, F.H.-T. Allain, Molecular basis for the wide range of affinity found in Csr/Rsm protein-RNA recognition, Nucleic Acids Res. 42 (2014) 5332–5346. https://doi.org/10.1093/nar/gku141.

[10] E. Krieger, G. Koraimann, G. Vriend, Increasing the precision of comparative models with YASARA NOVA–a self-parameterizing force field, Proteins 47 (2002) 393–402. https://doi.org/10.1002/prot.10104.

Author Contributions

Dr. Martin Empting conceived the concept of the study, interpreted the data of the presented idea and wrote parts of the manuscript.

Valentin Jakob wrote the manuscript, has done CsrA expression, performed the assays and involved in peptide synthesis and analytics.

Ben G. E. Zoller synthesized the peptides, performed the analytics and wrote parts of the manuscript.

Julia Rinkes involved in peptide synthesis and analytics and wrote some parts of the SI.

Yingwen Wu was involved in CsrA expression and in some assays, especially *E. coli* CsrA.

Dr. Alexander F. Kiefer contributed to peptide synthesis, data evaluation and processing and wrote parts of the manuscript.

Prof. Dr. Michael Hust and Saskia Helmsing taught us how to perform Phage Display and gave continuous support.

Andreas Siebert gave us an introduction how to handle the peptide synthesizer and was always a great help during the experiments with it.

The group of Prof. Dr. Uli Kazmaier provided the peptide synthesizer.

Dr. Andrew M. White synthesized compounds 6a, 6b and 6c, made parts of the analytics of these and wrote some parts of the SI.

Dr. Thomas Durek and Prof. David J. Craik were involved in planning and wrote parts of the manuscript.

Dr. Peta J. Harvey generated and deposited the NMR structures.

2. Publication Title:

Establishment of an In Bacterio assay for the Assessment of Carbon Storage Regulator A (CsrA) inhibitors

Authors:

Yingwen Wu, Ben G.E. Zoller, Mohamed Ashraf Mostafa Kamal, Sven-Kevin Hotop, Claus-Michael Lehr, Mark Brönstrup, Petra Dersch and Martin Empting

Bibliographic Data:

ChemBioChem

Research Article
doi.org/10.1002/cbic.202300369

**Chemistry
Europe**
European Chemical
Societies Publishing

www.chembiochem.org

Establishment of an In Bacterio Assay for the Assessment of Carbon Storage Regulator A (CsrA) Inhibitors

Yingwen Wu,[a, f] Ben G. E. Zoller,[a] Mohamed Ashraf Mostafa Kamal,[b, c] Sven-Kevin Hotop,[d] Claus-Michael Lehr,[b, c] Mark Brönstrup,[d] Petra Dersch,[e] and Martin Empting*[a, c, f]

Polymicrobial infections involving various combinations of microorganisms, such as *Escherichia*, *Pseudomonas*, or *Yersinia*, can lead to acute and chronic diseases in for example the gastrointestinal and respiratory tracts. Our aim is to modulate microbial communities by targeting the posttranscriptional regulator system called carbon storage regulator A (CsrA) (or also repressor of secondary metabolites (RsmA)). In previous studies, we identified easily accessible CsrA binding scaffolds and macrocyclic CsrA binding peptides through biophysical screening and phage display technology. However, due to the lack of an appropriate *in bacterio* assay to evaluate the cellular effects of these inhibitor hits, the focus of the present study is to establish an *in bacterio* assay capable of probing and quantifying the impact on CsrA-regulated cellular mechanisms. We have successfully developed an assay based on a luciferase reporter gene assay, which in combination with a qPCR expression gene assay, allows for the monitoring of expression levels of different downstream targets of CsrA. The chaperone protein CesT was used as a suitable positive control for the assay, and in time-dependent experiments, we observed a CesT-mediated increase in bioluminescence over time. By this means, the cellular on-target effects of non-bactericidal/non-bacteriostatic virulence modulating compounds targeting CsrA/ RsmA can be evaluated.

Introduction

Antimicrobial resistance (AMR) is a concerning worldwide health issue and inappropriate infection prevention and control is one factor for the steadily increasing occurrence of resistant microbes.[1] Without any counteractions, the spread of AMR will assumedly result in the deaths of 10 million people per year globally by 2050.[1]

Addressing and combatting AMR is very challenging, especially considering Gram-negative multi-drug-resistant (MDR) pathogens. Along these lines, carbapenem-resistant *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacteriaceae* are the top three on the WHO global priority pathogens list published in 2017.[2] Antibiotics or anti-infectives with alternative modes-of-action against these bacteria are therefore urgently needed.

In general, a healthy human microbiome contains diverse communities of microbes, which are stable and provide protection against colonization by pathogenic species. To suppress the growth of pathogens, commensal bacteria produce their own antimicrobial compounds such as peptides.[3] Dysbiotic communities, on the contrary, are typically less diverse and more dominated by few pathogenic species. These interactions lead to polymicrobial infections and in case these microbes are pathogenic, this can enhance the virulence of each of them. There are several examples of this effect including inhibition of competing microbes (so-called microbial interference), the mutual supply of nutrients in particular carbon sources, or subversion of immunity.[3–5]

Many acute and chronic diseases are associated with infections of the respiratory and gastrointestinal tract, where polymicrobial interactions are paramount.[4] Impacting and modulating these complex communities in order to reestablish or protect the commensal balance by small molecular entities might provide an attractive new approach for the discovery of anti-infectives. In this context, the CsrA (RsmA) protein could be considered a promising drug target. The Csr (carbon storage regulator) or Rsm (regulator of secondary metabolites) system is a post-transcriptional regulatory system, which affects mRNA

[a] Y. Wu, B. G. E. Zoller, Dr. M. Empting
     Department of Antiviral & Antivirulence Drugs (AVID)
     Helmholtz Institute for Pharmaceutical Research Saarland (HIPS)
     Saarland University, 66123 Saarbrücken (Germany)
     E-mail: martin.empting@helmholtz-hips.de

[b] M. A. M. Kamal, Prof. Dr. C.-M. Lehr
     Department of Drug Delivery (DDEL)
     Helmholtz Institute for Pharmaceutical Research Saarland (HIPS)
     Saarland University, 66123 Saarbrücken (Germany)

[c] M. A. M. Kamal, Prof. Dr. C.-M. Lehr, Dr. M. Empting
     Department of Pharmacy
     Saarland University, 66123 Saarbrücken (Germany)

[d] Dr. S.-K. Hotop, Prof. Dr. M. Brönstrup
     Department of Chemical Biology
     Helmholtz Centre for Infection Research
     German Center for Infection Research (DZIF)
     38124 Braunschweig (Germany)

[e] Prof. Dr. P. Dersch
     Institute of Infectiology
     Center for Molecular Biology of Inflammation (ZMBE)
     University of Münster, 48149 Münster (Germany)

[f] Y. Wu, Dr. M. Empting
     Cluster of Excellence RESIST (EXC 2155), Hanover Medical School
     Carl-Neuberg-Straße 1, 30625 Hanover (Germany)

🖥 Supporting information for this article is available on the WWW under
     https://doi.org/10.1002/cbic.202300369

ChemBioChem

Research Article
doi.org/10.1002/cbic.202300369

Chemistry
Europe
European Chemical
Societies Publishing

translation and/or stability, thereby regulating a multitude of cellular processes in response to environmental cues.[6–9] CsrA (RsmA) is a homodimer with two identical RNA-binding surfaces, which recognizes and binds to the GGA motifs in mRNAs. This highly conserved RNA-binding protein is widespread among Gram-negative pathogens.[6–10] For example the homology between CsrA from *Y. pseudotuberculosis* and CsrA from *E. coli* is 95%.[6] CsrA homologs can also be found in a variety of bacterial animal and plant pathogens.[6]

Early evidence showed that CsrA is not only essential for fundamental physiological properties and metabolism, but also for regulation of virulence factors required for host infection.[6] This was confirmed in previous studies by weakened virulence in murine models of e.g. *Y. pseudotuberculosis* using CsrA knock-out strains.[6,7] CsrA's activity includes for example modulation of carbon metabolism, motility, biofilm development, and quorum sensing.[7–11]

Regulation of Csr system

The complex regulation circuits of the Csr system of *E. coli* have been described in detail in reviews.[9,10] To illustrate the composition and function of the system, a simplified version is shown in Figure 1 including the essential steps relevant for the present study. In the following, the innate antagonists of CsrA will be described more in detail.

The activity of CsrA is controlled by the sequestration of the inhibitory sRNAs CsrB and CsrC (~350 nt long). Furthermore, the amount of CsrB and CsrC determines the level of free, functional CsrA, which is available for binding target mRNAs. The reason for the considered high affinity towards CsrB is for example the existence of 22 potential binding sites, which are able to sequester ~9 CsrA dimers (Figure 2). The binding element for CsrA is suggested to be the hairpin loop motif 5′-CA**GGA**UG-3′.[9,10,12] Experiments with △*csrB/C E. coli* strains and *csrB/C* overexpressing strains showed that its absence or increased abundance caused pleiotropic effects on bacterial physiology. Furthermore, expression of downstream targets regulated by CsrA such as *glgC* (responsible for glycogen biosynthesis) gene, and *flhDC* (operon for biosynthesis of flagella) is similarly affected. For example, glycogen accumulation and non-motile appearance of *Yersinia csrA* mutant strain have been observed.[6] Overall the regulatory RNAs allow the bacteria to fine-tune CsrA.[9,10,12]

Apart from the sRNA-mediated antagonism present in most CsrA/RsmA systems, there are some organisms that use innate proteins to modulate the activity of CsrA. For example, in Enteropathogenic *E. coli* (EPEC) a recently identified chaperon protein called CesT (Figure 3) binds to CsrA leading to alterations in virulence and metabolic gene expression.[9,10,13]
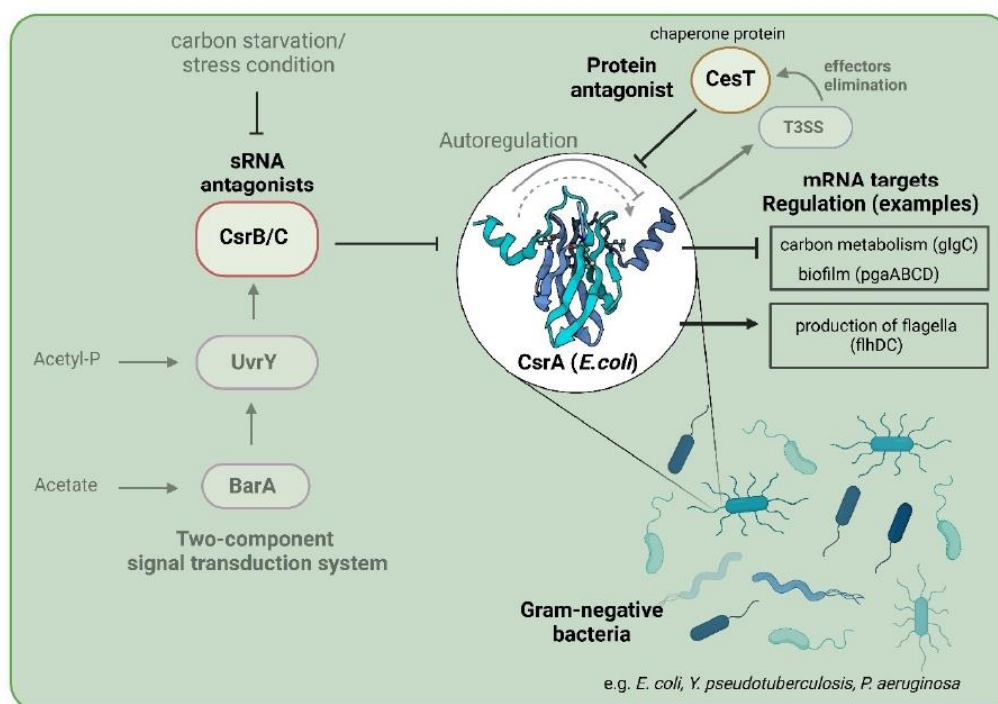
Figure 1. Simplified regulation circuit of CsrA in *E. coli*: CsrA is antagonized by sRNAs CsrB/C and chaperone protein CesT. The antagonists are controlled by other feedback cycles and regulatory circuits.[9,10] CsrA itself regulates e.g. the carbon metabolism and biofilm development by repressing the glgC (responsible for glycogen biosynthesis) gene and pgaABCD (operon for biosynthesis and secretion of biofilm polysaccharide adhesin) genes. Furthermore, CsrA activates the expression of flhDC (master operon for flagellum biosynthesis) genes to facilitate the production of flagella.

ChemBioChem

Research Article
doi.org/10.1002/cbic.202300369

Chemistry
Europe
European Chemical
Societies Publishing
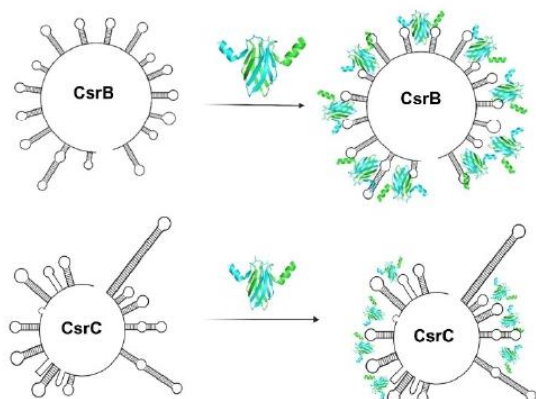
Figure 2. sRNA antagonists CsrB and CsrC of CsrA (PDB:1vpz): The affinity of CsrA for CsrB is ~ 10 fold higher than for CsrC (reported $Kd = 8.7 \pm 0.6$ nM for CsrC) in E. coli,[12] because of the large amount of the binding sites and the co-operative interaction between CsrA and CsrC transcript. Nevertheless, these two sRNAs share a similar mechanism for antagonizing the activity of CsrA. Furthermore, both CsrB and CsrC have a short half-life (~ 2 min) which indicates that CsrA is able to respond rapidly to changes in CsrB/CsrC levels.[12]
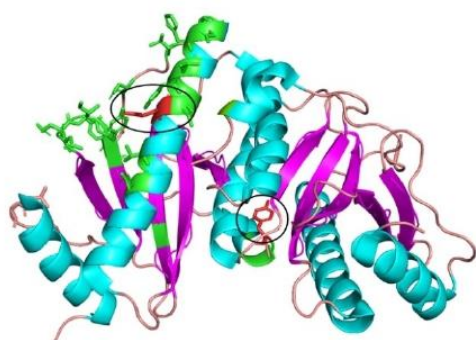
Figure 3. Structure of CesT with CsrA binding sites (PDB: 5Z38). CesT is a dimeric protein and one monomer consists of 5 β-sheets (magenta); 3 α-helices (cyan) and loops are shown in salmon. CsrA binding regions are highlighted in green and located mainly at the C-terminal region. Tyr152 and Glu121 are the important binding residues highlighted in red and encircled.[13,14]

The function of CesT is to stabilize and translocate virulence factors (effectors) that are secreted by the type three secretion system (T3SS) and are required for pathogenicity and survival in the host environment.[9,13,14] Furthermore, previous studies showed that during a T3SS-mediated bacterial infection, free CesT (not bound to T3SS effector proteins) binds to the CsrA regulator after injecting the effectors into the host cells. This results in the repression of CsrA-dependent T3SS proteins, which leads to a decrease in T3SS activity and an accumulation of the effectors that sequester CesT. These findings suggest that CsrA and T3SS activities regulate each other indirectly in a negative-feedback loop (Figure 1), while inhibition of CsrA by exogenic substances should lead to a reduction of T3SS activity.[13,14]

Development of an in bacterio assay for assessing CsrA inhibition

The aim of previous studies was to find novel inhibitors of CsrA, which are capable to disrupt the CsrA-RNA interaction.[7,8] Until now, some interesting CsrA inhibitor scaffolds have been identified using different biophysical screening methods as well as phage display. The discovered hit structures have been tested in a fluorescence polarization (FP) assay for their ability to replace the RNA from CsrA.[8] One of the most active synthetic compounds, which is a triazole peptide, showed an IC_{50} value in a single-digit micromolar range.[8]

However, the biophysical assay reflects the impact on the protein-RNA interaction in a cell-free setup. Since CsrA is a target for pathoblocker compounds, which ideally lack any bactericidal or bacteriostatic effects, typical antibacterial assays like minimum inhibitory concentration (MIC) assays are not suitable. Finding an appropriate *in bacterio* assay, which enables to probe and quantify the impact on CsrA-regulated cellular mechanisms, is challenging. To this end, we employed a (combination of qPCR and) luminescence-based assay setup towards the establishment of an *in bacterio* CsrA inhibition assay.

Results and Discussion

To study inhibitory effects on target protein levels in real-time, the bioluminescence of bacterial luciferases can be exploited. These enzymes emit light in the presence of the substrate luciferin (reduced riboflavin phosphate), which is oxidized to a long-chain aldehyde.[15] Expression of the bacterial-derived *luxCDABE* operon leads to cells emitting detectable light at 490 nm. This operon encodes the luciferase (LuxAB) and the substrate-producing enzymes (LuxCDE).[15] For our assay we used a vector (pvBE3) containing the *glgC-luxCDEAB* reporter fusion harboring the entire promoter region of *glgC* (which is negatively regulated by CsrA). As a consequence, in the presence of functional CsrA inhibitors, the bioluminescence signal is expected to increase due to the upregulation of target gene (*glgC*) expression.

In order to evaluate a *glgC-lux*-based assay results, we explored the suitability of the innate antagonists as shown in Figure 1 as control. Since CsrB is the main sRNA antagonist of CsrA in *E. coli*, it was the first positive control that we considered. Using a *lactose*-inducible *csrB* expression plasmid, we induced overexpression of CsrB by IPTG (500 μM) treatment. However, unlike what we expected, the bioluminescence decreased after 4 h incubation time (Figure 4A). To gain a better understanding of this outcome, we compared the transcript levels of the individual Csr components by a qPCR gene expression assay after overexpression of CsrB or CsrC (Figure 4B).

We investigated the level of expression of both sRNAs in different strains 4 h after IPTG induction. We could confirm the successful increase of sRNA levels (in a range between 100 to

ChemBioChem

Research Article
doi.org/10.1002/cbic.202300369

Chemistry
Europe
European Chemical
Societies Publishing
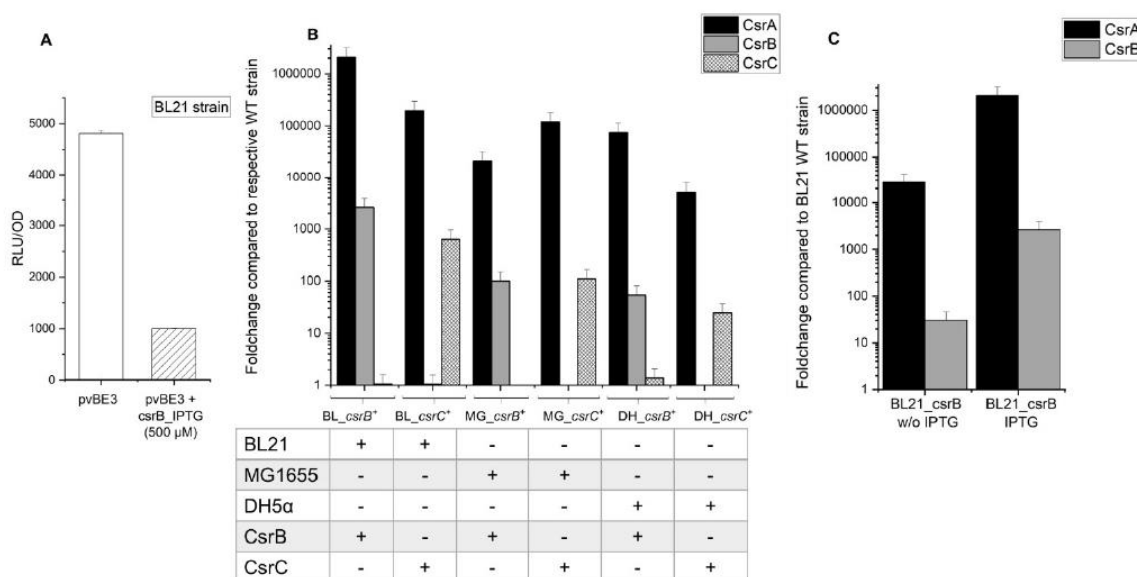
Figure 4. Influence of CsrB overexpression on the luciferase reporter gene (A) and validation of sRNAs' expressions via qPCR assays (B, C). Error bars represent the standard deviation of four replicates. (A): *E. coli* strains BL21 with only the plasmid pvBE3 (*glgC-lux*) or with additional plasmid pET28a(+) harboring the *csrB*+ were grown at 37 °C until the exponential phase was reached (OD = 0.6). The expression of CsrB was induced with 500 μM IPTG. After 4 h of induction the relative light units (RLU) were almost 5-fold less compared to the RLU of pvBE3. All data were normalized over OD_{600}. (B): *E. coli* strains BL: BL21 + pvBE3, MG: MG1655 with a constituent T7 promoter + pvBE3, DH: DH5α + pvBE3 with or without the plasmid pET28a(+) (*csrB*+) and (*csrC*+) were grown at 37 °C until exponential phase was reached (OD = 0.6). The wild type strains are mentioned *E. coli* strains without any plasmids. The expression of CsrB or CsrC was induced with 500 μM IPTG. After 4 h of induction with IPTG, total RNA was extracted from the cultures and a qRT-PCR analysis was performed. The results showed that both CsrB and CsrC RNA were increased, but also the expression of CsrA (black bars) was induced to a far greater extent in all strains. C) Results of qPCR expression assay with and without inducing CsrB: *E. coli* strains BL21 pvBE3, with the plasmid pET28a(+) (*csrB*+) were grown at 37 °C until the exponential phase was reached (OD = 0.6). The expression of *csrB* was not induced (w/o) or induced with 500 μM IPTG. After 4 h of induction the relative light units (RLU) were determined. All data were normalized over OD_{600}. The basal expression level of *csrA* was higher than in the wild type strain and exceeded the level of CsrB without IPTG. These results suggest that an increased amount of antagonistic sRNAs leads to a feedback mechanism resulting in higher production of CsrA.

3000 fold). However, also CsrA-encoding transcripts were drastically increased (in a range between 5000 to 20.000.000 fold), indicating that CsrB and CsrC overexpression was overcompensated by a 100- to 1000-fold higher *csrA* expression. We further found that strains harboring the *csrB* overexpression plasmid have a significantly higher *csrA* transcript level than in the wild type strain even without IPTG-mediated induction (Figure 4C p = 0.0128, calculated using the t-test over the data from BL21 strain harboring csrB without IPTG compared to BL21 strain harboring csrB with IPTG). Ultimately, these qPCR results explained the observation we had from the reporter gene assay.

The results implicated that the induction of sRNA expression triggered an unknown autoregulatory control circuit of the Csr system. Autoregulation of the Csr components has been described, but these studies only report the successful inhibition of the CsrA activity by CsrB/C shown as a *csrA* knock-out phenotype and changes in glucose consumption and free fatty acids production.[16,17] Due to the observed interdependency of the Csr components, using CsrB/C as positive controls proved to be difficult. Therefore, we decided to examine the protein antagonist CesT. This chaperone is reported to inhibit CsrA's activity and should not have an impact on the *csrA* transcript

level.[13,14] To ensure that the induction of CesT does not increase CsrA expression, the qPCR gene expression assay described above was applied (Figure 5B). A first observation was, that IPTG addition did not increase *cesT* expression over the basal (uninduced) levels. Importantly, in comparison to the results for CsrB/C, CsrA levels were less, but still affected compared to the wild type in the presence of IPTG (~10-fold), whereas no influence was observed in the absence of IPTG. This effect of the thio-sugar derivative IPTG on *csrA* expression might be linked to the involvement of CsrA in the post-transcriptional control of sugar metabolism.[9,10,17]

The promising results for the *cesT*-harboring plasmid-bearing strain in the absence of IPTG, encouraged us to rely on basal (uninduced) expression in follow-up experiments. We performed the reporter gene assay using the same condition and could observe a convincing increase in bioluminescence, indicating a derepression of the *glgC-lux* fusion in the presence of the *cesT*+ plasmid (Figure 5A). This suggested that the assay setup might be suitable for the identification and investigation of CsrA inhibitors. As a next step, we performed time-resolved experiments to gain insights into the kinetics of CesT-driven inactivation of CsrA in order to identify the most suitable

99

ChemBioChem

Research Article
doi.org/10.1002/cbic.202300369

Chemistry
Europe
European Chemical
Societies Publishing
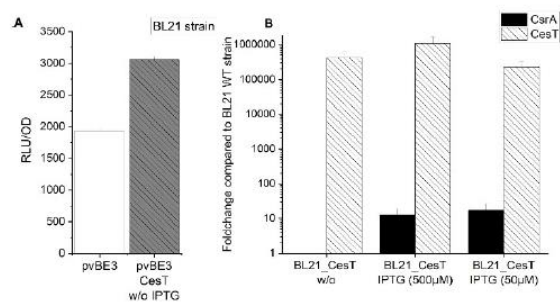
Figure 5. Influence of CesT overexpression on the luciferase reporter gene (A) and validation of *cesT* expression on *glgC-lux* transcription by qPCR gene expression assays (B). Error bars represent the standard deviation of four replicates. (A): *E. coli* strains BL21 pvBE3 with or without the plasmid pNS6236 (*cesT*+) were grown at 37 °C until the exponential phase was reached (OD=0.6) and the relative light units (RLU) were determined. All data were normalized over OD_{600}. (B): *E. coli* strains BL21 pvBE3 with or without the plasmid pNS6236 (*cesT*+) were grown at 37 °C until the exponential phase was reached (OD=0.6). The expression of *cesT* was induced with 50 or 500 μM IPTG. After 4 h of induction with IPTG, total RNA was extracted from the cultures and a qRT-PCR analysis was performed. The presence of the p*cesT*+ had no impact on *csrA* expression in the absence of IPTG.

incubation time for yielding marked effects enabling facile detection of inhibitory activities (Figure 6).

A steady increase of RLU was observed in the *cesT*+ strain BL21 pvEB3, pNS6236 over the course of 5 h, while values of the reference strain BL21 pvBE3 remained essentially unchanged (Figure 6A, also Figure S3). At the end of the five-hour experiment, we determined the most prominent effect, where the RLU of the *cesT*+ strain was ~3-fold higher than the RLU of the control strain. In parallel, we also performed qPCR to monitor the expression levels of *cesT* and *csrA* over the time course of the experiment. The expression level of *csrA* increased 2–40-fold within 2–5 h (Supplementary Figure S2). However, the transcript level of *cesT* was about 10,000-fold higher, and no impact on *glgC-lux* expression has been observed. In order to have enough samples for both assays, we started with a volume of 100 mL culture. However, for efficient compound testing, this large amount of culture is not suitable, because consequently high amounts of potential inhibitors are required. For this reason, we decreased the starting volume from 100 mL over 10 mL down to 200 μL, which was suitable for 96-well format (Figure 6B, C). Results were reproducible. In case of the 10 ml format, the expression patterns looked more defined and the induction of the reporter was more pronounced over time. However, the 200 μL format also yielded well-defined reproducible results clearly enabling to discriminate basal expression from the positive control (Figure 6C). Hence, we consider this assay fit-for-purpose for subsequent compound library screening in the future using the small 96-well plate format.

Based on the data gathered so far, we decided to use 10 mL cultures for the testing of compounds that were previously reported to disrupt the CsrA-RNA interaction in a cell-free environment for the first time.[8] We used disulfide- and triazole-macrocyclized peptidic CsrA inhibitors identified in our previous
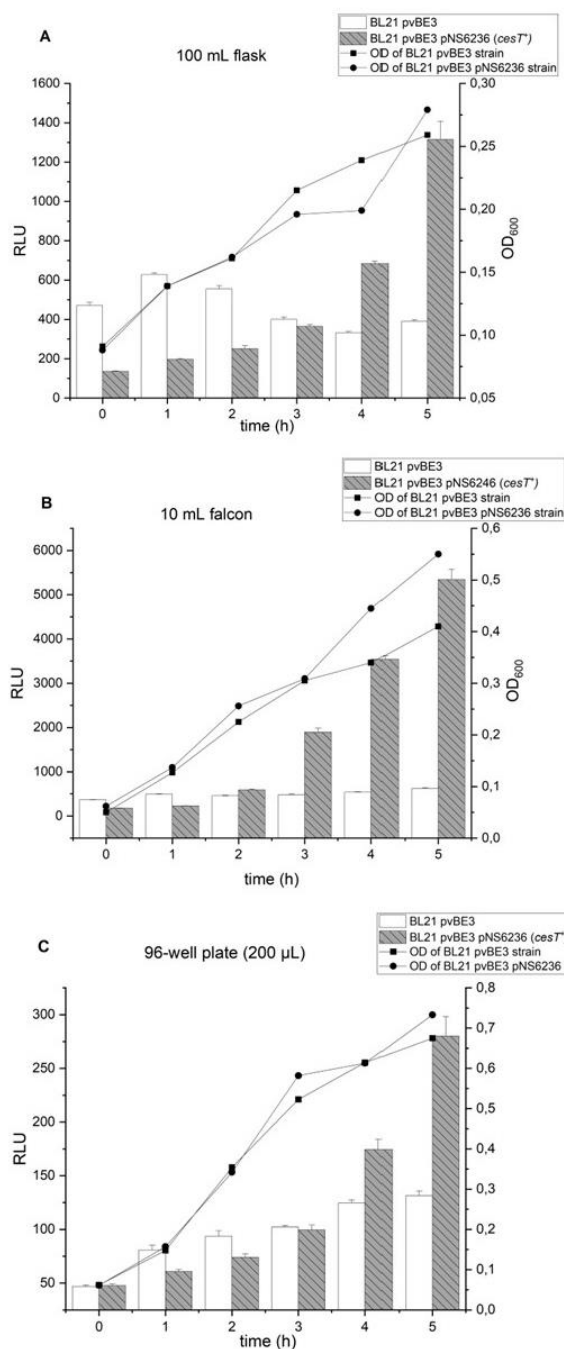
Figure 6. Time-dependent reporter gene assay including the positive control CesT and using decreasing assay volumes. Error bars represent the standard deviation of four replicates. (A) 100 ml assay volume run in 300 mL flasks: *E. coli* strains BL21 pvBE3 with or without the plasmid pNS6236 (*cesT*+) were grown at 37 °C for 5 hours and the relative light units (RLU) and OD600 were determined. (B) 10 mL assay volume run in 50 mL falcon tubes. (C) 200 μL assay volume run in 96-well microtiter plates (for details see material and methods section).

ChemBioChem

Research Article
doi.org/10.1002/cbic.202300369

Chemistry
Europe
European Chemical
Societies Publishing

study as these were among the most active compounds showing IC$_{50}$ values in a single-digit micromolar range in a fluorescence polarization assay.[8] In contrast to the CesT-expressing positive control, the addition of the inhibitory peptides did not lead to an increase in RLU values, and thus *glgC-lux* expression after 5 hours (Figure 7). We hypothesized that the inability of the peptides to inhibit CsrA in the *in bacterio* assay was due to their difficulty to penetrate the Gram-negative cell membranes and enter the cytoplasm to reach the target protein. This was verified by subcellular quantification of uptake in *E. coli* (supporting information). We observed that the triazole-based peptide **1** reached only nanomolar levels in the cytoplasm, although an extracellular concentration of 28 μM was applied (Figure S4). The intracellular concentrations were clearly insufficient to disrupt the CsrA-RNA interaction given the micromolar potency of the peptide in the target-based assay.

Nevertheless, we consider the general reporter gene assay concept now fit-for-purpose to facilitate quantitative compound evaluation with the aim to identify novel inhibitors with cellular efficacy.

Conclusions

In summary, we could establish an *in bacterio* assay, which directly measures the inhibition of CsrA based on a luciferase reporter gene assay. We could show that the expression of the chaperone protein CesT can be used as a suitable positive control for the assay because it acts as a natural intracellularly expressed antagonist of CsrA, which does not cause compensatory feedback effects. We could monitor a CesT-mediated increase of the bioluminescence over time with the most convincing effect being detectable after 5 hours of incubation.

Another interesting finding from our study is the enhancement of *csrA* transcript levels in sRNAs overexpression strains. Even without IPTG-mediated induction of *csrB* and *csrC* expression, we found a higher level of *csrA*-encoding transcripts. We suggest that a yet unknown autoregulatory control circuit of the Csr system causes this feedback mechanism. However, previous studies by other research groups showed the successful inhibition of CsrA's activity by the sRNAs through phenotypical results and changes in downstream targets of CsrA.[16,17] Hence, the interdependency of the Csr components observed in the frame of this study deserves further investigation.

One of the advantages of the established *glgC-lux* luciferase reporter assay setup is for instance the direct readout of potential CsrA inhibition. Even though the regulation of the Csr system is complex and contains multiple feedback mechanisms, this *in bacterio* assay has a well-detectable and stable read-out in the presence of the natural antagonist CesT. That means once CsrA is less active (due to inhibition), we can directly monitor its consequence and impact through this assay. Thus, it is sensitive towards the activity of potential pathoblockers. In addition, quantitative evaluation of the cellular effect (determination of EC$_{50}$ values) of promising new inhibitors should be possible via concentration-dependent experiments as well as potentially gaining insights into the regulatory kinetics with the time-dependent measurement setup. Moreover, in combination with a qPCR expression gene assay, we can even observe the expression levels of different downstream targets of CsrA.

Downscaling of the required culture volume to 96-well format was successful enabling high throughput testing of potential inhibitors. Using this reporter gene assay set up for phenotypic screening from commercial synthetic or natural product libraries is highly favorable and is one of the next major steps towards tackling this challenging virulence-modulating target. The previously reported disruption of CsrA/RsmA-RNA interactions *in vitro* using the target protein from multiple species holds promise for the identification of anti-infectives/virulence modulators with broader anti-Gram-negative activity.[8]

Figure 7. Influence of peptidic CsrA inhibitors on the expression of the CsrA-dependent *glgC-lux* fusion: *E. coli* strain BL21 pvBE3 in the presence of 5% DMSO, 125 μM triazole peptide **1**, 125 μM disulfide peptide **2** or the plasmid pNS6236 (*cesT⁺*) were grown at 37 °C for 5 hours and the relative light units (RLU) were determined each hour. Error bars represent the standard deviation of four replicates.

ChemBioChem

Research Article
doi.org/10.1002/cbic.202300369

**Chemistry
Europe**
European Chemical
Societies Publishing

Experimental Section

Bacterial strains and culture conditions

All strains and plasmids used in this study are described in the supplementary Table S2. Unless otherwise indicated, bacterial strains were routinely grown in LB medium at 37 °C containing the following antibiotics with respective final concentrations: ampicillin (100 μg/mL) and kanamycin (50 μg/mL).

Luciferase reporter gene assay

E. coli (BL21, DH5alpha, MG1655) pvBE3 with respective plasmids of inducible genes was grown in different vessels ranging from 10 mL (in falcon tubes) to 100 mL (in shaking flasks) at 37 °C to exponential phase ($OD_{600} = 0.6$). Subsequently, IPTG (500 μM) was added to induce the expression of CsrB, CsrC, or CesT. After 4 hours the cultures were added into the wells of a microtiter plate (200 μL per well) and luminescence (relative light units – RLU) was measured in triplicate. In addition to that, optical density (OD) at 600 nm was measured in 1:10 dilution.

Time-dependent measurement

Flask and falcon tube format: *E. coli* BL21 pvBE3 and *E. coli* BL21 pvBE3 with an additional plasmid carrying the *cesT* gene were grown at 37 °C to exponential phase ($OD_{600} = 0.6$). Subsequently, compound (500 μM; 250 μM; 125 μM, final concentrations) and DMSO (5%) were added each to *E. coli* BL21 pvBE3. After 5 min the first measurement was performed (time point 5 min). The culture was added to the measuring plates (200 μL per well) and luminescence (relative light units – RLU) was determined in triplicate. Afterwards, measurements were done every hour via the same procedure. In addition to that, optical density (OD) at 600 nm was measured in 1:10 dilution.

Microtiterplate format: *E. coli* strains BL21 pvBE3 with or without the plasmid pNS6236 (cesT+) were grown at 37 °C until OD = 0,6 is reached. Afterwards, cultures were diluted in LB medium to OD = 0.06. 100 μL of diluted cultures were transferred into the 96-well plates preloaded with 100 μL LB and 5% DMSO per well. (200 μL per well in total). The relative light units (RLU) and OD600 were determined directly from one plate every hour.

Isolation of total RNA

The total amount of the cellular RNA from each culture was isolated using the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. To avoid DNA contaminations, DNA digestions were conducted with DNase (Qiagen, Hilden, Germany) for 15 min. RNA was quantified by its absorbance at 260 nm and 280 nm using NanoDrop™. RNA samples were stored at −20 °C for only one-time usage.

Reverse transcription PCR (RT-PCR)

Reverse transcription was conducted using Applied Biosystems™ High-Capacity cDNA Reverse Transcription Kit (fisher scientific, USA). The reaction mixture (20 μL) contained 100 ng of RNA and master mix with reverse transcriptase. The conditions for the PCR were: 25 °C–10 min, 37 °C–120 min, 85 °C–5 min. cDNA products were either used directly for qPCR or stored at −20 °C.

Quantitative PCR (qPCR)

The qPCR was performed using SYBR Green master mix (Thermofisher Scientific, Germany) and respective primers listed in the supplementary information (Table S3). The samples consisted of 10 μL master mix, 0,5 μL cDNA product, 7,5 μL H_2O and 2 μL primers. Reactions for each sample were performed with StepOnePlus™ Real-Time PCR system (Thermofisher Scientific, Germany). The conditions for the qPCR were: 50 °C–2 min, 90 °C–2 min (holding stage), 95 °C–15 sec, 60 °C–1 min (40 cycles in the cycling stage), 95 °C–15 sec, 60 °C–1 min, 95 °C–15 sec (melt curve stage). The difference in cycle threshold (△CT) between control samples (wildtype MG1655, BL21 and DH5alpha strains) and treated samples (strains with plasmids containing inducible *csrB*, *csrC*, and *cesT* genes) was calculated using the Comparative Cτ (△△Cτ) Quantification method. Expression of individual genes was normalized against the *rpoD* and *opgD* genes. All the results were calculated and analyzed using Excel (Microsoft). The resulting values represent the mean expression level of duplicates from one qPCR assay.

Supporting Information

The authors have cited additional references within the Supporting Information.[18–19]

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

[1] S. A Strathdee, S. C Davies, J. R Marcelin, *Lancet* **2020**, 1050–1052.
[2] E. Tacconelli, E. Carrara, A. Savoldi, S. Harbarth, M. Mendelson, D. L. Monnet, C. Pulcini, G. Kahlmeter, J. Kluytmans, Y. Carmeli, M. Ouellette, K. Outterson, J. Patel, M. Cavaleri, E. M. Cox, C. R. Houchens, M. L. Grayson, P. Hansen, N. Singh, U. Theuretzbacher, N. Magrini, A. O. Aboderin, S. S. Al-Abri, N. Awang Jalil, N. Benzonana, S. Bhattacharya, A. J. Brink, F. R. Burkert, O. Cars, G. Cornaglia, O. J. Dyar, A. W. Friedrich, A. C. Gales, S. Gandra, C. G. Giske, D. A. Goff, H. Goossens, T. Gottlieb, M. Guzman Blanco, W. Hryniewicz, D. Kattula, T. Jinks, S. S. Kanj, L. Kerr, M.-

P. Kieny, Y. S. Kim, R. S. Kozlov, J. Labarca, R. Laxminarayan, K. Leder, L. Leibovici, G. Levy-Hara, J. Littman, S. Malhotra-Kumar, V. Manchanda, L. Moja, B. Ndoye, A. Pan, D. L. Paterson, M. Paul, H. Qiu, P. Ramon-Pardo, J. Rodríguez-Baño, M. Sanguinetti, S. Sengupta, M. Sharland, M. Si-Mehand, L. L. Silver, W. Song, M. Steinbakk, J. Thomsen, G. E. Thwaites van der Meer, Jos WM, N. van Kinh, S. Vega, M. V. Villegas, A. Wechsler-Fördös, H. F. Wertheim, E. Wesangula, N. Woodford, F. O. Yilmaz, A. Zorzet, *The Lancet Infect. Dis.* **2018**, *18*, 318–327.

[3] A. L. Welp, J. M. Bomberger, *Front. Cell. Infect. Microbiol.* **2020**, *10*, 213.

[4] K. A. Brogden, J. M. Guthmiller, C. E. Taylor, *The Lancet* **2005**, *365*, 253–255.

[5] E. M. Selleck, M. S. Gilmore, *mBio* **2016**, *7*, 1–3.

[6] A. K. Heroven, K. Böhme, P. Dersch, *RNA Biol.* **2012**, *9*, 379–391.

[7] C. K. Maurer, M. Fruth, M. Empting, O. Avrutina, J. Hoßmann, S. Nadmid, J. Gorges, J. Herrmann, U. Kazmaier, P. Dersch, R. Müller, R. W. Hartmann, *Future Med. Chem.* **2016**, *8*, 931–947.

[8] V. Jakob, B. G. Zoller, J. Rinkes, Y. Wu, A. F. Kiefer, M. Hust, S. Polten, A. M. White, P. J. Harvey, D. J. Craik, A. Siebert, U. Kazmaier, M. Empting, *Eur. J. Med. Chem.* **2022**, *231*, 114148.

[9] T. Romeo, P. Babitzke, *Microbiol. Spectrum* **2018**, *6*, 1–13

[10] C. A. Vakulskas, A. H. Potts, P. Babitzke, B. M. Ahmer, T. Romeo, *Microbiol. Mol. Biol. Rev.* **2015**, *79*, 193–224.

[11] V. Berndt, M. Beckstette, M. Volk, P. Dersch, M. Broenstrup, *Sci. Rep.* **2019**, *9*, 1–15.

[12] T. Weilbacher, K. Suzuki, A. K. Dubey, X. Wang, S. Gudapaty, I. Morozov, C. S. Baker, D. Georgellis, P. Babitzke, T. Romeo, *Mol. Microbiol.* **2003**, *48*, 657–670.

[13] N. Katsowich, N. Elbaz, R. R. Pal, E. Mills, S. Kobi, T. Kahan, I. Rosenshine, *Science* **2017**, *355*, 735–739.

[14] M. Yadav, M. Srinivasan, N. K. Tulsian, Y. X. Liu, Q. Lin, I. Rosenshine, J. Sivaraman, *Protein Sci.* **2021**, *30*, 2433–2444.

[15] F. Uliczka, F. Pisano, A. Kochut, W. Opitz, K. Herbst, T. Stolz, P. Dersch, *PLoS One* **2011**, *6*, 1–12.

[16] Tony Romeo, *Mol. Microbiol.* **1998**, *29*, 1320–1330.

[17] A. E McKee, B. J Rutherford, D. C Chivian, E. K Baidoo, D. Juminaga, D. Kuo, P. I Benke, J. A Dietrich, S. M Ma, A. P Arkin, C. J Petzold, P. D Adams, J. D Keasling, S. R Chhabra, *Microb. Cell Fact.* **2012**, *11*, 1–12.

[18] H. Prochnow, V. Fetz, S.-K. Hotop, M. A. García-Rivera, A. Heumann, M. Brönstrup, *Anal. Chem.* **2019**, *91*, 1863–1872.

[19] V. R. Matias, A. Al-Amoudi, J. Dubochet, T. J. Beveridge, *J. Bacteriol.* **2003**, *185*, 6112–6118.

ChemBioChem

Supporting Information

Establishment of an In Bacterio Assay for the Assessment of Carbon Storage Regulator A (CsrA) Inhibitors

Yingwen Wu, Ben G. E. Zoller, Mohamed Ashraf Mostafa Kamal, Sven-Kevin Hotop, Claus-Michael Lehr, Mark Brönstrup, Petra Dersch, and Martin Empting*
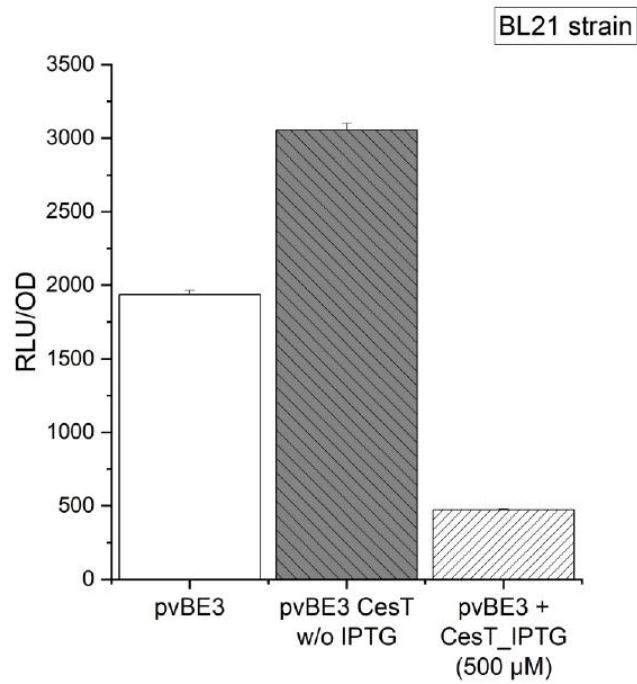
Figure S1. Analysis of *glgC-lux* expression in the presence and absence of CesT: *E. coli* strains BL21 pvBE3 with or without the plasmid pNS6236 (*cesT*+) were grown at 37°C until the exponential phase was reached (OD = 0.6) with or without IPTG and the relative light units (RLU) were determined. A strong reduction in bioluminescence was observed with BL21 pvBE3 overexpressing *cesT*. This confirmed the results of the qPCR analysis which showed an increased *csrA* transcript level. Error bars represent the standard deviation of four replicates
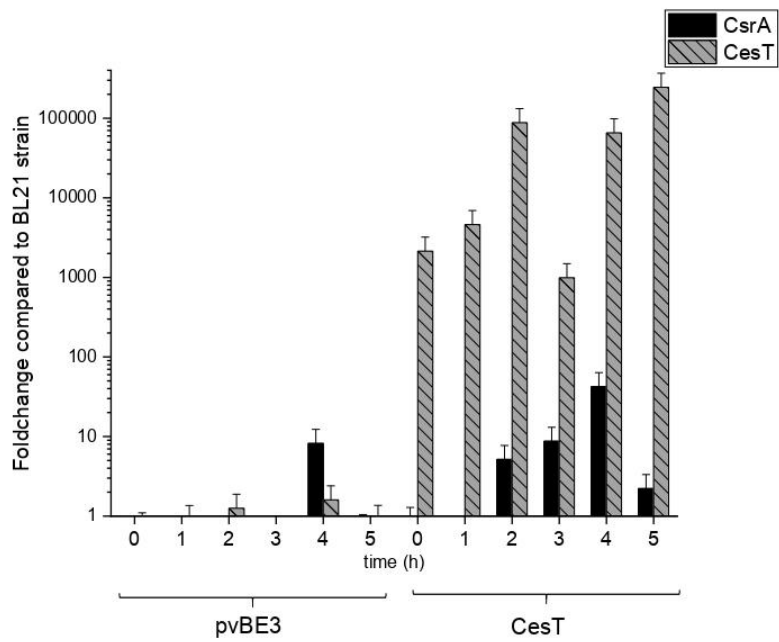
Figure S2. Monitoring the transcript levels of CsrA and CesT over 5 hours via qPCR: *E. coli* strains BL21 pvBE3 with or without the plasmid pNS6236 (*cesT*⁺) were grown at 37°C for 5 hours with or without IPTG and the relative light units (RLU) were determined each hour. The expression level of *csrA* was increased after 2-5 h of 2-40-fold *cesT* transcript levels were much higher (~1000-fold). Error bars represent the standard deviation of four replicates
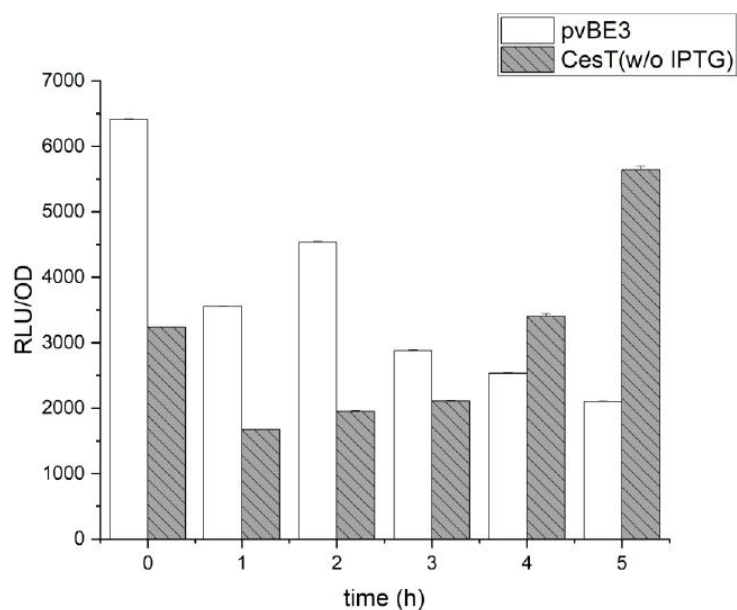
Figure S3. Time-dependent *glgC-lux* expression in the presence and absence of CesT. *E. coli* strains BL21 pvBE3 with or without the plasmid pNS6236 (*cesT*+) were grown at 37°C for 5 hours without IPTG and the relative light units (RLU) were determined. This depicts the same data as Figure 6A in the main text, but RLU values were normalized with OD_{600}. RLU values of the control strain (BL21 pvBE3) were reduced, whereas the RLU values of BL21 pvBE3 harboring plasmid pNS6236 (*cesT*+) increased with time. Error bars represent the standard deviation of four replicates.

Cellular uptake experiments:

E coli strain DSM1116 ATCC9637 was grown in Mueller Hinton Broth. The procedure of the subcellular fractionation was done according to the literature (15). The applied concentration of the compounds was 28 µM. For LC-MS analysis, protein precipitation was performed in deep 96 well plates. 80 µl of the sample was mixed with 80 µl of 1% formic acid, 120 µl acetonitrile and 120 µl methanol. Then the plate was centrifuged for 60 min at 2250 g at 4 °C. 320 µl of the supernatant were transferred, dried, and resuspended in 40 µl with caffeine as internal standard. The samples were analyzed using Agilent AdvanceBio Peptide Map 2.1 x 100 mm 2.7-Micron + AdvanceBio Peptide Map Guard 2.1x5 mm 2.7 Micron columns on Agilent 1290 UHPLC (Agilent Technologies, Santa Clara, CA, USA) coupled with AB Sciex QTrap 6500 triple quadrupole mass spectrometer (AB Sciex Germany GmbH, Darmstadt, Germany).

Table S1: Triple quadrupole MS-MS parameters

	Q1 (mass) [g/mol]	Q3 (mass) [g/mol]	Declustering potential [V]	Collision energy [V]	Collision cell exit potential [V]
Caffeine (IS)					
quantifier	195.116	138.1	81	27	10
qualifier	195.116	110.1	81	31	6
BZO153					
quantifier	895.409	692.3	196	41	44
qualifier	895.409	878.4	196	29	18
BZO164					
quantifier	881.384	864.2	201	25	18
qualifier	881.384	678.2	201	37	48
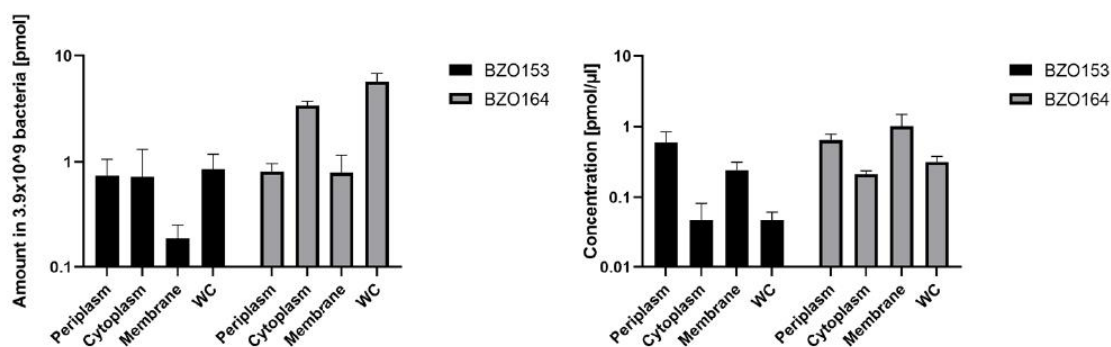
Figure S4: Subcellular quantification of the triazole peptides BZO153 and BZO164 in *E. coli*. Whole cell extracts (WC) were prepared and fractionated into a periplasm, cytoplasm, and membrane fraction, followed by peptide concentration measurements using LC/MS/MS. The graphs show (a) the amount (log scale) of the compounds in different bacterial compartments and (b) the concentration (log scale) of the compounds in different bacterial compartments normalized to the volumes as published.[18,19] The concentration of BZO153 in cytoplasm is 0.047 ± 0.035 µM and for BZO164 the concentration was 0.21 ± 0.023 µM. Error bars represent the standard deviation of three replicates.

1   **3**

Ac-V-[Pra-S-E-L-Aza]$_{cyclic}$-W-NH$_2$ Ac-V-[Aza-S-E-L-Pra]$_{cyclic}$-W-NH$_2$
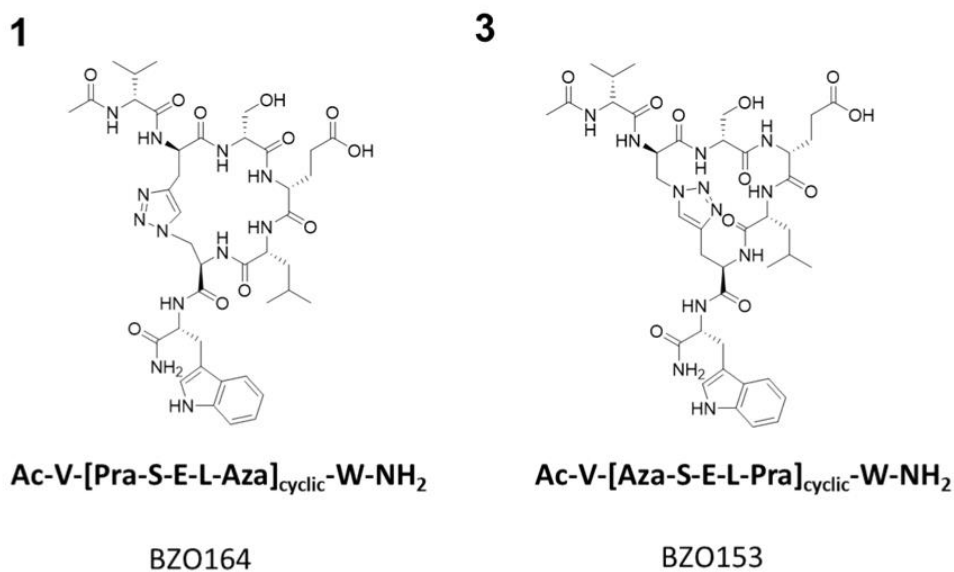
BZO164   BZO153

Figure S5: Chemical structures of the triazole peptides BZO164 **1** and BZO153 **3**.

Table S2. Bacterial strains

Strain	Abbreviation	Description	Notes [Reference from the main text]
BL21		F⁻*ompT hsd*S$_B$ (r$_B$⁻, m$_B$⁻) *gal dcm* (DE3)	
DH5α		F⁻ φ80*lacZ*ΔM15 Δ(*lacZ*YA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r$_K$⁻, m$_K$⁺) *pho*A *sup*E44 λ⁻*thi*-1 *gyr*A96 *rel*A1	
MG1655 Δ*csrB* Δ*csrC*		MG1655 with unmarked *csrB/C* deletion	[12]
MG1655-pvBE3-*csrB*	MG_CsrB	pvBE3 and pET28a(+)_*csrB* transformed into MG1655	This study
MG1655-pvBE3-*csrC*	MG_CsrC	pvBE3 and pET28a(+)_*csrC* transformed into MG1655	This study
BL21-pvBE3	BL_p	pvBE3 transformed into BL21	This study
BL21-pvBE3-*csrB*	BL_CsrB	pvBE3 and pET28a(+)_*csrB* transformed into BL21	This study
BL21-pvBE3-*csrC*	BL_CsrC	pvBE3 and pET28a(+)_*csrC* transformed into BL21	This study
BL21-pvBE3-*cesT*	BL_CesT	pvBE3 and pNS6236 transformed into BL21	This study
DH5α-pvBE3-*csrB*	DH_CsrB	pvBE3 and pET28a(+)_*csrB* transformed into DH5α	This study
DH5α-pvBE3-*csrC*	DH_CsrC	pvBE3 and pET28a(+)_*csrC* transformed into DH5α	This study
DSM1116 ATCC9637			DSMZ

Table S3. Plasmids and primers

Plasmid Name	Description/Sequence	Notes [Reference from the main text]
pvBE3	pFU53 + *glgC* (upstream region including putative CsrA-binding sites; template from *E. coli* K-12 CC16) – Amp^r	[15], Volker Berndt (HZI group CBIO)
pET28a(+)_csrB	*csrB* gene cloned into the *NcoI-XhoI* sites of pET28a(+) – Kan^r	BioCat GmbH
pET28a(+)_csrC	*csrC* gene cloned into the *NcoI-XhoI* sites of pET28a(+) – Kan^r	BioCat GmbH
pNS6236	CesT expression plasmid under control of the P$_{tac}$ promoter – Kan^r	[13,14]
CsrA_forward CsrA_reverse	TAGGGGAATTGTGAGCGGAT AGCCGGATCTCAGTGGTGGT	
CsrB_forward CsrB_reverse	GATTCGGTGGGTCAGGAAGG GTTCGTTTCGCAGCATTCCA	
CsrC_forward CsrC_reverse	CAGGAGGCGAAGACAGAGGA ACGGGTCTTACAATCCTTGC	
CesT_ forward CesT_ reverse	CTGTCAGACCAAGTTTACTC ACTCTTCCTTTTTCAATATTATTGAAG	

IV. Chapter 4: Targeting IGF2BP2/IMP2 for cancer therapy

1. Publication Title:

First Small-Molecule Inhibitors Targeting the RNA-Binding Protein IGF2BP2/IMP2 for Cancer Therapy

Authors:

Charlotte Dahlem*, Ali Abuhaliema*, Sonja M. Kessler*, Tarek Kröhler, Ben G. E. Zoller, Shilpee Chanda,Yingwen Wu, Simon Both, Fabian Müller, Konstantin Lepikhov, Susanne H. Kirsch, Stephan Laggai, Rolf Müller, Martin Empting and Alexandra K. Kiemer

* These authors contributed equally.

Articles

First Small-Molecule Inhibitors Targeting the RNA-Binding Protein IGF2BP2/IMP2 for Cancer Therapy

Charlotte Dahlem, Ali Abuhaliema, Sonja M. Kessler, Tarek Kröhler, Ben G. E. Zoller, Shilpee Chanda, Yingwen Wu, Simon Both, Fabian Müller, Konstantin Lepikhov, Susanne H. Kirsch, Stephan Laggai, Rolf Müller, Martin Empting, and Alexandra K. Kiemer*

Cite This: *ACS Chem. Biol.* 2022, 17, 361−375

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: The RNA-binding protein IGF2BP2/IMP2/VICKZ2/p62 is overexpressed in several tumor entities, promotes tumorigenesis and tumor progression, and has been suggested to worsen the disease outcome. The aim of this study is to (I) validate IMP2 as a potential target for colorectal cancer, (II) set up a screening assay for small-molecule inhibitors of IMP2, and (III) test the biological activity of the obtained hit compounds. Analyses of colorectal and liver cancer gene expression data showed reduced survival in patients with a high IMP2 expression and in patients with a higher IMP2 expression in advanced tumors. *In vitro* target validation in 2D and 3D cell cultures demonstrated a reduction in cell viability, migration, and proliferation in IMP2 knockout cells. Also, xenotransplant tumor cell growth *in vivo* was significantly reduced in IMP2 knockouts. Different compound libraries were screened for IMP2 inhibitors using a fluorescence polarization assay, and the results were confirmed by the thermal shift assay and saturation-transfer difference NMR. Ten compounds, which belong to two classes, that is, benzamidobenzoic acid class and ureidothiophene class, were validated *in vitro* and showed a biological target specificity. The three most active compounds were also tested *in vivo* and exhibited reduced tumor xenograft growth in zebrafish embryos. In conclusion, our findings support that IMP2 represents a druggable target to reduce tumor cell proliferation.
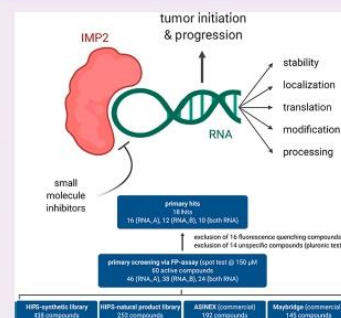
INTRODUCTION

RNA-binding proteins (RBPs) play an important role in diverse physiological functions. They are involved in mRNA maturation, stability, localization, and translation of mRNA targets.[1] Insulin-like growth factor 2 (*IGF2*) mRNA binding proteins (IGF2BPs/IMPs/VICKZs) are highly expressed during fetal development and maturation in different tissues. Their expression decreases in most tissues after birth.[2]

IMPs share structure and function similarities: they comprise two RNA recognition motifs (RRM1−2) in the N-terminal region and four KH homology domains (KH1−4) in their C-terminal region.[3] Several RNA recognition elements have been described to be recognized by IMP family members.[4−7]

IMP2 plays a distinct role in cancer progression and responsiveness to chemotherapy.[8,9] In different cancers, IMP2 has been shown to be more frequently amplified, and its expression is higher than those of IMP1 and IMP3.[8,10] While IMP1 and IMP2 are generally regarded as having a clear oncofetal expression pattern, data on IMP2 are somewhat conflicting.[2,3]

The expression and translation of different oncogenes are controlled directly or indirectly by IMP2. As a consequence, IMP2 increases cell proliferation, growth rate, migration, and invasion, promotes epithelial mesenchymal transition, and affects cell metabolism.[9,11−14] *IMP2* knockout mice show a reduction in size, total weight, and linear growth compared to wild types but are generally healthy.[15] Their metabolic activity and energy expenditure are improved, and the development of fatty liver disease and malignancy is reduced.[15] *Vice versa*, hepatocyte-specific overexpression of p62, a shortened splice variant of IMP2 containing its identical RNA binding domains, induces steatosis[16] and amplifies steatohepatitis and hepatocarcinogenesis.[2,9,17−19] In addition to hepatocellular carcinoma, IMP2 was suggested to play a role in another gastrointestinal tumor, which in fact represents the most frequent malignancy in the gastrointestinal system, that is, colon cancer.[20,21]

Due to the important role of IMP2 in carcinogenesis and tumor progression, we hypothesized that the inhibition of the activity of IMP2 might be a novel and attractive therapeutic approach for cancer therapy.

We therefore (I) undertook *in vitro* and *in vivo* approaches for target validation, (II) set up a fluorescence polarization (FP)
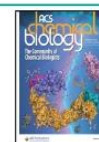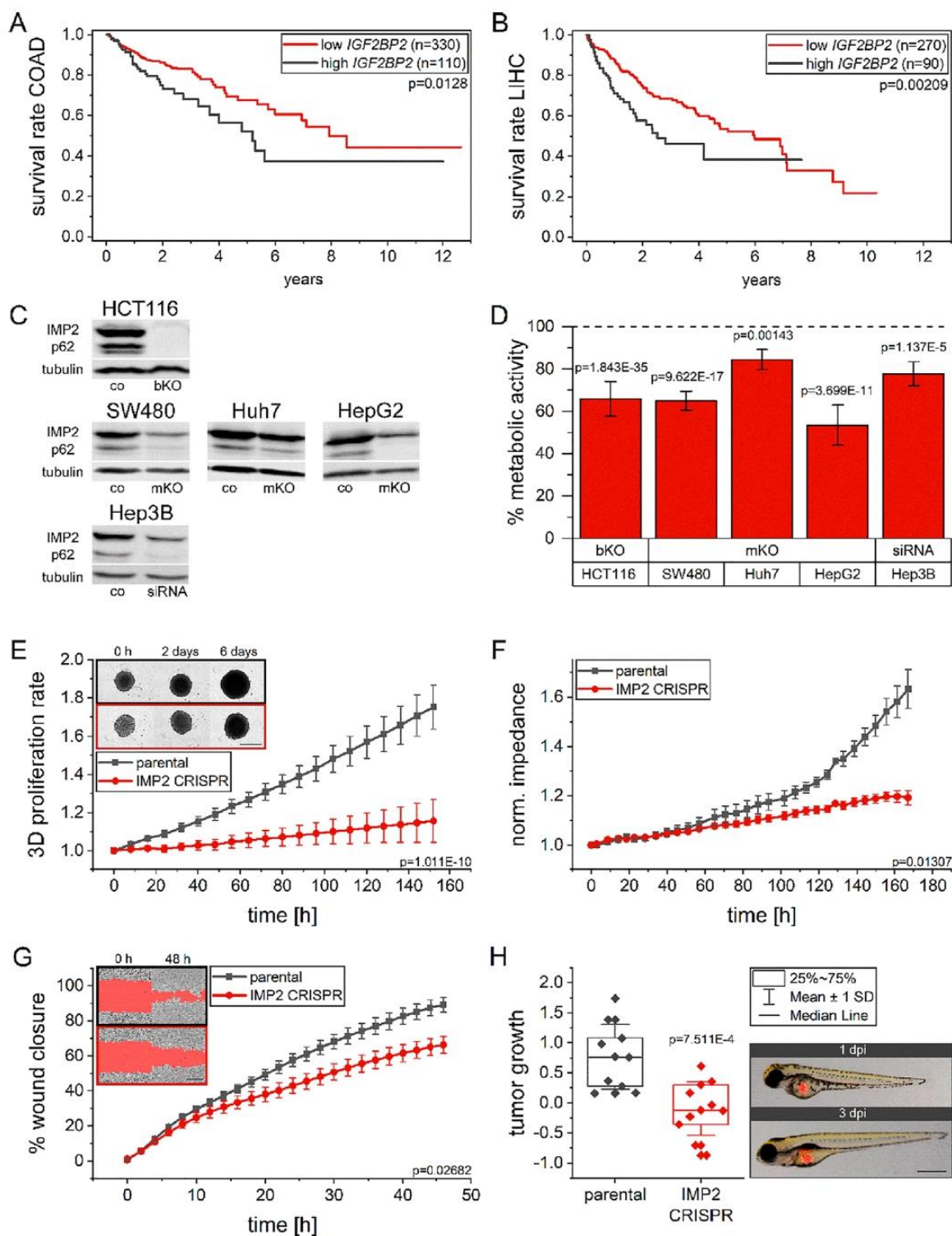
Figure 1. Clinical prognosis and target validation *in vitro* and *in vivo*. (A, B) Patient survival based on *IMP2/IGF2BP2* mRNA expression in the TCGA datasets COAD [(A), colon adenocarcinoma] and LIHC [(B), liver hepatocellular carcinoma]. The upper quartile is defined as a high expression. (C−H) IMP2 knockout/knockdown performed by CRISPR/Cas9 in HCT116, SW480, and Huh7 cells and by siRNA transfection in Hep3B cells. (C)

114

Figure 1. continued

Western blots for IMP2 and its splice variant p62 in complete, biallelic (bKO) knockout HCT116 cells, partial, monoallelic (mKO) knockout SW480, Huh7, and HepG2 cells, and siRNA-knockdown Hep3B cells. Parental cells or random siRNA-transfected cells were used as controls (co). The quantification is presented in Supporting Information Figure S3. (D) Metabolic activity of IMP2 knockout/knockdown cells measured by the MTT assay 96 h after seeding. Data were normalized to their respective controls (dashed line), $n = 4$ (triplicates). (E) After HCT116 spheroid formation for 3 days, the spheroid areas were monitored by an IncuCyte system. The area was normalized to 3 day old spheroids (0 h). Representative pictures show spheroids at the starting point (0 h), 2, and 6 days after initiation of measurements; scale bar = 500 μm; $n = 2$ (quadruplicates). (F) Impact of IMP2 knockout on HCT116 cell impedance; $n = 2$ (triplicates). (G) Migratory activity of IMP2 knockout HCT116 cells analyzed in an IncuCyte system. Representative pictures demonstrate the wound area in red at the starting point (0 h) and 48 h after wounding; scale bar = 400 μM; $n = 3$ (quadruplicates). (E−G) p-values were calculated for the respective last time points acquired. (H) Zebrafish embryo xenograft of HCT116 IMP2 knockout cells. Individual values of tumor growth quantification are presented in a box blot. Representative images of one parental xenotransplanted embryo at 1 dpi and 3 dpi are shown; scale bar = 1 mm.

screening assay to identify hits from several compound libraries, and (III) validated potential inhibitors *in vitro* and *in vivo*.

■ RESULTS AND DISCUSSION

Clinical Prognosis and Target Validation *in Vitro* and *in Vivo*. Overexpression of IMP2 in colorectal cancer *versus* healthy colon tissue has been reported in the past.[20,22] However, to the best of our knowledge, its potential implication in clinical prognosis has only been reported in a small dataset of 19 patients.[23] Thus, we analyzed the connection between *IGF2BP2* and prognosis in colon cancer in a large TCGA (The Cancer Genome Atlas) dataset. In colon adenocarcinoma patients, survival was significantly reduced in individuals with a high *IGF2BP2* expression (Figure 1A). Also, analyses of the hepatocellular carcinoma TCGA dataset confirmed previous findings of a poor prognosis of patients with a high *IGF2BP2* expression[9,13] (Figure 1B). Advanced colon tumor stages were associated with a significantly increased *IGF2BP2* expression (Supporting Information Figure S2). Due to the smaller availability of tumor classification data in the liver dataset, analyses on individual stages were not performed.

Comparing the gene expression levels of IMP2 with those of IMP1 and IMP3 at different developmental stages in murine and human tissues revealed a similar predominant fetal expression pattern (Supporting Information Figure S1).

Two different CRISPR/Cas9 approaches as well as an siRNA-facilitated knockdown were used to reduce the expression of IMP2 and its splice variant p62 in colorectal and hepatocellular carcinoma cell lines.[10,14] Sanger sequencing (Supporting Information Figure S3 A−C) and Western blots (Figure 1C, quantification in Supporting Information Figure S3D) confirmed a complete knockout in HCT116 cells upon CRISPR/Cas9, while the knockout was only partial in SW480, Huh7, and HepG2 cells. Clonal selection resulted in single-cell clones; their genetic editing is summarized in Supporting Information Figure S3C. Multiple trials of CRISPR/Cas9 knockout did not lead to a biallelic deletion in SW480, Huh7, and HepG2 cells, supporting that IMP2 is essential for cell proliferation. In Hep3B cells, IMP2 levels were knocked down by siRNA (Figure 1C). Metabolic activity in MTT assays, which typically correlates with cell numbers, was significantly reduced upon IMP2 knockout/knockdown in all tested cell lines (Figure 1D).

Since IMP2 has been suggested to act only partially on proliferation in 2D cultures[14] and CRISPR phenotypes in 3D *in vitro* cultures recapitulate more accurately those of *in vivo* tumors,[24] we assessed the proliferation of parental and IMP2 knockout HCT116 cells in 3D spheroids. Live-cell analysis revealed a strongly reduced proliferation of IMP2 knockout spheroids (Figure 1E). To exclude possible clonal artifacts, 3D

proliferation was assessed for multiple HCT116 IMP2 knockout clones harboring different gene edits obtained from different knockout approaches (Supporting Information Figure S3C). They all showed comparable proliferation, which was significantly lower than the proliferation of parental cells (Supporting Information Figure S4).

Target specificity of IMP2 knockout in HCT116 cells was further validated by rescue experiments, demonstrating a restored metabolic and proliferative activity of knockout cells, in which IMP2/p62 was overexpressed (Supporting Information Figure S5A−D). The knockout cells displayed a reduced expression of the tumor-promoting drivers *MYC* and the long noncoding RNA *DANCR* as described targets of IMP2[10,25,26] (Supporting Information Figure S5E,F). Their expression was partially restored when IMP2/p62 was overexpressed in knockout cells (Supporting Information Figure S5G,H). Since IMP2 facilitates its action on potentially thousands of mRNA targets *via* different actions, for example, regulating their stability, translation, or localization,[27,28] one would not expect that one single target is responsible for all IMP2 actions.

The major action of IMP2 on 3D growth has been suggested to be linked to its action on cell adhesion in hepatocellular carcinoma cells.[14] Electrical cell−substrate impedance sensing (ECIS) allows estimating cell adhesion in real time.[29] When this method was employed to assess cell responses toward CRISPR-facilitated IMP2 knockout, we found that the results were very similar to 3D growth (Figure 1F).

Furthermore, cell migration was significantly reduced in the absence of IMP2, as determined in a scratch wound assay (Figure 1G).

HCT116 parental and IMP2 knockout cells were then characterized for *in vivo* tumor growth in a zebrafish embryo xenograft model. Zebrafish embryo xenografts represent an excellent tool to perform *in vivo* experiments adhering to 3R rules to reduce animal experiments. Their usefulness for pharmacological studies has been extensively characterized, especially for studies on colorectal cancer.[30] In this model, tumor growth was significantly lower in IMP2 knockout cells compared to the parental control cells (Figure 1H).

IMP2 promotes tumorigenesis and tumor progression and is the most frequently amplified and the most highly expressed IMP in most cancer entities.[10] Our *in vitro* and *in vivo* data, together with published data on IMP2 in liver cancer,[9,13,17] validated IMP2 as an interesting target for the treatment of gastrointestinal tumors. *In vitro* and *in vivo* inhibition of proliferation by CRISPR-mediated knockout confirmed data from the literature using shRNA knockdown in colon cancer cells.[31,32] Xu *et al.* had reported similar findings in a pancreatic
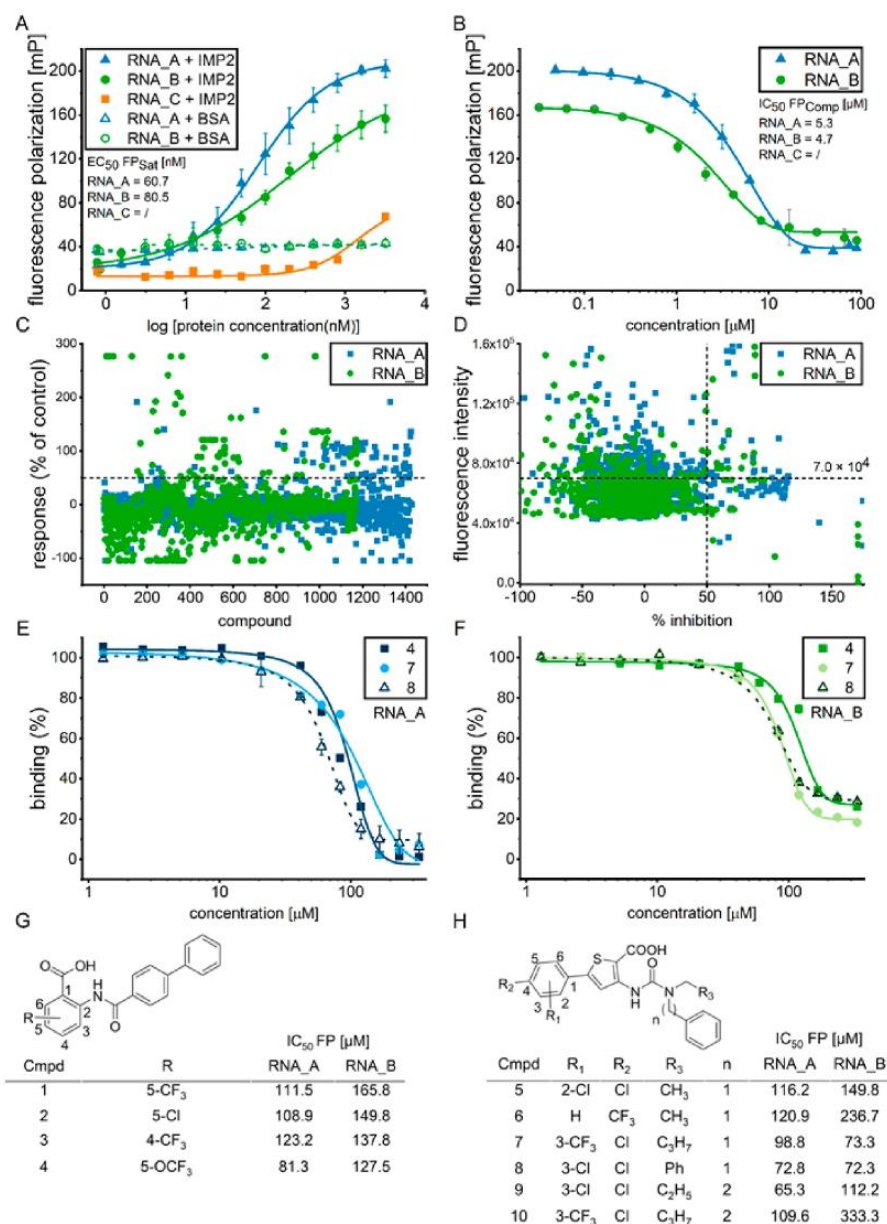
Figure 2. FP assay establishment and compound library screening. (A) FLC-labeled target RNA sequences (RNA_A/B) and the control sequence (RNA_C) were used in saturation experiments by titrating 1 nM RNA with a serial dilution of IMP2 protein or the unrelated protein BSA. (B) Competition experiments were conducted using fixed concentrations of 1 nM labeled RNA_A/B, 200 nM IMP2 and varying concentrations of the respective nonlabeled RNA as a competitor. Half-maximal effective concentrations in saturation assays [EC_{50} FP_{sat}, (A)] and half-maximal inhibitory concentrations in competition assays (IC_{50} FP_{comp}) were calculated using nonlinear regression analysis. Data are represented as means of FP values $\pm$ standard deviation (SD), $n = 2$ (duplicates). (C) 1428 compounds for RNA_A and 1175 compounds for RNA_B were screened in an FP assay at a final compound concentration of 150 μM. The scatter plot represents FP mean values of responses normalized to the response of nonlabeled RNAs used as the positive control. The dashed line indicates 50% inhibition of IMP2. (D) Mean fluorescence intensities of screened compounds were plotted against % inhibition. The dashed lines indicate the hit threshold. (E, F) Dose−response studies were performed with three representative hit compounds (compounds **4**, **7**, and **8**) against (E) RNA_A and (F) RNA_B in the FP-based competition assay. Competition assays were conducted using fixed concentrations of RNA_A/B (1 nM) and IMP2 (200 nM) and varying concentrations of hit compounds. Data are represented as means of FP values $\pm$ SD, $n = 2$ (duplicates). (G, H) Chemical structures and IC_{50} values of hit compounds from class A and class B. Descriptive compound data are summarized in Supporting Information Tables S1 and S2. IC_{50} values were calculated based on the competition assay; $n = 2$ (duplicates).

116

tumor xenograft model.[33] These encouraging results prompted us to screen for small-molecule inhibitors of IMP2.

Establishment of an FP Assay and Compound Screening. We established an FP screening assay to detect potential hits capable of inhibiting IMP2/RNA interactions using several published potential binding motifs.[27,34−36] The purity, size, and integrity of IMP2 protein were confirmed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot analysis (Supporting Information Figure S6A,B). Since the established FP assay is based on the use of small, labeled RNA sequences, we assured the absence of any RNase activity in the protein preparation under assay conditions (Supporting Information Figure S6C). Circular dichroism spectrometry (CDS) confirmed the pronounced presence of α-helical and β-sheet secondary structure elements indicating the correct folding of IMP2 (Supporting Information Figure S6D).

Serial dilutions of IMP2 were titrated against 1 nM of three different RNA sequences, that is, two sequences based on published binding motifs of IMP2 (RNA_A and RNA_B)[27,34−36] and one unspecific control sequence (RNA_C). IMP2 showed high affinity to RNA_A and RNA_B with EC_{50} values of 60.7 and 80.5 nM, respectively (Figure 2A). No affine binding interactions occurred with control RNA_C, and no binding was detected with bovine serum albumin (BSA), which was used as a negative control protein and titrated against RNA_A/B. Nonlabeled RNA_A and RNA_B oligos were used to test the displacement of labeled RNAs in a competitive FP assay. The IC_{50} values for RNA_A and RNA_B were 5.3 and 4.7 μM, respectively (Figure 2B).

For further experimental setup for library screening, dimethyl sulfoxide (DMSO) tolerance was tested using 1 nM RNA_A or RNA_B with IMP2 at DMSO concentrations between 0 and 10% v/v since all library compounds were dissolved in DMSO. The FP signal was stable in DMSO concentrations up to 10% (Supporting Information Figure S7A,B). A final concentration of 5% DMSO was selected for screening experiments. FP signals were measured at different time points: the FP signals were stable between 1.5 and 4 h of incubation at room temperature. Thus, the 1.5 h time point of incubation was considered appropriate for further screening experiments (Supporting Information Figure S7C,D). The Z' value was 0.9 for both RNAs and confirmed that the FP assay was robust and appropriate for further competitive screening (Supporting Information Figure S8A−C).

We used the established FP assay to screen compounds from four libraries. In total, 1428 compounds were screened for RNA_A and 1175 compounds for RNA_B (Figure 2C). Forty-six compounds achieved more than 50% inhibition of IMP2 with RNA_A, while 38 compounds showed activity against RNA_B (Figure 2D). Twenty-four hits obtained from screenings against RNA_A and 16 from RNA_B were excluded due to quenching of the fluorescence intensity (FI) or due to their autofluorescent nature (Figure 2D). The addition of Pluronic (concentration 0.013%) was used to identify false positives due to aggregation of compounds on IMP2. A total of 13 compounds were excluded by this means (Supporting Information Figure S8D). Finally, 16 compounds for RNA_A and 12 compounds for RNA_B were considered as confirmed hits. Ten of these compounds were able to inhibit binding of both RNAs to IMP2. These 10 most promising hits belonged either to the benzamidobenzoic acid class (class A) or the ureidothiophene class (class B) (Figure 2G,H).

The compounds' IC_{50} values were based on binding inhibition, measured by FP assay, and were found to be in the range between 65.3 and 120.9 μM for RNA_A and between 72.3 and 333.3 μM for RNA_B (Figure 2E−H).

Interaction of Inhibitors and IMP2. To validate protein binding by hit compounds, a thermal shift assay (TSA, also referred to as differential scanning fluorimetry) was used. The melting temperature (T_m) of IMP2 was 43 °C, and the change in T_m after the addition of hit compounds was calculated. All hit compounds showed shifts in T_m by −4.6 to −1.7 °C (Figure 3A,B). Tested against another RBP, that is, CsrA,[37] hit compounds **4**, **6**, and **9** demonstrated only marginal effects on CsrA/RNA interaction at very high concentrations, suggesting a
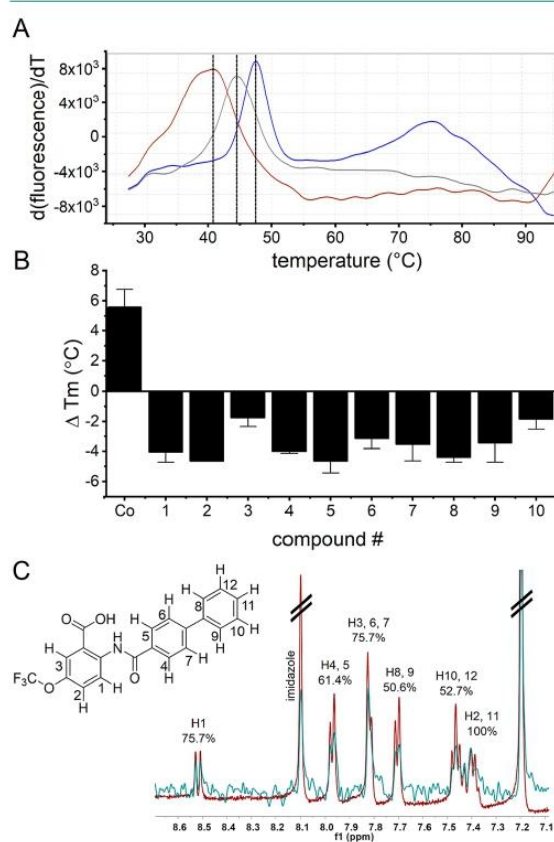
Figure 3. Hit compound confirmation *via* TSA and STD-NMR. (A) TSAs were performed at fixed concentrations of IMP2 (4.5 μM), nonlabeled RNA (100 μM), or hit compounds (100 μM), measuring the fluorescence of SYPRO orange. Representative melting curves demonstrate a shift in the IMP2 melting temperature (T_m) resulting from the binding to either RNA (blue) or hit compound **8** (red) compared to the control (gray). (B) Melting temperature shifts (ΔT_m) resulting from compound interactions were quantified and compared to the nonlabeled RNA control. Data are presented as means ± SD, $n = 2$ (one replicate). (C) STD-NMR was performed at fixed concentrations of IMP2 (5 μM) and compound **4** (500 μM). The reference spectrum without protein is shown in red, and the STD difference spectrum of the IMP2/compound **4** complex is shown in green. Overlaid STD off-resonance and STD effect spectra were normalized to the signal of H2 and H11.

117

specificity of the compounds (Supporting Information Figure S9).

IMP2 has many flexible loops rendering the crystallization of full-length IMP2 a huge challenge. As an alternative, we applied saturation-transfer difference (STD)-NMR to gain coarse-grained insight into the binding mode. This enabled not only to confirm ligand/protein interactions by a nonfluorescence-dependent (orthogonal) method but also to gather information on the binding orientation of a ligand.[38−40] The proximity of ligand protons to the bound protein influences the degree of saturation transfer: the nearest protons are most prominently saturated, while the more distant protons have the weakest signals in the STD spectra.[41] Typically, the proton with the strongest STD effect is used to normalize the signals of the other protons, resulting in values between 0 and 100%.

STD-NMR was successfully performed with compounds **1−6** as well as **13** and **14** (Figure 3C; Supporting Information Figure S10). Among this set of STD-NMR-investigated compounds were four compounds belonging to class A (**1−4**) and four compounds belonging to class B (**5, 6** and **13, 14**).

For the latter two compounds (**13, 14**), which inhibit only RNA_A, the overall STD effect was not very prominent and did not allow for a conclusive interpretation. Importantly, hit **4** was characterized by the strongest on-target effect in the FP assay. Hence, we used its STD-NMR results in order to inform subsequent docking experiments to gain insight into the mode of inhibition. A first observation was that protons H-1, H-2, and H-3 of the benzoic acid ring interact strongly with the protein (H-1 75.7%, H-2 100% observed together with H-11, H-3 75.7% observed together with H-6/7; Figure 3C). Second, within the middle ring, chemically equivalent protons H-6 and H-7 (75.7%; observed together with proton H-3) are in closer proximity to IMP2 than H-4 and H-5 (61.4%). Finally, the terminal phenyl moiety seems to interact less strongly overall (H-8/9 50.6%, H-10/12 52.7%, H-11 100% observed together with proton H-2).

With the aim to derive a plausible binding pose for class A compounds (representative **4**), we performed docking experiments with the available structure of the KH34 domain[7] as well as a homology model of the RRM1 domain considering the STD-NMR observations. In order to do this, we first generated the homology model of the IMP2 RRM1 domain in complex with the ACAC RNA binding motif[34] using the homologous IMP3 RRM12 structure (Supporting Information Figure S11A,B). Based on the assumption that our IMP2/RNA-interaction inhibitors act in an RNA-competitive manner, we docked compound **4** to the RNA-binding sites of both domains (RRM1 and KH34). The highest-ranked docking poses reflected some key observations from the STD-NMR experiment and is shown in Supporting Information Figure S11C−F. In both hypothetical ligand−target complexes, the benzoic acid head group interacts strongly with the protein. In the case of the RRM1 docking pose, the carbonyl function is involved in a salt bridge with the nearby Arg90 sidechain, with H-3 being the most solvent-exposed (less interacting) proton in this ring. Protons H-4 to H-7 showed a mixed solvent exposure profile, which is in agreement with the observed STD effect. A similar outcome can be observed for the docking pose of compound **4** to the KH34 domain. Here, the carboxylic function forms a salt bridge to Arg576 and Lys583. Furthermore, the mixed solvent-exposure profile is also evident for protons H-4 to H-7. In both the docking poses, the terminal phenyl ring is a part of the molecule, which is mostly exposed to the solvent. In conclusion, these docking poses suggest that for future optimization efforts,

enlargement of the identified hit scaffold (class A) should be possible at the terminal phenyl ring in order to improve efficacy. As the STD-NMR data for class B were not as conclusive as for class A, we did not put up a docking-based binding hypothesis. For that, more experimental information is needed to enable confidence-driven docking pose identification. We would like to stress that both the docking poses are hypothetical and that the current data do not allow to judge their validity beyond their agreement to the STD-NMR results. Ideally, X-ray crystallography will facilitate structure-guided optimization in the future.

Biological Activity of Hit Compounds. Different cancer cell lines were used to analyze the biological activity of the hit compounds. Colorectal (HCT116 and SW480) and hepatocellular (HepG2, Huh7, and Hep3B) carcinoma cells express a high amount of IMP2, whereas MCF7 cells do not express IMP2 and were therefore used as a control cell line (Cancer Cell Line Encyclopedia,[42] Western blot not shown). The lowest biological activity was noticed in MCF7 cells, supporting the specificity of the hits for IMP2 (Figure 4, Supporting Information Table S3).
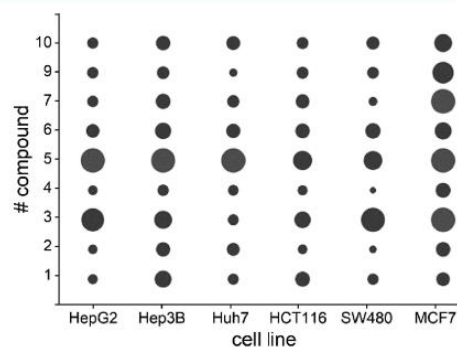
Figure 4. Biological activity (IC_{50} values) of final screening hits. The biological activity of screening hits was assessed *via* MTT assay in cancer cell lines expressing high IMP2 levels or lacking IMP2 (MCF7). Dots represent IC_{50} values (largest circle ≥ 80 μM, smallest circle 18.2 μM). $n = 2−5$ (triplicates). Specific values are listed in Supporting Information Table S3.

Compound **4** from class A and compound **9** from class B showed the highest potency on cells expressing IMP2, which was in concordance with the FP assay dose−response studies. Compound **3** from class A and compound **5** from class B showed the lowest potency.

Colorectal HCT116 and hepatocellular HepG2 cancer cells were selected to compare the potency of hit compounds on parental and IMP2 knockout cells. After 96 h treatment, parental cells demonstrated a significantly higher sensitivity toward hit compounds compared to CRISPR-modified cells (Figure 5A,B). Also, on measuring cell impedance, which proved to be a sensitive readout parameter for IMP2-facilitated actions, the hit compounds showed significantly stronger effects in the parental cells (Figure 5C,D; Supporting Information Figure S12). None of these compounds were optimized for target affinity, yet this selectivity is astonishing.

In addition to effects on cell metabolism and adhesion, compound treatment of HCT116 and SW480 cells resulted in a reduced expression of the IMP2 targets *DANCR*, *MYC*, and *HMGA1* in 2D and 3D cell cultures, as also observed for IMP2 knockout cells (Figure 6), further supporting the inhibitory action of the compounds on IMP2.
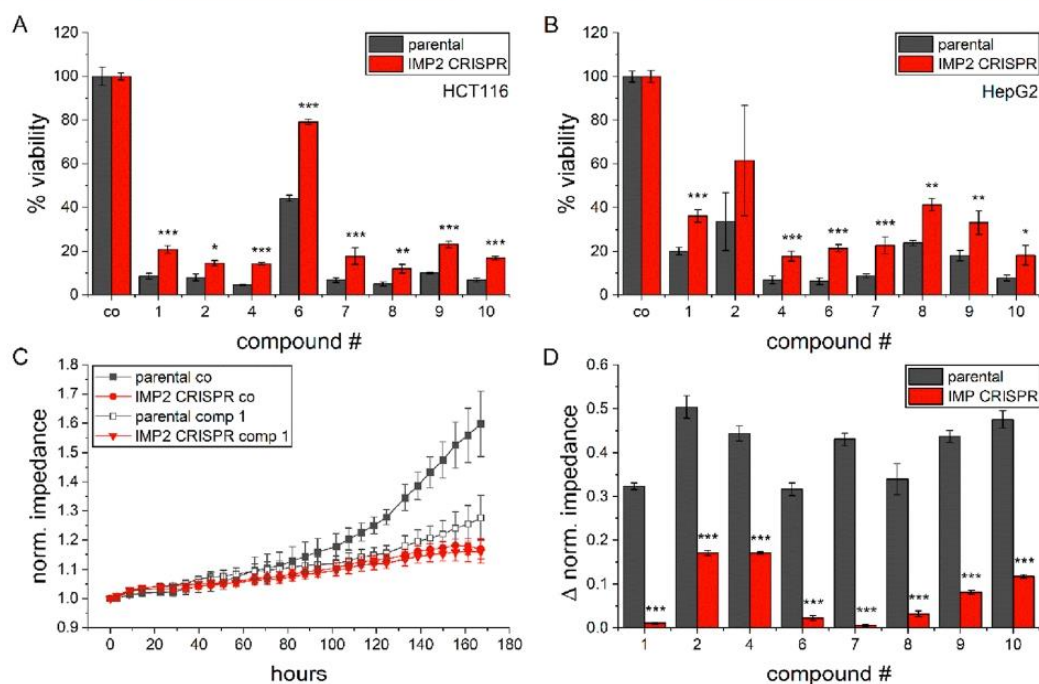
Figure 5. Target specificity of hit compounds. (A, B) Parental, biallelic HCT116, and monoallelic HepG2 IMP2 knockout cells were treated with 80 μM compounds or the respective DMSO solvent control. Metabolic activity was assessed 96 h after treatment *via* MTT assay and normalized to the respective controls; $n = 2–5$ (triplicates). (C) Cell impedance was assessed as a readout parameter for cell density and adhesion. One representative graph for compound **1** is shown. (D) Bars indicate impedance differences between compound and control treatment for parental and knockout cells, calculated for the last acquired timepoint; $n = 2$ (triplicates).
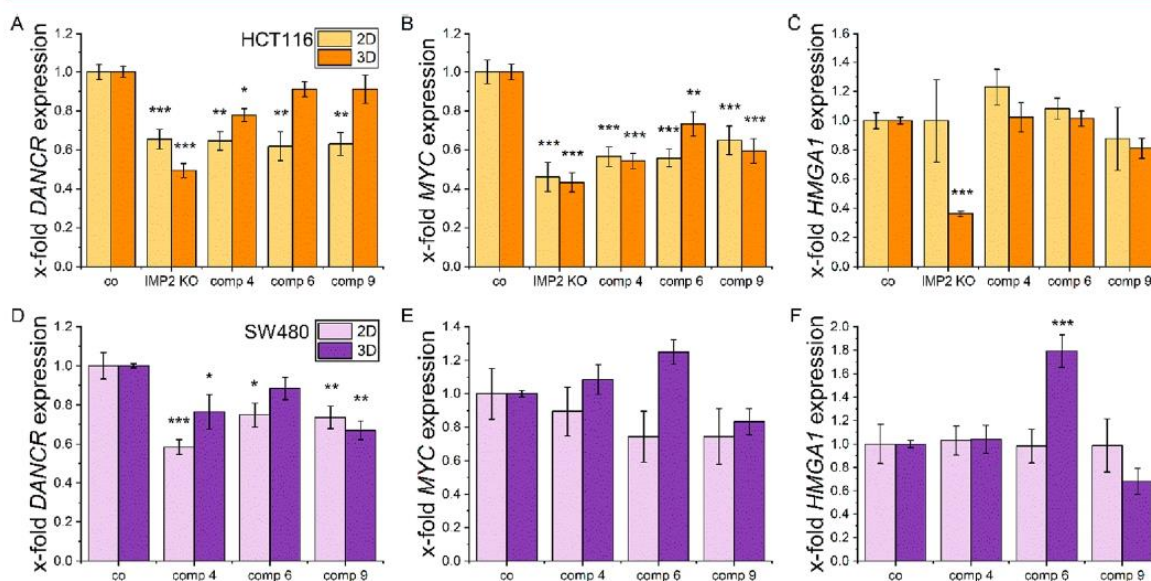
Figure 6. Expression of tumor-promoting *DANCR*, *MYC*, and *HMGA1* upon compound treatment. (A, D) *DANCR*, (B, E) *MYC*, and (C, F) *HMGA1* gene expression of HCT116 (A−C) and SW480 (D−F) cells after treatment with compound **4** (40 μM), **6** (50 μM), or **9** (50 μM) for 24 h, as determined by qPCR. Cells were cultured in 2D or 3D spheroid cultures. Values were normalized to the housekeeping gene *RNA18S*. Data are represented as means $\pm$ SEM; $n = 3$ (triplicates).

The antiproliferative properties of the IMP2 knockout were primarily observed in 3D models. Therefore, it is not surprising that the gene expression of the knockout cells also shows differences between 2D and 3D cultures, as seen for *HMGA1*.

119

The fact that these three target mRNAs of IMP2 are regulated also by other pathways might explain the cellular differences between HCT116 and SW480 cells, as well as the increase of *HMGA1* upon compound **6** treatment in SW480 cells, which might also represent a secondary effect.

Compound exposure resulted in no significant changes of *IMP2* gene expression in HCT116 and SW480 cells (Supporting Information Figure S13A). The same was observed on protein level, except for compound **4**, which reduced IMP2 in SW480 cells (Supporting Information Figure S13B).

The attenuated expression of the tumor-promoting IMP2 targets *DANCR*, *MYC*, and *HMGA1*, as well as the selective effects in IMP2 expressing *versus* knockout cells strongly support an on-target mode-of-action.

Compound treatment further inhibited tumor cell proliferation of HCT116, SW480, and Huh7 cells in 2D (Supporting Information Figure S14) and 3D (Supporting Information Figure S15) cultures. Only high compound concentrations induced cell death, as indicated by membrane permeability staining. Parallel measurements of apoptosis suggested that the observed cell death was induced *via* a caspase-3-dependent pathway (Supporting Information Figure S16).

Hit Compound Activity on Differentiated Huh7 Cells. Long-term cultivation of Huh7 cells in human serum (HS) has been described to induce cell differentiation, leading to a change in morphology and metabolic activity toward a more normal healthy state.[43] We therefore wanted to test whether the hit compounds affected the viability of these cells modeling normal cells. The cells' differentiation was confirmed according to their cobble-stone morphology typical for hepatocytes (Supporting Information Figure S17A) and an altered metabolism as shown by the increased albumin (*ALB*) expression (Supporting Information Figure S17B).[43]

In this model, compounds **6**, **7**, **8**, **9**, and **10** affected the viability of Huh7 but not of differentiated Huh7 cells. Compound **4** significantly reduced the viability of both differentiated and nondifferentiated Huh7 cells compared to the controls. Only compound **3** showed a higher effect on differentiated Huh7 cells (Figure 7). These results suggest a preferred compound activity against proliferating, undifferentiated cells.

The identified hit scaffolds were initially synthesized as antibacterial agents, and both compound classes were designed
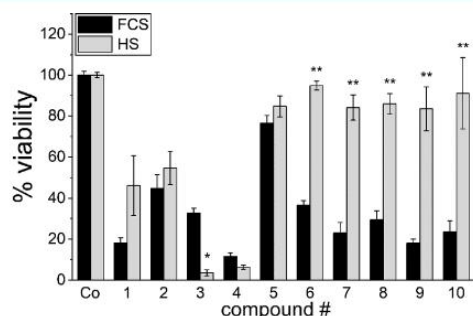
as inhibitors of bacterial RNA polymerase (RNAP).[44,45] Hence, they were envisioned as potential broad-spectrum antibiotics as RNAP is highly conserved among bacteria.[46] The amino acid sequences and the architecture of bacterial RNAP differ fundamentally from the eukaryotic RNAP, so that an undesired inhibition of eukaryotic RNAP is unlikely. No amino acid sequence similarities were found between either RNAP or IMP2. In fact, none of our compounds of class B impaired the viability of differentiated cells, and only one of class A congeners showed a higher effect in differentiated *versus* nondifferentiated cells.

***In Vivo* Action of Hit Compounds.** In order to assess whether the selected hit compounds affect tumor growth *in vivo*, compounds **4**, **6**, and **9** were administered in a zebrafish embryo xenograft model. In this model, compounds **4** and **6** significantly inhibited tumor growth (Figure 8A). Embryos showed no compound-induced toxicity after 3 days of treatment. Tumor-bearing and 50 μM compound-injected embryos showed normal development, but compound **4** caused a somewhat hunched body axis in 22.7% of the embryos at 3 dpi (Figure 8B).

CONCLUSIONS

In conclusion, we validated the RBP IMP2 as a druggable anti-cancer target using different *in vitro* and *in vivo* approaches. By establishing an FP-based screening platform to identify inhibitors of IMP2/RNA interactions, we found 10 hit compounds against IMP2.

Our hits from two different chemical classes serve as a suitable starting point for further optimization steps with the aim to generate more potent and specific compounds. Furthermore, we laid out a clear methodological path for the screening of additional libraries in order to identify structure-divergent compound series. Moreover, the described methodology will allow to include other RNA motifs, offering a valuable resource for the discovery of new compounds targeting RBPs.

The molecular mechanisms of action, that is, which pathways are specifically addressed by IMP2 inhibition/knockout, are the subject of ongoing studies. IMP2 has been suggested to bind to thousands of targets, which is why we expect that there are multiple layers of action.

METHODS

Materials. Kanamycin sulfate, cOmplete protease inhibitor cocktail tablets (#04693124001, Roche), diethylpyrocarbonate (DEPC), isopropyl-β-D-thiogalactoside (IPTG), Pluronic, lysozyme, BSA, and salts were purchased from Merck (Darmstadt, Germany). Water was treated with 0.1% v/v DEPC in all experiments handling with RNA.

Cell Lines. The IMP2 expressing cell lines SW480, HCT116, Hep3B, HepG2, and Huh7 and nonexpressing MCF7 cells (Cancer Cell Line Encyclopedia; Ghandi *et al.*, 2019) were used in cell-based experiments. SW480, HCT116, and MCF7 cells were maintained in Dulbecco's modified Eagle's medium (DMEM), and HepG2, Hep3B, and Huh7 cells were maintained in RPMI 1640 medium. Media were supplemented with 10% fetal calf serum (FCS), 1 mM glutamine, 100 U/mL penicillin, and 100 μg/mL streptomycin. Cells were cultured at 37 °C and 5% CO_2. Cell line authentication for hepatocellular cell lines was conducted by DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH) and for HCT116 by STR/DNA profiling. Mycoplasma testing was performed regularly *via* PCR.

Zebrafish Studies. AB wild-type zebrafish embryos were used for xenograft models. Zebrafish husbandry was conducted as described previously.[47] Zebrafish husbandry and all experiments were performed in accordance with the European Union Directive on the protection of animals used for scientific purpose (Directive 2010/63/EU) and the

Figure 7. Hit compound activity on differentiated Huh7 cells. Huh7 cells were either cultured under standard FCS conditions or differentiated in HS. Cells were treated for 96 h with 50 μM hit compounds or DMSO solvent control. Cell metabolic activity was determined by MTT; $n = 2$ (triplicates). *$p < 0.05$ and **$p < 0.01$ compared to values of standard FCS conditions.
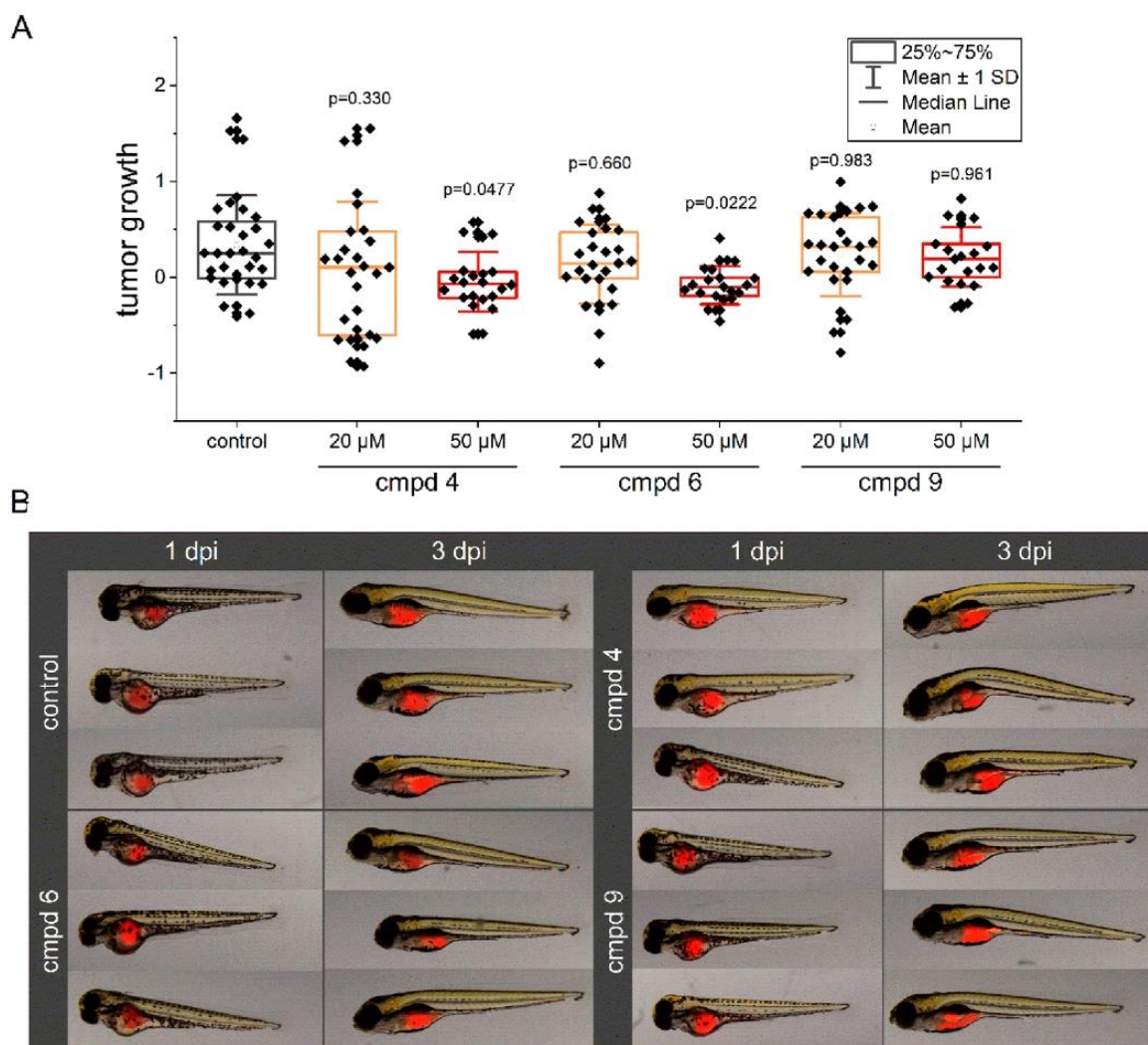
Figure 8. *In vivo* actions of compounds in a xenograft zebrafish embryo model. Fluorescence-labeled HCT116 cells suspended in compounds (cmpds) **4**, **6**, and **9** (20 and 50 μM) containing PBS were injected into the yolk sac 2 days post fertilization (dpf) zebrafish embryos. Embryos were imaged at 1 and 3 dpi, and the tumor growth was calculated based on the tumor area changes. (A) Individual values of tumor growth quantification are presented in a box blot. (B) Representative pictures of embryos are shown.

German Animal Welfare Act (§11 Abs. 1 TierSchG). Embryos were euthanized not later than 5 days post fertilization (dpf).

IMP2 siRNA Knockdown. A pool of four different HPLC-purified double-stranded RNA oligonucleotides was used for *IGF2BP2* knockdown in Hep3B cells (Flexitube Gene Solution, Qiagen, Hilden, Germany). Random siRNA was used as the control (Qiagen, Hilden, Germany). RNA oligo sequences are shown in Supporting Information Table S4. *IGF2BP2* knockdown was performed in 96-well plates by forward transfection (12,750 cells/well) employing INTERFERin Polyplus (Illkirch, France), as recommended by the manufacturer. Knockdown was confirmed by Western blot, and cell viability was measured using MTT assay 3 days after transfection.

IMP2 CRISPR-Mediated Knockout. The CRISPR/Cas9 technique was adapted to disrupt the gene of human *IGF2BP2* in two human colorectal (HCT116, SW480) and two hepatocellular carcinoma (Huh7, HepG2) cell lines by ribonucleoprotein delivery.

A validated single guide RNA (TrueGuide synthetic guide RNA, Thermo Fisher Scientific, Munich, Germany) targeting IMP2 (5′-

GATGGACTTTTGGCTCAATA-3′) and a recombinant Cas9 protein (TrueCutCas9 Protein v2, #A36496, Thermo Fisher Scientific, Munich, Germany) were delivered into the cells using Lipofectamine CRISPRMAX Cas9 transfection reagent (#CMAX00001, Thermo Fisher Scientific, Munich, Germany) according to the manufacturer's instructions. 80,000 cells were seeded in RPMI or DMEM media as described above, without antibiotics into a 12-well plate, incubated overnight, and transfected the next morning at a confluency of 30−70%. After 48 h of incubation time, the cells were detached, counted, and seeded into 96-well plates at a concentration of 0.8 cells/well for limiting dilution cloning. The remaining cells were used for gDNA extraction and verification of editing efficiency *via* the T7E1 mismatch assay.

Clones were cultured for downstream experiments until knockout of *IGF2BP2*, or at least reduced expression (monoallelic editing of the target region as assessed by Sanger sequencing) could be confirmed by Western blot. Clones that showed reduced IMP2 expression in Western blots underwent the whole procedure again until knockout was

Table 1. Sequences of the Components for the Assembly of pegRNAs[a]

pegRNA	abbreviation	spacer sequence	3′ extension	PBS length (nt)	RT template length (nt)
pegRNA1_GGtoT_11ntRT	pegRNA1_11	GACCACTCTTCCCGGGAGCA	GCGTGGATTGCTCCCGGGAAGAG	13	11
pegRNA1_GGtoT_15ntRT	pegRNA1_15	GACCACTCTTCCCGGGAGCA	AGGGGCGTGGATTGCTCCCGGGAAGAGT	13	15
pegRNA4_−GG_10ntRT	pegRNA4_10	AGAGCCATGGAGAAGCTAAG	TGATGCGCTTAGCTTCTCCATGG	13	10
pegRNA4_−GG_16ntRT	pegRNA4_16	AGAGCCATGGAGAAGCTAAG	TCAAACTGATGCGCTTAGCTTCTCCATGG	13	16
pegRNA6_+TA_10ntRT	pegRNA6_10	ATGCCCGCTTAGCTTCTCCA	AGCCTAATGGAGAAGCTAAGCGG	13	10
pegRNA6_+TA_16ntRT	pegRNA6_16	ATGCCCGCTTAGCTTCTCCA	TTTCAGAGCCTAATGGAGAAGCTAAGCGG	13	16

[a]These are composed of a spacer linked to the scaffold and 3′ extension. All sequences are displayed in 5′ to 3′ direction. PBS: primer binding site.

achieved. At least two rounds of CRISPR/Cas9 editing did not induce a biallelic knockout in SW480, Huh7, and HepG2 cells.

Prime Editing. *Design of Prime Editor System 2.* The prime editor 2 system was used to achieve *IGF2BP2* knockout in HCT116 cells. pCMV-PE2-P2A-GFP (Addgene plasmid #132776) and pU6-pegRNA-GG-acceptor (Addgene plasmid #132777) were a gift from David Liu.[48] Vectors were chosen to deliver the prime editor and pegRNA component of system 2. Golden Gate cloning was used to insert designed pegRNAs into the latter construct. Three different spacers targeting different loci of exon 6 served as a basis for the pegRNA assembly. Desired mutations were planned to disrupt the protospacer adjacent motif of the spacer sequences. As recommended, the length of the primer binding site was kept equal with 13 nt, but the size of the reverse transcriptase (RT) template varied between 10 and 16 nt (compare Table 1). The sequence 5′-AGAGCTAGAAATAG-CAAGTTAAAATAAGGCTAGTCCGTTATCAACTT-GAAAAAGTGGCACCGAGTCG-3′ was used as a scaffold for all pegRNAs.

Cell Line Development. 100,000 cells/well were seeded into 24-well plates overnight and transfected the next morning (after 16−24 h) at a confluency of approximately 60% with an equimolar ratio of the two vectors (2 μg total DNA content), employing 2 μL of lipofectamine 3000 (#L3000008, Thermo Fisher Scientific) according to the manufacturer's instructions. 48−72 h post transfection, the cells were washed with 1× PBS, detached with trypsin (#T3924, Merck), and resuspended to obtain a single cell suspension. Single GFP-positive cells were picked manually with a microneedle under a microscope and transferred into collagen-coated 60 mm dishes (collagen from rat tail, #C766160, Merck) into the squares of a grid (0.5 cm distance of the lines to each other) that had been drawn on the bottom of the dishes in advance. Periodically, colony formation of the single clones was surveyed until stable colonies were established that did not get in touch with each other and could be transferred into 96-well plates for further expansion and downstream analysis.

Kaplan−Meier Analyses. For survival analyses, TCGA patient survival data and *IGF2BP2* mRNA expression values were obtained from OncoLnc (http://www.oncolnc.org/). Patients were grouped into low and high (upper quartile) *IGF2BP2* expressers. Kaplan−Meier analyses were performed using OriginPro 19 b (OriginLab Corporation, Northhampton, MA, USA), and the log rank test was conducted to test equality over groups.

Western Blot. Western blots were performed as previously reported.[9] Antibodies used were specific for IMP2/p62[49] and α-tubulin (#T9026, Merk). IRDye680-conjugated anti-rabbit IgG (#926-68071, LI-COR Bioscience, Bad Homburg, Germany) and IRDye800-conjugated anti-mouse IgG (#926-32210, LI-COR Biosciences, Bad Homburg, Germany) were used as secondary antibodies. Signal intensities for IMP2 and its splice variant p62 were determined by using the Odyssey near-infrared imaging system from LI-COR Bioscience (Bad Homburg, Germany). Western blot signal intensities were quantified by Studio lite software (LI-COR Bioscience, Bad Homburg, Germany).

qPCR. Total RNA was isolated using the High Pure RNA Isolation Kit (#11828665001, Roche). For the analysis of 3D cultures, five spheroids were pooled. Concentration of isolated RNA was quantified

by NanoDrop (Thermo Fisher Scientific), and RNA with an A_{260}/A_{280} ratio higher than 1.8 was used for further experiments. RNA was transcribed using the High-Capacity cDNA Reverse Transcription Kit (#4368813, Thermo Fisher Scientific) in the presence of an RNase inhibitor (#10777-019, Invitrogen) according to the manufacturer's instructions. cDNA was analyzed using 5× HotFirePol EvaGreen qPCR Mix (#08-24-00020, Solis BioDyne, Tartu, Estonia) and the following primers: *RNA18S*_for 5′-AGGTCTGTGATGCCCTTAGA-3′, *RNA18S*_rev 5′-GAATGGGGTTCAACGGGTTA-3′, *DANCR*_for 5′-GCTCCAGGAGTTCGTCTCTTAC-3′, *DANCR*_rev 5′-TGCGCTAAGAACTGAGGCAG-3′, *MYC*_for 5′-AGCCACAGCA-TACATCCTGTCC-3′, *MYC*_rev 5′-CTCGTCGTTTCCGCAA-CAAGTC-3′, *HMGA1*_for 5′-CTAATTGGGACTCCGAGCCG-3′, *HMGA1*_rev 5′-GTAGCAAATGCGGATGCCTT-3′, *ALB*_for 5′-CACGCCTTTGGCACAATGAA-3′, *ALB*_rev 5′-ATCTCGAC-GAAACACACCCC-3′.

FP-Based Screening Assay. *Protein Purification.* Histidine-tagged IMP2 was expressed using a pET-28a(+) plasmid (Addgene plasmid #166997) in BL-21 *Escherichia coli* at 18 °C; the expression was induced with 0.4 mM IPTG. Bacteria were lysed, and the protein was isolated on a HisTrap HP nickel−sepharose column (#17524801, cytiva, Marlborough, MA, USA) followed by an imidazole buffer elution. Cells were harvested by centrifugation at 6733g and resuspended in binding buffer (20 mM Tris−HCl, pH 7.4, 2 mM MgCl$_2$, 150 mM NaCl, 10% v/v glycerol, 2 mM mercaptoethanol, 40 mM imidazole). Ethylenediamine tetraacetic acid -free protease inhibitor cocktail (cOmplete) was added fresh to the binding buffer. Cells were lysed using a French press homogenizer (two passages), and cell debris was removed by centrifugation at 42,858g at 4 °C for 1 h.

The supernatant was applied to a 5 mL HisTrap HP nickel−sepharose column at 0.75 mL/min on an ÄKTAxpress system. The column was equilibrated with binding buffer in a 10 × column volume or until the UV signal was stable. The column was washed later with 15 × column volume with binding buffer and with 15 × column volumes of high salt at 3 mL/min (Tris−HCl, pH 7.4, 2 mM MgCl$_2$, 1 M NaCl, 2.5% glycerol v/v, 2 mM mercaptoethanol, 40 mM imidazole). Subsequently, the column was washed with 15 × column volumes of binding buffer containing 100 mM imidazole.

Then, a gradient washing was done with a linear gradient to 100% of 500 mM imidazole buffer (Tris−HCl, pH 7.4, 2 mM MgCl$_2$, 150 mM NaCl, 2.5% glycerol v/v, 2 mM mercaptoethanol, 500 mM imidazole) in 10 × column volumes at 3 mL/min.

The protein was eluted in 0.7−1 M imidazole buffer (Tris−HCl, pH 7.4, 2 mM MgCl$_2$, 150 mM NaCl, 2.5% glycerol v/v, 2 mM mercaptoethanol, 700−1000 mM imidazole). Protein purity and identity were assessed by SDS-PAGE, Western blot, and CDS. The eluted IMP2 fractions were concentrated *via* centrifugal filtration using Vivaspin columns (30,000 MWCO, Sartorius). Buffer exchange into storage buffer (Tris−HCl, pH 7.4, 2 mM MgCl$_2$, 150 mM NaCl, 2 mM dithithreitol (DTT), 10% glycerol v/v, in DEPC-treated H$_2$O) was performed using multistep dilution inside Vivaspin columns. Protein concentrations were measured by both UV spectroscopy ($\varepsilon = 280$ nm) and Pierce assay (according to the manufacturer's instructions). Protein aliquots were stored at −80 °C.

The absence of potential RNase contamination in the protein preparations was assessed by mixing 15 μL of human RNA (470 ng/μL) isolated from MCF7 cells with either 15 μL of storage buffer or 15 μL of IMP2 protein (22.1 μM) and incubated on ice or at room temperature for 1.5 h, and then the samples were subjected to agarose gel electrophoresis.

Probe Design. Based on published IMP2 target sequences, two different RNA oligonucleotides were designed as IMP2 binding partners. The sequence of RNA_A was based on the published 4−5 bp motifs CAUC,[27] ACAC,[34] CCCC,[35] and ACACA[36] and contained a 3′ nucleotide extension after the fluorescein (FLC) label: FLC-AUG<u>CAUCCCC</u>GCAGCU<u>ACACACACA</u>CAACA. RNA_B was designed based on the binding motif UUCACGUUCAC and contained a 7-nucleotide extension in front of the tandem repeat sequence CCCCC<u>CUUUCACGUUCAC</u>UCUGUCU-FLC originally described in Nielsen *et al.*, 1999. A third RNA_C sequence (FLC-GAAAAAAAGAUUUAUUUAUUUAAGA) was reported to bind to AU rich element binding proteins and was used to detect the specificity of the fluorescent probe binding to the target.[35] FLC-labeled or nonlabeled HPLC-purified single-stranded RNA oligomers were purchased from Merck.

FP Assay. Lyophilized RNA oligomers were dissolved in the FP assay buffer (Tris−HCl, pH 7.4, 2 mM MgCl$_2$, 150 mM NaCl, 10% glycerol, 2 mM DTT, 10% glycerol v/v, in DEPC treated H$_2$O) to have 100 μM stock solutions of RNAs, further diluted into 100 nM aliquots, and stored at −80 °C. Saturation experiments were performed to detect direct binding of different RNA oligomers to either IMP2 or BSA. Thereby, a constant concentration (1 nM final) of each FLC-labeled RNA with a constant concentration (1 nM) was titrated with serial dilutions in the range of 0.15 nM to 3 μM final concentration in the FP assay buffer. In the competitive FP experiments, IMP2 was used in excess (2−3 folds above the EC$_{50}$ values). Based on the saturation experiments, a final IMP2 concentration of 200 nM was selected to be used in subsequent competition assays with RNA_A and RNA_B.

All competitive experiments included 1 h incubation of IMP2 with nonlabeled RNA oligo or compounds in 384-well black microplates before the addition of the labeled RNAs. Competition experiments were done at constant concentrations of RNA_A and RNA_B (1 nM) and a fixed IMP2 concentration (200 nM) by titration against serial dilutions of unlabeled RNAs (0.32 nM to 100 μM final). DMSO tolerance was evaluated by incubating different DMSO concentrations (0−10% v/v) and RNA_A or RNA_B at 1 nM, either with or without IMP2. The stability of FP values was assessed over time by measuring the FP at different time points (every 30 min until 4 h).

FP and FI were measured using a CLARIOstar Plus microplate reader (BMG LABTECH, Ortenberg, Germany) with an excitation at 485−495 nm and an emission at 520−530 nm. In general, the focal height and gain adjustments were done before starting each measurement to reach the maximum sensitivity.[50] The FI values of any compound deviating more than ±20% from the values of the controls were excluded from further procedure.[37] Each sample was tested in duplicate, and the FP values are reported in millipolarization units (mP).

The assay robustness was verified by assaying low controls (LCs) in 192 samples with RNA_A or RNA_B (1 nM final concentration) in 5% DMSO in the FP assay buffer and a high control (LC) plate containing 192 samples in the same plate in addition to 100 nM IMP2. Z′ value was calculated according to the formula Z′ = 1 − [(3SD × HC) + (3SD × LC)]/|(HC mean − LC mean).[51]

Screening Library. Compounds from four different libraries were screened: 838 compounds were a synthetic in-house library from the Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), 253 compounds were from the in-house natural products library at HIPS, 192 compounds were from a commercial library provided by ASINEX (Winston-Salem, NC, USA; https://www.asinex.com/), and 145 compounds were from the commercial Maybridge library (small molecular weight chemical fragments, Thermo Fischer Scientific): https://www.maybridge.com/portal/alias__Rainbow/lang__en/tabID__177/DesktopDefault.aspx.

Compounds were dissolved in DMSO to a stock concentration of 5 or 10 mM depending on their solubility.

FP-Based Screening. Compounds were diluted in the FP assay buffer to concentrations of 450 μM. A total of 10 μL in duplicates (15% DMSO) from each compound was added into 384-well microtiter plates using an electronic Eppendorf Xplorer 12-channel pipette (Eppendorf, Hamburg, Germany). In addition, 10 μL of 600 nM IMP2 in FP assay buffer was added into the same plate, and the samples were incubated for 1 h by shaking at room temperature. Afterward, 10 μL of a 3 nM RNA_A/B solution in the FP assay buffer was added to the mixture and incubated for an additional 1.5 h at room temperature in the dark. Accordingly, the final assay concentrations were 1 nM RNA_A/B, 200 nM IMP2, and 150 μM compounds in 5% DMSO. Any compound that enhanced or quenched the total FI more than 20% of the FI of the controls was excluded. The percentage of binding inhibition was calculated as % inhibition = (mean of HC − read compound value)/(mean of HCs − mean of LC) × 100%. In addition, Pluronic was added to the FP buffer to a final concentration of 0.0137% in the competitive assay to rule out any false-positive results from aggregation.

FP-Based Dose−Response Measurement. Two-fold serial dilutions of a 333 μM starting concentration of hit compounds were prepared in FP buffer in 5% DMSO and titrated in the presence or absence of IMP2. The experiments were performed twice independently using duplicates.

Competition binding assays using CsrA from *Yersinia pseudotuberculosis* were performed as previously described.[37]

TSA. The shift of the melting point of IMP2 in the presence or absence of hit compounds was recorded in 96-well plates using an Applied Biosystems StepOnePlus Real-Time PCR System from Thermo Fisher Scientific. Thermal shift experiments were performed at fixed concentrations of IMP2 (4.5 μM) and 100 μM hit compound. SYPRO orange (Merck) was used in 2.5-fold concentration (stock 5000-fold) in a total volume of 20 μL. The heating gradient started at 25 °C, and the temperature was increased by 0.5 °C/min until 95 °C, detecting the fluorescence of SYPRO orange. Melting curve plots of fluorescence *versus* temperature were converted into melting peaks, and melting temperatures were calculated subsequently by Protein Thermal Shift Software v1.3-Thermo Fisher Scientific.

STD-NMR. ¹H-STD-NMR experiments were conducted using a Bruker Fourier spectrometer (500 MHz) (MA, USA), and the probe temperature was kept at 283 K.

The final compound concentration was 500 μM based on the solubility limit in 10% DMSO d_6. A volume of 25 μL from each compound solution was diluted with 25 μL of Tris buffer pH 7.4, 150 mM NaCl in D$_2$O. IMP2 protein/ligand was used in a 1:100 ratio with a final concentration of 5 μM. A control spectrum was recorded under the same conditions without a protein to test for artifacts.

The STD-NMR experiments were carried out with a carrier set at −1 ppm for the on-resonance and −40 ppm for the off-resonance irradiation. Selective IMP2 protein presaturation was carried out at 0.5 s by using a train of 50 ms Gauss-shaped pulses.

The STD effect resulting from the difference in signal intensity after saturation transfer was quantified using the formula STD$_{effect}$ = (I_0 − I_{sat})/I_0. This provides insights into the relative proximity of the respective protons to the protein surface. I_0 represents the intensity of one signal in the off-resonance or reference NMR spectrum, and I_{sat} represents the intensity of a signal in the on-resonance NMR spectrum.

The STD-NMR spectrum of the IMP2/ligand sample was subtracted from the respective STD spectrum of the ligand alone using the same NMR conditions to eliminate any artefacts arising from the ligand. Protons of the residual imidazole contamination from the elution step appear at 7.2 and 8.1 ppm.

In Silico **Studies.** The homology modeling and molecular docking experiments were performed with MOE 2019.01 (Molecular Operating Environment). A homology model for IMP2 RRM1 was generated using a reported X-ray structure of IMP3 in complex with RNA (IMP3 RRM12, pdb ID: 6GX6).[35] The sequence of IMP2 RRM1 as well as 6GX6 pdb atom coordinates were loaded into MOE, and the homology model was generated using the built-in homology model function with standard parameters, AMBER10:EHT force field, and the RNA atoms

123

of 6GX6 as the environment. In the case of docking to the KH34 domain, we used the reported NMR structure (PDB ID: 6ROL).[7]

Before docking, a QuickPrep step was applied to the respective protein structure, including the RNA (ACAC) ligand in the case of the RRM1 homology model. Docking was done for the promising hit compound 4, where conclusive STD-NMR data were available in order to evaluate the plausibility of the docking poses. For the RRM1 homology model, the docking site was defined by the selection of protein residues in 4.5 Å proximity to RNA atoms, while for the KH34 domain, reported residues interacting with RNA were employed to define the docking site.[7] The built-in docking feature of MOE was used with standard parameters, induced-fit refinement, and AMBE10:EHT force field. The highest ranked docking pose was further refined by applying a QuickPrep step (standard parameters, AMBER10:EHT force field), and the resulting ligand−protein complex was analyzed using the ligand interaction tool of MOE.

Testing Hits for Inhibition of Cell Viability. *MTT.* Cells were seeded into 96-well plates at a density of 2500−10,000 cells/well based on the cell type (ability to reach 95−100% confluency in the control treatment at the time of measurement). Cell metabolic activity was measured using MTT 96 h after treatment of hit compounds (1−80 μM). For each compound, the inhibition of cell activity was calculated for each concentration normalized to its respective DMSO control or nontreated control (if DMSO in the respective concentration showed viability above 90%). The dot plot was generated with Python 3.8.1 software.

Kinetic Apoptosis/Necrosis Measurements. For the time-dependent analysis of cell death, the cells were analyzed in an IncuCyte S3 System (Sartorius). The day after seeding, supernatants were replaced by the respective media containing the IncuCyte Cytotox Red (#4632, Sartorius) and Caspase-3/7 Green (#4440, Sartorius) reagents according to the manufacturer's instructions. Cells were treated with different concentrations of the respective compounds or DMSO vehicle control, and cell confluency as well as apoptotic and necrotic events were monitored for 3 days. Fluorescent signals from apoptotic (caspase 3/7 active) and necrotic (permeable membrane) cells were normalized to cell confluency and the time point of treatment (0 h).

ECIS Assay. ECIS represents a powerful tool to assess cell proliferation in real time, but cell impedance is also responsive toward changes in cell adhesion,[29] which has been shown to be affected by IMP2.[14] HCT116 parental and IMP2 CRISPR knockout cells were seeded (7000 cells/well) into 96 (96W10E+)-well plates coated with rat tail collagen (#C766160, Merck, 30 μg/mL in 0.2% acetic acid). The cells were seeded directly after the compounds were added into the plate to reduce cell stress. Cells were treated with 25 μM of hit compounds in triplicate. Effects of hit compounds on proliferation were evaluated relative to (0.25%) DMSO controls. Cell impedance was assessed in an ECIS Zθ (theta) instrument (Applied BioPhysics Inc., NY, USA). Measurements were started immediately after cell seeding and were taken every 450−900 s for each well.

IncuCyte 2D and 3D Proliferation and Migration. *Proliferation.* For the 2D kinetic proliferation analysis, 5000 cells were seeded per well into 96-well plates. The next day, cells were treated with the respective compounds, and cell confluency was monitored in an IncuCyte S3 system. Cell confluency was analyzed using IncuCyte basic analyzer software, and confluency was normalized to the starting point.

For the 3D proliferation analysis, 3000 cells were seeded per well into low-attachment U-bottom 96-well plates. After spheroid formation for 3 days, spheroids were treated with the respective compounds, and monitoring in an IncuCyte S3 system was started, if not indicated otherwise. The spheroid area was analyzed using spheroid IncuCyte software, and the area was normalized to the first measuring time point after treatment (0 h).

Migration. Cell migration was measured using an IncuCyte S3 system. 100,000 cells per well were seeded into ImageLock 96-well plates. The next day, scratches were performed by the Woundmaker tool (IncuCyte Migration Kit). Cells were washed twice with 2% fetal bovine serum (FBS)-containing media, which was also used for further cultivation. Cells were treated with the respective compounds, and the

migration was monitored for 48 h. The cell-covered wound area was analyzed and quantified using IncuCyte migration software.

Differentiation of Huh7 Cells. Huh7 cells were seeded into 96-well plates (3000 cells/well) in full RPMI growth medium containing 10% FBS. After 24 h, the medium was aspirated and changed to RPMI medium containing 2% HS in order to induce cell differentiation as described.[43,52] Fresh medium was added twice a week for 21 days. Differentiated cells were treated with hit compounds, and the cell viability was assessed *via* MTT assay 96 h after treatment.

Zebrafish Xenograft Model and *in Vivo* Proliferation Measurement. Parental and *IMP2* knockout HCT116 cells were used for proliferation studies in a zebrafish embryo xenograft model. 2 × 10^6 cells were suspended in 1 μL of 0.1% BSA/PBS. At 2 days post fertilization (dpf), 2 nL of cell suspension was injected into the yolk sac by a FemtoJet microinjector (Eppendorf). Before injection, the tumor cells were stained with the cell tracker orange dye (#C34551, Thermo Fisher Scientific) according to the manufacturer's protocol. For the analysis of compound-induced effects on tumor growth, the tumor cells were suspended in 0.1% BSA/PBS containing the compounds in the indicated concentrations just before injection. Single embryos were placed into 96-well plates and imaged the next day with a Leica M205 FCA fluorescence stereomicroscope (1 day post injection, dpi). Tumor growth was determined at 3 dpi by analysis of the fluorescent tumor area and quantification by ImageJ. The growth rate was calculated as follows: (tumor area 3 dpi − tumor area 1 dpi)/tumor area 1 dpi. The effects of compound injection on the zebrafish embryo development and viability were assessed by microscopic observation of the eye, heart, and body axis formation, heartbeat, and pigmentation.

Statistical Analysis. Data analysis was done by Microsoft Excel, and statistics were performed using OriginPro. IC$_{50}$ and EC$_{50}$ values were calculated using nonlinear regression analysis with Origin pro version 19 software. Data are represented as means ± SEM if not indicated otherwise. Shapiro−Wilk test was performed to analyze the data distribution. Depending on whether the data were normally distributed and on the group size, statistical differences were calculated using one-way ANOVA, Student's *t*-test, Mann−Whitney *U* test, or Kolmogorov−Smirnov test. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acschembio.1c00833.

> Additional data related to the clinical prognosis of the oncofetal protein IMP2; IMP2-knockout clones and target validation; additional data related to the FP assay screening development and validation; additional data related to hit compound activity and specificity; differentiation process of Huh7 cells in HS; descriptive data of compounds; IC$_{50}$ values of hit compounds; and sequences of siRNA oligonucleotides used in IMP2 knockdown (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Alexandra K. Kiemer − *Department of Pharmacy, Pharmaceutical Biology, Saarland University, Saarbrücken 66123, Germany;* ⓞ orcid.org/0000-0002-7224-9900; Email: pharm.bio.kiemer@mx.uni-saarland.de; Fax: +49 681 302 57302

Authors

Charlotte Dahlem − *Department of Pharmacy, Pharmaceutical Biology, Saarland University, Saarbrücken 66123, Germany;* ⓞ orcid.org/0000-0001-9983-0802

Ali Abuhaliema − *Department of Pharmacy, Pharmaceutical Biology, Saarland University, Saarbrücken 66123, Germany*

Sonja M. Kessler − *Department of Pharmacy, Pharmaceutical Biology, Saarland University, Saarbrücken 66123, Germany; Institute of Pharmacy, Experimental Pharmacology for Natural Sciences, Martin Luther University Halle-Wittenberg, Halle 06108, Germany*

Tarek Kröhler − *Department of Pharmacy, Pharmaceutical Biology, Saarland University, Saarbrücken 66123, Germany*

Ben G. E. Zoller − *Department of Drug Design and Optimization, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Saarland University, Saarbrücken 66123, Germany*

Shilpee Chanda − *Department of Pharmacy, Pharmaceutical Biology, Saarland University, Saarbrücken 66123, Germany*

Yingwen Wu − *Department of Drug Design and Optimization, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Saarland University, Saarbrücken 66123, Germany*

Simon Both − *Department of Pharmacy, Pharmaceutical Biology, Saarland University, Saarbrücken 66123, Germany*

Fabian Müller − *Center for Bioinformatics, Saarland University, Saarbrücken 66123, Germany*

Konstantin Lepikhov − *Department of Genetics, Saarland University, Saarbrücken 66123, Germany*

Susanne H. Kirsch − *Department of Microbial Natural Products, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research, Saarland University, Saarbrücken 66123, Germany*

Stephan Laggai − *Department of Pharmacy, Pharmaceutical Biology, Saarland University, Saarbrücken 66123, Germany*

Rolf Müller − *Department of Microbial Natural Products, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research and Department of Pharmacy, Saarland University, Saarbrücken 66123, Germany*

Martin Empting − *Department of Drug Design and Optimization, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI) and Department of Pharmacy, Saarland University, Saarbrücken 66123, Germany*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acschembio.1c00833

Author Contributions

C.D., A.A., and S.M.K. have provided equal contribution. Initiation and direction of the study: A.K.K.; conceptualization: S.M.K., M.E., and A.K.K.; formal analysis: C.D. and A.A.; funding acquisition: A.K.K.; investigation: C.D., A.A., T.K., B.G.E.Z., Y.W., S.B., and F.M.; resources: K.L., S.H.K., and R.M.; supervision: S.M.K., S.L., M.E., and A.K.K.; visualization: C.D., A.A., and F.M.; writing—original draft: S.M.K., C.D., A.A., M.E., and A.K.K.; writing—review & editing: T.K., B.G.E.Z., S.H.K. S.L., and R.M.

Notes

The authors declare no competing financial interest.

■ **REFERENCES**

(1) Mohibi, S.; Chen, X.; Zhang, J. Cancer the "RBP"eutics-RNA-binding proteins as therapeutic targets for cancer. *Pharmacol. Ther.* **2019**, *203*, 107390.

(2) Czepukojc, B.; Abuhaliema, A.; Barghash, A.; Tierling, S.; Naß, N.; Simon, Y.; Körbel, C.; Cadenas, C.; van Hul, N.; Sachinidis, A.; et al. IGF2 mRNA Binding Protein 2 Transgenic Mice Are More Prone to Develop a Ductular Reaction and to Progress toward Cirrhosis. *Front. Med.* **2019**, *6*, 179.

(3) Bell, J. L.; Wächter, K.; Mühleck, B.; Pazaitis, N.; Köhn, M.; Lederer, M.; Hüttelmaier, S. Insulin-like Growth Factor 2 mRNA-Binding Proteins (IGF2BPs): Post-Transcriptional Drivers of Cancer Progression? *Cell. Mol. Life Sci.* **2013**, *70*, 2657−2675.

(4) Chao, J. A.; Patskovsky, Y.; Patel, V.; Levy, M.; Almo, S. C.; Singer, R. H. ZBP1 recognition of β-actin zipcode induces RNA looping. *Genes Dev.* **2010**, *24*, 148−158.

(5) Patel, V. L.; Mitra, S.; Harris, R.; Buxbaum, A. R.; Lionnet, T.; Brenowitz, M.; Girvin, M.; Levy, M.; Almo, S. C.; Singer, R. H.; et al. Spatial Arrangement of an RNA Zipcode Identifies mRNAs under Post-Transcriptional Control. *Genes Dev.* **2012**, *26*, 43−53.

(6) Schneider, T.; Hung, L.-H.; Aziz, M.; Wilmen, A.; Thaum, S.; Wagner, J.; Janowski, R.; Müller, S.; Schreiner, S.; Friedhoff, P.; et al. Combinatorial Recognition of Clustered RNA Elements by the Multidomain RNA-Binding Protein IMP3. *Nat. Commun.* **2019**, *10*, 2266.

(7) Biswas, J.; Patel, V. L.; Bhaskar, V.; Chao, J. A.; Singer, R. H.; Eliscovich, C. The Structural Basis for RNA Selectivity by the IMP Family of RNA-Binding Proteins. *Nat. Commun.* **2019**, *10*, 4440.

(8) Kessler, S. M.; Lederer, E.; Laggai, S.; Golob-Schwarzl, N.; Hosseini, K.; Petzold, J.; Schweiger, C.; Reihs, R.; Keil, M.; Hoffmann, J.; et al. IMP2/IGF2BP2 Expression, but Not IMP1 and IMP3, Predicts Poor Outcome in Patients and High Tumor Growth Rate in Xenograft Models of Gallbladder Cancer. *Oncotarget* **2017**, *8*, 89736−89745.

(9) Kessler, S. M.; Pokorny, J.; Zimmer, V.; Laggai, S.; Lammert, F.; Bohle, R. M.; Kiemer, A. K. IGF2 mRNA Binding Protein p62/IMP2-2 in Hepatocellular Carcinoma: Antiapoptotic Action Is Independent of IGF2/PI3K Signaling. *Am. J. Physiol.: Gastrointest. Liver Physiol.* **2013**, *304*, G328−G336.

(10) Dai, N.; Ji, F.; Wright, J.; Minichiello, L.; Sadreyev, R.; Avruch, J. IGF2 mRNA Binding Protein-2 Is a Tumor Promoter That Drives Cancer Proliferation through Its Client mRNAs IGF2 and HMGA1. *Elife* **2017**, *6*, No. e27155.

(11) Huang, S.; Wu, Z.; Cheng, Y.; Wei, W.; Hao, L. Insulin-like Growth Factor 2 mRNA Binding Protein 2 Promotes Aerobic Glycolysis and Cell Proliferation in Pancreatic Ductal Adenocarcinoma via Stabilizing GLUT1 mRNA. *Acta Biochim. Biophys. Sin.* **2019**, *51*, 743−752.

(12) Janiszewska, M.; Suvà, M. L.; Riggi, N.; Houtkooper, R. H.; Auwerx, J.; Clément-Schatlo, V.; Radovanovic, I.; Rheinbay, E.; Provero, P.; Stamenkovic, I. Imp2 controls oxidative phosphorylation and is crucial for preserving glioblastoma cancer stem cells. *Genes Dev.* **2012**, *26*, 1926−1944.

(13) Kessler, S. M.; Laggai, S.; Barghash, A.; Schultheiss, C. S.; Lederer, E.; Artl, M.; Helms, V.; Haybaeck, J.; Kiemer, A. K. IMP2/p62 Induces Genomic Instability and an Aggressive Hepatocellular Carcinoma Phenotype. *Cell Death Dis.* **2015**, *6*, No. e1894.

(14) Xing, M.; Li, P.; Wang, X.; Li, J.; Shi, J.; Qin, J.; Zhang, X.; Ma, Y.; Francia, G.; Zhang, J.-Y. Overexpression of p62/IMP2 can Promote Cell Migration in Hepatocellular Carcinoma via Activation of the Wnt/β-Catenin Pathway. *Cancers* **2019**, *12*, 7−16.

(15) Dai, N.; Zhao, L.; Wrighting, D.; Krämer, D.; Majithia, A.; Wang, Y.; Cracan, V.; Borges-Rivera, D.; Mootha, V. K.; Nahrendorf, M.; et al. IGF2BP2/IMP2-Deficient Mice Resist Obesity through Enhanced

Translation of *Ucp1* mRNA and Other mRNAs Encoding Mitochondrial Proteins. *Cell Metab.* **2015**, *21*, 609−621.

(16) Tybl, E.; Shi, F.-D.; Kessler, S. M.; Tierling, S.; Walter, J.; Bohle, R. M.; Wieland, S.; Zhang, J.; Tan, E. M.; Kiemer, A. K. Overexpression of the IGF2-mRNA binding protein p62 in transgenic mice induces a steatotic phenotype. *J. Hepatol.* **2011**, *54*, 994−1001.

(17) Simon, Y.; Kessler, S. M.; Bohle, R. M.; Haybaeck, J.; Kiemer, A. K. The Insulin-like Growth Factor 2 (*IGF2*) mRNA-Binding Protein p62/IGF2BP2-2 as a Promoter of NAFLD and HCC? *Gut* **2014**, *63*, 861−863.

(18) Laggai, S.; Kessler, S. M.; Boettcher, S.; Lebrun, V.; Gemperlein, K.; Lederer, E.; Leclercq, I. A.; Mueller, R.; Hartmann, R. W.; Haybaeck, J.; Kiemer, A. K. The *IGF2* mRNA Binding Protein p62/IGF2BP2-2 Induces Fatty Acid Elongation as a Critical Feature of Steatosis. *J. Lipid Res.* **2014**, *55*, 1087−1097.

(19) Kessler, S.; Simon, Y.; Gemperlein, K.; Gianmoena, K.; Cadenas, C.; Zimmer, V.; Pokorny, J.; Barghash, A.; Helms, V.; Van Rooijen, N.; et al. Fatty Acid Elongation in Non-Alcoholic Steatohepatitis and Hepatocellular Carcinoma. *Int. J. Mol. Sci.* **2014**, *15*, 5762−5773.

(20) Ye, S.; Song, W.; Xu, X.; Zhao, X.; Yang, L. IGF2BP2 promotes colorectal cancer cell proliferation and survival through interfering with RAF-1 degradation by miR-195. *FEBS Lett.* **2016**, *590*, 1641−1650.

(21) Li, T.; Hu, P.-S.; Zuo, Z.; Lin, J.-F.; Li, X.; Wu, Q.-N.; Chen, Z.-H.; Zeng, Z.-L.; Wang, F.; Zheng, J.; et al. METTL3 Facilitates Tumor Progression via an M6A-IGF2BP2-Dependent Mechanism in Colorectal Carcinoma. *Mol. Cancer* **2019**, *18*, 112.

(22) Liu, W.; Li, Z.; Xu, W.; Wang, Q.; Yang, S. Humoral Autoimmune Response to *IGF2* mRNA-Binding Protein (IMP2/p62) and Its Tissue-Specific Expression in Colon Cancer. *Scand. J. Immunol.* **2013**, *77*, 255−260.

(23) Bigagli, E.; De Filippo, C.; Castagnini, C.; Toti, S.; Acquadro, F.; Giudici, F.; Fazi, M.; Dolara, P.; Messerini, L.; Tonelli, F.; et al. DNA Copy Number Alterations, Gene Expression Changes and Disease-Free Survival in Patients with Colorectal Cancer: A 10 Year Follow-Up. *Cell. Oncol.* **2016**, *39*, 545−558.

(24) Han, K.; Pierce, S. E.; Li, A.; Spees, K.; Anderson, G. R.; Seoane, J. A.; Lo, Y.-H.; Dubreuil, M.; Olivas, M.; Kamber, R. A.; Wainberg, M.; et al. CRISPR Screens in Cancer Spheroids Identify 3D Growth-Specific Vulnerabilities. *Nature* **2020**, *580*, 136−141.

(25) Hu, X.; Peng, W.-X.; Zhou, H.; Jiang, J.; Zhou, X.; Huang, D.; Mo, Y.-Y.; Yang, L. IGF2BP2 Regulates DANCR by Serving as an N6-Methyladenosine Reader. *Cell Death Differ.* **2020**, *27*, 1782−1794.

(26) Wang, Y.; Lu, J.-H.; Wu, Q.-N.; Jin, Y.; Wang, D.-S.; Chen, Y.-X.; Liu, J.; Luo, X.-J.; Meng, Q.; Pu, H.-Y.; et al. LncRNA LINRIS Stabilizes IGF2BP2 and Promotes the Aerobic Glycolysis in Colorectal Cancer. *Mol. Cancer* **2019**, *18*, 174.

(27) Hafner, M.; Landthaler, M.; Burger, L.; Khorshid, M.; Hausser, J.; Berninger, P.; Rothballer, A.; Ascano, M.; Jungkamp, A.-C.; Munschauer, M.; et al. Transcriptome-Wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell* **2010**, *141*, 129−141.

(28) Cao, J.; Mu, Q.; Huang, H. The Roles of Insulin-like Growth Factor 2 mRNA-Binding Protein 2 in Cancer and Cancer Stem Cells. *Stem Cells Int* **2018**, *2018*, 4217259.

(29) Stolwijk, J. A.; Wegener, J. Impedance-Based Assays along the Life Span of Adherent Mammalian Cells in Vitro: From Initial Adhesion to Cell Death. *Bioanal. Rev.* **2019**, *2*, 1−75.

(30) Fior, R.; Póvoa, V.; Mendes, R. V.; Carvalho, T.; Gomes, A.; Figueiredo, N.; Ferreira, M. G. Single-Cell Functional and Chemosensitive Profiling of Combinatorial Colorectal Therapy in Zebrafish Xenografts. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E8234−E8243.

(31) Ye, Y.; Saburi, W.; Odaka, R.; Kato, K.; Sakurai, N.; Komoda, K.; Nishimoto, M.; Kitaoka, M.; Mori, H.; Yao, M. Structural Insights into the Difference in Substrate Recognition of Two Mannoside Phosphorylases from Two GH130 Subfamilies. *FEBS Lett.* **2016**, *590*, 828−837.

(32) Lu, S.; Han, L.; Hu, X.; Sun, T.; Xu, D.; Li, Y.; Chen, Q.; Yao, W.; He, M.; Wang, Z.; Wu, H.; Wei, M. N6-Methyladenosine Reader IMP2 Stabilizes the ZFAS1/OLA1 Axis and Activates the Warburg Effect:

Implication in Colorectal Cancer. *Journal of Hematology & Oncology* **2021**, *14* (1), 123.

(33) Xu, X.; Yu, Y.; Zong, K.; Lv, P.; Gu, Y. Up-Regulation of IGF2BP2 by Multiple Mechanisms in Pancreatic Cancer Promotes Cancer Proliferation by Activating the PI3K/Akt Signaling Pathway. *J. Exp. Clin. Cancer Res.* **2019**, *38*, 497.

(34) Conway, A. E.; Van Nostrand, E. L.; Pratt, G. A.; Aigner, S.; Wilbert, M. L.; Sundararaman, B.; Freese, P.; Lambert, N. J.; Sathe, S.; Liang, T. Y.; et al. Enhanced CLIP Uncovers IMP Protein-RNA Targets in Human Pluripotent Stem Cells Important for Cell Adhesion and Survival. *Cell Rep.* **2016**, *15*, 666−679.

(35) Jia, M.; Gut, H.; Chao, J. A. Structural Basis of IMP3 RRM12 Recognition of RNA. *RNA* **2018**, *24*, 1659−1666.

(36) Van Nostrand, E. L.; Freese, P.; Pratt, G. A.; Wang, X.; Wei, X.; Xiao, R.; Blue, S. M.; Chen, J.-Y.; Cody, N. A. L.; Dominguez, D.; et al. A Large-Scale Binding and Functional Map of Human RNA-Binding Proteins. *Nature* **2020**, *583*, 711−719.

(37) Maurer, C. K.; Fruth, M.; Empting, M.; Avrutina, O.; Hoßmann, J.; Nadmid, S.; Gorges, J.; Herrmann, J.; Kazmaier, U.; Dersch, P.; et al. Discovery of the first small-molecule CsrA-RNA interaction inhibitors using biophysical screening technologies. *Future Med. Chem.* **2016**, *8*, 931.

(38) Viegas, A.; Manso, J.; Nobrega, F. L.; Cabrita, E. J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Educ.* **2011**, *88*, 990−994.

(39) Becker, W.; Bhattiprolu, K. C.; Gubensäk, N.; Zangger, K. Investigating Protein-Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy. *ChemPhysChem* **2018**, *19*, 895−906.

(40) Kirsch, P.; Jakob, V.; Oberhausen, K.; Stein, S. C.; Cucarro, I.; Schulz, T. F.; Empting, M. Fragment-Based Discovery of a Qualified Hit Targeting the Latency-Associated Nuclear Antigen of the Oncogenic Kaposi's Sarcoma-Associated Herpesvirus/Human Herpesvirus 8. *J. Med. Chem.* **2019**, *62*, 3924−3939.

(41) Mayer, M.; Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 1784−1788.

(42) Ghandi, M.; Huang, F. W.; Jané-Valbuena, J.; Kryukov, G. V.; Lo, C. C.; McDonald, E. R.; Barretina, J.; Gelfand, E. T.; Bielski, C. M.; Li, H.; et al. Next-Generation Characterization of the Cancer Cell Line Encyclopedia. *Nature* **2019**, *569*, 503.

(43) El-Shamy, A.; Eng, F. J.; Doyle, E. H.; Klepper, A. L.; Sun, X.; Sangiovanni, A.; Iavarone, M.; Colombo, M.; Schwartz, R. E.; Hoshida, Y.; et al. A Cell Culture System for Distinguishing Hepatitis C Viruses with and without Liver Cancer-Related Mutations in the Viral Core Gene. *J. Hepatol.* **2015**, *63*, 1323−1333.

(44) Hinsberger, S.; Hüsecken, K.; Groh, M.; Negri, M.; Haupenthal, J.; Hartmann, R. W. Discovery of Novel Bacterial RNA Polymerase Inhibitors: Pharmacophore-Based Virtual Screening and Hit Optimization. *J. Med. Chem.* **2013**, *56*, 8332−8338.

(45) Sahner, J. H.; Groh, M.; Negri, M.; Haupenthal, J.; Hartmann, R. W. Novel small molecule inhibitors targeting the "switch region" of bacterial RNAP: Structure-based optimization of a virtual screening hit. *Eur. J. Med. Chem.* **2013**, *65*, 223−231.

(46) Murakami, K. Structural Biology of Bacterial RNA Polymerase. *Biomolecules* **2015**, *5*, 848−864.

(47) Dahlem, C.; Siow, W. X.; Lopatniuk, M.; Tse, W. K. F.; Kessler, S. M.; Kirsch, S. H.; Hoppstädter, J.; Vollmar, A. M.; Müller, R.; Luzhetskyy, A.; et al. Thioholgamide A, a New Anti-Proliferative Anti-Tumor Agent, Modulates Macrophage Polarization and Metabolism. *Cancers* **2020**, *12*, 1288.

(48) Anzalone, A. V.; Randolph, P. B.; Davis, J. R.; Sousa, A. A.; Koblan, L. W.; Levy, J. M.; Chen, P. J.; Wilson, C.; Newby, G. A.; Raguram, A.; et al. Search-and-Replace Genome Editing without Double-Strand Breaks or Donor DNA. *Nature* **2019**, *576*, 149−157.

(49) Lu, M.; Nakamura, R. M.; Dent, E. D.; Zhang, J.-Y.; Nielsen, F. C.; Christiansen, J.; Chan, E. K. L.; Tan, E. M. Aberrant Expression of

126

Fetal RNA-Binding Protein p62 in Liver Cancer and Liver Cirrhosis. *Am. J. Pathol.* **2001**, *159*, 945−953.

(50) Rossi, A. M.; Taylor, C. W. Analysis of Protein-Ligand Interactions by Fluorescence Polarization. *Nat. Protoc.* **2011**, *6*, 365−387.

(51) Zhang, J.-H.; Chung, T. D. Y.; Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screening* **1999**, *4*, 67−73.

(52) Steenbergen, R.; Oti, M.; ter Horst, R.; Tat, W.; Neufeldt, C.; Belovodskiy, A.; Chua, T. T.; Cho, W. J.; Joyce, M.; Dutilh, B. E.; Tyrrell, D. L. Establishing Normal Metabolism and Differentiation in Hepatocellular Carcinoma Cells by Culturing in Adult Human Serum. *Sci. Rep.* **2018**, *8*, 11685.

Supporting Information

The first small molecule inhibitors targeting the RNA binding protein IGF2BP2/IMP2 for cancer therapy

Charlotte Dahlem[a,#], Ali Abuhaliema[a,#], Sonja M. Kessler[a,b,#], Tarek Kröhler[a], Ben G. E. Zoller[c], Shilpee Chanda[a], Yingwen Wu[c], Simon Both[a], Fabian Müller[d], Konstantin Lepikhov[e], Susanne H. Kirsch[f], Stephan Laggai[a], Rolf Müller[f,g], Martin Empting[c,g], Alexandra K. Kiemer[a,*]

[a] Department of Pharmacy, Pharmaceutical Biology, Saarland University, Saarbrücken, 66123, Germany

[b] Institute of Pharmacy, Experimental Pharmacology for Natural Sciences, Martin Luther University Halle-Wittenberg, Halle, 06108, Germany

[c] Department of Drug Design and Optimization, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Saarland University, Saarbrücken, 66123, Germany

[d] Center for Bioinformatics, Saarland University, Saarbrücken, 66123, Germany

[e] Department of Genetics, Saarland University, 66123 Saarbrücken, 66123, Germany

[f] Department of Microbial Natural Products, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research, Saarland University Campus, 66123 Saarbrücken, Germany

[g] Department of Pharmacy, Saarland University, Saarbrücken, 66123, Germany

[#] equal contribution

[*] corresponding author

Alexandra K. Kiemer
Saarland University
Department of Pharmacy
Pharmaceutical Biology
Campus C2 3
D-66123 Saarbrücken, Germany
Fax: +49 681 3021 57302
pharm.bio.kiemer@mx.uni-saarland.de

Supplementary Figures
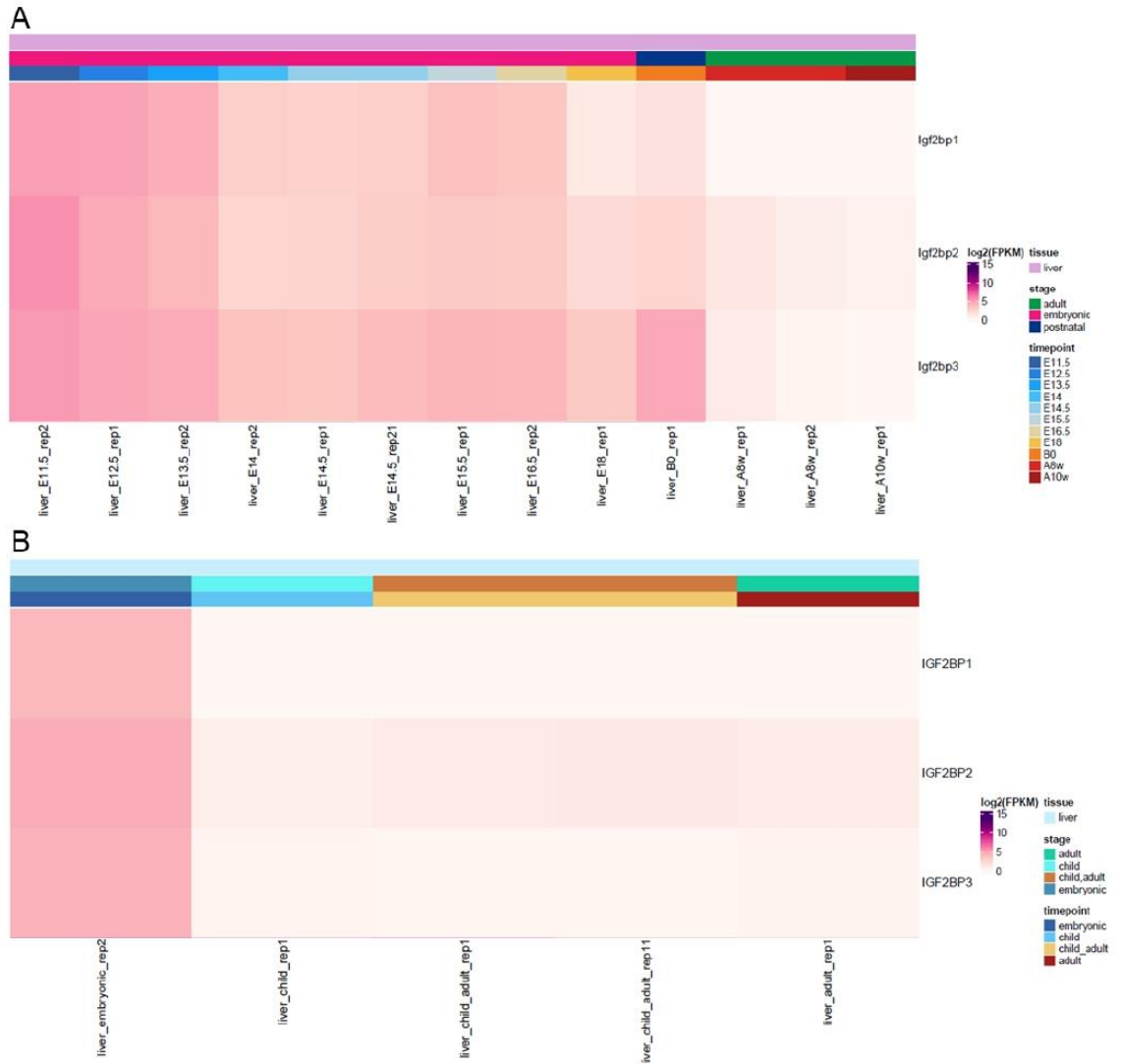
Supplementary Figure S1. Comparison of *IGF2BP2* expression with *IGF2BP1* and *IGF2BP3* in murine and human embryonic and adult livers

Gene expression levels (log2 FPKM) for IGF2BPs 1-3 in (A) mouse liver and (B) human liver in different developmental stages. Data originated from a mouse developmental atlas and the ENCODE project.
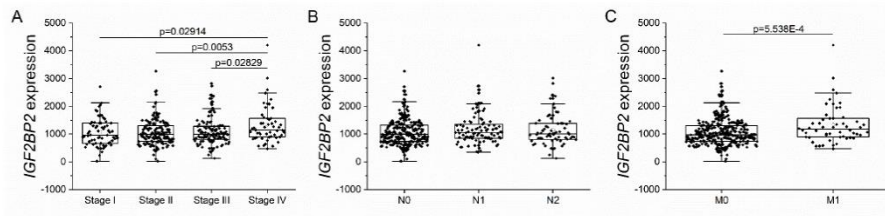
Supplementary Figure S2. *IGF2BP2* **expression of COAD patients**

Analyses of the *IGF2BP2* expression in the COAD dataset tumors characterized by (A) AJCC neoplasm disease stage and (B, C) pathology T/N/M stage.

"KO #1" HCT116 clone 47-1	Biallelic A insertion
"KO #3" HCT116 clone 47-2	Allele 1: A insertion Allele 2: 10 bp deletion
"KO #4" HCT116 clone 47-6	Allele 1: 10 bp deletion Allele 2: G insertion
"KO #2" HCT116 clone PE 12-1	Biallelic 2 bp deletion
Huh7 clone	Monoallelic A insertion
HepG2 clone	Monoallelic mutation: CGG instead of ATA
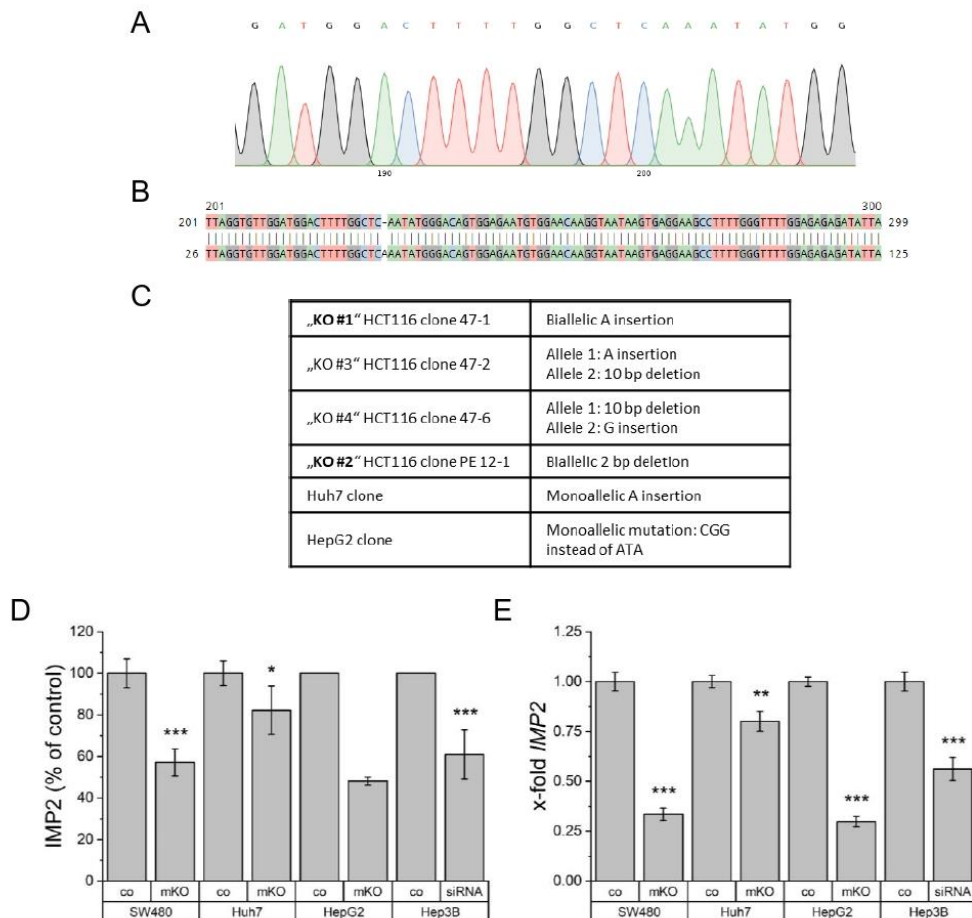
Supplementary Figure S3. Sanger sequencing of IMP2 CRISPR clones

(A) Representative Sanger sequencing result of the IMP2 CRISPR HCT116 clone 47-1 that was used for target validation and compound testing (e.g., Figure 1 and Figure 5).
(B) Sequence alignment of IMP2 CRISPR HCT116 clone 47-1, demonstrating a bi-allelic A insertion.
(C) Table summarizing the editing of clones used in this study. HCT116 clone "KO #2" was generated using prime editing.
(D) Quantification of IMP2 protein levels in partial IMP2 knockdown cells compared to parental cells. Data are represented as means ± SEM, n=2-6.
(E) Quantification of *IMP2* gene levels in partial IMP2 knockdown cells compared to parental cells. Data are represented as means ± SEM, n=3.

130
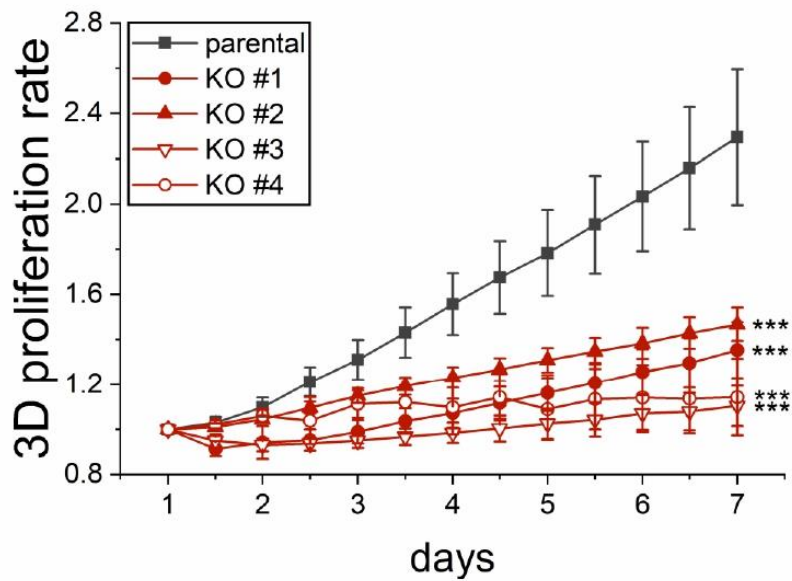
Supplementary Figure S4. 3D growth comparison of HCT116 knockout clones

The spheroid growth of different HCT116 IMP2 knockout clones showing different gene edits (see Supplementary Figure S2) was monitored by automated live-cell microscopy, starting after spheroid formation for 24 h. Spheroid area was analyzed using the IncuCyte® S3 system and was normalized to 1-day old spheroids. Data are represented as means ± SEM, n=3 (quadruplicates). Statistical analysis was performed for the last acquired time point (7 days). Asterisks represent values for the comparisons between the growth of parental and respective knockout cells. p values comparing the growth of different clones were > 0.05.
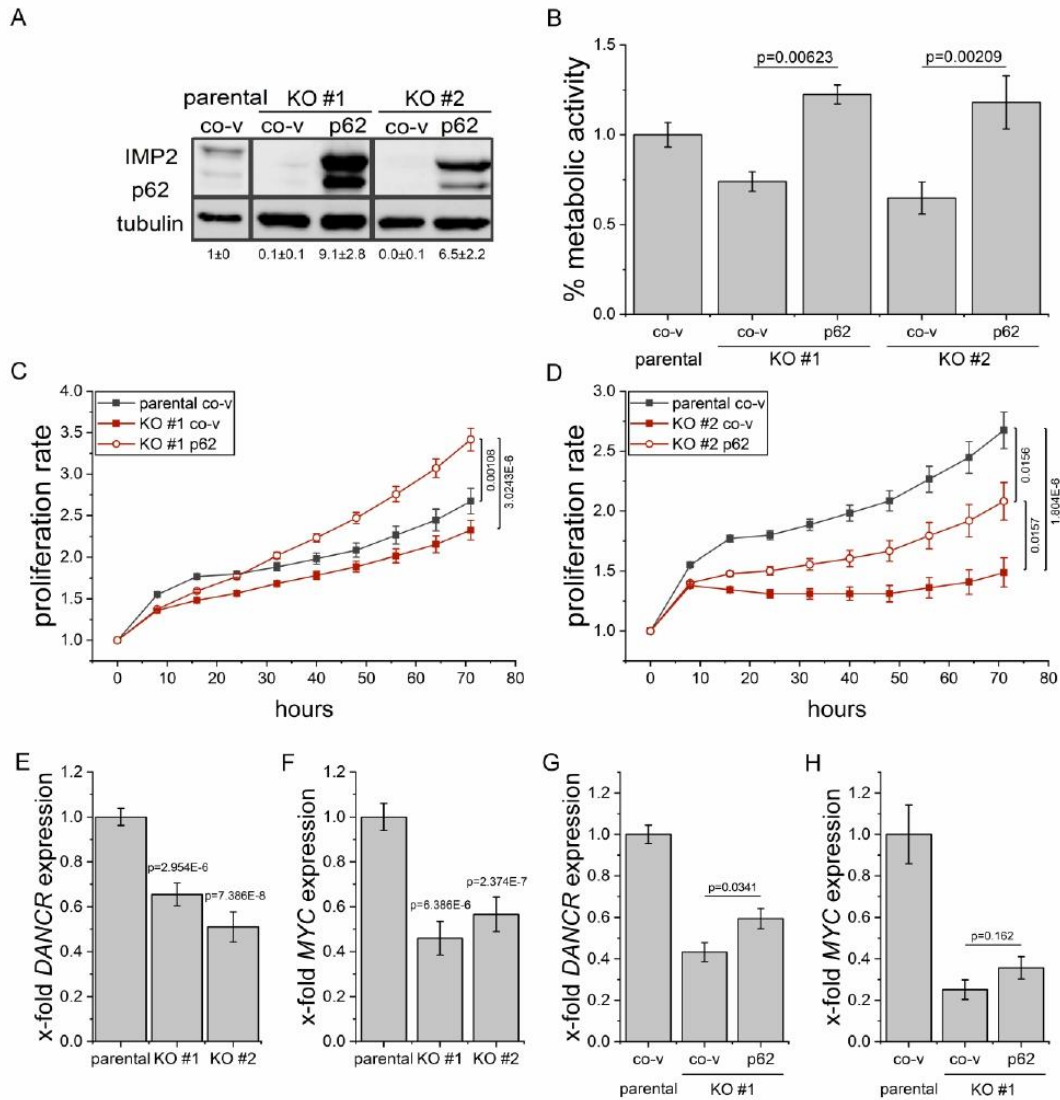
Supplementary Figure S5. Target specificity of IMP2 knockout in HCT116 cells

(A-D) HCT116 parental and IMP2 knockout HCT116 cells (CRISPR/Cas9 clone KO #1, prime editing clone KO #2) were transfected with p62/IMP2 or control vector (co-v).

(A) Transfection efficiency and p62/IMP2 overexpression was controlled by Western blot 3 days post transfection.

(B) Metabolic activity was measured *via* MTT assay 3 days post transfection.

(C, D) Cell confluency was monitored using the IncuCyte® S3 system over 3 days. Confluency was normalized to the time point of transfection (0 h). Data are represented as means ± SEM, n=2 (quadruplicates).

(E, F) *DANCR,* and *MYC* gene expression was determined in HCT116 IMP2 knockout clones and (G, H) p62/IMP2 overexpressing parental and knockout cells by qPCR. Values were normalized to the housekeeping gene *RNA18S,* n=3 (triplicates). Data are represented as means ± SEM.

A

M: marker; 1: supernatant; 2: column flow through;
3: binder buffer wash; 4: high salt wash; 5: 300 mM
imidazole; 6: 700 mM imidazole; 7: 1 M imidazole.

B

M: marker; 1-3: IMP2 elution fractions.

C

M: 1Kb marker; 1: RNA + storage buffer; 2: RNA
+ IMP2; 3: RNA + IMP2 after 1.5 h at RT.

D

	ratio
Helix	37.5
Beta	28.5
Turn	5.8
Random	28.2
Total	100
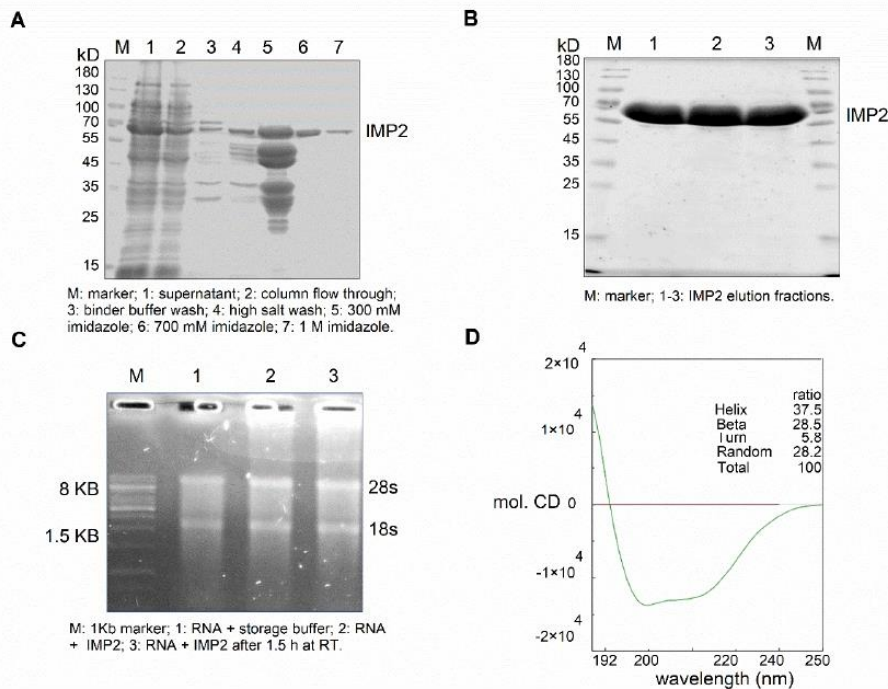
Supplementary Figure S6. IMP2 isolation and characterization

Histidine-tagged IMP2 was overexpressed in *E. coli* and isolated *via* affinity chromatography using a HisTrap HP Nickel–Sepharose column. Protein was eluted in an imidazole buffer with increasing imidazole concentrations.

(A) Fractions were collected and run on an SDS-PAGE (lanes 5-7). A 10-180 kD protein ladder marker (M), the unpurified cell lysate (1), the column flow-through (2), and washing buffers (3-4) were also run on the gel. The gel was stained with Coomassie blue and revealed the pure IMP2 protein in the 300 mM and 700 mM imidazole fractions (lanes 6 and 7). IMP2 containing fractions were combined and concentrated.

(B) The identity of the 67 kDa protein IMP2 was confirmed by Western blot analysis.

(C) The absence of RNase activity was confirmed *via* RNA integrity measurement of MCF7 RNA in the presence of eluted protein, as visualized on an agarose gel. M: 1 kb marker, 1: RNA incubated with storage buffer for 1.5 h as a control, 2: RNA incubated with IMP2 for 1.5 h on ice, 3: RNA incubated with IMP2 for 1.5 h at room temperature (RT).

(D) Circular dichroism (CD) spectroscopy was used to verify the correct protein folding of the purified protein.

S6

133
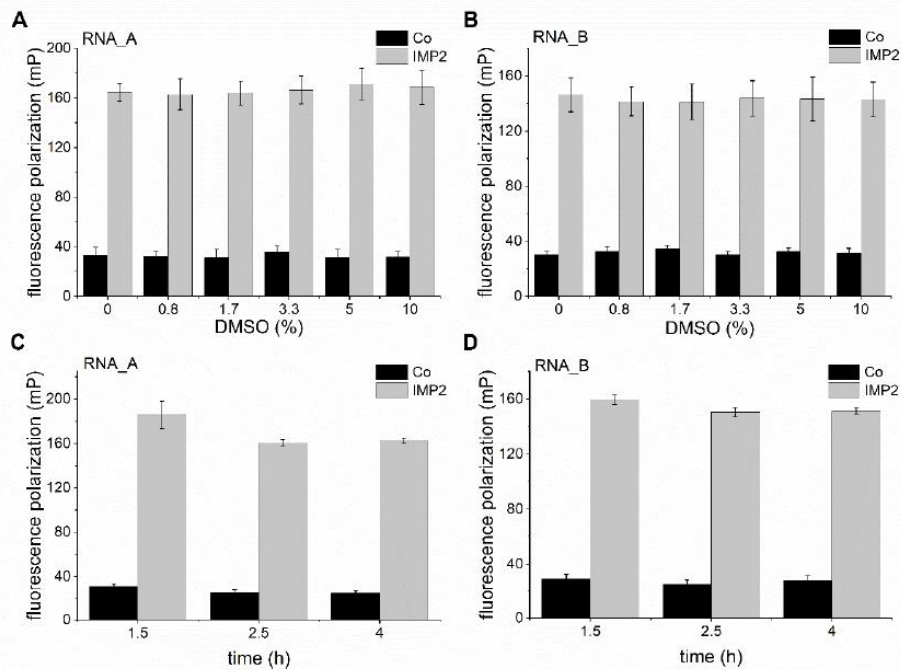
Supplementary Figure S7. Fluorescence polarization assay development and validation

(A, B) The DMSO tolerance of the FP assay was determined by using 1 nM of either the (A) RNA_A or the (B) RNA_B probe, IMP2 (120 nM for RNA_A and 160 nM for RNA_B), and varying concentrations of DMSO v/v. Unlabeled RNA was used as a control.

(C, D) The stability of the protein-RNA complex was assessed for 5% DMSO at different time points. Data are represented as means ± SD, n=2 (triplicates).

S7

134
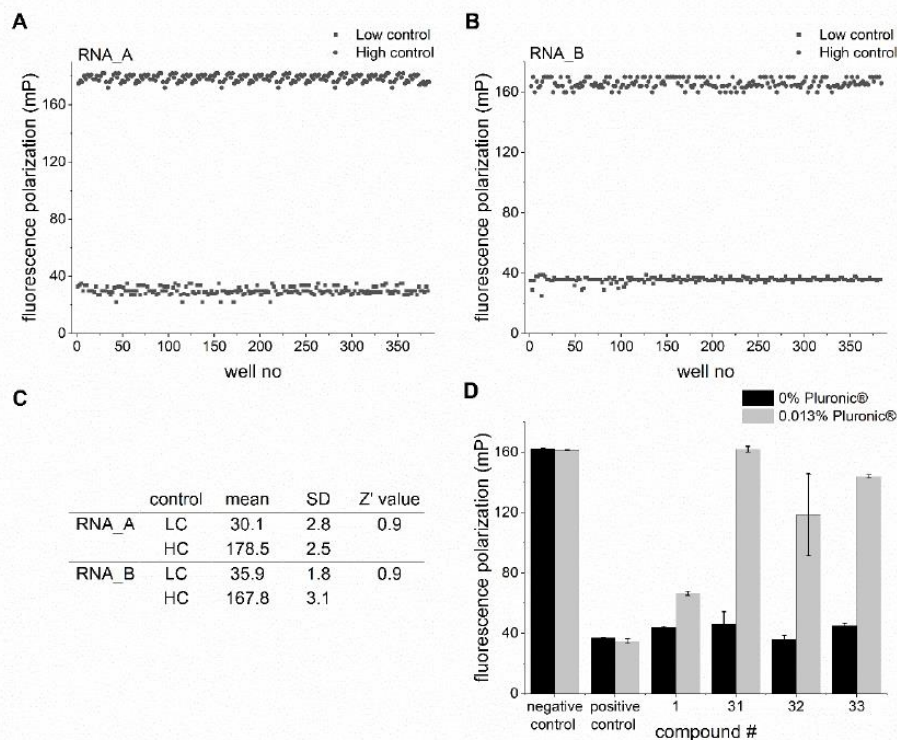
Supplementary Figure S8. The robustness of the FP assay

(A, B) To determine the robustness of the FP assay, 192 samples of low controls (LC), containing 1 nM (A) RNA_A or (B) RNA_B without IMP2, and 192 samples of high controls (HC) containing additionally 120 nM and 160 nM IMP2 for RNA_A and RNA_B, respectively, were assessed at 5% DMSO in the FP assay after 1.5 h incubation.

(C) Z'-factors were calculated based on the obtained data.

(D) To minimize unspecific aggregation and, therefore, false-positive results, 0.013% Pluronic® were added to FP buffer. The inhibitory effect of compounds **31 – 33** was lost after addition of Pluronic®, but not for compound **1**. Data are represented as means ± SD, n=1 (duplicates).

S8

135
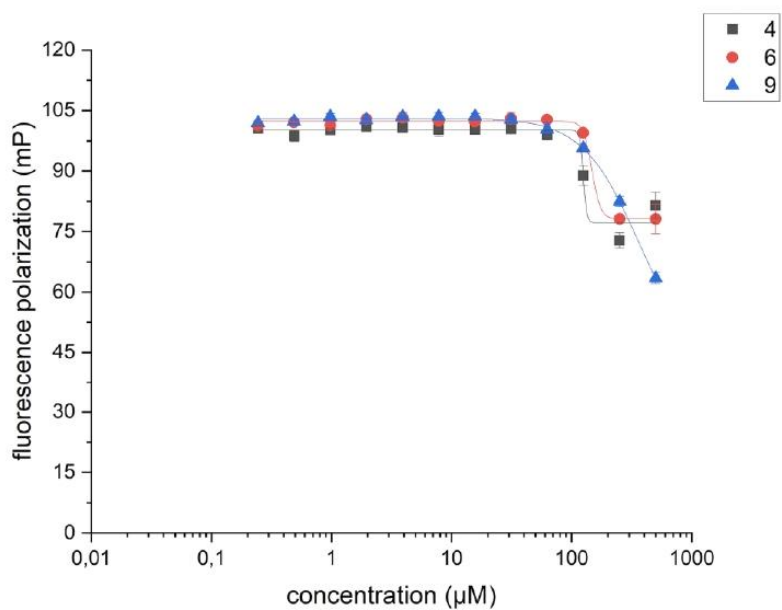
Supplementary Figure S9. Hit compounds tested against the RBP CsrA

Competition binding assays were performed by FP assay using CsrA from *Y. pseudotuberculosis* to test the specificity of IMP2:compound interactions. Hit compounds **4**, **6** and **9** were used in concentrations up to 500 µM to compete with the fluorescence labelled target RNA (15 nM) for CsrA (400 nM) binding.

S9

136
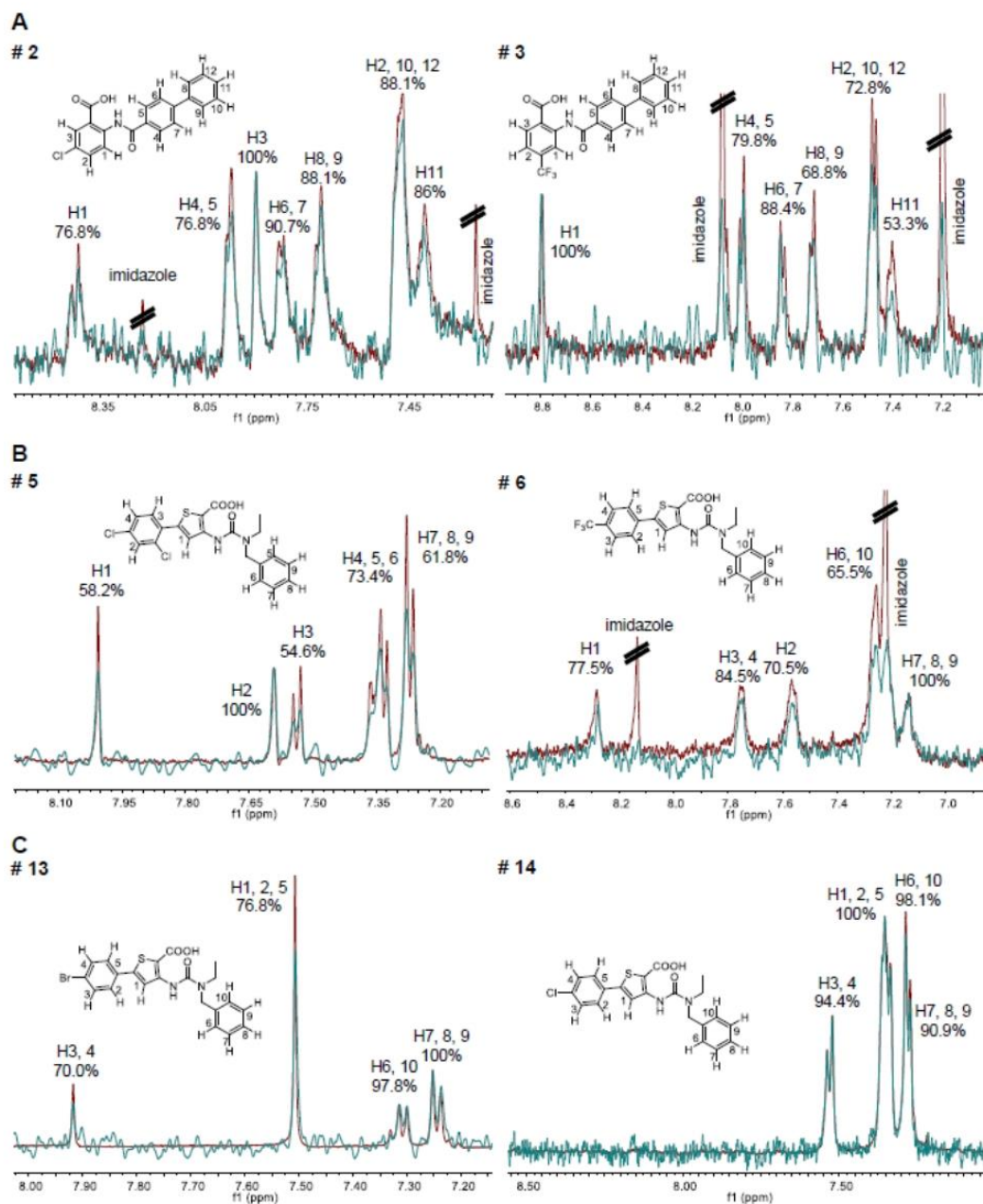
Supplementary Figure S10. STD-NMR analysis

STD-NMR experiments were performed at fixed concentrations of 2.5-5 µM IMP2 and either 250 µM for compound **2** and **3** or 500 µM for compounds **5-6** and **13-14** based on the solubility limit in 10% DMSO D_6 (molar ratio of protein to ligand was 1:100). Compounds **2** and **3** represent class A hit compounds (A), compounds **5** and **6** class B compounds (B), and compounds **13** and **14**, selective RNA_A inhibitors (C). The reference spectrum without protein is shown in red, and the STD difference spectrum of the IMP2/compound complexes is shown in green. Overlaid STD off-resonance and STD effect spectra were normalized to the signal of the highest proton signal.
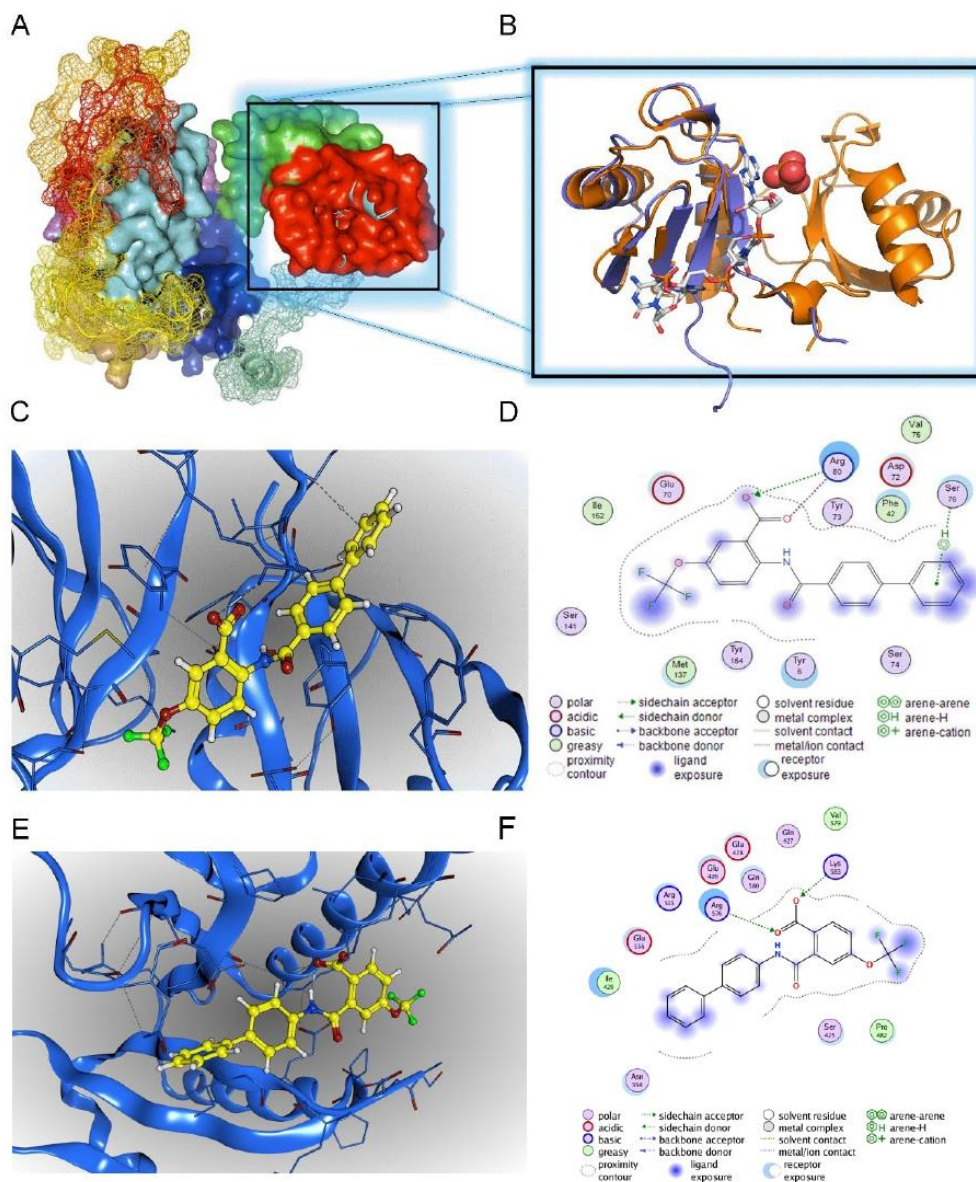
Supplementary Figure S11. Molecular docking analysis of IMP complex

(A) IMP2 structure prediction based on IMP2 homology modeling.

(B) Overlay of IMP2 RRM1 (blue), and IMP3 RRM12 (orange) crystals show 3D structure similarity.

(C) 3D depiction of the docking-derived binding hypothesis for hit compound **4**. RNA binding sites on the IMP2 RRM1 homology model were identified based on IMP3 RRM12-binding RNA coordinates and used as the docking site.

(D) Ligand interaction scheme for the docking pose of compound **4** in complex with IMP2 RRM1.

(E) 3D depiction of the docking-derived binding hypothesis for hit compound **4** to the IMP2 KH34 domain. RNA binding sites on the IMP2 KH34 domain are reported in the literature and were used as docking site (Biswas et al., 2019).

(F) Ligand interaction scheme for the docking pose of compound **4** in complex with the IMP2 KH34 domain.

S11

138
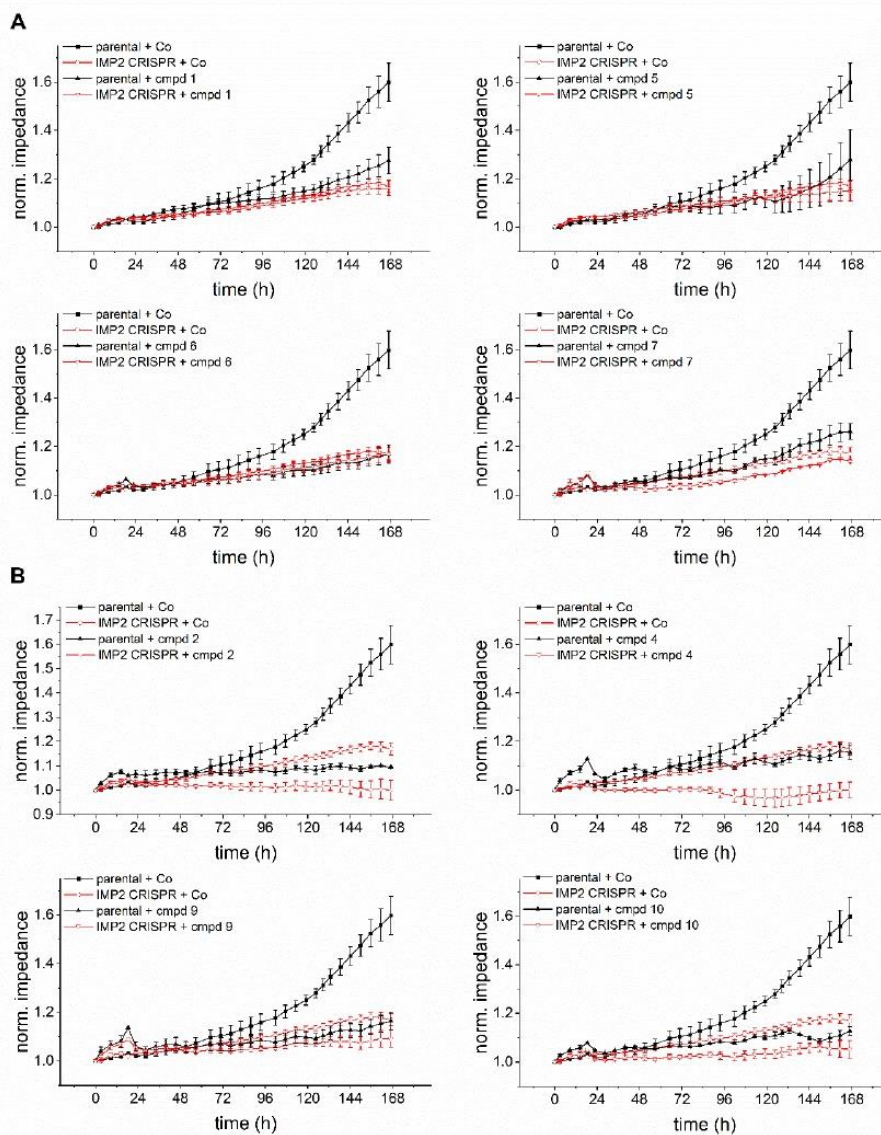
Supplementary Figure S12. Action of hit compounds in the absence of the target on cell impedance changes

Cell impedance was assessed as readout parameter for cell density and adhesion. HCT116 parental and IMP2 CRISPR/Cas9 knockout cells were seeded in equal numbers and treated with 25 µM of the respective compound or DMSO solvent control (co). Hit compounds demonstrated effective anti-proliferative effects in HCT116 parental cells but (A) no or (B) lower effects in IMP2 CRISPR cells. Data are represented as means ± SEM, n=2 (triplicates).

S12
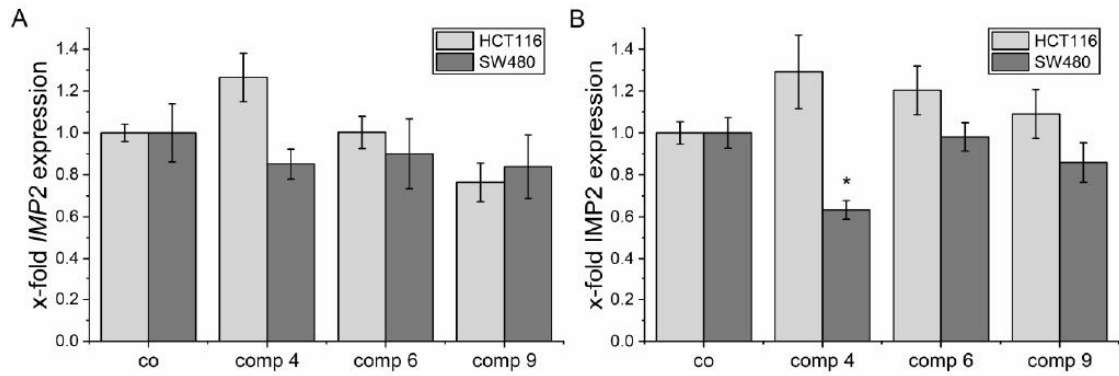
Supplementary Figure S13. Expression of IMP2 upon compound treatment

(A) Quantification of *IMP2* gene levels of HCT116 and SW480 cells, treated with compound **4** (40 μM), **6** (50 μM), or **9** (50 μM) for 24 h. Data are normalized to *RNA18S* and are represented as means ± SEM, n=3.

(B) Quantification of IMP2 protein levels of HCT116 and SW480 cells, treated with compound **4** (40 μM), **6** (50 μM), or **9** (50 μM) for 24 h. Data are normalized to tubulin and are represented as means ± SEM, n=2.
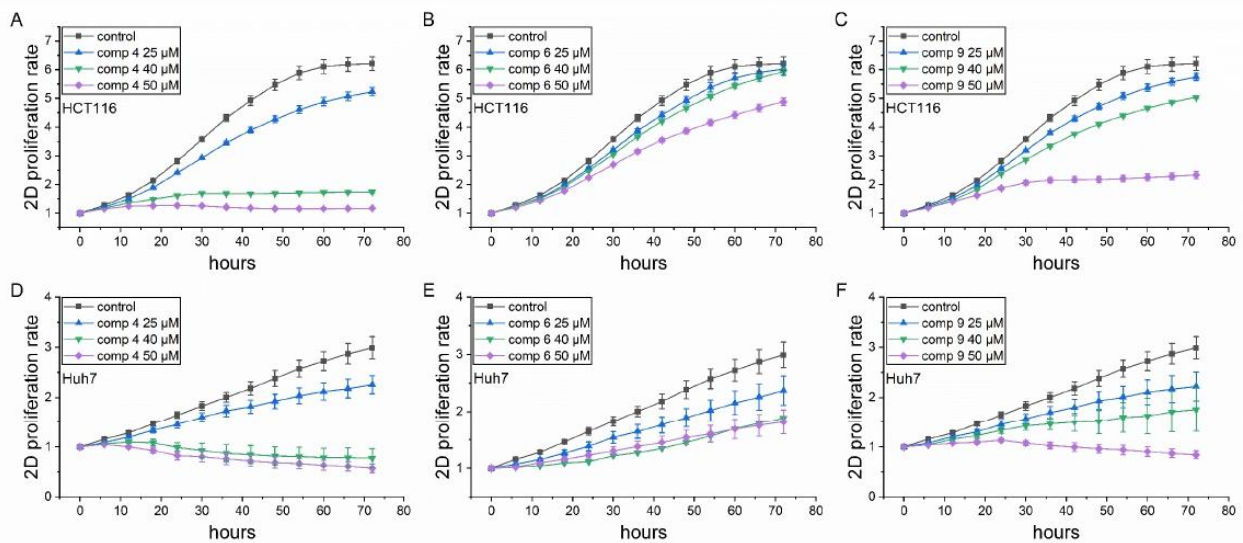
Supplementary Figure S14. *In vitro* actions of compounds on 2D tumor cell proliferation

Live-cell microscopy-based analysis of compound-induced anti-proliferative activity. Confluency of (A-C) HCT116 and (D-F) Huh7 cells was monitored in an IncuCyte® S3 system during compound (**4**, **6**, **9**) or control treatment over 72 h and normalized to the point of treatment (0 h). Data are represented as means ± SEM, n=3 (quadruplicates).

140
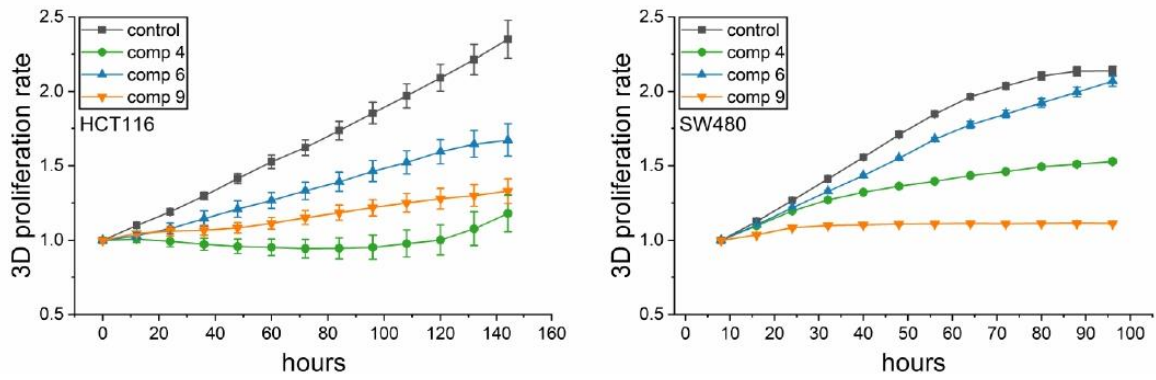
Supplementary Figure S15. *In vitro* actions of compounds on tumor cell proliferation

Live-cell microscopy-based analysis of compound-induced anti-proliferative activity. 3-day old HCT116 (left panel) and SW480 (right panel) spheroids were treated with 50 µM of the respective compounds and the spheroid area was monitored by the IncuCyte® S3 system. The spheroid area was normalized to the first measuring point after treatment. Data are represented as means ± SEM, n=3 (quadruplicates).
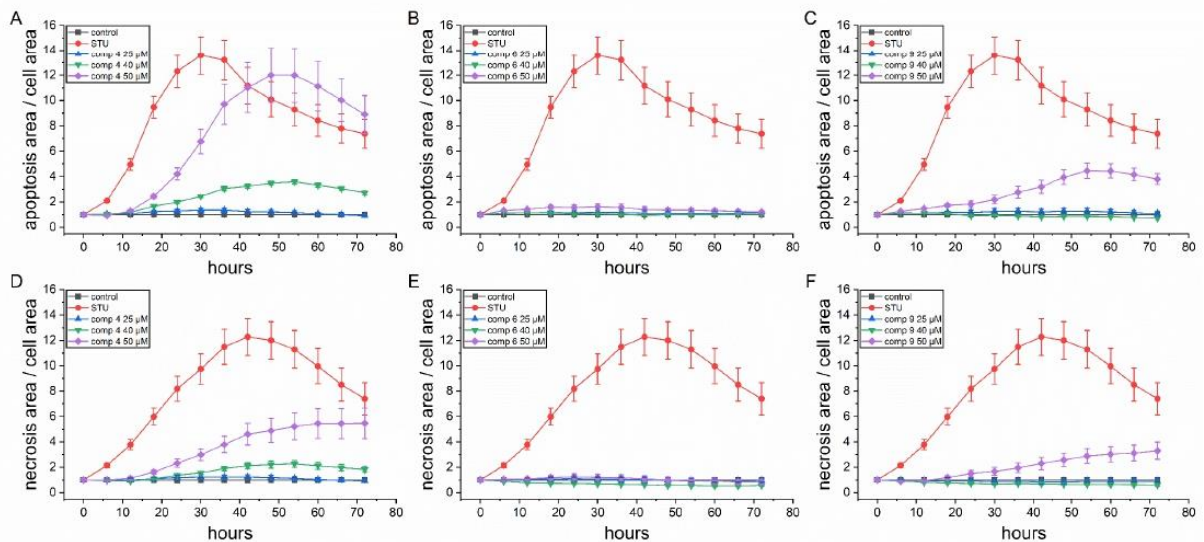
Supplementary Figure S16. Evaluation of compound-induced cell death

Live cell microscopy-based analysis of cell death upon compound treatment. HCT116 cells were stained for (A–C) caspase 3/7 activity and (D-F) cell membrane permeability and monitored in an IncuCyte® S3 system during compound (**4**, **6**, **9**) or vehicle control treatment over 72 h. The apoptosis inducer staurosporine (STU, 1 µM) was used as positive control. Fluorescent signals from apoptotic (caspase 3/7 active) and necrotic (permeable membrane) cells were normalized to cell confluency and the time point of treatment (0h). Data are represented as means ± SEM, n=3 (quadruplicates).

141
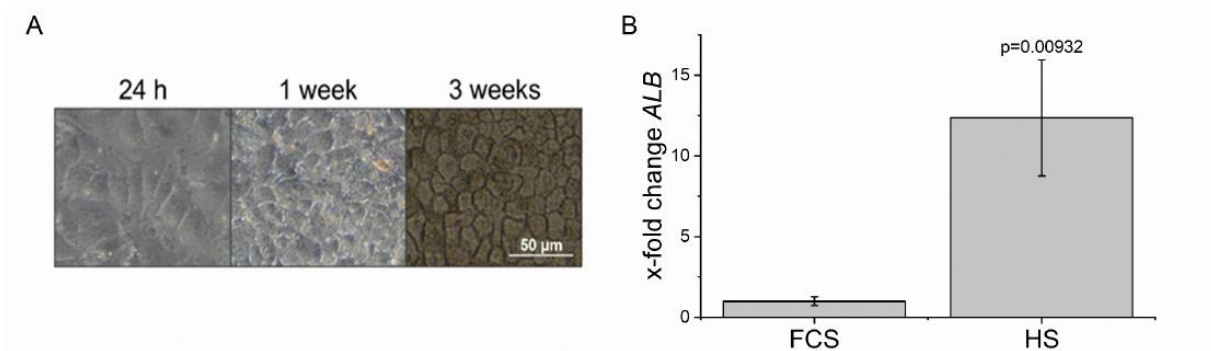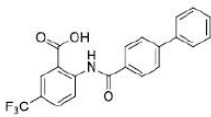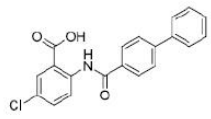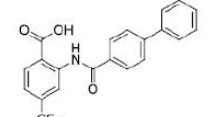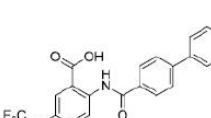
Supplementary Figure S17. Differentiation process of Huh7 in human serum

Huh7 cells were differentiated in media supplemented with 2% human serum (HS) for 3 weeks.
(A) Cell morphology was monitored microscopically 24 h, 1 week, and 3 weeks after medium change.
(B) Gene expression of albumin (*ALB*) in differentiated and FCS-cultured cells. RNA was isolated 3 weeks after medium change and gene expression was assessed by qPCR. Values were normalized to the housekeeping gene *RPS11*. Data (x-fold of values for FCS-cultured cells) are represented as means ± SEM, n=4 (triplicates).

S17

142

Supplementary Table 1. Descriptive data of class A compounds

Class A compounds: Molecular weights, chemical structures, and analytical data. Abbreviations: carbon-NMR (^{13}C NMR); coupling constant (J); deuterated dimethyl sulfoxide (DMSO.d$_6$); doublet peak (d); liquid chromatography–mass spectrometry (LC/MS); melting point (mp); multiplet peak (m); parts per million (ppm); proton NMR (^{1}H NMR); quartet peak (q); retention time (t$_R$); singlet peak (s); triplet peak (t).

#	Reference	MW	Structure	Analytical Data
1	Hinsberger et al., 2014	385.34		^{1}H NMR (500 MHz, DMSO-d_6) δ ppm 12.41 (s, 1H) 8.93 (d, J = 8.8 Hz, 1H) 8.28 (d, J = 1.9 Hz, 1H) 8.03 - 8.06 (m, 2H) 8.01 (dd, J = 8.8, 1.9 Hz, 1H) 7.87 - 7.92 (m, 2H) 7.73 - 7.78 (m, 2H) 7.47 - 7.54 (m, 2H) 7.40 - 7.46 (m, 1H). ^{13}C NMR (126 MHz, DMSO-d_6) δ ppm 168.90, 164.66, 144.27, 144.01, 138.75, 132.65, 130.85 (q, J_{CF} = 3.7 Hz) 129.06, 128.35, 127.94 (q, J_{CF} = 3.7 Hz), 127.84, 127.20, 126.95, 123.80 (q, J_{CF} = 271.3 Hz), 122.76 (q, J_{CF} = 33.0 Hz) 120.36, 116.81. LC/MS: m/z = 386, 771; t$_R$ = 14.60 min; 99.76% pure (UV). White solid; mp. 264.8-266.0°C.
2	Hinsberger et al., 2014	351.78		^{1}H NMR (500 MHz, DMSO-d_6) δ ppm 12.14 (s, 1H) 8.74 (d, J = 9.1 Hz, 1H) 8.02 - 8.05 (m, 2H) 8.00 (d, J = 2.8 Hz, 1H) 7.86 - 7.91 (m, 2H) 7.75 - 7.79 (m, 2H) 7.73 (dd, J = 9.0, 2.7 Hz, 1H) 7.48 - 7.54 (m, 2H) 7.41 - 7.47 (m, 1H). ^{13}C NMR (126 MHz, DMSO-d_6) δ ppm 168.77, 164.38, 143.80, 139.88, 138.82, 133.87, 132.92, 130.39, 129.08, 128.32, 127.75, 127.18, 126.96, 126.53, 121.81, 118.55. LC/MS: m/z = 352, 354; t$_R$ = 14.32 min; 99.33% pure (UV). White solid; mp. 273.8-278.9°C.
3	Hinsberger et al., 2014	385.34		1H NMR (500 MHz, DMSO-d6): δ ppm 12.35 (s, 1H) 9.10 (d, J = 1.6 Hz, 1H) 8.26 (d, J = 8.2 Hz, 1H) 8.04 - 8.08 (m, 2H) 7.89 - 7.95 (m, 2H) 7.75 - 7.81 (m, 2H) 7.57 (dd, J = 8.2, 1.6 Hz, 1H) 7.49 - 7.55 (m, 2H) 7.41 - 7.47 (m, 1H). 13C NMR (126 MHz, DMSO-d6): δ ppm 169.00, 164.78, 143.98, 141.41, 138.79, 133.38 (q, JCF = 32.0 Hz), 132.67, 132.45, 129.08, 128.35, 127.80, 127.23, 126.97, 122.88 (q, JCF = 273.1 Hz), 119.15 (q, JCF = 3.7 Hz), 116.25 (q, JCF = 3.7 Hz). LC/MS: m/z = 386, 427, 771; tR = 14.11 min; 100.00% pure (UV). White solid; mp. 244.9-245.8°C.
4	Hinsberger et al., 2014	401.34		^{1}H NMR (500 MHz, DMSO-d_6): δ ppm 12.17 (s, 1H) 8.81 (d, J = 9.1 Hz, 1H) 8.03 - 8.06 (m, 2H) 7.92 (dd, J = 3.0, 0.8 Hz, 1H) 7.88 - 7.91 (m, 2H) 7.74 - 7.78 (m, 2H) 7.68 - 7.74 (m, 1H) 7.49 - 7.53 (m, 2H) 7.41 - 7.46 (m, 1H). ^{13}C NMR (126 MHz, DMSO-d_6): δ ppm 168.61, 164.48, 143.86, 142.66 (q, J_{CF} = 1.8 Hz), 140.12, 138.82, 132.88, 129.08, 128.33, 127.80, 127.20, 127.04, 126.96, 123.31, 121.90, 118.55, 120.04 (q, J_{CF} = 256.0 Hz). LC/MS: m/z = 402, 803; t$_R$ = 14.58 min; 100.00% pure (UV). White solid; mp. 244.6-245.9°C.

Supplementary Table 2. Descriptive data of class B compounds

Class B compounds: Molecular weights, chemical structures, and analytical data. Abbreviations: carbon-NMR (^{13}C NMR); coupling constant (J); deuterated dimethyl sulfoxide (DMSO.d$_6$); doublet peak (d); liquid chromatography–mass spectrometry (LC/MS); melting point (mp); multiplet peak (m); parts per million (ppm); proton NMR (^{1}H NMR); quartet peak (q); retention time (t$_R$); singlet peak (s); triplet peak (t).

#	Reference	MW	Structure	Analytical Data
5	Sahner et al., 2013	449.35		^{1}H-NMR (300 MHz, DMSO-d_6): δ ppm 1.15 (t, J = 7.1, 3H), 3.39 (q, J = 7.1, 2H), 4.58 (s, 2H), 7.23 – 7.37 (m, 5H), 7.53 (dd, J = 8,5, 2,1, 1H), 7.68 (d, J = 8.5, 1H), 7.79 (d, J = 2.1, 1H), 8.27 (s, 1H), 10.04 (s, 1H), 13.51 (br. s., 1H). ^{13}C NMR (75 MHz, DMSO-d_6) δ ppm 13.1, 41.8, 49.3, 108.8, 122.1, 127.1, 128.1, 128.5, 130.1, 130.5, 132.1, 132.4, 134.3, 138.0, 142.3145.6, 153.0, 165.6. LC/MS: m/z = [M+H$^+$] 448.55, [2M+H$^+$] 898.57 t$_R$ = 15.14 min, 99.3% pure (UV). mp: 212 - 214°C.
6	Sahner et al., 2013	448.46		^{1}H-NMR (300 MHz, DMSO-d_6): δ ppm 1.16 (t, J = 7.1, 3H), 3.40 (q, J = 7.1, 2H), 4.59 (s, 2H), 7.24 – 7.38 (m, 5H), 7.79 (d, J = 8.4, 2H), 7.91 (d, J = 8.1, 2H), 8.41 (s, 1H), 10.05 (s, 1H), 13.47 (br. s., 1H). ^{13}C NMR (75 MHz, DMSO-d_6) δ ppm 13.1, 41.8, 49.3, 108.2, 124.0 (q, $^1J_{CF}$ = 272.0 Hz),126.2 (q, $^3J_{CF}$ = 3.7 Hz), 126.4, 127.2, 128.4, 128.5, 129.3 (q, $^2J_{CF}$ = 32.0 Hz) 136.5, 138.1, 145.3, 146.7, 153.0, 165.6. LC/MS: m/z = [M+H$^+$] 448.68, [2M+H$^+$] 896.93 t$_R$ = 14.59 min, 98.02% pure (UV). mp: 193 – 194°C.
7	Sahner et al., 2013	510.96		^{1}H-NMR (300 MHz, DMSO-d_6): δ ppm 0.88 (t, J = 7.5, 3H), 1.25 – 1.37 (m, 2H), 1.52 – 1.63 (m, 2H), 3.32 (t, J = 7.5, 2H), 4.59 (s, 2H), 7.23 – 7.38 (m, 5H), 7.77 – 7.80 (s, 1H), 7.98 – 8.00 (m, 2H), 8.38 (s, 1H), 10.08 (s, 1H), 13.56 (br. s, 1H). LC/MS: m/z = [M+H$^+$] 510.85; t$_R$ = 16.16 min, 100% pure (UV).
8	Sahner et al., 2013	511.42		^{1}H-NMR (300 MHz, DMSO-d_6): δ ppm 4.64 (s, 4H), 7.26 – 7.38 (m, 10H), 7.65 – 7.71 (m, 2H), 7.94 (s, 1H), 8.35 (s, 1H), 10.08 (s, 1H), 12.66 (br. s, 1H). ^{13}C NMR (75 MHz, DMSO-d_6) δ ppm 49.9, 50.1, 118.8, 128.9, 127.1, 127.3, 127.3, 128.6, 128.6, 128.9, 130.0, 131.4, 131.6, 131.8, 132.1, 133.3, 137.2, 144.2, 146.2, 153.5, 165.4. LC/MS: m/z = [M+H$^+$] 511.35; t$_R$ = 16.36 min, 95.70% pure (UV). mp: 190 – 191°C.
9	Sahner et al., 2013	477.40		^{1}H-NMR (300 MHz, DMSO-d_6): δ ppm 0.88 (t, J = 7.4, 3H), 1.60 (q, J = 7.4, 2H), 2.89 (t, J = 7.4, 2H), 3.25 (t, J = 7.4, 2H), 3.50, J = 7.4, 2H), 7.19 – 7.32 (m, 5H), 7.65 – 7.71 (m, 2H), 7.93 (d, J = (dd, J = 1.1, 1H), 8.33 (s, 1H), 10.03 (s, 1H), 13.56 (br. s, 1H). ^{13}C NMR (75 MHz, DMSO-d_6) δ ppm 11.0, 21.2, 34.1, 49.0, 49.1, 197.7, 118.7, 125.9, 126.2, 127.3, 128.4, 128.8, 131.4, 131.6, 132.1, 133.3, 138.7, 144.4, 146.7, 152.6, 165.8. LC/MS: m/z = [M+H$^+$] 476.89, [2M+H$^+$] 954.65 t$_R$ = 16.58 min, 98.83% pure (UV). mp: 178 – 179°C.
10	Sahner et al., 2013	524.98		^{1}H-NMR (300 MHz, DMSO-d_6): δ ppm 0.88 (t, J = 7.3, 3H), 1.25 – 1.37 (m, 2H), 1.50 – 1.60 (m, 2H), 2.89 (t, J = 7.8, 2H), 3.29 (t, J = 8.1, 2H), 3.50 (t, J = 8.2, 2H), 7.18 – 7.37 (m, 5H), 7.78 – 7.81 (m, 1H), 7.99 – 8.01 (m, 2H), 8.38 (s, 1H), 10.04 (s, 1H), 13.61 (br. s., 1H). LC/MS: m/z = [M+H$^+$] 524.86 t$_R$ = 16.85 min, 96.95% pure (UV).

Supplementary Table 3. IC$_{50}$ values of hit compounds

Metabolic activity was determined by MTT assay 96 h after treatment with hit compounds or DMSO solvent control. IC$_{50}$ values were calculated using non-linear regression analysis. n=2-5 (triplicates).

IC$_{50}$ values [µM]						
	MCF7	HCT116	SW480	HepG2	Huh7	Hep3B
1	43.6	48.1	35.1	30.7	34.7	54.7
2	46.7	29.9	22.5	29.1	40.9	45.5
3	> 80	53.2	78.5	75.0	34.5	58.0
4	48.1	31.0	18.2	29.6	33.3	35.5
5	> 80	62.6	61.2	> 80	> 80	> 80
6	55.1	46.9	49.0	42.9	45.4	52.6
7	> 80	44.7	27.1	35.0	38.8	47.7
8	50.3		37.1	33.6	43.5	
9	70.0	37.8	36.8	35.7	24.9	39.8
10	58.0	37.5	41.4	35.0	44.3	46.2

Supplementary Table 4. Sequences of siRNA oligonucleotides used in IMP2 knockdown
Individual sequences of the 4-oligo siRNA mixture purchased from Qiagen to knock down IMP2 and the random RNA oligo sequence.

Oligo RNA	Sequence (5'- 3')
oligo 1	CAGGGCGTTAAATTCACAGAT
oligo 2	TCCGCTAGCCAAGAACCTATA
oligo 3	CAGCGAAAGGATGGTCATCAT
oligo 4	CCCGGGTAGATATCCATAGAA
random	AACACGTCTATACGC

145

2. Publication Title:

Integration of in vivo pooled CRISPR and direct capture Perturb-seq screens identifies synthetic dosage lethal interactions of Polo-Like Kinase 1 in cancer cells

Authors:

Chelsea E. Cunningham*, Frederick S. Vizeacoumar*, Yue Zhang*, Liliia Kyrylenko*, Peng Gao, Vincent Maranda, Ashtalakshmi Ganapathysamy, He Dong, Simon Both, Konrad Wagner, Yingwen Wu, Jared Price, Tetiana Katrii, Ben G. E. Zoller, Karthic Rajamanickam, Prachi Walke, Li Hui, Hardikkumar Patel, Alain Morejon Morales, Kalpana K. Bhanumathy, Hussain Elhasasna, Renuka Dahiya, Omar Abuhussein, Tanya Freywald, Erika Prando Munhoz, Anand Krishnan, Eytan Ruppin, Joo Sang Lee, Katharina Rox, Behzad Toosi, Raju Datla, Sunil Yadav, Yuliang Wu, Martin Empting, Alexandra K. Kiemer, Andrew Freywald and Franco J. Vizeacoumar

Integration of *in vivo* pooled CRISPR and direct capture Perturb-seq screens identifies synthetic dosage lethal interactions of Polo-Like Kinase 1 in cancer cells

Chelsea E. Cunningham[1]*, Frederick S. Vizeacoumar[1,2]*, Yue Zhang[1]*, Liliia Kyrylenko[1]*, Peng Gao[3,4], Vincent Maranda[1,2], Ashtalakshmi Ganapathysamy[1], He Dong[1], Simon Both[5], Konrad Wagner[6], Yingwen Wu[6], Jared Price[1,2], Tetiana Katrii[1,2], Ben G. E. Zoller[6], Karthic Rajamanickam[1], Prachi Walke[1,7], Li Hui[1], Hardikkumar Patel[1,2], Alain Morejon Morales[2], Kalpana K. Bhanumathy[1], Hussain Elhasasna[1], Renuka Dahiya[2], Omar Abuhussein[8], Tanya Freywald[1], Erika Prando Munhoz[1], Anand Krishnan[8], Eytan Ruppin[9], , Joo Sang Lee[9,10], Katharina Rox[11], Behzad Toosi[12] Raju Datla[3], Sunil Yadav[1], Yuliang Wu[13], Martin Empting[6,$], Alexandra K. Kiemer[5,$], Andrew Freywald[2,$] and Franco J. Vizeacoumar[1,14,$]

[1]Division of Oncology, Cancer Cluster, College of Medicine, University of Saskatchewan, Saskatoon, S7N 5E5 Canada. [2]Department of Pathology and Laboratory Medicine, University of Saskatchewan, Saskatoon, SK S7N 5E5, Canada. [3]Global Institute for Food Security, University of Saskatchewan, Saskatoon, SK, S7N 4L8, Canada. [4]Agriculture and Agri-Food Canada, Saskatoon Research and Development Centre, 107 Science Place, Saskatoon, SK, S7N 0X2, Canada. [5]Department of Pharmacy, Pharmaceutical Biology, Saarland University, Saarbrücken 66123, Germany. [6]Antiviral & Antivirulence Drugs (AVID), Helmholtz Institute for Pharmaceutical Research, Saarland (HIPS), Helmholtz Centre for Infection Research (HZI) and Department of Pharmacy, Saarland University, Saarbrücken 66123, Germany. [7]Department of Anatomy, Physiology, and Pharmacology, University of Saskatchewan, and Cameco MS Neuroscience Research Centre, 701 Queen St., Saskatoon, S7K 0M7, SK, Canada. [8]College of Pharmacy, University of Saskatchewan, 104 Clinic Place, Saskatoon, S7N 2Z4, Canada. [9]Cancer Data Science Laboratory, National Cancer Institute, National Institute of Health, Bethesda, MD 20892, USA; Center for Bioinformatics and Computational Biology & Department of Computer Sciences, University of Maryland, College Park, MD 20742, USA. [10]Department of Precision Medicine, School of Medicine & Department of Artificial Intelligence, Sungkyunkwan University, Suwon, 16419, Republic of Korea. [11]Department of Chemical Biology (CBIO), Helmholtz Center for Infection Research (HZI), Inhoffenstrasse 7, 38124 Braunschweig, Germany. [12]Western College of Veterinary Medicine, University of Saskatchewan, Room 2343, 52 Campus Drive, Saskatoon, S7N 5B4 Canada. [13]Department of Biochemistry, Microbiology and Immunology, University of Saskatchewan, Saskatoon, SK S7N 5E5, Canada. [14]Cancer Research, Saskatchewan Cancer Agency, 107 Wiggins Road, Saskatoon, S7N 5E5, Canada

* These authors contributed equally

$ Correspondence to:
Franco J. Vizeacoumar: franco.vizeacoumar@usask.ca
Andrew Freywald: andrew.freywald@usask.ca
Alexandra K. Kiemer: pharm.bio.kiemer@mx.uni-saarland.de
Martin Empting: martin.empting@helmholtz-hips.de

Running Title: Targeting PLK1 by SDL

Key Words: Synthetic dosage lethality, Tumor heterogeneity, PLK1, Chromosomal instability, Perturb-seq, Single cell CRISPR screening, *in vivo* CRISPR screen, IGF2BP2, IMP2.

Abstract

Tumor heterogeneity represents a major therapeutic challenge that underlies treatment resistance. The polo-like kinase 1 (PLK1) is overexpressed across all cancers, inducing

chromosomal instability (CIN). As CIN promotes intra-tumor heterogeneity, and direct PLK1 inhibition has not yielded clinical advances, we aimed to target its genetic interactions by using the synthetic dosage lethality (SDL) approach to overcome the challenges associated with heterogeneity and ensure selective killing of PLK1 overexpressing tumor cells. We performed a genome-wide *sh*RNA screen, selected 105 SDL candidates using computational strategies and evaluated them by an *in vivo* pooled CRISPR and a matching *in vitro* arrayed CRISPR screen in a patient-derived xenograft model of breast cancer. We next used direct capture Perturb-seq to characterize individual SDL hits that can eliminate PLK1-overexpressing single cells. Our unbiased strategy identified IGF2BP2/IMP2 as a top SDL hit. We also found that IGF2BP2 suppression, either genetically or pharmacologically, downregulates PLK1 and ultimately, reduces tumor growth. Taken together our observations strongly suggest that targeting genetic interactions of PLK1 represents a promising therapeutic avenue. As PLK1 is overexpressed in multiple cancers, our work is likely to trigger broad therapeutic implications.

Introduction

Tumor heterogeneity is an enormous clinical challenge as it provides selective evolutionary advantages to subsets of cancer cells, leading to the establishment of aggressive clones that are metastatic and treatment-resistant (*1-3*). The current drug development programs focus on co-targeting multiple pathways within cancer cells. For example, the clonal heterogeneity in colorectal cancer was overcome by simultaneous co-inhibition MEK and EGFR kinases (*4*). A promising approach to overcome intra-tumor heterogeneity is based on targeting factors that directly contribute to genetic diversity and intra-tumor heterogeneity (*5*). This approach should limit the acquisition of multi-drug resistance and treatment failure.

Chromosomal instability (CIN) is one of the key driving forces of genetic diversity within tumors, and at the same time remains as a key underlying feature of genetically diverse malignancies (*6-13*). CIN arises due to aberrant mitotic divisions or defective double strand break repair or replication stress or ineffective telomere maintenance (*7, 8, 10, 12, 14-19*). The polo-like kinase 1 (PLK1) is a serine/threonine protein kinase and is a central player in controlling CIN (*20-22*). On the molecular level, PLK1 contributes towards the stability of the genome by signaling the initiation of mitosis, centrosome maturation, bipolar spindle formation and chromosome segregation, and cytokinesis (*20, 23-26*). Constitutive overexpression of PLK1 leads to CIN and aneuploidy, a common salient feature of most cancers (*22, 27*). In fact, tumor cells upregulate genes like PLK1 to support their survival and propagation (*28-31*). Previously, we have shown that changes in the expression patterns of genes like PLK1, involved in the maintenance of CIN and factors that remodel tumor microenvironment, represent some of the earliest events of tumor evolution (*32*). Consistent with this idea, PLK1 is overexpressed in a wide range of cancers, including breast (*33*), colon (*34*), pancreatic (*35*), gliomas (*36*), lung (*37*) and prostate (*38*). However, even after 30 years of its discovery (*39*) and 20 years after recognition of its pro-malignant properties (*40*), bringing PLK1 inhibitors into clinical applications remains extremely challenging. The lackluster effectiveness of PLK1 inhibitors is due to the difficulty of achieving PLK1-specific inhibition (*24*). Unwanted inhibition of closely-related members of the polo-like family can lead to toxicities in nervous system (*41*), or those that might interfere with the hypoxic response and promote angiogenesis (*42*). Additionally, PLK1 itself plays key roles in controlling a multitude of cellular processes regulating CIN. Complete loss-of-function of this protein could be detrimental to normal cells (*24, 43*), and may defeat the tumor-selective basis of treatment. Thus, considering the diverse functions of PLK1 and its family members there are still major challenges in using PLK1 inhibitors in cancer therapy and alternative strategies are required.

To take advantage of frequent PLK1 overexpression in tumors, we applied the genetic approach called Synthetic Dosage Lethality (SDL), where an overexpression of a gene, like PLK1, is lethal only when another, normally non-lethal, genetic alteration is also present (*44, 45*). SDL is still a largely untapped area in cancer research. Since tumor cells upregulate genes like PLK1, discovery of SDL interactions is a valuable method to reveal new therapeutic targets for cancer treatment (*46-48*). Unlike a direct inhibition of PLK1, where cancer cells may undergo cell death or cell cycle arrest or increased aneuploidy, suppression of molecules that exhibit SDL interactions with PLK1 should result only in cell death **(Figure 1A)**. Here, we report the integration of multiple, unbiased platforms, including genome-wide pooled *sh*RNA screening with subsequent validation of SDL targets using pooled *in vivo* and arrayed *in vitro* CRISPR/Cas9 screens in a patient-derived xenograft (PDX) model. As PLK1 overexpression leads to cellular heterogeneity, there is also a need to confirm if suppression of its SDL partner(s) truly eliminates PLK1 overexpressing cells. Therefore, we evaluated the effect of the top SDL targets, at the single cell level, using a single cell CRISPR/Cas9 screen (Perturb-seq) with direct capturing of guide RNAs (*49*). This work convincingly identified IGF2BP2 as a potential new therapeutic target and a novel regulator of PLK1 expression. Subsequent studies using a newly characterized pharmacological inhibitor of IGF2BP2, was found to affect the expression of PLK1 and preferentially eliminates PLK1 overexpressing cells and tumors.

Methods
Cell lines, transfections, transductions and antibodies used in the study.
All cell lines were purchased from Cedarlane labs, unless indicated (Burlington, Ontario, Canada), a Canadian distributor for American Type Culture Collection (ATCC) or from MilliporeSigma. Cell lines purchased from ATCC/Sigma were passaged for less than three months at a time following resuscitation and therefore, no additional authentication was performed. BT549 cells were cultured in RPMI media with 850ng/ml Insulin, 10% fetal bovine serum and 1% penicillin/streptomycin at 37°C and 5% CO_2. MDA-MB-231 and MCF7 cells were cultured in DMEM media with 10% fetal bovine serum and 1% penicillin/streptomycin at 37°C and 5% CO_2. HCC1143 cells were cultured in RPMI medium with 10% fetal bovine serum and 1% penicillin/streptomycin at 37°C and 5% CO_2. Hs578Bst cells were cultured in ATCC Hybri-Care Medium (Catalog No. 46-X) with 30 ng/ml mouse EGF, 10% fetal bovine serum and 1% penicillin/streptomycin at 37°C and 5% CO_2. HCT116-PLK1 cells were (a kind gift from Pierre-Fabre) cultured in McCoy's 5A media supplemented with 5% fetal bovine serum and 1% penicillin/streptomycin at 37°C and 5% CO_2 with 200 µg/mL hygromycin (to allow the maintenance of Plk1 expression vectors) and 10 µg/mL blasticidin (to allow inducible TREX system selection) and induced with 4 ng/mL doxycycline or 500 ng/mL tetracycline. HCI-010 PDX cells (a kind gift from Welhm lab) were cultured using DMEM-F12 (Gibco) supplemented with 50 µg/mL gentamicin, B27 supplement (Gibco), 20 ng/mL human epidermal growth factor (Stemcell Technologies), 20 ng/mL basic fibroblast growth factor (Stemcell Technologies), 10 µg/mL insulin, 500 ng/mL hydrocortisone, and 0.02% heparin (Stemcell Technologies) in low-attachment tissue culture plates (Corning) at 5% CO_2. Lentiviral particles were generated by transfecting HEK293T cells with psPAX2, pMD2.G, and pLKO.1-shRNA 9:1:10 plasmid mixture using Xtremegene 9 transfection reagent (Sigma) and Optimem Serum Free media (Gibco). Media was replaced with DMEM containing 2% (w/v) bovine serum albumin 18 hours post-transfection, and lentivirus-containing media was harvested at 48- and 72-hours post-transfection. Transfections with siRNA for Hs578Bst were carried out using RNAiMax transfection reagent (Life Technologies) and Optimen Serum-free media to a final siRNA concentration of 50 nM. Cas9-expressing HCI-010 cells were generated by transducing HCI-

010 with Cas9-blast lentiviral particles (addgene #52962), replacing the lentiviral media after 24 hours, and adding 5 ug/mL blasticidin selection 48 hours post-transduction for 21 days. The antibodies GAPDH (sc-25778) and PLK1 (sc-5585) were obtained from Santa Cruz, Cas9 (ab191468), IGF2BP2 (ab128175) from Abcam, tubulin (#T902) from Merck, and PARP (9542S) from Cell Signaling.

Genome-wide pooled shRNA screening and data analysis
Screening and microarray scoring was performed as previously described (*50*). HCT116-PLK1 cells were transduced at an MOI of 0.3 and 24 hours after transduction cells were treated with 2 µg/mL puromycin for 48 hours. Puromycin-selected cells were divided into two populations and one population was induced with doxycycline (PLK1-IN) at day 0. Cells were passaged over 16 days with 200x hairpin representation, and samples were collected at day 0, 8, and 16. Genomic DNA was extracted for each timepoint and *sh*RNA sequences PCR-amplified using 98°C for 3 minutes, followed by 98°C for 10 seconds, 55°C for 15 seconds, 72°C for 15 seconds for 29 cycles. Amplified hairpins were XhoI digested, and the stable half hairpins gel extracted and purified and used for probe hybridization on UT-GMAP 1.0 microarrays. The signal intensities for the microarray were normalized with quantile normalization and *sh*RNAs with signal below the background (i.e. $\log_2$ scale of less than 8) at initial timepoint T0 and signals below 0 in timepoints T8 and T16 were removed prior to further analyses to compute the fitness score, The weighted differential cumulative change (WDC$_h$) for each *sh*RNA between the doxycycline induced cells (PLK1-IN) and its isogenic uninduced cells (PLK1-UN) was calculated for each consecutive timepoints using the formula:

$$WDC_h = \sum_{t=0}^{T} \frac{\varepsilon}{t+1}\left(x_{t+1,r}^{PLK1-IN} - x_{t,,r}^{PLK1-IN}\right) - \sum_{t=0}^{T} \frac{\varepsilon}{t+1}(x_{t+1,r}^{PLK1-UN} - x_{t,r}^{PLK1-UN})$$

where $x_{t,r}^{PLK1-IN}$ is the normalized signal intensity of the PLK1-IN cells at time point $t \in (0,..T)$ in replicates $r \in (1..N)$. Similarly, $x_{t,r}^{PLK1-UN}$ are for PLK1-UN cells. ε is a constant that determines the weight between each time point so that *sh*RNA drop at earlier time points are ranked before the *sh*RNA that drop at later time points. Gene level WDC$_{gene}$ was computed as the average of the top two *sh*RNA with most negative value for that gene using the formula below.

$$WDC_{gene} = average(arg_{h,h'}^{min}\left[WDC_{gene,h}\ ; WDC_{gene,h'}\right])$$

To identify *sh*RNA and their corresponding genes that are significantly different between the between PLK-IN and the PLK1-UN cells, Student t-test was used in combination with the permutation test p-value by estimating the frequency of randomized, shuffled WDC with more negative values in comparison with the observed gene level WDC value, as previously described (*50*). Bayesian analysis of gene essentiality algorithm was used to evaluate the performance of the screens (*51*).

Computational analysis and datasets
Reactome pathway enrichment was done by using the Reactome database to assign genes into pathways (https://reactome.org/) (*52*) and then running gene enrichment analysis on the query gene set of 960 PLK1 screen hits using idep software package (*53*). Gene expression analysis was performed using The Cancer Genome Atlas database downloaded from Genomic Data Commons data portal (https://portal.gdc.cancer.gov/) for 33 different cancer types. The RNASeq Expectation Maximization (RSEM) normalized mRNA expression data was transformed to $\log_2$ scale and the Spearman-rank correlation between PLK1 and every single of the 960 SDL hits was calculated in each of the 33 different malignant patient samples. The

resulting correlation coefficient (r) was then clustered by calculating the Euclidean distance between different genes.

The gene expression profile from the Cancer Cell Line Encyclopedia database (*54*) was used to rank cell lines by PLK1 expression to classify them as two groups of cell lines that over- or under-express PLK1. For the essentiality datasets from Project Achilles (*55*) as well as Marcotte et al.,(*56*), the top 25% and bottom 25% of the ranked cell lines were assigned as PLK1-overexpressing and under expressing respectively. To determine if a PLK1 screen hit was classified as essential or non-essential, the fold change of the essentiality scores between the PLK1-overexpressing and PLK1-underexpressing groups, as well as the p-value significance was calculated using the Mann-Witney U-test.

ToppGene candidate gene prioritization (*57*) was performed using "PLK1" as the training gene set and the list of 960 PLK1 screen hits as the test gene set. ToppGene compiles functional annotations and network analysis from gene ontology for molecular function, biological process, and cellular component, gene expression, pathways, protein domains, transcription factor binding sites, miRNA targets, drug-disease interactions, disease-drug interactions, and interactions published in NCBI search engine PubMed. Gene expression data from the TCGA was used to divide patients into two groups based on PLK1 expression and PLK1 screen hit expression for each of the 960 PLK1 screen hits. The two groups are illustrated in **Figure 2D**. The median gene expression value for PLK1 and for the screen hit was determined and used as the cut-off value to divide samples into PLK1 low and PLK1 high. TCGA clinical data was then used for standard Kaplan-Meier analysis between the patients exhibiting natural SDL and those who do not exhibit natural SDL.

Drug response analysis was performed by using the Genomics of Drug Sensitivity in Cancer (GDSC) dataset. GDSC contains the gene expression as well as the drug-response for various drugs treated in several cancer cell lines. We grouped the cell lines based on the expression of PLK1 as top 1% and bottom 1% cell lines. Then we computed the p-value significance using Mann-Whitney U-test of the IC50s of the inhibitors of SDL hit present in GDSC between the high and low PLK1 cells.

Pooled *in vivo* CRISPR screening

Mice were housed in sterile conditions at the University of Saskatchewan and all animal protocols were reviewed and approved by the University of Saskatchewan Animal Research Ethics Board. Digital caliper measurements were used to monitor tumor size and tumor volume was calculated by the formula A/2*B2 (where A and B are the long and short diameters respectively). The sgRNA sequences used in the imaging experiment were pooled and used to generate a pooled lentiviral library. HCI-010 and HCI-010+Cas9 PDX-derived cells (*58*) were transduced at an MOI of 0.3 and after 24 hours were selected using 2 µg/mL puromycin. After 48 hours of selection, 3 million viable cells mixed 1:1 with Matrigel (Corning) in a total volume of 100 µL were injected into the mammary fat pad of 6- to 8-week-old female mice and allowed to grow for 3 weeks. Tumors were harvested and flash frozen using liquid nitrogen. Genomic DNA was extracted using mortar and pestle and a DNA Blood Maxi kit (Qiagen). Sequencing library was constructed as described previously (*59*) with a small modification to primers of the first PCR to match the vector backbone (*60*) (Forward primer 5-caaaatacgtgac gtagaaagtaataatttcttgggtag-3' and reverse primer 5'-gcgtaaaattgacgcatgt gttttatcggtctgtatatcgag-3'). The fastq files were aligned to the sequence library using Bowtie2 alignment package. The alignment with scores above the default threshold score is mapped to the library. The count matrix data were analyzed as described in pooled screen analyses above.

Microscopy and imaging

HCI-010 and HCI-010+Cas9 were seeded in 96 well low attachment plates at a concentration of 10 000 cells per well with 8 µg/mL polybrene and 100 µL of sgRNA-lentiviral targeted pool (2 sgRNA for each target gene, one gene per well) in 200 µL total volume. After 24 hours, media was replaced with media containing 2 µg/mL. After 48-hour puromycin selection, 4 sites per well were imaged at 4x objective using brightfield setting for several days (Day0, Day4 and Day 6). After imaging, the images were analyzed using MetaXpress custom module where the cells were carefully identified by eliminating any artifacts. The custom algorithm was also used to measure the area, perimeter and number of cells per well. The increase in cell confluency each day was compared to baseline (Day 2) and then the HCI-010+Cas9 was compared to HCI-010 to determine if the gene knockouts caused a decrease in cell proliferation.

Direct-capture of guide RNA using Perturb-seq screen

10x Genomics Chromium Next GEM Single Cell 3'Reagent Kits compatible sgRNA library including 130 sgRNAs (targeting 65 genes), 15 non-targeting controls, 16 positive controls were generated by MilliporeSigma. Samples for Direct-capture Single-cell Perturb-seq were prepared by transducing MCF7-Cas9 cells with the single-cell perturb-seq sgRNA library Lentiviral particles and harvested 2 days after 48hrs of puromycin selection. 3'Gene expression library and CRISPR screening library were constructed followed by the instruction of Chromium Next GEM Single Cell 3'Reagent Kits v3.1 (10x Genomics). Both libraries were sequenced through NovaSeq S1 Flow Cell platform (Illumina). The raw data from the NovaSeq platform was extracted as fastq.gz files. These files were mapped to GRC38 reference genome and aligned using cell ranger (v6.1.2) count algorithm run on 32 core cluster computing. The samples were split into three batches with the cell expectancy estimated around 10,000 cells per batch. The analyses of three batches were pooled together by using the cell ranger aggregate algorithm. Upon further investigation, it was found that several cells had more than one sgRNA. Hence, cells with single sgRNA were curated and recomputed using cell ranger re-analyze pipeline. The resulting output was then imported into Loupe Browser (v.6.0) to generate various visualization outputs.

Colony formation

The cells ($1x10^3$/well for MCF, MDA-MB-231, $3x10^3$/well for HCC1143, $5x10^3$/well for BT-549) were seeded in 6-well plates and incubated at 37 °C under 5% CO_2 for 10 days, replacing the media every 3 day. The plates then were washed with PBS, fixed with 4% formaldehyde and stained with 0.5% crystal violet. Colonies were counted using ImageJ software, and images of colonies were scanned.

Genomic cleavage detection

Primers for genomic cleavage detection were designed for the region around the sgRNA target sequence with the forward primer ~200 bp upstream and the reverse primer ~300 bp downstream using Primer3 v. 0.4.0 (http://bioinfo.ut.ee/primer3-0.4.0/). The primer sequences are given in **Supp. Table S9.** Cells were lysed, sgRNA target region PCR amplified, and mismatches enzyme digested by T7 endonuclease using GeneArt Genomic Cleavage Detection kit as per manufacturer's directions (Invitrogen).

Droplet digital PCR

cDNA templates for droplet digital PCR were prepared by High-Capacity cDNA Reverse

Transcription Kit (Thermo Scientific). Each 20µL quantitative polymerase chain reaction system was performed using 10µL ddPCR™ Supermix for Probes (Bio-Rad Laboratories), 1µL target gene probe (*IGF2BP2*: FAM-MGB, Hs00538954_g1, *PLK1*: FAM-MGB, Hs00983227_m1, Thermo Scientific), 1µL reference probe (*GAPDH*: VIC-MGB, Hs02786624_g1, Thermo Scientific), 50ng cDNA templates. Droplets were generated from QX100 ddPCR Droplet generator (Bio-Rad Laboratories). Quantitative polymerase chain reaction was done by CFX96 Real-Time PCR Thermocycler (Bio-Rad Laboratories) using the following program: 95℃ 10 minutes, 45 cycles of: 94℃ 30 seconds, 60℃ 90 seconds, 98℃ 10 minutes after 45 cycles, 4℃ on hold. Signals were detected by QX100 ddPCR Droplet reader (Bio-Rad Laboratories) and absolute quantification analysis was performed by QuantaSoft software (Bio-Rad Laboratories).

Cell viability assay

The viability-inhibiting activity of chemical compounds was assessed using a resazurin reduction assay. MDA-MB-231, MCF7 and BT-549 cell lines were treated with increasing concentrations of the dimers (from 5 to 75 µM) and incubated for 72 h before resazurin (R&D Systems) was added. The fluorescence was detected by a Varioskan LUX Plate Reader (ThermoFisher). Each experiment was conducted thrice, each with six replicates. The dose–response curves were generated, and the 50% inhibitory concentrations (IC_{50}) were calculated using Prism GraphPad 9.

Proliferation assay in tumorspheres

Cells were seeded into 24-Ultra-Low attachment plates ($4x10^3$/well) in a complete Mammocult medium (STEMCELL, Canada) and allowed to propagate in tumourspheres for 7 days. For each replicate of IGF2BP2-KD, and corresponding control (shRFP), tumourspheres from 4 independent wells were combined and dissociated using mechanical dissociation (each sample was pipetted 10 times by P1000), and proliferation was assessed by cell counting using a hemocytometer. Each sample was assessed in triplicates and three independent experiments were performed. Pictures were taken using EVOS fl Microscope (Invitrogen, Carlsbad, CA, United States). Images were converted to 8-bit binary images and were analyzed using the ImageJ software.

Xenograft and pharmacokinetics studies

All animal experimental procedures were reviewed and approved by the University of Saskatchewan Animal Research Ethics Board. Mice used in the present study were from our established colony of immunodeficient female NOD.Cg-*Prkdc*[scid]*Il2rg* (NOD-SCID) mice at the Laboratory Animal Services Unit (LASU), University of Saskatchewan. Cells were trypsinized and re-suspended in ice-cold PBS. 2×10^6 MDA-MB-231 cells in a total volume of 100 µL were injected subcutaneously into the mammary fat pad of 6 to 8 weeks old NOD/SCID mice. Where indicated, treatment with IGF2BP2 inhibitor (Compound 4) was initiated the following day after injection of MDA-MB-231 cells. IGF2BP2 inhibitor dissolved in DMSO with excipient 2-hydroxypropyl-β-cyclodextrin, were administered intraperitoneally (IP) at the dosage of 10 mg/kg for each injection. Injections were given for five consecutive days, followed by two days of rest, for a total of 30 days. Tumors were measured every three to four days using a digital caliper, and the tumor volume was calculated using the formula

$(A*B^2)/2$, where A and B represent the long and the short diameter of the tumor, respectively.

For pharmacokinetic studies, outbred male CD-1 mice (Charles River, Germany), 4-weeks-old, were used. The animal studies were conducted in accordance with the recommendations of the European Community. All animal procedures were performed in strict accordance with the German regulations of the Society for Laboratory Animal Science (GV-SOLAS) and the European Health Law of the Federation of Laboratory Animal Science Associations (FELASA). Animals were excluded from further analysis if sacrifice was necessary according to the human endpoints established by the ethical board. All experiments were approved by the ethical board of the Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit, Oldenburg, Germany. The compounds C1, C4, C6 and C9 were dissolved in 1.6 % DMSO, 20 % PEG400, 20 % Tris 1 % pH 9.0 and 58.4 % 0.9 % isotonic NaCl-solution. Mice were administered C1, C4, C6 and C9 in a cassette PK format at 1 mg/kg IV per compound. About 20 µl of whole blood was collected serially from the lateral tail vein at time points 0.25, 0.5, 1, 2, 4, and 8 h post administration. After 24 h mice were euthanized, and blood was collected from the heart. Whole blood was collected into Eppendorf tubes coated with 0.5 M EDTA. One part was immediately spun down at 13.000 rpm for 10 min at 4°C. The plasma was transferred into a new Eppendorf tube and then stored at −80°C until analysis. Additionally, blood cells were preserved and used for analysis as well as whole blood. Furthermore, C4 and C6 were tested each in a single dose PK study at 10 mg/kg IP. C4 and C6 were dissolved in 10 % DMSO and 90 % Tris 1 % pH 9.0. About 20 µl of whole blood was collected serially from the lateral tail vein at time points 0.25, 0.5, 1, 6, 24 and 48 h post administration for C4 and at time points 0.25, 0.5, 1, 2, 4 and 8 h post administration for C6. After 72 h for C4 and after 24 h for C6 mice were euthanized and blood was collected from the heart as well as liver tissue Whole blood was collected into Eppendorf tubes coated with 0.5 M EDTA and immediately spun down at 13.000 rpm for 10 min at 4°C. The plasma was transferred into a new Eppendorf tube and then stored at −80°C until analysis. Moreover, spontaneous urine was collected for the cassette and the single dose studies. Liver samples were homogenized in isotonic sodium chloride solution using a Polytron® homogenizer.

All PK plasma samples were analyzed via HPLC-MS/MS using an Agilent 1290 Infinity II HPLC system and coupled to an AB Sciex QTrap6500plus mass spectrometer. First, a calibration curve was prepared by spiking different concentrations of C1, C4, C6 and C9 into mouse plasma, mouse whole blood, mouse blood cells and mouse urine from CD-1 mice. Caffeine was used as an internal standard. In addition, quality control samples (QCs) were prepared for C1, C4, C6 and C9 in the respective matrices. same extraction procedure was used: 7.5 µl of a plasma or whole blood sample or 5 µl blood cell sample + 7.5 µl isotonic sodium chloride solution or 10 µl urine sample (calibration samples, QCs or PK samples) was extracted with 37.5 µl of methanol containing 12.5 ng/ml of caffeine as internal standard for 10 min at 2000 rpm on an Eppendorf MixMate® vortex mixer. Then samples were spun down at 13.000 rpm for 10 min at 4 °C. Supernatants were transferred to standard HPLC-glass vials. 50 µl of a homogenized liver sample (adjusted to a final concentration 300 mg/ml; calibration samples, QCs or PK samples) was extracted with 50 µl methanol containing 12.5 ng/ml caffeine as internal standard for 10 min at 800 rpm on an Eppendorf MixMate® vortex mixer. Then samples were spun down at 4,000 rpm for 10 min at 4°C. Supernatants were transferred to 96well-V-bottom-Greiner plates and sealed. HPLC conditions were as follows: column: Agilent Zorbax Eclipse Plus C18, 50x2.1 mm, 1.8 µm; temperature: 30°C; injection volume: 1 µl; flow rate: 700 µl/min; solvent A: water + 0.1 % formic acid; solvent B: acetonitrile + 0.1 % formic acid; gradient: 99 % A at 0 min and until 0.1 min 99 % - 0% A from 0.1 min to 4.0 min, 0 % A until

4.5 min, 0 % - 99 % A from 4.5 min to 4.7 min; Mass spectrometric conditions were as follows: Scan type: MRM, positive and negative mode; Q1 and Q3 masses for caffeine, C1, C4, C6 and C9 can be found in **Table S10**; peak areas of each sample and of the corresponding internal standard were analyzed using MultiQuant 3.0 software (AB Sciex). PK parameters were determined using a non-compartmental analysis with PKSolver (*61*).

Analyses of PDX models

Tumor growth pattern of 174 PDX models from various tissue types such as breast, colorectal, non-small cell lung carcinoma, melanoma and pancreas that had match RNA-Seq data were taken from previously published studies (*62*). The samples of these 174 PDX models that were not exposed to any drug treatment were chosen for further analysis. The slope of the tumor growth was calculated using Xeva package (v.1.99.20) (*63*). The doubling time of the tumor was taken from the original publication (*62*). The grouping of the PDX models were based on their RNA-Seq expression in FPKM (Fragments per kilo base of transcript per million mapped fragments) units. FPKM measure of <20 was considered low expression and >30 was considered as high expression for both PLK1 and IGF2BP2, as most PDX samples had high expression of PLK1. Accordingly, the PDX models were grouped into two categories as PLK1 high expression and IGF2BP2 low expression (representing SDL) and both PLK1 and IGF2BP2 with high expression. Finally, the slope and doubling time of each group were plotted in violin plots and their significance was calculated using unpaired t-test with Welch's correction.

Results

Unbiased genome-wide screening captured novel SDL interactions of the polo-like kinase 1.

CIN is a hallmark of cancer (*64*) and PLK1 overexpression has been shown to induce this response (*21, 22*). To identify SDL interactions of PLK1 and exploit them for cancer therapeutics, we chose an inducible system (*65*) based on the overexpression of a constitutively active PLK1-S137D mutant in HCT116 colorectal cancer cells **(Figure 1B),** a model that has been described in our previous publication (*27*).The S137D mutant form of PLK1 has been previously shown to be defective in spindle assembly checkpoint (*66*), and checkpoint defects had been well linked to CIN (*14*). HCT116 also represents an ideal model, as it is a chromosomally stable, near-diploid, colon cancer cell line that has an intact DNA damage checkpoint (*67, 68*).

To identify gene knockdowns that cause lethality only when PLK1 is overexpressed, the HCT116-PLK1 cells were transduced with a lentiviral pooled library of 90,000 *sh*RNA sequences targeting ~18,000 different genes with ~5 independent hairpins per gene. Library transduction was performed at a scale of ~300-fold representation in two distinct populations (induced vs uninduced) as previously described (*50*) **(Figure 1C)** and this identified 960 hits **(Figure 1D)** (with at least 2-fold decrease and p<0.05). The complete list of 960 significant hits, their corresponding weighted differential cumulative change (WDC) scores, and biological functions are detailed in **Table S1,** and these genes are hereafter referred to as 'PLK1-SDL' hits. Although the replicates of the screen showed high correlation (Pearson correlation coefficient, r>0.9 between replicates) **(Figure S1A),** we compared our screen results to sets of essential and non-essential genes using a previously published framework to ensure our screening reliably identified true SDL hits (*69*). This approach measured a good performance by calculating the accuracy score from the precision and recall test (F-measures > 0.75) **(Figure 1E)**. The cumulative signal used to calculate the fitness score of all PLK1-SDL hit

genes displayed a significantly greater magnitude of 2.2-fold decrease in the induced population from T0 to T16 (Kolmogorov-Smirnov test p<0.0001) as opposed to uninduced population from T0 to T16 with 1.3-fold decrease (Kolmogorov-Smirnov test, p<0.0001) **(Figure 1F)**. Consistent with this, the overlay of all individual signals from all dropouts was found to be highly represented in the induced compared to the uninduced population (**Figure S1B)**. To illustrate the dropout of signal over time in the induced population alone, we present the analysis of one of the top hits, PPP2R5D, a regulatory subunit of the protein phosphatase 2A (PP2A) complex, that we previously reported to exhibit SDL with PLK1 (*27*) **(Figure 1G)**, reiterating the confidence in our screening approach. SDL interactions are functionally coherent (*44*) and as expected, comparison of our PLK1-SDL hits with previously published mitosis-related screens (*70-77*), found several hits to be associated with mitosis, DNA repair, and cell cycle-related pathways, apart from several completely novel PLK1 partners **(Figure 1H and Table S2)**. Consistent with these findings, Reactome pathway analyses indicated that PLK-SDL hits are enriched in cell cycle (adj. p 1e-03), mitosis and checkpoint related pathways (adj. p < 6e-03), and RNA metabolism pathways (adj. p 1e-03) **(Figure 1I)**. To gain insights into the functional relevance of SDL hits, a thorough literature analyses of the 960 genes were performed using Cytoscape STRING analyses. While this categorized the SDL hits into multiple biological functions including, cell cycle progression, centrosome amplification or cytokinetic components, the presence of extensive interactions among these components indicates that most of the SDL interactions are also functionally related **(Figure S1C; Table S1)**.

While the enrichment analyses increased confidence in our findings, we also used drug response data from Genomic of Drug Sensitivity in Cancer (GDSC) (https://www.cancerrxgene.org), as we found several of the SDL hits had chemical inhibitors. This revealed few inhibitors of SDL hits selectively suppressing PLK1 overexpressing cells **(Figure S2A)**. Thus, our screen identified GSK3A and its inhibitor CHIR-99021 to preferentially affect PLK1 overexpressing cells. Similarly, ALK receptor tyrosine kinase was found amongst the SDL hits and ALK inhibitor, NVP-TAE684, appeared to selectively target PLK1 overexpressing cells **(Figure S2A).** Given that ALK inhibition activates spindle assembly checkpoint causing mitotic delay (*78*), and that PLK1 overexpressing cells may be defective in spindle assembly checkpoint (*66*), we speculate that inhibition of ALK in PLK1 overexpressing cells may lead to mitotic catastrophe. These completely independent cross-validations by chemical-genetics, and testable hypothesis, confirm the quality of our screens and their potential to identify new potentially therapeutically relevant targets.

Systematic prioritization of PLK1-SDL hits for further validation studies.
Genome-wide screens tend to produce some false positive hits that can confound SDL identity. Most genome-wide studies cherry-pick one or two hits to serve as validated proofs of principles, but the bulk of data often remains unvalidated and therefore, underutilized by the research community. Obviously, validating all hits from a large-scale screen, as the one conducted here and pointing towards 960 SDL hits, is also practically challenging. Therefore, we applied three distinct strategies that allowed to prioritize a subset of PLK1-SDL hits for experimental validation, while also maintaining the strength of a large-scale unbiased screening approach.

Our first strategy to prioritize PLK1-SDL hits was based on the rationale that SDL genes that are differentially up-regulated, when PLK1 is overexpressed, represent co-regulatory mechanisms that become essential within the molecular context of elevated PLK1 levels (*79*). Therefore, we used patient data from The Cancer Genome Atlas (TCGA)

(https://portal.gdc.cancer.gov/) and asked how many of the PLK1-SDL hits are co-upregulated with PLK1, across 33 different cancer types. There, we found expression of a subset of PLK1-SDL hits to positively correlate with PLK1 levels across cancer types, suggesting that these hits may represent specific genetic dependencies of PLK1 overexpressing cells **(Figure 2A)**. After filtering out essential genes (*56, 80*) and genes that are likely not to be expressed in HCT116 cells (based on CCLE data (*54*); log2 expression score <5.0), we chose 20 genes whose expression strongly correlated with PLK1 levels (Spearman-rank correlation coefficient, r ≥ 0.4 in a majority of cancer types) **(Figure 2A).** In our second strategy, we investigated, how many of the 960 PLK1 SDL hits were 'pre-validated' in independently published essentiality screens (*55, 56*). This was done by classifying all the cell lines from these published studies into two groups, those that have high expression of PLK1 (within top 10%) and those with lower PLK1 expression (representing the lowest 10%), and then evaluating if our PLK1-SDL hits were more essential in the naturally PLK1-overexpressing cell line group. We found 30 SDL hits in Marcotte et. al. data **(Figure 2B)** and 19 hits in Project Achilles data **(Figure 2C)** that had higher essentiality scores in PLK1-overexpressing cells (Wilcoxon rank sum test, p<0.05) (*55, 56*)**.** Complete list with p-values is provided in **Table S3.**

Our third approach assessed SDL interactions based on patient prognoses associated with the expression level of the screen hits (*79*). Briefly, patient samples were queried for the 'naturally occurring SDL interactions', depending on expression levels of PLK1 and its SDL partner **(Figure 2D).** Genes for which naturally occurring SDL expression patterns positively correlated with significantly improved patient survival (37 genes; Kaplan-Meier log rank, p value <0.05), were selected for further validation **(Figures 2D and S2B; Table S4)**. Apart from the above three approaches, we also included 21 genes from the chromosome 19 short arm (19p13.2-3), as our screen picked many genes from this locus **(Table S5)**. Given that this region has been associated with macrocephaly (*81*) and that PLK1 and its centrosome functions have been previously linked to both micro- and macrocephaly (*82, 83*), we expected to see a functional crosstalk between PLK1 and this vulnerable locus. We also included 25 functionally relevant PLK1-SDL hits, as determined through gene prioritization algorithms (*57*) **(Table S5),** and finally, 30 potential cell surface SDL targets since they may represent targetable vulnerabilities for advanced strategies such as antibody-based inhibitors **(Table S5)**. Overall, after removing hits overlapping among these approaches, we selected 134 genes for further studies. Of these, 29 genes failed guide RNA cloning and therefore, we focused our efforts on validating the remaining 105 genes **(Figure 2E).**

Combination of in vivo ***pooled CRISPR and*** in vitro ***arrayed CRISPR screens in a PLK1-overexpressing breast cancer PDX model.***
Following the prioritization of 105 candidate SDL hits, we made a systematic effort to identify the best SDL hit as a potential anti-tumor targets. To this end, we selected 210 sgRNA sequences for 105 genes and queried the essentiality of the prioritized PLK1-SDL hits in a previously described breast cancer PDX model, HCI-010, that was found to closely mimic tumor properties (*58*). The derived cells display PLK1 overexpression compared to a non-malignant breast cell line Hs578Bst **(Figure 3A).** We engineered HCI-010 cells to stably express Cas9 and used them in a CRISPR-based arrayed screen following lentiviral transduction and high-throughput imaging **(Figure 3B)**. To evaluate each SDL hit, HCI-010 cells with and without Cas9 expression **(Figure 3B)** were transduced with pooled lentiviral particles expressing two independent sgRNAs in a single well. The cells were monitored over time to determine the viability of PLK1 overexpressing HCI-010 after gene knockout by automated image analyses **(Figure 3C)**. The confluency of each well was calculated using

Molecular Devices MetaXpress image analysis software and used as a proxy measurement for cell viability. In total, 60 of the 105 genes queried reduced confluency by ~40% or more (Student t-test, p<0.0.5) **(Figures 3D, FS3A and Table S6)**.

In parallel, we also generated a pooled lentiviral library containing all 210 sgRNA sequences and transduced HCI-010 PDX cells with and without Cas9 in a manner mimicking the genome-wide pooled screen **(Figure 3E)**. Following puromycin selection, the sgRNA transduced HCI-010 cells, were introduced into the mammary fat pad regions of female mice with over 4000-fold representation per sgRNA (3 million cells injected containing 210 library sgRNAs). Once the tumors were grown for three weeks they were harvested, and the genomic DNA extracted. The sgRNA sequences were PCR amplified with Illumina adapter sequences and sequenced to determine the dropout of sgRNAs from the HCI-010 Cas9-expressing samples, while HCI-010 tumors without Cas9-expression served as a baseline control. Replicates of the screen showed high correlation (Pearson correlation coefficient, r > 0.9) **(Figure S3B)** between seven tumors in each condition. This *in vivo* pooled CRISPR screen identified 15 SDL hits (p<0.05) **(Figure 3F and Table S7)**. Overall, by these two complementing approaches, we validated 65 PLK1-SDL hits in the HCI-010 PDX model, with 10 SDL hits overlapping in *in vivo* and *in vitro* experiments **(Figure 3G)**. Cas9 genome editing was confirmed in these experiments to eliminate off target effects by using a cleavage detection assay **(Figure S3C)**.

Shortlisting top candidates using Perturb-seq with direct guide RNA capture.

PLK1 overexpression is known to induce CIN and heterogeneity of tumor cells (*21, 22, 84, 85*). Thus, overexpression of PLK1 can be considered as a common feature of most cancer cells, beyond sub-clonal intra-tumor heterogeneity. Therefore, we next asked which of the 65 SDL hits, eliminated most of the PLK1 overexpressing single cells. The direct-capture Perturb-seq methodology not only queries the changes within the transcriptome at single cell level (*49*), but can also be used to evaluate the survival of individual CRISPR knock-outs at single cell level, akin to our negative selection genome-wide screen **(Figure 1C and 1D)**. To take advantage of this approach, we constructed a new sgRNA library, targeting the 65 experimentally validated genes, into a Chromium Single Cell Gene Expression v3.1 (Next GEM) compatible LV13 vector backbone (U6-gRNA:EF1a-Puro-2a-BFP, MilliporeSigma) with 10x Genomics compatible capture sequence 1 (CS1: 5'-TTGCTAGGACCGGCCT TAAAGC-3') at the stem-loop position. In this system, the functionally expressed sgRNA incorporates a capture sequence directly into the guide scaffold to allow a direct capture of sgRNAs during single-cell RNAseq (*49*). Sixteen guide sequences targeting housekeeping genes such as RPL3, RPL11, RPL13 and RPL18 were used as positive controls, as described previously (*69*) and 15 non-targeting guide sequences were used as negative controls **(Table S8)**. This sub library of 161 guides targeting 65 genes and the relevant controls were used to transduce innately PLK1 overexpressing MCF7 breast cancer cells (*86*) **(Figure 4A)**.

Following transduction and selection, pooled knockout cells were cultured for four days and single cell gel beads in emulsion (GEMs) with barcoded polyadenylated mRNA primers and sgRNA capture sequences were generated to read through the single cell transcriptomes and CRISPR libraries construction. After sequencing and standard data processing using 10x Genomics cell ranger (ver. 6.1.2) software package, the UMI matrix of cells from each replicate with single guides were extracted and subjected to quality control and normalization, and then, imported into Loupe browser (ver. 6.0) for downstream analysis and t-sne plot generation **(Figure 4B)**. We sequenced 22,041 single cells that yielded an average of 51,479 reads per cell with a median of ~4000 genes expressed per cell. Individual cells that had less than 5000

transcripts (<5000 UMI count per cell) or mapped to 0 or ≥ 2 guides in a single cell were eliminated. Overall, we considered 7434 single cells with at least 50 representative cells for most of the knockouts, when we mapped a single guide per cell **(Figure S4A).** The efficiency of each knockout was confirmed by comparing the expression of the corresponding knockout gene with negative controls **(Figure S4B).** Expression of six genes (LPL, CNN1, C19ORF35, CLPS, ATP4A and TMC) did not achieve a significant reduction and were excluded from further analyses **(Figure S4B).** The resulting t-SNE plot from the correlation of transcriptomic data across 65 gene knockouts outlined in 10 clusters **(Figure 4B)** with clusters 1 and 4 enriched for knockouts of eight genes (*TJP3, IGF2BP2, CCDC18, RCOR2, DOCK6, GOLM1, PPP2R5D* and *CLK2*), including positive controls **(Figure 4C)**. As the housekeeping positive controls are expected to cause cell death, co-enrichment of these knockouts in clusters 1 and 4 with the positive controls indicated that the loss of functions of these genes aligns with the expected SDL phenotype. Moreover, enrichment for the knockout of PPP2R5D within these cluster also increases the confidence in our approach, as we previously reported PPP2R5D to exhibit SDL with PLK1 (*27*). Interestingly, cluster 2 was enriched for not only PLK1-overexpressing cells, but also for Aurora Kinases-overexpressing cells **(Figure 4D)** and therefore, we mapped individual gene knockouts separately to determine their relation to cell populations with different PLK1 levels **(Figure 4E).** We found knockouts of four genes (*IGF2BP2, CRB3, DPP9 and TJP3*) that displayed a negative selection with decreased number of PLK1 overexpressing cells or increased lethality to cells with higher PLK1 levels **(Figure 4E)**. Overall, these results indicate that Perturb-seq screen with direct guide RNA capture, efficiently shortlisted few potential candidates for further analyses.

Loss of IGF2BP2 affects PLK1 expression and suppresses PLK1 overexpressing cells, tumorspheres and tumors.

As PLK1 overexpressing cancer cells may become addicted to high PLK1 activity, we tested, if the loss of IGF2BP2, *CRB3, DPP9 or TJP3* affects PLK1 mRNA levels. While CRB3, DPP9 and TJP3 knockouts did not affect PLK1 expression, we found the expression of PLK1 to be decreased in cells with IGF2BP2 knockout **(Figure 4F)**. To confirm this, we performed digital droplet PCR to get an absolute quantitation of both IGF2BP2 and PLK1 expression, after two days of IGF2BP2 knockdown. We found PLK1 expression to be significantly downregulated in response to IGF2BP2 silencing in PLK1 overexpressing MCF7 ($p<0.001$) and BT549 cells ($p<0.05$) **(Figures 5A and S5A, S5B and S5C)**. In contrast, PLK1 transcript levels increased in MDA-MB-231 cells ($P<0.001$), when IGF2BP2 was knocked down **(Figure 5A)**. Thus, it appears that loss of IGF2BP2 can regulate PLK1 mRNA differently in distinct biological contexts. Nevertheless, we observed a decrease in PLK1 protein levels **(Figure 5B)** following IGF2BP2 silencing in all tested cell lines, including MDA-MB-231. To determine if the functional relation between PLK1 and IGF2BP2 is restricted to specific cell line models, we monitored it at the protein level and also followed the correlation between PLK1 and IGF2BP2 mRNA levels in multiple tumor types using the cancer genome atlas (TCGA) data. We found a strong positive correlation between these two genes (Spearman-rank correlation coefficient, $r \geq 0.3$) in multiple cancers (**Figure 5C**). This suggested that the SDL interaction between PLK1 and IGF2BP2 is based on the reduction in PLK1 activity in the absence of IGF2BP2, which preferentially affects PLK1 overexpressing cells that develop addiction to high levels of this kinase. It is important to note that, unlike direct PLK1 inhibition that leads to the complete disruption of PLK1 function and therefore, affects normal cells, downregulation of PLK1 by targeting IGF2BP2 still retains PLK1 protein levels to a minimal amount, possibly without abrogating its functions in normal cells. To assess this, we confirmed that the loss of IGF2BP2 does not affect non-malignant

Hs578Bst cells that do not overexpress PLK1 **(Figures 3A and 5D)**. As Hs578Bst cells did not grow efficiently with lentiviral knockdown, we used siRNA to transfect these cells. The cells were transfected with siRNA and confluency was monitored using live-cell imaging S3-Incucyte®. Knockdown of IGF2BP2 did not affect cell growth compared to the matching control cells, when monitored over seven days **(Figure 5D)**. Using the same live cell imaging technique, we also used the HCT116-PLK1 inducible model cell line from the genome-wide screen and found that the loss of IGF2BP2 caused a decrease in viability only in the PLK1-induced condition over a period of 10 days, which matched the results of the genome-wide screen **(Figure 5E)**. Furthermore, knockdown of IGF2BP2 also produced a significant decrease in colony formation compared to the control transduced cells in the panel of breast cancer cell lines, including MDA-MB-231 that showed an increase in PLK1 mRNA and a decrease in PLK1 protein abundance upon IGF2BP2 knockdown **(Figures 5F and S5D)**. Finally, to investigate how knockdown of IGF2BP2 affects cancer stem cells, we performed tumorsphere analysis in selected breast cancer cell lines **(Figure 5G)**. Tumorsphere models were also employed, as they better simulate tumor biology than cells cultivated in monolayers and knockdown of IGF2BP2 effectively suppressed growth of breast cancer cells in tumorspheres. Following this finding, we next examined if IGF2BP2 loss reduces tumor growth in xenograft models. To test the ability of IGF2BP2 knockdown to reduce tumor size *in vivo*, we used individual shRNAs to silence IGF2BP2 in MDA-MB-231 cells and injected the cells into immunodeficient female NOD-SCID mice. Consistent with the effects observed in tumorsphere models, silencing of IGF2BP2 reduced growth of PLK1-overexpressing MDA-MB-231 tumors, representing a model of human triple-negative breast cancer **(Figure 5H)**.

Pharmacological inhibition of IGF2BP2 affects PLK1 expression and decreases expansion of PLK1 overexpressing cells and tumors.

Our team has recently described the first ever reported small molecule inhibitors of IGF2BP2 (*87*). Taking advantage of our previous work, we tested four of these new compounds **(Figure 6A)** and similar to IGF2BP2 silencing, they all showed a consistent reduction in PLK1 levels but not a complete elimination of this protein in cancer cells **(Figure 6B).** We also generated dose-response curves for these compounds in multiple PLK1-overexpressing breast cancer cell lines, to determine their IC_{50}. While compound C4 and C9 inhibited more effectively at lower concentrations **(Figure 6C),** compound C4 showed the best *in vivo* pharmacokinetic (PK) profile of the tested IGFBP2 inhibitors. It exhibited the highest exposures in plasma, whole blood, blood cells, and urine after intravenous application in mice and a half-life of around 22 hours. By contrast, C6 and C9 had a relatively low half-life of only 1 hour, whereas C1 had a half-life of around 9 hours. Moreover, C4 had a high C0 at approximately 1.5 µg/ml in plasma after a relatively low dose of 1 mg/kg IV. Additionally, C4 and C1 exhibited a low volume of distribution (~ 0.5 l/kg and 0.7 l/kg, respectively) and a low clearance of ~ 0.3 ml/min/kg and 0.9 ml/min/kg, respectively. Compounds C6 and C9 had a moderate to low clearance with 25 and 10 ml/min/kg, respectively **(Figure 6D)**.

As C4 exhibited the best PK parameters out of the four compounds tested in the cassette PK study, we explored the intraperitoneal route at 10 mg/kg to enable an administration over several days. Moreover, we tested C6 as it had a higher volume of distribution of around 2 l/kg after IV administration and we wanted to probe terminal compound levels in liver as well **(Figure 6E)**. C4 showed sustained plasma levels above the IC_{50} (range of 47-82 nM in different cell lines) with a C_{max} of 7.1 µg/ml and a mean residence time of around 15 hours. No accumulation was observed as plasma levels decreased until 72 hours. This suggests that a dosing interval > 24 hours could be used for efficacy studies for C4. Compound

C4 was also found in urine up to 24 hours. Compound C6 had a C_{max} of around 1.5 µg/ml with a T_{max} of 0.4 h and a mean residence time of around 1.9 h. Compound C6 had terminal liver levels of around 32 ng/g tissue after 24 hours, whereas compound C4 still exhibited terminal liver levels of around 8.4 ng/g after 72 hours. Moreover, C4 had a bioavailability of around 32 %, whereas C6 had a bioavailability of around 38 % after IP administration. Based on these PK features, we chose C4 for further studies.

Using C4, we next assessed PLK1 and IGF2BP2 protein levels after 72h treatments with compound C4 at doses ranging from 5 µM to 75 µM. PLK1 protein levels consistently decreased after treatment with this inhibitor, although IGF2BP2 protein levels remained unaltered **(Figure 6F).** Finally, the PARP cleavage experiment was used to determine if apoptotic pathways were activated in cells following treatment with C4. Indeed, we observed cleaved PARP in all tested cancer cells treated with higher doses of the compound **(Figure 6F).** To further evaluate the therapeutic potential of IGF2BP2 inhibition, we examined if treatment with C4 will reduce tumor development in xenograft models. These experiments revealed that C4 administration suppressed tumor growth in a triple-negative breast cancer xenografts generated with MDA-MB-231 cells **(Figure 6G).** Overall, these studies suggests that further optimization of IGF2BP2 inhibitors holds high therapeutic potential, as application of these compounds would allow to effectively tune down PLK1 protein levels and ultimately, inhibit PLK1-overexpressing cancer cells and tumors. To derive additional support for our findings, we took advantage of multiple previously described PDX models, representing different tissue types (*62, 63*). Briefly, we used tumor growth pattern of 174 PDX models that were not exposed to any drug treatment and classified them into two groups based on the expression pattern of PLK1 and IGF2BP2. The first group represents high PLK1 expression, but low in IGF2BP2, while the second group represents PDX tumors with both PLK1 and IGF2BP2 expression high. Analyzing doubling time and the slope of tumor growth, we found PDX models of Group 1 to have significantly higher doubling time and significantly decreased growth slope compared to Group 2 (p<0.05, unpaired t-test with Welch's correction) **(Figure 6H)**. These results revealed that in agreement with our observations in cultured cancer cells, IGF2BP2 also displays an SDL interaction and selectively suppresses growth of PLK1 overexpressing tumors in PDX models**.** In summary, our work indicates that PLK1 levels can be modulated by IGF2BP2, and its loss can decrease growth of tumors with elevated PLK1 expression.

Discussion

Multiple PLK1-targeting compounds have been identified and are currently being assessed in clinical trials (*88*). The most notable trials have been with BI2536, BI6727 (Volasertib), and GSK461364A, which are all ATP-competitive inhibitors of PLK1. However, monotherapy with BI2536 had to be terminated because of a low objective response rate and poor survival (*89*) and GSK461364 was associated with a high incidence of venous thrombotic emboli in clinical studies (*90*). Volasterib, a derivative of BI2536, had initial success, gaining FDA Breakthrough Therapy status, but has not shown significant promising results since (*91*). New inhibitors of PLK1, such as TAK960 and NMS-P937, are still at early stages (*88*). These attempts and not yet encouraging results indicate the importance of PLK1 targeting and highlight the challenges associated with its direct inhibition.

Here, we queried nearly the entire genome to identify genes that are selectively essential for the survival of PLK1 overexpressing cells. Our main goal is to take an unbiased systematic approach to pan the optimal target for future therapeutic approaches. Ideally, from an unbiased perspective, all PLK1 SDL hits identified in our initial genome-wide screen should be validated,

however testing 960 genes is practically not feasible. Therefore, several independent computational strategies were utilized to narrow down the most promising candidates from the list of the screening hits. For example, we shortlisted several candidates by using previously published genome-wide screens performed across multiple cell lines. Similarly, by using RNAseq data from TCGA, we shortlisted candidates associated with the improved patient survival in multiple cancer types. From a technical standpoint, we are also cognizant that genome editing tools have their shortcomings associated with concerns of off-target effects (*92, 93*). Our combined use of orthogonal *sh*RNAs and *sg*RNAs reagents in validation approaches has mitigated potential false positives associated with our screening and validation strategies. Finally, a variety of PLK1 overexpressing models were employed, as relying on any single model of cancer in a laboratory setting is always associated with a set of limitations. PDX-based models better capture the heterogeneity of real patient tumors, representing improved alternative to cell lines (*94*). Therefore, we tested out hits in a compendium of cancer cell lines apart from the PDX model. By using complimentary and independent validations with different gene-silencing vs gene-editing techniques and model systems, confounding effects were filtered out. We also demonstrate optimal strategies to apply pooled *in vivo* CRISPR screens, by initially prioritizing our hits. Prioritizing candidates can reduce libraries to a manageable size for both *in vitro* and *in vivo* experimental work, while achieving meaningful results, with a sufficient representation of guide RNAs.

Of the ten overlapping genes from the *in vivo* and *in vitro* PDX-based validation, a several have a role in cancer progression and regulating stemness. For example, in addition to its role in centrosome clustering and cytokinesis, KIF5B has been shown to regulate EMT and stemness of cancer cells (*95*). Similarly, FDPS has a role in maintaining tumor initiating cells (TICs) in glioma (*96*) and IGF2BP2 in maintaining TIC populations in both colon cancer and glioma (*97, 98*). To further shortlist the best candidate, in the context of PLK1 overexpression, we chose to use a Single-cell CRISPR screen (Perturb-seq) by direct guide RNA capture. The key reason for this is that PLK1 overexpression can induce CIN and the resulting intra-tumor heterogeneity. Hence, it is imperative to query the efficiency of the SDL target at a single cell level. This led eventually to the shortlisting of four best candidates. Among these, IGF2BP2 is an N6-methyladenosine (m6A) reader (monitoring the presence of a methyl group at the *N*6 position of adenines in poly(A)$^+$ RNA) that has been shown to bind m6A methylated mRNA molecules to enhance their stability and translation (*99*). As the mRNA of PLK1 has been previously reported to be m6A methylated (*100*), we speculate that one of the potential mechanisms by which IGF2BP2 can regulate PLK1 expression could be via the stabilization of the m6A methylated mRNA of PLK1 **(Figure 6I).** Consistent with this notion, while our work was in progress, a recent study showed that IGF2BP2 binds to m6A of PLK1 3' untranslated region and is involved in stabilizing PLK1 expression (*101*)**.** Interestingly, the loss of IGF2BP2 has not only affected PLK1 mRNA, but also reduced the abundance of the PLK1 protein in multiple models, including MDA-MB-231 breast cancer cells, where its inactivation does not decrease the abundance of PLK1 mRNA. Taken together, these data suggest that IGF2BP2 operates in cancer cells in a context-dependent manner, but its SDL relation with PLK1 universally relies on PLK1 suppression. This PLK1 inactivation causes eventually a selective elimination of PLK1-overexpressing cancer cells that developed dependence on elevated PLK1 activity.

It is interesting to note that both PLK1 and Aurora Kinases are overexpressed within the same single cells and loss of IGF2BP2 appears to eliminate most of these single cells. Aurora kinases represent a family of therapeutic targets, as they are highly overexpressed in cancer cells and functionally linked to PLK1 (*24*). Although there are over 20 Aurora kinase inhibitors

in different stages of clinical development, prolonged treatment of patients has been shown to develop drug resistance (*102*). From this perspective, targeting IGF2BP2 should provide a potential opportunity to eliminate these cells that co-express both PLK1 and Aurora kinases. It is also important to note that while the concentration of C4 used to inhibit IGF2BP2 is relatively high, further development of this molecule or strategies to define combination therapies should potentially lead to increased efficacy at the nanomolar range in combination with good pharmacokinetic properties. Additionally, the PK results for C4 at 10 mg/kg IP suggest that already with this first hit compound sufficient concentrations in plasma could be reached so that efficacy studies should be envisaged.

While we have focused on IGF2BP2, further exploration of the other three candidates is much needed to properly dissect their SDL relationships with PLK1. For example, CRB3 and TJP3 are involved in the establishment of tight junctions (*103, 104*) and may impact cytokinesis, as PLK1 also plays a key role in this process (*105*). In fact, among the pleiotropic defects caused by PLK1 overexpression (*20, 23-26*), failures in cytokinesis and abscission are reported to be the most abundant defects, resulting in aneuploidy (*22*). Further studies to explore the role of these proteins might also point towards new therapeutic venues. DPP9 on the other hand, is a peptidase that degrades most cytosolic proline containing peptides (*106*), has been shown to interact specifically with SUMO1 (*107*) and this may facilitate mitotic entry by affecting ubiquitination of the transcription factor Forkhead box protein M1b (FoxM1b), a known substrate of PLK1 (*108*). Thus, our work has led to several testable models towards discovery science, and we expect the research community to benefit from our extensive validation strategies.

Our systematic integration of multiple unbiased platforms identified a large resource of PLK1-SDL hits that can be targeted to suppress the proliferation of PLK1 overexpressing cells. As PLK1 overexpression can lead to cellular heterogeneity, there is also a need to confirm if suppression of its SDL partner(s) truly eliminates PLK1 overexpressing individual cells. Application of direct-capture Perturb-seq identified IGF2BP2, loss of which affects PLK1 protein levels, either genetically or pharmacologically. In this regard, we are actively working on optimizing our novel IGF2BP2 inhibitors to use them effectively in pre-clinical studies with an ultimate goal to progress them into clinical trials.

Abbreviations
ATCC   - American Type Culture Collection
BCA   - Bicinchoninic Acid
CIN   - Chromosomal Instability
GDSC   - Genomics of Drug Sensitivity in Cancer
GEM   - Gel bead Emulsion
GEO   - Gene Expression Omnibus
PDX   - Patient-Derived Xenograft
RSEM   - RNASeq Expectation Maximization
SDL   - Synthetic Dosage Lethality
TCGA   - The Cancer Genome Atlas
TIC   - Tumor Initiation Cells
WDC   - Weighted Differential Cumulative change

Declarations
Ethical Approval and Consent to participate
The studies were approved by the Animal Ethics Research Board of University of Saskatchewan and appropriate regulatory authorities (# 20150067) and the ethical board of

the Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit, Oldenburg, Germany.

Consent for publication
Not applicable

Availability of supporting data
The data files are uploaded in the GEO datasets (Gene Expression Omnibus) and its SuperSeries accession number is GSE203242. This reference consists of two SubSeries for both Microarray raw data (.CEL files) for the *sh*RNA screens GSE203237 and the deep sequencing data of CRISPR screens (.fastq.gz files) GSE203240.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
Conceptualization: F.S.V., M.E., A.K.K., A.F., and F.J.V.; Methodology: C.E.C., F.S.V., Y.Z., P.G., V,M., S.B., K. W., Y.W1., B. G. E. Z., K.R., K.K.B., H.E., L.K., H.P., R.D., O.A., P.W., H.D., T.F., E.P.M., J.S.L., and B.T., Writing, Review & Editing: C.E.C., F.S.V., Y.Z., V.M., P.W., H.D., B.T.,A.K., E.R., R.D., Y.W2., M.E., A.K.K., A.F. and F.J.V.; Funding Acquisition: A.F., A.K.K., M.E., and F.J.V.

Authors' information
Division of Oncology, Cancer Cluster, College of Medicine, University of Saskatchewan, Saskatoon, S7N 5E5 Canada.
Chelsea E. Cunningham, Frederick S. Vizeacoumar, Yue Zhang, Vincent Maranda, Ashtalakshmi Ganapathysamy, He Dong, Jared Price, Tetiana Katrii, Li Hui, Karthic Rajamanickam, Kalpana K. Bhanumathy, Hussain Elhasasna, Liliia Kyrylenko, Hardikkumar Patel, Prachi Walke, He Dong, Tanya Freywald, Erika Prando Munhoz, Sunil Yadav and Franco J. Vizeacoumar

Department of Pathology and Laboratory Medicine, University of Saskatchewan, Saskatoon, SK S7N 5E5, Canada.
Frederick S. Vizeacoumar, Vincent Maranda, Hardikkumar Patel, Jared Price, Tetiana Katrii,

Renuka Dahiya, Alain Morejon Morales and Andrew Freywald

Global Institute for Food Security, University of Saskatchewan, Saskatoon, SK, S7N 4L8, Canada.
Peng Gao and Raju Datla

Agriculture and Agri-Food Canada, Saskatoon Research and Development Centre, 107 Science Place, Saskatoon, SK, S7N 0X2, Canada.
Peng Gao

Department of Pharmacy, Pharmaceutical Biology, Saarland University, Saarbrücken 66123, Germany.
Simon Both and Alexandra K. Kiemer

Antiviral & Antivirulence Drugs (AVID), Helmholtz Institute for Pharmaceutical Research, Saarland (HIPS), Helmholtz Centre for Infection Research (HZI) and Department of Pharmacy, Saarland University, Saarbrücken 66123, Germany.
Konrad Wagner, Yingwen Wu, Ben G. E. Zoller and Martin Empting

College of Pharmacy, University of Saskatchewan, 104 Clinic Place, Saskatoon, S7N 2Z4, Canada.
Omar Abuhussein

Department of Anatomy, Physiology, and Pharmacology, University of Saskatchewan, and Cameco MS Neuroscience Research Centre, 701 Queen St., Saskatoon, S7K 0M7, SK, Canada.
Prachi Walke and Anand Krishnan

Cancer Data Science Laboratory, National Cancer Institute, National Institute of Health, Bethesda, MD 20892, USA; Center for Bioinformatics and Computational Biology & Department of Computer Sciences, University of Maryland, College Park, MD 20742, USA.
Eytan Ruppin and Joo Sang Lee

Department of Chemical Biology (CBIO), Helmoltz Centre for Infection Research (HZI), Inhoffenstrasse 7, 38124 Braunschweig, Germany.
Katharina Rox
Western College of Veterinary Medicine, University of Saskatchewan, Room 2343, 52 Campus Drive, Saskatoon, S7N 5B4 Canada.
Behzad Toosi

Department of Precision Medicine, School of Medicine & Department of Artificial Intelligence, Sungkyunkwan University, Suwon, 16419, Republic of Korea.
Joo Sang Lee

Department of Biochemistry, Microbiology and Immunology, University of Saskatchewan, Saskatoon, SK S7N 5E5, Canada.
Yuliang Wu

Cancer Research, Saskatchewan Cancer Agency, 107 Wiggins Road, Saskatoon, S7N 5E5, Canada
Franco J. Vizeacoumar

Figure Legends
Figure 1: Genome-wide *sh*RNA screening identifies synthetic dosage lethal partners of PLK1.

A. Schematic illustrating the expected outcome in PLK1-overexpressing cancer cells after inhibition of PLK1 or of PLK1 SDL target. PLK1 inhibition is expected to cause aneuploidy and potentially an increase in cell death. PLK1-SDL gene inhibition is expected to cause only cell death. **B.** Western blot analysis of the HCT116-PLK1 inducible cell line for PLK1 with and without induction with doxycycline. The upper band represents the constitutive phosphorylated form of PLK1-S137D mutant form. GAPDH is shown as a total protein loading control. **C.** Schematic illustration of the genome-wide screening workflow. Example microarray signal outcomes for non-essential (NE, green and yellow), synthetic dosage lethal (SDL, blue), and essential (E, red) genes are shown. **D.** Volcano plot representing results of the genome-wide pooled *sh*RNA screen. Negative genetic interactions, or genes that significantly ($p<0.05$ and WDC score < 2-fold) decreased the fitness scores in the PLK1 overexpressing population over the PLK1 non-overexpressing, are indicated in red. Positive genetic interactions are in blue. The total number significant hits came to 960 genes and the full list is provided in Table S1. **E.** Precision Vs Recall (PR) curve calculated by measuring the Bayes factor for the genes previously described as general essential and non-essential genes from published screens. F-measure > 0.75. **F.** Box plots summarizing microarray signals for all queried hairpin barcodes at different timepoints in PLK1-untreated (no induction) and PLK1-induced conditions. A decrease of 2.2-fold change in the induced population from T0 to T16 was observed (Kolmogorov-Smirnov test $p<0.0001$). **G.** Drop plots of microarray signal for best two individual hairpins targeting PPP2R5D gene over time in PLK1-uninduced and PLK1-induced. **H.** SDL-hits that were also picked up from published mitosis-related screens. Only the top 45 genes with maximum overlaps are shown. Rows and columns are sorted by total number of overlaps in descending order. References for each row are: 1:PMID: 20360068, 2:PMID: 15616564, 3:PMID: 20360735, 4, 5, 6:PMID: 24104479 in MUS81, BLM, and PTTG1 null cells respectively, 7:PMID: 27929715 in U2OS cells, 8:PMID: 14654843, 9:PMID: 27929715 in RPE1-hTERT cells, 10:PMID: 24104479 in PTEN null cells, 11:PMID: 16564017, and 12:PMID: 17001007 **I.** Dendrogram of the Reactome pathways that are significantly enriched for the 960 PLK1-SDL candidate genes. FDR adjusted p-values are indicated in the Figure.

Figure 2: Prioritization of the PLK1-SDL candidate genes using computational analyses.
A. Hierarchical clustering of the correlation for the expression of PLK1 with the expression of each SDL hit across 33 cancer types from The Cancer Genome Atlas (TCGA) patient data obtained from Genomic Data Commons (GDC). **B and C.** Violin plots of the difference in essentiality scores for PLK1 SDL hits in different cell lines grouped by low versus high PLK1 expression. Essentiality scores derived from published *sh*RNA screens data Marcotte et al., (Highlighted in green) and Project Achilles database (https://depmap.org/portal/achilles/) (highlighted in orange). The p-value significance was calculated using non-parametric Wilcoxon Rank sum test. **D.** Schematic of identifying naturally occurring SDL interactions using Kaplan-Meier survival curves in breast cancer (BRCA). Log rank p-values were computed to calculate significance. **E.** Schematic summarizing gene prioritization for experimental validation.

Figure 3: CRISPR-knockout validation of PLK1-SDL candidate genes in patient-derived breast cancer model.
A. Western blot analysis of PLK1 and Cas9 in HCI-010, HCI-010 stably transduced with a Cas9 expression vector, and in Hs578Bst non-malignant breast epithelial cells. GAPDH is shown as the total protein loading control. **B.** Schematic of the methodology for the CRISPR arrayed in vitro screening validation in PDX breast cancer cells with and without Cas9 using high-

throughput imaging automation to determine lethality. **C.** Display of sample images acquired using automated imaging over time and analysis for different PLK1-SDL candidate gene sgRNA in -Cas9 and +Cas9 HCI-010 cells. MetaXpress object masking overlay is shown in blue for the day 6 images. **D.** Bar graph quantification of the imaging analysis over the course of 4 days. n = 3, * p-value < 0.05, student's t-test. **E.** Schematic of the *in vivo* pooled CRISPR screening in a PDX breast cancer model with and without Cas9. Sequencing was done to capture *sg*RNA dropout. **F.** Volcano plot of log10 p-value versus WDC fitness score for all queried genes from the *in vivo* pooled CRISPR screen. P-value cut-off < 0.05 based on WDC permutation shuffling. Significant essential genes are colored red. **G.** The final list of SDL hits identified from the *in vitro* arrayed CRISPR screen and *in vivo* pooled CRISPR screen.

Figure 4: Shortlisting of top PLK1-SDL candidates using direct guide RNA capture Perturb-seq screening.

A. Schematic overview of the Perturb-seq screening with direct capture of guide RNA workflow. Briefly, 10X compatible, guide library with direct capture sequence 1 in the stem loop region, was transduced in MCF7 cells and the cells were grown for four days following puromycin selection. Gene expression library and CRISPR KO library were prepared following barcoding and indexing single cells and sequenced with Novaseq. **B.** Single-cell k-means cluster projection of t-sne embedded pooled sgRNA screen showing 10 clusters (n=7434 cells). **C.** Hierarchal clustering using Pearson correlation of the percentage of knockout cells for each gene in the 10 different clusters. Knockouts of few genes that are enriched in clusters 1 and 4, along with positive controls are presented. **D.** Expression analyses of PLK1, AURKA and AURKB in individual cells in log2 scale (n=7434 cells) shows that they are highly co-expressed in cluster 2. **E.** Cells with high expression of PLK1 (marking cluster 2 with a box), were tracked for each of the 65 single knockouts. Knockouts of CRB2, DPP9, IGF2BP2, TJP3 are associated with lowest numbers of cells with high PLK1 levels (negative selection). **F.** Violin plots showing PLK1 levels in the aggregate of all single cells from 65 knockouts. The graph was generated by 10x Genomics Loupe Browser software.

Figure 5: IGF2BP2 affects the expression of PLK1 and tumor growth.

A. Absolute quantification results of ddPCR presented as Target/GAPDH ratios. After knocking down of IGF2BP2, PLK1 expression level was significantly downregulated in MCF7 (P<0.001) and BT549 cells (P<0.05). In contrast, PLK1 transcript levels increased in MDA-MB-231 cells (P<0.001). **B.** Western blot showing the levels of PLK1 following the knockdown of IGF2BP2. GAPDH was used as a loading control. **C.** Expression patterns of PLK1 and IGF2BP2 in multiple cancer types from TCGA show positive correlations. Spearman rank correlations were calculated using the RNA-Seq Expectation Maximization (RSEM) normalized expression of PLK1 and IGF2BP2 in log2 scale. The frequency plots along the x-axis (top) show the frequency of PLK1 expression and the frequency plots along the y-axis (right) shows the frequency of IGF2BP2 expression in each tissue type. Each patient is represented by a blue dot. Spearman correlation co-efficient (r), number of patients (n), p-value significance (p) are included in the graphs. The blue line represents the best linear fit of the distribution. **D.** Sample images with cell masking shown in orange and box plot of Hs578Bst breast epithelial cell confluency before and after transfection with IGF2BP2 targeting *si*RNA. No significant (n.s.) decrease in confluency was observed relative to the non-targeting *si*RNA control, and HS578Bst cells showed normal expansion. **E.** Sample images with cell masking shown in orange and quantitation of PLK1-inducible HCT116 cells, measured using S3- Incucyte® following knockdown of IGF2BP2. **F.** Colony formation assay performed with the indicated

breast cancer cells following knockdown of IGF2BP2. shRFP was used as a control. Colonies were quantified using ImageJ software for the area of the colonies representing the overall colony abundance compared to area of the control in MCF7 (p<0.01), MDA-MB-231 (p<0.01), BT549 (p<0,05), and HCC1143 (p<0,01). **G.** Representative images of tumorspheres formed by MCF7, MDA-MB-231, and BT549 cells (scale bar, 1000 μm). Images of tumorspheres were taken using an EVOS m5000 images, and tumorsphere area was calculated using ImageJ software. The graph represents area of tumorspheres after knocking down of IGF2BP2 compared to matching control in MCF7 (p<0.01), MDA-MB-231 (p<0.01), and BT549 (p<0,05). **H.** Graph representing effect of IGF2BP2 knockdown on tumor volume. MDA-MB-231 cells (2x10^6) transduced with shIGF2BP2 or shRFP (control) were injected into mammary fat pads of immunodeficient female NOD-SCID mice (n=10/group) and tumor volume was measured every 3 to 4 days. The graph shows the mean tumor volume (± standard error) at different time points post-injection. The IGF2BP2 knockdown group showed a significant decrease in tumor volume compared to control group (p<0.001, two-way ANOVA).

Figure 6: Pharmacological inhibition of IGF2BP2 reduces tumor growth.
A. Chemical structures of inhibitors of IGF2BP2. **B.** Western blot showing the levels of PLK1 after treatment with IGF2BP2 inhibitors. Cells were treated with 30 µM concentration for four compounds over 24 hrs and subjected to western blot analyses. The bottom panel shows quantitation of the PLK1 levels as a ratio of tubulin. **C.** Dose-response curves of the compounds (C1, C4, C6 and C9) that show IGF2BP2 inhibition in PLK1-overexpressing cell lines. **D.** Concentration of compounds C1, C4, C6 and C9 in plasma, whole blood, blood cells at 0.25, 0.5, 1, 2, 3, 8, and 24 h as well as in urine at 1, 2 and 24 hours after IV administration (1 mg/kg). **E.** Concentration of compound C4 in plasma at 0.25, 0.5, 1, 6, 24, 48 and 72, in liver at 72 and at 0.25, 0.5, 1, 6 and 24 hours in urine after IP administration (10 mg/kg), Concentration of compound C6 in plasma at 0.25, 0.5, 1, 2, 4, 8 and 24, in liver at 24 and at 0.5, 1 and 2 hours in urine after IP administration (10 mg/kg). **F.** Western blot showing the levels of PLK1, IGF2BP2 and PARP after treatment with IGF2BP2 inhibitor (C4) in MCF7, MDA-MB-231 and BT-549 cell lines at doses ranging from 5 to 75 μM for 72h. GAPDH was used as a loading control. **G.** Effect of IGF2BP2 inhibitor (C4) on tumor growth. MDA-MB-231 cells (2x10^6) were injected into mammary fat pads of immunodeficient female NOD-SCID mice. Treatment with IGF2BP2 inhibitor at the dosage of 10 mg/kg was initiated the following day after the injection of cells. IGF2BP2 inhibitor and a matching volume of DMSO were dissolved in 2-hydroxypropyl-β-cyclodextrin and administered intraperitoneally (IP) for five consecutive days, followed by two days of rest, for a total of 30 days. Tumor volumes were measured with digital caliper in control group treated with DMSO (n=10) and group treated with IGF2BP2 inhibitor (C4) (n=10) every 3 to 4 days after starting treatment. The graph shows the mean tumor volume (± standard error of the mean). The group treated with inhibitor showed a significant decrease in tumor volume compared to the control group (p<0.001, two-way ANOVA) **H.** PDX models were classified based on the expression levels of PLK1 and IGF2BP2. Following this tumor doubling times (left) and slopes of tumor growth (right) are presented for the two categories. **I.** Schematic model showing the regulation of PLK1 expression by IGF2BP2.

Supplementary Figure Legends
Supplementary Figure 1:
A. Correlation between replicates of uninduced and induced samples from the genome-wide *sh*RNA screens. Correlations between multiple timepoints are presented. The left panel shows

168

correlation between uninduced samples, and the right panel shows correlation between induced samples. **B.** Magnitudes of dropouts between uninduced and induced samples at different time points for each hairpin from the genome-wide *sh*RNA screen are plotted. **C.** Cytoscape network of all the 960 SDL hits identified by the genome-wide *sh*RNA screens. Each node represents a hit from the screen. They are color coded based on their gene ontology. Blue edges represent, previously published interactions downloaded from STRING database, showing crosstalk among the SDL hits.

Supplementary Figure 2:
A. IC50 curves of drug inhibitors of some of the SDL hits identified in the screen. The red sigmoidal curve represents the IC50 curve in PLK1 overexpressing cell lines and the blue sigmoidal curves are IC50 of cell lines with low PLK1 expression. **B.** Few representative examples of Kaplan-Meier plots for colon and pancreatic cancer patients displaying natural SDL expression pattern are presented.

Supplementary Figure 3:
A. Display of representative images acquired using automated imaging over time for different PLK1-SDL candidate knockout in Cas9⁻ and Cas9⁺ HCI-010 cells. MetaXpress object masking overlay is shown in blue for the day 6 images. **B.** Correlation plots between the replicates of the *in vivo* pooled CRISPR screen. The left side panel is for the Cas9 negative samples, and the right-side panel is for the Cas9 positive samples. **C.** Representative cleavage assay confirming individual knockouts.

Supplementary Figure 4:
A. Bar graph showing number of cells with each individual knockout analyzed in the single-cell direct capture Perturb-seq screen. **B.** Knockout efficiency in the single cell CRISPR screen was confirmed by comparing the expression of the corresponding target gene between cells from negative controls and knockouts. The percentage of knockout efficiency for each target is presented besides the heatmap.

Supplementary Figure 5:
A. Absolute quantification (concentration in copies/μL) of GAPDH (VIC labeled, green) and PLK1 (FAM labeled, blue) for 3 cell line shRFP control and shIGF2BP2 knockdown samples. **B.** Representative 1D plot showing the positive (blue) and negative (gray) droplets based on IGF2BP2 probe (FAM labeled) in shRFP control and shIGF2BP2 knockdown cell line samples. **C.** Representative 1D plot showing the positive (green) and negative (gray) droplets based on GAPDH probe (VIC labeled) in shRFP control and shIGF2BP2 knockdown cell line samples. **D.** Colony formation assay performed with the indicated breast cancer cells following knockdown of IGF2BP2. shRFP was used as a control. Colonies were quantified using ImageJ software for number of colonies.

Supplementary Table Information
Supplementary Table 1: List of SDL hits from genome-wide *sh*RNA screen with significant weighted differential cumulative changes in PLK-induced versus PLK1-uninduced cells.

Supplementary Table 2: Comparison of all 960 hits with previously published screens associated with mitosis and cell cycle progression.

Supplementary Table 3: List of genes with validated scores based on Marcotte et. al. or Project Achilles screens used to prioritize PLK1-SDL interactions. Significant p-values indicates SDL interactions of PLK1.

Supplementary Table 4: List of genes prioritized using clinical data to capture "naturally occurring" SDL interactions in patients across multiple cancer types.

Supplementary Table 5: List of genes prioritized from 1) ToppGene prioritization, 2) localized to the plasma membrane, and/or 3) were located on the cytoband 19p13.2-3.

Supplementary Table 6: List of SDL hits validated by *in vitro* CRISPR screening approach.

Supplementary Table 7: List of SDL hits validated by *in vivo* CRISPR screening approach.

Supplementary Table 8: List of sgRNAs used as positive and negative controls in Perturb-seq screening

Supplementary Table 9: List of primer sequences used for genomic cleavage detection.

Supplementary Table 10: Output from mass spectrometric experiments.

References

1. R. A. Burrell, N. McGranahan, J. Bartek, C. Swanton, The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338-345 (2013).
2. R. A. Burrell, C. Swanton, The evolution of the unstable cancer genome. *Current opinion in genetics & development* **24**, 61-67 (2014).
3. N. McGranahan, C. Swanton, Biological and Therapeutic Impact of Intratumor Heterogeneity in Cancer Evolution. *Cancer Cell* **27**, 15-26 (2015).
4. S. Misale *et al.*, Blockade of EGFR and MEK intercepts heterogeneous mechanisms of acquired resistance to anti-EGFR therapies in colorectal cancer. *Sci Transl Med* **6**, 224ra226 (2014).
5. C. E. Cunningham *et al.*, Targeting the CINful genome: Strategies to overcome tumor heterogeneity. *Progress in biophysics and molecular biology*, (2019).
6. D. P. Cahill, K. W. Kinzler, B. Vogelstein, C. Lengauer, Genetic instability and darwinian selection in tumours. *Trends Cell Biol* **9**, M57-60 (1999).
7. S. M. Gollin, Mechanisms leading to chromosomal instability. *Semin Cancer Biol* **15**, 33-42 (2005).
8. D. J. Gordon, B. Resio, D. Pellman, Causes and consequences of aneuploidy in cancer. *Nat Rev Genet* **13**, 189-203 (2012).
9. M. Grigorova, J. M. Staines, H. Ozdag, C. Caldas, P. A. Edwards, Possible causes of chromosome instability: comparison of chromosomal abnormalities in cancer cell lines with mutations in BRCA1, BRCA2, CHK2 and BUB1. *Cytogenetic and genome research* **104**, 333-340 (2004).
10. P. V. Jallepalli, C. Lengauer, Chromosome segregation and cancer: cutting through the mystery. *Nat Rev Cancer* **1**, 109-117 (2001).
11. G. J. Kops, B. A. Weaver, D. W. Cleveland, On the road to cancer: aneuploidy and the mitotic checkpoint. *Nat Rev Cancer* **5**, 773-785 (2005).
12. M. A. Nowak *et al.*, The role of chromosomal instability in tumor initiation. *Proc Natl Acad Sci U S A* **99**, 16226-16231 (2002).
13. S. L. Thompson, D. A. Compton, Examining the link between chromosomal instability and aneuploidy in human cells. *J Cell Biol* **180**, 665-672 (2008).
14. J. M. Schvartzman, R. Sotillo, R. Benezra, Mitotic chromosomal instability and cancer: mouse modelling of the human disease. *Nat Rev Cancer* **10**, 102-115 (2010).

15.   J. Luo *et al.*, A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell* **137**, 835-848 (2009).

16.   H. Rajagopalan, M. A. Nowak, B. Vogelstein, C. Lengauer, The significance of unstable chromosomes in colorectal cancer. *Nat Rev Cancer* **3**, 695-701 (2003).

17.   A. V. Roschke, E. Rozenblum, Multi-layered cancer chromosomal instability phenotype. *Frontiers in oncology* **3**, 302 (2013).

18.   B. A. Weaver, D. W. Cleveland, Does aneuploidy cause cancer? *Current opinion in cell biology* **18**, 658-667 (2006).

19.   B. A. Weaver, D. W. Cleveland, The aneuploidy paradox in cell growth and tumorigenesis. *Cancer Cell* **14**, 431-433 (2008).

20.   C. E. Cunningham *et al.*, The CINs of Polo-Like Kinase 1 in Cancer. *Cancers* **12**, (2020).

21.   L. Gheghiani *et al.*, PLK1 Induces Chromosomal Instability and Overrides Cell-Cycle Checkpoints to Drive Tumorigenesis. *Cancer Res* **81**, 1293-1307 (2021).

22.   G. de Carcer *et al.*, Plk1 overexpression induces chromosomal instability and suppresses tumor development. *Nature communications* **9**, 3012 (2018).

23.   V. Archambault, D. M. Glover, Polo-like kinases: conservation and divergence in their functions and regulation. *Nat Rev Mol Cell Biol* **10**, 265-275 (2009).

24.   S. M. Lens, E. E. Voest, R. H. Medema, Shared and separate functions of polo-like kinases and aurora kinases in cancer. *Nat Rev Cancer* **10**, 825-841 (2010).

25.   F. A. Barr, H. H. Sillje, E. A. Nigg, Polo-like kinases and the orchestration of cell division. *Nat Rev Mol Cell Biol* **5**, 429-440 (2004).

26.   S. Iliaki, R. Beyaert, I. S. Afonina, Polo-like kinase 1 (PLK1) signaling in cancer and beyond. *Biochemical pharmacology* **193**, 114747 (2021).

27.   C. E. Cunningham *et al.*, Therapeutic relevance of the protein phosphatase 2A in cancer. *Oncotarget*, (2016).

28.   G. Chun *et al.*, Polo-like kinase 1 enhances survival and mutagenesis after genotoxic stress in normal cells through cell cycle checkpoint bypass. *Carcinogenesis* **31**, 785-793 (2010).

29.   Y. Mao, L. Xi, Q. Li, X. Zhang, C. Yu, Polo-like kinase 1 is involved in apoptosis, invasion, and metastasis of pancreatic ductal adenocarcinoma. *Transl Cancer Res* **9**, 6672-6682 (2020).

30.   Y. Kong *et al.*, The kinase PLK1 promotes the development of Kras/Tp53-mutant lung adenocarcinoma through transcriptional activation of the receptor RET. *Science signaling* **15**, eabj4009 (2022).

31.   Z. Gao *et al.*, PLK1 promotes proliferation and suppresses apoptosis of renal cell carcinoma cells by phosphorylating MCM3. *Cancer Gene Ther* **27**, 412-423 (2020).

32.   Jing A *et al.*, Exclusively expression-based analyses indicates mutational changes within the genome precede induction of chromosomal instability in triple negative breast cancer. *npj Systems biology and Applications* **2018, 4:38; doi:10.1038/s41540-018-0074-z**, (2018).

33.   V. Maire *et al.*, Polo-like kinase 1: a potential therapeutic option in combination with conventional chemotherapy for the management of patients with triple-negative breast cancer. *Cancer Res* **73**, 813-823 (2013).

34.   T. Takahashi *et al.*, Polo-like kinase 1 (PLK1) is overexpressed in primary colorectal cancers. *Cancer science* **94**, 148-152 (2003).

35.   W. Weichert *et al.*, Overexpression of Polo-like kinase 1 is a common and early event in pancreatic cancer. *Pancreatology : official journal of the International Association of Pancreatology* **5**, 259-265 (2005).

36.   K. Dietzmann, E. Kirches, B. von, K. Jachau, C. Mawrin, Increased human polo-like kinase-1 expression in gliomas. *J Neurooncol* **53**, 1-11 (2001).

37.   G. Wolf *et al.*, Prognostic significance of polo-like kinase (PLK) expression in non-small cell lung cancer. *Oncogene* **14**, 543-549 (1997).

38.   A. Deeraksa *et al.*, Plk1 is upregulated in androgen-insensitive prostate cancer cells and its inhibition leads to necroptosis. *Oncogene* **32**, 2973-2983 (2013).

39. C. E. Sunkel, D. M. Glover, polo, a mitotic mutant of Drosophila displaying abnormal spindle poles. *Journal of cell science* **89 (Pt 1)**, 25-38 (1988).

40. M. R. Smith *et al.*, Malignant transformation of mammalian cells initiated by constitutive expression of the polo-like kinase. *Biochem Biophys Res Commun* **234**, 397-405 (1997).

41. D. P. Seeburg, M. Feliu-Mojer, J. Gaiottino, D. T. Pak, M. Sheng, Critical role of CDK5 and Polo-like kinase 2 in homeostatic synaptic plasticity during elevated activity. *Neuron* **58**, 571-583 (2008).

42. Y. Yang *et al.*, Polo-like kinase 3 functions as a tumor suppressor and is a negative regulator of hypoxia-inducible factor-1 alpha under hypoxic conditions. *Cancer Res* **68**, 4077-4085 (2008).

43. L. Y. Lu *et al.*, Polo-like kinase 1 is essential for early embryonic development and tumor suppression. *Molecular and cellular biology* **28**, 6870-6876 (2008).

44. S. Parameswaran *et al.*, A Road Map to Personalizing Targeted Cancer Therapies Using Synthetic Lethality. *Trends in cancer* **5**, 11-29 (2019).

45. J. M. Paul, S. D. Templeton, A. Baharani, A. Freywald, F. J. Vizeacoumar, Building high-resolution synthetic lethal networks: a 'Google map' of the cancer cell. *Trends Mol Med* **20**, 704-715 (2014).

46. S. Duffy *et al.*, Overexpression screens identify conserved dosage chromosome instability genes in yeast and human cancer. *Proc Natl Acad Sci U S A* **113**, 9967-9976 (2016).

47. E. S. Kroll, K. M. Hyland, P. Hieter, J. J. Li, Establishing genetic interactions by a synthetic dosage lethality phenotype. *Genetics* **143**, 95-102 (1996).

48. R. J. Reid *et al.*, A Synthetic Dosage Lethal Genetic Interaction Between CKS1B and PLK1 Is Conserved in Yeast and Human Cancer Cells. *Genetics* **204**, 807-819 (2016).

49. J. M. Replogle *et al.*, Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat Biotechnol* **38**, 954-961 (2020).

50. J. M. Paul *et al.*, Targeting synthetic lethality between the SRC kinase and the EPHB6 receptor may benefit cancer treatment. *Oncotarget*, (2016).

51. T. Hart, J. Moffat, BAGEL: a computational framework for identifying essential genes from pooled library screens. *BMC Bioinformatics* **17**, 164 (2016).

52. B. Jassal *et al.*, The reactome pathway knowledgebase. *Nucleic Acids Res* **48**, D498-D503 (2020).

53. S. X. Ge, E. W. Son, R. Yao, iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics* **19**, 534 (2018).

54. J. Barretina *et al.*, The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-607 (2012).

55. G. S. Cowley *et al.*, Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. *Sci Data* **1**, 140035 (2014).

56. R. Marcotte *et al.*, Essential gene profiles in breast, pancreatic, and ovarian cancer cells. *Cancer Discov* **2**, 172-189 (2012).

57. J. Chen, E. E. Bardes, B. J. Aronow, A. G. Jegga, ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* **37**, W305-311 (2009).

58. Y. S. DeRose *et al.*, Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. *Nat Med* **17**, 1514-1520 (2011).

59. S. C. Chafe *et al.*, Genome-wide synthetic lethal screen unveils novel CAIX-NFS1/xCT axis as a targetable vulnerability in hypoxic solid tumors. *Sci Adv* **7**, (2021).

60. E. Metzakopian *et al.*, Enhancing the genome editing toolbox: genome wide CRISPR arrayed libraries. *Scientific reports* **7**, 2244 (2017).

61. Y. Zhang, M. Huo, J. Zhou, S. Xie, PKSolver: An add-in program for pharmacokinetic and pharmacodynamic data analysis in Microsoft Excel. *Comput Methods Programs Biomed* **99**, 306-314 (2010).

62. H. Gao *et al.*, High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat Med* **21**, 1318-1325 (2015).

63. A. S. Mer *et al.*, Integrative Pharmacogenomics Analysis of Patient-Derived Xenografts. *Cancer Res* **79**, 4539-4550 (2019).

64. S. F. Bakhoum, L. C. Cantley, The Multifaceted Role of Chromosomal Instability in Cancer and Its Microenvironment. *Cell* **174**, 1347-1360 (2018).

65. A. G. Renner *et al.*, A functional link between polo-like kinase 1 and the mammalian target-of-rapamycin pathway? *Cell cycle* **9**, 1690-1696 (2010).

66. B. C. van de Weerdt *et al.*, Uncoupling anaphase-promoting complex/cyclosome activity from spindle assembly checkpoint control by deregulating polo-like kinase 1. *Molecular and cellular biology* **25**, 2031-2044 (2005).

67. T. Waldman, C. Lengauer, K. W. Kinzler, B. Vogelstein, Uncoupling of S phase and mitosis induced by anticancer agents in cells lacking p21. *Nature* **381**, 713-716 (1996).

68. T. Waldman, K. W. Kinzler, B. Vogelstein, p21 is necessary for the p53-mediated G1 arrest in human cancer cells. *Cancer Res* **55**, 5187-5190 (1995).

69. T. Hart, K. R. Brown, F. Sircoulomb, R. Rottapel, J. Moffat, Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. *Mol Syst Biol* **10**, 733 (2014).

70. J. S. Andersen *et al.*, Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* **426**, 570-574 (2003).

71. J. R. Hutchins *et al.*, Systematic analysis of human protein complexes identifies chromosome segregation proteins. *Science* **328**, 593-599 (2010).

72. R. Kittler *et al.*, An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. *Nature* **432**, 1036-1040 (2004).

73. J. Moffat *et al.*, A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283-1298 (2006).

74. M. Mukherji *et al.*, Genome-wide functional analysis of human cell-cycle regulators. *Proc Natl Acad Sci U S A* **103**, 14819-14824 (2006).

75. B. Neumann *et al.*, Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* **464**, 721-727 (2010).

76. M. Sokolova *et al.*, Genome-wide screen of cell-cycle regulators in normal and tumor cells identifies a differential response to nucleosome depletion. *Cell cycle* **16**, 189-199 (2017).

77. F. J. Vizeacoumar *et al.*, A negative genetic interaction map in isogenic cancer cell lines reveals cancer cell vulnerabilities. *Mol Syst Biol* **9**, 696 (2013).

78. S. Munira, R. Yuki, Y. Saito, Y. Nakayama, ALK Inhibitors-Induced M Phase Delay Contributes to the Suppression of Cell Proliferation. *Cancers* **12**, (2020).

79. L. Jerby-Arnon *et al.*, Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. *Cell* **158**, 1199-1209 (2014).

80. T. Hart *et al.*, High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163**, 1515-1526 (2015).

81. R. Jorge, C. Silva, S. Agueda, S. Doria, M. Leao, Intellectual disability and overgrowth-A new case of 19p13.13 microdeletion syndrome with digital abnormalities. *Am J Med Genet A* **167A**, 2839-2843 (2015).

82. P. L. Chavali, M. Putz, F. Gergely, Small organelle, big responsibility: the role of centrosomes in development and disease. *Philos Trans R Soc Lond B Biol Sci* **369**, (2014).

83. J. Gonzalez-Martinez *et al.*, Genetic interaction between PLK1 and downstream MCPH proteins in the control of centrosome asymmetry and cell fate during neural progenitor division. *Cell death and differentiation*, (2022).

84. C. F. Aspinall, D. Zheleva, A. Tighe, S. S. Taylor, Mitotic entry: Non-genetic heterogeneity exposes the requirement for Plk1. *Oncotarget* **6**, 36472-36488 (2015).

85. Z. Li *et al.*, Polo-like kinase 1 (Plk1) overexpression enhances ionizing radiation-induced cancer formation in mice. *The Journal of biological chemistry*, (2017).

86. M. Wierer *et al.*, PLK1 signaling in breast cancer cells cooperates with estrogen receptor-dependent gene transcription. *Cell reports* **3**, 2021-2032 (2013).

87. C. Dahlem *et al.*, First Small-Molecule Inhibitors Targeting the RNA-Binding Protein IGF2BP2/IMP2 for Cancer Therapy. *ACS chemical biology* **17**, 361-375 (2022).

88.	R. E. Gutteridge, M. A. Ndiaye, X. Liu, N. Ahmad, Plk1 Inhibitors in Cancer Therapy: From Laboratory to Clinics. *Mol Cancer Ther* **15**, 1427-1435 (2016).

89.	K. Mross *et al.*, A randomised phase II trial of the Polo-like kinase inhibitor BI 2536 in chemo-naive patients with unresectable exocrine adenocarcinoma of the pancreas - a study within the Central European Society Anticancer Drug Research (CESAR) collaborative network. *British journal of cancer* **107**, 280-286 (2012).

90.	D. Olmos *et al.*, Phase I study of GSK461364, a specific and competitive Polo-like kinase 1 inhibitor, in patients with advanced solid malignancies. *Clin Cancer Res* **17**, 3420-3430 (2011).

91.	W. M. Stadler *et al.*, An open-label, single-arm, phase 2 trial of the Polo-like kinase inhibitor volasertib (BI 6727) in patients with locally advanced or metastatic urothelial cancer. *Cancer* **120**, 976-982 (2014).

92.	A. J. Aguirre *et al.*, Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov* **6**, 914-929 (2016).

93.	M. Kosicki, K. Tomberg, A. Bradley, Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol* **36**, 765-771 (2018).

94.	U. Ben-David *et al.*, Patient-derived xenografts undergo mouse-specific tumor evolution. *Nat Genet* **49**, 1567-1575 (2017).

95.	A. Moamer *et al.*, A role for kinesin-1 subunits KIF5B/KLC1 in regulating epithelial mesenchymal plasticity in breast tumorigenesis. *EBioMedicine* **45**, 92-107 (2019).

96.	H. Y. Kim *et al.*, Farnesyl diphosphate synthase is important for the maintenance of glioblastoma stemness. *Exp Mol Med* **50**, 1-12 (2018).

97.	K. E. Hamilton *et al.*, IMP1 promotes tumor growth, dissemination and a tumor-initiating cell phenotype in colorectal cancer cell xenografts. *Carcinogenesis* **34**, 2647-2654 (2013).

98.	M. Janiszewska *et al.*, Imp2 controls oxidative phosphorylation and is crucial for preserving glioblastoma cancer stem cells. *Genes Dev* **26**, 1926-1944 (2012).

99.	H. Huang *et al.*, Recognition of RNA N(6)-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat Cell Biol* **20**, 285-295 (2018).

100.	H. Luo *et al.*, METTL3-mediated m(6)A modification regulates cell cycle progression of dental pulp stem cells. *Stem Cell Res Ther* **12**, 159 (2021).

101.	S. Tatekawa *et al.*, N(6)-methyladenosine methylation-regulated polo-like kinase 1 cell cycle homeostasis as a potential target of radiotherapy in pancreatic adenocarcinoma. *Scientific reports* **12**, 11074 (2022).

102.	D. S. Boss, J. H. Beijnen, J. H. Schellens, Clinical experience with aurora kinase inhibitors: a review. *The oncologist* **14**, 780-793 (2009).

103.	C. Lemmers *et al.*, CRB3 binds directly to Par6 and regulates the morphogenesis of the tight junctions in mammalian epithelial cells. *Mol Biol Cell* **15**, 1324-1333 (2004).

104.	L. Gonzalez-Mariscal, A. Betanzos, A. Avila-Flores, MAGUK proteins: structure and role in the tight junction. *Semin Cell Dev Biol* **11**, 315-324 (2000).

105.	M. Petronczki, P. Lenart, J. M. Peters, Polo on the Rise-from Mitotic Entry to Cytokinesis with Plk1. *Dev Cell* **14**, 646-659 (2008).

106.	R. Geiss-Friedlander *et al.*, The cytoplasmic peptidase DPP9 is rate-limiting for degradation of proline-containing peptides. *The Journal of biological chemistry* **284**, 27211-27219 (2009).

107.	E. Pilla *et al.*, A novel SUMO1-specific interacting motif in dipeptidyl peptidase 9 (DPP9) that is important for enzymatic regulation. *The Journal of biological chemistry* **287**, 44320-44329 (2012).

108.	J. Zhang, C. Yuan, J. Wu, Z. Elsayed, Z. Fu, Polo-like kinase 1-mediated phosphorylation of Forkhead box protein M1b antagonizes its SUMOylation and facilitates its mitotic function. *The Journal of biological chemistry* **290**, 3708-3719 (2015).

Fig 1

Fig 2

Fig 3

Fig 5

Fig 4

179

Fig 6

A

B

C

PLK1

Legend

Translation & Ribosome Biogenesis	DNA Damage Repair & Nucleic Acid Metabolism		
mRNA Processing	Chromatin & Transcription	Metabolism	Immune Process
Cytoskeleton	Signal Transduction	Unknown	Miscellaneous
Bipolar Spindle Formation & Chromosome Alignment	Developmental Process	Cell Communication & Adhesion	Intracellular Trafficking & Vesicle Formation
Centrosome & Microtubule Formation	Non-coding and Pseudogenes	Protein Modifcation & Proteolysis	Protein Transport & Folding
Cell Cycle Progression & DNA Replication	Apoptosis & Cell Fate	Ion & Small Molecule Transport	Enzyme Activity

△ Published   ◇ Gene Ontology

A

CHIR-99021
targets GSK3

SB-505124
targets ACBR1V/ALK4

rTRAIL agonist
for multiple members

NVP-TAE684
targets ALK

GNF-2
targets ABL1

B

COAD

PAAD

— Naturally occuring SDL
— No SDL

Fig S3

183

Fig S4

A MCF7

B IGF2BP2 levels

C GAPDH levels

D

V.   Chapter 5: Final Discussion

As elucidated in previous chapters, RNA-binding proteins (RBP) play a crucial role in diverse biological processes within both bacterial and mammalian cells. Through their selective recognition and binding to specific RNA sequences, RBPs are able to regulate gene expression and influence the RNA metabolism.

Discovery of inhibitors targeting RBPs including cellular validation of these potential hit scaffolds was the aim of this thesis. The first part (**Chapter III.1 and 2**) was focused on CsrA, which is widely spread and highly conserved in diverse gram-negative pathogens. Moreover, CsrA serves as a key regulator of gene expression, coordinating cellular responses to environmental changes and influencing important bacterial phenotypes, including carbon metabolism, biofilm formation and virulence. Especially as a promising target for anti-virulence therapy, it is important to find cell-active substances disrupting these CsrA-RNA interactions.

The second part (**Chapter IV.1 and 2**) was focused on Insulin-like growth factor 2 mRNA binding proteins (IMPs), particularly IMP2, which has been demonstrated to be overexpressed in various tumor types including liver and colon cancer. Furthermore, this RBP facilitates tumorigenesis, tumor progression and has also been implicated in contributing to worsen the disease outcome. Targeting IMP2 for anti-cancer therapy presents a viable approach to reduce tumor cell proliferation. Similar to the CsrA project, the discovery of inhibitors that interfere the IMP2-RNA associations is crucial in the study of IMP2. These inhibitors should have the potential to modulate IMP2's function and downstream regulatory effects, offering new avenues for therapeutic interventions and furthering the understanding of IMP2-mediated processes.

For the sake of clarity, the compound codes mentioned in the following discussion section are derived from the capital letter of each chapter followed by the number within the respective manuscripts they belong to.

1. Screening for potential inhibitors against CsrA

Based on the previous findings, the FP-based assay proofed to be robust and reliable

methodology for screening new lead structures and determination of the inhibitory activity against labeled RNA in a cell-free setup.[31] Furthermore, our initial approach of using phage display technology to screen for cyclic peptides as a novel inhibitor scaffold was successful (**Chapter III.1, A**).[30] In both studies the CsrA target was from *Yersinia pseudotuberculosis* and included a specific biotin-modification within the construct. Biotinylated CsrA enabled the immobilization for the purpose of phage display. Details of the phage display-based screening have been already discussed in the thesis of Dr. Valentin Jakob[39], therefore this work will only discuss the most important findings that are relevant for the discussion of the following chapter.

Phage display offered the distinct advantage of having a vast library containing millions of compound variants. In our study, we designed a phage library comprising 2.48 million disulfide-constrained heptapeptide variants, encompassing a mass range spanning from 548 to 1193 Da. Out of 32 clones we could identify one sequence as a potential CsrA binder. This disulfide-bridged peptide **A1** was chosen due to the presence of only one tryptophan. Higher amount of tryptophan usually leads to unspecific binding. The glutamic acid residue within the peptide structure is beneficial as well for binding the positively charged surface of CsrA. Further structure-activity relationships were determined through techniques such as Ala Scan, which resulted in the conclusion that both the macrocycle itself and the serine/tryptophan sidechains are essential for high affinity and activity. Together with the collaborators we solved the solution structure of the disulfide peptide by NMR, followed by a docking experiment based on the *Yersina pseudotuberculosis* CsrA homology model. The result of the docking experiment were in accordance with our SAR findings.[30]

In order to achieve intracellular activity, the reductive linearization of the cysteine bridge in **A1** was prevented by biomimetic replacement with non-natural amino acids bearing alkyne and azide functions in their sidechain for click chemistry-based macrocyclization. Both 1,4-disubsituted triazole bridging motifs **A5a,b** maintained their activity against the CsrA protein from *Yersinia pseudotuberculosis* and additionally demonstrated activity against *E. coli* and *Pseudomonas aeruginosa* protein as well. Particularly in *E. coli*, they exhibited IC_{50} values in the low micromolar range. These findings have encouraged us to proceed with the evaluation of these compounds in a cellular environment.

Since **A1** and **A5a,b** are most active against *E. coli* CsrA, the development of an *in bacterio* assay for assessment of these hit compounds was the next logical step. (**Chapter III.2, B**) This study was the main focus of the current PhD.

2. Development of an in bacterio assay for assessment of potential inhibitors

Given that CsrA is a target for pathoblocker compounds, which are ideally devoid of bactericidal or bacteriostatic effects, conventional antibacterial assays such as minimum inhibitory concentration (MIC) assays are unsuitable. Thus far, a dedicated assay tailored for this purpose remains unavailable.

The biggest challenge was to establish an *in bacterio* assay that not only allows for investigating but also enables quantification of the impact on CsrA-regulated cellular mechanisms. Initially, our choice was to employ a luminescence-based assay setup. The utilization of a luciferase reporter gene assay offers the advantage of enabling real-time investigation of inhibitory effects. This assay system allows for direct and notable changes in bioluminescence, facilitating the rapid and precise assessment of inhibitory activity.

In case of targeting the CsrA, we used a designed vector called pvBE3 harboring the *glgC-luxCDEAB* reporter fusion. This fusion construct encompassed the complete promoter region of *glgC*, which is negatively regulated by CsrA as described in **chapter I.3.2**. In the presence of active pathoblocker compounds, an upregulation of *glgC-lux* expression is expected resulting in enhancement of the bioluminescence.

The difficulty we had to face was the validation and interpretation of the results obtained from this inhibition assay. To ensure the functionality of the assay, a positive control is required. Initially, natural intracellularly expressed antagonists of CsrA were considered as suitable for this purpose.

As outlined in **chapter III.2 B,** CsrA exhibits a higher affinity for sRNA antagonist CsrB compared to its own target mRNAs, therefore we prioritized the examination of CsrB being the positive control first. Inducing the overexpression of sRNA within different *glglux* fusion strains lead to a decline in luminescence unlike the expected results. Our

findings in qPCR gene expression assay proved that the overexpression of CsrB caused a significant increase in CsrA-encoding transcripts (p = 0,0128). Experiments involving CsrC resulted in a comparable outcome. Even in the absence of IPTG-mediated induction, the overexpression of sRNA antagonists was overcompensated by a 100- or 1000-fold upregulation in *csrA* expression. These results obtained from the qPCR analysis provided an explanation for the observations made in luciferase reporter gene assay. Furthermore, this indicated the sensitivity of the reporter gene assay towards the *csrA* transcripts level.

The reasons for this (auto)regulatory mechanism that induces *csrA* upregulation through overexpression of the sRNA antagonists will be elucidated in the following section. As described in **chapter I.3.3**, the regulatory circuitry surrounding the Csr system in *E. coli* is undoubtedly complex, vast and not fully understood yet (Figure 18). According to previous studies, Csr system utilizes several negative feedback loops[32,34,60], which are beneficial for acceleration of regulatory responses[32,60,61], generation of graded responses thus effectively minimizing cell-cell variability[32,60,62] and reducing background noise.[32,60,63] Previous investigations of the Csr system in *E. coli* demonstrated the utilization of negative feedback to achieve reduction of response time.[32,64]
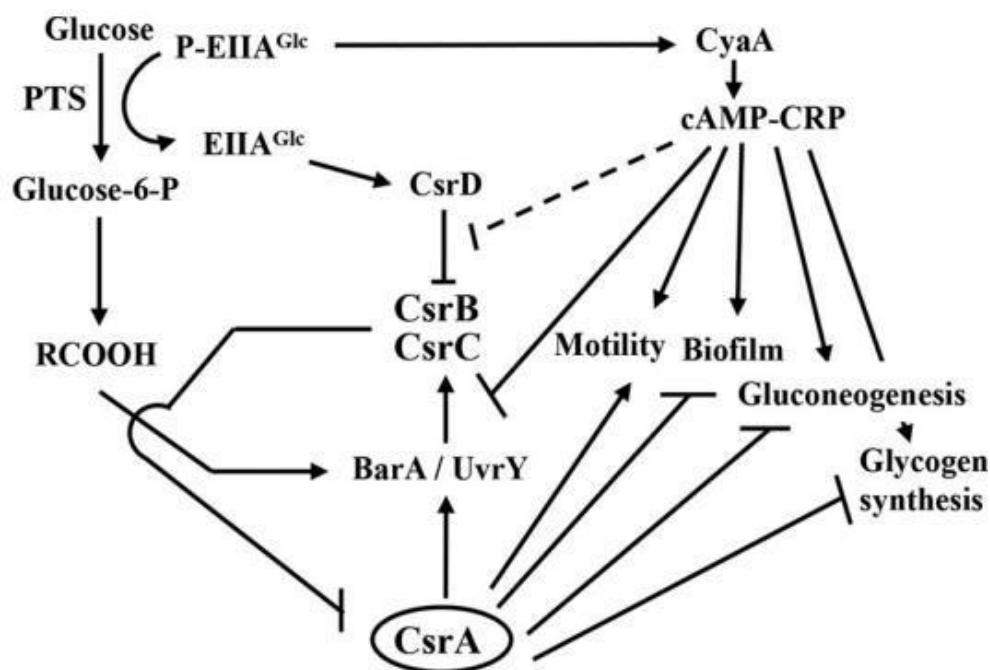

Figure 18: illustration of the simplified regulation circuit in Csr (*E. coli*) system, originated from [65]. Regulatory interactions are labeled either in arrowhead (activation) or flathead (inhibition).Dash line symbolized the weak effect (2-fold) of cAMP-CRP on the turnover of sRNAs.[65]

A main negative feedback loop within the Csr system in *E. coli* involves the indirect activation of CsrB/C transcription through CsrA via BarA-UvrY two-component signal transduction system (TCS).[32,38,40,60] For this purpose, CsrA positively influences UvrY expression and facilitates BarA kinase activity.[32,34] In this context, it is important to take into account the influence of carbon nutrients. This is because the BarA-UvrY signaling pathway is additionally activated by short-chain fatty acids (SCFAs), including formate and acetate, which are metabolic end products of carbon metabolism (Figure 18).[32,38,40,65] The depletion of the preferred carbon substrate (glucose) and accumulation of its end products lead to upregulation of CsrB/C expression.[65] In return, by sequestering CsrA through sRNAs, rapid reduction of CsrA's activity can be achieved without relying on dilution through growth.[32,64]

Apart from the BarA-UvrY TCS, other factors such as nutrient starvation and stress conditions positively regulate CsrB/C transcription as well.[32,60] During stringent response, the intracellular concentration of alarmone molecules (nucleotide secondary messenger guanosine tetraphosphate, ppGpp) increases leading to activation of 10-fold CsrB/C expression in order to downregulate CsrA's activity (Figure 18).[32,60] ppGpp and CsrA also have opposite effects on the mRNA targets *glgC* (glycogen synthesis) and *flhDC* (motility), which will be intensified due to the downregulation of CsrA's activity.[60]

All the investigations mentioned above confirmed the findings that upregulation of sRNAs leads to the inhibition of CsrA's activity. This has been supported by observed decreases in the expression of glycolytic genes and increases in the expression of genes related to gluconeogenesis and glycogen biosynthesis among others.[41,65,66] Additionally, physiological switch from the exponential phase to the stationary phase of growth and a stress-resistant phenotype are facilitated by the downregulation of CsrA.[60,65,67] None of these studies described the effect we observed and found in **chapter III.2, B**. regarding the induced upregulation of CsrA's expression by sRNAs' overexpression, which is contradictory in this context.

In order to provide a potential explanation for our finding, it is necessary to revisit the regulatory circuitry and take into account the involvement of the second protein in the Csr system, CsrD. The role of CsrD will be described in the following section, and the

potential correlation to our findings will be elucidated.

This predicted membrane protein has similar features of a signaling protein[32,67] containing an EAL domain(it is called after its conserved sequence motif Glu-Ala-Leu) that facilitates binding with glucose-specific enzyme IIA (EIIA^{Glc}) and triggering the decay of CsrB/C (Figure 18).[65,67] Specifically, the dephosphorylated state of EIIA^{Glc}, which is predominant during glucose transport via the phosphotransferase system (PTS), enables binding to the EAL domain of CsrD.[32,65,67] The turnover of CsrB/C involves the participation of endonuclease RNase E and the 3'- to 5'-exonuclease polynucleotide phosphorylase (PNPase).[32,65,67] In the absence of CsrD-EIIA^{Glc}, RNase E is unable to trigger the decay of the sRNAs, due to the protective effect of CsrA, which binds to CsrB blocking the RNAse cleavage sites. The role of CsrD demonstrates that even in the presence of CsrA, the turnover of CsrB/C can be still activated by the availability of glucose and other carbon sources.[32,65,67] However, it also has been reported that CsrA represses CsrD as in a sperate negative feedback loop. Nevertheless, EIIA^{Glc}–dependent regulatory pathways allow glucose to simultaneously stimulate both the turnover and synthesis pathways of CsrB/C.[60,68]

Based on the background information mentioned above, a possible explanation is that the overexpression of CsrB/C leads to the upregulation of CsrA, which in turn affects the availability of glucose and activates another regulatory pathway. As CsrA represses gluconeogenesis, the absence of preferred carbon source such as glucose results in the phosphorylated form of EIIA^{Glc} (P-EIIA^{Glc}) binding to adenylate cyclase (CyaA) (Figure 18). [65] This binding facilitates the synthesis of cyclic AMP (cAMP) and the formation of cAMP-cAMP receptor protein (CRP) complex.[65] The cAMP-CRP complex directly represses the transcription of CsrC and indirectly affects the transcription of CsrB.[65] It is not to be excluded that the overexpression of the sRNAs trigger other regulatory pathways to regulate and maintain the balance between CsrA and CsrB/C levels in the cells. Perhaps even as an defense mechanism aimed at conserving resources, considering that both turnover and synthesis of sRNAs require carbon sources. In summary, the proposed explanation for our findings emphasize the complexity of these regulatory circuits surrounding CsrA and the tightly interplay among the Csr components. It is crucial to recognize that regulatory mechanisms within the Csr system can vary between different bacterial species, further adding to

the complexity of understanding this intricate network. Due to the challenges encountered in utilizing CsrB/C as positive control, as discussed earlier, an alternative approach was adopted. We chose to investigate the protein antagonist CesT, a chaperone protein known to inhibit CsrA's activity, while not affecting the transcript level of CsrA.[42,43] With this positive control, we were able to start assessing the first inhibitor hits discovered in **chapter III.1**. The peptides **B1** and **B2** encountered difficulty in penetrating the Gram-negative cell membranes, hindering their ability to enter the cytoplasm and effectively target the desired protein. As a result, these peptides were unable to inhibit CsrA in the *in bacterio* assay. This was verified through subcellular quantification of uptake in *E. coli*. Despite an extracellular concentration of 28 μM being applied, the intracellular levels of **B1** were only detected at nanomolar concentrations. These intracellular concentrations proved to be inadequate for disrupting the CsrA-RNA interaction, considering the micromolar potency of the peptide demonstrated in the target-based FP assay.

Nevertheless, the general reporter gene assay concept is now fit-for-purpose to facilitate quantitative compound evaluation aiming to identify novel inhibitors with cellular efficacy. Furthermore, high throughput screening of potential inhibitors can be achieved by this *in bacterio* assay as well using the 96-well format.

3. Concluding remarks and outlook of CsrA project

In the first part of this dissertation, we discovered a macrocyclic peptide that serves as a novel inhibitor scaffold specifically targeting CsrA. This discovery was achieved through a screening method utilizing phage display. The subsequent characterization involved conducting fluorescence polarization-based functional activity tests and binding assays using MST (microscale thermophoresis). The results demonstrated promising potency, with low micromolar IC_{50} values even against CsrA homologs in other gram-negative species. Furthermore, we conducted additional SAR (structure-activity relationship) research to optimize the compounds and enhance their stability towards reductive linearization.

We have successfully developed an *in bacterio* assay based on a luciferase reporter gene assay. As a positive control for the assay, we utilized the chaperone protein CesT,

which demonstrated the expected increase in bioluminescence in subsequent experiments. This approach allows us to assess the on-target effects of these compounds within bacterial cells. While we validated the peptide scaffolds identified through our initial phage display-based screening, we encountered a limitation regarding their penetration in this *in bacterio* assay.

Further peptide optimizations to aim at improving penetration as one of the next steps to be considered. The idea is to use nanoparticles encapsulating the peptides, improving the targeted delivery.

Furthermore, we are in the process of preparing a manuscript that highlights novel small molecule scaffolds discovered from a commercial compound library using *in silico* and fluorescence polarization (FP)-based screening. These hit compounds demonstrate a greater degree of diversity compared to the peptides, and their compact size and suitable physicochemical properties render penetration into bacterial cells possible. Additionally, these small molecules can be readily optimized to possess desired drug-like properties, including stability, solubility, metabolic stability, and low toxicity. The cellular efficacy of these potential inhibitors will be validated using the established reporter gene assay.

The purpose of the *in bacterio* assay is not only validation but also presents a favorable opportunity to conduct phenotypic screening using commercial synthetic or natural product libraries. This is one of the major steps towards addressing the complexities of this virulence-modulating target.

Another perspective for the CsrA project is to establish additional cell-based assays such as a motility assay for providing more comprehensive assessment of the inhibitor's efficacy against flagellum biosynthesis. However, also identification of possible off-target effects and overall a better understanding of the mode of action in CsrA downstream regulation can be achieved.

It would be also worth to optimize the expression in order to obtain enough yield for solving the (co-)crystal structures of *E. coli* CsrA, in order to unambiguously elucidate the binding mode of identified inhibitors. This includes key binding residues and

structural features that contribute to binding affinity and specificity. Moreover, the crystal structures would be valuable templates for further virtual screening and ligand-based design approaches accelerating the development of effective and selective inhibitors.

4. Screening for potential inhibitors against IMP2

Chapter IV.1, C focused on achieving three primary goals. Firstly, the validation of IMP2 as a potential target for colorectal cancer, followed by the establishment of a screening assay specifically designed for the identification of small molecule inhibitors targeting IMP2. Lastly, the assessment of the identified hit compounds regarding their biological activity.

One of the challenges in this work was establishing a FP-based assay for screening purposes. Unlike the CsrA project, a specific labeled RNA sequence had not been identified yet. Therefore, two sequences were selected based on published binding motifs of IMP2. To determine the biological efficacy of these sequences, a saturation experiment was conducted. The resulting EC_{50} values revealed that RNA_A exhibited a lower concentration required for potency leading to a higher affinity to IMP2 (EC_{50} of 60.7 nM) compared to RNA_B (EC_{50} of 80.5 nM). Nevertheless, both RNA sequences were used in the FP-based screening as well as FP-based *in vitro* hit validation.

In the FP-based screening of different compound libraries, two classes of compounds emerged as potential inhibitors with biological target specificity: the benzamidobenzoic acid class and the ureidothiophene class. To validate the screening results, additional assays were employed, including the thermal shift assay (TSA) and saturation transfer difference NMR (STD-NMR). These complementary techniques confirmed the binding affinity and target specificity of the identified hit compounds, especially the three most active ones (**C4,C6** and **C9**). The discovery of these compound classes provides promising starting points for further optimization and development of specific inhibitors against IMP2.

In addition, as outlined in **chapter IV.2, D**, it has been demonstrated that IMP2

possesses the ability to regulate the expression of polo-like kinase 1 (PLK1), which is commonly found to be overexpressed in various types of cancer. This dysregulation of PLK1 can lead to chromosomal instability (CIN), a characteristic feature of cancer cells. Through the Inhibition of IMP2 using the three hit compounds (**D4, D6** and **D9)** identified from **chapter IV.1**, a notable reduction in the expansion of PLK1 overexpressing cells and tumors was observed. The workflow employed in this side project shares similarities with the CsrA project, but it incorporates robust biological test systems for *in vivo* validation. For example, the MTT assay, a widely used cell viability assay, was employed to assess the impact of the identified hit compounds on cell viability and proliferation. Moreover, xenograft models (for example zebrafish embryos or triple-negative breast cancer (TNBC)) were utilized to evaluate the efficacy and potential anti-tumor effects of the scaffolds in a more complex and physiological setting.

Besides *in vivo* validation, pharmacokinetic (PK) studies were conducted in **chapter IV.2** in order to evaluate the absorption, distribution, metabolism, and excretion (ADME) properties of the identified compounds. After intravenous application in mice and a half-life of around 22 hours **D4** exhibited the best PK profile the highest exposures in plasma, whole blood, blood cells, and urine. The findings from these studies suggest that further optimization of inhibitors targeting IMP2 holds substantial therapeutic potential, representing a significant step towards advancing anti-cancer therapy.

5. Unpublished investigations in the IMP2 project

Building upon the previous research described in the last chapter, our investigation continued in identifying novel inhibitor scaffolds, with a specific emphasis on targeting the KH34 domains. As outlined in **chapter I.4.1**, the C-terminal region of the KH34 domains, along with the variable loop regions, play a crucial role in specific RNA recognition and stabilization. Instead of using the full length protein, further screening experiments and following assays were based on isolated KH34 domains.

The subsequent section highlights the challenges encountered during these investigations. Our initial focus was to assess the binding affinity of KH34 domains towards the selected RNA_A and RNA_B sequences. However, the titration

experiments revealed a reduced affinity towards both RNA sequences with an Kd value in low micromolar range. This observation already indicated the difference between using KH34 domains alone and the full length protein. As described in **chapter I.4.1**, both RRM domains and all four KH domains are involved in stabilization of RNA-protein interaction. Consequently, the observed loss of affinity could be explained as followed: While the KH34 domains are crucial for RNA recognition and stabilization, it is likely that the presence of other domains and their intricate interactions also play a important role. This emphasizes the importance of considering the collective contribution of all domains in the IMP2 and their coordinated actions in maintaining proper RNA-protein interactions.

Despite the decreased affinity observed with the isolated KH34 domains, we aimed to evaluate whether the three most active hit compounds (**C4, C6**, and **C9**) exhibited inhibitory activity against these domains alone as well. The results from the FP competition assay validated our previous findings with the full-length protein. The IC_{50} values obtained were consistent within low micromolar range observed in **chapter IV.1**. However, we encountered difficulty with the stability of the KH34 domains over an extended period of six months. Additionally, issues related to the expression arose, including low yield, loss of RNA binding affinity and decrease of Z-factor for the robustness of FP assay.

The observed low yield of the domains could be attributed to inappropriate expression and purification conditions, leading to solubility issues. In this case, the temperature used for purifying the KH34 domains may not have been optimal (alternating between room temperature and 4 degrees Celsius). As highlighted in **chapter I.4.1**, the anti-parallel pseudo-dimer conformation is essential for effective RNA interaction, and any disruptions in folding can impact the selectivity of RNA targets. The observed decrease in binding affinity towards the RNA sequences may occur due to the impaired folding and solubility issues of the KH34 domains.

These findings also indicate that isolated KH34 domains are sensitive, particularly regarding their stability and activity. Another possible explanation for the loss of RNA binding affinity is the potential influence of the MBP tag on the conformation of the protein. While the construct used for the isolated KH34 domains is structurally identical

to the full-length protein (as shown in the figure 19), the presence of the MBP tag may introduce conformational changes that can affect the recognition, stabilization, and binding of RNA molecules. In contrast, the additional domains present in the full-length protein may contribute to stabilizing each other and preventing conformational changes.
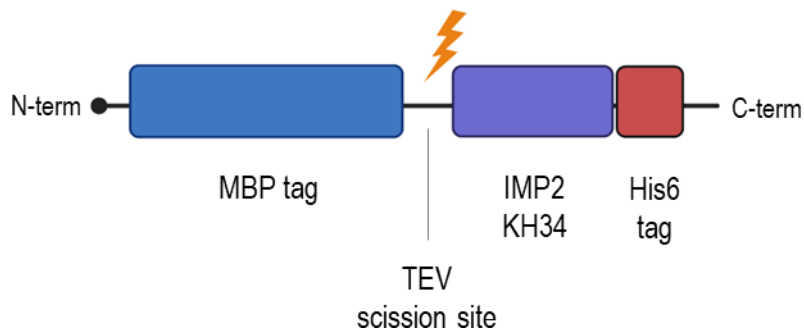

Figure 19: IMP2 protein domains (KH34) were inserted in pMAL vector. The construct contains an MBP-tag (N- terminal) and a 6XHis tag (C-terminal). The MBP-tag is followed by a TEV protease cleavage site. The expression system used was Rosetta 2 *E. coli.*

While FP assays have various advantages, they also have some limitations to consider. These biophysical assays require ligand, in our case the RNA sequence, to be labeled with a fluorescence probe. The addition of the fluorescent label can potentially alter the binding affinity or behavior of the molecule. We also encountered many times that companies were not able to provide us with the labeled RNA sequences due to difficulties in RNA synthesis. In general, FP assays can be sensitive to environmental conditions, such as pH, temperature, and ionic strength. Variations in these factors can affect the polarization signal, potentially impacting the accuracy and reproducibility of the assay.[69] This could be one of the reasons why the quality of our FP assay decreased.

However, the robustness of the FP assay mainly depends on the properties of the protein and the fluorescence probe used. Therefore, the focus of optimizations should center on these key elements to stabilize the assay's reliability, sensitivity, and accuracy. While environmental factors may still play a role and have to be taken care of, emphasizing improvements in the protein expression/purification and fluorescence probe components will be critical for achieving a robust FP assay and reliable results.

6. Concluding remarks and outlook for IMP2 project

The second part of this dissertation highlighted IMP2 as a promising anti-cancer target,

given its overexpression in cancer cells and its role in promoting tumorigenesis and tumor progression. Through a well-established FP-based screening assay, we successfully identified three active hit compounds from two inhibitor classes, namely benzamidobenzoic acid and ureidothiophene. Subsequent *in vitro* and *in vivo* assessments confirmed their target specificity, inhibitory activity, and cellular efficacy, revealing their significant therapeutic potential. Overall, these findings underscore the importance of exploring novel IMP2 inhibitor scaffolds. This leads to advancing the development of effective anti-cancer therapy.

The next crucial step in this project involves evaluating the decision to focus only on the isolated KH34 domains. To ensure comprehensive validation of the hit compounds, it may be beneficial to continue verifying them with the full length protein together with the isolated KH34 domains. Another approach worth considering is to explore the activities of the hit compounds towards the other four domains (RRM12 and KH12). This investigation could provide valuable insights into the mode of action and interactions between these domains. Another idea is to investigate the activities of the hit compounds towards other four domains. This might be helpful for further understanding of the mode of interaction between the domains.

To proceed with further (screening) experiments, it is important to optimize the expression and purification of the KH34 domains to obtain stable protein. Addressing temperature changes during purification is a crucial aspect of the optimization process, as all purification steps require a constant 4 degrees Celsius. Additionally, to investigate whether the MBP tag impacts the protein's conformation and stability, the cleavage of the tag using tobacco etch virus (TEV) protease should be conducted after the MBP purification step. Nonetheless, removing the MBP tag could potentially impact the outcomes of the FP assay due to the reduction in protein since after the MBP tag cleavage. It is important to additionally keep this in mind.

As our collaborators in Switzerland have already initiated the (co-)crystallization process, it is worth to continue with these efforts to gain a deeper insight into the binding mechanism and interactions between the full length protein or the isolated domains and the hit compounds. The structural information obtained from the (co-) crystallization experiments will provide invaluable details about the precise binding

sites and confirm the mode of action of the inhibitors. This knowledge enables us to fine-tune the compounds for enhanced efficacy and selectivity.

Another further perspective of the IMP2 project is the inclusion of IMP1 and IMP3. By incorporating these related proteins into the study, we can assess the inhibitory effects of the hit compounds across the IMP protein family, gaining a more comprehensive understanding of their selectivity. Additionally, investigating the possibility of dual-inhibitors, which can simultaneously target multiple IMP proteins, presents an potential opportunity to enhance therapeutic outcomes. This approach could open up new possibilities for developing combination therapies and broadening the scope of potential applications for the identified compounds.

References

[1] Sir Alexander Fleming **1964**.

[2] "Antimicrobial resistance", to be found under https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance, **2021**.

[3] A. R. Burmeister, *Evol Med Public Health* **2015**, *2015*, 193–194.

[4] E. M. Darby, E. Trampari, P. Siasat, M. S. Gaya, I. Alav, M. A. Webber, J. M.A. Blair, *Nature reviews. Microbiology* **2023**, *21*.

[5] J. L. Martinez, *Drug discovery today. Technologies* **2014**, *11*.

[6] Sebastian Walesch, Joy Birkelbach, Gwenaëlle Jézéquel, F P Jake Haeckl, Julian D Hegemann, Thomas Hesterkamp, Anna K H Hirsch, Peter Hammann, Rolf Müller, *EMBO reports* **2023**, *24*, e56033.

[7] K. Iskandar, J. Murugaiyan, H. D. Hammoudi, S. E. Hage, V. Chibabhai, S. Adukkadukkam, C. Roques, L. Molinier, P. Salameh, D. M. Van, *Antibiotics (Basel, Switzerland)* **2022**, *11*.

[8] Jim O'Neill, *Review on Antimicrobial Resistance* **2016**.

[9] Honor Whiteman, *Medical News Today* **2014**.

[10] Allison L. Welp, Jennifer M. Bomberger, *Front. Cell. Infec. Microbiol.* **2020**, 1–18.

[11] M. G. E.M.Selleck, *mBio* **2016**.

[12] a) Siham Hattab, Anna-Maria Dagher, Robert T. Wheeler, *Infection and Immunity* **2022**, *90*, 1–15; b) K. A. Brogden, J. M. Guthmiller, C. E. Taylor, *Lancet (London, England)* **2005**, *365*, 253–255;

[13] W. H. Tay, K. K. L. Chong, K. A. Kline, *Journal of molecular biology* **2016**, *428*, 3355–3371.

[14] John Bottrell, "Cystic Fibrosis", to be found under https://respiratorytherapycave.blogspot.com/2010/12/cystic-fibrosis.html, **2010**.

[15] Catherine Gara, "Piecing Together Cystic Fibrosis", to be found under https://www.hopkinsmedicine.org/research/advancements-in-research/fundamentals/in-depth/piecing-together-cystic-fibrosis.

[16] Oleh Admin, "Gastroenteritis - 6 Signs Of Gastroenteritis And How To Relieve It : Typical symptoms include abdominal cramps, diarrhoea and vomiting.", to be found under https://omar-vientodeleste.blogspot.com/2021/05/gastroenteritis-6-signs-of.html, **2021**.

[17] P. Zimmermann, N. Curtis, *The Journal of infection* **2019**, *79*, 471–489.

[18] P. C. Barko, M. A. McMichael, K. S. Swanson, D. A. Williams, *Journal of veterinary internal medicine* **2018**, *32*, 9–25.

[19] G. L. Trakman, S. Fehily, C. Basnayake, A. L. Hamilton, E. Russell, A. Wilson-O'Brien, M. A. Kamm, *Journal of gastroenterology and hepatology* **2022**, *37*, 237–245.

[20] WHO, "WHO publishes list of bacteria for which new antibiotics are urgently needed", to be found under https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed.

[21] Matthew B. Calvert, Varsha R. Jumde, Alexander Titz, *Beilstein Journal Organisc Chemistry* **2018**, *14*, 2607–2617.

[22] Robert Richter, Mohamed.A.M. Kamal, Mariel A. García-Rivera, Jerome Kaspar, Maximilian Junk, Walid A.M. Elgaher, Sanjay Kumar Srikakulam, Alexander Gress, Anja Beckmann, Alexander Grißmer et al., *Materials Today Bio* **2020**, *8*.

[23] M. Masi, M. Réfregiers, K. M. Pos, J.-M. Pagès, *Nature microbiology* **2017**, *2*, 17001.

[24] S.I. Miller, *mBio* **2016**, *7*.

[25] Chiara Rezzoagli, Martina Archetti, Ingrid Mignot, Michael Baumgartner, Rolf Kümmerli, *PLOS Biology* **2020**, *18*, e3000805.

[26] A. E. Clatworthy, E. Pierson, D. T. Hung, *Nature chemical biology* **2007**, *3*.

[27] Martin Empting, *Nachrichten aus der Chemie* **2018**, *66*, 290–294.

[28] Matthew B Calvert, Varsha R Jumde, Alexander Titz, *Beilstein Journal of Organic Chemistry* **2018**, *14*, 2607.

[29] A. K. Heroven, K. Böhme, P. Dersch, *RNA biology* **2012**, *9*, 379–391.

[30] V. Jakob, B. G. E. Zoller, J. Rinkes, Y. Wu, A. F. Kiefer, M. Hust, S. Polten, A. M. White, P. J. Harvey, T. Durek et al., *European journal of medicinal chemistry* **2022**, *231*, 114148.

[31] C. K. Maurer, M. Fruth, M. Empting, O. Avrutina, J. Hoßmann, S. Nadmid, J. Gorges, J. Herrmann, U. Kazmaier, P. Dersch et al., *Future medicinal chemistry* **2016**, *8*, 931–947.

[32] P. B. T. Romeo, *Microbiol. Spectrum* **2018**, *6*, 1–13.

[33] C. A. Vakulskas, A. H. Potts, P. Babitzke, B. M. M. Ahmer, T. Romeo, *Microbiol. Mol. Biol. Rev.* **2015**, *79*, 193–224.

[34] T. Weilbacher, K. Suzuki, A. K. Dubey, X. Wang, S. Gudapaty, I. Morozov, C. S. Baker, D. Georgellis, P. Babitzke, T. Romeo, *Molecular microbiology* **2003**, *48*, 657–670.

[35] C. Rife, R. Schwarzenbacher, D. McMullan, P. Abdubek, E. Ambing, H. Axelrod, T. Biorac, J. M. Canaves, H.-J. Chiu, A. M. Deacon et al., *Proteins* **2005**, *61*, 449–453.

[36] Christine Walt, Master thesis, Universität des Saarlandes, Saarbrücken, **2018**.

[37] J. Timmermans, M. L. Van, *Journal of bacteriology* **2009**, *191*.

[38] J. Timmermans, L. van Melderen, *Cell. Mol. Life Sci.* **2010**, *67*, 2897–2908.

[39] V. Jakob, Saarländische Universitäts- und Landesbibliothek, **2021**.

[40] T. Romeo, C. A. Vakulskas, P. Babitzke, *Environmental microbiology* **2013**, *15*, 313–324.

[41] Tony Romeo, *Molecular microbiology* **1998**, *29*, 1321–1330.

[42] M. Yadav, M. Srinivasan, N. K. Tulsian, Y. X. Liu, Q. Lin, I. Rosenshine, J. Sivaraman, *Protein science : a publication of the Protein Society* **2021**, *30*, 2433–2444.

[43] N. Katsowich, N. Elbaz, R. R. Pal, E. Mills, S. Kobi, T. Kahan, I., *Science* **2017**, *355*, 734–739.

[44] C. Dahlem, A. Abuhaliema, S. M. Kessler, T. Kröhler, B. G. E. Zoller, S. Chanda, Y. Wu, S. Both, F. Müller, K. Lepikhov et al., *ACS chemical biology* **2022**, *17*, 361–375.

[45] Jeetayu Biswas, Vivek L. Patel, Varun Bhaskar, Jeffrey A. Chao, Robert H. Singer, Carolina Eliscovich, *nature communications* **2019**, *10*, 1–10.

[46] J. Wang, L. Chen, P. Qiang, *Cancer cell international* **2021**, *21*.

[47] J. L. Bell, K. Wächter, B. Mühleck, N. Pazaitis, M. Köhn, M. Lederer, S. Hüttelmaier, *Cellular and molecular life sciences : CMLS* **2013**, *70*.

[48] K. Wächter, M. Köhn, N. Stöhr, S. Hüttelmaier, *Biological chemistry* **2013**, *394*.

[49] Jeffrey A. Chao, Yury Patskovsky, Vivek Patel, Matthew Levy, Steven C. Almo, Robert H. Singer, *Genes & development* **2010**, *24*, 148–158.

[50] Min Jia, Heinz Gut, Jeffrey A. Chao, *RNA* **2018**, *24*, 1659–1666.

[51] D. Weidensdorfer, Nadine Stöhr, Anne Baude, M. Lederer, Marcel Köhn, A. Schierhorn, S. Buchmeier, E. Wahle, S. Hüttelmaier, *RNA* **2008**.

[52] S. Hüttelmaier, D. Zenklusen, M. Lederer, J. Dictenberg, M. Lorenz, X. Meng, G. J. Bassell, J. Condeelis, R. H. Singer, *Nature* **2005**, *438*.

[53] Ning Dai, Joseph Rapley, Matthew Angel, M. Fatih Yanik, Michael D. Blower, Joseph Avruch, *Genes Dev.* **2011**, *25*, 1159–1172.

[54] M. D. Hall, A. Yasgar, T. Peryea, J. C. Braisted, A. Jadhav, A. Simeonov, N. P. Coussens, *Methods and applications in fluorescence* **2016**, *4*.

[55] O. D. Hendrickson, N. A. Taranova, A. V. Zherdev, B. B. Dzantiev, S. A. Eremin, *Sensors (Basel, Switzerland)* **2020**, *20*.

[56] F. Uliczka, F. Pisano, A. Kochut, W. Opitz, K. Herbst, T. Stolz, P. Dersch, *PloS one* **2011**, *6*, e20425.

[57] Aisha J. Syed, James C. Anderson, *Chem. Soc. Rev.* **2021**, *50*, 5668–5705.

[58] Cindy J. Smith, A. Mark Osborn, *FEMS Microbiology Ecology* **2009**, *67*, 6–20.

[59] D. Zucha, M. Kubista, L. Valihrach, *Cells* **2021**, *10*.

[60] Adrianne N. Edwards, Laura M. Patterson-Fortin, Christopher A. Vakulskas, Jeffrey W. Mercante, Katarzyna Potrykus, Daniel Vinella, Martha I. Camacho, Joshua A. Fields, Stuart A. Thompson, Dimitris Georgellis et al., *Molecular microbiology* **2011**, *80*, 1561–1580.

[61] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, M. B. Elowitz, *Science (New York, N.Y.)* **2005**, *307*.

[62] D. Nevozhay, R. M. Adams, K. F. Murphy, K. Josić, G. Balázsi, *Proceedings of the National Academy of Sciences* **2009**, *106*, 5123–5128.

[63] A. Becskei, L. Serrano, *Nature* **2000**, *405*, 590–593.

[64] D. N. Adamson, H. N. Lim, *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 13120–13125.

[65] A. Pannuri, C. A. Vakulskas, T. Zere, L. C. McGibbon, A. N. Edwards, D. Georgellis, P. Babitzke, T. Romeo, *Journal of bacteriology* **2016**, *198*.

[66] A. E. McKee, B. J. Rutherford, D. C. Chivian, E. K. Baidoo, D. Juminaga, D. Kuo, P. I. Benke, J. A. Dietrich, S. M. Ma, A. P. Arkin et al., *Microbial cell factories* **2012**, *11*.

[67] C. A. Vakulskas, Y. Leng, H. Abe, T. Amaki, A. Okayama, P. Babitzke, K. Suzuki, T. Romeo, *Nucleic acids research* **2016**, *44*.

[68] K. Suzuki, P. Babitzke, Kushner, SR, T. Romeo, *Genes & development* **2006**, *20*.

[69] X. Huang, A. Aulabaugh in *High Throughput Screening*, Humana Press, New York, NY, **2016**.