Free Energy Prediction Of Biomolecular Systems Using Ensembles Of Structures

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
an der Naturwissenschaftlich–Technischen Fakultät II
der Universität des Saarlandes

von

Alexander Benedix

Saarbrücken

2009

Tag des Kolloquiums: 5.11.2009

Dekan:
Univ.-Prof. Dr. rer. nat. Christoph Becher

Mitglieder des
Prüfungsausschusses:
Univ.-Prof. Dr. rer. nat A. Ott
Univ.-Prof. Dr. rer. nat. L. Santen
Dr. rer. nat. R. Böckmann
PD Dr. rer. nat. P. Huber

Abstract

Knowledge about the underlying free energy landscape of biomolecules is crucial for a basic understanding of the inner workings of proteins. Its fast and accurate calculation is indispensable for conformational analysis, structure–based protein design or for protein docking. On the one hand, existing rigorous methods like free energy perturbation or thermodynamic integration are time–consuming and cannot be used for large scans required for protein or vaccine design. On the other hand, fast treatments rely on empirical or statistical data and deliberately neglect protein flexibility and are therefore limited in accuracy.
In this thesis, a novel method for the estimation of free energy changes upon mutation is proposed combining a physical effective energy function with an efficient sampling of available conformational space. The energy function is based on physical chemistry and an efficient continuum solvent approach. It is averaged over alternative protein conformations fulfilling geometric constraints. The main advantage of our method is its inclusion of full protein flexibility, which dramatically improves the prediction quality for protein-protein binding affinities. Due to its hundredfold gain in speed with respect to conventional methods the method enables e.g. a full mutant scan of protein–protein interfaces. The method was successfully applied to the study of mutational effects on protein–protein and protein–peptide binding.

Zusammenfassung

Die Kenntnis der Freie–Energielandschaft von Proteinen ist essentiell für ein tiefergehendes Verständnis ihrer Funktionsweise. Die schnelle und präzise Bestimmung der Freien Energie ist wichtig für strukturbasierte Analysen, Proteindesign oder für das Proteindocking. Methoden wie die Freie Energie Störungsrechnung liefern physikalisch korrekte Beschreibungen, die allerdings mit hohem Rechenaufwand verbunden sind und daher für umfängliche Untersuchungen ungeeignet sind. Schnelle Methoden stützen sich dagegen auf empirische oder statistische Daten, vernachlässigen dabei die Proteinflexibilität und weisen daher eine eingeschränkte Genauigkeit auf.
Diese Arbeit stellt eine neu entwickelte Methode zur Berechnung von Freien Energieänderungen durch Mutationen vor, die eine physikalisch effektive Energiefunktion mit einer effizienten Abtastung des verfügbaren Konformationsraumes kombiniert.
Die über alternative Proteinstrukturen gemittelte Energiefunktion basiert auf physikalischer Chemie und einer effizienten Behandlung des Lösungsmittels. Der größte Vorteil unserer Methode ist die Berücksichtigung der vollen Flexibilität, welche die Vorhersagequalität von Bindungsaffinitäten deutlich steigert. Durch einen hundertfachen Geschwindigkeitszuwachs im Vergleich zu konventionellen Methoden werden Studien ermöglicht, die den vollen Mutationsraum von Protein–Protein Grenzflächen abdecken. Die Methode wurde erfolgreich angewandt zur Analyse der Protein–Protein sowie der Protein-Peptid Bindung.

Acknowledgements

I would like to thank everybody who contributed directly or indirectly to the success of this Ph.D. thesis.

Many thanks to my advisor Dr. Rainer Böckmann for entrusting me with this interesting topic, which finally lead to a Nature Methods article! Thanks also for energizing discussions, support in the last years and for giving me the opportunity to present my work at various conferences. He is also not innocent when it comes to my addiction to coffee...

I wish to thank the complete *Theoretical and Computational Membrane Biology* group for a great working atmosphere, inspiring coffee breaks, enlightening discussions and exhausting but amusing soccer matches: it was a great time!

Especially I want to mention my (former) room mates Caroline Becker, Simon Leis and Daniele Narzi.
While I was giving the finishing touches to the folding free energy estimation, Caroline started working on adopting the presented method for binding free energy predictions. The resulting procedures of this cooperation were presented and utilized in this thesis in Chapter 4. I also want to thank her for the flawless teamwork when building the Concoord/PBSA web interface in three weeks' time.
What would I have done without my *Problembär* Simon? I don't know — I should consult Simon and tell him that I don't know what I would have done without him. Maybe I should invent a room with one Simon in it to talk to about all the problems you ran along such a thesis... ah, that's it! Thanks for the tip!
A lot of questions about MHC complexes, some more about molecular dynamics simulations, and maybe even more about pKa calculations have been no problem for Daniele. *Grazie!* Apart from topics concerning this thesis I want to thank him for his tutorials in Italian dishes and curses (maybe this is also the right place for excuses to God, Jesus, his mother and various saints...) and for numerous jam sessions.

Thanks to Shirley Siu for taking a lot of nice pictures when I was in need for a photographer!

Many thanks also to Drs. Wei Gu, Michael Hutter, Sam Ansari and Tihamer Geyer from the working group of Prof. Dr. Helms. They have been of great help in particular with computer problems. Here, their advise, screw drivers, patch cables and helping hands have successfully shorten the half–life period of hardware and software bugs. I'm thanking Wei also for his support in understanding the GYF domain and teaching me to cook *Dumblings* (sorry Wei, I'm afraid I already forgot again). Thanks to Sam and Tihamer for introducing me to the Mac. Also thanks to Prof. Dr. Volkhard Helms for several discussions.

I appreciate the help and support of Prof. Dr. Bert de Groot concerning problems with his program Concoord, too. He also provided me with supplementary material concerning Concoord.

Thanks to Prof. Dr. Karin Jacobs for her seminar about job applications and for helping me with mine — I finally found a job!

To Prof. Dr. Norbert Graf and Dr. Gabriele Wevers–Donauer: I think this is the right place to say thank you.

I want to thank my parents for keeping an eye on my daughter Henrike at countless occasions so that I could keep on writing this thesis.

Not just for moral support, but also for a lot of explanations concerning biological and immunological topics that I encountered during my work I want to express my gratitude to my wife Julia.

And thanks to the smallest yet most persisting distraction imaginable — my daughter Henrike Emilia. For seven months she has been my *boss* and has kindly given me the opportunity to develop and test the Concoord/GBSA part while being asleep.

List of Figures

List of Tables

Notation

Abbreviations

cg	Conjugate Gradient
Concoord	Constrained Coordinates
Coul	Coulomb
es	electrostatic
FEP	Free Energy Perturbation
FF	Force Field
G53a6	GROMOS 53a6
GB	Generalized Born
(l–)BFGS	(low Memory) Broyden–Fletsher–Goldman–Shano algorithm
LIE	Linear Interaction Energy
LJ	Lennard–Jones
MD	Molecular Dynamics
MM	Molecular Mechanics
(MM/CC)/GBSA	(MM/Concoord)/Generalized Born Surface Area
(MM/CC)/PBSA	(MM/Concoord)/Poisson–Boltzmann Surface Area
OPLS(–AA)	Optimized Potentials for Liquid Simulations (– All Atom)
PB	Poisson–Boltzmann
PBE	Poisson–Boltzmann Equation
PPIS	Protein–Protein Interaction Surface
RF	Reaction Field
rmsf	root mean square fluctuation
SA	Surface Area
SASA	Solvent Accessible Surface Area
SDEC	Standard Deviation of the Error of Calculation
s_e	standard error
SES	Solvent Excluded Surface
steep	Steepest Descent
TI	Thermodynamical Integration
vdW	van der Waals

The symbol for standard deviation σ is also used for the standard deviation for the error of calculation (SDEC) when a mix up is precluded.

Amino Acids

One letter and three letter codes of amino acids:

Short	Abbrev.	Amino Acid
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic Acid
E	Glu	Glutamic Acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine

Short	Abbrev.	Amino Acid
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Proteins

A mutation at position `n` is abbreviated `XnY`, with `X` being the One-Letter-Code of the wild type amino acid and `Y` denoting the mutated amino acid. If needed the letter of the peptide chain used in the pdb file is placed as prefix: `C_XnY`. For example, `B_A30G` describes an alanine to glycine mutation at position 30 of chain B.
In general, the wild type is abbreviated WT and the mutant MUT, respectively.

Contents

Introduction

A wealth of different methods has been developed to shed light on the inner workings of biomolecular systems that are the basis for cellular tasks. While biochemical experiments may yield the influence of single functional groups, structure determining techniques allow the analysis of interactions between them. Computer–aided studies using molecular mechanics, quantum chemistry methods or data mining techniques supplement these studies by e.g. elucidating protein dynamics, or enabling the study of chemical reactions *in silico*. They also open up the possibility to predict the function of a protein or even its atomic structure aiming to replace difficult, time consuming and expensive experiments.
Disciplines like protein–structure analysis, structure–based protein design and protein docking rely on accurate and fast computation of protein free energies. Rigorous treatments based on physical effective energy functions involve computationally expensive methods such as free energy perturbation, which are time–consuming and are thus incompatible with the need to perform extensive scans. Commonly used fast methods, in turn, involve empirically derived scoring functions and usually do not include protein flexibility or are based on statistical potentials and are therefore highly dependent on the availability of case–dependent experimental training data. Hence, such methods are inherently limited in accuracy and applicability.
Here a structure–based, computational approach named *Concoord/Poisson–Boltzmann Surface Area* (Concoord/PBSA) for protein free energy prediction is proposed. The method can be used for both fast and quantitative estimation of the folding free energy of mutants, that is, for measuring their conformational stability and for predicting the effect of mutations on protein–protein binding affinity [1].

On a molecular scale the building blocks of life consist mainly of amino acids, nucleic acids and fatty acids [2–4]. The latter are responsible for zoning the interior of a cell and delimiting it to the outside by forming membranes. Both, DNA, storing information for the construction of other cell components, and RNA, involved in translating the genetic code to proteins and in gene regulation, consist of nucleic acids. Most of the remaining tasks occurring in a cell are covered by proteins. These working units, made up of amino acids, have many distinct functions that comprise catalysis, cell signalling, cell division, immune response, and also structural and mechanical

tasks in cell adhesion, the cytoskeleton or muscles.
Detailed knowledge about the molecular mechanism involved in the workings of a protein is the key not only to gain insight into the cellular machinery, but also opens up the lane towards a directed functional design of proteins. Also, it facilitates the rational design of drugs with defined properties.
Most of the above topics are typically addressed by experiments, e.g. by structural studies, mutation experiments, or biophysical measurements yielding, for example, binding strengths between molecules, turnover rates of enzymes, or information about flexibility and conformational changes of proteins. These studies are increasingly supplemented and partially replaced by *in silico* experiments. The capabilities of the latter experienced an unprecedented increase during the last three decades, not only due to the increase in computational speed, but also due to algorithmic development and theoretical advancements.

Proteins
Proteins consist of amino acids that form linear chains by connecting their amino and carboxyl groups via peptide bonds. Amino acid chains fold into a sequence–dependent three dimensional structure typically referred to as the native state fold. The protein fold is essential for its function. In general, twenty different natural amino acids are available for the construction of proteins. The amino acid sequence that underlies the three dimensional structure is termed the primary structure. Repeating local arrangements of hydrogen bonds between the amino acid backbone are called secondary structure. α–helices and β–sheets are most common. The overall three dimensional arrangement of secondary structure elements in the folded protein is termed tertiary structure. The quaternary structure of proteins describes the arrangement of protein complexes. Figure 1 sketches the different representations of a protein.
The basis to all functions assigned to proteins is the interaction of the protein with water molecules, ions, drugs, other proteins, nucleic acids, membranes, or other organic compounds found in the cytoplasm. The numerous activities associated with proteins are potential targets in pharmaceutical research. With the goal to magnify, reduce or inhibit the resulting outcome, the interactions are usually altered by introducing drugs to these systems.

Protein Folding and Stability
Protein folding describes the process of the amino acid chain to develop the fully functional three dimensional structure starting from an unfolded con-

formation (see Figure 2) [7]. *In vivo* folding occurs during and after ribosomal synthesis. Proteins fold in solution or with the help of chaperones. The folding process can be completed on a microsecond timescale, but it also may last for several hours depending on the system.
The state of an unfolded amino acid chain cannot be described by a single conformation, but as an ensemble consisting of random coil configurations and residual secondary and tertiary structure [9, 10]. As every amino acid side chain bears distinct chemical features such as polarity, hydrophobicity, aromaticity or protonation, interactions with each other and with its surroundings eventually fold the protein into its native state located at the minimum in free energy [11]. Most native conformations show a stabilizing protected hydrophobic core surrounded by polar or charged amino acids at the surface with their side chains reaching into the solvent or building salt bridges. The main driving force is the burying of nonpolar side chains [12]. The folding process starts with the setup of secondary structure elements — the so–called *molten globule* state [13] — followed by the arrangements of α–helices and β–sheets, and ends with the positioning of the remaining parts. Even for small molecules, the folding is not limited to a single pathway. Smaller proteins often fold in a single step, whereas larger proteins tend to have one or more intermediate states. Although being a steered process, a complete, detailed understanding of the general underlying framework for protein folding is still missing.
Comparison of the folding timescale with the number of microstates (known as the Levinthal Paradox [14]) demonstrates that the folding process is indeed directed: Assuming an unfolded state of a protein with one hundred amino acids, where every amino acid takes one of three different possible conformations (e.g. rotameric states), a total number of 3^{100} configurations can be sampled. Assuming a random search for the folded state in configurational space, and that the unfolded protein has a sampling rate of $10^{13}\,\mathrm{s}^{-1}$, the folding would last 10^{27} years. This is not only at variance with the normal time scale for folding, but also with the life span of living organisms.
While under physiological conditions protein folding is a reversible process, also the Law of Mass Action applies, i.e. both states exist in equilibrium and the reaction rates for folding and unfolding are equal. The thermodynamic stability of a reversibly folding and unfolding protein is described by the difference in Gibbs free energy between both states

$$\Delta F = F_{\text{folded}} - F_{\text{unfolded}}.$$

Anfinson's dogma [11] states that the native conformation corresponds to a minimum in Gibbs free energy in a physiological milieu.

The environment, i.e. the solvent, salt concentration, temperature, pressure, pH, and molecular crowding [15] influence protein folding. Transmembrane portions of proteins, for example, are by themselves more stable in a lipid environment than in water. Also chaperones [16] and denaturants change the folding behavior. While the former assists in folding by e.g. preventing misfolding and aggregation under stress conditions like heat exposure, denaturants disrupt the native state by pushing the equilibrium towards the unfolded state.
Misfolded proteins can lead to divergent illnesses [19]. Proteins that are able to misfold into an extraordinary stable state that is immune against digestion are called prions [20]. This unfolded state is prone to aggregation and is probably coupled to or even invokes diseases like the Creutzfeldt–Jakob disease or bovine spongiform encephalopathy (BSE). Neurodegenerative diseases like Alzheimer's or Parkinson's disease also produce misfolded proteins that aggregate to amyloid fibrils, as depicted in Figure 3. It is still unknown, whether the misfolded protein compound is the cause or the outcome of the incapability to degrade proteins in Alzheimer's disease [21].

In vivo or *in vitro* folding studies analyze the assistance of other proteins that support folding of specific proteins. In folding experiments, the protein is exposed either to high temperature, to high or low pH values or to denaturants like urea or guanidine hydrochloride. Refolding is initiated by returning to physiological conditions. With spectral analysis, e.g. circular dichroism (CD) or fluorescence spectroscopy, it is possible to measure equilibrium concentrations of the native and denatured state of the protein, or to determine kinetic (un–) folding rates in order to deduce free energy changes upon folding [22, 23].
Being able to measure free energy changes, not only the stability of specific mutants can be inspected, but also more complex questions may be addressed by introducing systematic mutations. ϕ–value analysis [24], for example, yields information about the transition state in a two state folding process. Here, every amino acid is mutated (one per studied mutant) to an equally charged, smaller residue, and, thus, functional interactions are removed. The differences in free energies are interpreted in terms of an intermediate state that is mapped onto the amino acid sequence. This method helps to elucidate folding pathways.

Binding and Affinity
Protein–ligand binding takes over a crucial role in cell signalling pathways but is also important for the regulation of physiological activities via hormones or drugs [2–4]. Also, the binding of energy carriers in biological cells,

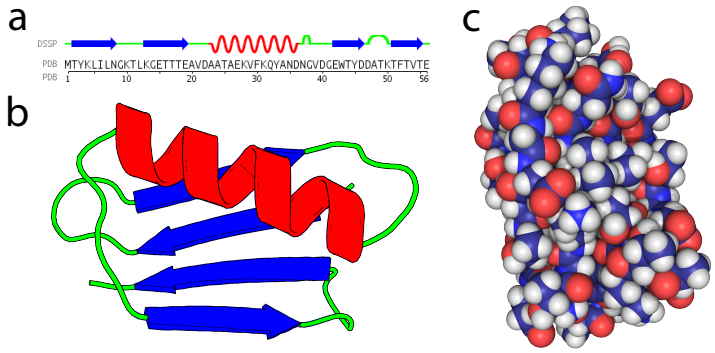

Figure 1: (a) Primary and secondary structure of the B1 immunoglobulin–binding domain of streptococcal protein G (PDB code 1PGA [5]). The illustration is taken from the protein data bank web site `http://www.pdb.org` [6].
(b) Cartoon of the secondary structure projected on the tertiary structure.
(c) Representation of the tertiary structure with atoms depicted as spheres.

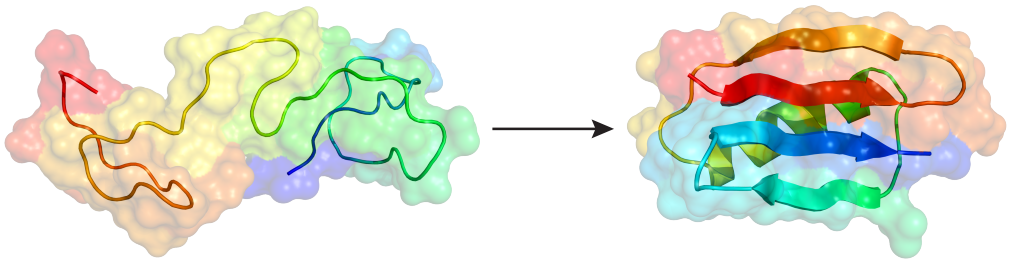

Figure 2: Denatured (left) and native (right) state of a protein (1PGA [5], unfolded structure provided by Daniele Narzi [8]) in cartoon representation.

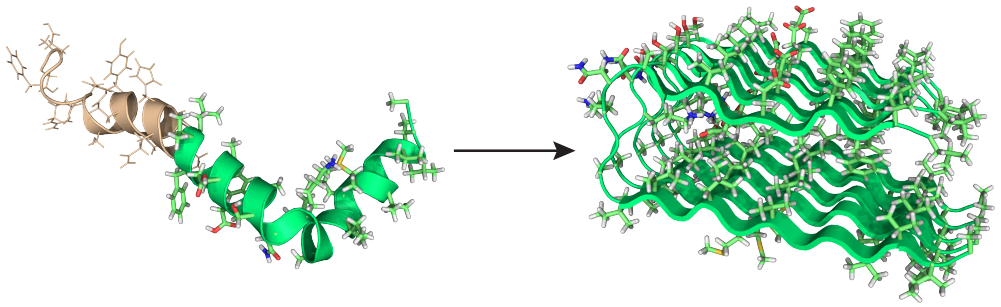

Figure 3: The green–colored fragment of the helical, native shape of the amyloid beta A4 protein (1IYT) [17] misfolds to the fibril forming beta sheets (2BEG) [18].

like ATP and GTP, is essential for the transformation of chemical energy in translational (e.g. myosin) or rotational motion (e.g. F_1F_0–ATP synthase [25]). The association between protein and ligand is usually non–covalent and reversible and may result in a conformational change. Depending on the process, binding may act as a switch enabling or disabling the functionality of a protein.

Protein–protein binding is crucial for the arrangement of functional protein complexes and thus it is the basis for quaternary structure formation, e.g. in the regulation (e.g. by protease inhibitors that inhibit protease), in triggering immune responses (binding of antibody and antigen), for motor proteins like kinesins *walking* along microtubules transporting other molecules, or in pathological protein aggregation as mentioned above. Other proteins, e.g. insulin, are fully functional in their monomeric form only but oligomerize for self–regulation.

Protein–protein complexes usually show large interaction surfaces and thus a large number of interactions between the chemical groups of both partners [26]. Moreover the binding is highly specific revealing a lock–and–key concept [27], an induced fit mechanism [28], or conformational selection [29]. Similar to the folding of a protein, no single elementary principle for protein–protein binding could be identified due to the complex network of interactions between the interacting partners. Thus, the analysis of protein–protein binding is usually performed for specific systems [30]. But, still, the binding process is energy driven, and the binding affinity between two proteins is expressed by the difference in free energy of the complex and the two separated proteins

$$\Delta F = F_{\text{complex}} - F_{\text{partner A}} - F_{\text{partner B}}.$$

While *in vivo* studies mainly aim at the identification of new interacting partners in the cell, *in vitro* experiments additionally allow the quantification of the binding affinity. Similar to the protein stability, the dissociation constant K_D is frequently measured by spectroscopic methods. The binding affinity can directly be expressed by this constant, and easily yield the binding free energy via (k_B = Boltzmann constant)

$$\Delta F = k_B \ln K_D.$$

Systematic mutation studies like the alanine scanning approach [31] yield information about how single amino acid side chains influence the binding behavior. Hereby, amino acids of interest are successively mutated to alanine.

For investigating small binding regions consisting of a short peptide sequence, spot synthesis [32] yields a qualitatively more detailed picture since

every possible single–point mutant (including every other amino acid as a mutation) is tested for binding. The tested sequences are synthesized on a cellulose membrane and brought in contact with its binding partner. The density of the bound protein to the anchored sequence is measured afterwards by chemiluminescent methods.

Structure Determination

Next to folding and binding experiments, structure determination is essential for many studies. X–ray crystallography [22, 33] as well as NMR spectroscopy [34] provide a detailed picture at atomic resolution and therefore enable a structure based analysis.

Combined with ϕ–value analysis or alanine scanning the three dimensional structure leads to an improved understanding of interatomic interactions and the folding or binding behavior of proteins.

Also, experimentally derived protein structures are the basis for accurate theoretical and computational structure–based studies (see below).

***In silico* experiments**

Investigating large biomolecular systems *in silico* is a common task these days. The accuracy in the prediction of e.g. free energy differences is restricted mainly by the available computing power and of both the precision of the protein structure and the accuracy of the underlying theoretical model. The level of *in silico* investigations is steadily broadened by the continuously increasing computational power. Not only faster CPUs, faster networks and larger memories emerge, also new technologies like multi core processing units speed up computations facilitating classical molecular mechanics studies of one million atom systems [35] or microsecond simulations of smaller biomolecular systems [36, 37]. Theoretical advancements like the mixed quantum and classical mechanics simulations enable to treat small parts of a system quantum mechanically (QM/MM) [38], replica exchange molecular dynamics (REMD) simulations [39] allow the folding of small proteins *in silico*. Even the internet contributes to large scale computations by making projects like Folding@Home [40] possible.

All biomolecular processes are governed by the underlying free energy landscape. Knowledge of the free energy landscape is the key to understand protein folding and unfolding or to predict binding affinities. Additionally, free energy predictions are crucial in protein or drug design studies and in molecular docking studies. The latter aims at identifying high–affinity binders and binding sites for drugs.

A variety of methods has been developed in the past to predict the free en-

ergy of a biomolecular system. These may be divided into structure–based methods based on physico–chemical forcefields (see below) (molecular mechanics, quantum mechanics, QM/MM), into knowledge–based methods relying on statistical, empirical data and on the protein structure only and on machine–learning approaches (that rely on empirical data). While the former are more accurate and can be used for *ab initio* predictions, the latter are computationally cheap and, therefore, can easily be applied, e.g. in large–scale mutation studies.
Molecular dynamics (MD) simulations use parametrized potentials for interatomic interactions known as force fields [41]. Iteratively solving Newton's equations of motion for all nuclei yields the positions of all atoms as a function of time, termed trajectory. With statistical ensembles as outcome, observables and thermodynamic potentials (e.g. the Gibbs free energy) are well defined from statistical thermodynamics and can be compared to experimental measurements.
Depending on the addressed problem, different simulation–based procedures are at hand for the prediction of free energy changes with varying computational effort [42–47]. Although also being titled as *computational alchemy* [43] due to their sometimes involved transitions between non–physical states, statistical mechanics methods like Free Energy Perturbation (FEP) [48, 49] or Thermodynamic Integration (TI) [50] yield the most accurate results with the lowest inherent statistical error of typically less than 1 kcal/mol [46]. Both methods slowly migrate the system from an initial state (e.g. the wild type) to a defined final state (e.g. the mutant) and directly yield the free energy difference between the two states. These methods are based on the free energy F which is related to the partition function Z via

$$F = -k_B T \ln Z.$$

Techniques like Molecular Mechanics/Poisson–Boltzmann Surface Area (MM/PBSA) [51] or Linear Interaction Energy (LIE) [52] only consider the end states (initial and final state). Here, free energy differences are based on the decomposition in enthalpic (ΔU)[1] and entropic ($T\Delta S$) contributions

$$\Delta F = \Delta U - T\Delta S,$$

which are calculated on trajectories of the end states, only. Solvation effects and changes in entropy are only approximated.
Thus, these procedures are computationally less demanding with the drawback of an enlarged error [46].

[1] U denotes the internal energy. The correct term for the enthalpy is $H = U + pV$ and is used in the Gibbs free energy G = H-TS.

All of the above techniques make use of explicit solvent MD simulations that are followed by the evaluation of physical energy functions. Depending on the system size, adequate MD simulations of a fully solvated protein complex can take several weeks on multiple processors.
The accurate prediction of free energies enables the prediction of protein stabilities or binding affinities between proteins and ligands, rendering it possible to find new drugs prohibiting, diminishing or enhancing binding, or to design e.g. peptide–based vaccines to build up immune response against tumor antigens [53]. Also, redesigning the surface of hydrophobic membrane proteins like G protein–coupled receptors with a retained function could be possible [54]. This may allow the design of water soluble analogues significantly simplifying biochemical experiments.
Due to the immense computational effort the above simulation based methods cannot be applied to extensive mutational studies. Mutating a protein in every MD snapshot instead of simulating the mutant leads to faster approaches like the computational alanine scanning [55] or the so–called virtual mutagenesis method [56], which is the computational analogue to the experimental spot synthesis. Inherently, these unrelaxed mutations neglect conformational changes and flexibility adaptions as response to the mutation that are, however, both crucial for the correct energetic prediction [57]. When studying only a small number of mutation sites, also simulations of the full single–point mutational set [58] are possible.
Knowledge–based fast methods, that neglect the structural flexibility of a biological system have been developed, too. For example, Fold-X [59] takes conformational data from one structure only and applies an empirically derived potential to obtain folding and binding free energies, and EGAD [60, 61] uses a rotamer optimized configuration and physical free energy functions for the design of novel proteins.
A third category of methods makes use of machine learning algorithms [62, 63] or simple scoring functions trained on experimental data sets. Thereby, the obtained results strongly depend on the quality of the training set. Due to the complexity of the (non–) linear functions no physical interpretions are possible. Frequently, only the amino acid sequence acts as an input for machine learning algorithms (e.g. artificial neural networks, SVM), which makes this method class also applicable to cases where no atomic structure is available.
To increase the mutant space of expensive, structure–based methods, bioinformatics methods can be combined with computational alanine scanning or the virtual mutagenesis method [58]. Using MD results of hundreds of mutations as training data for these methods, the prediction of billions to trillions of (multiple) mutants is easily achieved. Interesting results can, in

turn, be further analyzed by means of molecular mechanics.

Objectives and Organization

The main objective of this work was to develop a fast albeit structure–based technique that is capable of reproducing experimental folding or binding free energy differences upon mutation. Physical free energy functions were developed and evaluated on a structural ensemble generated using geometric restraints only, replacing time–demanding MD simulations.

Chapter 1
Chapter 1 gives a short overview of statistical mechanics, with a focus on the concept of free energy. Additionally, methods for the calculation of energies of biomolecular systems are reviewed.
Molecular mechanics force fields, representing the internal energy of a biomolecular system, as well as the methods of molecular dynamics simulation and energy minimization are introduced.
Both, MD–based statistical methods as well as continuum solvent approaches are presented. As end state calculations are in need of an explicit entropy estimate, different techniques for the approximation of the system's entropy are described. In addition to MD approaches, various other fast prediction methods based on the crystal structure alone or on protein sequences are presented.

Chapter 2
As the comparison to experiment is crucial for method development, the related experimental methods are sketched in Chapter 2. As this thesis would not be possible without three dimensional structures, the two common methods for resolving protein structures at atomic resolution — X–ray crystallography and NMR–spectroscopy — are briefly described. Besides, the experimental analogues to free energy computations with respect to protein stability and protein–protein binding are covered.

Chapter 3
In Chapter 3 of this thesis a new approach is presented for the calculation of free energies using structure ensembles. Structures are generated using Concoord [64], which samples conformational space around a given input structure using geometrical constraints, only. Thereby, the computational efficiency is increased hundredfold as compared to simulation–based approaches.

As a first test case a set of more than five hundred mutants from five different proteins was considered. The method was parameterized in order to reproduce experimental stabilities with high accuracy.
As the Concoord algorithm was used on protein structures *in vacuo*, continuum solvent approaches were required to include protein–water interactions. For the development the slow yet precise Poisson–Boltzmann approach was chosen, hence, the name Concoord/PBSA — Concoord / Poisson–Boltzmann Surface Area.
A web interface for the online calculation of free energy differences using the Concoord/PBSA method is reported. With the help of the Concoord/PBSA web interface the reproducibility and the time consumption were analyzed.

Chapter 4
Chapter 4 includes the enhancement to compute mutational effects on protein–protein and protein–lignad binding affinities [30]. This chapter also reports applications for the prediction of mutational free energy differences using Concoord/PBSA. Next to the study of protein–peptide binding of the GYF domain, as well as on the folding and dimerization of insulin, performance comparisons to other methods are presented utilizing the TEM1–BLIP complex or the p53–MDM2 complex as test cases.

Chapter 5
Conclusions and an outlook are given in Chapter 5. Further development and possible applications are discussed.

Chapter 1

Free Energy Calculations, Approximations and Predictions

1.1 Statistical Mechanics

Statistical mechanics [42–44, 65] distinguishes several thermodynamic ensembles. For example the canonical ensemble (also termed *NVT* ensemble due to constant particle number N, constant volume V and constant temperature T) is coupled to a heat bath that allows exchange of energy. Experiments are frequently performed in the *NpT* ensemble (constant pressure p instead of constant volume). In the following we will concentrate on the *NVT* ensemble. The results are easily transferable to other ensembles.
A molecular system can be described by the means of a Hamilton operator or function $H(\mathbf{p},\mathbf{q})$ of the generalized coordinates $\mathbf{q}$ and their conjugate generalized momenta $\mathbf{p}$. Given the Cartesian coordinates $\mathbf{q} = (q_1, q_2, \ldots, q_{3N})$, momenta $\mathbf{p} = (p_1, p_2, \ldots, p_{3N})$ and masses $m_1, m_2, \ldots, m_N$ of a classical system with N atoms, the Hamiltonian is given by

$$H(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{3N} \frac{p_i^2}{2m_i} + \phi(\mathbf{q}), \tag{1.1.1}$$

where $\phi(\mathbf{q})$ is the interatomic potential. $\mathbf{q}$ and $\mathbf{p}$ span the phase space Γ that holds all possible states of a system. An important quantity directly linked to all possible states is the partition function Z given by

$$Z = \iint e^{-\beta H(\mathbf{p},\mathbf{q})} \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}}, \tag{1.1.2}$$

with the minimal phase space volume h^{3N}. If treating indistinguishable particles an additional prefactor $\frac{1}{N!}$ occurs in Equation (1.1.2) [43].
Based on the energies of the microstates, the partition function yields information about the partitioning of probabilities for all possible states. Z stands for the German 'Zustandssumme', as for countable microstates it is a 'sum over states'.

The probability to find the system in a microstate (**q**, **p**) can be expressed by the density function ρ

$$\rho(\mathbf{p},\mathbf{q}) = \frac{e^{-\beta H(\mathbf{p},\mathbf{q})}}{Z}. \tag{1.1.3}$$

The integral over the whole phase space volume Γ is equal to one

$$\iint \rho(\mathbf{p},\mathbf{q}) \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}} = \frac{Z}{Z} = 1. \tag{1.1.4}$$

The observable $\langle A \rangle$ of a system property A can be written by means of its canonical ensemble average

$$\langle A \rangle = \iint \rho(\mathbf{p},\mathbf{q}) A(\mathbf{p},\mathbf{q}) \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}}. \tag{1.1.5}$$

Applied to the Hamiltonian H the ensemble average is equivalent to the internal energy U

$$U = \langle H \rangle = \iint \rho(\mathbf{p},\mathbf{q}) H(\mathbf{p},\mathbf{q}) \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}}. \tag{1.1.6}$$

The internal energy consists of the kinetic energy including translational, rotational and vibrational contributions and the potential energy due to inter– and intramolecular interactions. The Hamiltonian and its ensemble average as found in biomolecular systems will be discussed in more detail together with molecular force fields in subsection 1.3.1.

Another important quantity describing a thermodynamical system is its entropy S

$$S = -k_B \langle \ln \rho \rangle. \tag{1.1.7a}$$

The entropy is a measure for the phase space volume that is within reach of a phase trajectory at constant macroscopic conditions like in the *NVT* or the *NpT* ensembles. If only one single state ψ is accessible, $\rho_\psi = 1$ and $\rho = 0$ otherwise and, thus, $S = 0$. With more microstates ρ adopts values in the range of $0 < \rho < 1$, and as $x \ln x < 0$ applies for $0 < x < 1$ the entropy increases. Thus, the entropy may be used as a measure for the disorder of a system.

Every closed system strives for a maximum in entropy. It is increased over time until the system reaches equilibrium.

Inserting equation (1.1.3) in (1.1.7a) leads to

$$\begin{aligned} S &= -k_B \iint \rho(\mathbf{p},\mathbf{q}) \ln \rho(\mathbf{p},\mathbf{q}) \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}} \\ &= -k_B \iint \rho(\mathbf{p},\mathbf{q}) \ln \frac{e^{-\beta H(\mathbf{p},\mathbf{q})}}{Z} \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}} \\ &= -k_B \iint \rho(\mathbf{p},\mathbf{q}) \left(-\beta H(\mathbf{p},\mathbf{q}) - \ln Z\right) \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}} \\ &= \frac{1}{T}\langle H \rangle + k_B \ln Z. \end{aligned} \tag{1.1.7b}$$

The emerging term $k_B T \ln Z$ allows the connection of the statistical mechanics of microscopic systems with the (Helmholtz) free energy of (macroscopic) thermodynamics

$$F = -k_B T \ln Z. \tag{1.1.8a}$$

The free energy may also be written as a difference of the ensemble average of the internal energy and temperature times entropy

$$F = \langle H \rangle - TS = U - TS. \tag{1.1.8b}$$

While the entropy of closed systems (*NVT*, *NpT*,...) tends to its maximum, the free energy adopts its minimum in equilibrium. In an *NpT* ensemble Gibbs Free Energy

$$G = U + pV - TS \tag{1.1.9}$$

is considered.

1.2 Free Energy and Rate Constants of Biomolecular Processes

The free energy landscape determines the direction and rate of biomolecular processes. In the following, the link of the free energy to rate constants is established on the example of two–state systems.
Biomolecular systems show a minimum in free energy for the preferred state according to the second law of thermodynamics.
In chemical equilibrium, the binding of two molecules A and B to the compound AB

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{AB}$$

is described by the thermodynamic binding constant

$$K = \frac{a_{\mathrm{AB}}}{a_{\mathrm{A}} a_{\mathrm{B}}}, \tag{1.2.1}$$

where a is the chemical activity. The chemical activity a can be approximated at low concentrations (or when other perturbing reactions can be neglected) with the dimensionless concentration

$$a \approx \frac{c}{1\frac{mol}{l}}. \tag{1.2.2}$$

As c_{AB} is the concentration for the bound state and the product $c_{\mathrm{A}} c_{\mathrm{B}}$ a measure for all possible combinations for the unbound molecules, their ratio equals the ratio of the states' partition functions

$$K = \frac{c_{\mathrm{AB}}}{c_{\mathrm{A}} c_{\mathrm{B}}} = \frac{Z_{\mathrm{AB}}}{Z_{\mathrm{A+B}}}. \tag{1.2.3}$$

Thus, this relation can be used to calculate the difference in free energy between the two states

$$\begin{aligned} \Delta F = F_{\mathrm{AB}} - F_{\mathrm{A+B}} &= -k_B T \ln Z_{\mathrm{AB}} - (-k_B T \ln Z_{\mathrm{A+B}}) \\ &= -k_B T \ln \frac{Z_{\mathrm{AB}}}{Z_{\mathrm{A+B}}} = -k_B T \ln K. \end{aligned} \tag{1.2.4}$$

This relation couples the microscopic free energy difference to the thermodynamic binding constant accessible by biochemical experiments. It allows to study protein stability, solubility, protein–ligand and protein–protein binding (see Figure 1.1), aggregation, conformational changes or the protonation of a protein by analysis of the microscopic initial and final states. Here, the interesting question arises how these properties change upon a perturbation of the system, either by the introduction of a mutation, the insertion of a drug, or by a change in environment. The change in free energy differences upon mutation is of special interest for the studied systems in this work.
In addition to the end states, free energies are often evaluated along a reaction path (potential of mean force, PMF, subsection 1.4.3). Here, e.g. the permeability and ion flux of an ion channel can be investigated (see Figure 1.1).
Estimates of observables of biomolecular systems from microscopic structures require the generation of proper statistical ensembles of the system (see e.g. Equation (1.1.5)). Often these are obtained via Monte Carlo or Molecular Dynamics simulations (see Section 1.3.2) of the biomolecular system. Since the free energy is a function of state, i.e. it vanishes along closed paths,

also unphysical pathways may be exploited in order to evaluate the free energy difference between physical states.

1.3 Internal Energy, Hamiltonian and its Evolution In Time

The analysis of the properties of a biomolecular system requires knowledge about the respective statistical ensembles. This is determined by the partition function and thus by the Hamiltonian of the system.
The most rigorous treatment would involve a full inclusion of quantum mechanics [65]. However, solving the Schrödinger equation is limited to small systems consisting only of a few particles. Approximations of the time–dependent Schrödinger equation lead to a classical mechanics approach, termed Molecular Mechanics (MM), where the atoms are treated as classical point masses moving in a semi–empirical potential according to Newton's equations of motion. Thereby, the nucleic and electronic degrees of freedom are separated (Born–Oppenheimer approximation). Typical sizes of biomolecular systems using Molecular Mechanics range between 10,000 and 1,000,000 atoms [35].

1.3.1 Molecular Mechanics Force Fields

In molecular dynamics (MD) simulations the interatomic interaction potentials are approximated by so called force field energy functions. They consist of a set of functions for the potential energies and the parameters used in this functions. The potentials are usually divided into bonded and non–bonded interactions. The first type accounts for chemical bonds, angles between bonds and dihedral angles that describe the rotation around bonds. For the non–bonded interactions the Pauli repulsion and the van der Waals attraction are approximated by e.g. a Lennard–Jones potential. Additionally, atoms interact via their partial atomic charges (Coulomb interaction).
As an example, the OPLS–AA (optimized potentials for liquid simulations – all atom) [68] force field is presented. The potential energy function

$$\phi_{\mathrm{FF}}(\mathbf{r}) = E_{\mathrm{bond}} + E_{\mathrm{angle}} + E_{\mathrm{dihedral}} + E_{\mathrm{LJ}} + E_{\mathrm{Coulomb}} \tag{1.3.1}$$

contains a bond stretching term (see Figure 1.2 a), a1))

$$E_{\mathrm{bond}} = \sum_{ij} \frac{1}{2} k_b^{ij} (r_{ij} - r_{ij}^0)^2 \tag{1.3.2}$$

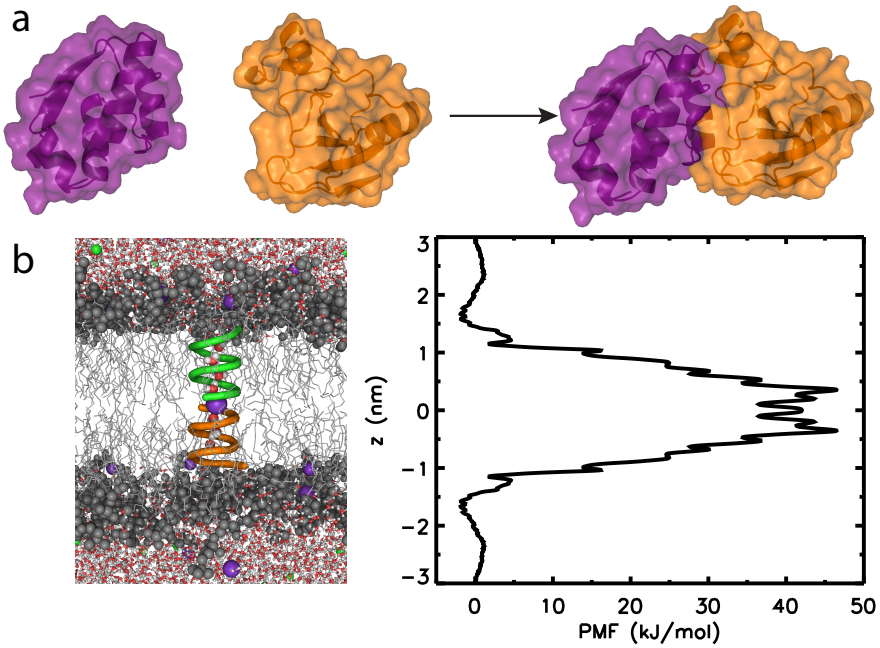

Figure 1.1: Examples for linking free energy to biology:
(a) Protein-Protein binding of Barnase-Barstar (1BRS) [66].
(b) Ion permeation through a membrane channel with the corresponding free energy profile. Coordinates and PMF-data provided by Shirley Siu [67].

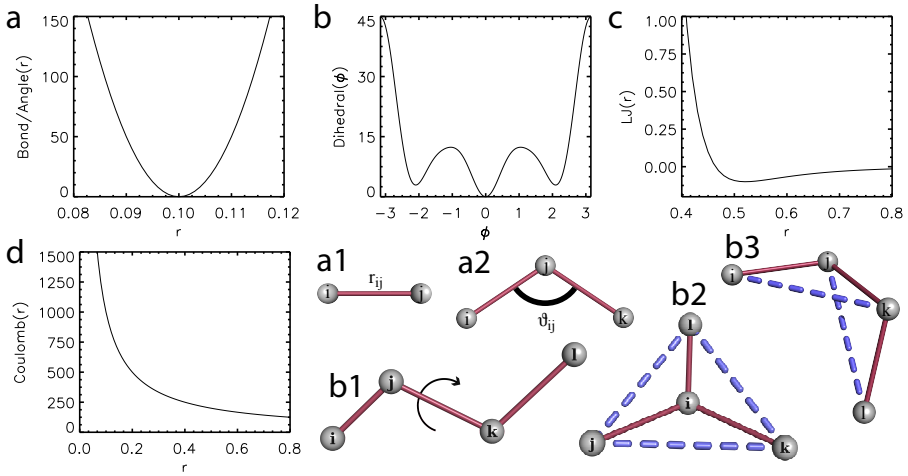

Figure 1.2: Force Field Contributions: The harmonic potential (a) is used to describe bonded atoms (a1) and angle bending (a2), the Ryckaert-Bellemans potential (b) for proper (b1) and improper (b2/3) dihedrals, and the Lennard-Jones (c) and Coulomb (d) potentials for non-bonded pairs.

and an angle bending contribution (see Figure 1.2 a), a2))

$$E_{\text{angle}} = \sum_{ijk} \frac{1}{2} k_a^{ijk} (\vartheta_{ijk} - \vartheta_{ijk}^0)^2 \tag{1.3.3}$$

between two bonds, both being approximated by a harmonic potential function.
Besides two– and three–body interactions, also four–body interactions are considered using the Ryckaert–Bellemans potential [69]

$$E_{\text{dihedral}} = \sum_{ijkl} \sum_{n=0}^{5} (C_n \cos(\phi_{ijkl}))^n. \tag{1.3.4}$$

The dihedral angle ϕ_{ijkl} is defined as the angle between the planes (i,j,k) and (j,k,l). In addition to the normal, proper dihedral interaction, also improper dihedrals are considered. They keep aromatic rings planar and conserve chirality (Figure 1.2 b), b1), b2), b3)).
The nonbonded terms include the Coulombic potential

$$E_{\text{Coul}} = \sum_{ij} \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{\varepsilon_r r_{ij}}, \tag{1.3.5}$$

with the dielectric permittivity of vacuum ε_0, and the Lennard-Jones potential

$$E_{\text{LJ}} = \sum_{ij} 4\varepsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right), \tag{1.3.6}$$

where ε_{ij} and σ_{ij} depend on the considered atoms i and j (Figure 1.2). The potential energy (Equation (1.3.1)) together with the kinetic energy $\sum_{i=1}^{N} \frac{p_i^2}{2m_i}$ these energy contributions approximate the true Hamiltonian of the system.
Other popular force fields are e.g. the AMBER force field [70], GROMOS [71], or the CHARMM FF [72]. While AMBER, CHARMM and OPLS–AA describe every atom explicitly, OPLS–UA [73] or GROMOS only implicitly consider non–polar hydrogen atoms by so–called united atoms.
The above force fields all serve the same purpose, but use slightly different functions, parameterizations and approximations for the interatomic interactions in order to obtain global agreement with known observables.
The parameterization for the chosen interaction functions of a force field is done by fitting to several molecular properties like geometric characteristics, dynamical behavior, dielectric permittivity or to energetic transitions for small molecules against experimentally derived or quantum-mechanically

computed data [71]. Non–bonded interactions for OPLS–AA were parameterized using Monte Carlo simulations and bonded parameters were taken from the AMBER FF [70], which are based on combinations of experiments and MD simulations. The GROMOS force field is based on free enthalpies of hydration and apolar solvation [71].
In general, these force fields can be taken for any quantity of a biomolecular system. But also force fields serving a special purpose have been developed, e.g. the Egad FF [61] for energy minimization of rotamers and the evaluation of free energies.
The above force fields allow to generate statistical ensembles via MD or MC simulation, and thereby e.g. to analyze free energy differences between different states of a system.

1.3.2 Molecular Dynamics Simulations

With the method of molecular dynamics simulation it is possible to follow the time evolution of a system in order to obtain an *NpT*, *NVT*, or other ensembles. In the MD simulation, the trajectory is determined by solving Newton's equations of motion

$$\vec{F}_i = m_i \frac{\partial^2}{\partial t^2} \vec{r}_i. \tag{1.3.7}$$

The forces $\vec{F}_i$ on atom i are the negative derivatives of the potential function $\phi(\vec{r}_1, \vec{r}_2, \ldots \vec{r}_N)$

$$\vec{F}_i = -\frac{\partial}{\partial \vec{r}_i} \phi(\vec{r}_1, \vec{r}_2, \ldots \vec{r}_N) \tag{1.3.8}$$

with respect to the position $\vec{r}_i$ of atom i with mass i.
The numerical solution of Equations (1.3.7, 1.3.8) typically requires a time step of 1 to 5 fs. Reaching microsecond time scale is thus computationally demanding and takes several weeks to years depending on the system size and the amount of processors available.

Leap Frog Algorithm

A popular algorithm for the integration of equations (1.3.7, 1.3.8) in time is obtained as follows: The second–order differential equation (1.3.7) can be split into two first–order differential equations resulting in expressions for the particle positions $\vec{r}_i$ and velocities $\vec{v}_i$

$$\frac{\partial}{\partial t}\vec{v}_i = \frac{\vec{F}_i}{m_i} \tag{1.3.9a}$$

$$\frac{\partial}{\partial t}\vec{r}_i = \vec{v}_i. \tag{1.3.9b}$$

Using the Taylor expansion of $\vec{r}_i$ for a small time step Δt yields

$$\vec{r}_i(t+\Delta t) = \vec{r}_i(t) + \frac{d\vec{r}_i}{dt}\Delta t + \frac{1}{2}\frac{d^2\vec{r}_i}{dt^2}\Delta t^2 + \mathcal{O}(\Delta t^3) \tag{1.3.10a}$$

$$= \vec{r}_i(t) + \vec{v}_i(t)\Delta t + \frac{1}{2}\frac{\vec{F}_i}{m_i}\Delta t^2 + \mathcal{O}(\Delta t^3) \tag{1.3.10b}$$

$$= \vec{r}_i(t) + \left(\vec{v}_i(t) + \frac{\vec{F}_i}{m_i}\frac{\Delta t}{2}\right)\Delta t + \mathcal{O}(\Delta t^3). \tag{1.3.10c}$$

Here, the terms in brackets are the first two orders of the Taylor expansion of $\vec{v}_i$ at the time point $t + \frac{\Delta t}{2}$

$$\vec{v}_i\left(t+\frac{\Delta t}{2}\right) = \vec{v}_i(t) + \frac{d\vec{v}_i}{dt}\frac{\Delta t}{2} + \frac{1}{2}\frac{d^2\vec{v}_i}{dt^2}\frac{\Delta t}{2}^2 + \mathcal{O}(\Delta t^3) \tag{1.3.11a}$$

$$= \vec{v}_i(t) + \frac{\vec{F}_i}{m_i}\frac{\Delta t}{2} + \mathcal{O}(\Delta t^2). \tag{1.3.11b}$$

Substituting equation (1.3.11b) in (1.3.10c) yields

$$\vec{r}_i(t+\Delta t) = \vec{r}_i(t) + \vec{v}_i\left(t+\frac{\Delta t}{2}\right)\Delta t + \mathcal{O}(\Delta t^3), \tag{1.3.12}$$

which is used to calculate $\vec{r}_i(t+\Delta t)$ using $\vec{r}_i(t)$ and $\vec{v}_i\left(t+\frac{\Delta t}{2}\right)$.
While $\vec{r}_i(t+\Delta t)$ is evaluated at times $(t+\Delta t)$ the particle's velocities are calculated in time steps $\left(t+\frac{\Delta t}{2}\right)$ in between. The corresponding formula is obtained by a second Taylor expansion at time $\left(t-\frac{\Delta t}{2}\right)$

$$\vec{v}_i\left(t-\frac{\Delta t}{2}\right) = \vec{v}_i(t) - \frac{d\vec{v}_i}{dt}\frac{\Delta t}{2} + \frac{1}{2}\frac{d^2\vec{v}_i}{dt^2}\frac{\Delta t}{2}^2 + \mathcal{O}(\Delta t^3) \tag{1.3.13a}$$

$$= \vec{v}_i(t) - \frac{\vec{F}_i}{m_i}\frac{\Delta t}{2} + \mathcal{O}(\Delta t^2). \tag{1.3.13b}$$

Solving equation (1.3.13b) for $\vec{v}_i(t)$ and substituting in (1.3.11) yields

$$\begin{aligned}\vec{v}_i\left(t+\frac{\Delta t}{2}\right) &= \vec{v}_i\left(t-\frac{\Delta t}{2}\right)+\frac{d\vec{v}_i}{d\,t}\Delta t+\mathcal{O}(\Delta t^2)\\ &= \vec{v}_i\left(t-\frac{\Delta t}{2}\right)+\frac{\vec{F}_i}{m_i}\Delta t+\mathcal{O}(\Delta t^2),\end{aligned} \tag{1.3.14}$$

the second equation for the leap frog scheme. The name comes from the alternating calculation of positions and velocities.

Temperature and Pressure Coupling

Using the above equations should ideally lead to a microcanonical ensemble with constant particle number N, constant volume V and constant energy E. However, in most cases it is more convenient for the studied system to use temperature and pressure as independent variables in favor of volume or energy. This can be realized using different temperature or pressure coupling approaches [74–77].

Constraints

In order to allow larger time steps when simulating, fast bond vibrations are often removed by using constraints especially on the length of covalent bonds to hydrogen atoms.
Two commonly applied constrain algorithms are SHAKE [78] and LINCS [79]. Although not contributing to the free energy itself, the SHAKE method is presented here, as a similar procedure is used later on for random structure generation.
The goal is to correct a normal, unconstrained set of coordinates $\vec{r}_i\,'$ to a new set $\vec{r}_i$ that fulfills previously defined constraints σ_k

$$\sigma_k(\vec{r}_1,\vec{r}_2,\ldots\vec{r}_N)=0 \qquad (k=1,\ldots K), \tag{1.3.15}$$

for example

$$\sigma_k(\vec{r}_\alpha,\vec{r}_\beta)=(\Delta\vec{r}_k)^2-d_k^2=0, \tag{1.3.16}$$

with

$$\Delta\vec{r}_k=\vec{r}_{k\alpha}-\vec{r}_{k\beta}. \tag{1.3.17}$$

Here, $k\alpha$ and $k\beta$ denote the two particles α and β involved in the constraint k.
Introducing Lagrange multipliers leads to the altered equations of motion

$$m_i \frac{\partial^2}{\partial t^2}\vec{r}_i = -\frac{\partial}{\partial \vec{r}_i}\phi(\vec{r}_1,\vec{r}_2,\ldots\vec{r}_N) - \sum_{k=1}^{K}\lambda_k \frac{\partial}{\partial \vec{r}_i}\sigma_k, \tag{1.3.18}$$

with

$$\frac{\partial}{\partial \vec{r}_i}\sigma_k = 2\Delta\vec{r}_k\left(\delta_{k\alpha,i} - \delta_{k\beta,i}\right), \tag{1.3.19}$$

where $\delta_{a,b}$ denotes the Kronecker delta.
Integrating twice with respect to time using the leap frog scheme yields a corrected displacement

$$\vec{r}_i(t+\Delta t) = \underbrace{\vec{r}_i(t) + \vec{v}_i\left(t+\frac{\Delta t}{2}\right)\Delta t}_{=\vec{r}_i{}'(t+\Delta t)} - 2\frac{(\Delta t)^2}{m_i}\sum_{k=1}^{K}\lambda_k\Delta\vec{r}_k\left(\delta_{k\alpha,i} - \delta_{k\beta,i}\right). \tag{1.3.20}$$

As the correction due to constraints have to satisfy

$$\sigma_k(\vec{r}_\alpha,\vec{r}_\beta,t+\Delta t) = \left(\vec{r}_\alpha(t+\Delta t) - \vec{r}_\beta(t+\Delta t)\right)^2 - d_k^2 = 0, \tag{1.3.21}$$

inserting (1.3.20) yields K equations that have to be solved for the Lagrange multipliers λ_k

$$\begin{aligned}\sigma_k(t+\Delta t) = \\ \left(\Delta\vec{r}_k{}'(t+\Delta t) - 2(\Delta t)^2\sum_{l=1}^{K}\lambda_l\Delta\vec{r}_l\left[\frac{\left(\delta_{l\alpha,k\alpha}-\delta_{l\beta,k\alpha}\right)}{m_{k\alpha}} - \frac{\left(\delta_{l\alpha,k\beta}-\delta_{l\beta,k\beta}\right)}{m_{k\beta}}\right]\right)^2 \\ - d_k^2 = 0,\end{aligned} \tag{1.3.22}$$

with $\Delta\vec{r}_k{}'(t+\Delta t) = \vec{r}_{k\alpha}{}'(t+\Delta t) - \vec{r}_{k\beta}{}'(t+\Delta t)$.
Two approximations made in the SHAKE algorithm lead to an iterative procedure to find a constrained solution for the coordinates [78, 80]. First, the quadratic terms in the Lagrange multipliers are neglected linearizing the system of equations. Second, the K constraints are treated independently of each other by assuming that the atoms involved in constraint k do not contribute to any other constraint. Thus, equation (1.3.22) becomes

$$\begin{aligned}\sigma_k(t+\Delta t) &= \left(\Delta\vec{r}_k{}'(t+\Delta t)\right)^2 \\ &- 4\Delta\vec{r}_k{}'(t+\Delta t)\Delta\vec{r}_k(t)(\Delta t)^2\lambda_k\left[\frac{1}{m_{k\alpha}} + \frac{1}{m_{k\beta}}\right] - d_k^2 = 0.\end{aligned} \tag{1.3.23}$$

And the multipliers are given by

$$\lambda_k = \frac{\left(\Delta\vec{r}_k\,'(t+\Delta t)\right)^2 - d_k^2}{4\Delta\vec{r}_k\,'(t+\Delta t)\Delta\vec{r}_k(t)(\Delta t)^2\left[\frac{1}{m_{k\alpha}}+\frac{1}{m_{k\beta}}\right]}, \tag{1.3.24}$$

with the corresponding coordinate displacements

$$\vec{r}_{k\alpha}(t+\Delta t) = \vec{r}_{k\alpha}\,'(t+\Delta t) - 2(\Delta t)^2\lambda_k\frac{\Delta\vec{r}_k(t)}{m_{k\alpha}} \tag{1.3.25}$$

and

$$\vec{r}_{k\beta}(t+\Delta t) = \vec{r}_{k\beta}\,'(t+\Delta t) + 2(\Delta t)^2\lambda_k\frac{\Delta\vec{r}_k(t)}{m_{k\beta}}. \tag{1.3.26}$$

These formulas are taken as iteration scheme using the corrected coordinates $\vec{r}_{k\alpha}$ and $\vec{r}_{k\beta}$ as unconstrained input for the next constraints or the next iteration step. Thus, it is possible that already fulfilled constraints become violated again. The iteration cycle is terminated as soon as the constraints lie within a given tolerance ε

$$\frac{|\Delta\vec{r}_k(t+\Delta t) - d_k|}{d_k} \leq \varepsilon. \tag{1.3.27}$$

Simulation Suites

Freely available software packages for MD simulations include GROMACS (available at `http://www.gromacs.org`) [81–83], NAMD (available at `http://www.ks.uiuc.edu/Research/namd/`) [84] or TINKER [85] (`http://dasher.wustl.edu/tinker/`). Gromos [86], Amber [87] and CHARMm [72, 88] are examples for commercial simulation suites.

1.3.3 Energy Minimization

In addition to propagate a system's trajectory, it is also possible with a force field at hand to find local energy minima. Minimizations are applied to remove e.g. van der Waals overlaps or distortions in crystal structures, or to optimize mutated crystal structures. In the following we shortly discuss the widely applied methods for the energy minimization.

Steepest Descent

The negative derivative of the energy function with respect to the atom's coordinates results in the forces acting on every single atom

$$\vec{F}_i = -\vec{\nabla}_i \phi. \tag{1.3.28}$$

Small displacements along the gradient of the potential

$$\begin{aligned} \Delta\vec{r}_i &= -\varepsilon\vec{\nabla}_i\phi && \text{(1.3.29a)} \\ &= \varepsilon\vec{F}_i && \text{(1.3.29b)} \end{aligned}$$

with step size ε give a small step towards the local energy minimum. Repeating this procedure on the previous step is called the steepest descent algorithm [89]. The iteration stops after a fixed number of calculations or when a convergence criteria is reached, typically a given maximum force that is allowed.
Using the gradient vector $\mathbf{g}$

$$\mathbf{g} = \left(\frac{\partial\phi}{\partial q_1}, \quad \frac{\partial\phi}{\partial q_2}, \quad \cdots \quad \frac{\partial\phi}{\partial q_{3N}} \right)^T \tag{1.3.30}$$

which contains the first derivatives of the potential function ϕ with respect to the generalized coordinates $\mathbf{q}$,the steepest descent algorithm can be written in the form

$$\mathbf{q}_{i+1} = \mathbf{q}_i - \varepsilon_i \mathbf{g}_i. \tag{1.3.31}$$

Here, the gradient vector $\mathbf{g}_i$ is evaluated at iteration step i. The step size ε_i can be held constant or determined per iteration step by different means. While the former converges slower, the latter method is computationally more demanding.

Conjugate Gradient

The conjugate gradient method [89] uses a displacement that is a linear combination of the gradient and the previous displacement

$$\Delta\mathbf{q}_i = \varepsilon_i \left(-\frac{\mathbf{g}_i}{|\mathbf{g}_i|} + \gamma_i \Delta\mathbf{q}_{i-1} \right) \tag{1.3.32}$$

with

$$\gamma_i = \frac{\mathbf{g}_i^T \mathbf{g}_i}{\mathbf{g}_{i-1}^T \mathbf{g}_{i-1}}. \tag{1.3.33}$$

The conjugate gradient implementation in the Gromacs package utilizes one steepest descent optimization step every n conjugate gradient steps to ensure a fast convergence.

Newton-Raphson Optimization and l-BFGS

Next to the simple steepest descent and the conjugate gradient method, the limited memory Broyden-Fletcher-Goldfarb-Shanno (l-BFGS) technique [90, 91] is presented here, as it is used later on in this thesis.
While the gradient $\mathbf{g}$ holds the first derivatives of the potentials, the Hessian matrix $\mathbf{H}$ consists of the second derivatives

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 \phi}{\partial q_1^2} & \frac{\partial^2 \phi}{\partial q_1 \partial q_2} & \cdots & \frac{\partial^2 \phi}{\partial q_1 \partial q_{3N}} \\ \frac{\partial^2 \phi}{\partial q_1 \partial q_2} & \frac{\partial^2 \phi}{\partial q_2^2} & \cdots & \frac{\partial^2 \phi}{\partial q_2 \partial q_{3N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \phi}{\partial q_1 \partial q_{3N}} & \frac{\partial^2 \phi}{\partial q_2 \partial q_{3N}} & \cdots & \frac{\partial^2 \phi}{\partial q_{3N}^2} \end{pmatrix}. \tag{1.3.34}$$

The quadratic Taylor expansion of the potential energy ϕ around the generalized coordinates $\mathbf{q}_j$ can thus be written as

$$\phi(\mathbf{q}_j + \Delta\mathbf{q}_j) = \phi(\mathbf{q}_j) + \mathbf{g}_j^T \Delta\mathbf{q}_j + \frac{1}{2} \Delta\mathbf{q}_j^T \mathbf{H}_j \Delta\mathbf{q}_j, \tag{1.3.35}$$

where j denotes the iteration number of the optimization and

$$\mathbf{q}_{j+1} = \mathbf{q}_j + \Delta\mathbf{q}_j. \tag{1.3.36}$$

The gradient vector $\mathbf{g}_j$ and the Hessian $\mathbf{H}_j$ are both evaluated at $\mathbf{q}_j$. As the Hessian is symmetric for twice continuously differentiable functions (which is assumed to be the case here), the derivative of equation (1.3.35) reads

$$\nabla\phi(\mathbf{q}_j + \Delta\mathbf{q}_j) = \mathbf{g}_j + \mathbf{H}_j \Delta\mathbf{q}_j. \tag{1.3.37}$$

If we assume that $\mathbf{q}_{j+1}$ minimizes ϕ, equation (1.3.37) becomes zero

$$0 = \mathbf{g}_j + \mathbf{H}_j \Delta\mathbf{q}_j. \tag{1.3.38}$$

and we obtain the iterative formula used in the Newton-Raphson optimization [89]

$$\Delta\mathbf{q}_j = -\mathbf{H}_j^{-1}\mathbf{g}_j. \tag{1.3.39}$$

As it is computationally expensive to calculate the inverse Hessian, the main concept of the so called quasi-Newton methods is to approximate the inverse Hessian iteratively using matrices fulfilling

$$\lim_{j\to\infty} \mathbf{A}_j = \mathbf{H}^{-1}. \tag{1.3.40}$$

Based on the Broyden-Fletcher-Goldfarb-Shanno (BFGS) [89, 90] algorithm

$$\mathbf{A}_{\mathrm{j+1}} = \underbrace{\left(\mathbf{I} - \frac{\Delta\mathbf{g}_k \Delta\mathbf{q}_k^T}{\Delta\mathbf{g}_k^T \Delta\mathbf{q}_k}\right)^T}_{=\mathbf{U}_k^T} \mathbf{A}_j \underbrace{\left(\mathbf{I} - \frac{\Delta\mathbf{g}_k \Delta\mathbf{q}_k^T}{\Delta\mathbf{g}_k^T \Delta\mathbf{q}_k}\right)}_{=\mathbf{U}_k} + \underbrace{\frac{\Delta\mathbf{q}_k \Delta\mathbf{q}_k^T}{\Delta\mathbf{g}_k^T \Delta\mathbf{q}_k}}_{=\mathbf{V}_k}, \tag{1.3.41}$$

Nocedal introduced the limited memory BFGS updating scheme [90, 91]

$$\begin{aligned}
\mathbf{A}_{\mathrm{j+1}} &= \left(\mathbf{U}_k^T \cdot \mathbf{U}_{k-1}^T \cdots \mathbf{U}_{k-m}^T\right) \mathbf{A}_0 \left(\mathbf{U}_{k-m} \cdot \mathbf{U}_{k-m+1} \cdots \mathbf{U}_k\right) \\
&+ \left(\mathbf{U}_k^T \cdot \mathbf{U}_{k-1}^T \cdots \mathbf{U}_{k-m+1}^T\right) \frac{\Delta\mathbf{q}_{k-m} \Delta\mathbf{q}_{k-m}^T}{\Delta\mathbf{g}_{k-m}^T \Delta\mathbf{q}_{k-m}} \left(\mathbf{U}_{k-m+1} \cdot \mathbf{U}_{k-m+2} \cdots \mathbf{U}_k\right) \\
&+ \left(\mathbf{U}_k^T \cdot \mathbf{U}_{k-1}^T \cdots \mathbf{U}_{k-m+2}^T\right) \frac{\Delta\mathbf{q}_{k-m+1} \Delta\mathbf{q}_{k-m+1}^T}{\Delta\mathbf{g}_{k-m+1}^T \Delta\mathbf{q}_{k-m+1}} \left(\mathbf{U}_{k-m+2} \cdot \mathbf{U}_{k-m+3} \cdots \mathbf{U}_k\right) \\
&\vdots \\
&+ \frac{\Delta\mathbf{q}_k \Delta\mathbf{q}_k^T}{\Delta\mathbf{g}_k^T \Delta\mathbf{q}_k}
\end{aligned} \tag{1.3.42}$$

using only information of the last m steps. Instead of solving the inverse Hessian explicitly, an efficient algorithm for calculating the product of the inverse Hessian and the gradient $\mathbf{H}_j^{-1}\mathbf{g}_j$ is used. As initial guess for the inverse Hessian a scaled identity matrix $\mathbf{I}$ is used.

1.4 Computational Alchemy & Statistical Physics Methods

Various methods have been developed in the past to compute free energy (differences) of biomolecular systems based on statistical thermodynamics (Equation (1.1.8a)). In the following we will briefly discuss methods for the calculation of the free energy differences between an initial and final state of a system, as well as for the computation of free energy along a reaction path (PMF).

1.4.1 Thermodynamic Integration

For two well defined states A and B a continuos coupling parameter λ can be introduced with $\lambda = 0$ describing state A, $\lambda = 1$ state B, respectively, and $0 < \lambda < 1$ for states in between. Thus, the Hamiltonian H of the end states can be described with λ such that $H_\mathrm{A} = H(\lambda = 0)$ and $H_\mathrm{B} = H(\lambda = 1)$ while passing smoothly from H_A to H_B with λ varying continuously from 0 to 1. The coupling parameter λ can be chosen in a way that describes non-physical, alchemical reactions like making a methanol molecule vanish while an ethane molecule appears or the mutagenesis of whole amino acids. The Hamiltonian H along the path is given by

$$H(\lambda) = (1-\lambda)H_\mathrm{A} + \lambda H_\mathrm{B}. \tag{1.4.1}$$

Depending on the addressed problem also thermodynamic variables like the temperature, pressure or a spatial coordinate can be chosen as a coupling parameter λ.
As the Hamiltonian is a function of λ, also the free energy depends on the coupling parameter, and the change in free energy is given as an integral over the first derivative of Equation (1.1.8a) with respect to λ

$$\begin{aligned}
\frac{d}{d\lambda}F(\lambda) &= -k_B T \frac{\partial}{\partial\lambda} \ln \iint e^{-H(\mathbf{p},\mathbf{q},\lambda)/k_B T} \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}} \\
&= \frac{\iint \frac{\partial}{\partial\lambda} H(\mathbf{p},\mathbf{q},\lambda) e^{-H(\mathbf{p},\mathbf{q},\lambda)/k_B T} \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}}}{\iint e^{-H(\mathbf{p},\mathbf{q},\lambda)/k_B T} \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}}} \\
&= \left\langle \frac{\partial}{\partial\lambda} H(\mathbf{p},\mathbf{q},\lambda) \right\rangle_\lambda
\end{aligned} \tag{1.4.2}$$

and yields

$$\Delta F = F(\lambda_B) - F(\lambda_A) = \int_{\lambda_A}^{\lambda_B} \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda. \tag{1.4.3}$$

Evaluating the ensemble average of the derivative of the Hamiltonian with respect to λ (1.4.2) from MD simulations carried out at a series of λ values between 0 and 1 and subsequent numerical integration (of Equation (1.4.3)) yields the free energy change between the initial (A) and the final (B) state. This method is called *thermodynamic integration* (TI) [50]. In the *slow growth* method the system is changed in a single simulation with a continuously varying coupling parameter from the initial to the final state [92, 93]. Here the problem arises that the system is never in equilibrium as the Hamiltonian is changing in every time step. Recent applications of TI can be found in Rodriguez et al. [94] and Schwab et al. [95]. TI with extended heptapeptides serving as unfolded reference states was previously also used to estimate folding free energies [96, 97] (compare to Section 3.2.2).

1.4.2 Free Energy Perturbation

The *free energy perturbation* approach (FEP) [48, 49] is an alternative to thermodynamic integration. Again, the free energy difference is calculated with the help of the coupling parameter λ. Here, every (intermediate) state λ is perturbed by a small change $\Delta\lambda$.
The free energy difference between state λ and its perturbation $\lambda + \Delta\lambda$ reads

$$\begin{aligned} \Delta F_\lambda &= F(\lambda+\Delta\lambda) - F(\lambda) = -k_B T \ln \frac{Z_{\lambda+\Delta\lambda}}{Z_\lambda} \\ &= -k_B T \ln \frac{\iint e^{-H(\mathbf{p},\mathbf{q},\lambda+\Delta\lambda)/k_B T} \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}}}{\iint e^{-H(\mathbf{p},\mathbf{q},\lambda)/k_B T} \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}}}. \end{aligned} \tag{1.4.4}$$

Multiplying the integrand of the partition function in the numerator with the identity $e^{-H(\mathbf{p},\mathbf{q},\lambda)/k_B T} e^{+H(\mathbf{p},\mathbf{q},\lambda)/k_B T}$ yields the change in free energy

$$\begin{aligned} \Delta F_\lambda &= -k_B T \ln \frac{\iint e^{-\frac{H(\mathbf{p},\mathbf{q},\lambda)}{k_B T}} e^{+\frac{H(\mathbf{p},\mathbf{q},\lambda)}{k_B T}} e^{-\frac{H(\mathbf{p},\mathbf{q},\lambda+\Delta\lambda)}{k_B T}} \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}}}{\iint e^{-H(\mathbf{p},\mathbf{q},\lambda)/k_B T} \frac{d\mathbf{p}d\mathbf{q}}{h^{3N}}} \\ &= -k_B T \ln \left\langle e^{-(H(\lambda+\Delta\lambda)-H(\lambda))/k_B T} \right\rangle_\lambda. \end{aligned} \tag{1.4.5a}$$

Summing over every intermediate free energy difference results in the change of free energy between the initial and the final state

$$\Delta F = F(\lambda_B) - F(\lambda_A) = \sum_{\lambda=\lambda_A}^{\lambda_B-\Delta\lambda} \Delta F_\lambda. \tag{1.4.5b}$$

For reasonable results a significant overlap between the states λ and $\lambda+\Delta\lambda$ is required. In practical applications, therefore, one chooses small perturbations $\Delta\lambda$ to sum over an appropriate number of intermediate steps.
Recent adaptions of the free energy perturbation method can be found in [98] and [99]. FEP was used for stability free energy computations with a tripeptide as the unfolded reference state [100] (compare to Section 3.2.2).

1.4.3 Potential of Mean Force and Umbrella Sampling

Potential of Mean Force

The potential of mean force of an N particle system (developed by Kirkwood [50]) is the mean force $\overline{\vec{F}}$ acting on a particle i for a fixed set of n molecules averaged over all possible conformations of the $n+1\ldots N$ free particles

$$\overline{\vec{F}_i} = -\vec{\nabla}_i \phi_{\mathrm{PMF}} = \frac{\int\ldots\int \left(-\vec{\nabla}_i\phi\right) e^{-\phi/k_BT} d\mathbf{q}_{n+1}\ldots d\mathbf{q}_N\, d\mathbf{p}_{n+1}\ldots d\mathbf{p}_N}{\int\ldots\int e^{-\phi/k_BT} d\mathbf{q}_{n+1}\ldots d\mathbf{q}_N\, d\mathbf{p}_{n+1}\ldots d\mathbf{p}_N}. \tag{1.4.6}$$

As the potential ϕ is independent of the generalized momenta $\mathbf{p}$, the integration of $d\mathbf{p}$ yields the same result in the numerator and in the denominator and can therefore be canceled.
The formalism can also be applied to obtain free energy changes along a specified reaction coordinate R [44, 101]. The reaction coordinate R is a hypersurface in configurational space and therein a function of the particle's coordinates $R = R(\mathbf{q})$. To restrain the system with respect to the reaction coordinate the constraint

$$R' = R(\mathbf{q}) \tag{1.4.7}$$

must be applied. Thus, instead of integrating with respect to $d\mathbf{q}_{n+1}\ldots d\mathbf{q}_N$, the whole set of (infinitesimal) generalized coordinates (denoted by $d\mathbf{q}$) is taken and the restraining of one or more particles (or e.g. only one coordinate of one particle) is done by introducing $\delta(R(\mathbf{q}) - R')$ in the integral, which yields

$$
\begin{aligned}
-\vec{\nabla}_i \phi_{\mathrm{PMF}} &= \frac{\int \delta(R(\mathbf{q})-R')\left(-\vec{\nabla}_i\phi\right) e^{-\phi/k_BT} d\mathbf{q}}{\int \delta(R(\mathbf{q})-R') e^{-\phi/k_BT} d\mathbf{q}} \\
&\overset{1}{=} k_BT \frac{\vec{\nabla}_i \int \delta(R(\mathbf{q})-R') e^{-\phi/k_BT} d\mathbf{q}}{\int \delta(R(\mathbf{q})-R') e^{-\phi/k_BT} d\mathbf{q}} \\
&= k_BT \vec{\nabla}_i \ln \int \delta(R(\mathbf{q})-R') e^{-\phi/k_BT} d\mathbf{q}. \qquad (1.4.8)
\end{aligned}
$$

As we have now the derivative of a logarithm of a function $\partial_x \ln f(x)$, we can multiply any constant to the function $f(x)$ as it will cancel out by applying the chain rule of derivatives on $\ln f(x)$. Hence, dividing the argument of the logarithm with $\int e^{-\phi/k_BT} d\mathbf{q}$ we obtain a formula for the mean force containing an average distribution function $\langle \rho(R') \rangle$

$$
-\vec{\nabla}_i \phi_{\mathrm{PMF}} = k_BT \vec{\nabla}_i \ln \underbrace{\left(\frac{\int \delta(R(\mathbf{q})-R') e^{-\phi/k_BT} d\mathbf{q}}{\int e^{-\phi/k_BT} d\mathbf{q}} \right)}_{=\langle \rho(R') \rangle}. \qquad (1.4.9)
$$

Evaluation of the integral results in the free energy of the system at a stationary reaction coordinate R

$$
F(R) = -k_BT \ln \langle \rho(R) \rangle + \text{constant}. \qquad (1.4.10)
$$

The relative frequency, taken from a normal simulation, to find a system at a reaction coordinate $R' = R(\mathbf{q})$ yields the probability $\langle \rho(R') \rangle$. Evaluating the reaction coordinate at any value R', a free energy profile can be established. Recently, the potential of mean force methodology was used also for structure prediction and verification [102].

The application of Equation (1.4.10) requires averaging over a statistical ensemble. However, reaction coordinates R' corresponding to conformations with high energy will not be sampled properly and will, therefore, not give reliable results.

[1] For a δ distribution the relation $\partial_x^n \int \delta f\, dx = \int \delta\, \partial_x^n f\, dx = (-1)^n \int \partial_x^n \delta f\, dx$ holds.

Umbrella Sampling

To overcome the sampling problem of high energy configurations, Torrie and Valleau [103] restrained the distribution function in order to foster the sampling of energetically unfavorable conformations. This can also be achieved, for example, by adding a harmonic potential [104]

$$\phi_{\mathrm{U}}(R(\mathbf{q}),R_0) = k_{\mathrm{U}}(R(\mathbf{q})-R_0)^2 \tag{1.4.11}$$

to the Hamiltonian H to restrain the position of the particle to the reaction coordinate R_0 [44, 101].
Inserting the biased potential $\phi+\phi_{\mathrm{U}}$, the mean force (Equation (1.4.9)) yields

$$-\vec{\nabla}_i \phi^b_{\mathrm{PMF}} = k_B T \vec{\nabla}_i \ln\left(\frac{\int \delta(R(\mathbf{q})-R')\, e^{(-\phi-\phi_{\mathrm{U}}(R))/k_B T} d\mathbf{q}}{\int e^{(-\phi-\phi_{\mathrm{U}}(R))/k_B T} d\mathbf{q}}\right) \tag{1.4.12}$$

$e^{-\phi_{\mathrm{U}}(R)/k_B T}$ will become a constant $e^{-\phi_{\mathrm{U}}(R')/k_B T}$ in the numerator because of the δ–distribution and we can write the term outside the integral. In addition we multiply with the identity $\int e^{-\phi/k_B T} d\mathbf{q} / \int e^{-\phi/k_B T} d\mathbf{q}$ in the denominator and obtain

$$\begin{aligned} -\vec{\nabla}_i \phi^b_{\mathrm{PMF}} &= k_B T \vec{\nabla}_i \ln\left(\frac{e^{-\phi_{\mathrm{U}}(R')/k_B T}}{\frac{\int e^{(-\phi-\phi_{\mathrm{U}}(R))/k_B T} d\mathbf{q}}{\int e^{-\phi/k_B T} d\mathbf{q}}} \frac{\int \delta(R(\mathbf{q})-R')\, e^{-\phi/k_B T} d\mathbf{q}}{\int e^{-\phi/k_B T} d\mathbf{q}}\right) \\ &= k_B T \vec{\nabla}_i \ln\underbrace{\left(\frac{e^{-\phi_{\mathrm{U}}(R')/k_B T}}{\left\langle e^{-\phi_{\mathrm{U}}(R)/k_B T}\right\rangle}\left\langle \rho(R')\right\rangle\right)}_{=\left\langle \rho(R')\right\rangle^b} \end{aligned} \tag{1.4.13}$$

with the biased distribution function $\left\langle \rho(R')\right\rangle^b$. In order to obtain the unbiased potential of mean force, the extra term $e^{-\phi_{\mathrm{U}}(R')/k_B T}/\left\langle e^{-\phi_{\mathrm{U}}(R)/k_B T}\right\rangle$ has to cancel out. Rewriting Equation (1.4.13)

$$-\vec{\nabla}_i \phi_{\mathrm{PMF}} = k_B T \vec{\nabla}_i \ln\left(\frac{\left\langle e^{-\phi_{\mathrm{U}}(R)/k_B T}\right\rangle}{e^{-\phi_{\mathrm{U}}(R')/k_B T}}\left\langle \rho(R')\right\rangle^b\right). \tag{1.4.14}$$

and integration finally yield

$$F(R') = -k_B T \ln \langle \rho(R') \rangle^b - \phi_U(R') - k_B T \ln \left\langle e^{-\phi_U(R)/k_B T} \right\rangle + \text{constant.} \quad (1.4.15)$$

Besides this harmonic umbrella potential ansatz that gives a weighted potential of mean force, other techniques exist [105], e.g. the weighted histogram analysis method (WHAM) [106] or an approach using a weighted distribution function [107].
Umbrella sampling is well suited for exploring free energy profiles or landscapes. Multiple simulations using constraints at different values for the reaction coordinate R allow e.g. to depict free energy barriers moving, for example, ions through lipid bilayers or through ion channels [67] (see Figure 1.1).

1.4.4 Jarzynski's Equality

The presented methods above apply all for (quasi–) equilibrated systems, whereas Jarzynski's relation [108, 109]

$$\Delta F = -k_B T \ln \overline{e^{-W/k_B T}} \quad (1.4.16)$$

couples the free energy difference ΔF between two states in equilibrium and the work W of non–equilibrium processes to drive the system from state A to state B. The Jarzyinski equality opens up the way for free energy calculations of stressed systems.
States A and B are again described by $\lambda = 0$ or $\lambda = 1$, where λ could be the previously defined coupling parameter or a normalized reaction coordinate. The work performed on the system in the initial state A to reach the final state B (at time t_s) is obtained by

$$W = \int_0^{t_s} \frac{\partial \lambda}{\partial t} \frac{\partial H}{\partial \lambda} (\mathbf{q}(t), \mathbf{p}(t))\, dt \quad (1.4.17)$$

where $(\mathbf{q}(t), \mathbf{p}(t))$ represents a molecular dynamics trajectory. An experimental verification of the Jarzynski equation was reported by Liphardt et al. [110].

1.5 Continuum Solvent Approaches

One factor making regular molecular dynamics simulations computationally expensive is the explicit use of solvent molecules in order to model appropriate surroundings for the studied biomolecules.

However, for some cases like for the analysis of protonation states of titratable amino acid side chains (pK_a value), implicit water models have been shown to yield reasonable results [111]. Continuum solvent approaches make use of simplified models for the van der Waals and Coulombic interaction between the solute and the solvent, neglecting explicit solvent molecules. With the drawback of limited accuracy, continuum models were also used in early MD simulations of biomolecules or recently to simulate folding of small peptides *in silico*.

1.5.1 Electrostatics

In classical electrostatics [112] the energy E of a point charge q placed in an electrostatic potential ϕ reads

$$E = q\phi. \tag{1.5.1}$$

Starting from a charge distribution $\rho(\vec{r})$ in vacuum the electric field $\vec{E}$ (not to confuse with the scalar energy E) is given by

$$\vec{E}(\vec{r}) = -\vec{\nabla}\phi(\vec{r}) \tag{1.5.2}$$

and the electrostatic potential $\phi(\vec{r})$ is obtained through Poisson's Equation

$$\Delta\phi(\vec{r}) = -\frac{\rho(\vec{r})}{\varepsilon_0}, \tag{1.5.3}$$

where Δ denotes the Laplace Operator $\Delta = \vec{\nabla}\cdot\vec{\nabla}$ and ε_0 the permittivity of vacuum. For a point charge q_1 placed at $\vec{r}_1$ the solution of the Poisson Equation (1.5.3) adopts the simple form

$$\phi_1(\vec{r}) = \frac{q_1}{4\pi\varepsilon_0|\vec{r}-\vec{r}_1|}. \tag{1.5.4}$$

Thereby, the interaction energy E (Equation 1.5.1) for a second charge q_2 at $\vec{r}_2$ with q_1 is given by

$$E = \frac{q_1 q_2}{4\pi\varepsilon_0|\vec{r}_2-\vec{r}_1|}. \tag{1.5.5}$$

As the superposition principle holds for the electrostatic potentialthe electrostatic potential ϕ and the interaction energy E can be written in terms of a charge distribution $\rho(\vec{r})$

$$\phi(\vec{r}) = \frac{1}{4\pi\varepsilon_0}\int\frac{\rho(\vec{r}_i)}{|\vec{r}-\vec{r}_i|}dV \tag{1.5.6}$$

and

$$E(\vec{r}) = \frac{1}{8\pi\varepsilon_0} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} dV_1 dV_2, \tag{1.5.7}$$

where the additional factor $\frac{1}{2}$ accounts for counting charge–charge interactions twice in the integral.

For a charge distribution ρ embedded in a dielectric medium, like for example water or methane, dielectric screening has to be considered. E.g., dipolar molecules exposed to an external electric field will orient along the field vector and result in an antagonizing polarization field $\vec{P}$. This dielectric response leads to a weakened field by a factor of $\frac{1}{\varepsilon}$, where ε is the relative dielectric constant.

Thus, the Poisson equation in a dielectric medium reads

$$\Delta\phi(\vec{r}) = -\frac{\rho(\vec{r})}{\varepsilon\varepsilon_0}. \tag{1.5.8}$$

Additionally, the so-called electronic polarization, separating the centers of positive and negative charges of an atom, leads to low dielectric constants between $1.5 - 2.5$ [113].

For water one obtains a relative dielectric constant of $\varepsilon \approx 78$, while the dielectric permittivity of proteins lies in the range between 1 and 40 as discussed in numerous publications, see e.g. [113, 114]. The dielectric constant in proteins does also rely on the structure and amount of polar residues, thus, different proteins may bear different dielectric permittivities. A model for the calculation of a local dielectric permittivity of a protein proposed by Voges and Karshikoff [115] considers polarizable sites and freely rotating dipoles.

Due to this difference in dielectric constants of solute and solvent the Poisson equation takes the more complex shape

$$-\vec{\nabla}\left[\varepsilon(\vec{r})\vec{\nabla}\phi(\vec{r})\right] = \frac{\rho(\vec{r})}{\varepsilon_0}. \tag{1.5.9}$$

As different polarizations occur in different media, the induced dipoles cancel in the material, but not at the surface, and, therefore, lead to induced surface charges. This can also be seen from the Poisson equation (Equation (1.5.9)) after applying the product rule for derivatives and insertion of Equation (1.5.2)

$$-\varepsilon(\vec{r})\Delta\phi(\vec{r}) - \vec{\nabla}\varepsilon(\vec{r})\vec{\nabla}\phi(\vec{r}) = \frac{\rho(\vec{r})}{\varepsilon_0}$$

$$\Longleftrightarrow \qquad \varepsilon(\vec{r})\vec{\nabla}\vec{E}(\vec{r}) = -\vec{\nabla}\varepsilon(\vec{r})\vec{E}(\vec{r}) + \frac{\rho(\vec{r})}{\varepsilon_0}. \tag{1.5.10}$$

Here, sources of the electric field $\vec{E}$ may result from discontinuities in the dielectric permittivity as would emerge at interfaces. Analytical solutions to problems with simple geometry can be found in textbooks, e.g. [112]. A numerical algorithm for calculating the electrostatic potential using finite differences is presented in Section 1.5.3. Next to numerical solutions other methods are available like the Generalized Born approach [116] (see page 44), the Modified Image Electrostatic approximation [117] or the Tanford–Kirkwood electrostatics [118].
The electrostatic potential ϕ can be split into ϕ_C having its sources in the initial charge distribution $\rho(\vec{r})$ and in a reaction field contribution $\phi_{RF}(\vec{r}, \varepsilon_{solvent})$[2] originating from the induced surface charges at the dielectric boundary. The product of the charge distribution and the reaction field integrated over space yields the reaction field energy E_{RF}

$$E_{RF}(\varepsilon_{solvent}) = \frac{1}{2}\int \rho(\vec{r}, \varepsilon_{solvent})\phi_{RF}\, dV, \tag{1.5.11}$$

which gives directly the interaction energy between the charges in the solute and the continuum dielectric solvent. The factor $\frac{1}{2}$ is due to polarization work[3].
The free energy for hydration, putting the solute from vacuum ($\varepsilon_{solute} = 1$) into water with $\varepsilon_{solute} = 78$, is simply the difference of the interaction energies

$$\Delta E_{hyd} = E_{RF}(\varepsilon_{water} = 78) - E_{RF}(\varepsilon_{vacuum} = 1). \tag{1.5.12}$$

Therefore, the Coulombic interactions inside the solute cancel.

1.5.2 Poisson–Boltzmann Equation

Debye and Hückel [119] suggested a model for the electrostatic free energy of spherical ions in an ionic solution, which builds the basis of the Poisson–

[2] The reaction field potential also depends on the dielectric permittivity of the solute, which is typically held constant during computations, while the solvent may be exchanged.

[3] Assuming a linear response for the reaction field to a charge $\phi_{RF} = cq$, the work for charging the particle from 0 to q is $W = \int_0^q \phi_{RF} dq' = \int_0^q cq' dq' = \frac{c}{2}q^2 = \frac{1}{2}q\phi_{RF}$.

Boltzmann formalism combining Poisson's Equation with a Boltzmann distribution for mobile ions.

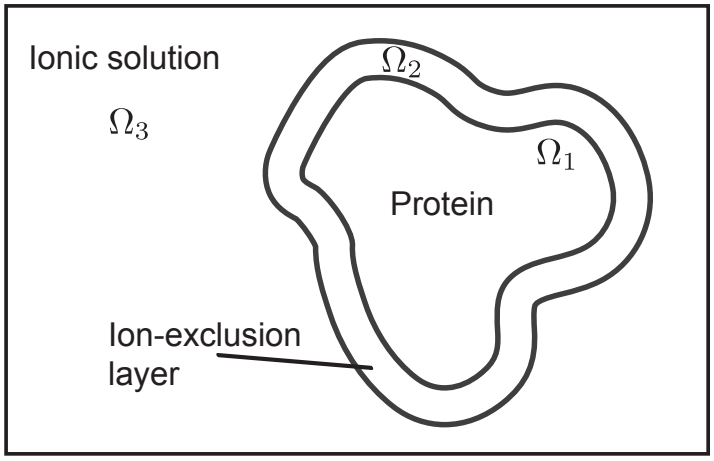

Figure 1.3: Schematic representation of the Debye-Hückel model showing the three regions inside the protein (Ω_1) and in solvent without ions (Ω_2) near the protein and with ions (Ω_3) far from the protein.

The solute is placed in region Ω_1 with a dielectric constant ε_1 (see Figure 1.3). The solvent containing mobile ions is located in region Ω_3 with the dielectric permittivity of the solvent ε_3. Region Ω_2 in between is called the ion–exclusion layer prohibiting ions to come closer to the solute. As Ω_2 is in the solvent phase its dielectric constant is $\varepsilon_2 = \varepsilon_3$.
The following derivation assumes single charged ions, carrying either the positive or negative charge of an electron e_c.
For each region Ω_k with $k = 1,2,3$ the Poisson equation applies for the electrostatic potential

$$\Delta\phi_k(\vec{r}) = -\frac{\rho_k(\vec{r})}{\varepsilon_k \varepsilon_0}. \tag{1.5.13}$$

A requirement for solving these equations is a given charge distribution $\rho_k(\vec{r})$.
For the protein interior region Ω_1 we have a non–vanishing charge distribution $\rho_1(\vec{r})$ leading to a potential (Equation (1.5.6)= and in region Ω_2 we find a vanishing charge distribution $\rho_2 = 0$. Region Ω_3 contains mobile ions. In the following, far from region Ω_1, the ion concentration c_∞ is assumed to be constant for univalent ions with charge $\pm e_c$. Near region Ω_1, the concentration of positive ions c_+ will usually be different from the concentration

c_- of negative ions. The basic assumption of Debye and Hückel is that the Boltzmann distribution applies for the ratio of concentrations

$$\frac{c_\pm}{c_\infty} = e^{-W_\pm(\vec{r})/k_BT}. \tag{1.5.14}$$

$W_\pm(\vec{r})$ is the work required to move one ion from infinity to $\vec{r}$

$$W_\pm(\vec{r}) = \pm e_c \phi(\vec{r}). \tag{1.5.15}$$

Inserting (1.5.15) in (1.5.14) yields the solution for the charge distribution in region Ω_3

$$\begin{aligned} \rho_3(\vec{r}) = c_+ e_c - c_- e_c &= c_\infty e_c e^{-e_c\phi(\vec{r})/k_BT} - c_\infty e_c e^{e_c\phi(\vec{r})/k_BT} \\ &= -2c_\infty e_c \sinh\left(\frac{e_c\phi(\vec{r})}{k_BT}\right). \end{aligned} \tag{1.5.16}$$

With the piecewise definition of the dielectric constant ε

$$\varepsilon(\vec{r}) = \begin{cases} \varepsilon_1, & \text{if } r \in \Omega_1 \\ \varepsilon_3, & \text{else} \end{cases} \tag{1.5.17}$$

and the introduction of the (modified[4]) Debye-Hückel parameter κ^2

$$\kappa^2(\vec{r}) = \begin{cases} 0, & \text{if } r \in \Omega_1 \\ \dfrac{2c_\infty e_c^2}{\varepsilon_0 k_B T}, & \text{else} \end{cases} \tag{1.5.18}$$

the resulting Poisson–Boltzmann equation can be written as

$$-\vec{\nabla}\left[\varepsilon(\vec{r})\vec{\nabla}\phi(\vec{r})\right] = -\kappa^2\left(\frac{k_BT}{e_C}\right)\sinh\left(\frac{e_c\phi(\vec{r})}{k_BT}\right) + \frac{\rho_1(\vec{r})}{\varepsilon_0} \tag{1.5.19}$$

or in its linearized[5] form

$$-\vec{\nabla}\left[\varepsilon(\vec{r})\vec{\nabla}\phi(\vec{r})\right] = -\kappa^2\phi(\vec{r}) + \frac{\rho_1(\vec{r})}{\varepsilon_0}. \tag{1.5.20}$$

[4]The unmodified parameter can be obtained by division with the dielectric constant $\overline{\kappa} = \kappa/\varepsilon$.

[5]$\sinh(ax) \approx ax + \mathcal{O}(x^3)$.

For the two interfaces between the three distinguished regions the continuity conditions

$$\phi_1|_{\Omega_1\cap\Omega_2} = \phi_2|_{\Omega_1\cap\Omega_2}, \qquad \varepsilon_1\vec{\nabla}\phi_1|_{\Omega_1\cap\Omega_2}\vec{n} = \varepsilon_2\vec{\nabla}\phi_2|_{\Omega_1\cap\Omega_2}\vec{n}, \tag{1.5.21a}$$

$$\phi_2|_{\Omega_2\cap\Omega_3} = \phi_3|_{\Omega_2\cap\Omega_3}, \qquad \varepsilon_2\vec{\nabla}\phi_2|_{\Omega_2\cap\Omega_3}\vec{n} = \varepsilon_3\vec{\nabla}\phi_3|_{\Omega_2\cap\Omega_3}\vec{n}, \tag{1.5.21b}$$

apply, where $\vec{n}$ is the normal vector of length 1, and $\phi(\infty) = 0$ as the boundary condition.

1.5.3 Numerical Solution of the Linearized PBE via Finite Difference Method

The numerical solution is usually obtained either by the finite difference method [120–123] or by the boundary element method [124]. The former method will be described in detail.

Finite difference methods are based on replacing derivative expressions by difference quotients by decomposing the region of interest in grid points. In the case of the Poisson-Boltzmann equation the solute molecule(s) and the surrounding dielectric continuum solvent have to be translated onto a three dimensional cubic grid. Atomic charges are not represented anymore by point charges, but fractional charges mapped on the grid:

$$q_f = \left(1 - \frac{|\Delta x|}{h}\right)\left(1 - \frac{|\Delta y|}{h}\right)\left(1 - \frac{|\Delta z|}{h}\right)q. \tag{1.5.22}$$

The atomic charges are smeared on the nearest eight grid points, with a grid spacing h and $|\Delta x|$,$|\Delta y|$,$|\Delta z|$ denoting the distance between particle and grid point respectively in x, y and z direction.

Next to the mapping of charges, it is also important to define dielectric regions on the grid. In general, two possibilities exist for the dielectric boundary: the van der Waals surface (see Figure 1.4 a) as the union of spheres with atomic van der Waals radii and the solvent excluded surface [125], also including regions that are inaccessible to solvent molecules (see Figure 1.4 c). The solvent excluded surface can be obtained by first building the so–called solvent accessible surface [126] (see Figure 1.4 b). This is achieved by an increase of the van der Waals radii by a probe radius — usually 1.4 Å as this approximates the size of a water molecule — and by merging the enlarged spheres. For the solvent excluded surface a ball with the same probe radius is rolled over the initial vdW surface, but its center is not allowed to leave the solvent accessible surface [127]. Every point that cannot be reached by the rolling ball is not accessible for the solvent, and the boundary surface is

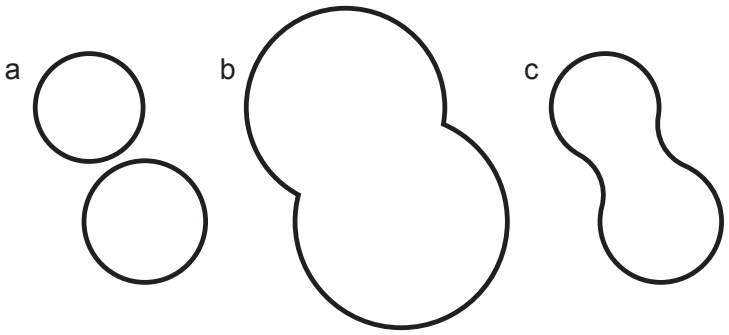

Figure 1.4: Two dimensional representations of the van der Waals (a), the solvent accessible (b) and the solvent excluded (c) surface of two neighboring atoms.

called the solvent excluded surface. A grid based method for generating the solvent excluded surface is presented in Reference [123].
Besides the molecular properties like charge and boundary surface, also the dielectric permittivity ε for solvent and solute as well as the modified Debye–Hückel parameter κ can be assigned to the lattice.
In finite difference methods each grid point represents the appropriate average over the volume that surrounds the grid point, thus we find at the ith grid point

$$\iiint_i f\,dx\,dy\,dz = f_i h^3, \tag{1.5.23}$$

where the integration is done over the spatial points that are closer to grid point i than to any other point. Consequently, the integral is taken over a cube centered at the grid point i and with side length h.
Integrating the linearized Poisson-Boltzmann equation (1.5.20) over volume yields

$$-\int \vec{\nabla}\left[\varepsilon(\vec{r})\vec{\nabla}\phi(\vec{r})\right] dV = -\int \kappa^2\phi(\vec{r})\,dV + \int \frac{\rho(\vec{r})}{\varepsilon_0}\,dV. \tag{1.5.24}$$

The second integral can be approximated with $-\kappa_i^2\phi_i h^3$ and the third one with q_i/ε_0, where the index i indicates the affiliation with the ith grid point. Using Gauss' theorem on the first integral results in a surface integral and equation (1.5.24) reads

$$-\oint \varepsilon(\vec{r})\vec{\nabla}\phi(\vec{r})\,d\vec{A} = -\kappa_i^2\phi_i h^3 + \frac{q_i}{\varepsilon_0}, \tag{1.5.25}$$

where $d\vec{A}$ is an infinitesimal surface element with an orientation perpendicular to the surface. Again applying finite difference formalism on the surface integral yields

$$\begin{aligned}\oint \varepsilon(\vec{r})\vec{\nabla}\phi(\vec{r})\,d\vec{A} \quad \rightarrow \quad & \varepsilon_1^i \frac{\phi_1^i - \phi_i}{h}\vec{e}_x h^2 \vec{e}_x + \varepsilon_2^i \frac{\phi_2^i - \phi_i}{h}(-\vec{e}_x)h^2(-\vec{e}_x) \\ & + \varepsilon_3^i \frac{\phi_3^i - \phi_i}{h}\vec{e}_y h^2 \vec{e}_y + \varepsilon_4^i \frac{\phi_4^i - \phi_i}{h}(-\vec{e}_y)h^2(-\vec{e}_y) \\ & + \varepsilon_5^i \frac{\phi_5^i - \phi_i}{h}\vec{e}_z h^2 \vec{e}_z + \varepsilon_6^i \frac{\phi_6^i - \phi_i}{h}(-\vec{e}_z)h^2(-\vec{e}_z) \\ = \quad & \sum_{j=1}^{6} \varepsilon_j^i \left(\phi_j^i - \phi_i\right) h, \end{aligned} \tag{1.5.26}$$

where ϕ_j^i denotes the potential of the jth grid point surrounding i and $\varepsilon_j^i = \sqrt{\varepsilon_i \varepsilon_j}$. Insertion in Equation (1.5.25) yields

$$-\sum_{j=1}^{6} \varepsilon_j^i \left(\phi_j^i - \phi_i\right) h = -\kappa_i^2 \phi_i h^3 + \frac{q_i}{\varepsilon_0}, \tag{1.5.27}$$

and, eventually, after solving for ϕ_i:

$$\phi_i = \frac{\dfrac{q_i}{h\varepsilon_0} + \sum\limits_{j=1}^{6} \varepsilon_j^i \phi_j^i}{\kappa_i^2 h^2 + \sum\limits_{j=1}^{6} \varepsilon_j^i}. \tag{1.5.28}$$

This equation is solved for every grid point based on an initial guess for ϕ and the whole grid is updated according to the solution. This procedure is repeated until a predefined termination criterion is reached.
Important to such calculations are the boundary conditions, where the most simple possibility is to set the potential boundary points to zero, or to the Debye–Hückel expression

$$\phi_i^{\text{boundary}} = \frac{\sum_{j=1}^{n} q_j e^{-\kappa r_{ij}}}{\varepsilon_{\text{solvent}} r_{ij}}. \tag{1.5.29}$$

The Debye–Hückel potential is an approximation of the analytical solution for the electrostatic potential far from the protein.
The numerically determined potential can be used to calculate the electrostatic energy according to

$$E_{\text{grid}} = \frac{1}{2} \sum_{i \in \text{grid}} q_i \phi_i \tag{1.5.30}$$

resulting in the so-called grid energy.
As the molecule is mapped onto the grid and the point charges are smeared, numerical artifacts are introduced depending on the grid size and the relative position of the molecule to the grid.

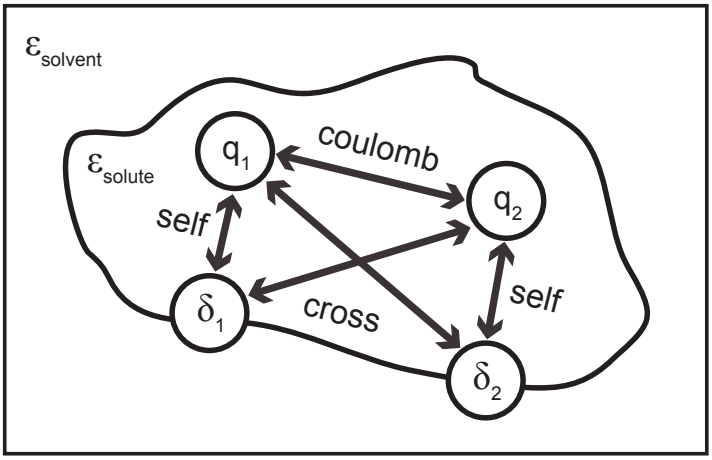

Figure 1.5: Coulomb, cross and self contributions of the total electrostatic energy are cartooned. Two real charges q_1 and q_2 lead to induced surface charges δ_1 and δ_2, respectively, due to the dielectric boundary. Although the induced charges are distributed on the whole surface, δ_1 and δ_2 are depicted in a localized fashion to clarify the existing electrostatic interactions.
Next to the Coulomb interaction between the real charges q_1 and q_2, there exist additionally the cross interactions between a real charge and the surface charges induced by other real charges and the self contribution between a charge and its induced surface charge.

Another problem is that part of the potential is generated by the charge carrier itself. It's *own* contribution[6] would result in an infinite energy, but because of the smudged charges the grid energy remains finite with the drawback of another numerical artifact (the resulting potential is grid dependent). Other contributions to the total electrostatic grid energy that allow for a more detailed analysis of electrostatic interactions are depicted in Figure 1.5. To cancel numerical artifacts and get rid of the *own* contribution, a second computation of E_{grid} with $\varepsilon_{\text{solvent}} = \varepsilon_{\text{solute}}$ has to be subtracted

[6]This contribution is entitled as *own* contribution to distinguish from the *self* reaction field of a charge due to its induced surface charge (see Figure 1.5).

(thereby, the internal Coulombic interaction cancels, too). The neutralization holds also for the calculation of hydration free energies according to equation (1.5.12) — the internal Coulomb interactions and the artificial grid contributions are canceled and no further treatment is necessary.
An alternative to the computation of the reaction field energy E_{RF} via a sum over the grid point energies is the direct calculation via induced surface charges at the dielectric boundary [123]

$$E_{\mathrm{RF}} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{q_i\delta_j}{\varepsilon_{\mathrm{solute}}r_{ij}}, \tag{1.5.31}$$

where n is the number of real point charges, m the number of boundary points hosting induced surface charges and δ_i denotes the surface charge at point i. For the atomic partial charges the actual position of the atoms are inserted. Thus, only the surface charge is grid–dependent. As the dielectric response in the form of the induced surface charges will be calculated explicitly, the dielectric constant is dropped and Poisson's equation (1.5.3) can be used to obtain an effective charge distribution, containing real and induced charges:

$$-\Delta\phi(\vec{r}) = \frac{\rho_{\mathrm{real}}(\vec{r}) + \rho_{\mathrm{induced}}(\vec{r})}{\varepsilon_0}. \tag{1.5.32}$$

Using the same finite difference formalism as for the linearized Poisson-Boltzmann equation yields

$$-\sum_{j=1}^{6}\left(\phi_j^i - \phi_i\right)h = \frac{q_i + \delta_i}{\varepsilon_0}. \tag{1.5.33}$$

At the boundary points we find for the induced surface charges[7]

$$\delta_i = -\varepsilon_0\sum_{j=1}^{6}\left(\phi_j^i - \phi_i\right)h - q_i. \tag{1.5.34}$$

With only one calculation of the potential needed in contrast to the grid energy method this formalism is computationally faster by a factor of ~ 2. Besides it is more accurate using point charges instead of the smeared ones. In order to take account of salt effects the grid method is still necessary and two computations are required: one with the de–ionized solvent and one with a salt concentration I. The resulting salt contribution is then given by

$$\Delta E_{\mathrm{grid}}^{\mathrm{salt}} = E_{\mathrm{grid}}(I) - E_{\mathrm{grid}}(I=0) \tag{1.5.35}$$

[7] In the medium itself applies $\frac{\rho_{\mathrm{real}}}{\varepsilon} = \rho_{\mathrm{real}} + \rho_{\mathrm{induced}}$.

which has to be added to the reaction field energy at $I = 0$ in order to obtain the complete solute–solvent interaction energy

$$E_{\mathrm{RF}}(I) = E_{\mathrm{RF}}(I = 0) + \Delta E^{\mathrm{salt}}_{\mathrm{grid}}(I). \tag{1.5.36}$$

Several software packages are freely available for the numerical solution of the Poisson–Boltzmann equation using the finite difference method like DelPhi (`http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:DelPhi`) [121–123], UHBD (`http://adrik.bchs.uh.edu/uhbd.html`) [128, 129] or BALL (`http://www.ball-project.org`) [130]. Another popular program is apbs (`http://apbs.sourceforge.net`) [131] using a parallel adaptive finite element method to solve the PBE.

1.5.4 Generalized Born Model

Solving the Poisson–Boltzmann equation numerically is still computational expensive and, therefore, prohibitive for use in MD simulations. One successful approach for substituting the Poisson–Boltzmann formalism is the use of the Generalized Born model [116].

Born [132] reported an analytical formula for the solvation free energy of ions

$$\Delta G_{\mathrm{Born}} = \frac{1}{4\pi\varepsilon_0}\left(1 - \frac{1}{\varepsilon}\right)\frac{q^2}{2a} \tag{1.5.37}$$

in a dielectric medium with a permittivity of ε, where q and a are the charge and radius of an ion.

For a system consisting of n charged particles, pairwise separated at distances much larger than the sum of their radii, the total electrostatic free energy G_{es} (sum of Coulomb interactions and Born solvation terms) can be expressed as

$$G_{\mathrm{es}} = \frac{1}{4\pi\varepsilon\varepsilon_0}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{q_i q_j}{r_{ij}} - \frac{1}{4\pi\varepsilon_0}\left(1 - \frac{1}{\varepsilon}\right)\sum_{i=1}^{n}\frac{q_i^2}{2a_i}. \tag{1.5.38}$$

The first term can be split up into the Coulombic interaction energy in vacuum and an expression similar to the second term

$$G_{\text{es}} = \underbrace{\frac{1}{4\pi\varepsilon_0}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{q_i q_j}{r_{ij}}}_{=G_{\text{Coulomb}}^{\text{vacuum}}} \underbrace{-\frac{1}{4\pi\varepsilon_0}\left(1-\frac{1}{\varepsilon}\right)\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{q_i q_j}{r_{ij}} - \frac{1}{4\pi\varepsilon_0}\left(1-\frac{1}{\varepsilon}\right)\sum_{i=1}^{n}\frac{q_i^2}{2a_i}}_{=G_{\text{solvation}}}. \tag{1.5.39}$$

The goal of the Generalized Born model is to rewrite the latter sum $G_{\text{solvation}}$ in the form

$$G_{\text{solvation}} = -\frac{1}{4\pi\varepsilon_0}\left(1-\frac{1}{\varepsilon}\right)\sum_{i=1}^{n-1}\sum_{j=i}^{n}\frac{q_i q_j}{f_{ij}^{\text{GB}}}, \tag{1.5.40}$$

with a simple approximation for the effective Born radius f_{ij}^{GB} applicable for standard molecular biology systems.
Still et al. [116] proposed

$$f_{ij}^{\text{GB}} = \sqrt{r_{ij}^2 + \alpha_{ij}^2 e^{-r_{ij}^2/4\alpha_{ij}^2}}, \tag{1.5.41}$$

with

$$\alpha_{ij} = \sqrt{\alpha_i \alpha_j}, \tag{1.5.42}$$

where α_i denotes the Born radii of the atom i, which not only depend on the radius of the atom, but also on the positions and radii of every other atom in the system. The Born radius α_i is obtained by evaluating the solvation free energy $G_{\text{solvation}}^i$ for atom i

$$G_{\text{solvation}}^i = -\frac{1}{4\pi\varepsilon_0}\left(1-\frac{1}{\varepsilon}\right)\frac{q_i^2}{2\alpha_i}. \tag{1.5.43}$$

All other charges are assumed to be neutral, while they replace the dielectric medium of the solvent with the solute's, at the same time. The most accurate way to analyze the solvation free energy is to solve the Poisson–Boltzmann equation numerically. But this procedure would have a similar or even worse time consumption as solving the Poisson–Boltzmann equation for the fully charged molecule directly, and, therefore, is not generally applicable.
In the model suggested by Still et al. [116], the effective Born radius is obtained by solving Born's equation (1.5.37) piecewise for concentric shells around atom i with thickness T. The contribution of each shell is assumed

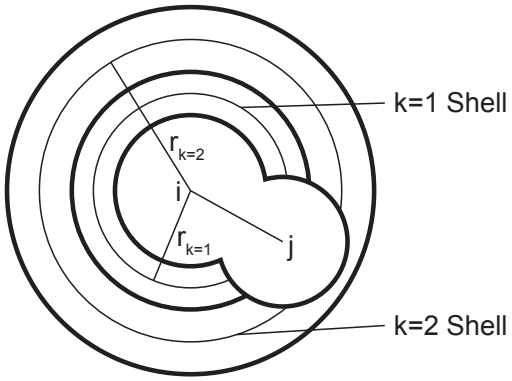

Figure 1.6: Schematic representation of Still's GB method: Shells around an atom of a diatomic molecule for the numerical calculation of the effective Born radius.

to be the Born energy of the shell times the fraction of the shell area outside the van der Waals volume of the molecule

$$a_k = \frac{A_k}{4\pi r_k^2}. \tag{1.5.44}$$

A_k is the surface area of the kth shell outside the van der Waals surface of the whole molecule. The shell is isolated by the thick lines in Figure 1.6, while the surface area is evaluated in the middle of the shell (light circular lines) in between the concentric shell borders. The surface area can be numerically achieved as described in the next section.
If a shell surrounds the complete molecule ($k = n+1$), the Born equation is used directly for calculating the energy contribution of the remaining dielectric space.
Thus, the solvation free energy for atom i reads

$$G^i_{\text{solvation}} = -\frac{1}{4\pi\varepsilon_0}\left(1-\frac{1}{\varepsilon}\right) q_i^2 \left[\left(\sum_{k=1}^{n} a_k \left(\frac{1}{r_k - \frac{1}{2}T_k} - \frac{1}{r_k + \frac{1}{2}T_k}\right)\right) + \frac{1}{r_{n+1}}\right], \tag{1.5.45}$$

with $r_{k+1} = r_k + \frac{1}{2}(T_k + T_{k+1})$. For efficiency reasons, Still et al. [116] used $T_{k+1} = \frac{3}{2}T_k$ starting with $T_1 = 1$Å and $r_1 = r_{FF} - 0.09$Å (r_{FF} denotes the atomic radius derived from the force field).
Combining equations (1.5.45) and (1.5.43) the effective radius α_i is obtained which obviously does neither depend on charge nor on the dielectric permittivity. The Generalized Born solvation free energy is also independent of

the molecule's dielectric constant, which could be mimicked by changing the coefficient containing the permittivity to

$$\left(\frac{1}{\varepsilon_{\text{solute}}} - \frac{1}{\varepsilon_{\text{solvent}}}\right). \tag{1.5.46}$$

Also other methods for finding effective Born radii have been proposed [133–139] and been reviewed in [140].

1.5.5 Non–polar Solvation Contributions

The non–polar Lennard–Jones interactions between solute and solvent molecules are typically approximated to be proportional to the solvent accessible surface area A_{SASA} [141, 142] (see Figure 1.4)

$$G_{\text{SASA}} = \gamma A_{\text{SASA}} + b. \tag{1.5.47}$$

The surface tension γ and the constant b are empirically derived.
This surface area is obtained by the union of the atomic spheres with enlarged atomic radii by addition of a probe radius approximating a virtual solvent molecule radius [126].
Although not capable of reproducing the experimental hydration free energies of cyclic alkanes or alkenes [143] this simple approximation yields reasonable results when combined with polar contributions [142] or when compared to MD simulations [144, 145]. The model was also successfully applied in combination with other contributions to reproduce protein stabilities or binding affinities, e.g. in combination with force field contributions and entropy estimates in the EGAD program [60, 61].
The solvent accessible surface area can be computed in slow but correct analytical ways [146–149], using approximate analytical approaches [116, 150] or numerically [126, 151].
The grid–based technique developed by Shrake and Rupley [151] distributes reference points on an atomic sphere via

$$\begin{pmatrix} \cos\varphi_i \sin\vartheta_i \\ \sin\varphi_i \sin\vartheta_i \\ \cos\vartheta_i \end{pmatrix} a, \tag{1.5.48}$$

where ϑ_i varies from 0 to π and φ_i from 0 to 2π. a is the atomic radius.
Subsequently every reference point is tested whether it is buried or exposed and the fraction of exposed points yields the atomic solvent accessible surface area

$$A = \frac{n_{\text{exposed}}}{n_{\text{total}}} 4\pi a^2. \tag{1.5.49}$$

Other methods for the nonpolar solvation also include a volume term [152]

$$G_{\text{SAV}} = pV, \tag{1.5.50}$$

where p is a constant parameter, or an integral [143, 152] of the form

$$G_I = \rho \sum_{i=1}^{n} \int_{\text{solvent}} I_i(\vec{r}_i, \vec{x}) dx^3, \tag{1.5.51}$$

where ρ is the density of the solute and $I_i(\vec{r}_i, \vec{x})$ contains the attractive (r^{-6}) contributions of the Lennard–Jones interaction (see Equation (1.3.6)) between atom i at position $\vec{r}_i$ and water at position $\vec{x}$.

1.6 Entropy

Apart from the internal energy also the entropy of a system contributes to the overall free energy crucial for protein stabilities or binding affinities. The following subsections present both, the change of entropy derived systematically by thermodynamic integration between two states and two widely applied approximations for the entropy of a solute that is not directly accessible by MD simulations.

1.6.1 Thermodynamic Integration

Although being included implicitly in the free energy obtained e.g. via thermodynamic integration, it can be desirable to calculate the entropy explicitly [44].
The entropy is obtained as the difference of internal energy and free energy (Equation (1.1.8b))

$$TS = U - F. \tag{1.6.1}$$

Similar as for the free energy (see Equation (1.4.2), the first derivative with respect to the coupling parameter λ of the internal energy

$$U = \langle H \rangle = \frac{\iint e^{-\beta H(\mathbf{p},\mathbf{q},\lambda)} H(\mathbf{p},\mathbf{q},\lambda) d\mathbf{p} d\mathbf{q}}{\iint e^{-\beta H(\mathbf{p},\mathbf{q},\lambda)} d\mathbf{p} d\mathbf{q}} \tag{1.6.2}$$

is calculated and yields

$$\begin{aligned}\frac{dU}{d\lambda} &= \frac{\iint \left(-\beta e^{-\beta H}\frac{\partial H}{\partial \lambda}H + e^{-\beta H}\frac{\partial H}{\partial \lambda}\right) d\mathbf{p}d\mathbf{q} \iint e^{-\beta H} d\mathbf{p}d\mathbf{q}}{\left(\iint e^{-\beta H} d\mathbf{p}d\mathbf{q}\right)^2} \\ &+ \frac{\iint e^{-\beta H} H d\mathbf{p}d\mathbf{q} \iint -\beta e^{-\beta H}\frac{\partial H}{\partial \lambda} d\mathbf{p}d\mathbf{q}}{\left(\iint e^{-\beta H} d\mathbf{p}d\mathbf{q}\right)^2} \\ &= \left\langle -\beta H \frac{\partial H}{\partial \lambda} \right\rangle_\lambda + \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda - \langle H \rangle_\lambda \left\langle \beta \frac{\partial H}{\partial \lambda} \right\rangle_\lambda . \end{aligned} \tag{1.6.3}$$

Differentiating equation (1.6.1) leads to

$$T\frac{dS}{d\lambda} = \frac{dU}{d\lambda} - \frac{dF}{d\lambda}. \tag{1.6.4}$$

Substituting the derivatives of the internal energy U (Equation (1.6.3)) and of the free energy (Equation (1.4.2)) finally yields

$$\frac{dS}{d\lambda} = \frac{1}{k_B T^2}\left(-\left\langle H\frac{\partial H}{\partial \lambda} \right\rangle_\lambda + \langle H \rangle_\lambda \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda \right). \tag{1.6.5}$$

Integrating with respect to λ results in the change in entropy between initial and final state [44]. Thus, this method relies on an extensive sampling of configurational space. A very precise ensemble is required since not only the difference of the Hamiltonian between different states, but the full Hamiltonian enters in the calculation. Therefore, simplified approaches have been developed.

1.6.2 Normal Mode Analysis

The entropy of a protein is often approximated by the entropy of a quantum–harmonic oscillator (quasi–harmonic approximation) [153]. For a 1–dimensional quantum–harmonic oscillator the entropy is given by [154]

$$S_{\text{qho}} = \frac{k_B \alpha}{e^\alpha - 1} - k_B \ln\left(e^\alpha - 1\right) \tag{1.6.6}$$

with

$$\alpha = \frac{\hbar\omega}{k_B T}. \tag{1.6.7}$$

Approximating the entropy of every (internal) degree of freedom of a protein by the entropy of a quantum mechanical oscillator the total configurational entropy due to the vibrational modes can be written as

$$S_{\text{vib}} = \sum_{i=1}^{3N-6} \left[\frac{k_B \alpha_i}{e_i^{\alpha} - 1} - k_B \ln\left(e_i^{\alpha} - 1\right) \right]. \tag{1.6.8}$$

For the remaining six translational and rotational degrees of freedom, that are treated semi–classical, the entropy is given by [155]

$$S_{\text{trans}} = k_B \left[\frac{5}{2} + \frac{3}{2} \ln\left(\frac{m k_B T}{2\pi\hbar^2} \right) - \ln \rho \right] \quad \text{and} \tag{1.6.9}$$

$$S_{\text{rot}} = k_B \left[\frac{3}{2} + \frac{1}{2} \ln(I_1 I_2 I_3) + \frac{3}{2} \ln\left(\frac{2 k_B T}{\hbar^2} \right) - \ln \sigma \right]. \tag{1.6.10}$$

ρ is the number density at a concentration of 1M, I_i are the three principal moments of inertia and σ is a symmetry factor that equals 1 for non–symmetric molecules and 2 for symmetric ones like dimers.
Every rotational and translational degree of freedom additionally contributes an amount of $\frac{1}{2} k_B T$ to the absolute free energy.
A number of similar quasi–harmonic approaches using normal mode analysis (NMA) to obtain vibrational frequencies of the protein have been developed in the past [155–160]. Normal modes of an oscillating system are collective vibrations with all particles moving with the same resonant frequency. Every harmonic oscillation of a system can be described as a linear combination of its normal modes.
For molecules, the normal modes can be obtained starting from a local energy minimum with coordinates $\mathbf{r}^0$. A local Taylor expansion of the potential ϕ yields

$$\phi(\mathbf{r}) = \phi(\mathbf{r}^0) + \sum_{i=1}^{3N} \left(r_i - r_i^0\right) \left.\frac{\partial \phi}{\partial r_i}\right|_{\mathbf{r}^0} + \frac{1}{2} \sum_{i=1}^{3N} \sum_{j=1}^{3N} \left(r_i - r_i^0\right)\left(r_j - r_j^0\right) \left.\frac{\partial^2 \phi}{\partial r_i \partial r_j}\right|_{\mathbf{r}^0} + \mathcal{O}(3). \tag{1.6.11}$$

Since the potential has a minimum at $\mathbf{r}^0$ the second term equals zero and the Hamiltonian H can be written in a harmonic approximation around the minimum $\mathbf{r}^0$

$$H(\mathbf{r}) = \frac{1}{2} \sum_{i=1}^{3N} m_i \dot{r}_i^2 + \frac{1}{2} \sum_{i=1}^{3N} \sum_{j=1}^{3N} \left(r_i - r_i^0\right)\left(r_j - r_j^0\right) \left.\frac{\partial^2 \phi}{\partial r_i \partial r_j}\right|_{\mathbf{r}^0} + \text{const.} \tag{1.6.12}$$

Introducing mass–weighted coordinates via

$$R_i = \sqrt{m_i}\left(r_i - r_i^0\right) \tag{1.6.13}$$

the Hamiltonian writes

$$H(\mathbf{r}) = \frac{1}{2}\sum_{i=1}^{3N}\dot{R}_i^2 + \frac{1}{2}\sum_{i=1}^{3N}\sum_{j=1}^{3N} R_i R_j \left.\frac{\partial^2\phi}{\partial R_i \partial R_j}\right|_{\mathbf{R}^0}. \tag{1.6.14}$$

To determine the normal modes, the Hamiltonian is rewritten as a superposition of $3N-6$ independent harmonic oscillators[8]

$$H(\mathbf{q}) = \frac{1}{2}\sum_{i=1}^{3N-6}\dot{q}_i^2 + \frac{1}{2}\sum_{i=1}^{3N-6}\omega_i^2 q_i^2. \tag{1.6.15}$$

The internal coordinates $\mathbf{q}$ are related to the coordinates $\mathbf{R}$ through the transformation

$$\mathbf{R} = \mathbf{M}\mathbf{q}, \tag{1.6.16}$$

with $\mathbf{M}^T\mathbf{M} = \mathbf{I}$. The first term of the Hamiltonian can now be expressed as

$$\sum_{i=1}^{3N}\dot{R}_i^2 = \dot{\mathbf{R}}^\mathbf{T}\dot{\mathbf{R}} = \dot{\mathbf{q}}^\mathbf{T}\mathbf{M}^\mathbf{T}\mathbf{M}\dot{\mathbf{q}} = \dot{\mathbf{q}}^\mathbf{T}\dot{\mathbf{q}}, \tag{1.6.17}$$

and the second one as

$$\sum_{i=1}^{3N}\sum_{j=1}^{3N} R_i R_j \left.\frac{\partial^2\phi}{\partial R_i \partial R_j}\right|_{\mathbf{R}^0} = \mathbf{q}^\mathbf{T}\mathbf{M}^\mathbf{T}\mathbf{H}_{\mathbf{q}^0}\mathbf{M}\mathbf{q}, \tag{1.6.18}$$

where $\mathbf{H}_{q^0}$ denotes the Hessian matrix (1.3.34) evaluated at the minimum $\mathbf{q}^0$. Comparing equations (1.6.17, 1.6.18) with (1.6.15) we find

$$\mathbf{q}^\mathbf{T}\mathbf{M}^\mathbf{T}\mathbf{H}_{\mathbf{q}^0}\mathbf{M}\mathbf{q} = \sum_{i=1}^{3N-6}\omega_i^2 q_i^2. \tag{1.6.19}$$

Introducing

$$\Omega = \mathrm{diag}\left(\omega_1^2, \omega_2^2, \ldots \omega_{3N-6}^2\right) \tag{1.6.20}$$

Equation (1.6.19) can be written as

$$\mathbf{q}^\mathbf{T}\mathbf{M}^\mathbf{T}\mathbf{H}_{\mathbf{q}^0}\mathbf{M}\mathbf{q} = \mathbf{q}^\mathbf{T}\Omega\mathbf{q}, \tag{1.6.21}$$

which results in the eigenvalue problem

$$\mathbf{M}^\mathbf{T}\mathbf{H}_{\mathbf{q}^0}\mathbf{M} = \Omega. \tag{1.6.22}$$

[8]There are only $3N-6$ internal degrees of freedom as three degrees of freedom each are due to the translation and rotation of the system, respectively.

The solution yields a set of normal modes with eigenvectors a_n and their associated resonant frequencies ω_n.
In principle, the frequencies ω could be obtained from the covariance matrix. However in Cartesian coordinates the covariance matrix is singular, rendering its application in the above expression for S_{vib} (Equation 1.6.8) impossible. The singularities can be avoided by transformation to internal coordinates, the so–called normal mode approach.
Entropy approximation via normal mode analysis makes use of a single structure at a local energy minimum. Thus, an excessive energy minimization has to precede the above presented procedure, which renders the analysis computationally quite demanding. Also, the normal modes will only be a rough approximation of protein fluctuations at a physiological temperature.

1.6.3 Schlitter's Approach

In 1993 Schlitter [154] presented a modified approach analyzing absolute entropies from the covariance matrix of atomic fluctuations.
Starting from a system with discrete states $n = 0, 1, \ldots$ and associated energies ε_n the partition function takes the form

$$Z = \sum_n e^{-\beta\varepsilon_n}. \tag{1.6.23}$$

The probabilities of the states are given by

$$\rho_n = \frac{-\beta\varepsilon_n}{Z} \tag{1.6.24}$$

and are related to the entropy according to

$$S = -k_B \sum_n \rho_n \ln \rho_n. \tag{1.6.25}$$

In the following we consider a one dimensional system with particle mass m and position x. We assume that

$$\langle x \rangle = 0 \quad \text{and} \tag{1.6.26}$$

$$\langle x^2 \rangle = \sum_n \rho_n \langle x^2 \rangle_n. \tag{1.6.27}$$

$\langle \cdot \rangle_n$ denotes the expectation value belonging to the individual state n.
For a given (covariance) $\langle x^2 \rangle$ the upper limit for the entropy is found using the derivatives of the entropy with respect to the probabilities ρ_n, including

the Lagrange multipliers λ and μ for the constraints of Equation (1.6.27) and for the normalization condition $\sum \rho_n = 1$. A stationary point is found at

$$\frac{d}{d\rho_n} S = -k_B \left(\ln \rho_n + 1 \right) + \lambda \langle x^2 \rangle_n + \mu = 0. \tag{1.6.28}$$

It is a maximum because

$$\frac{d^2}{d\rho_n^2} S = -k_B \frac{1}{\rho_n} < 0. \tag{1.6.29}$$

Substituting (1.6.24) in (1.6.28) and rearrangement yields

$$\beta \varepsilon_n - \frac{\lambda}{k_B} \langle x^2 \rangle_n = \frac{\mu}{k_B} - \ln Z + 1. \tag{1.6.30}$$

To find a maximum to the total entropy this equation has to be valid for every state n. Thus, the left hand site of equation (1.6.30) has to cancel and the energies and variances in each state have to be proportional to each other.
This is true for a quantum-harmonic oscillator with energy eigenvalues

$$\varepsilon_n = m\omega_n^2 \langle x^2 \rangle_n. \tag{1.6.31}$$

Thus the entropy of a quantum–harmonic oscillator S_{qho} (1.6.6) is an upper limit for the true entropy

$$S \leq S_{\mathrm{qho}} = \frac{k_B \alpha}{e^\alpha - 1} - k_B \ln \left(e^\alpha - 1 \right). \tag{1.6.32}$$

A classical treatment of the harmonic oscillator would result in a Gaussian distribution for x. At a given variance the Gaussian distribution gives the largest entropy as compared to other distributions. For an entropy analysis based on MD simulations the variance is replaced by the classical variance $\langle x^2 \rangle_c$. In the classical limit $\hbar\omega \ll k_B T$ the equipartition theorem

$$m\omega^2 \langle x^2 \rangle_c = k_B T \tag{1.6.33}$$

is used to obtain the frequency ω.
Schlitter introduced another function being an upper limit to the entropy of a quantum-harmonic oscillator

$$S \leq S_{\mathrm{qho}} < S' = \frac{1}{2} k_B \ln \left(1 + \frac{e^2}{\alpha} \right), \tag{1.6.34}$$

which can directly be expressed as a function of the variance

$$S < S' = \frac{1}{2} k_B \ln\left(1 + \frac{k_B T e^2}{\hbar} m \langle x^2 \rangle_c\right). \tag{1.6.35}$$

For the generalization to many–particle systems Schlitter made use of the covariance matrix

$$\sigma_{ij} = \left\langle (x_i - \langle x_i \rangle)\left(x_j - \left\langle x_j \right\rangle\right)\right\rangle, \tag{1.6.36}$$

obtained from a MD trajectory. Introducing mass weighted coordinates

$$X_i = \sqrt{m_i} x_i \tag{1.6.37}$$

the mass–weighted covariance matrix is given by

$$\sigma' = \mathbf{M}^{\mathbf{T}} \sigma \mathbf{M}, \tag{1.6.38}$$

with the $3N \times 3N$ mass matrix

$$\mathbf{M} = \mathrm{diag}\left(\sqrt{m_1}, \sqrt{m_1}, \sqrt{m_1}, \sqrt{m_2}, \sqrt{m_2}, \sqrt{m_2}, \ldots \sqrt{m_N}\right). \tag{1.6.39}$$

Diagonalizing the mass weighted covariance matrix yields the classical variances of the resulting new coordinates q_i in the diagonal elements. As the fluctuations are independent and uncorrelated the one dimensional approximation of the entropy can easily be applied to every coordinate separately, with the sum yielding the total entropy

$$S < S' = \frac{1}{2} k_B \sum_{n=1}^{3N} \ln\left(1 + \frac{k_B T e^2}{\hbar} \langle q_i^2 \rangle_c\right) \tag{1.6.40a}$$

$$= \frac{1}{2} k_B \ln \prod_{n=1}^{3N} \left(1 + \frac{k_B T e^2}{\hbar} \langle q_i^2 \rangle_c\right) \tag{1.6.40b}$$

The product in the logarithm can be written in terms of a determinant of a diagonal matrix with elements $\left(1 + \frac{k_B T e^2}{\hbar} \langle q_i^2 \rangle_c\right)$. As a determinant is invariant under orthogonal transformations — like the one used to diagonalize the mass weighted covariance matrix — the entropy estimate can directly be obtained from the mass weighted covariance matrix

$$S < S' = \frac{1}{2} k_B \ln \det\left(1 + \frac{k_B T e^2}{\hbar} \mathbf{M}^{\mathbf{T}} \sigma \mathbf{M}\right). \tag{1.6.41}$$

Schlitter's approach has the big advantage that the covariances in cartesian coordinates can be used for the entropy estimate. Although the covariance matrix may be singular, Equation (1.6.41) will yield reasonable results —

different from the normal mode approach. With the inclusion of every structure of a conformational ensemble for the approximation of the upper limit of the configurational entropy, Schlitter's method is expected to yield more reliable entropy estimates than the normal mode approach based on the crystal structure only.

1.6.4 Solvent Entropy

One entropic contribution to the free energy not yet considered is due to the increased ordering of solvent molecules at the surface of proteins with respect to bulk water. It may be approximated by a term proportional to the solvent–accessible surface.
For an explicit treatment of the solvent entropy the computational costly TI method can be applied or the faster permutation–reduced approach based on Schlitter's approximation as recently proposed by Reinhard and Grubmüller [161] may be used.

1.7 Thermodynamic End States Methods

In addition to the already presented statistical physics based methods using the system's Hamiltonian (Chapter 1.4) other techniques explicitly sum up the various enthalpic (and entropic) contributions of different states. Without the need to sample configurational space along a pathway but, instead, sampling around the initial and final state only, these methods are considerably faster than the former ones. Some approximation schemes even aim at sampling one state only and obtaining a conformational ensemble for the other state by crudely trimming the trajectory.
Due to thermodynamic fluctuations, end states cannot be described accurately by single conformations. As a protein's conformational flexibility causes a distribution of states around energy minima in the free energy landscape, a single structure can only be used either for empirical and statistical approximations or as a starting point for a structural sampling to cover larger parts of the conformational space. The latter can e.g. be achieved by MD or MC methods.

1.7.1 Linear Interaction Energy (LIE)

Åqvist et al. [52] developed a semi–empirical method for the calculation of binding free energies using molecular dynamics simulations. Based on the

assumption of a linear response for electrostatic and van der Waals interactions upon protein–ligand binding, the free energy function takes the simple form

$$\Delta G_{\text{binding}} = \alpha \left\langle \Delta E^{\text{vdW}} \right\rangle + \beta \left\langle \Delta E^{\text{es}} \right\rangle . \tag{1.7.1}$$

The electrostatic and van der Waals energies only take interactions between the ligand and its surroundings, i.e. solvent or a solvated, bound protein into account, internal energies are neglected. $\langle\cdot\rangle$ denotes the ensemble average (averaging over snapshots of a MD simulation).
Two trajectories are computed using MD simulations. One with the ligand bound to the protein in a water box, and the second contains only the freed ligand in a water box. ΔE in equation (1.7.1) denotes the change in interaction energy between the ligand and its surroundings with respect to the two trajectories.
While the electrostatics coefficient β is fixed to 0.5 due to the linear response approximation [162], empirical values for α range between 0.165 and 0.181 [52, 163]. $\beta = 0.5$ does not hold for non–ionic compounds and values between 0.33 and 0.5 were reported depending on the type of molecule [163]. Zoete et al. [164] also found different, sometimes even negative parameters depending on the system.
Different approaches including continuum solvent representations were suggested [163, 164] speeding further up the method. Huang and Caflisch [165] additionally substituted the time consuming MD simulations by an energy minimization of the ligand.
The LIE method is not universal as it needs a training data set for each investigated protein in form of experimental data in order to fit the system–dependent coefficients α and β.

1.7.2 MM/PBSA

A popular method for the calculation of free energies is the *Molecular Mechanics / Poisson–Boltzmann Surface Area* method (MM/PBSA) [51, 166–168]. In this method, the free energy is obtained as a sum of molecular mechanics force field terms (G_{MM}), the solute–solvent interaction obtained by the numerical solution of the Poisson–Boltzmann equation (G_{PB}), a non–polar solvation term proportional to the surface area approximating the solute–solvent van der Waals interactions (G_{SA}) and an estimate for the conformational entropy ($-TS$):

$$\overline{G}_{\text{MM/PBSA}} = \overline{G}_{\text{MM}} + \overline{G}_{\text{PB}} + \overline{G}_{\text{SA}} - TS, \tag{1.7.2}$$

with

$$G_{\mathrm{MM}} = G_{\mathrm{bond}} + G_{\mathrm{angle}} + G_{\mathrm{dihedral}} + G_{\mathrm{LJ}} + G_{\mathrm{Coulomb}}. \tag{1.7.3}$$

The energetic contributions are averaged over a set of snapshots (with removed solvent) [166]. The entropy is estimated using NMA techniques.
MM/PBSA is frequently applied for the calculation of binding affinities according to

$$\Delta G_{\mathrm{binding}} = G_{\mathrm{compound}} - G_{\mathrm{partner\ 1}} - G_{\mathrm{partner\ 2}}. \tag{1.7.4}$$

Usually, only the trajectory of the compound is sampled and the free energies of the isolated partners are analyzed on extracted structures from the compound trajectory (see e.g. Reference [169]). This scheme on the one hand is computationally inexpensive. On the other hand, the above free energy function probably has defiances in estimating free energy changes due to conformational changes induced by unbinding (see also [170]).
A variant of the MM/PBSA method uses the Generalized Born formalism instead of the Poisson–Boltzmann equation for the polar solvation interaction (MM/GBSA) [171]. In contrast to the slower but more accurate MM/PBSA, MM/GBSA allows also the energy decomposition per atom. Thus, enabling a residue–resolved resolution of the energetic contributions to binding.
Compared to the LIE method both variants, MM/PBSA and MM/GBSA, are universal as these are parameter–free methods.
Kuhn and Kollman [172] compared the LIE approach and the MM/PBSA method to experimental data computing binding affinities between non–peptide ligands and avidin and between a hexapeptide and streptadivin. While MM/PBSA results were in excellent agreement to experiment (correlation $r = 0.92$), the LIE method performed poor with a correlation of $r = 0.55$ (after fitting its parameters to the system).
The MM/PBSA method may also be used for the prediction of mutation–induced changes in folding free energy, e.g. Zoete and Meuwly [173].

1.8 Statistical, Empirical Approaches

Statistical, empirical approaches for the structure–based free energy estimates of biomolecular systems derive their potentials via statistical analysis of e.g. the relative frequency of contacting residues, combined with physical effective energy functions and structural descriptors [174]. The energy terms are subsequently weighted to reproduce experimental data. Usually, these methods rely on only one structure per folded state.

Next to the Fold-X method [59] presented in the next subsection, examples for these empirical potentials are, the Eris method [175, 176] or an approach developed by Bordner and Abagyan [174] introducing a parameterized denatured state.

1.8.1 Fold-X

A popular method is the empirical Fold-X energy function, developed by Guerois et al. [59] based on data from protein stability experiments applied on static configurations that are only for optimized hydrogen bond networks. The mutational change in folding free energy in Fold-X is given by

$$\begin{aligned}\Delta G = {} & W_{\mathrm{vdW}}\Delta G_{\mathrm{vdW}} + W_{\mathrm{solvH}}\Delta G_{\mathrm{solvH}} + W_{\mathrm{solvP}}\Delta G_{\mathrm{solvP}} + \Delta G_{\mathrm{wb}} \\ & + \Delta G_{\mathrm{hbond}} + \Delta G_{\mathrm{el}} + W_{\mathrm{mc}} T\Delta S_{\mathrm{mc}} + W_{\mathrm{sc}} T\Delta S_{\mathrm{sc}}.\end{aligned} \tag{1.8.1}$$

The W_x denote weighting factors, *vdW* the van der Waals contributions, *solvH* and *solvP* the solvation of hydrophobic and polar groups, respectively, *wb* water bridges (water molecules with more than one hydrogen bond to the protein leading to stabilization), *hbond* intra–molecular hydrogen bonds, *el* electrostatic interactions, *mc* main chain (backbone), and *sc* side chain contributions to the entropy. This energy function is computed for a single native wild type and the mutated conformation, respectively.
Solvent exposure effects are included by applying an additional scaling to the atomic contributions. In contrast to the weighting factors W_x these scaling parameters are not fitted to experimental data but constructed using the atomic occupancies

$$\mathrm{Occ}_i = \sum_{j, d_{ij}<6\text{Å}} V_j e^{\frac{d_{ij}^2}{2\sigma^2}}, \tag{1.8.2}$$

with the fragmental volume V_j of atom j (taken from a predefined table), d_{ij} is the distance between atoms i and j, and $\sigma = 3.5\,\text{Å}$ (corresponding roughly to the minimum of the van der Waals potential for two heavy atoms). The scaling factor itself writes

$$S^i_{\mathrm{fact}} = \frac{\mathrm{Occ}_i - \mathrm{Occ}^i_{\mathrm{min}}}{\mathrm{Occ}^i_{\mathrm{max}} - \mathrm{Occ}^i_{\mathrm{min}}}. \tag{1.8.3}$$

The minimal and maximal reference values for the occupancy have been statistically derived from a protein structure database. If the calculated occupancy is less than the minimal value, S^i_{fact} is set to 0, if its value is larger

than the maximum, the scaling factor is set to 1. S^i_{fact} is applied to any energy term the atom i is involved in.

While the energy values per atom for van der Waals contributions, solvation and hydrogen bonds were determined from an experimental data set, the electrostatic contributions are calculated according to Coulomb's Law including ionic screening effects via

$$E_{ij} = \frac{332 q_i q_j}{\varepsilon d_{ij}} e^{-d_{ij}\kappa}, \tag{1.8.4}$$

with the Debye–Hückel parameter κ (see Equation 1.5.18 and associated footnote).

The water bridge contribution includes a prediction for water positions followed by the evaluation of an energy term combining hydrogen bond energy, solvation costs for water burial, and several entropic contributions:

$$\Delta G_{\text{wb}} = N_{\text{hb}} \Delta G_{\text{hb}} + S_{\text{fact}} \Delta G_{\text{solvW}} + \delta S_{\text{prot}} + (1 - S_{\text{fact}}) S^{\text{max}}_{\text{wat}} + S_{\text{fact}} S^{\text{min}}_{\text{wat}} . \tag{1.8.5}$$

Here, N_{hb} denotes the number of hydrogen bonds, ΔG_{hb} the hydrogen bond energy, S_{fact} the previously obtained scaling factor regarding exposure, ΔG_{solvW} the solvation free energy contribution for water burial, δS_{prot} the entropic cost for fixing the backbone or the side chain involved in the water bridge. The minimal and maximal water entropies $S^{\text{min}}_{\text{wat}}$ and $S^{\text{max}}_{\text{wat}}$ account for the entropic cost of fixing the water molecule fully buried or fully exposed, respectively. The water bridge term is only added if it is smaller than 0.

If van der Waals clashes occur in the structure a correction energy

$$\Delta G_{\text{clash}} = S_{\text{fact}} \left(R_i + R_j - 0.35\,\text{Å} - d_{ij} \right) \frac{\text{kcal}}{\text{mol}\,\text{Å}} \tag{1.8.6}$$

is added, where R_i is the corresponding atomic radius of atom i.

Unfolded states are not explicitly taken into account. However assuming similar properties for the unfolded states of single point mutants with respect to the denatured wild type, they are included implicitly with the use of the weighting factors W_{x}.

The involved parameters and the five weighting factors were trained on a data set of 339 mutants and tested on 667 mutants resulting in a correlation coefficient of 0.83 and a standard deviation 0.81 $\frac{\text{kcal}}{\text{mol}}$ upon neglect of 5% outliers.

1.9 Bioinformatic Techniques

Algorithms lend from computer sciences are also used for sometimes crude estimations of free energies. Both qualitative and quantitative estimates may be obtained by machine learning algorithms like the artificial neural network (ANN) or support vector machines (SVM) that are trained on experimental data and e.g. combined with the protein sequence or structure [62]. Similar to the above presented, structure–based statistical methods, weights are fitted to an energy function. However, in the case of machine learning techniques the energy function is not a linear combination of energy terms but consists of more complex nonlinear terms. These terms are often lacking a physical basis and make use of a large amount of descriptors. Thus, these methods are capable of the prediction of free energies, however, without supplying knowledge of the underlying physico–chemical processes. With a given suitable training set, machine learning algorithms can also be used to predict free energies for cases where physical functions fail, e.g. due to missing crystal structures. Another advantage of ANNs or SVMs is the small computational effort allowing even the prediction of e.g. 10^{12} data points. However, over–fitting of parameters is a frequently occurring problem.

Examples of machine–learning based methods are I-Mutant (`http://gpcr2.biocomp.unibo.it/~emidio/I-Mutant/I-Mutant.htm`) [62, 177] or MU-Pro (`http://www.ics.uci.edu/~baldig/mutation.html`) [63].

I-Mutant, for example, predicts the folding free energy upon mutation with only the protein sequence as input to an SVM. It achieves a correlation of $r = 0.62$ and a root mean square standard error of 1.45 kcal/mol for a data set comprising 2048 mutants from 64 proteins. The method shows best agreement to experimental data for conservative mutations of apolar amino acids.

Chapter 2

Structure Determination and Experimental Free Energy Measurements

2.1 Structure Determination

The experimental determination of protein structures (as found in the protein data bank `http://www.pdb.org` [6]) usually involves complex and difficult procedures. Especially the crystallization of membrane proteins for X–ray diffraction experiments is tedious and often even impossible. This chapter provides a brief introduction to structure determination by X–ray crystallography and nuclear magnetic resonance (NMR) spectroscopy as well as to experimental methods applied in the study of protein stabilities and protein–ligand binding affinities.

2.1.1 X–ray Crystallography

In X–ray crystallography, the atomic structure is solved after obtaining the electron density through X–ray diffraction measurements [22, 33].
When placing a crystal in an X–ray beam, the atoms' electrons inside the crystal scatter the incoming electromagnetic waves resulting in a reflection picture. A regular arrangement of particles within the crystal leads to a regular scattering. While most interferences are destructive, spots can be found in the diffraction pattern. Bragg's law

$$2d \sin \vartheta = n\lambda \tag{2.1.1}$$

applies for these maxima, where d is the separation of grid planes responsible for the interference, ϑ denotes the angle between the incoming beam and the reflecting surfaces (the incoming X–ray is diffracted at an angle of 2ϑ), n is an integer and λ the wavelength of the monochromatic X–ray.
In order to obtain enough data to construct the three dimensional structure,

the protein crystal has to be rotated during the measurement and a diffraction pattern has to be recorded for every orientation.
After the unit cell of the crystal is determined, the collected data is indexed: every spot in the diffraction pattern corresponds to a set of reflection sheets that can be described by Miller indices (h,k,l). The indices correspond to the inverse intersections with the lattice vectors.
Information on the desired crystal structure is not just hidden in the position of the spots, but also in the spots' intensities $I(h,k,l)$. Being proportional to the squared amplitudes $F(h,k,l)$

$$I(h,k,l) \propto F^2(h,k,l) \tag{2.1.2}$$

the intensity does not comprise any information concerning the phase $\alpha(h,k,l)$ of the electromagnetic wave. However, this property is crucial for the reconstruction of the protein structure.
Via Fourier transformation, the three dimensional electron density ρ can be written as

$$\rho(x,y,z) = \frac{1}{V} \sum_{h,k,l} F(h,k,l) e^{i\alpha(h,k,l)} e^{-2\pi i(hx+ky+lz)}, \tag{2.1.3}$$

with the still unknown phase $\alpha(h,k,l)$ and the volume V of the unit cell.
The phase is refined starting from an initial guess. Next to the direct method using known relations between reflexes or utilizing a similar molecule to fit on the imperfect electron density, changing the molecular composition can help to solve the phase problem. Including either heavy metal atoms or replacing the sulfur in methionine by selene (mutating to selene–methionine), anomalies are introduced in the scattering pattern. The inner electrons of these atoms also contribute to scattering and introduce a known phase. Also, diffraction patterns for three wavelengths — one near the adsorption maximum of the heavy atoms, one above, one below may be used to determine the phase.
From this initial phase an initial model structure is fitted to the electron density. This model leads to a calculated set of complex spot amplitudes $F(h,k,l)e^{\alpha(h,k,l)}$ which in turn yield a new electron density. This iterative procedure is repeated until the largest possible correlation is obtained.
If some residues exist in various conformations, the average electron density is smeared over a large area not detectable any more by X–ray diffraction. If an atom takes over a small number of distinct positions, it will appear multiple times in the density map.
The final structures show different resolutions. While structures with a resolution of ≤ 2Å can reliably be used e.g. in MD simulations, lower resolu-

tions introduce errors that can lead to wrong conformations of amino acid side chains. Depending on the quality of the crystal, even the assignment of hydrogen atoms to the electron density may be possible (resolution < 1 Å).
Prior to the X–ray diffraction experiment, the most difficult task is to crystallize the protein. Hereby the endeavor becomes more demanding with growing size and decreasing solubility of the molecule.
Ideally, the crystal is pure, without defects, of high regularity, and, obviously, the protein should remain folded in the crystal.Different crystallization conditions have to be screened for each protein. Hundreds of surroundings are tested: different temperatures, pH, different salts at different concentrations, additives that stabilize the protein fold, substances that help to grow crystals, and many more.
Depending on the realization of the experiment, the crystal is cooled down with liquid nitrogen, reducing radiation damage and lowering noise due to thermal motions.
In theoretical studies based on crystal structures care has to be taken since the protein surface may reflect artifacts due to crystal packing. Contacts in the crystal between two molecules may alter the conformation of side chains or of exposed loop regions. In general, crystallization conditions do not represent physiological states. This problem may be (partially) circumvented by a simulation of a single structure in a water box at physiological settings.
Minor conformational changes due to interactions with additives, or a change in pH, salts or temperature, can be corrected by carefully preparing the system for simulations.

2.1.2 NMR Spectroscopy

Nuclear magnetic resonance spectroscopy of biomolecules is used to solve a protein structure by locating atoms with a half–integer spin like ^{1}H, ^{13}C and ^{15}N [22, 23, 34, 178].
A particle with a spin quantum number I shows an absolute value for its spin of

$$J = \sqrt{I(I+1)}\hbar. \tag{2.1.4}$$

In an external magnetic field $\vec{B} = (0,0,B_z)$ a particle with a spin will orient along the magnetic field and precess. An atom with spin $I = \frac{1}{2}$ can adopt two magnetic quantum numbers $m = \pm\frac{1}{2}$ with a spin contribution in z–direction of

$$J_z = m\hbar. \tag{2.1.5}$$

Due to the linkage to the magnetic moment via the material–specific gyromagnetic ratio γ

$$\mu_z = m\gamma\hbar \tag{2.1.6}$$

the two quantum states have different electric energies

$$E_m = -m\gamma\hbar B_z. \tag{2.1.7}$$

The difference in energy

$$\Delta E = \gamma\hbar B_z \tag{2.1.8}$$

determines the distribution of quantum states according to the Boltzmann distribution

$$\frac{N_{-\frac{1}{2}}}{N_{\frac{1}{2}}} = e^{\frac{\gamma\hbar B_z}{k_B T}} \tag{2.1.9}$$

with the number $N_\pm$ of particles found in the quantum states + and −,
and the resonant frequency — also called Larmor frequency —

$$\omega_0 = \frac{\Delta E}{\hbar} = \gamma B_z. \tag{2.1.10}$$

The bound and unbound neighborhood of an atom induces a shift in the Larmor frequency due to its inherent magnetic field (*chemical shift*).
In the course of an experiment a short magnetic pulse in x–direction acts on the spin systems of the protein, forcing the spins to rotate by 90° along the x–axis and reorient in y–direction.
Assuming a first order kinetic law, the relaxation of the magnetization $\vec{M}$ obeys Bloch's equations:

$$\frac{d}{dt}M_{x,y} = \gamma(\vec{M}\times\vec{B})_{x,y} - \frac{M_{x,y}}{T_2} \tag{2.1.11a}$$

$$\frac{d}{dt}M_z = \gamma(\vec{M}\times\vec{B})_z + \frac{M_0 - M_z}{T_1}, \tag{2.1.11b}$$

with the equilibrium magnetization M_0, the spin–lattice relaxation time T_1 and the spin–spin relaxation time T_2.
Studying the T_2–relaxation after a 90° pulse, the Fourier transformation of the so–called free induction decay yields information about the neighborhood of an atom that can be used later on to solve the structure. As many peaks can overlap, assigning protons or other atoms may be difficult or even impossible. In multi–dimensional NMR spectroscopy — different puls sequences with varying intervals — improve the assignment.

After the signals are assigned to amino acid protons, so–called cross peaks provide information about intramolecular distances and dihedral angles. Based on distance and angle restraints derived from the signals and using other general information about non–measured atom types, protein structures are modelled fulfilling these restraints. Since the distance restraints do not define one unique protein structure, typically 20 structures for one protein extracted from the same NMR data are deposited in the protein data bank.
In contrast to crystallographic approaches, the protein is in solution for NMR measurements. In order to measure carbon or nitrogen atoms, these particles have to be substituted by their less common, heavier isotopes ^{13}C and ^{15}N with spin $\frac{1}{2}$. Structure determination using NMR is especially demanding for large proteins as signals are overlapping and since the relaxation times are shorter for larger molecules.

2.2 Stability

2.2.1 Thermal Unfolding

Protein stability is frequently measured in differential scanning calorimetry experiments (DSC) [179–182] used during thermal unfolding. With increasing temperature, the heat absorption of proteins is measured. Parallel to the solvated sample, a reference chamber only containing the solvent with the same volume is mounted. Both test samples are exposed to a constant increase in temperature $\frac{dT}{dt}$.
Covering the transition from folded to the denatured conformation by thermal unfolding, the resulting DSC scan yields some thermodynamic quantities:
The primal outcome is the specific heat c_p as a function of the temperature T. Integrating the heat capacity yields the change in calorimetric enthalpy ΔH_{cal}

$$\Delta H_{\text{cal}} = \int_{T_1}^{T_2} c_p \, dT. \tag{2.2.1}$$

When choosing T_1 and T_2 as initial and final temperature for the transition peak, ΔH_{cal} is the change in enthalpy due to protein denaturation. If the unfolding is a two–state process, the enthalpy can be refined using the van't Hoff enthalpy [180]

$$\Delta H_{\mathrm{vH}} = \frac{4 k_B T_m^2 c_p^{\max}}{\Delta H_{\mathrm{cal}}}, \tag{2.2.2}$$

with the denaturation mid–point temperature T_m, at which half of the protein concentration is unfolded, and the maximum excess heat capacity $c_p^{\max}$. The folding free energy at temperature T is then found with the Gibbs–Helmholtz equation

$$\Delta G(T) = \Delta H_m \left(\frac{T_m - T}{T_m} \right) - \Delta c_p \left(T_m - T + T \ln \frac{T}{T_m} \right), \tag{2.2.3}$$

where Δc_p is the change in heat capacity during denaturation.

2.2.2 Chemical Denaturation

In chemical denaturation experiments, denaturants like urea or guanidine hydrochloride are used to shift the chemical equilibrium towards the unfolded state. Using the most common techniques like absorbance or fluorescence spectroscopy as well as circular dichroism experiments [22, 23] the equilibrium concentrations of the native and denatured state can be measured. Combined with *stopped flow* [22] methods, also the folding rates can be quantified.

Equilibrium Measurements

Assuming a proportionality between the folding free energy ΔG and the denaturant concentration c_{D} via

$$\Delta G(c_{\mathrm{D}}) = \Delta G_{\mathrm{H_2O}} - m c_{\mathrm{D}}, \tag{2.2.4}$$

with the folding free energy in water $\Delta G_{\mathrm{H_2O}}$ and the proportionality coefficient m, the folding free energy is obtained by measuring concentrations of folded and unfolded proteins as a function of the denaturant concentration as described below.
At the denaturation midpoint with equal amounts of folded and unfolded proteins, the free energy change upon folding is 0 and the denaturant density c_{D}^m yields the stability in pure water

$$\Delta G_{\mathrm{H_2O}} = m c_{\mathrm{D}}^m. \tag{2.2.5}$$

Denaturation curves are fitted to this equation in order to obtain both m and c_{D}.

While m–values can vary for mutations and possibly provide information about similarities in unfolded states, averaged m–values $\overline{m}$ from all mutants may be used, as the resulting equation

$$\Delta G = \overline{m} c_{\mathrm{D}}^{m} \tag{2.2.6}$$

holds also if the linearity assumed in equation (2.2.4) does not.
A more general formula is given by

$$\Delta G(c_{\mathrm{D}}) = m \left(c_{\mathrm{D}}^{m} - c_{\mathrm{D}} \right) \tag{2.2.7}$$

and may be used to measure the folding free energy as a function of the denaturant density.

Stopped Flow Method

In *stopped flow* experiments a solution containing folded proteins in water is rapidly mixed with a solvated denaturant within 1 ms. Also refolding experiments in a similar manner are possible.
The folding rates are measured with e.g. the circular dichroism method described below.
Similar to equation (2.2.4) a linearity is assumed between the logarithm of the (un–) folding rate $k_{\mathrm{f/u}}$[1] and the denaturant concentration c_{D}:

$$\ln k_{\mathrm{f/u}}(c_{\mathrm{D}}) = \ln k_{\mathrm{f/u}}^{\mathrm{H_2O}} + m_{\mathrm{f/u}} c_{\mathrm{D}}, \tag{2.2.8}$$

with the proportionality factor $m_{\mathrm{f/u}}$.
Fitting the experimental feasible range to this equation yields $\ln k_{\mathrm{f/u}}^{\mathrm{H_2O}}$ and $m_{\mathrm{f/u}}$. The free energy change upon folding is given by

$$\Delta G = -k_B T \ln \frac{k_f^{\mathrm{H_2O}}}{k_u^{\mathrm{H_2O}}}. \tag{2.2.9}$$

If the mutants measured in the experiment show only small fluctuations in the folding or unfolding proportionality factor $m_{\mathrm{f/u}}$, no extrapolation for the folding rates is necessary, as these terms cancel when calculating the change in folding free energy $\Delta\Delta G$ upon mutation

$$\Delta\Delta G = -k_B T \left(\ln \frac{k_f^{\mathrm{mut}}(c_{\mathrm{D}})}{k_u^{\mathrm{mut}}(c_{\mathrm{D}})} - \ln \frac{k_f^{\mathrm{wt}}(c_{\mathrm{D}})}{k_u^{\mathrm{wt}}(c_{\mathrm{D}})} \right), \tag{2.2.10}$$

[1] The indices f and u denote the folded and unfolded state, respectively, throughout the entire chapter.

where the upper index *wt* or *mut* refer to the wild type or mutant, respectively. Thereby, different denaturant concentrations for unfolding and refolding are allowed.

Absorbance Spectroscopy

The absorbance A of incoming light is measured in absorbance spectroscopy to obtain information about protein concentrations applying the Lambert-Beer law [22, 23]

$$A(\lambda) = \log_{10} \frac{I_{\text{in}}}{I_{\text{out}}} = \varepsilon(\lambda) c d, \tag{2.2.11}$$

with the intensities of incoming and outgoing light $I_{\text{in/out}}$, the sample concentration c, the thickness of the sample d and the extinction coefficient $\varepsilon(\lambda)$.
In the case of a mixed solution consisting of native and denatured proteins, equation (2.2.11) can be rewritten as

$$A(\lambda) = \left(\varepsilon_f(\lambda) c_f + \varepsilon_u(\lambda) c_u\right) d. \tag{2.2.12}$$

Thus from a fully recorded spectrum the concentrations of folded and unfolded proteins can be deduced.
A prerequisite for this method is the absence of light scattering and photoreactions as well as an homogenous distribution of the solute.

Fluorescence Spectroscopy

Fluorescence is found in materials where a photon is absorbed and, after a short time of several nanoseconds, a photon with lower energy is emitted again. Upon absorption, an electron is raised from its ground state to an excited state and by emitting light it falls back to its ground state or to a low lying excited state. Due to relaxing oscillations the outgoing light typically has a longer wavelength than the incoming one.
In contrast to phosphorescence, the spin of the involved electron never changes. Tyrosine, phenylalanine, and the stronger fluorophore tryptophane may be used as absorbers and emitters in fluorescence spectroscopy [22, 23].
With the quantum efficiency ϕ

$$\phi = \frac{\text{number of absorbed photons}}{\text{number of emitted photons}} \tag{2.2.13}$$

the fluorescence intensity F can be written as

$$F = 2.303 I_{\text{in}} \phi \varepsilon c d, \tag{2.2.14}$$

with the sample concentration c, the thickness of the sample d and the extinction coefficient $\varepsilon(\lambda)$.
The quantum efficiency can also be described by excited state decay rates

$$\phi = \frac{k_f}{\sum_i k_i}, \tag{2.2.15}$$

with the decay rate due to fluorescence k_f and the sum of all possible decay rates $\sum_i k_i$.
The surroundings of the fluorophore can shift the absorption curve and its maxima. A fully buried tryptophane will give a different spectrum as when being exposed. Thus, with spectra of fully denatured and native proteins analyzed, the midpoint spectrum yields the concentration of both shapes.

Circular Dichroism

Circular dichroism exploits the chirality of chemical groups found in proteins [22]. Circular polarized light is absorbed differently depending on whether it is right or left circular polarized light yielding different extinction coefficients ε_r and ε_l.
Using incoming linearly polarized light, that is a linear combination of right and left circular polarized light, the ellipticity θ can be measured which yields the concentrations of folded and denatured proteins (with known extinction coefficients) via

$$\theta(\lambda) = \text{const.}\,(\varepsilon_r - \varepsilon_l)\, c d. \tag{2.2.16}$$

Again analyzing a full spectrum of native and unfolded conformations as well as the denaturation midpoint yields the desired concentrations (e.g. at the midpoint) of folded and unfolded proteins.

2.3 Affinity

For quantitative affinity measurements many experimental instruments are at hand. While kinetic binding rates may be measured with the methods described above, other techniques like surface plasmon resonance or isothermal titration calorimetry can be used for measuring protein binding affinities.

Surface Plasmon Resonance

Reflecting light at the surface of a metal film shows a reduced intensity at a certain angle [183]. When using a prism as coupling medium, this angle mainly depends on the refraction index of the surface not facing the prism. In surface plasmon resonance experiments, proteins are attached to this surface and a solution with its binding partner flows underneath it. Protein–protein binding leads to a change in the refraction angle and, thus, the angle of maximum absorbance changes. The rate directly yields the binding rate k_{on}, and subsequently using a solution without the binding partner, the dissociation rate k_{off} can be determined.
The binding free energy is then given by

$$\Delta G = k_B T \ln \frac{k_{\text{on}}}{k_{\text{off}}}. \tag{2.3.1}$$

Isothermal Titration Calorimetry

From a titration curve obtained by isothermal titration calorimetry the binding affinity constant K_a can be directly derived [184]. Via

$$\Delta G = -k_B T \ln K_a \tag{2.3.2}$$

the binding free energy is directly accessible, too.
Two cells are kept at the same temperature via an external heater. One cell contains the sample solution, in which the ligand is titrated, the other one only includes the solute. The addition of ligands leads to peaks in the heater's power supply, which due to a feedback keeps both cells at the same temperature. Positive or negative peaks point to endothermic or exothermic binding, respectively, and give the titration curve.

Chapter 3

Method Development for Protein Stability Calculations

Concoord/PBSA, the approach presented here, is based on a modified version of the MM/PBSA energy function. Instead of applying the physical effective free energy function to snapshots taken from a computationally expensive MD trajectory, ensembles of structures fulfilling specific geometric constraints deduced from the input structure were taken here.
MM/PBSA methods are rarely used for mutational stability studies, as they also require assumptions and approximations for the unfolded state [173]. For the development of the Concoord/PBSA procedure the problem of protein stability was deliberately chosen as a first test case to overcome the lack of a fast method for the prediction of stability free energies taking protein flexibility into account.
The chapter is organized as follows:

- **Materials: Selection of Data Set** reports the experimental data used in this chapter to adjust the developed technique to experiment.
- **Preliminary Methods** that are crucial for the novel method but that are not mentioned before: The sampling of conformational space that is used here, and the application of a thermodynamic cycle for the calculation of folding free energy changes upon mutation.
- **Methods: Concoord/PBSA** is a novel protocol to predict folding free energy changes upon mutation starting from a crystal or NMR structure.
- The **Results** section reports on performance of Concoord/PBSA on the test set.
- The **Concoord/PBSA Web Interface** is presented in the following section (3.5). Performance issues are also considered here.
- A **Comparison to Fold-X** is followed by a variety of possible

- **Alternatives and Variations** of the method that came up during development.
- A **Discussion** closes the chapter.

3.1 Materials: Selection of Data Set

Protein X-Ray Structures and Experimental Free Energies

As experimental data set we used the (pseudo–) wildtype X–ray structures (Protein Data Bank codes in brackets) of colicin immunity protein Im7 (1AYI), chymotrypsin inhibitor–2 (1YPC), B1 domain of protein L (1HZ6), B1 immunoglobulin–binding domain from streptococcal protein G (1PGA), Staphylococcal Nuclease (1STN), Bacteriophage T4 Lysozyme (2LZM), and of Salmonella typhimurium CheY (3CHY) combined with folding free energies of mutants of these proteins.
References of the crystal structures and the experimental free energies, as well as the number of mutations considered for each protein are listed in Table 3.1. The folding free energies for every mutation used in the folding study are reported in Table A.1.
A total number of 582 mutations has been investigated in this work including 425 conservative mutations preserving charges, 154 non-conservative ones (charged to neutral changes), 326 solvent exposed and 253 buried locations.
The data set comprises experimental folding free energies measured by different techniques under different conditions. The test set is composed partly of the training data set used by Guerois et al. [59] for the development of the Fold-X method. Other proteins and mutations with known structures and experimental energies were added to enlarge the diversity of used proteins.
A more systematic way of choosing a test set according to specific search criteria applied to the protherm data bank [185] as previously done by Capriotti et al. [62], Saraboji et al. [186], and partly by Guerois et al. [59] proved to be inefficient. Narrowing the list of mutations with respect to temperature, pH, denaturing conditions or experimental methods yields lists, that have to be carefully combed through since often only crystal structures with large gaps, bad resolution, ligands or mutations are available, rendering a structure–based prediction inaccurate or even impossible for these mutants. The few remaining mutants would be an inappropriate data set for further method development.
The need to shorten the mutation list from Guerois et al. [59] was due to similar problems mentioned above, and due to the restrictions to small proteins

(with less than 200 amino acids) during method development.

Table 3.1: Proteins and Mutations used as test set for the development of Concoord/PBSA

PDB	Resolution	PDB-Reference	Mutation-References	#Mutations
1AYI	2.0Å	[187]	[188]	26
1HZ6	1.7Å	[189]	[190]	68
1PGA	2.07Å	[5]	[191]	30
1STN	1.7Å	[192]	[182, 193–196]	265
1YPC	1.7Å	[197]	[198]	76
2LZM	1.7Å	[199]	[180, 200–223]	82
3CHY	1.66Å	[224]	[225]	35

3.2 Preliminary Methods

3.2.1 Sampling of Conformational Space

The idea of using distance geometry to calculate a structure was first proposed by Crippen [226], which proved to be a major tool in solving protein structures from NMR distance restraints. Later the concept was picked up to reproduce native backbone conformations [227] or whole protein configurations [64]. In this work the program CONCOORD (`http://www.mpibpc.gwdg.de/groups/de_groot/concoord/`) developed by de Groot et al. [64] was used for conformational space sampling.
In the generation of random structure ensembles using CONCOORD three steps can be distinguished (the workflow of Concoord is depicted in Figure 3.1):

Blueprint
In the first step, the distance of every occurring atom pair is measured and classified with respect to its interaction type and intramolecular function (e.g. an atom pair that is part of a quadruple defining a dihedral angle or part of a secondary structure element). Different interactions lead to distinct fluctuations of distances. To roughly mimic the structural flexibility at physiological temperature a distance range is assigned by adding and subtracting a type–dependent distance limit D. These distance restraints defined by upper and lower boundaries should not be violated by the resulting structures.

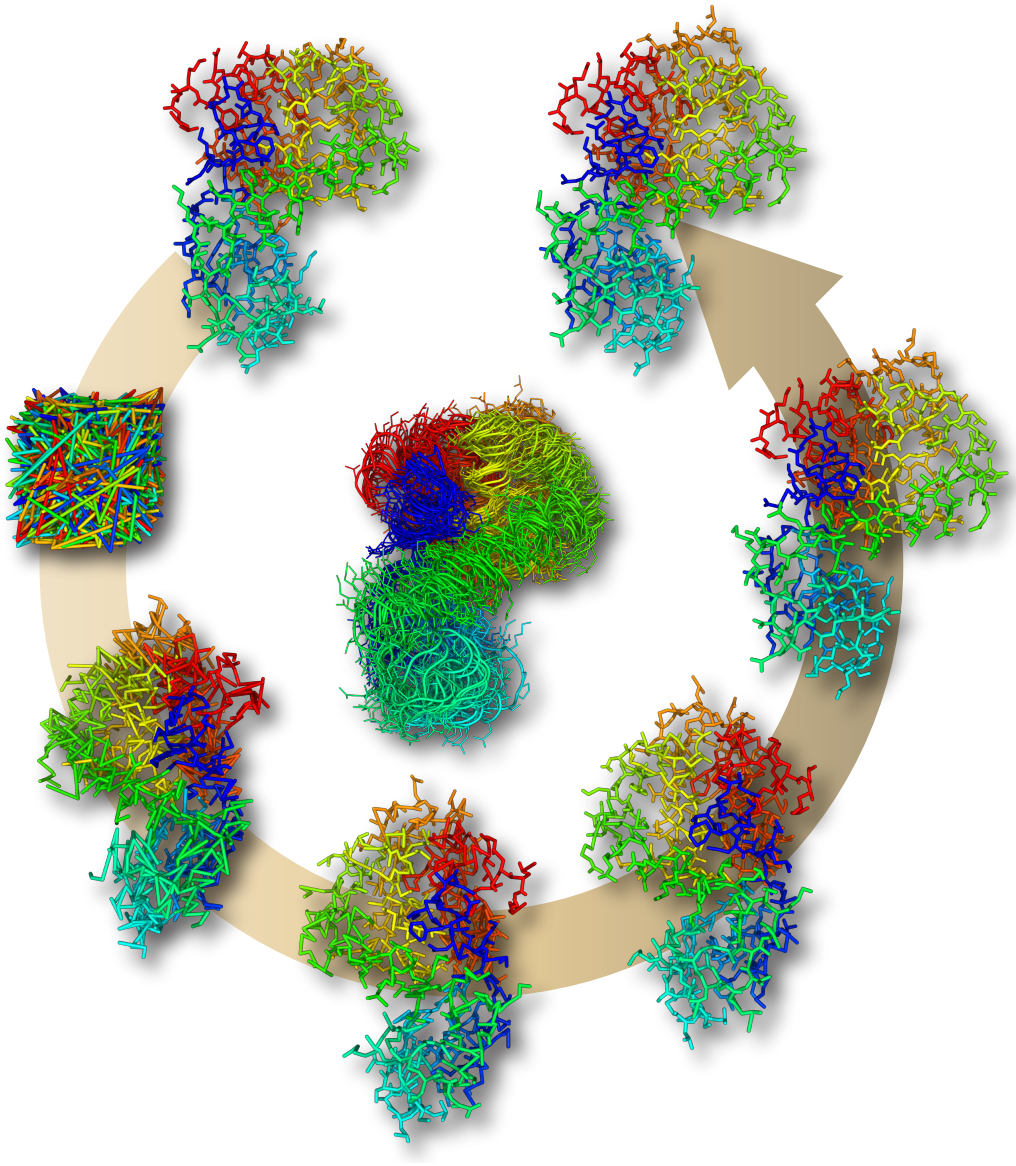

Figure 3.1: Sampling of conformational space using the Concoord method. The input structure (2LZM [199]; at the beginning of the circular arrow) is used for the generation of distance constraints. The cubic shape is the initially generated random structure. The next three structures along the arrow represent the first three iterations quickly converging towards a vague mirror image of the input structure. In this example, the eleventh step (next structure in row) then showed the correct chirality, until in the 88th step (resulting conformation the arrow points to) all distances fulfilled the given distance criteria. In this way a set of independent conformations can be generated (center)

Besides covalently bound atom pairs (*1–2 restricted pairs*), atoms connected by two (*1–3 restricted pairs*) or three (*1–4 restricted pairs*) bonds, and also, restraints for general, non–bonded atom pairs are distinguished. Depending on the underlying dihedral angle, the 1–4 interactions are split into different types shown in Table 3.2. Other restraints are for example *ring restrictions* applied to atom pairs found in rings (e.g. in the tryptophane side chain) or for hydrogen bonded pairs.
The different classifications and their appropriate distance limits are given in Table 3.2.

Table 3.2: Concoord distance classifications and margins D. Omega, phi and psi denote the dihedral angles found in the backbone of a peptide chain. Data taken from de Groot et al. [64]. Classifications for the latest version of Concoord exceed the number of 50 and are not shown.

Classification	Margin D[nm]
1-2	0.002
1-3	0.005
Ring	0.01
double bond 1-4	0.01
Omega 1-4	0.01
Tight phi/psi 1-4	0.02.
Loose phi/psi 1-4	0.04
Other phi/psi 1-4	0.03
Other 1-4	0.04
Secondary structure	0.05
Salt bridges	0.075
Hydrogen bonds	0.05
Tight hydrophobic	0.05
Loose hydrophobic	0.1
All other pairs	0.5

The resulting list of constrained distances can be seen as a blueprint or a construction plan for the protein structure, strongly depending on the input structure. Optimized, high resolution input structures usually yield better constraints in terms of convergence and a faster structure generation.

Random Structure Generation
Structure generation based on the before computed distance restraints starts with initial random assignment of coordinates within the limits of a defined

cubic box (see the cubic shaped random conformation in Figure 3.1).

Iterative Correction
The non–physical random structures are iteratively corrected to fulfill the distance constraints in a SHAKE–like manner [228] (see Page 22 ff. for a description of the SHAKE algorithm [78]). The resulting structures fulfill all distance constraints. And as they rely only on the predefined constraints, they are independent of each other. Neglecting energetic barriers, these sampled conformations (based on geometrical considerations only) largely cover the accessible conformational space close to the input structure.

With this procedure, representative ensembles are generated by 2 to 3 orders of magnitude faster than with MD methods. For example, the MD simulation of an MHC–peptide complex solvated in explicit water with a total of ca. 100,000 atoms takes 25 days for 20 ns on eight Intel Xeon 3.2 GHz processors using Gromacs [81]. In contrast, one may sample 500 structures of the complex using CONCOORD within three days on one processor of the same type.
For free energy calculations it turned out that 300–600 sampled conformations are sufficient (shown in Section 3.4.3). Also, a required short optimization of the random configurations renders this method computationally still less expensive than MD.
For the development of Concoord/PBSA an unofficial build of CONCOORD version 2.1 (build date April 21 2006) provided by Bert de Groot was used.

3.2.2 Denatured State Approximation and Thermodynamic Cycle

The folding free energy of a protein ΔG^{fold} is given as a difference between the free energy of the folded (*native*) state G^{native} and of the denatured state $G^{\text{denatured}}$:

$$\Delta G^{\text{fold}} = G^{\text{native}} - G^{\text{denatured}}. \tag{3.2.1}$$

However, while the native state may be described by available crystal structures or homology models, structural information for the characterization of the denatured state is lacking [229]. Therefore, in previous theoretical studies, the free energy of the denatured state was either approximated to be linearly related to the free energy of the folded state [230] or single contributions to the free energy were obtained based on the protein sequence or fragments of the folded state (see [231, 232] for polar contributions and

[233–235] for non–polar solvation contributions, and [117, 222, 236–238] for entropic contributions). In other approaches, unfolded conformations were constructed by *pumping up* the native conformation [239], by Monte Carlo sampling [240], or approximated by random coil models [241].
It is difficult or even impossible to obtain absolute folding free energies ΔG^{fold} by physical means. In contrast, for the calculation of relative changes in stability upon mutation simple models can be used. Here, not the total energy of the unfolded state, but the difference between two denatured proteins differing in only one amino acid (for single–point mutants) contribute. For the calculation of folding free energy changes upon mutation, i.e. the difference between the folding free energies of the mutant $\Delta G^{\text{fold}}_{\text{Mutant}}$ and the wildtype $\Delta G^{\text{fold}}_{\text{Wild Type}}$,

$$\Delta\Delta G^{\text{fold}}_{\text{mutation}} = \Delta G^{\text{fold}}_{\text{Mutant}} - \Delta G^{\text{fold}}_{\text{Wild Type}}. \tag{3.2.2}$$

four states (folded/denatured wildtype/mutant) have to be considered as shown in the thermodynamic cycle

$$\begin{array}{ccc} G^{\text{denatured}}_{\text{Wildtype}} & \xrightarrow{\Delta G^{\text{fold}}_{\text{Wildtype}}} & G^{\text{native}}_{\text{Wildtype}} \\ \Big\downarrow{\scriptstyle \Delta G^{\text{mutate}}_{\text{denatured}}} & & \Big\downarrow{\scriptstyle \Delta G^{\text{mutate}}_{\text{native}}} \\ G^{\text{denatured}}_{\text{Mutant}} & \xrightarrow[\Delta G^{\text{fold}}_{\text{Mutant}}]{} & G^{\text{native}}_{\text{Mutant}} \end{array} \tag{3.2.3}$$

Using that a closed path in the thermodynamic cycle does not lead to any change in free energy

$$0 = \Delta G^{\text{fold}}_{\text{Wildtype}} + \Delta G^{\text{mutate}}_{\text{native}} - \Delta G^{\text{fold}}_{\text{Mutant}} - \Delta G^{\text{mutate}}_{\text{denatured}}, \tag{3.2.4}$$

the change in stability upon mutation can either be calculated from the difference between the folding free energies of the mutant and the wildtype (3.2.2) or as the difference in energy for mutating the native ($\Delta G^{\text{mutate}}_{\text{native}}$) and denatured ($\Delta G^{\text{mutate}}_{\text{denatured}}$) state

$$\begin{aligned} \Delta\Delta G &= \Delta G^{\text{fold}}_{\text{Mutant}} - \Delta G^{\text{fold}}_{\text{Wildtype}} \\ &= \Delta G^{\text{mutate}}_{\text{native}} - \Delta G^{\text{mutate}}_{\text{denatured}}. \end{aligned} \tag{3.2.5}$$

The occurring energy difference between the unfolded states

$$\Delta G^{\text{mutate}}_{\text{denatured}} = G^{\text{denatured}}_{\text{Mutant}} - G^{\text{denatured}}_{\text{Wildtype}} \tag{3.2.6}$$

is easier to approximate than the total free energy of a single unfolded state, e.g. by assuming a similar shape of the denatured conformations of wild type and mutant. One of the easiest models to approximate the unfolded state is an extended tripeptide `GXG` — the amino acid of interest `X` surrounded by glycine residues `G` on each side. A more complicated model is, for example, the random sequence structure model developed by Pokala and Handel [61], who showed that longer chains with thirteen amino acids in random sequence projected on protein fragments yield smaller errors than shorter ones when comparing to experimental data. First tests with longer chains when prototyping the Concoord/PBSA method did not yield a significant improvement in correlation between experimental data and predicted free energies (extended chains and protein fragments of 5 or 7 residues have been considered, data not shown).

3.3 Methods: Concoord/PBSA

The Concoord/PBSA method is a composition of different procedures described in the following. A rough overview is sketched as a workflow in Figure 3.2.
In order to obtain the desired change in folding free energy upon mutation all four states shown in the thermodynamic cycle (Equation (3.2.3)) were considered and their corresponding free energies $G^{\text{native/denatured}}_{\text{Wildtype/Mutant}}$ were obtained in the same way as outlined below.

Structure Preparation and Mutant Modeling
Prior to structure generation, the input structure is corrected and minimized:
In a first step missing heavy side chain atoms in the crystal structures were added with the *corall* routine of the program WHAT IF [242] and atoms not belonging to the protein, e.g. water, ions and ligands are removed.
After a short minimization, the wild type crystal structure is mutated with the program WHAT IF [243] and energy minimized, again. Both minimizations follow the protocol described below.
Contingent pseudo wild type crystal structures were mutated to match the original wild type sequence.

Energy Minimization Protocol
All minimizations were carried out with the GROMACS simulation suite (version 3.3.1) [81] using the GROMOS 53a6 [71] force field. The minimizations were performed in vacuum without explicit water molecules. Solvent

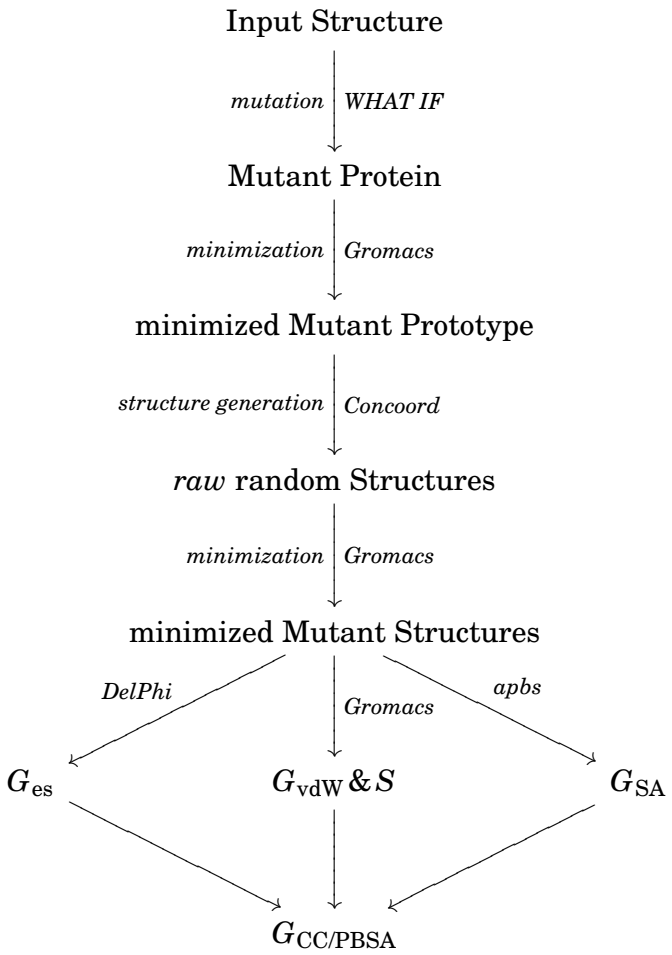

Figure 3.2: Concoord/PBSA workflow: The processed input structure is mutated and optimized, the resulting prototype is randomly replicated, the resulting *raw* structures are optimized and their energy is evaluated and averaged over the whole set. The four contributions to the Concoord/PBSA energy functions are the electrostatic energy G_{es}, the intramolecular Lennard–Jones interactions G_{vdW}, the surface area term approximating solute–solvent van der Waals interactions G_{SA} and the entropy estimate S.
Next to the working steps the used programs are given.

effects were implicitly taken into account by a distance–dependent dielectric permittivity of $\varepsilon(r) = 40\,\mathrm{nm}^{-1}\,r$ [70]. The l–BFGS algorithm [90, 91] (see Page 26 ff.) was applied with an initial step size of 0.01 nm and a gradient tolerance of $100\,\mathrm{kJ\,mol}^{-1}\,\mathrm{nm}^{-1}$. Both Lennard–Jones and electrostatic interactions were computed without a cut–off.

Protonation
Missing hydrogen atoms were added to the structures using the program *pdb2gmx* that is part of the Gromacs package. Here, protonation states for all titratable groups were chosen according to their model pK_a value at pH 7 [244]. Thus, the N– and C–termini as well as the side chains of Asp, Glu, Arg and Lys were considered charged and His uncharged (single protonation). The termini of the tripeptides were uncharged to avoid artificial effects due to charged cappings.
Protonation changes in amino acids that are not target of the mutagenesis were omitted. In our model we assumed the same protonation states for all titratable amino acids in the denatured conformations as chosen for the folded conformation. Also, possible mutant–induced protonation charges concerning other amino acids were not considered here.

Structure Sampling
Starting from the minimized mutated structures, random conformations were generated with the program CONCOORD [64]. In that way an ensemble of independent random structures (e.g. 300 in numbers) were sampled for the wild type and mutant both in their folded and denatured states (e.g. a total of $4 \times 300 = 1200$ conformations). All structures were subsequently minimized.

Van der Waals Contributions to Free Energy
Inter– and intramolecular van der Waals interactions were approximated by a Lennard–Jones potential (1.3.6), parameterized in the chosen force field. Gromacs was used for the evaluation of the force field energies.
Non–polar solute–solvent interactions were estimated to be proportional to the solvent accessible surface area using Equation (1.5.47). The *APOLAR* routine of apbs [131] was applied for the surface calculations. 0.5 Å was used as probe size for surface area calculations. The radii of various atom types were obtained via the Lennard–Jones potential: the minimum of the Lennard–Jones potential for an atom pair of the same atom type i yields the atomic radius for atom type i:

$$R_{ii} = 2^{\frac{1}{6}} \sigma_{ii}. \tag{3.3.1}$$

In cases of small radii (≤ 0.1 nm), we used a minimum atom radius of 0.1 nm to avoid numerical artifacts.

Electrostatics Contribution
The electrostatic interaction of the solute with the solvent was estimated by the solution of the Poisson–Boltzmann equation (see Section 1.5.2). For

the numerical solution of the linearized PBE the program DelPhi [123] was used. DelPhi utilizes a finite–difference algorithm (see Page 39 ff.).
A dielectric constant of 78 was chosen for water, and the relative dielectric permittivity of the protein interior was set to 2.
GROMOS 53a6 atomic partial charges and the previously calculated atomic radii (also based on the GROMOS 53a6 Lennard–Jones parameterization) were used. (In the OPLS–AA [68] test case used later on, OPLS charges were used. Radii based on the OPLS–AA FF were also obtained according to Equation (3.3.1).
The cubic simulation box was chosen in a way that the proteins longest linear dimension filled 60% of the lattice's linear dimension with two grid points per Å. Dipolar boundary conditions were applied. Ionic concentration was set to 0 mM. Thereby, only the Poisson equation (1.5.3) was solved.
Since biomolecular force fields neglect electrostatic interactions between atom pairs that are connected by less than three covalent bonds, DelPhi was also used for the calculation of the solute–solute Coulombic energies to retain consistency.

Entropy Approximation
An upper boundary for the entropy was estimated using Schlitter's method [154]. The mass weighted covariance matrix was diagonalized applying the *g_covar* program of the Gromacs package.

Mean Energy of the Structural Ensemble
All energy contributions for each state were evaluated for the complete set of generated structures and the arithmetic average was used in the following. Thus, a single energy contribution A was obtined via

$$\Delta\Delta A = \left(\overline{A}^{\text{folded}}_{\text{mutant}} - \overline{A}^{\text{folded}}_{\text{wild type}}\right) - \left(\overline{A}^{\text{denatured}}_{\text{mutant}} - \overline{A}^{\text{denatured}}_{\text{wild type}}\right). \tag{3.3.2}$$

3.4 Results

3.4.1 Concoord/PBSA Energy Function

After generating an ensemble of 600 random structures for each thermodynamic end state of our test set with 582 mutants the mutational change in physical free energy

$$\Delta\Delta G_{\text{CC/PBSA}} = \alpha\Delta\Delta G_{\text{es}} + \beta\Delta\Delta G_{\text{LJ}} + \gamma\Delta\Delta A_{\text{SA}} - \tau T\Delta\Delta S \tag{3.4.1}$$

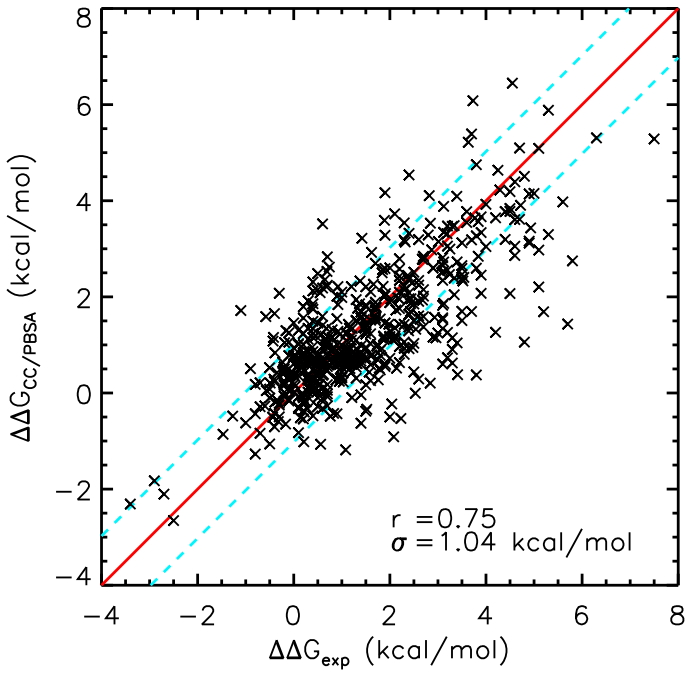

Figure 3.3: Concoord/PBSA results I: The calculated changes in free energy upon mutation are plotted against experimentally determined values. Fitting the weights of the single contributions leads to a correlation of $r = 0.75$ and a standard deviation of $\sigma = 1.04 \frac{\text{kcal}}{\text{mol}}$. The solid line represents y=x, the dashed lines are drawn at $y = x \pm \sigma$.

was evaluated on all generated structures.

Four weighting factors (α, β, γ, τ) for the considered electrostatic ($\Delta\Delta G_{\text{es}}$), Lennard-Jones ($\Delta\Delta G_{LJ}$), surface area ($\Delta\Delta A_{SA}$), and entropic ($T\Delta\Delta S$) contributions were introduced and fitted to maximize the agreement to experimental data. Five–fold cross validation [245] was applied yielding $\alpha = 0.224$, $\beta = 0.217$, $\gamma = 16.6\,\text{cal}\,\text{mol}^{-1}\text{Å}^{-2}$ and $\tau = 0.0287$ (at $T = 298\,\text{K}$) (see Fig. 3.3). The individual energy contributions for mutational folding free energy differences are listed in Table A.1 of Appendix A.

In the five–fold cross validation procedure, the mutants were randomly divided into five sets. One set was left out for later validation, while the remaining four were used for a regression fit to obtain the scaling factors. The energy function that resulted out of it was applied to the validation sample. This procedure was repeated five times using each of the five sets as validation sample while the remaining four served as training data. The averages of the results were then used as final parameter set.

As a measure of the predictive power of Concoord/PBSA, the Pearson corre-

lation coefficient defined as [246, 247]

$$r = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2 \sum_{i=1}^{N}(y_i - \overline{y})^2}} \tag{3.4.2}$$

was calculated (x_i: predicted data; y_i: experimental data). The average of the five validation sets is $r = 0.748 \pm 0.018$ and the standard deviation of the error of calculation (SDEC)[1] [246, 248]

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2} \tag{3.4.3}$$

achieved in the same fashion is $\sigma = (1.04 \pm 0.03)\,\mathrm{kcal\,mol^{-1}}$. Neglecting mutants with a deviation from experiment larger than two times the standard deviation — $\left|\Delta\Delta G^i_{\mathrm{exp}} - \Delta\Delta G^i_{\mathrm{CC/PBSA}}\right| > 2\sigma$ — the resulting correlation is $r = 0.82$ and the standard deviation $\sigma = 0.82\,\mathrm{kcal\,mol^{-1}}$. These outliers are shown in Table 3.3 and are discussed in Section 3.4.4.

The attained accuracy of the Concoord/PBSA results indicates the method's potential for the prediction of relative stability free energies of proteins.

Figures 3.4, 3.5, and 3.6 show subsets of the single proteins used as training and validation sets in this study. Other selected sets include mutations that conserve the charge of the mutation site (*conservative mutations*) or change the charge of the mutation site (*non–conservative mutations*), solvent exposed mutations and mutations of buried sites. Also, the data set excluding the outliers is plotted. The last two figures (Figure 3.6) show datasets with and without alanine as mutational target.

While the mutational stability predictions show a very good agreement for the majority of proteins (1AYI, 1STN, 1YPC, 2LZM and 3CHY), the largest deviations were obtained for the B1 immunoglobulin–binding domain from streptococcal protein G (PDB code 1PGA).

Predictions of folding free energies of charge conserving mutations are comparable to the whole data set with respect to correlation and SDEC. The non–conservative mutants yield calculated free energy differences with a worse but still acceptable correlation. As our model neglects dissimilarities

[1] In the following the term *standard deviation* alone is often used to describe the SDEC. Depending on the type of study, the quantity may also be termed *standard deviation of the error of prediction* (SDEP)

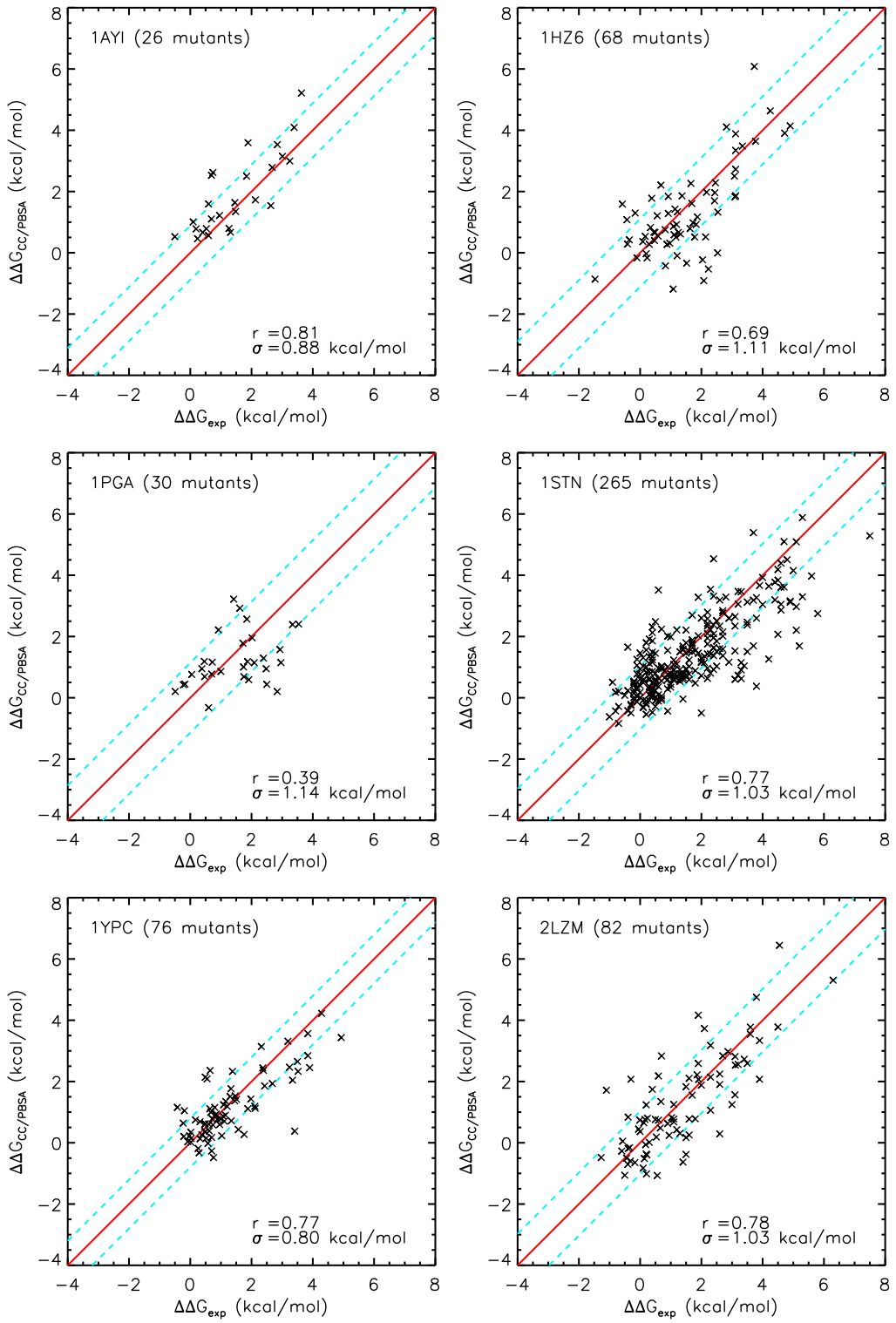

Figure 3.4: Concoord/PBSA results II: Subsets of single proteins.

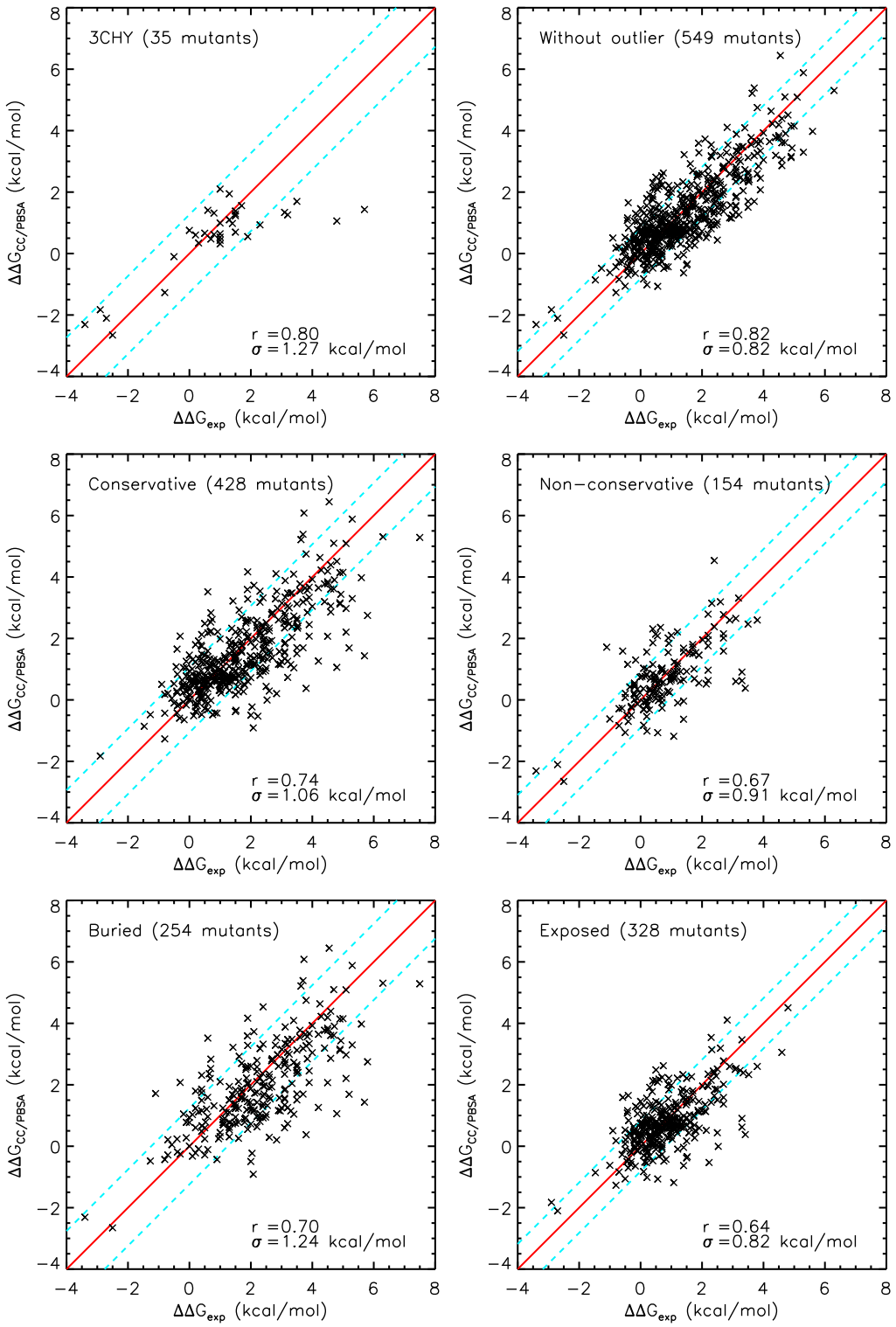

Figure 3.5: Concoord/PBSA results III: Subsets of a single protein, conservative, non-conservative, buried and exposed mutations and the dataset without outliers are shown.

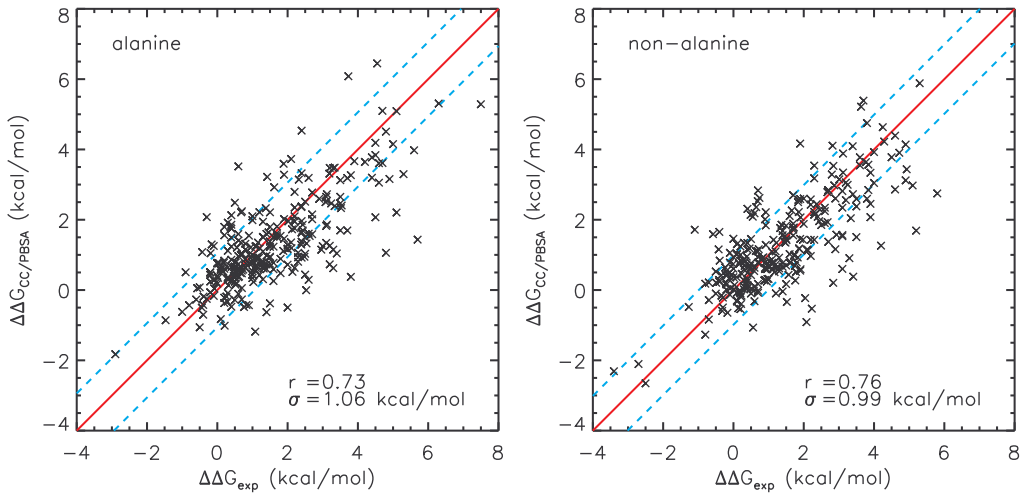

Figure 3.6: Concoord/PBSA results IV: Subsets with alanine or non–alanine mutations.

in the denatured state, the reason for deviations from experiment are probably due to long range interactions in the unfolded state between charged groups.

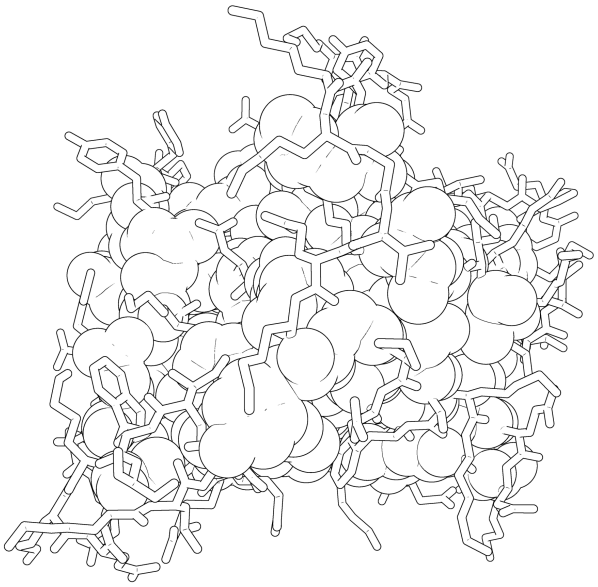

Figure 3.7: Buried and exposed portions of a protein (1STN). The interior is shown by the outlines of spheres, the solvent exposed part is depicted by the outlines of a stick model.

Predictions of buried mutations outperformed those of exposed mutants slightly with respect to the correlation, while free energy predictions that involved exposed site mutants revealed a smaller SDEC than those of the buried ones. Buried and exposed mutants were distinguished by the ratio between the contribution to the solvent accessible surface area (calculated with a probe size of 1.4 Å) of the whole protein and when placed in a tripeptide `GXG`. If the ratio exceeded 0.2 the amino acid was considered to be exposed. This value was determined empirically by visual inspection. The resulting regions for the 1STN wild type are depicted in Figure 3.7.

As can be seen in Figure 3.6, our method is capable of predicting folding free energy changes of alanine and non–alanine mutants at a similar accuracy. None of the subsets that are presented in Figures 3.4, 3.5, and 3.6 shows anomalies that could hint to systematic errors. For more than two third of the outliers probable explanations are discussed in Section 3.4.4.

The unscaled energetic contributions to the predicted mutational changes in protein stability are given as a function of the experimental stabilities in Figure 3.8. The correlations to experiment for the different energetic contributions (electrostatic top left, Lennard–Jones top right, molecular surface term, bottom left and entropy bottom right) vary between 0.1 and 0.62, significantly smaller than the overall correlation of the weighted sum (3.4.1). The necessity of the scaling factors and their interaction is discussed in detail on Page 113 and following.

In addition to the averaged change in mutational free energies, the distribution of energies in the generated structural ensemble were analyzed. Examples are discussed in the following: a charged–apolar mutation, a polar–nonpolar mutation, and a charged–polar mutation.

As an example, Figure 3.9 shows the free energy contributions for the mutant `K116A` of Staphylococcal Nuclease (1STN). Here, the entropy is left out, since it was calculated via the covariance matrix of the complete set of structures only.

The electrostatic contributions — reaction field energy and Coulomb energy — almost compensate each other: In this example, the Coulomb energy is more favorable by 22 kcal/mol for the folded mutant as compared to the native wild type, while the reaction field energy is more favorable for the wild type ($\approx$ 21 kcal/mol). A similar effect is seen for the tripeptides (right column in Figure 3.9).

For mutating polar to nonpolar side chains, or vice versa, a large difference in the Coulomb energy of the native state is counterbalanced by a similar difference in the tripeptides. For example, the mutation `N118A` of the protein 1STN (see Figure 3.10) shows a difference in Coulomb energy for the native state between the mutant and the wildtype of $\Delta G_{\text{coul}}^{native} = 26.4$ kcal/mol

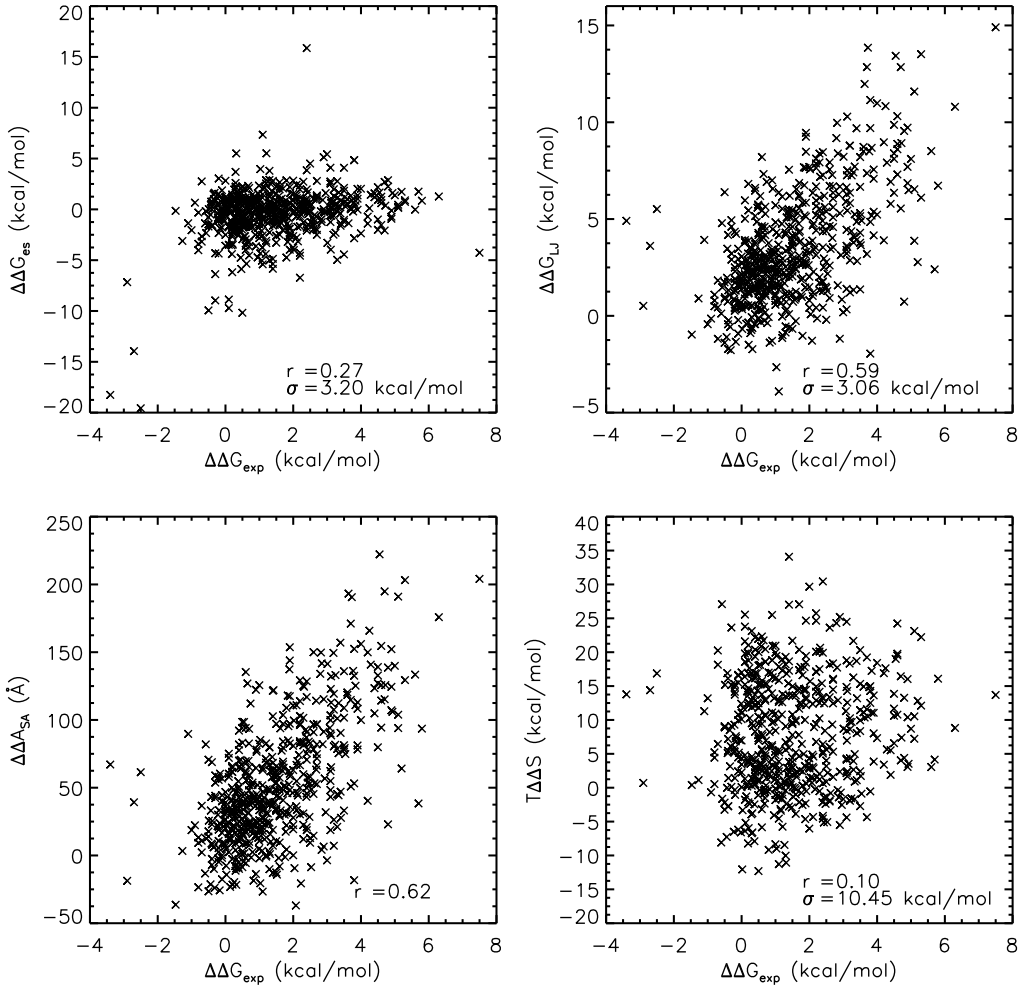

Figure 3.8: Single Concoord/PBSA energy contributions compared to experiment.

while the difference in the reaction field contribution was only $\Delta G_{\mathrm{RF}}^{native} = 0.857\,\mathrm{kcal/mol}$. In this case, however, the Coulomb free energy difference of the native states is compensated by a similar contribution of the tripeptides representing the unfolded mutant and wildtype protein($\Delta G_{\mathrm{coul}}^{denatured} = 24.3\,\mathrm{kcal/mol}$ and $\Delta G_{\mathrm{RF}}^{denatured} = 1.81\,\mathrm{kcal/mol}$).

For mutations involving both charged and polar groups, all four contributions are crucial. For example, the energies for N118D of 1STN (see Figure 3.11) are $\Delta G_{\mathrm{coul}}^{native} = 3.86\,\mathrm{kcal/mol}$, $\Delta G_{\mathrm{RF}}^{native} = 12.67\,\mathrm{kcal/mol}$, $\Delta G_{\mathrm{coul}}^{denatured} = 20.4\,\mathrm{kcal/mol}$ and $\Delta G_{\mathrm{RF}}^{denatured} = -7.44\,\mathrm{kcal/mol}$.

Apart from the individual energetic contributions, Figures 3.9 – 3.11 addi-

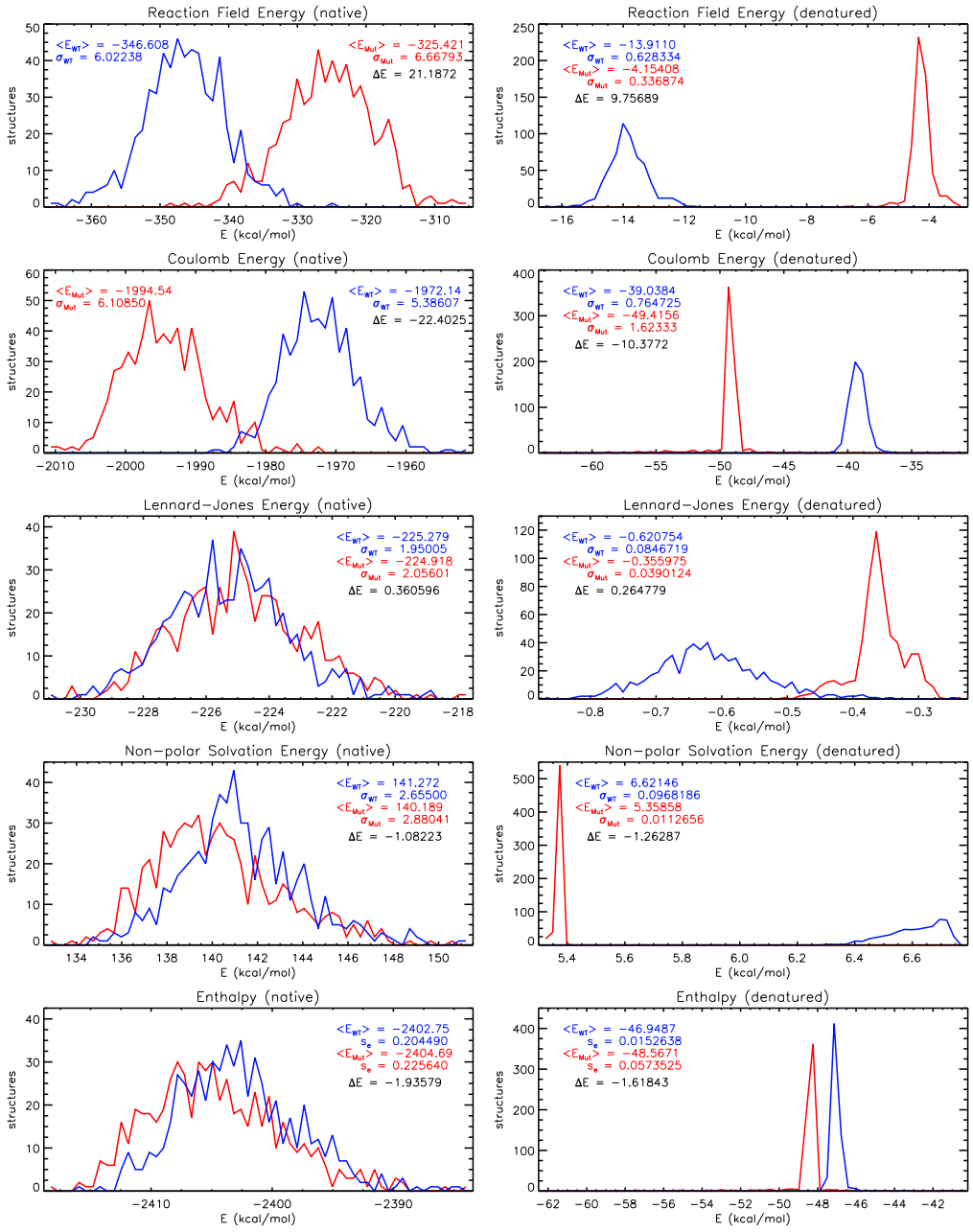

Figure 3.9: Concoord/PBSA energy distributions are shown for the structural ensemble of the native (left) and denatured (right) shapes of the wild type (blue) protein (1STN) and the non–conservative K116A mutant (red). All energies are given in kcal/mol.

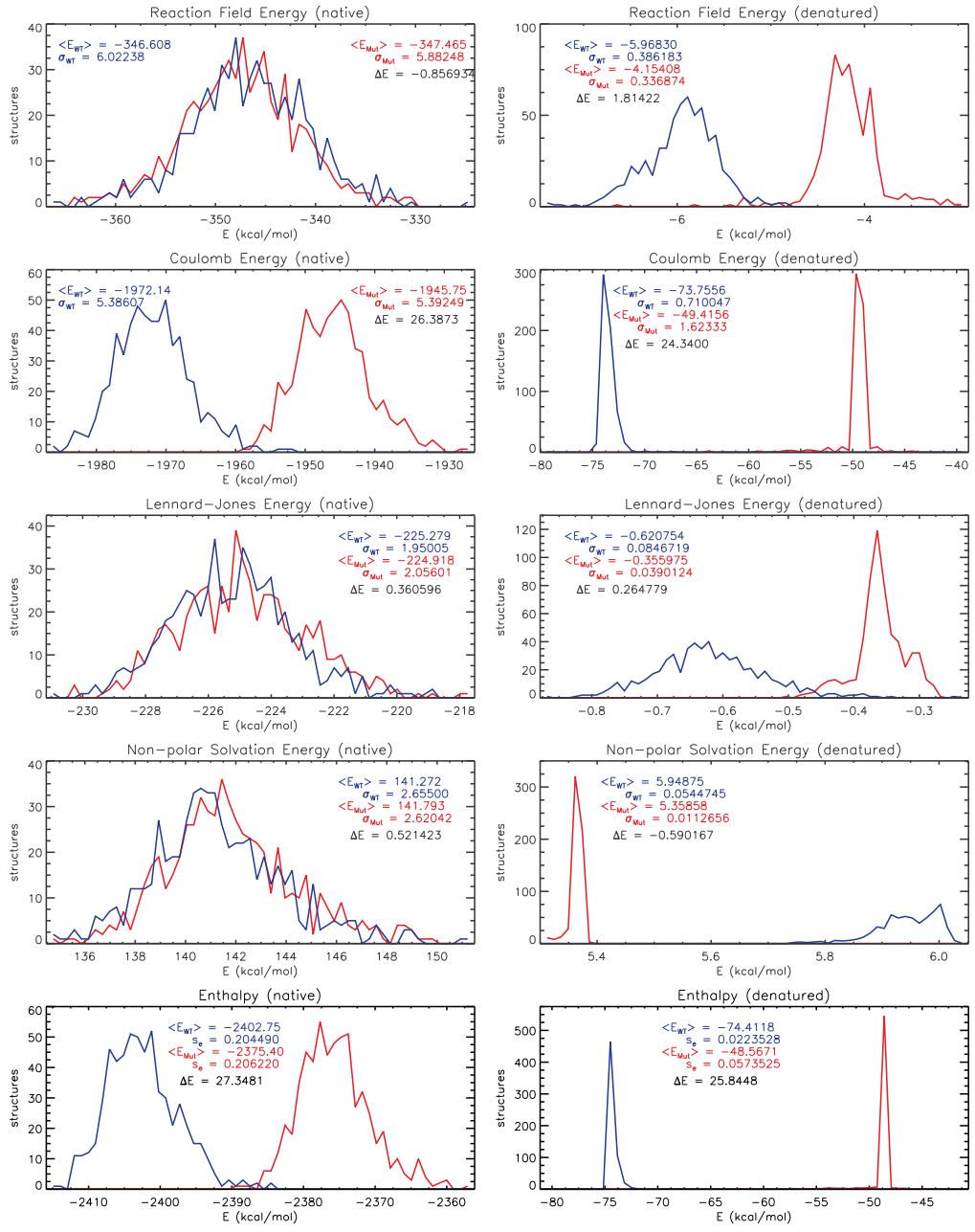

Figure 3.10: Concoord/PBSA energy distributions are shown for the structural ensemble of the native (left) and denatured (right) shapes of the wild type (blue) protein (1STN) and the non–conservative N118A mutant (red). All energies are given in kcal/mol.

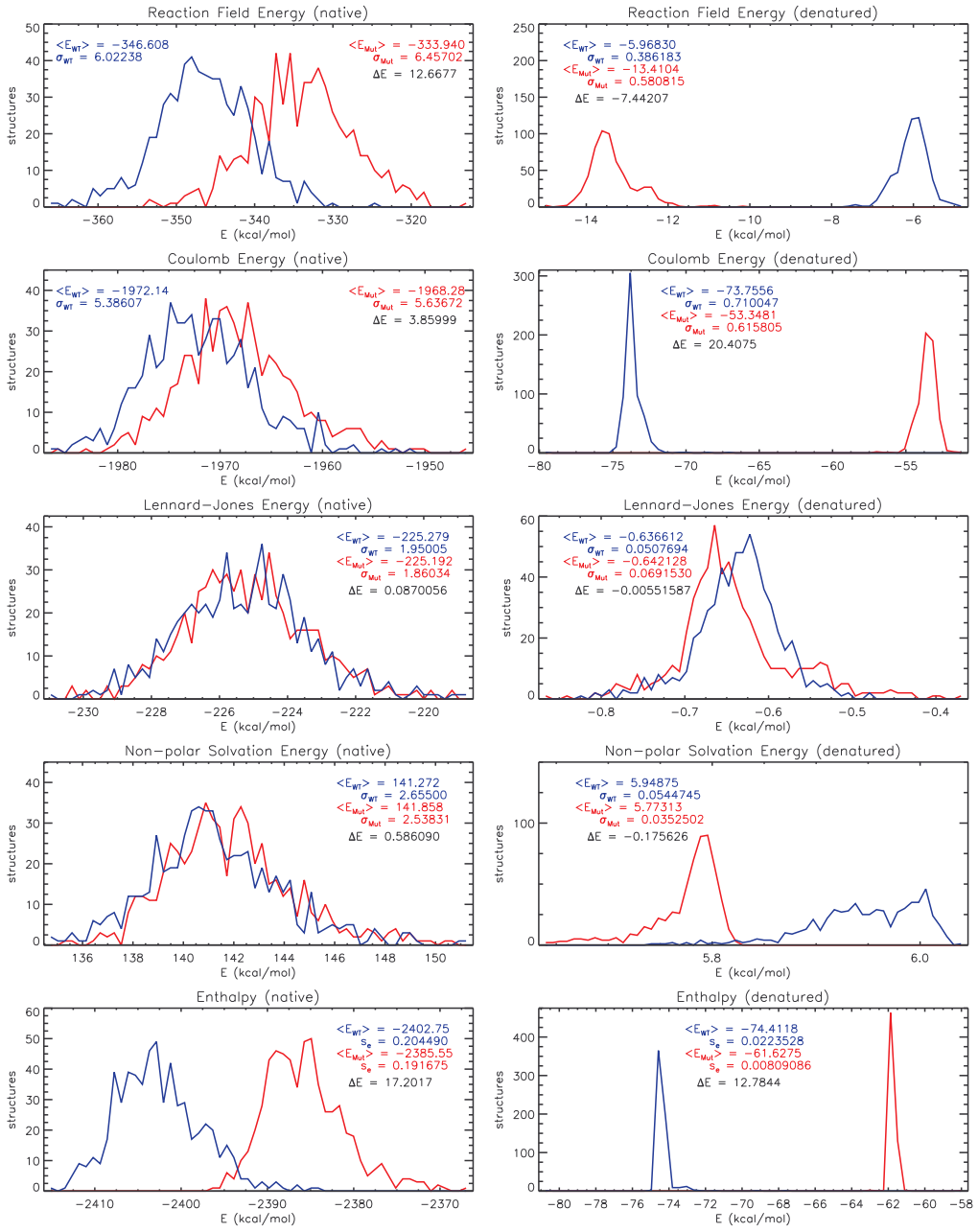

Figure 3.11: Concoord/PBSA energy distributions are shown for the structural ensemble of the native (left) and denatured (right) shapes of the wild type (blue) protein (1STN) and the non–conservative N118D mutant (red). All energies are given in kcal/mol.

tionally show the distribution of the enthalpy H^2, i.e. the sum of the free energy contributions (bottom row). The standard error for the enthalpy is given as

$$s_e(H) = \sqrt{\frac{\sum_{i=1}^{N}\left(H_i - \overline{H}\right)}{N(N-1)}}. \tag{3.4.4}$$

The resulting error for the mutational change in enthalpy was analyzed according to

$$s_e(\Delta\Delta H) = \sqrt{s_e^2(H^{\text{folded}}_{\text{mutant}}) + s_e^2(H^{\text{folded}}_{\text{wild type}}) + s_e^2(H^{\text{denatured}}_{\text{mutant}}) + s_e^2(H^{\text{denatured}}_{\text{wild type}})} \tag{3.4.5}$$

and is listed for the calculated mutations in Table A.1. The calculated error shows a dependence on the size of a protein ranging from 0.14 kcal/mol to 0.4 kcal/mol. The error in the entropic contribution to the free energy was neglected.

3.4.2 Importance of Considering Structural Flexibility

The importance for considering the structural flexibility of proteins was addressed by applying a similar free energy function as used for Concoord/PBSA to the minimized (mutated) crystal structure only. Thus, leaving out the energetic averaging over a Concoord – generated structural ensemble. The single–point energies of the wildtype crystal structure and mutant structure substitute the mean values, and Schlitter's entropy estimate was replaced by an NMA analysis (see Chapter 1.6.2).

The scaling factors were determined again by five–fold cross validation yielding a correlation of $r = 0.57$ and an SDEC of $\sigma = 1.37$ kcal/mol (parameters: $\alpha = 0.0965$, $\beta = 0.0187$, $\gamma = 23.55\,\text{cal}\,\text{mol}^{-1}\,\text{Å}^{-2}$ and $\tau = 8.2{\cdot}10^{-6}$). The comparison to experiment is depicted in Figure 3.12. A similar result was obtained neglecting the entropic contributions.

The poor performance using only crystal structures underlines the need for explicit consideration of structural flexibility in order to estimate mutational free energy changes. However, the computational effort scales linearly with the number of structures.

[2]The term *enthalpy* has to be handled with care here, as entropic contributions may be included implicitly.

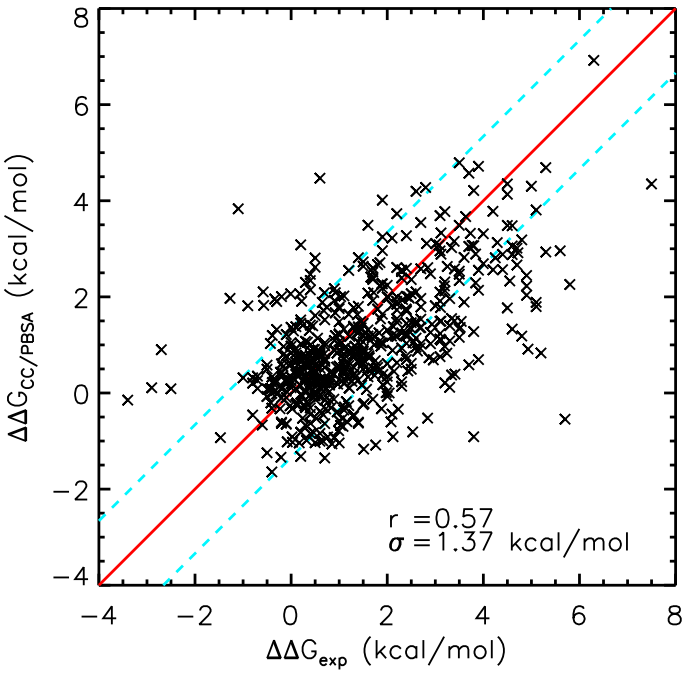

Figure 3.12: Concoord/PBSA results using only crystal structures as input instead of ensembles generated using Concoord.

3.4.3 Convergence of Concoord/PBSA

The convergence of the proposed method with respect to its correlation and standard deviation (SDEC) to experiment was analyzed as shown in Figure 3.13. 600 data points were sampled in total for each mutant, and both correlation and SDEC to experiment were analyzed as a function of the number of input structures. Both values are given as averages over respective subsets of all generated structures, e.g. the correlation and the standard deviation to experiment for 50 structures are mean values of twelve independent subsets.

The entropic contribution was left out for the convergence analysis and a five–fold cross validation fit for the remaining three parameters was used ($r = 0.731$, $\sigma = 1.05$ kcal/mol, $\alpha = 0.241$, $\beta = 0.179$ and $\gamma = 15.57\,\mathrm{cal\,mol^{-1}\AA^{-2}}$). They have been evaluated using 600 structures and were kept fixed for the convergence analysis.

Both correlation and standard deviation to experiment show a monotonous increase (decrease) with the number of generated structures. Consideration of 50 concoord structures was sufficient to obtain better agreement to experiment than with the crystal structure ($r = 0.609$ and SDEC $\sigma = 1.35$ kcal/mol)

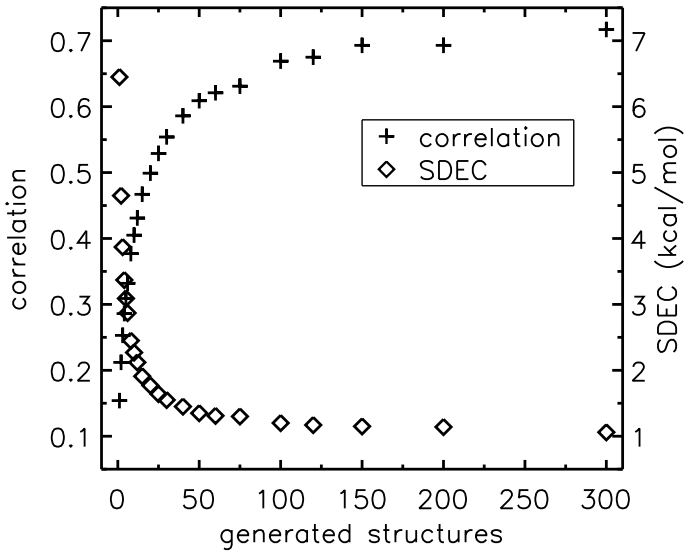

Figure 3.13: Convergence of the correlation and the standard deviation of the error of calculation with respect to the number of generated structures.

alone.
The complete ensemble of 600 Concoord structures yielded improved results ($r = 0.731$, $\sigma = 1.05$ kcal/mol) with respect to 300 conformations ($r = 0.717$, $\sigma = 1.06$ kcal/mol). This is an increase of 2.0% in correlation compared to a doubled computational cost. Using the full energy function including entropic contributions for 300 and 600 structures the difference is even less than 1%: $r = 0.741$, $\sigma = 1.04$ kcal/mol for 300 and $r = 0.748$, $\sigma = 1.04$ kcal/mol for 600 conformations. Thus, we suggest the averaging over 300 structures as a compromise between accuracy and computational effort.

3.4.4 Outliers

33 mutations shown in Table 3.3 led to a Concoord/PBSA energy that deviated more than two times the standard deviation from the experimental value.
In nine cases this discrepancy could be attributed to a different behavior of the unfolded state between wild type and mutant. As their m-value [249] (see also Section 2.2.2) normalized to the wild type is smaller than 0.8 or larger than 1.2, a large change in solvent accessible surface area in the unfolded state is assumed to occur [250] that is only poorly described by the tripeptide model. For the protein 3CHY, for example, López-Hernández and Serrano [225] found only two mutations with an m-value below 0.8 with re-

Table 3.3: Analysis of outliers: Data points deviating by more than 2σ from their experimental value are shown with their experimental and calculated free energy differences (in kcal/mol). Possible explanations are mentioned. * denotes that the experimental and calculated free energies are taken relative to a pseudo wild type.

PDB	Mutation	$\Delta\Delta G_{\text{exp}}$	$\Delta\Delta G_{\text{calc}}$	possible explanation
1HZ6	E32I	1.08	-1.1835	
1HZ6	K41A	-0.58	1.5894	Mutants are in the proximity of the mutation in the crystal structure Y47W that was reversed for calculations.
1HZ6	G45A	2.23	-0.5297	
1HZ6	F62V	3.73	6.0822	
1HZ6	G15V	2.53	-0.0044	General problem for mutating glycin.
1HZ6	G24A	2.08	-0.9119	
1HZ6	G55A	2.04	-0.2295	
1PGA	G41A	2.84	0.2048	
1STN	D77A	3.10	0.6014	$m = 0.75$
1STN	D95A	3.30	0.9063	
1STN	I72A	5.10	2.9746	$m = 1.29$
1STN	K133G	3.30	0.6233	
1STN	L108A	5.80	2.7486	$m = 0.77$
1STN	L38G	0.60	3.5184	
1STN	N100A	5.20	1.6906	$m = 0.80$
1STN	N100G	5.10	2.2059	$m = 0.71$
1STN	N118D	2.40	4.5357	
1STN	V111A	4.20	1.2623	$m = 0.64$
1STN	V23T	3.20	0.7316	$m = 1.23$
1STN	V74T	3.80	0.3745	
1STN	V99T	3.30	1.1498	
1STN	Y54L	3.40	1.0415	
1STN	Y93F	2.00	-0.4976	Mutants are either not correctly modelled or they influence the unfolded state.
1STN	Y93G	7.50	5.2864	
1STN	Y93L	4.50	2.0697	
1YPC	D52A	3.41	0.3808	
2LZM	E11A	-1.10	1.7152	
2LZM	I3A	0.70	2.8336	
2LZM*	F67A	1.90	4.1671	
2LZM*	T152S	2.60	0.2912	Mutants are in the proximity of the pseudo wild type mutation C97A.
2LZM*	F153L	-0.30	2.0754	
3CHY*	V10T	5.70	1.4345	$m = 0.61$
3CHY*	V54T	4.80	1.0595	$m = 0.77$

spect to the pseudo wildtype — these two mutants were both identified as outliers. Analysis of another five mutants were not expected to yield accurate results, as they were found near an amino acid that was mutated either for crystallization or for stability measurements.

Substitutions of glycine were in general not well described. Only for two mutants out of eight good results were obtained, while four mutants deviated more than two times the standard deviation from their experimental reference values.
The change in stability for all three mutations of `Y93` from 1STN was underestimated by more than 2 kcal/mol. This probably hints to a specific functional importance of this tyrosine for the folding of this protein.
No specific reasons could be attributed to the remaining 12 outliers. Possible reasons could be a residual fold of either the wildtype or the mutant, inaccuracies in the experimental measurements, or comparatively large conformational changes upon mutation.

3.4.5 Flexibility

Apart from the prediction of stability changes upon mutation, Concoord/PBSA additionally yields a prediction of protein flexibility. Especially the change in flexibility upon mutation may be crucial also for the function of a protein.
The root mean square fluctuation rmsf for atom i [247]

$$\mathrm{rmsf}_i = \frac{1}{\sqrt{N}} \sqrt{(x_i - \overline{x})^2 + (y_i - \overline{y})^2 + (z_i - \overline{z})^2}. \tag{3.4.6}$$

provides information about the flexibility of site i. This atomic rmsf is averaged over all atoms of one amino acid in order to describe the collective fluctuation of whole residues.
Two examples taken from the mutant set (`F50A` of 1YPC and `D12A` of 3CHY) have been analyzed exemplarily. The cartoons at the top of Figure 3.14 show the color–coded relative changes in flexibility upon mutation projected on the wild type structure. An increase in flexibility upon mutation is colored red, a decrease blue. The wild type amino acid that is subject to the mutation is colored green. Neighboring side chains are shown in stick representation. In the second panel, the relative change in rmsf is plotted with the residues sorted according to their distance from the mutation site. For error estimation, the structures are divided into six subsets for the evaluation of the rmsf. The error bars are depicted in gray.
The bottom panel shows the rmsf per residue for both wild type (blue) and mutant (red) as a function of the sequence residue number.
The stability changes of the shown mutants were both accurately predicted:
$\Delta\Delta G_{\mathrm{exp}}(\mathrm{1YPC,\ F50A}) = 3.84\,\frac{\mathrm{kcal}}{\mathrm{mol}}$, $\Delta\Delta G_{\mathrm{CC/PBSA}}(\mathrm{1YPC,\ F50A}) = 3.57\,\frac{\mathrm{kcal}}{\mathrm{mol}}$
and
$\Delta\Delta G_{\mathrm{exp}}(\mathrm{3CHY,D12A}) = -2.50\,\frac{\mathrm{kcal}}{\mathrm{mol}}$, $\Delta\Delta G_{\mathrm{CC/PBSA}}(\mathrm{3CHY,D12A}) = -2.66\,\frac{\mathrm{kcal}}{\mathrm{mol}}$.

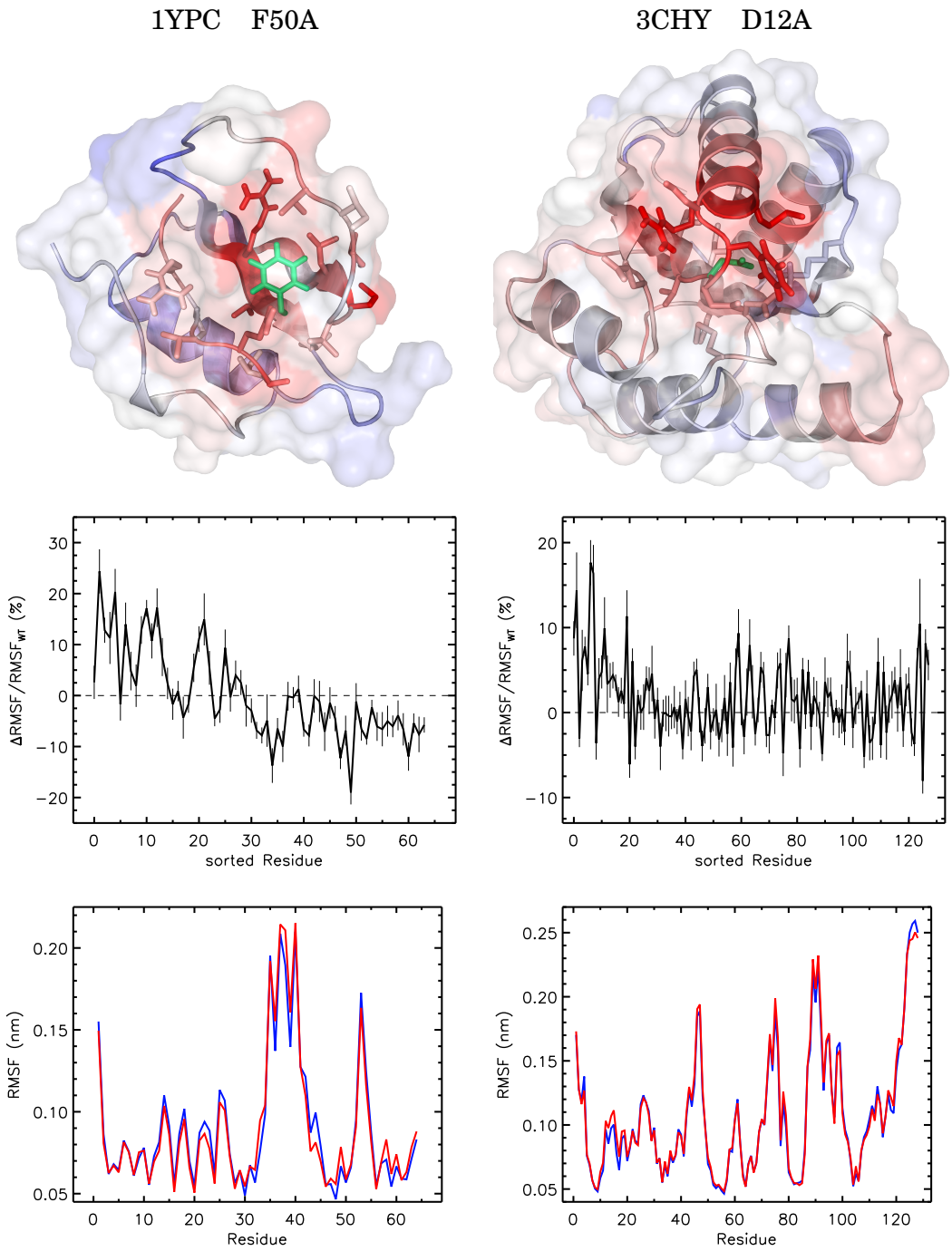

Figure 3.14: CC/PBSA flexibility prediction. For explanations see text.

As the mutations involve a replacement of an amino acid side chain by a smaller side chain the neighborhood can sample a larger conformational space. Thus, the flexibility of amino acid side chains in the vicinity of the mutation site is increased. Depending on the space that is freed by mutation, also solvent molecules can fill the arising cavity. Interestingly, also protein sites distal from the mutation site may exhibit significantly decreased or increased flexibility.

3.5 Concoord/PBSA Web Interface

A Concoord/PBSA web interface was set up to make the method easily available to the public. It is accessible under `http://ccpbsa.bioinformatik.uni-saarland.de` and provides the possibilities to predict changes in stability upon mutation, as well as the binding between proteins presented in the next chapter.
The web interface was used to study the reproducibility of the results as well as the time consumption of the method.

3.5.1 Reproducibility

The results of Concoord/PBSA depend on randomly generated structures. Thus, the results differ when repeating the procedure with a different initial random seed. The averaged standard deviation of predicted results for the respective experimental value as obtained for selected mutations of the chymotrypsin inhibitor–2 (1YPC) is 0.256 kcal/mol on average (see Table 3.4). ITables A.2 to A.9 of the Supplementary show a more detailed analysis for the individual mutants and of the different contributions to the Concoord/PBSA energy function.

Table 3.4: Reproducibility of Concoord/PBSA results: The experimental mutational folding free energy, the Concoord/PBSA free energy difference, and an averaged Concoord/PBSA free energy difference (10 calculations) for selected mutants of chymotrypsin inhibitor–2 (1YPC) are shown. Additionally, the standard deviation for the ten results obtained from the web service for each mutant is displayed, too. All quantities shown are in kcal/mol.

	A16G	D45A	E15Q	F50A	N56D	S12A	T39D	V63T
$\Delta\Delta G_{\text{exp}}$	1.09	0.80	0.47	3.84	1.21	0.89	-0.02	1.15
$\Delta\Delta G_{\text{CC/PBSA}}$	1.24	0.894	0.359	3.57	1.19	0.766	0.228	0.900
$\Delta\Delta\overline{G}^{\text{web}}_{\text{CC/PBSA}}$	1.12	0.712	0.331	2.95	1.26	0.624	-0.121	0.903
σ^{web}	0.261	0.253	0.331	0.222	0.182	0.288	0.274	0.238

3.5.2 CPU Time for Concoord/PBSA

The CPU time consumption of the Concoord/PBSA method was measured using the public Concoord/PBSA web interface (see Table 3.5). For every protein of the mutational test set, eight mutants have been randomly selected. Thus, a total number of nine (eight mutants plus one wild type protein) times 300 structures was generated and evaluated for every protein. The fast evaluation of the tripeptide is neglected, as the bottleneck for its computation is the peripheral hardware like hard disks and the network bandwidth. Besides, the tripeptides had to be evaluated only once and could be used for all calculations afterwards. For the first calculated mutant the time is doubled due to the necessary wild type calculation.

Apart from the crystal structures 1HZ6 and 2LZM the CPU times ranged between 108 and 249 minutes for the remaining five proteins. For the former two protein crystal structures, the Concoord procedure experienced convergence problems, indicating (local) stress, by e.g. van der Waals overlaps of atoms. A single awkward mutation may already raise the required CPU time. The computation time also depends on external parameters like the overall load of the master node in the compute cluster.

From the small number of studied proteins, no general law for the time consumption could be derived. On average, Concoord/PBSA requires 3.5 CPU minutes per residue.

Table 3.5: Concoord/PBSA CPU time consumption: The average computation time for mutational folding free energy differences (300 structures) on a single processor (Intel Xeon, 3GHz) is presented together with the number of amino acids of the wild type structure.

PDB	# AA	CPU time per mutant (min)
1AYI	87	169
1HZ6	72	582
1PGA	56	166
1STN	149	249
1YPC	64	108
2LZM	164	1026
3CHY	128	200

3.6 Comparison to Fold-X

For comparison, folding free energies for the mutation set used for Concoord/PBSA have additionally been derived using Fold-X [59]. The most recent Fold-X version (at the time of writing 3.0 Beta) obtained from `http://foldx.crg.es/` was used to mutate and to predict stabilities (unpublished). Comparisons to experiments and to Concoord/PBSA are shown in Figure 3.15.

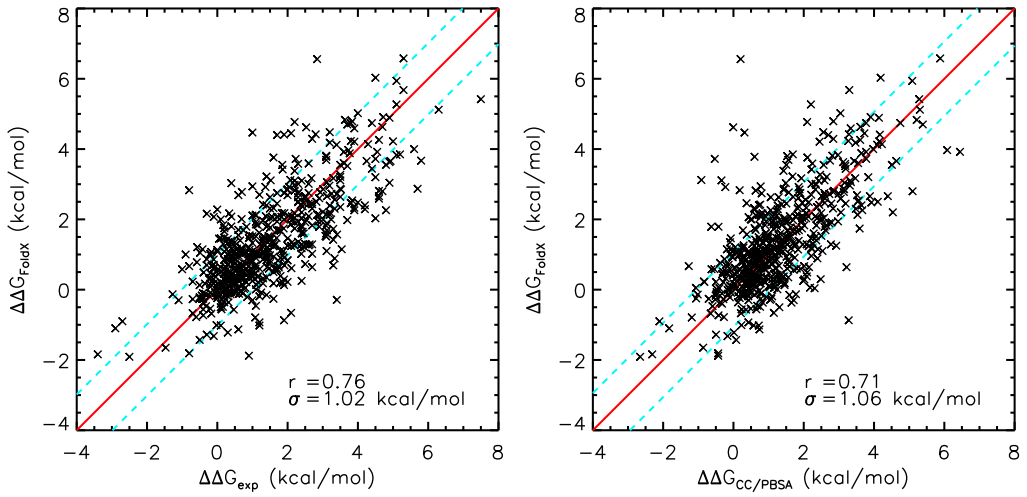

Figure 3.15: Folding free energies predicted with Fold-X are plotted against the corresponding experimental free energies (left) and against the Concoord/PBSA results (right).

The overall correlation and SDEC for Fold-X ($r = 0.76$, $\sigma = 1.02$ kcal/mol) were comparable to the Concord/PBSA energy function ($r = 0.75$, $\sigma = 1.02$ kcal/mol)[3].
Both methods show a comparable performance with respect to the accuracy. Also, some similarities can be seen in the outliers discussed in Section 3.4.4. Table 3.6 shows 14 common outliers for both methods. The two glycine substitutions are overestimated by Fold-X and underestimated by Concoord/PBSA. The remaining outliers are comparable in their deviation from experiment for both methods. The Y93 mutants of 1STN were also reported by Bordner and Abagyan [174] as outliers for their method.
Fold-X requires roughly one minute of computation time per mutation on a 3GHz Intel Xeon machine. In this time, it has to mutate and evaluate

[3]The Concoord/PBSA correlation and SDEC have been obtained without 5-fold cross validation in this section for comparison to Fold-X.

Table 3.6: Common outliers for Concoord/PBSA and Fold-X (data points deviating by more than 2σ from their experimental value). Experimental and calculated energies are in kcal/mol. * denotes that the experimental and calculated free energies are taken relative to a pseudo wild type.

PDB	Mutation	$\Delta\Delta G_{\text{exp}}$	$\Delta\Delta G_{\text{CC/PBSA}}$	$\Delta\Delta G_{\text{Fold-X}}$
1HZ6	G15V	2.53	-0.0044	4.62
1PGA	G41A	2.84	0.2048	6.56
1STN	D95A	3.30	0.9063	0.69
1STN	L108A	5.80	2.7486	3.67
1STN	L38G	0.60	3.5184	2.69
1STN	V74T	3.80	0.3745	1.46
1STN	Y54L	3.40	1.0415	-0.29
1STN	Y93G	7.50	5.2864	5.42
1STN	Y93L	4.50	2.0697	2.03
2LZM	I3A	0.70	2.8336	2.78
2LZM*	F67A	1.90	4.1671	4.41
2LZM*	F153L	-0.30	2.0754	2.16
3CHY*	V10T	5.70	1.4345	2.87
3CHY*	V54T	4.80	1.0595	2.26

the energies of both wild type and mutant. Its time consumption beats Concoord/PBSA by two to three orders of magnitude. For the generation and energy evaluation of three hundred structures, Concoord/PBSA requires 2-5 CPU hours depending on the size of the protein and the quality of the structure (see Section 3.5).

With the prediction of conformational ensembles using Concoord it is possible to analyze protein configurations as a function of the respective energies, or the combined influence of mutations on the folding free energy, structure and flexibility. It is also possible to choose conformations from the generated ensemble applying different criteria (e.g. energy, clustering according to rmsd) for further treatment (e.g. structure prediction). These possibilities are not given using Fold-X. The comparability of the predictive power of Fold-X and of Concoord/PBSA suggests a limited influence of mutational flexibility changes on the folding stability.

While stabilities obtained by Concoord/PBSA are comparable to Fold-X folding free energies, CC/PBSA outperforms Fold-X with respect to the prediction of mutational effects on protein–protein binding affinities (see Chapter 4.2 and Benedix et al. [1]).

3.7 Alternatives and Variations

During the development phase of the Concoord/PBSA method, the protocol and parameters were adjusted numerous times. The most important parameters and methods are discussed below. Results for the variation of the dielectric permittivity or the probe size are exemplarily presented in more detail. Additional variations that proved to have no or only small effects on the outcome are briefly mentioned.
As a fast adaption, the Concoord/GBSA method is introduced substituting the slow numerical solution of the Poisson-Boltzmann equation by the faster Generalized Born method.

3.7.1 Dielectric Permittivity

One major variable for the Concoord/PBSA procedure is the dielectric permittivity ε for the interior of the protein. Due to the high computational cost only a few values have been tested: 1, 2 and 4. The permittivity $\varepsilon = 2$ used in Concoord/PBSA yielded only slightly more accurate results as compared to 1 or 4 as shown in Table 3.7. The scaling factor of the electrostatic contribution to the total free energy (α) was roughly doubled for a doubled dielectric permittivity, as expected. Interestingly, the best prediction accuracy obtained for $\varepsilon = 2$ comes with an almost identical scaling for both the electrostatic and the van der Waals contribution.

Table 3.7: Parameters and accuracy of Concoord/PBSA as a function of the dielectric permittivity of the protein interior. ([a]) σ in kcal/mol, [b]) γ in $\mathrm{cal\,mol^{-1}\AA^{-2}}$.)

ε	r	$\sigma^{a)}$	α	β	$\gamma^{b)}$	τ
1	0.745	1.042	0.109	0.204	17.68	0.0261
2	0.748	1.037	0.224	0.217	16.64	0.0287
4	0.738	1.052	0.391	0.216	16.56	0.0360

3.7.2 Probesize

For the non–polar solute–solvent interaction contribution to the free energy different approaches and approximations are at hand, e.g. a linear relationship to the molecular surface area (Equation (1.5.47)), to the molecular volume (Equation (1.5.50)) or a volume integral ansatz (Equation (1.5.51)) as shown in Section 1.5.5. The three presented approaches have been tested

for the usage in Concoord/PBSA. Both, the integral and volume term, as incorporated in apbs, did not lead to an improvement of the correlation and standard deviation of the Concoord/PBSA method (not shown) but raised the computational cost and have not been considered further.
In addition to testing technical approaches, also the impact of different molecular surfaces (defined by their probe size) on the correlation and the standard deviation of the predicted Concoord/PBSA folding free energies to the experimental free energies have been studied (see Figure 3.16). The surface area term in Concoord/PBSA is an approximation that assumes a linear relationship between the strength of the solute–solvent interactions and the solute–solvent interface. Therefore, scaling of the probe size to obtain an improved correlation to experiment appears justified.
Figure 3.16 shows the dependency of the correlation and the SDEC, as well as of the scaling factors on the probe size. The best agreement for Concoord/PBSA to experiment was obtained for a probe size of 0.5 Å. This probe size maximizes the correlation and minimizes the SDEC. The frequently chosen probe size of 1.4 Å, i.e. the approximate size of a water molecule, yields a significantly decreased correlation of Concoord/PBSA to experiment and led to a negative scaling factor for the surface area term (see lower panel in Figure 3.16).
Interestingly, for a probe size of 1.2 Å the best correlation was achieved for a vanishing surface term ($\gamma \approx 0$). While the scaling factors for electrostatics and entropy were hardly affected by the probe size, the Lennard–Jones coefficient β showed a strong dependency on the probe size, opposite to the surface tension γ. Both reached an extremum at a probe size of 0.5 Å.
One possible explanation for the excellent performance of this comparably small probe size is the appearance of internal cavities. However, explicitly considering cavity–related energetic contributions proportional to the cavity volume did not yield a significant improvement for the prediction of mutational changes in the folding free energy (data not shown). A possible further refinement of Concoord/PBSA may be achieved by a separate treatment of water–exposed protein surfaces and of buried cavity surfaces.
The interesting interaction between the Lennard–Jones contribution and the surface area term is further discussed in section 3.8.

3.7.3 Minimization Method

Next to the chosen l–bfgs optimization, the steepest descent and conjugate gradient methods were tested as alternatives (Figure 3.17).
Tolerances and step–sizes were unchanged for all methods. The conjugate gradient method is supported every ten steps by a steepest descent opti-

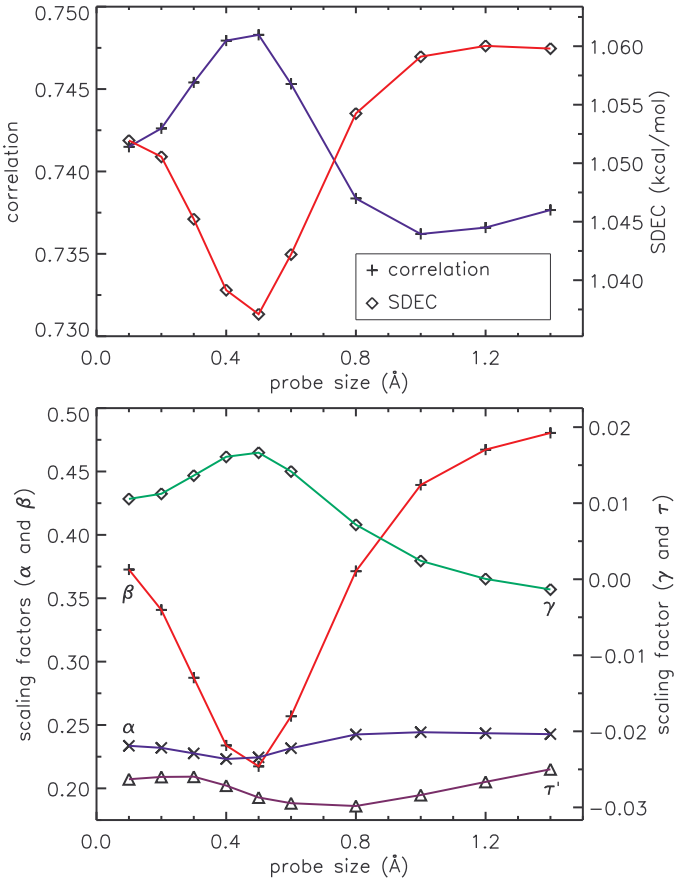

Figure 3.16: Dependency of correlation and standard deviation of Concoord/PBSA to experiment on the chosen probe size for the non–polar solute–solvent interaction (upper panel). The scaling factors in Concoord/PBSA were adjusted to the respective probe size (lower panel). Instead of τ, $\tau' = -\tau$ is shown. Thus, the negative sign of the entropic contribution is correctly expressed (adjusting TS).

mization step for faster convergence. Correlation, SDEC and scaling factors were obtained using five–fold cross validation.

Table 3.8 holds the scaling factors for all three methods. While conjugate gradient yielded similar results as compared to l–bfgs, the steepest descent method is obviously not suitable for the optimization of Concoord structures due to its slow convergence and and its comparably bad performance.

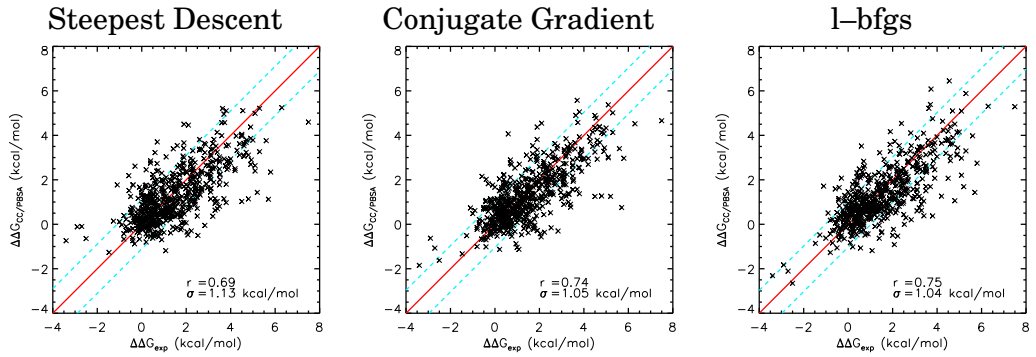

Figure 3.17: Concoord/PBSA results using different minimization methods.

Table 3.8: Scaling factors for different minimization approaches. For every set three hundred structures were considered. [a] σ in kcal/mol, [b] γ in cal mol^{-1}Å^{-2}.

Minimization Method	r	σ[a]	α	β	γ[b]	τ
l-bfgs	0.741	1.045	0.224	0.217	16.64	0.0287
conjugate gradient	0.737	1.038	0.220	0.343	6.34	0.0149
steepest descent	0.695	1.108	0.171	0.335	5.71	0.00725

3.7.4 OPLS-AA force field vs. Gromos G53a6 force field

The OPLS-AA all atom force field [73] was the first choice in the initial developmental stages of the Concoord/PBSA method. Due to the explicit consideration of all hydrogen atoms an increased accuracy was expected. However, in first comparative tests, G53a6 [71] outperformed OPLS both in accuracy and efficiency.

For OPLS-AA, five–fold cross validation yielded $r = 0.670$ and $\sigma = 1.13$ kcal/mol with $\alpha = 0.186$, $\beta = 0.123$, $\gamma = 11.90$ cal mol^{-1} Å^{-2} and $\tau = 3.5 \cdot 10^{-4}$ (see Figure 3.18).

One possible reason for the comparably low accuracy of Concoord/PBSA using the OPLS-AA force field is the strong sensitivity of the minimization observed for OPLS protein structures. The optimization converged to distinct conformations upon small translations of the same molecule. A similar effect was also observed for the GROMOS force field, however with a clearly reduced impact. This problem may be reduced or eliminated by an increased numerical accuracy with the drawback of a significantly increased computational effort.

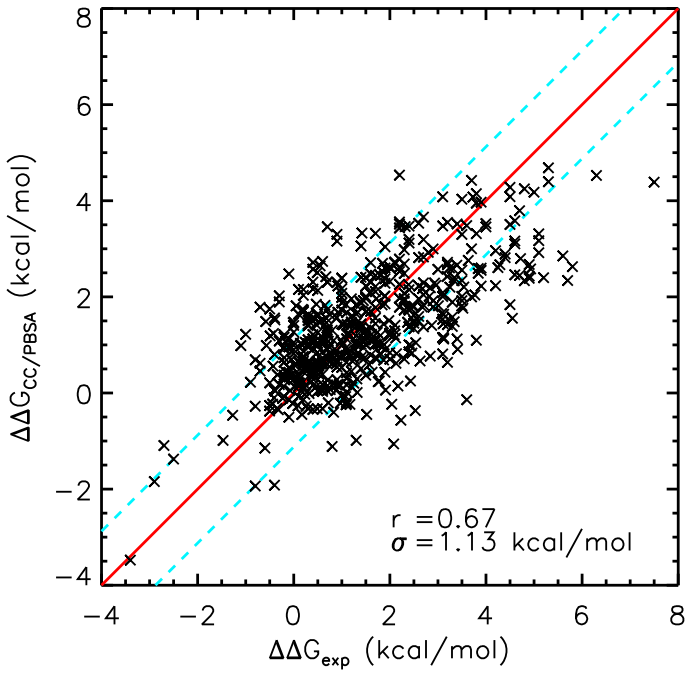

Figure 3.18: Concoord/PBSA results using the OPLS-AA force field.

3.7.5 Inapplicable Contributions

Entropy from Normal Mode Analysis

The entropy estimate as proposed by Schlitter [154] (see Section 1.6.3) is based on the covariance matrix. Here, the time consumption only has a slight influence on the overall computational effort. Therefore, Schlitter's approximation was preferred over the normal mode approach (see Section 1.6.2). The latter requires expensive energy minimizations to the local minimum that would render the inclusion of every generated structure impossible for a large data set. Therefore, only the entropy for the Concoord/PBSA input structure was estimated using NMA analysis. Coupled with the enthalpy of six hundred structures this approach yielded $r = 0.729$, $\sigma = 1.06\,\text{kcal/mol}$ with $\alpha = 0.241$, $\beta = 0.181$, $\gamma = 15.47\,\text{cal}\,\text{mol}^{-1}\,\text{Å}^{-2}$ and $\tau = 1.92 \cdot 10^{-5}$ (via five–fold cross validation, see left panel of Figure 3.19). Neglection of this NMA based entropic energy contribution had a negligible effect on the correlation of Concoord/PBSA to experiment and, thus, renders this approach inappropriate.

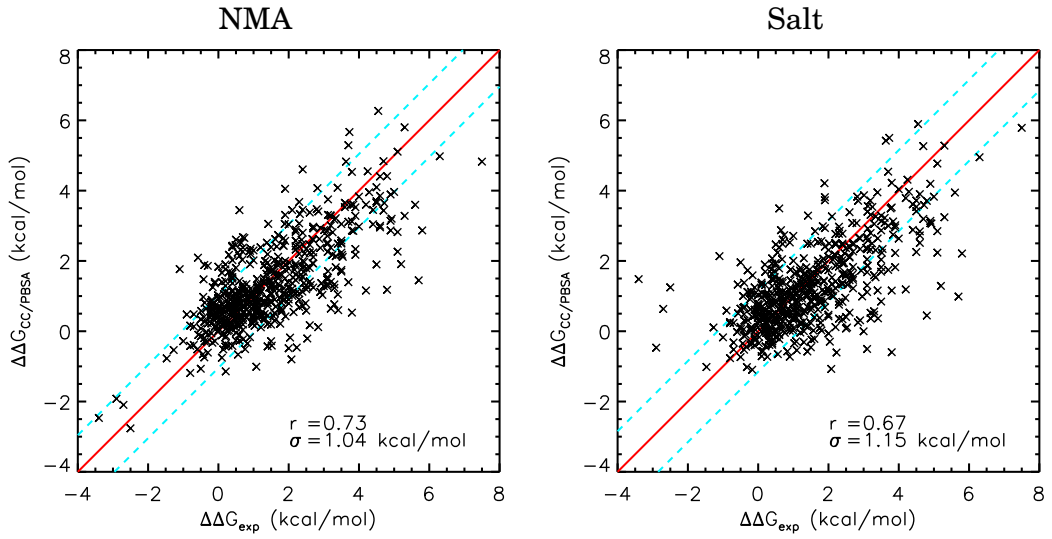

Figure 3.19: Concoord/PBSA results applying a normal mode analysis on Concoord structures (left panel) and energies choosing a physiological salt concentration (right panel).

Salt effects to polar solute-solvent interactions
In the Concoord/PBSA protocol salt effects on the reaction field energy were neglected, thus, i.e. the Poisson equation is solved instead of the more general Poisson–Boltzmann equation. For a physiological salt concentration of $I = 0.15\,\mathrm{M}$ of monovalent ions, the correlation decreased to $r = 0.665$ with an increased SDEC of $\sigma = 1.17\,\mathrm{kcal/mol}$ as shown in Figure 3.19 (right panel). The scaling factors determined using five–fold cross validation are $\alpha = 0.000265$, $\beta = 0.0818$, $\gamma = 25.34\,\mathrm{cal\,mol^{-1}\,\AA^{-2}}$ and $\tau = 0.0442$. The strongly decreased electrostatic contribution to the free energy hints to problems in the description of the unfolded state.

3.7.6 Local Dielectric Permittivity

The use of a single constant to describe the dielectric permittivity of a protein is a rough estimate. A transition region between the high permittivity solvent and the low dielectric in the protein core can be taken into account by assigning an increased intermediate permittivity to exposed groups which are frequently identified by their increased flexibility (larger B–factor). This concept of two different dielectric regions was successfully applied to pKa calculations [111] in the past.
While Voges and Karshikoff [115] gave a more detailed picture for the lo-

cal dielectricity of a protein by assigning a local dielectric constant to each amino acid, two distinct approximations are much more easier to handle:
A simple comparison of the solvent–accessible surface of residues in proteins with that of tripeptides in order to decide whether a residue is buried or exposed could serve for the classification into a high (exposed) or low (buried) dielectric region (see Figure 3.7) with predefined dielectric constants.
Also the Concoord/PBSA flexibility of amino acids could be used to distinguish low and high dielectric permittivity regions corresponding to low and high flexibility, respectively.
Two problems arise due to the usage of different dielectric regions: The regions may adopt different shapes upon mutation. This would lead to malformed electrostatic contributions to the free energy function. Also, the dielectric permittivity of the tripeptides resembling the unfolded states must match the folded counterpart.
Therefore, at this developmental stage, we refrained from the usage of the different dielectric regions inside the protein.

3.7.7 Concoord/GBSA

One of the computational bottlenecks of the Concoord/PBSA method is the numerical solution of the Poisson–Boltzmann equation for every sampled conformation. Substituting the Poisson–Boltzmann formalism by the Generalized Born theory may significantly speed up the analysis of the solvation free energy of proteins. To that end, the original finite difference method developed by Still et al. [116] (see Section 1.5.4) was implemented using the surface algorithm by Shrake and Rupley [151] (see Section 1.5.5). The implementation was tested on a non–redundant set of pdb files entitled `cullpdb_pc40_res2.0_R0.25_d080530_chains5847` that was obtained from `http://dunbrack.fccc.edu/PISCES.php` [251].
Figure 3.20 shows a very high correlation between the solution of the PBE with that of the GB theory. Unfavorable calculation times were obtained only for proteins with more than 10,000 atoms. This number exceeds the number of atoms in the considered proteins. The Concoord/PBSA test set consists of proteins with less than 1,800 atoms at most, and, evidently, the time consumption for proteins with less than 2,000 atoms is at least reduced by a factor of 5 applying the Generalized Born theory over the numerical solution of the Poisson–Boltzmann equation.
Replacing the Poisson–Boltzmann calculations in the Concoord/PBSA protocol by the Generalized Born model leads to a slightly lowered accuracy ($r = 0.711$, $\sigma = 1.11\,\text{kcal/mol}$) as compared to the Poisson-Boltzmann variant, however, with a significantly lowered computational cost. With a dielectric

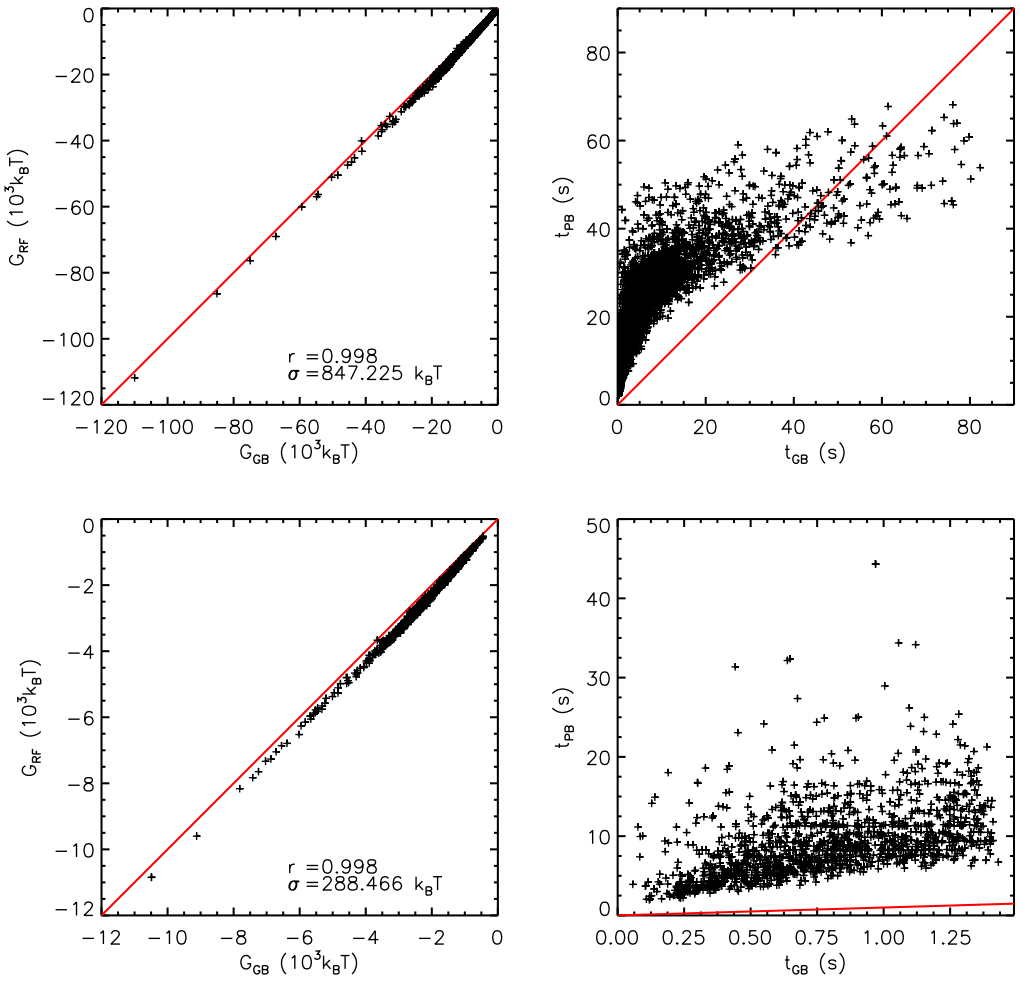

Figure 3.20: Comparing the performance and outcome of the Generalized Born technique with the Poisson-Boltzmann methodology. The reaction field energy correlates well with the GB energy (correlation coefficient of $r = 0.998$ (left)). The right panel compares the time consumption between the two methods. The whole set consisting of 5246 structures is shown on top, results for a subset containing proteins with less than two thousand atoms (1394 structures) are shown in the bottom panel.

permittivity of $\varepsilon = 1$ for the protein the scaling factors obtained through five–fold cross validation are $\alpha = 0.0863$, $\beta = 0.157$, $\gamma = 21.43\,\mathrm{cal\,mol^{-1}\,\AA^{-2}}$ and $\tau = 0.0310$. The scaled energies for 600 structures per mutant are compared to experimental data in Figure 3.21.
Due to explicitly taking the dielectric permittivity $\varepsilon_{\mathrm{solute}}$ of the protein into account via Equation (1.5.46)

$$\left(\frac{1}{\varepsilon_{\mathrm{solute}}} - \frac{1}{\varepsilon_{\mathrm{solvent}}}\right)$$

several dielectric permittivities can be scanned with a small computational effort. Concoord/GBSA showed most suitable results at $\varepsilon_{\mathrm{solute}} = 1$ as shown in Figure 3.22, compared to $\varepsilon_{\mathrm{solute}} = 2$ for the Concoord/PBA approach (see Table 3.7).
As stated by Still et al. [116], the outcome of the Generalized Born formula is roughly ten times more sensitive to small conformational alterations than for changes in the effective Born radii. This led to the idea of an ultra fast energy scanning for Concoord/GBSA by calculating the effective Born radii only once for the mutated and minimized Concoord input structure. The calculation of the polar solute–solvent interactions of 600 Concoord structures via GB applying fixed effective Born radii proved to be less time consuming than the determination of the effective Born radii of the crystal structure of the same protein (what had been calculated 600 times in the previous case). While the comparison with Poisson–Boltzmann shows a slightly decreased performance as shown in Figure 3.23 (right panel), surprisingly, the overall performance of Concoord/GBSA with fixed Born radii (left panel) was slightly improved with respect to its slower GB counterpart ($r = 0.717$, $\sigma = 1.08\,\mathrm{kcal/mol}$, $\alpha = 0.0964$, $\beta = 0.160$, $\gamma = 20.05\,\mathrm{cal\,mol^{-1}\,\AA^{-2}}$ and $\tau = 0.0373$). A scan of the dielectric permittivity gave similar results as for the *correct* Concoord/GBSA method (not shown).
An additional alteration in favor of Generalized Born that can be tested in the near future is the replacement of the distance–dependent dielectric permittivity applied for the geometry optimization in Concoord/PBSA by Generalized Born electrostatics to mimic the dielectric screening. At the time of writing this feature was not yet implemented in the official GROMACS package.

3.8 Discussion

The random structure generation based on geometrical considerations only was successfully combined with a physical free energy function using im-

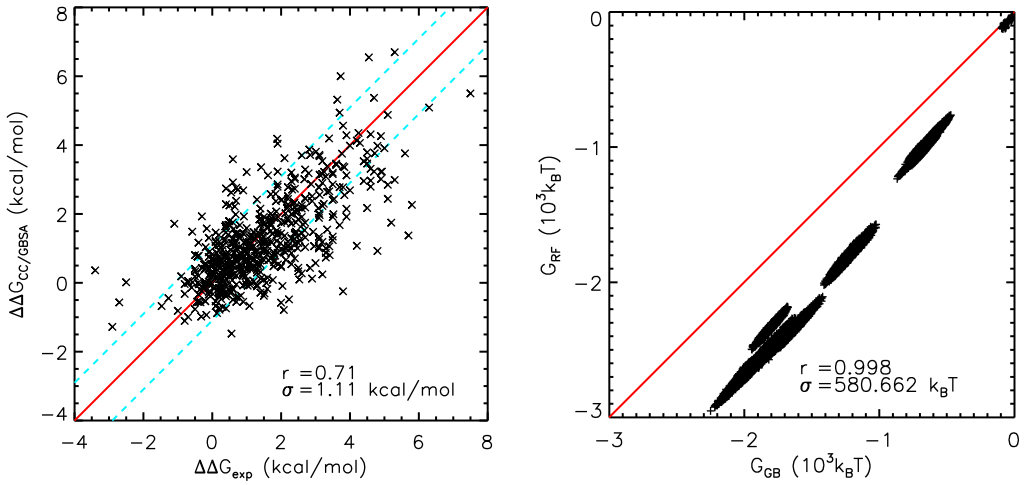

Figure 3.21: Concoord/GBSA results: Calculated free energy differences using Generalized Born are plotted against experimental values (left panel). The reaction field energies obtained by the Generalized Born model and the Poisson–Boltzmann formalism are compared for more than 360,000 structures of the test set (right panel).

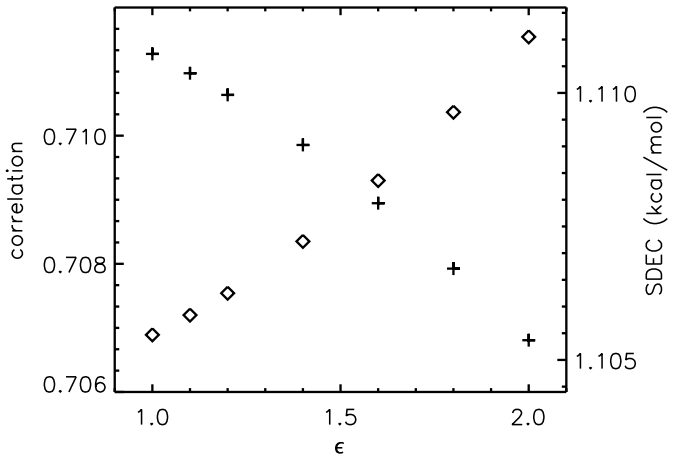

Figure 3.22: Concoord/GBSA correlation (+) and SDEC (◇) as a function of the dielectric permittivity of the protein.

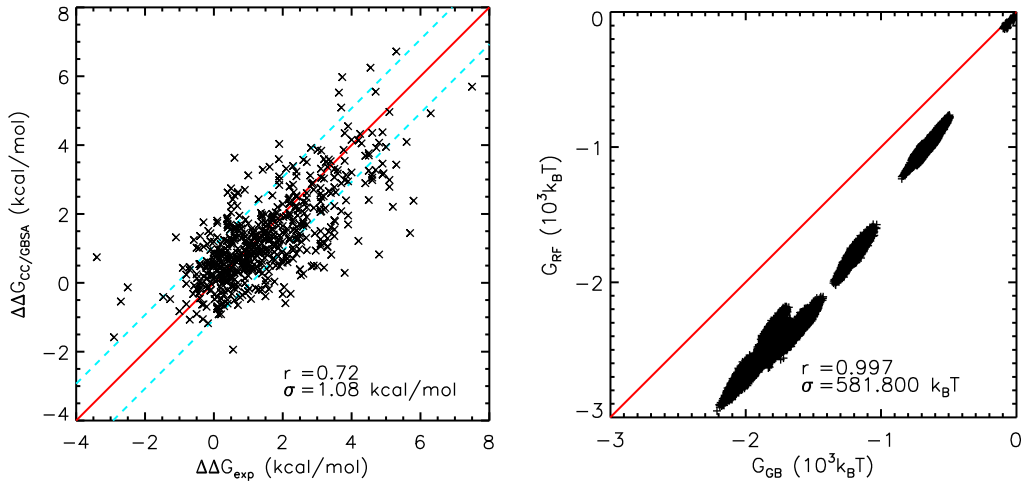

Figure 3.23: Concoord/GBSA results for fixed effective Born radii.

plicit solvent models. The resulting Concoord/PBSA method is capable of accurately calculating changes in folding free energies upon mutation.

Similar to MM/PBSA, explicit intramolecular interactions inside the protein expressed by molecular mechanics force field energies were coupled with continuum solvent approximations. Slow MD simulations were replaced by Concoord–generated random structures that sample the conformational space in the vicinity of the starting structure. The single energy contributions were adjusted by introducing four scaling factors to reproduce experimental energies at high accuracy.

The development of the Concoord/PBSA method was based on a test set of five proteins with experimentally known folding free energies for 582 mutants and with known crystal structure of the (pseudo) wild type. The analysis of the full mutation set led to a correlation $r = 0.75$ to experiment and a standard deviation (SDEC) of $\sigma = 1.04$ kcal/mol.

Within this approach, it is shown that inclusion of conformational flexibility via averaging of energies computed on an ensemble of structures is crucial for a reliable prediction of folding free energy changes due to mutation. A number of 300 structures was found as a compromise between accuracy and computation time.

An analysis of the same set of mutants using Fold-X revealed a similar accuracy ($r = 0.76$, $\sigma = 1.02$ kcal/mol). Different to Fold-X, Concoord/PBSA additionally yields a structural ensemble that can be used as a starting point for further studies. Clustering algorithms [252, 253] or methods determining

the quality of structures [242, 254] possibly further increase the accuracy. The single energy rated structures could also be used for conformational prediction in homology modelling or as a structural basis set for the docking of ligands.
As a fast alternative to the numerical solution of the Poisson–Boltzmann equation the Generalized Born electrostatics was implemented and tested.
Different from the MM/PBSA method, Concoord/PBSA makes use of scaling factors for the different energetic contributions. Probable reasons for these scaling factors are both due to the different structure generation and due to the neglect and simplification of energetic contributions, as outlined in the following.

Structural ensemble
One important approximation with respect to MM/PBSA is the use of minimized Concoord structures in place of snapshots taken from explicit solvent molecular dynamics trajectories. While the latter provide genuine statistical ensembles in equilibrium allowing to obtain potential and kinetic energies as ensemble averages, the random structures obtained via Concoord/PBSA represent local energetic minima.
However, Concoord was reported to sample a similar conformational subspace as compared to MD [64] for proteins residing in a single preferred conformation. For proteins or mutants expected to undergo conformational transitions, explicit MD simulations are required for the sampling of the (available) conformational space.

Unfolded state
Another assumption concerns the treatment of the unfolded state. The simple tripeptide model mainly covers mutation–induced changes of the solvation free energy of the unfolded state. Differences in the long range interactions that are caused by mutations are not considered in this method. Possible improvements in the description of the unfolded state maybe achieved e.g. by the inclusion of long range electrostatic interactions using a Gaussian chain model [231] or a more comprehensive model of the unfolded state. The latter can be thought of as an ensemble of random coils without regular secondary structure. Similar to the method developed by Elcock [239] it should be possible to use Concoord with altered distance restraints for non-bonded pairs to sample unregular structures. However, the unfolded structure of a protein may well encompass a non–random structure [255, 256] which further complicates the energetic estimate of the unfolded state.

Continuum solvent approaches

A further key element of the Concoord/PBSA technique is the use of continuum solvent approaches at various states of the computation while using a MM force field that was partially parametrized for explicit solvent usage.
In the optimization process a distance dependent dielectric permittivity $\varepsilon(r) = 40\,\mathrm{nm}^{-1} r$ mimics solvent screening. This screening is essential for preventing incorrect modelling of solvent exposed polar or charged amino acid side chains. While energy minimizations with explicit solvent models are too time consuming to meet the requirements for a fast method, minimizations using the Generalized Born method may serve as an alternative (soon available in Gromacs).
Implicit solute–solvent interactions approximated by the solution of the Poisson–Boltzmann equation, the Generalized Born model, or by the molecular surface area approach are partly based on empirical parameters like the dielectric constant of a protein or the surface tension that show large differences in literature. Although the most suitable parameters were chosen for the whole test set, it is possible that these parameters vary slightly from protein to protein, e.g. depending on the polarity of the protein surface.
A general problem is the smoothly varying dielectric permittivity between the protein interior ($\varepsilon \approx 2$) and the bulk ($\varepsilon \approx 78$). For the numerical solution of the PB equation $\varepsilon = 2$ was chosen. However, changing the dielectric constant to $\varepsilon = 1$ results in a doubling of the Coulomb interaction and the reaction field shows a scaling behavior similar to $\frac{1}{\frac{1}{\varepsilon}-\frac{1}{78}}$. The electrostatics weighting factor $\alpha = 0.224$ may hint to a dielectric constant of about $\frac{\varepsilon}{\alpha} = 9$ for $\alpha = 1$, although calculations with $\varepsilon = 2$ yield the overall largest accuracy.

Schlitter's entropy estimation

Schlitter's method yields an upper limit for the entropy of the solute. Therefore, a scaling is required within the context of the whole energy function.

Experimental conditions

Experimental folding free energies used for the parameterization of Concoord/PBSA were based on a variety of experimental conditions and methods with differing accuracies. A refinement of Concoord/PBSA could therefore be achieved by a method or condition dependent scaling of parameters for the mutational test, i.e. results of Concoord/PBSA would be dependent on the experimental method or condition chosen for comparison.

Energy Function and Scaling Factors

The Concoord/PBSA free energy is a linear combination of the considered energy contributions. A regression fit applied to find suitable scaling factors is independent of the sources of the individual terms, i.e. by fitting energetic terms neglected in the free energy function may implicitly be considered.
A similar scaling approach was previously also applied to MM/PBSA calculations [257]. As the fitting procedure was applied to one protein with different ligands only, the obtained weighting factors are probably highly system dependent.

Table 3.9: Importance of the different Concoord/PBSA contributions: Correlation, SDEC and scaling coefficients for different forms of the energy function when neglecting up to two contributions are shown. [a)] σ in kcal/mol, [b)] γ in $\mathrm{cal\,mol^{-1}\,\AA^{-2}}$.

Energy Function	r	σ [a)]	α	β	γ [b)]	τ
$\Delta G_{\mathrm{CC/PBSA}}$	0.748	1.04	0.224	0.217	16.64	0.0287
$\Delta\Delta G_{\mathrm{es}} + \Delta\Delta G_{\mathrm{SA}} - T\Delta\Delta S$	0.741	1.05	0.209	—	29.87	0.0248
$\Delta\Delta G_{\mathrm{es}} + \Delta\Delta G_{\mathrm{LJ}} - T\Delta\Delta S$	0.737	1.06	0.244	0.468	—	0.0266
$\Delta\Delta G_{\mathrm{es}} + \Delta\Delta G_{\mathrm{LJ}} + \Delta\Delta G_{\mathrm{SA}}$	0.731	1.05	0.241	0.179	15.57	—
$\Delta\Delta G_{\mathrm{LJ}} + \Delta\Delta G_{\mathrm{SA}} - T\Delta\Delta S$	0.665	1.16	—	0.080	25.46	0.0438
$\Delta\Delta G_{\mathrm{es}} + \Delta\Delta G_{\mathrm{SA}}$	0.727	1.06	0.226	—	26.88	—
$\Delta\Delta G_{\mathrm{es}} + \Delta\Delta G_{\mathrm{LJ}}$	0.721	1.07	0.258	0.417	—	—
$\Delta\Delta G_{\mathrm{SA}} - T\Delta\Delta S$	0.665	1.17	—	—	30.32	0.0419
$\Delta\Delta G_{\mathrm{LJ}} - T\Delta\Delta S$	0.633	1.21	—	0.462	—	0.0424
$\Delta\Delta G_{\mathrm{LJ}} + \Delta\Delta G_{\mathrm{SA}}$	0.622	1.21	—	0.004	24.76	—
$\Delta\Delta G_{\mathrm{es}} - T\Delta\Delta S$	0.270	1.69	0.229	—	—	-0.115
$\Delta\Delta G_{\mathrm{CC/PBSA}} - \Delta\Delta G_{\mathrm{coul}}$	0.665	1.17	0.0009	0.078	25.61	0.0429
$\Delta\Delta G_{\mathrm{CC/PBSA}} - \Delta\Delta G_{\mathrm{RF}}$	0.665	1.17	0.0001	0.081	25.42	0.0439
$\Delta\Delta G_{\mathrm{CC/PBSA}} + \Delta\Delta G_{\mathrm{bonded}}$	0.739	1.05	0.210	0.041	27.38	0.0297

The significance of the individual energetic terms for the full free energy function was determined by neglecting contributions and applying a five–fold cross validation procedure to the reduced set. All possible permutations including two or three contributions are shown in Table 3.9.
Seemingly, the Lennard–Jones energies and the molecular surface term contain redundant data for the calculation of folding free energies. Neglecting one contribution leads to a roughly doubled scaling of the other while the scaling of electrostatics (α) and of entropy (τ) remained almost unaffected. The redundancy of both energies was examined by a comparison as shown in Figure 3.24. Surprisingly, the SA energies showed a high correlation of

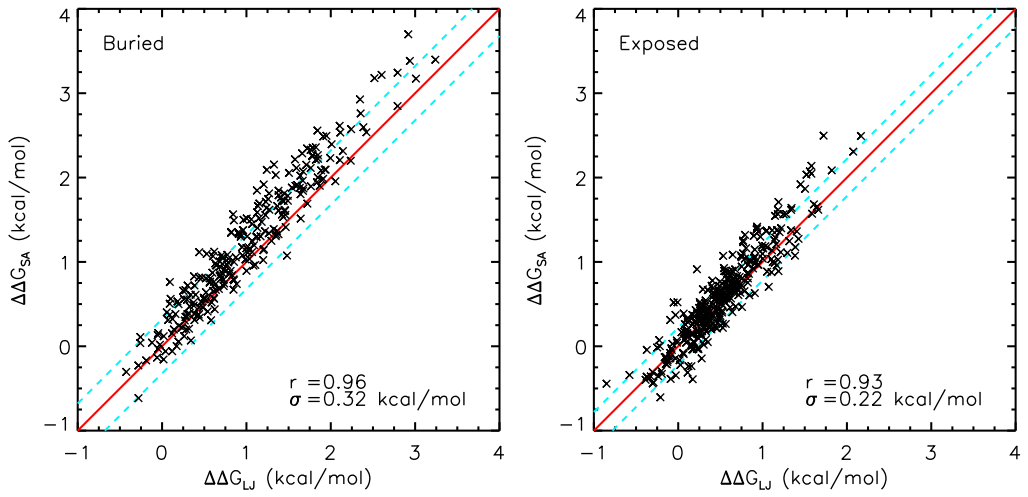

Figure 3.24: The scaled energy terms (taking the full Concoord/PBSA energy function) of the non-polar solvation contribution and the Lennard-Jones interaction are compared for mutations at buried and exposed sites revealing a surprisingly high correlation.

$r = 0.96$ with a SDEC of $\sigma = 0.27\,\text{kcal/mol}$ to the LJ energies. While exposed mutation sites are expected to show different behavior in the surface area than buried ones, the molecular surface term and Lennard–Jones contributions were compared for exposed and buried mutations separately, both resulting in a similar picture. One has to keep in mind, however, that only the mutation–induced energy differences are similar. The total free energies for these two contributions differ considerably (see Figure 3.9).

Also the importance for the inclusion of the full electrostatics consisting of the Coulombic and reaction field contributions is seen (Table 3.9). Neglect of either Coulombic or reaction field term resulted in a vanishing contribution of the second electrostatic contribution.
The stability of the electrostatic scaling parameters was tested by introducing individual scaling parameters for all four states of the thermodynamic cycle (Equation (3.2.3)), i.e. the folded and unfolded wild type and mutant states. All scaling factors were determined to approximately ±0.22 with signs correctly predicted in all cases (not shown).
The value for the surface surface tension $\gamma = 16.64 \frac{cal}{mol\,\text{Å}}$ is significantly larger than the commonly used $5\,\text{cal}\,\text{mol}^{-1}\,\text{Å}^{-2}$ introduced by Sitkoff et al. [142]. This frequently applied value in MM/PBSA calculations [51] was directly fitted to hydration free energies of sidechain analogues and therefore finds its use in unscaled energy functions.

The last energy function analyzed in Table 3.9 is the Concoord/PBSA energy function that now includes the full force field energies (bond, angle, dihedral,...). The obtained accuracy is comparable to the set without any force field contributions at all. A probable reason is the use of minimized input structures. The steep energy functions of the bonded contributions may lead to large errors if bond lengths or angles are far from equilibrium.

Chapter 4

Protein binding affinities

The existing Concoord/PBSA method was adopted for the prediction of protein–protein binding affinities. This chapter provides a brief description of the Concoord/PBSA approach for the prediction of mutational effects on binding affinities as well as its application to protein–protein and protein–ligand binding.

For an unbiased comparison of the Concoord/PBSA prediction for mutational effects on the binding free energies of protein–protein complexes to other well–established, fast methods the TEM1–BLIP complex was chosen (Chapter 4.2). For this complex about 100 data points for mutations to alanine are known from experiment.

The size of this complex is, however, prohibiting for a comparison to the simulation–based MM/PBSA approach (Chapter 1.7.2). Here, the considerably smaller p53–MDM2 complex was selected since this system was studied previously using MM/PBSA [166] (Chapter 4.3).

Apart from the above systems, Concoord/PBSA was applied to complete mutational scans of proline–rich peptides binding to the GYF domain (Chapter 4.4) and of the insulin–insulin dimer interface in order to suggest mutations with a decreased binding affinity (Chapter 4.5). The latter is crucial to identify so–called fast insulins that are used in the treatment of diabetes mellitus.

4.1 Methods

4.1.1 Binding Affinity Predictions with Concoord/PBSA

Several changes have been made to the Concoord/PBSA approach for the calculation of mutational changes of binding affinities [1]. The most important one is due to an adapted thermodynamic cycle[1]:

[1]Although both binding partners are described as mutant in the mutant states (lower row) of the thermodynamic cycle, only one partner may truly carry a mutation.

$$\begin{array}{ccc} G^{\text{partner A}}_{\text{Wildtype}} + G^{\text{partner B}}_{\text{Wildtype}} & \xrightarrow{\Delta G^{\text{bind}}_{\text{Wildtype}}} & G^{\text{compound}}_{\text{Wildtype}} \\ \Big\downarrow \Delta G^{\text{mutate}}_{\text{unbound}} & & \Big\downarrow \Delta G^{\text{mutate}}_{\text{compound}} \\ G^{\text{partner A}}_{\text{Mutant}} + G^{\text{partner B}}_{\text{Mutant}} & \xrightarrow[\Delta G^{\text{bind}}_{\text{Mutant}}]{} & G^{\text{compound}}_{\text{Mutant}} \end{array} \tag{4.1.1}$$

The denatured state is irrelevant for binding calculations. The change in free energy upon binding, i.e. the binding free energy, is given by

$$\Delta G^{\text{bind}} = G^{\text{compound}} - \left(G^{\text{partner A}} + G^{\text{partner B}} \right). \tag{4.1.2}$$

While Concoord/PBSA ensembles were generated for the wild type and mutated complex structures, conformations of the respective binding partners were obtained from the complex Concoord/PBSA ensembles by removing the other partner, similar to the MM/PBSA approach. Thus, only interaction energies between both partners as found in the complex structures are determined, and bonded energies calculated from the force field cancel. Also by restricting the structure generation to the complex corresponding CPU time is decreased by more than a half.

As the structures were only generated for the complex conformations, the usually increased flexibility of the isolated proteins with respect to the complex is disregarded. Due to the weakness of implicit solvent free energy functions based on present–day force fields in distinguishing different conformational substates of proteins, this approach was reported to yield significantly improved results with respect to approaches explicitly considering the conformational flexibility of the isolated proteins [166].

The Concoord/PBSA energy function for binding free energy changes upon mutation has the form

$$\Delta\Delta G^{\text{binding}}_{\text{CC/PBSA}} = \alpha \Delta\Delta G_{\text{es}} + \beta \Delta\Delta G_{\text{LJ}} + G_{\text{PPIS}}, \tag{4.1.3}$$

introducing an additional cooperativity contribution

$$G_{\text{PPIS}} = \gamma I_{\text{wt}} + c \tag{4.1.4}$$

that is proportional to the protein–protein interaction surface of the wild type crystal structure I_{wt} (evaluated with a probe size of 1.4 Å). The weighting factors were determined using five–fold cross validation on a set of 367 mutants from 9 protein–protein complexes yielding $\alpha = 0.137$, $\beta = 0.258$, $\gamma = -0.768\,\text{cal}\,\text{mol}^{-1}\text{Å}^{-2}$ and $c = 2.574\,\text{kcal/mol}$.

Consideration of entropy did not show any significant effect on the results (data not shown).
For selected cases, changes in protonation states of titratable amino acids upon complex formation or/and upon mutation were considered using a correction term to $\Delta\Delta G^{\text{binding}}_{\text{CC/PBSA}}$ [258, 259] (based on pKa calculations, see next section):

$$\Delta G_{\text{pK}} = k_B T \ln(10) \cdot \left(\text{pK}_\text{a} - \text{pH}_{\text{exp}}\right). \tag{4.1.5}$$

The calculated binding free energies from the used test set data are in very good agreement to the experimental values with a correlation of $r = 0.80$ and a SDEC $\sigma = 1.17\,\text{kcal/mol}$ (Figure 4.1).

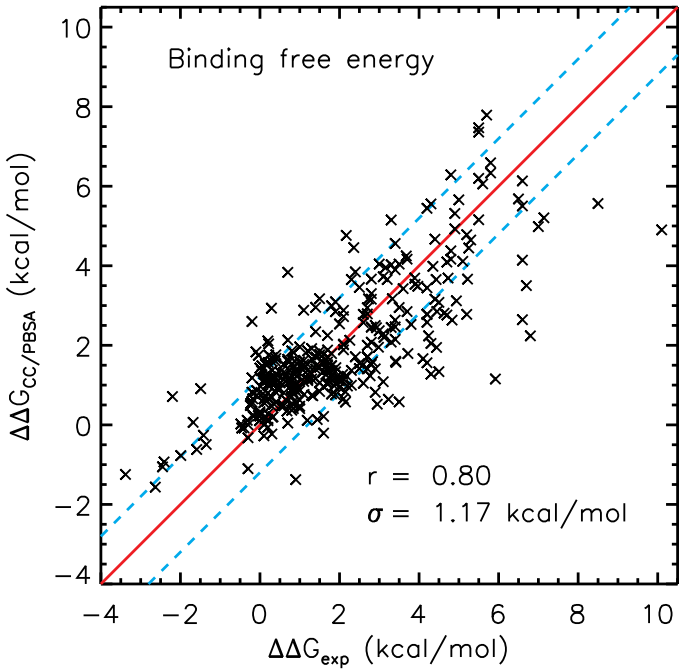

Figure 4.1: Concoord/PBSA binding free energy results: The calculated changes in binding free energy upon mutation are plotted against experimentally determined values. Fitting the weights of the single contributions leads to a correlation of $r = 0.80$ and a standard deviation of $\sigma = 1.17\,\frac{\text{kcal}}{\text{mol}}$.

4.1.2 pK_a Calculations

The calculation of pK_a values for titratable groups is based on the thermodynamic circle for the protonation/de–protonation of an acceptor A or donator

AH isolated in water and in its environment, i.e. in the protein [111, 260–262]:

$$\begin{array}{ccc} \text{AH in Water} & \xrightarrow{\Delta G^{\text{deprotonate}}_{\text{in Water}}} & \text{A in Water} \ +\text{H}^+ \\ \Big\downarrow \Delta G^{\text{env}}_{\text{AH}} & & \Big\downarrow \Delta G^{\text{env}}_{\text{A}} \\ \text{AH in Protein} & \xrightarrow[\Delta G^{\text{deprotonate}}_{\text{in Protein}}]{} & \text{A in Protein} \ +\text{H}^+ \end{array} \tag{4.1.6}$$

These calculations usually only take electrostatic contributions into account. While $\Delta G^{\text{deprotonate}}_{\text{in Water}}$ is experimentally known and listed in textbooks as the model pK^0_a for isolated titratable groups in water environment, the two energies ($\Delta G^{\text{env}}_{\text{AH}}$ and $\Delta G^{\text{env}}_{A}$) introducing the protonated and de–protonated state into the protein environment lead to an *intrinsic* $\text{pK}^{\text{int}}_\text{a}$. The environmental free energy contribution ΔG^{env} consists of the change in reaction field energy upon insertion in the dielectric environment of the uncharged protein, ΔG^{RF}, and the electrostatic interactions between the inserted amino acid and the permanent dipoles of the (now charged) rest of the protein ΔG^{dipole}

$$\Delta G^{\text{env}} = \Delta G^{\text{RF}} + \Delta G^{\text{dipole}}. \tag{4.1.7}$$

Every other titratable group is taken in its neutral state for calculating ΔG^{dipole}.

The intrinsic pK_a is now obtained via

$$\text{pK}^{\text{int}}_\text{a} = \text{pK}^0_\text{a} - \frac{\zeta}{k_B T \ln(10)} \left(\Delta G^{\text{env}}_{\text{A}} - \Delta G^{\text{env}}_{\text{AH}} \right). \tag{4.1.8}$$

with $\zeta = 1$ for acidic and -1 for basic groups. For molecules with a single titratable site the true pK_a equals the intrinsic pK^0_a. For multiple titratable sites, the interaction between all titratable groups has to be taken into account.

The change in free energy for charging two titratable groups $\Delta\Delta G_{ij}$ is given according to the thermodynamic cycle

$$\begin{array}{ccc} (0,0) & \longrightarrow & (0,1) \\ \Big\downarrow & & \Big\downarrow \\ (1,0) & \longrightarrow & (1,1) \end{array} \tag{4.1.9}$$

$$\Delta\Delta G_{ij} = G_{ij}(1,1) - G_{ij}(0,1) - G_{ij}(1,0) + G_{ij}(0,0). \tag{4.1.10}$$

Here 0 denotes a neutral state and 1 a charged state that is later on expressed by ξ. The pair (ξ_i, ξ_j) describes the charge state of both interacting partners i and j. The charge q_i on group i can be described as

$$q_i = -\xi_i \zeta_i, \tag{4.1.11}$$

with the sign of the charge defined by ζ as defined above.
For a given protonation state n of the whole protein with N titratable groups, the electrostatic free energy is given by [260]

$$\Delta G^n = \sum_{i=1}^{N} \left[q_i^n \left(2.3\, k_B T \left\{ pH - pK_a^{\text{int}} \right\} \right) + \xi_i^n \sum_{1 \le j < i} \xi_j^n \Delta\Delta G_{ij} \right]. \tag{4.1.12}$$

With a total of 2^N possible protonation states, the average charge $< q_i >$ of group i is given by

$$< q_i >= \frac{\sum_{n=1}^{2^N} -\xi_i^n \zeta_i e^{-\Delta G^n / k_B T}}{\sum_{n=1}^{2^N} e^{-\Delta G^n / k_B T}}. \tag{4.1.13}$$

As for large proteins the number of 2^N states is not feasible on a realistic time scale any more, Monte Carlo sampling techniques are frequently used to find the energetic minimum at a given pH corresponding to the most probable protonation fraction. A Henderson-Hasselbalch fit of the titration curve $< q_i >$ (pH) then yields the pK_a of a group i (midpoint of the fit). The accuracy is in the range of one pK_a unit.
While WHAT IF [242] uses the approach presented above for the calculation of a pK_a, the freely available MCCE program (`http://www.sci.ccny.cuny.edu/~MCCE/`) [263] follows a slightly different approach that also includes different sidechain conformations for the titratable groups.

4.2 Comparison of Concoord/PBSA to Fold-X and Robetta for the TEM1–BLIP complex

4.2.1 TEM1–BLIP complex

The study of the TEM1–BLIP complex is highly relevant for improving β–lactam antibiotics like penicillin [264]. β–lactamases such as the TEM–1 β–lactamase (TEM1), can be found in various bacteria. These enzymes are capable of hydrolyzing an important amide bond of β–lactamase antibiotics

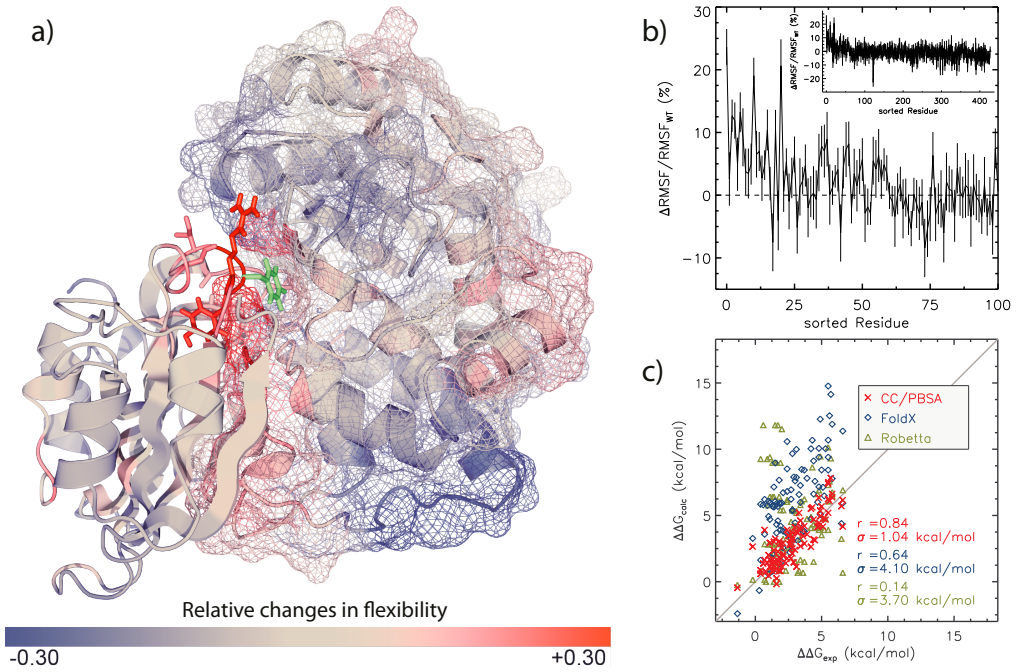

Figure 4.2: Effect of alanine mutations on the TEM1–BLIP complex. (a) Relative changes in flexibility of the TEM1–BLIP complex upon the F142A mutation (green). Side chains with large increase in flexibility are shown in stick representation. (b) Relative change in root mean square fluctuations (RMSF) upon the F142A mutation for all residues in the TEM1–BLIP complex, sorted according to their distance to the mutation site. (c) Calculated changes in binding free energy for TEM1–BLIP alanine mutants applying Fold-X, Robetta and Concoord/PBSA. For the comparison, parameters for CC/PBSA were fitted on the remaining dataset on other protein–protein complexes only. The diagonal line corresponds to ideal prediction.

rendering them ineffective. The β–lactamase inhibitory protein (BLIP) is a potent inhibitor of TEM1. However, still many bacterial organisms are capable of hydrolyzing antibiotics in the presence of BLIP. Mutational studies may eventually lead to the design of a more powerful β–lactamase inhibitory protein analogue that increases the number of bacteria sensitive to β–lactam antibiotics.

For the TEM1–BLIP complex, 96 (alanine) mutants and their effect on the binding affinity were collected from literature [265–267]. These served as an unbiased test set for a comparison of the prediction accuracy of Concoord/PBSA to other well–established methods, Fold-X [59] and Robetta (`http://robetta.bakerlab.org/`) [268]. The test set was restricted to mutations to alanine since Robetta does not allow for different target amino acids. All predictions were based on the same wild type crystal structure for the

TEM1–BLIP complex (1JTG) [269].

4.2.2 Results

Figure 4.2 c) shows the correlation of the calculated values of Robetta, Fold-X and of Concoord/PBSA with the experimentally determined mutational changes in binding free energies:Concoord/PBSA results are depicted in red crosses, Fold-X in blue diamonds and Robetta in green triangles. Clearly, Concoord/PBSA outperforms both Fold-X and Robetta. While the correlation to experiment is reasonable for Fold-X ($r = 0.64$), Robetta is not able to predict the influence of mutations on the binding affinity of the TEM1–BLIP complex ($r = 0.14$). The latter is probably due to the observation, that mutations of the TEM1–BLIP complex act in a highly cooperative manner [266], while Robetta assumes a linear superposition of mutational effects. It is therefore not able to cope with cooperative effects by construction. In contrast, the structure generation in Concoord/PBSA is based on the particular mutant structure and thus, apart from cooperative effects in the free energy function, explicitly considers mutation–induced differences in flexibility. This is shown in Figure 4.2 a): the flexibility of residues close to the mutant residue for the BLIP–`F142A` mutant was substantially enhanced (mutation site colored in green, increase in rmsd in red, decrease in blue). This effect is not taken into account by the Fold-X method or by Robetta.

4.3 Full mutational scan of p53 bound to MDM2 and comparison to MM/PBSA

4.3.1 Function and Importance

The multifunctional transcription factor p53 is responsible for cell cycle regulation and acts as a tumor suppressor, among others [270]. Mutations of p53 or down–regulation by overexpressed murine double minute 2 protein (MDM2) is found in most cancers. In order to restore p53 functionality in cancers that leave the p53 sequence unchanged, the interaction between MDM2 and p53 is a promising target for therapeutic studies. As p53 is only fully functional in its tetrameric form, cancers showing mutated p53 with intact binding interfaces may be targeted with altered p53 proteins. Fully functional, mutated p53 with a decreased oligomerization affinity towards the inactive p53 may be induced in cancer cells to restore the basal level and activity of p53 [271, 272].

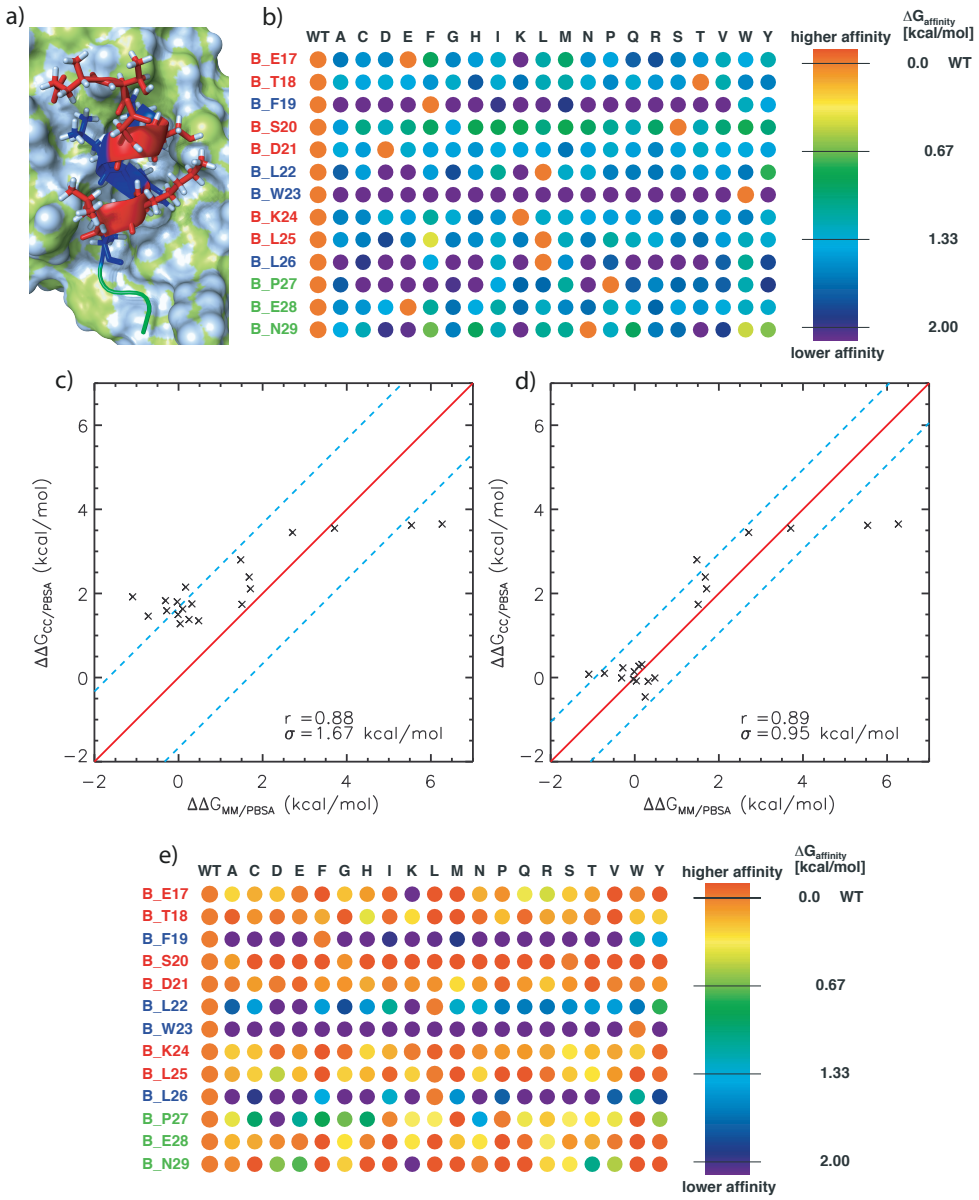

Figure 4.3: Concoord/PBSA full mutational scan of the p53 binding interface to MDM2 and comparison to MM/PBSA. a) Cartoon representation of the binding interface shows the sequence of p53 colored according to its solvent accessibility (blue inaccessible, red exposed and green exposed but no comparison to MM/PBSA possible). b) and e) show the full mutational scan of the p53 binding interface with full inclusion of the interface term (b) and with inclusion according to the solvent accessibility (e). The first column shows the wild type sequence colored as in (a). The spots represent predicted binding free energy changes upon a single–point mutation determined by row (mutation site) and column (mutant amino acid). Plots c) and d) show the correlation with MM/PBSA with and without full inclusion of the interface term.

The p53–MDM2 binding was studied in order to check whether Concoord/PBSA is able to identify known hot spots and for comparison to MM/PBSA calculations reported by Massova and Kollman [166].

4.3.2 Results

Figure 4.3 b) shows a complete mutational scan of p53 bound to human MDM2 (PDB entry 1YCR) [273]. The experimentally determined hot spots, `B_F19`, `B_W23` and `B_L26`, were clearly identified also by Concoord/PBSA. Almost all mutations of these amino acids result in a loss in binding affinity of more than 2 kcal/mol.
A direct comparison to the MM/PBSA results for p53 bound to human (PDB entry 1YCR) and frog MDM2 (PDB entry 1YCQ) (both crystal structures determined by Kussie et al. [273]) shows a significant deviation for those mutants that have a small effect on binding affinity as predicted by MM/PBSA (Figure 4.3 c)). A more detailed analysis revealed that these mutational sites are solvent–exposed, i.e. not at the interface between p53 and MDM2. For these mutants, the protein–protein interaction surface term G_{PPIS} was neglected for further analysis and resulted in significantly improved agreement (Figure 4.3 d)). The cartoon in Figure 4.3 a) shows a surface model of MDM2 with a bound p53. The p53 sequence in the cartoon as well as in the mutational scan is color–coded according to its solvent accessibility — blue buried, red solvent exposed — and the possible importance of the interaction surface term, respectively. The green colored, solvent exposed C–terminus of the p53 peptide chain has not been studied by Massova and Kollman [166]. The results underline that the interface term G_{PPIS} should only be considered for interface mutations. A possible distance–dependence of the interface term cannot be determined based on the limited available experimental data for protein–protein binding. The mutational scan when neglecting the interface term for solvent exposed mutation sites is shown in Figure 4.3 e).

4.4 Proline–rich peptide binding to the GYF domain

4.4.1 Function and Importance

GYF domains are responsible for the recognition of proline–rich sequences [274, 275] next to profilin, SH3, the WW, the EVH1, and the UEV domains. These intracellular domains assist in the coordinated assembly of multi–protein complexes. The GYF domain is named after the glycine–tyrosine–

phenylalanine tripeptide occurring in the binding interface that is important for the recognition of the sequence motif (R/K/G)XXPPGX(R/K) by CD2BP2 [276]. Kofler et al. [276] analyzed the binding affinity towards single–point mutants of the sequence SHRPPPPGHRV via spot synthesis [32] shown in Figure 4.4. The synthesized peptides are fixated and arranged as spots on a cellulose membrane. In a following step the binding partner is brought into contact with the cellulose membrane and can bind the immobilized peptide chain. Via chemiluminescent methods the spot intensities and radii are obtained. Both correlate with the binding affinity between the GYF–domain and the corresponding peptide chain [32, 277, 278].

4.4.2 Results

A full mutational scan of the SHRPPPPGHRV sequence, comparable to the spot analysis of Kofler et al. [276], has been evaluated using the Concoord/PBSA web interface based on the crystal structure of the CD2BP2-GYF domain (1L2Z) [279] that is in complex with the mentioned peptide sequence. The outcome of the calculation is shown in Table B.1. The results and a comparison to the spot synthesis are depicted in Figure 4.4. As a direct quantification of the experimental spot analysis is not possible due to the large noise and the low resolution of the picture, the peptides were sorted according to binders or non-binders by visual inspection. For many mutants no clear cut was possible. The calculated binding free energies can be sorted more easily by a well defined cut–off value. Based on the differentiation of the experimental data the cut–off was obtained by computationally scanning for the highest accuracy. For a cut–off energy difference of 1.97 kcal/mol an accuracy (quotient of correctly predicted energy differences and the total number of considered mutants) of 72.4% was achieved.
Although a high accuracy was obtained, two main problems exist:

- Alternative binding modes for the GYF domain have been observed using MD simulations [280]. Here, the peptide chain with the mutation G8W is shifted by one position while the binding site remains. Such conformational changes cannot be predicted by Concoord/PBSA. Also, a different binding site for SH3 domains has been identified by means of MD [281]. A similar behavior of the GYF domain cannot be excluded. Alternative binding modes may be considered by loosened constraints between the peptide and the protein combined with a structure–based clustering [252, 253]. This was, however, not considered here due to the limited accuracy of the experimental data.

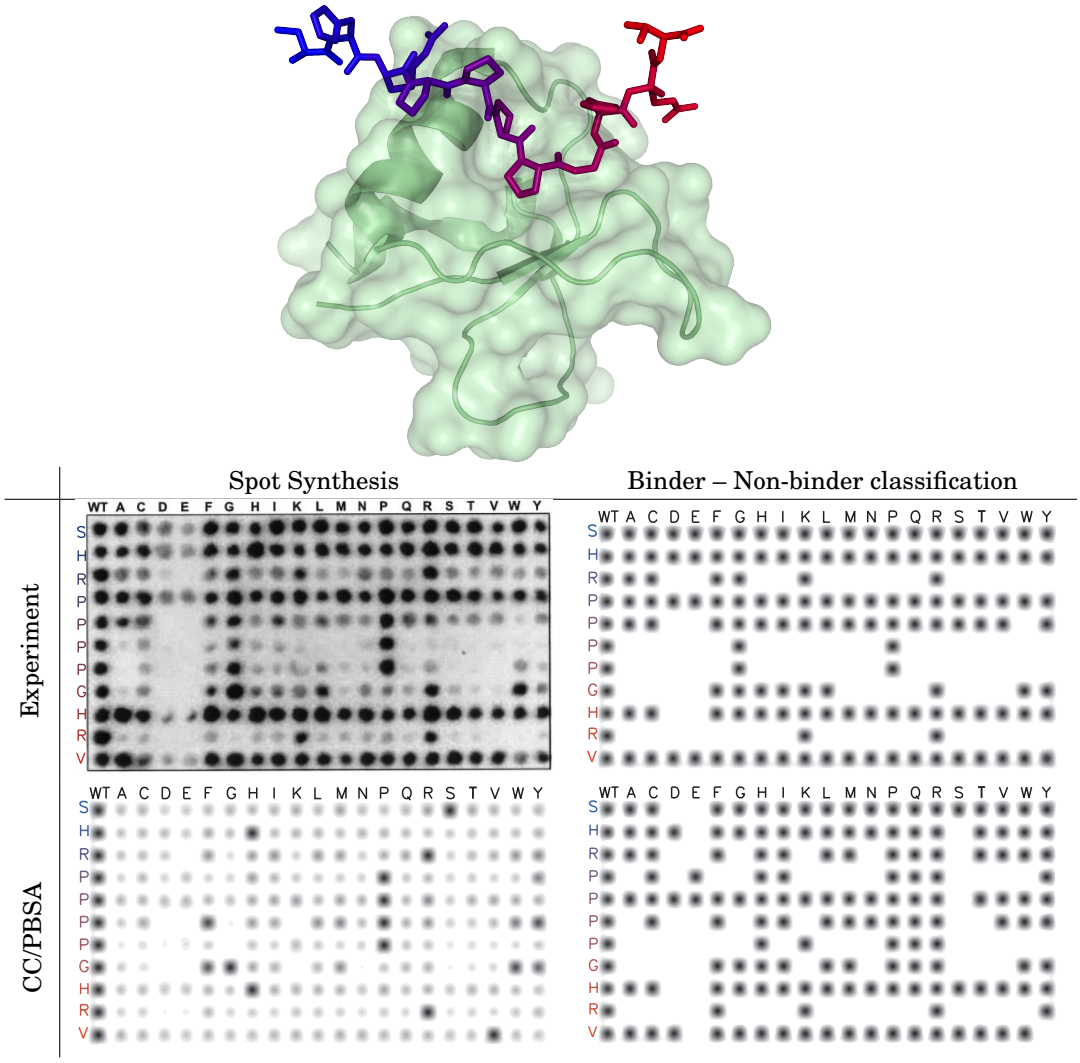

Figure 4.4: Concoord/PBSA energies for proline-rich peptide binding to the GYF domain: At the top the GYF domain is shown in light green with the peptide **SHRPPPGHRV** colored blue to red from the N to C–terminus, respectively. This also applies for every appearance of the sequence. The experimental spot synthesis was taken from [276], the classification according to binder or non-binder is an estimate done by visual inspection. The Concoord/PBSA results that are shown below are displayed in a similar fashion. The intensities and radii of the spots shown in the *Spot Synthesis* column correlate with the binding free energy. The achieved accuracy with respect to the classification is 72.4%.

- Extensive pK_a calculations of the whole mutational set using MCCE [263] revealed possible changes in the protonation state of the second histidine of the peptide chain upon binding. The calculated pK_a for the second histidine is raised from the model pK_a of 6.0 to 7.0 in the bound and 6.7 in the unbound state of the wild type (6.6 and 6.5 for the first histidine, respectively). As a consequence, the protonation states of these histidines are not well defined when assuming a pH of around 7.0. Another problem arising from the experimental data is the lack of reported pH conditions of the study. Thus, pH 7 is only an assumption. Due to the above mentioned uncertainties concerning a pK_a correction term ΔG_{pK} (see Equation (4.1.5)) was not considered in the presented results.

  Results of the pK_a calculations for the two histidines in the peptide chain are presented in Table B.1 in the Appendix.

4.5 Dimerization of Insulin

4.5.1 Function and Importance

Insulin is a hormone consisting of 51 amino acids distributed among two distinct chains that form the monomer [2]. A cartoon of an insulin dimer is shown in Figure 4.5 b). Insulin is crucial for the regulation of the physiological glucose level in blood. By binding to the insulin receptor the glucose uptake is activated. The malfunctioning of the regulatory process is known as *diabetes mellitus*.

Type 1 diabetes mellitus patients suffer from an autoimmune disease. The production of insulin is hindered by the destruction of insulin secreting β-cells of the pancreas. Here, an insulin replacement therapy medicating insulin analogues is applied [282–285]. A huge problem is to mimic a physiological insulin secretion. While fast acting insulin analogues show no physiological rapid rise in plasma insulin concentration shortly after a meal, long lasting insulin analogues provide no constant insulin level over night or between meals. With slight alterations to the human insulin it is tried to converge the activity profile of the analogues to the regular physiological secretion. The change in starting time, impact and decay are accomplished by exploiting the ability of insulin to self-associate.

The physiological active monomer forms dimers at micromolar concentrations. Dimers further associate to hexamers in the presence of zinc.

The idea behind rapid acting insulin is to reduce the propensity of insulin for the self–association that should eventually lead to faster absorption and

a shorter activity period. There are three fast–acting insulin analogues currently used as medication in Germany at the time of this thesis [286]:

- aspart insulin (B_P28D) [283],
- lispro insulin (B_P28K, B_K29P) [284],
- glulisine insulin (B_N3K, B_K29E) [285].

With only one or two mutations (mentioned in brackets) the insulin self–association is decreased. However, the resulting mutant has to retain its biological activity and binding affinity towards the insulin receptor. Experimentally determined residues that are important for receptor binding are A_G1, A_I2, A_V3, A_E4, A_Q5, A_Y19, A_N21, B_L6, B_G8, B_L11, B_V12, B_E13, B_Y16, B_Y17, B_G23, B_F24, B_F25 and B_Y26 [287]. Docking insulin to a three dimensional structure of the insulin receptor solved by electron microscopy, the amino acids A_E4, A_Q5, B_V12, B_Y16, B_Y17, B_F24 and B_Y26 were recognized to be in close contact with the receptor [288]. Docking analysis of altered crystal structures of insulin and the insulin receptor identified A_G1, A_I2, A_V3, A_E4, A_T8, A_Y19, A_N21, B_V12, B_B13, B_Y16, B_F24, B_F25 and B_K29 as (possible) key residues for interacting with the receptor [289].
The insulin stability and dimerization was extensively studied by Zoete et al. [169, 173, 290]. Unfortunately, the authors only considered mutations in one monomer of the dimer. Therefore, a direct comparison between the binding affinities reported using MM/PBSA [169] and Concoord/PBSA data was omitted, as only a full mutational analysis using Concoord/PBSA applied to both monomers at the same time was analyzed.

4.5.2 Results

For a Concoord/PBSA test study the direct contacts between the two insulin monomers were chosen as mutation sites. The full single–point mutation set comprising the positions B_S9, B_V12, B_E13, B_Y16, B_G20,B_E21, B_R22, B_G23, B_F24, B_F25, B_Y26, B_T27, B_P28, B_K29 and B_A30 was computed for the porcine insulin (4INS [291]) (human insulin was not available as X-ray structure in the correct conformation). The only difference to human insulin is a threonine at position 30 in chain B which is replaced by alanine in porcine insulin. Sample calculations revealed no noticeable difference between both insulin analogues and, thus, the full mutational scanning computation was applied to the unaltered porcine insulin.

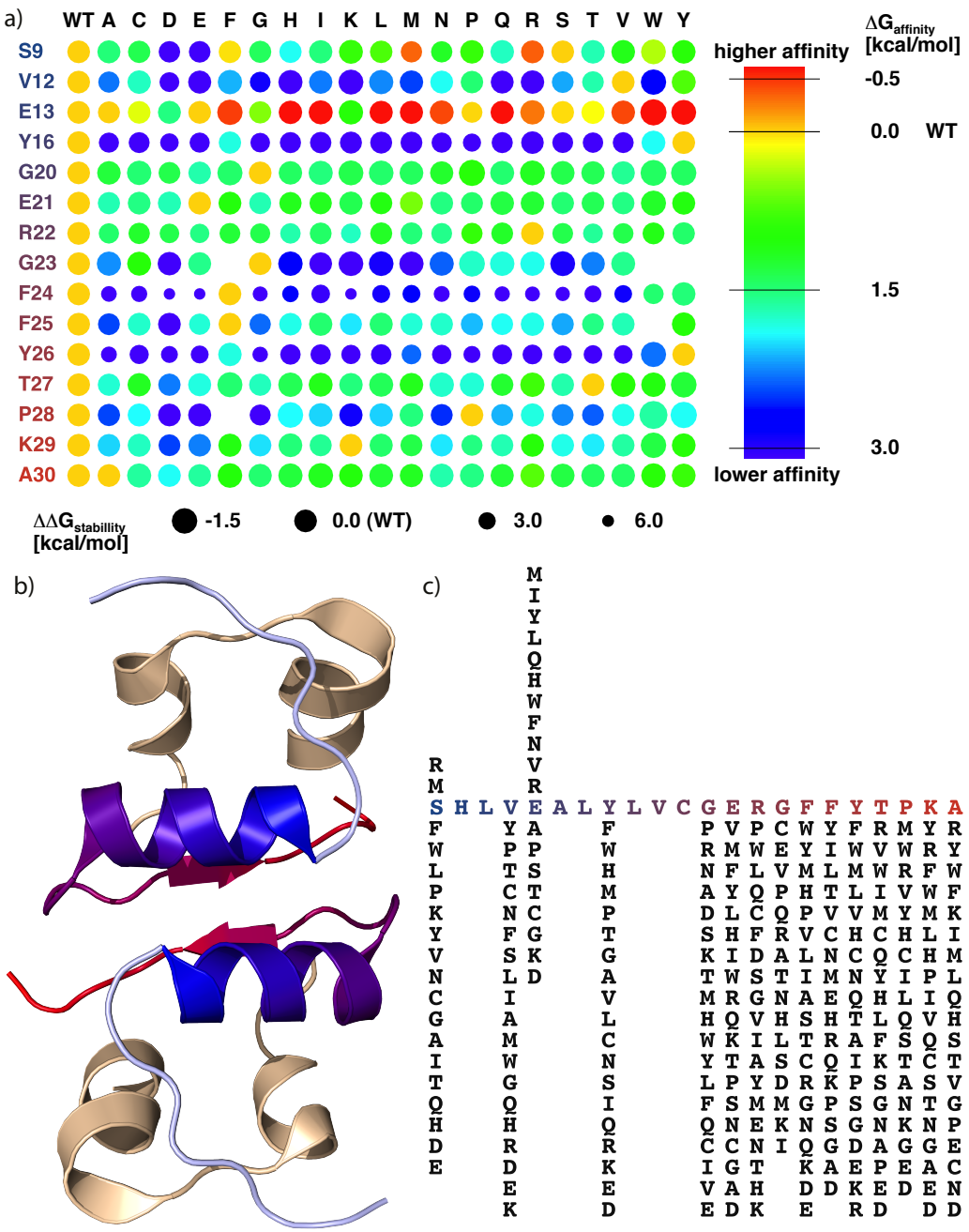

Figure 4.5: Concoord/PBSA energies for stability and dimerization analysis of insulin mutations. For explanations see text.

Results of the complete mutational scan for both stability and dimerization

analysis are depicted in Figure 4.5 a) and c). The free energies are shown in a combined representation for both folding and binding: The coloring corresponds to the computed change in binding free energies and the circles' radii are a measure for the change in folding free energies (see also legend). The mutation sites are colored from blue to red (N– to C–terminal) at every occurrence. Figure 4.5 b) shows an insulin dimer (4INS) in cartoon representation. The lower right panel lists all mutations sorted according to their effect on the binding affinity. The columns correspond to the mutations at a particular position in the sequence and are independent of each other (only single–point mutations considered). The effect of a mutation on the binding affinity is given by its height.
Modelling of the five mutations `B_G23F`, `B_G23W`, `B_G23Y`, `B_F25W` and `B_P28F` led to problems in the structure generation using Concoord. This was due to overlaps of atoms when inserting bulky amino acids. These were omitted from the analysis.

Table 4.1: *In vitro* alanine scanning results for dimerization and known analogues with single-point mutations are compared to Concoord/PBSA binding and stability free energies (in kcal/mol).

Mutation	$\Delta\Delta G^{\text{bind}}_{\text{CC/PBSA}}$	$\Delta\Delta G^{\text{fold}}_{\text{CC/PBSA}}$	occurrence *in vitro*
B_V12A	2.317	0.8270	Dimer [292]
B_Y16A	3.775	1.4402	Monomer [292]
B_F24A	3.13	3.8942	Monomer [292]
B_F25A	2.48	0.5117	Dimer [292]
B_Y26A	2.356	3.7926	Monomer [292]
B_T27A	1.797	0.2585	Dimer [292]
B_Y16H	3.324	0.9024	Monomer [293]
B_E13Q	-0.812	-0.3203	Dimer/Hexamer [284]
B_F25D	2.903	0.8648	Monomer [284]
B_V12E	10.504	-0.2038	most reduced self–association [283]
B_S9D	3.172	0.6974	*****| [283]
B_P28D	4.005	1.4015	****| [283]
B_Y26E	5.085	2.6839	***| [283]
B_V12I	2.298	-0.3551	**| [283]
B_T27E	1.847	-0.1883	least reduced self–association [283]

The outcome of Concoord/PBSA was compared to analogues with known increased or decreased self–association (Table 4.1). The first part is in good agreement with an experimental alanine scanning reported by Chen et al.

[292]. Only the prediction for the B_F25A mutant is in disagreement with experiment. However, this is probably due to the usage of porcine insulin in Concoord/PBSA: Calculations based on the human insulin (porcine insulin with mutation B_A30T) yielded an increased, destabilizing free energy of 3.44 kcal/mol for the B_Y26A mutation, while the dimerization free energy for B_F25A stayed constant (2.47 kcal/mol). Also, stabilization/destabilization of the monomer fold may have an influence on dimer formation, i.e. the dimer formation propensity may be decreased due to a decreased stability of the monomer.
The tendency of the next three insulin mutants, B_Y16H, B_E13Q and B_F25D have been estimated correctly, too. Especially the shift of the equilibrium towards the dimeric and hexameric form for the B_E13Q mutant is correctly predicted by a negative change in dimerization free energy.
Brange et al. [283] measured a decrease in self–association for several insulin analogues. Most importantly the lowest and highest affinity decreasing mutations were accurately predicted. However, the Concoord/PBSA free energy changes of three mutations show a wrong order. A repeated calculation using mutations to human insulin showed similar values, thus, method dependent fluctuations and the use of the wrong wild type can be canceled as reasons.

Table 4.2: Binding energies of medically relevant insulin analogues. The porcine insulin with the mutation B_A30T (*human mutation*) was used for the calculation. Energies in kcal/mol.

Analogue	$\Delta\Delta G_{\mathrm{CC/PBSA}}$	$\Delta\Delta G_{\mathrm{LJ}}$	$\Delta\Delta G_{\mathrm{es}}$	$\Delta\Delta G_{\mathrm{PPIS}}$
Aspart Insulin	3.92	0.0106	2.34	1.57
Glulisine Insulin	2.04	0.176	0.299	1.57
Lispro Insulin	2.71	-0.0293	1.17	1.57

Results for the medically used insulin analogues are reported in Table 4.2 showing a decrease in dimerization affinity for all three mutants.
The studies of Zoete and Meuwly [173] allow for a direct comparison between MM/PBSA and Concoord/PBSA concerning the relative folding free energies for insulin mutations. The comparison is shown in Figure 4.6 a) with their computed Fold-X values as a kind of third opinion. The three different calculations allow a comparison between the distinct methods. For a full alanine scan (except alanine, cysteine or glycine) of the insulin monomer correlation coefficients and SDECs were determined for the comparison of Concoord/PBSA with MM/PBSA ($r = 0.611$, $\sigma = 4.38$ kcal/mol), Concoord/PBSA

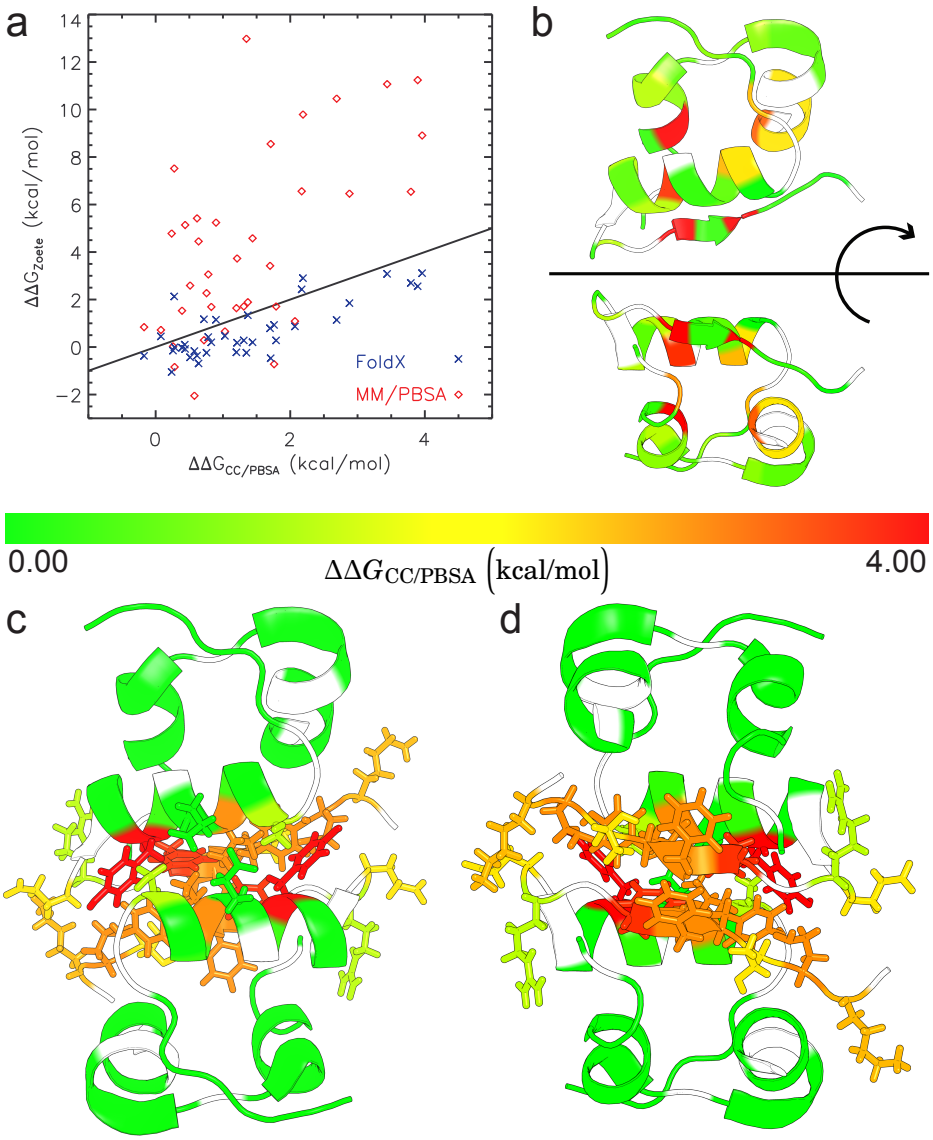

Figure 4.6: Concoord/PBSA alanine scan on insulin for probing stability and dimerization. A comparison was done between MM/PBSA (red), Fold-X (blue) and Concoord/PBSA on stability predictions (a). The loss in stability or affinity is projected on the cartoon representation of the monomer (b) and the dimer (c/d), respectively. The color–code from green (similar stability/affinity) to red (unstable) as shown in the legend is the same for both cases. The monomer is rotated around the shown axe to give a different point of view (b). The same points of view were chosen for the dimers (c/d), which also show interface residues in stick representation. Positions colored in white were not computed, as the wild type shows either already an alanine, a cysteine or a glycine.

Table 4.3: pKa corrections to insulin dimerization affinity calculations. All energies in kcal/mol. $_m$ denotes a change in protonation in the monomer and $\Delta G_{\mathrm{pK_a}}$ has been added two times as there are two monomers. AA denotes the position and amino acid which shows a different protonation state according to $\mathrm{pK_a}$ calculations with respect to pH = 7.5.

Mutation	$\Delta\Delta G_{\mathrm{CC/PBSA}}$	AA	$\mathrm{pK_a}$	$\Delta G_{\mathrm{pK_a}}$	$\Delta\Delta G^{\mathrm{pK_a}}_{\mathrm{CC/PBSA}}$
B_S9D	3.172	B_H10	8.444	-1.293	1.879
B_V12D	8.053	B_D12	9.636	-2.926	
		D_D12	9.157	-2.27	2.856
B_V12E	10.504	B_E12	8.167	-0.914	
		D_E12	10.746	-4.447	5.143
B_V12K	10.53	B_K12	6.041	-2.0	
		D_K12	1.539	-8.167	-0.373
B_R22H	1.697	D_H22	8.293	-1.086	
		$_m$B_H22	8.421	+2.524	3.134
B_G23E	1.581	B_E23	8.084	-0.8	0.781
B_F24E	6.657	B_E24	9.287	-2.448	4.209
B_F24K	3.785	B_K24	4.531	-4.068	
		D_K24	3.142	-4.183	-6.252
B_F24R	3.528	B_R24	6.38	-1.534	2.023

with Fold-X (r = 0.771, σ = 1.02 kcal/mol), and MM/PBSA with Fold-X (r = 0.496, σ = 4.91 kcal/mol). Again, the calculated Concoord/PBSA and Fold-X data show a high correlation, while the correlation to MM/PBSA data was significantly smaller for both Concoord/PBSA (0.61) and Fold-X (0.496). With the absence of experimental data, no interpretation on the accuracy of the three methods can be given. Therefore, it remains unclear which approach yields the best estimates for the insulin stability.
The alanine scanning with respect to both folding (b) and dimerization (c/d) is depicted in Figure 4.6. For affinity calculations the interface term G_{PPIS} was neglected for mutation sites not belonging to the interface (suggested by the outcome of the p53–MDM2 binding study, Section 4.3).

Additionally, $\mathrm{pK_a}$ calculations using the MCCE program [263] for the full mutational set on the dimer and on one monomer each were performed. Almost all mutants showed a standard protonation except those listed in Table 4.3. Here the correction term (4.1.5) was added. Due to large errors of 1 pK unit the corrected binding free energies have to be handled with care. A large affinity increase for the mutant `B_F24K` of more then 6 kcal/mol sug-

gests that the ΔG_{pK} contribution should be scaled similar to the Coulomb and reaction–field contributions in Concoord/PBSA. In addition, ΔG_{pK} does not account for structural rearrangements upon (de–) protonation.

From the mutational scan the mutant `B_P28` and is suggested as promising mutation target for further studies including receptor binding. The mutants `B_V12`, `B_Y16`, `B_G23`, and `B_Y26` show a decrease in binding affinity and can also be suggested for further studies. However, due to the importance to receptor binding [287–289], they may not be applicable in the treatment of diabetes mellitus. Although leading to decreased self–associations, mutations at position 24 of chain B do not seem to be suitable as they also decrease the stability of the monomer. The increased affinities of the `B_E13` mutations may hint to an important functional role of the wild type residue. Also these mutations may provide an appropriate insulin supply over night.
A combined docking and Concoord/PBSA approach could reveal implications of insulin mutations on insulin receptor binding.

Chapter 5

Conclusions and Outlook

A fast and accurate free energy calculation scheme that substitutes computationally expensive molecular dynamics simulations with a structure generation based on geometrical considerations only was proposed. It can be applied to study mutational effects on the protein folding stability, on protein–protein and on protein–ligand binding affinities. The method combines a physical effective free energy function averaged over structural ensembles with the efficient generation of conformational ensembles of the different states of the respective thermodynamic cycle.

The developed Concoord/PBSA free energy function is similar to the well established MM/PBSA [51] method combining molecular mechanics force field contributions with continuum solvent energies. Different from MM/PBSA, only electrostatic and van der Waals energies were considered, and the energetic terms were weighted by comparison to huge experimental data sets. For protein–protein binding, an additional energetic contribution taking account of cooperativity effects at interfaces was introduced.

The development of Concoord/PBSA was based on test sets of several hundred mutations. Following parameterization, the method was successfully applied for full mutational scans on different systems. For the prediction of mutational effects on folding free energies the accuracy of Concoord/PBSA is similar to other methods. For the prediction of protein–protein binding affinity changes Concoord/PBSA significantly outperforms well–established methods like Robetta or Fold-X. This improved prediction is related to the inclusion of flexibility. Especially, the mutation–dependent flexibility of Concoord/PBSA allowed to correctly predict cooperative mutational effects for the TEM1–BLIP protein complex.

The Concoord approach, however, does not cover large conformational changes as e.g. observed in MD simulations for the GYF [280] or the SH3 domain [281] exhibiting different binding modes of bound peptides. Different approaches for the prediction of conformational changes [294] as well as the structural prediction of bound peptide chains [295] using Concoord as basis have been developed recently. The tConcoord extension established by Seeliger et al. [294] estimates hydrogen-bond stabilities that lead to different

constraints enabling the prediction of larger conformational changes like biologically essential opened and closed structures of proteins. This extension has already been applied for the identification of transient binding pockets in proteins by Eyrisch and Helms [296, 297]. Another approach by Seibert [295] aims at loosening the distance constraints between the protein and a bound peptide while preserving experimentally known anchor positions. By introducing *constraint learning* using several crystal structures of peptide chains bound to the same protein important interactions are preserved. Here, a conformational prediction of any sequence with given length is possible. Next to these predicted structures Concoord/PBSA is a complementary method that enables an energetic estimate of the sampled configurations. Despite the limitations of Concoord for the sampling of significantly different conformations a good agreement with experiment was achieved for a large mutational scan of peptides bound to the GYF domain.
For biomolecular systems that are sensitive to small changes in the pH an extension of Concoord/PBSA encompassing extensive pK_a calculations was suggested. It is conceivable to combine the Concoord/PBSA conformational prediction with pK_a calculating schemes for an improved prediction of protonation states. While the MCCE method [263] considers side chain rotamers only, a Concoord–based prediction would additionally take the backbone flexibility into account. Although this approach would be computationally more expensive than conventional pK_a calculations, it could prove efficient since the sampled structures and part of the calculated energies are needed anyway for free energy estimates. For performance reasons, Poisson-Boltzmann solutions may be substituted by the considerably faster Generalized Born calculations.

The primary goal for the near future is the application of Concoord/PBSA on the immunologically relevant Major Histocompatibility Complexes (MHC). With the development of the Concoord/PBSA method presented in this thesis and the enhancements reported by Becker [30] and Seibert [295] for binding affinity and binding mode prediction a prediction of the binding strength of more than 10,000 peptides appears feasible. This data set is not sufficient to cover the sequence space of 9^{20} possibilities in the case of MHC bound peptides (that are typically nine amino acids long), but instead it can be used to substitute frequently missing experimental data in order to train machine learning algorithms that can predict the affinity for some orders of magnitude more peptide sequences. This combination of a reliable structure–based prediction with a statistical method will prove as a valuable tool for the design of peptide vaccines in the case of MHC proteins and more generally in the design of proteins.

Being limited on peptide compounds only, the inclusion of solvent molecules, ions and other non-peptide ligands will enlarge the applicability of Concoord/PBSA. With the combined fast structure generation and energy evaluation it also opens up the lane towards a full flexible docking technique.
With the Concoord/PBSA web interface being publicly available, stability and binding affinity predictions using the reported Concoord/PBSA procedure may easily be performed by the interested reader. The respective conformational Concoord/PBSA ensembles are also accessible for further studies.

Chapter 6

Author Contributions

Rainer A. Böckmann designed research. Alexander Benedix developed the Concoord/PBSA method, set up the test set, performed simulations and analysis. The adoption of Concoord/PBSA to binding affinity and the web interface was developed in co–operation with Caroline Becker in the context of her Master's thesis [30], advised by Alexander Benedix. Simulations for the comparison of binding affinity between Concoord/PBSA and Fold-X were performed by Caroline Becker.

Appendix A

Protein Stability Results

Table A.1: Concoord/PBSA results for stability calculations. Experimental and calculated differences in folding free energies (in kcal/mol) relative to the wild type (both experimental and calculated) are shown. Also single contributions of the computation are presented. * denotes that the experimental and calculated free energies are taken relative to a pseudo wild type.

PDB	Mutation	$\Delta\Delta G_{\text{exp}}$	$\Delta\Delta G_{\text{calc}}$		$\Delta\Delta G_{\text{RF}}$	$\Delta\Delta G_{\text{Coul}}$	$\Delta\Delta G_{\text{LJ}}$	$\Delta\Delta G_{\text{SA}}$	$-T\Delta\Delta S$
1AYI	A13G	0.60	0.5601	±0.2553	0.1007	0.0442	0.1754	0.1488	0.0911
1AYI	A28G	0.19	0.7808	±0.2563	-2.1241	1.1374	0.6986	0.8327	0.2362
1AYI	A77G	1.27	0.7952	±0.2532	-0.3600	0.1481	0.3659	0.3988	0.2424
1AYI	A78G	1.31	0.6711	±0.2451	0.3473	-0.1665	0.1557	0.0169	0.3177
1AYI	F15A	3.63	5.2159	±0.2335	-1.1149	0.8079	2.6032	3.2166	-0.2970
1AYI	F41L	1.89	3.5927	±0.2260	-1.6453	1.2238	1.6611	2.2866	0.0665
1AYI	I22V	2.13	1.7274	±0.2201	-1.9938	1.6072	1.0373	1.1593	-0.0827
1AYI	I44V	0.53	0.7856	±0.2208	-0.8839	0.4800	0.6172	0.5622	0.0101
1AYI	I54V	2.63	1.5377	±0.2226	0.0560	-0.1799	0.7512	0.8876	0.0228
1AYI	I68V	0.60	1.5933	±0.2318	-0.6934	0.3681	0.9892	1.0288	-0.0994
1AYI	I72V	0.41	0.6623	±0.2281	-1.5447	1.2690	0.5399	0.4120	-0.0139
1AYI	I7V	1.46	1.6385	±0.2227	-1.5114	1.1684	1.1061	0.9641	-0.0888
1AYI	L18A	3.01	3.1499	±0.2271	-1.0577	0.3183	1.7862	2.3529	-0.2498
1AYI	L19A	3.39	4.0922	±0.2442	-2.1251	2.0346	2.1057	2.6135	-0.5366
1AYI	L34A	1.84	2.5004	±0.2342	-2.5352	1.8163	1.6684	1.9637	-0.4129
1AYI	L37A	2.84	3.5325	±0.2315	-1.4392	1.1593	1.8233	2.3595	-0.3704
1AYI	L38A	2.68	2.7843	±0.2272	-2.1132	1.6817	1.5828	1.9007	-0.2677
1AYI	L3A	0.74	2.6208	±0.2238	-2.1330	1.8226	1.5041	1.8662	-0.4391
1AYI	L53A	3.25	2.9949	±0.2340	-0.4211	-0.0970	1.5761	2.2153	-0.2784
1AYI	T51S	0.96	1.2195	±0.2208	1.2429	-0.8859	0.3420	0.4038	0.1167
1AYI	V16A	1.48	1.3472	±0.2287	-1.3197	1.0826	0.6848	0.8159	0.0835
1AYI	V27A	-0.50	0.5258	±0.2315	-1.1799	1.0039	0.3757	0.2327	0.0934
1AYI	V33A	0.24	0.4607	±0.2248	0.0536	-0.4417	0.4104	0.3701	0.0682
1AYI	V36A	0.10	1.0071	±0.2282	-1.3101	0.9411	0.6262	0.5666	0.1833
1AYI	V42A	0.69	1.1035	±0.2276	-2.1862	1.8430	0.6759	0.7272	0.0436
1AYI	V69A	0.69	2.5279	±0.2305	0.5616	-0.3882	0.9603	1.3753	0.0190
1HZ6	A13P	-0.10	-0.1675	±0.1470	-0.0568	0.2051	-0.2003	-0.2934	0.1779
1HZ6	A13V	0.83	-0.4249	±0.1444	-0.1227	0.3005	-0.3010	-0.3088	0.0072
1HZ6	A20G	2.17	1.9766	±0.1852	-0.8456	0.8085	0.7725	1.2346	0.0066
1HZ6	A20V	-1.47	-0.8600	±0.1470	-0.3769	0.3439	-0.2107	-0.6048	-0.0115
1HZ6	A29G	2.54	1.3253	±0.1816	-0.0541	0.1571	0.4726	0.8127	-0.0630
1HZ6	A33G	3.10	1.8246	±0.1817	-0.0114	0.0559	0.7397	0.9144	0.1260
1HZ6	A35G	1.32	0.6115	±0.1777	-0.1323	0.1967	0.2171	0.2615	0.0686
1HZ6	A37G	3.12	1.8632	±0.1862	-0.1205	0.1746	0.7737	1.0839	-0.0484

PDB	Mutation	$\Delta\Delta G_{\text{exp}}$	$\Delta\Delta G_{\text{calc}}$		$\Delta\Delta G_{\text{RF}}$	$\Delta\Delta G_{\text{Coul}}$	$\Delta\Delta G_{\text{LJ}}$	$\Delta\Delta G_{\text{SA}}$	$-T\Delta\Delta S$
1HZ6	A52G	0.49	0.6651	±0.1819	-0.0872	-0.0795	0.3456	0.4545	0.0317
1HZ6	A8G	2.43	1.6962	±0.1795	-0.0597	0.0055	0.6436	0.9595	0.1473
1HZ6	D38A	1.21	-0.0952	±0.1472	0.9933	-1.7402	0.6426	0.4048	-0.3957
1HZ6	D38G	2.14	0.5166	±0.1694	0.8314	-1.5389	0.7424	0.7382	-0.2566
1HZ6	D50A	0.20	-0.0297	±0.1455	3.0007	-3.3558	0.3681	0.3048	-0.3475
1HZ6	E21A	0.59	0.5404	±0.1484	-5.7066	5.3348	0.6591	0.6927	-0.4395
1HZ6	E32G	1.19	0.9244	±0.1727	-3.3186	2.3335	1.0327	1.1869	-0.3101
1HZ6	E32I	1.08	-1.1836	±0.1346	-3.4161	2.2837	0.1817	0.0572	-0.2901
1HZ6	E46A	0.23	-0.1624	±0.1468	-0.7381	0.2225	0.5007	0.3591	-0.5067
1HZ6	F12A	3.12	3.8833	±0.1454	-1.2154	1.0719	2.2375	2.1992	-0.4100
1HZ6	F12L	0.68	2.2094	±0.1352	-0.6396	0.7497	1.0005	1.1623	-0.0635
1HZ6	F22A	4.25	4.6338	±0.1590	-0.3584	0.3827	2.3547	2.7607	-0.5059
1HZ6	F22L	3.12	2.7270	±0.1464	0.0309	0.1092	1.1821	1.4738	-0.0690
1HZ6	F26G	3.08	2.5028	±0.1762	-1.3187	1.2992	1.2805	1.5663	-0.3245
1HZ6	F26L	0.38	1.7796	±0.1378	-0.5183	0.6398	0.6764	0.9112	0.0705
1HZ6	F62L	3.34	3.4836	±0.1492	-0.6449	0.7240	1.6761	1.8708	-0.1423
1HZ6	F62V	3.73	6.0822	±0.1445	-0.5184	0.7187	3.0106	3.1729	-0.3015
1HZ6	G15A	1.52	-0.3397	±0.1796	-0.1995	-0.1001	0.1571	-0.2355	0.0384
1HZ6	G15V	2.53	-0.0044	±0.1699	0.0543	-0.3389	0.2787	-0.1439	0.1453
1HZ6	G24A	2.08	-0.9119	±0.1837	0.0127	-0.0098	-0.2792	-0.6147	-0.0208
1HZ6	G45A	2.23	-0.5297	±0.1831	-0.0180	-0.0930	-0.0287	-0.3437	-0.0462
1HZ6	G55A	2.04	-0.2295	±0.1831	-0.1894	0.1417	0.0178	-0.1561	-0.0435
1HZ6	I11A	1.37	1.8571	±0.1442	-0.5683	0.3520	1.1682	0.9612	-0.0560
1HZ6	I11V	0.47	0.3975	±0.1308	-0.3787	0.1804	0.4483	0.2176	-0.0700
1HZ6	I60A	4.72	3.9042	±0.1462	-0.1717	0.2549	1.9396	2.0936	-0.2121
1HZ6	I60V	1.69	1.6095	±0.1349	-0.2867	0.0481	1.0343	1.0375	-0.2238
1HZ6	I6A	4.90	4.1407	±0.1493	-0.8583	0.6503	2.1135	2.5337	-0.2985
1HZ6	I6V	0.56	1.2084	±0.1353	0.2734	-0.4428	0.7364	0.7446	-0.1032
1HZ6	K23A	0.88	0.2723	±0.1446	-15.4784	14.9159	0.7143	0.6067	-0.4862
1HZ6	K28G	-0.16	1.2971	±0.1697	-18.2562	17.8394	0.9205	1.0042	-0.2109
1HZ6	K41A	-0.58	1.5894	±0.1484	-23.8279	23.7892	1.0411	1.3644	-0.7774
1HZ6	K42A	-0.35	0.4221	±0.1439	-12.9648	13.2957	0.2351	0.1002	-0.2440
1HZ6	K54A	0.09	0.3662	±0.1449	-11.1482	11.3491	0.2398	0.3957	-0.4702
1HZ6	K61A	0.45	0.6771	±0.1433	-18.9566	17.8887	1.0550	1.1747	-0.4846
1HZ6	K7A	0.92	1.8398	±0.1414	-11.6481	11.8167	0.9740	1.2984	-0.6011
1HZ6	L10A	3.12	3.3422	±0.1423	-0.3246	0.5151	1.6332	1.8789	-0.3603
1HZ6	L40A	2.44	1.9611	±0.1430	-0.7887	0.4541	1.3480	1.3049	-0.3571
1HZ6	L58A	3.77	3.6404	±0.1439	-1.1118	1.0642	1.8470	2.2668	-0.4258
1HZ6	N14A	1.78	0.5149	±0.1479	-1.0652	1.2635	0.2105	0.2915	-0.1854
1HZ6	N44A	0.34	0.8270	±0.1428	-0.9788	0.6197	0.6807	0.7673	-0.2619
1HZ6	N59A	1.73	0.9768	±0.1466	-2.6702	2.5380	0.6210	0.8483	-0.3602
1HZ6	N9A	1.87	1.1695	±0.1438	-2.1320	1.7570	0.7912	1.0402	-0.2870
1HZ6	S16A	0.30	0.2733	±0.1444	-0.1138	-0.1612	0.3040	0.2854	-0.0411
1HZ6	S31A	-0.41	0.2857	±0.1419	0.0222	0.1023	0.1614	0.1466	-0.1468
1HZ6	S31G	0.82	0.7510	±0.1761	0.0266	0.1830	0.2708	0.3272	-0.0566
1HZ6	T17A	1.17	0.5087	±0.1413	-0.4447	0.2518	0.3316	0.3816	-0.0116
1HZ6	T19A	1.11	0.5899	±0.1492	0.0587	-0.4029	0.4844	0.5990	-0.1493
1HZ6	T25A	1.25	0.6667	±0.1455	-0.3058	-0.0014	0.4849	0.5151	-0.0262
1HZ6	T30A	1.09	0.8650	±0.1464	-0.6745	-0.1653	0.7987	0.8619	0.0443
1HZ6	T39G	0.17	0.5338	±0.1699	-0.1459	-0.1148	0.4909	0.3178	-0.0142
1HZ6	T48A	0.97	0.3118	±0.1437	-0.4021	-0.0006	0.5296	0.2187	-0.0338
1HZ6	T57A	1.83	0.9248	±0.1448	-0.4459	0.1342	0.6153	0.7636	-0.1425
1HZ6	T5A	1.63	0.7914	±0.1404	-0.7461	0.4503	0.6025	0.5536	-0.0689
1HZ6	V49A	0.92	1.2834	±0.1452	-0.0017	-0.0047	0.6317	0.7509	-0.0927
1HZ6	V4A	1.22	1.3426	±0.1445	-0.2647	0.1658	0.5960	0.8128	0.0327

PDB	Mutation	$\Delta\Delta G_{\text{exp}}$	$\Delta\Delta G_{\text{calc}}$		$\Delta\Delta G_{\text{RF}}$	$\Delta\Delta G_{\text{Coul}}$	$\Delta\Delta G_{\text{LJ}}$	$\Delta\Delta G_{\text{SA}}$	$-T\Delta\Delta S$
1HZ6	V51A	1.14	1.4277	±0.1465	-0.3181	0.4315	0.6088	0.7101	-0.0046
1HZ6	Y34A	2.82	4.1066	±0.2140	-0.8121	0.7679	2.1667	2.4927	-0.5086
1HZ6	Y36A	2.46	2.2851	±0.2117	-0.8959	0.8023	1.2214	1.3986	-0.2412
1HZ6	Y56A	1.66	2.2630	±0.2170	-0.9055	1.1427	1.1525	1.2870	-0.4137
1HZ6	Y56L	-0.43	1.0794	±0.2070	-0.7064	0.8066	0.4270	0.5515	0.0007
1PGA	A20G	2.39	1.2928	±0.1973	-0.2647	0.3917	0.3845	0.7199	0.0615
1PGA	A26G	2.96	1.1477	±0.1996	0.3069	0.0538	0.1832	0.5308	0.0731
1PGA	A34G	2.48	0.9462	±0.1965	-0.1419	0.2174	0.3288	0.4736	0.0685
1PGA	D22A	1.75	0.6876	±0.1646	-2.4317	2.5629	0.4086	0.5000	-0.3522
1PGA	D46A	1.74	1.0053	±0.1742	-1.0807	0.9005	0.8387	0.8581	-0.5114
1PGA	D47A	-0.49	0.2058	±0.1702	-3.3969	3.4021	0.2080	0.4282	-0.4356
1PGA	E15A	0.47	0.6919	±0.1694	-2.5704	2.3584	0.5198	0.7449	-0.3608
1PGA	F30L	1.42	3.2195	±0.1756	-0.6007	0.6431	1.3480	2.0304	-0.2013
1PGA	F52L	3.54	2.4051	±0.1615	-1.0620	1.1255	1.0702	1.2460	0.0254
1PGA	G41A	2.84	0.2048	±0.1933	-0.0500	-0.1669	0.3426	0.0689	0.0102
1PGA	I6A	2.09	1.1464	±0.1683	-0.2032	0.0413	0.8441	0.6042	-0.1401
1PGA	K28G	0.05	0.7607	±0.1921	-18.4931	18.4774	0.3974	0.7170	-0.3379
1PGA	K31G	2.02	1.9553	±0.1949	-21.7903	20.8941	1.6453	1.6566	-0.4505
1PGA	L7A	1.85	2.5673	±0.1664	-0.3608	0.4667	1.4664	1.4125	-0.4175
1PGA	N35G	2.50	0.4387	±0.1868	-0.8234	0.7603	0.3138	0.2130	-0.0250
1PGA	N37A	-0.17	0.4446	±0.1689	-1.5450	1.5063	0.2832	0.3965	-0.1964
1PGA	Q32G	1.00	0.8596	±0.1941	-1.4854	1.5801	0.4717	0.4317	-0.1384
1PGA	T11A	0.60	-0.3199	±0.1677	-0.5766	0.3568	0.0651	-0.0934	-0.0718
1PGA	T16A	0.38	0.9497	±0.1762	-0.8243	0.2519	0.7288	0.8229	-0.0296
1PGA	T18A	0.46	1.1855	±0.1725	-0.4932	0.1472	0.5332	1.0713	-0.0730
1PGA	T25A	-0.22	0.4244	±0.1652	-1.0224	0.8409	0.2614	0.4998	-0.1553
1PGA	T49A	0.72	1.1610	±0.1708	-2.0040	2.2167	0.4744	0.6735	-0.1996
1PGA	T51A	1.87	1.1826	±0.1715	-0.8600	0.6659	0.5970	0.8718	-0.0921
1PGA	T53A	1.91	0.5922	±0.1646	-0.3023	-0.2811	0.5969	0.5964	-0.0176
1PGA	V29A	0.70	0.7745	±0.1688	-0.2865	0.2876	0.3509	0.4160	0.0065
1PGA	V39A	1.72	1.7804	±0.1634	-1.0123	0.8355	0.9524	1.0786	-0.0738
1PGA	V54A	2.93	1.5742	±0.1705	-0.4462	0.3730	0.6820	0.9450	0.0204
1PGA	Y33A	0.92	2.2149	±0.2282	-1.0146	0.9977	1.4269	1.2732	-0.4682
1PGA	Y3L	1.62	2.9239	±0.2220	-1.9023	1.8883	1.4221	1.5621	-0.0463
1PGA	Y45L	3.34	2.3901	±0.2210	-2.3604	2.3793	1.1403	1.2569	-0.0260
1STN	A102G	1.30	1.5013	±0.3281	-0.5213	0.7949	0.4365	0.8157	-0.0244
1STN	A109G	1.00	0.4421	±0.3220	-0.2000	0.1425	0.1241	0.3235	0.0519
1STN	A112G	0.00	0.7729	±0.3273	0.0250	0.0699	0.3007	0.3795	-0.0022
1STN	A12G	2.40	0.9205	±0.3111	0.2246	-0.0611	0.2789	0.3479	0.1303
1STN	A130G	1.10	0.5677	±0.3188	-0.2209	0.2202	0.1743	0.2632	0.1309
1STN	A132G	3.70	1.6953	±0.3140	-0.5367	0.5651	0.6948	0.8474	0.1248
1STN	A17G	0.30	1.1174	±0.3240	-0.3059	0.5200	0.2785	0.6650	-0.0403
1STN	A58G	2.60	1.6058	±0.3278	0.3349	-0.1076	0.5648	0.7551	0.0587
1STN	A60G	1.40	0.5957	±0.3173	-0.2958	0.2480	0.2504	0.3138	0.0792
1STN	A69G	2.00	0.7216	±0.3221	0.4736	-0.1305	0.0908	0.3724	-0.0848
1STN	A90G	2.00	1.1758	±0.3124	0.2089	-0.1092	0.4389	0.4641	0.1730
1STN	A94G	2.40	1.6317	±0.3159	0.2320	0.2101	0.3797	0.8289	-0.0190
1STN	D19A	0.10	-0.3230	±0.2968	-26.2933	24.3017	1.0900	1.1984	-0.6198
1STN	D19G	0.50	0.2376	±0.3092	-26.5625	24.2793	1.2381	1.6400	-0.3572
1STN	D21A	-0.70	-0.2863	±0.2963	-23.2145	22.5008	0.3975	0.6109	-0.5811
1STN	D21G	-0.30	-0.2530	±0.3102	-23.6295	22.1987	0.6312	0.9819	-0.4352
1STN	D40A	-0.20	0.0106	±0.3051	-21.9956	21.7449	0.3703	0.3175	-0.4267
1STN	D40G	0.50	0.7988	±0.3057	-21.4423	21.3189	0.5540	0.6389	-0.2707
1STN	D77A	3.10	0.6014	±0.3190	-30.3757	31.2967	0.0717	0.3117	-0.7030
1STN	D77G	2.20	0.6310	±0.3298	-29.3614	29.8876	0.0911	0.5044	-0.4906

PDB	Mutation	$\Delta\Delta G_{\text{exp}}$	$\Delta\Delta G_{\text{calc}}$		$\Delta\Delta G_{\text{RF}}$	$\Delta\Delta G_{\text{Coul}}$	$\Delta\Delta G_{\text{LJ}}$	$\Delta\Delta G_{\text{SA}}$	$-T\Delta\Delta S$
1STN	D83A	3.80	2.5975	±0.3052	-30.4607	31.5486	1.0548	0.8850	-0.4301
1STN	D83G	2.70	2.7752	±0.3124	-31.7084	32.3364	1.3171	1.2457	-0.4157
1STN	D95A	3.30	0.9063	±0.3006	-28.0378	27.9876	0.8244	0.7541	-0.6219
1STN	D95G	2.70	1.5888	±0.3261	-28.0152	28.2600	0.7701	1.1004	-0.5266
1STN	E101A	1.90	1.4550	±0.2978	-23.8362	24.3302	0.5709	0.8046	-0.4146
1STN	E101G	3.10	2.6518	±0.3209	-24.0678	24.6816	0.9664	1.4982	-0.4266
1STN	E10A	1.30	0.6155	±0.2950	-29.4380	28.3161	0.9343	1.3257	-0.5225
1STN	E10G	1.80	1.5407	±0.3112	-29.0970	28.0329	1.2690	1.8127	-0.4768
1STN	E122A	0.40	1.1536	±0.2975	-26.2732	26.2230	0.7085	1.0796	-0.5844
1STN	E122G	2.20	1.7557	±0.3125	-26.1469	25.9943	0.9599	1.3769	-0.4286
1STN	E129A	0.40	2.3108	±0.3042	-28.9364	29.1813	1.1204	1.2929	-0.3474
1STN	E129G	2.20	2.7916	±0.3131	-28.5309	28.7857	1.2306	1.7127	-0.4065
1STN	E135A	0.70	0.9366	±0.3062	-25.6084	25.3472	0.6456	1.0861	-0.5339
1STN	E135G	1.70	1.8052	±0.3150	-25.3369	25.2288	0.9347	1.3979	-0.4193
1STN	E43A	-0.30	-0.4952	±0.2797	-20.7366	18.7243	1.0737	0.7047	-0.2613
1STN	E43G	-0.50	0.0777	±0.3012	-20.9342	18.7028	1.3883	1.1763	-0.2556
1STN	E52A	0.10	1.0293	±0.3028	-24.3149	23.5314	1.2965	1.2495	-0.7333
1STN	E52G	0.40	1.8576	±0.3141	-22.9614	22.4046	1.4366	1.5329	-0.5550
1STN	E57A	0.20	0.8297	±0.3020	-19.9217	20.0254	0.5690	0.5641	-0.4069
1STN	E57G	1.60	1.5130	±0.3196	-19.7466	20.2497	0.5760	0.7146	-0.2808
1STN	E67A	1.00	0.7028	±0.3006	-27.7204	27.6817	0.4892	0.7151	-0.4628
1STN	E67G	0.90	1.5461	±0.3124	-27.7952	28.0611	0.7459	1.0746	-0.5403
1STN	E73A	1.20	1.1026	±0.3060	-25.5534	24.8580	1.2021	1.1393	-0.5433
1STN	E73G	2.70	2.1511	±0.3196	-25.6342	25.0728	1.4545	1.6152	-0.3573
1STN	E75A	2.20	0.9688	±0.2981	-30.3462	28.8361	1.4525	1.6023	-0.5757
1STN	E75G	3.50	2.4666	±0.3262	-30.3258	29.3302	1.7823	2.2629	-0.5829
1STN	F34A	3.70	5.3888	±0.2991	-0.8624	1.1624	2.7929	2.8480	-0.5521
1STN	F61A	2.30	3.5398	±0.3006	-0.8591	1.0692	1.8211	2.0856	-0.5770
1STN	F61G	4.80	4.5079	±0.3016	-0.1626	0.6683	2.0755	2.3077	-0.3810
1STN	F76A	4.00	3.9248	±0.2949	-2.1878	1.6568	2.3871	2.5972	-0.5285
1STN	F76G	4.70	5.0978	±0.3074	-2.5148	2.0486	2.7928	3.2433	-0.4721
1STN	H121A	3.10	2.3434	±0.3119	-2.1560	1.8755	1.4425	1.7208	-0.5394
1STN	H121G	4.20	3.6445	±0.3299	-2.6064	2.2922	1.9491	2.4946	-0.4850
1STN	H124A	-0.40	1.6527	±0.3066	-1.9150	1.9765	0.9080	1.1464	-0.4632
1STN	H124G	0.50	2.4835	±0.3228	-1.1050	1.6109	0.9301	1.3854	-0.3379
1STN	H46A	0.50	1.4512	±0.2964	-2.1923	2.6222	0.9086	0.6303	-0.5176
1STN	H46G	0.40	2.1265	±0.3072	-3.2347	3.2294	1.4062	1.3625	-0.6369
1STN	H8A	0.40	0.3093	±0.3071	-1.0297	0.5409	0.6223	0.6474	-0.4717
1STN	H8G	0.80	0.8769	±0.3196	-1.0606	0.6735	0.8169	0.8658	-0.4187
1STN	I139A	3.50	2.3786	±0.3067	1.8742	-1.3990	0.9331	1.2988	-0.3285
1STN	I139G	4.40	3.1952	±0.3161	1.5130	-1.1617	1.2573	1.7452	-0.1586
1STN	I139V	1.50	0.7465	±0.3003	0.6896	-0.7918	0.5219	0.5502	-0.2234
1STN	I15A	2.70	1.3739	±0.2947	-0.9962	0.5396	1.1060	0.8769	-0.1524
1STN	I15G	3.30	3.4662	±0.3171	-0.4888	0.7388	1.6096	1.6827	-0.0760
1STN	I15V	0.80	0.6489	±0.2896	-0.0104	-0.0543	0.5292	0.2681	-0.0837
1STN	I18A	2.50	1.8509	±0.3037	-0.7693	0.8631	0.9524	0.9668	-0.1620
1STN	I18G	2.50	2.5124	±0.3143	-0.8870	0.9647	1.1719	1.3777	-0.1149
1STN	I18V	1.10	0.8976	±0.2979	-0.1916	0.1579	0.5418	0.5490	-0.1596
1STN	I72A	5.10	2.9746	±0.3097	-0.2773	0.3084	1.4220	1.7289	-0.2073
1STN	I72V	1.80	0.8877	±0.3037	-0.1768	0.0261	0.5875	0.5197	-0.0687
1STN	I92A	4.00	3.0453	±0.3044	0.3607	0.0949	1.1367	1.6631	-0.2101
1STN	I92V	0.50	0.5995	±0.2975	-0.1527	0.2525	0.3208	0.3880	-0.2092
1STN	K110A	1.30	0.9215	±0.2974	9.6488	-9.3489	0.5508	0.6251	-0.5542
1STN	K110G	2.70	2.0141	±0.3127	9.1903	-8.9055	0.9970	1.1936	-0.4613
1STN	K116A	-0.70	-0.8388	±0.3102	11.4308	-12.0254	0.0960	0.1807	-0.5210

PDB	Mutation	$\Delta\Delta G_{\text{exp}}$	$\Delta\Delta G_{\text{calc}}$		$\Delta\Delta G_{\text{RF}}$	$\Delta\Delta G_{\text{Coul}}$	$\Delta\Delta G_{\text{LJ}}$	$\Delta\Delta G_{\text{SA}}$	$-T\Delta\Delta S$
1STN	K116G	-1.00	-0.6213	±0.3258	11.2648	-11.7244	-0.0943	0.3116	-0.3789
1STN	K127A	-0.20	0.1590	±0.3152	11.7477	-11.9632	0.4145	0.3541	-0.3941
1STN	K127G	0.70	0.5503	±0.3131	11.8935	-11.9526	0.5199	0.3991	-0.3097
1STN	K133A	1.40	-0.0392	±0.3091	5.7398	-7.0522	0.7524	1.0562	-0.5354
1STN	K133G	3.30	0.6233	±0.3166	5.6631	-6.7778	0.8959	1.3092	-0.4671
1STN	K134A	-0.10	0.3531	±0.2981	6.2081	-6.4285	0.3521	0.6665	-0.4451
1STN	K134G	0.70	1.0108	±0.3102	7.3147	-7.0523	0.4139	0.6386	-0.3041
1STN	K136A	0.90	0.2869	±0.2987	5.1239	-5.1528	0.3908	0.2437	-0.3187
1STN	K136G	0.20	0.8347	±0.3191	5.1193	-4.7924	0.3254	0.5426	-0.3602
1STN	K16A	0.20	-0.5397	±0.2991	11.6700	-13.0584	0.5929	0.6768	-0.4210
1STN	K16G	0.70	0.0558	±0.3171	12.6190	-13.6365	0.6477	0.8000	-0.3745
1STN	K24A	0.20	1.0034	±0.2975	12.6060	-13.2086	1.1835	1.0263	-0.6037
1STN	K24G	1.20	2.2131	±0.3123	13.3991	-13.4035	1.3376	1.3766	-0.4967
1STN	K28A	0.70	-0.0767	±0.3017	5.7800	-6.5042	0.6009	0.6335	-0.5868
1STN	K28G	0.70	1.0379	±0.3059	5.8173	-6.1328	0.9065	0.9800	-0.5331
1STN	K45A	-0.30	-0.2155	±0.2922	7.3391	-8.3085	0.6994	0.7329	-0.6784
1STN	K45G	-0.20	0.4938	±0.3116	7.4318	-8.4091	0.9452	0.9642	-0.4382
1STN	K48A	-0.10	-0.1357	±0.2995	9.3867	-9.3990	0.1592	0.0027	-0.2854
1STN	K48G	-0.20	0.4301	±0.3158	9.6613	-9.5248	0.2844	0.1780	-0.1688
1STN	K49A	0.30	0.1018	±0.3062	5.5292	-5.8542	0.5157	0.5748	-0.6638
1STN	K49G	0.20	1.5598	±0.3225	5.6596	-5.4317	0.8447	1.0259	-0.5388
1STN	K53A	0.20	-0.1809	±0.2942	2.7480	-3.3022	0.2787	0.4557	-0.3612
1STN	K53G	0.30	0.3650	±0.3214	3.2742	-3.4853	0.4529	0.5340	-0.4108
1STN	K63A	0.50	0.0391	±0.3051	4.8028	-5.4630	0.6133	0.6601	-0.5741
1STN	K63G	1.50	0.9272	±0.3143	4.8553	-5.4446	0.9010	1.0449	-0.4293
1STN	K64A	-0.10	0.0394	±0.3018	7.2724	-7.3782	0.3453	0.2777	-0.4778
1STN	K64G	0.40	0.8612	±0.3168	7.2267	-7.0294	0.4951	0.5397	-0.3709
1STN	K70A	0.10	-0.2989	±0.3015	9.8046	-9.8046	0.1931	0.0409	-0.5329
1STN	K70G	0.50	0.4143	±0.3114	9.4285	-9.2030	0.3782	0.2468	-0.4362
1STN	K71A	0.40	0.2465	±0.3042	10.3521	-10.1245	0.3320	0.1723	-0.4854
1STN	K71G	1.10	1.2321	±0.3165	10.2666	-9.8169	0.6797	0.4725	-0.3698
1STN	K78A	0.60	0.6331	±0.3077	11.9663	-12.8597	1.0082	1.1589	-0.6407
1STN	K78G	1.10	0.9267	±0.3228	12.7764	-13.3501	0.9061	1.0490	-0.4547
1STN	K84A	-0.20	0.5424	±0.3088	9.8495	-10.3713	0.5431	0.9272	-0.4061
1STN	K84G	0.30	0.2938	±0.3184	9.9647	-10.4056	0.4670	0.6089	-0.3413
1STN	K97A	0.10	-0.0738	±0.3008	9.7106	-9.7721	0.4595	0.2106	-0.6825
1STN	K97G	1.70	1.5065	±0.3175	9.5331	-9.0734	0.8271	0.8043	-0.5844
1STN	K9A	1.40	0.6607	±0.2939	5.0377	-6.1310	1.1119	1.1374	-0.4953
1STN	K9G	1.90	1.4596	±0.3120	5.5774	-6.2323	1.1145	1.3751	-0.3750
1STN	L103A	4.60	4.3931	±0.3071	-0.4545	0.5924	2.2416	2.5737	-0.5601
1STN	L108A	5.80	2.7486	±0.3074	0.0922	0.0985	1.4616	1.5580	-0.4616
1STN	L125A	4.90	3.1211	±0.2981	-0.6993	0.5365	1.6724	2.0716	-0.4602
1STN	L137A	2.30	1.4284	±0.3024	-0.4526	0.1671	1.1217	0.9984	-0.4062
1STN	L137G	4.60	3.0542	±0.3123	-0.3677	0.5735	1.5874	1.5705	-0.3095
1STN	L14A	2.30	2.2228	±0.3012	-0.0147	0.0257	1.3245	1.3895	-0.5022
1STN	L14G	3.70	3.1842	±0.3124	-0.4292	0.4329	1.6019	1.9109	-0.3323
1STN	L25A	2.70	2.8423	±0.3003	-0.3262	0.6344	1.3823	1.6088	-0.4569
1STN	L25G	4.50	4.1893	±0.3185	-0.2839	0.7731	1.8238	2.2112	-0.3348
1STN	L36A	3.50	1.5066	±0.3021	0.2894	-0.5186	0.8215	1.3541	-0.4398
1STN	L36G	5.30	3.2976	±0.3179	0.3727	-0.2088	1.3272	2.1541	-0.3477
1STN	L37A	1.70	1.8877	±0.2997	-0.1365	-0.3202	1.3558	1.3231	-0.3344
1STN	L37G	3.80	3.2679	±0.3163	-0.8232	0.1857	2.1082	2.2073	-0.4102
1STN	L38A	1.70	2.6967	±0.3077	-0.0752	-0.0250	1.4921	1.7249	-0.4202
1STN	L38G	0.60	3.5184	±0.3100	-0.7579	0.6358	1.7830	2.2530	-0.3955
1STN	L7A	1.60	1.0227	±0.2906	-0.2662	-0.2944	0.9087	1.0652	-0.3906

PDB	Mutation	$\Delta\Delta G_{\mathrm{exp}}$	$\Delta\Delta G_{\mathrm{calc}}$		$\Delta\Delta G_{\mathrm{RF}}$	$\Delta\Delta G_{\mathrm{Coul}}$	$\Delta\Delta G_{\mathrm{LJ}}$	$\Delta\Delta G_{\mathrm{SA}}$	$-T\Delta\Delta S$
1STN	L7G	1.50	1.7933	±0.3054	-0.4642	-0.2722	1.1825	1.6166	-0.2694
1STN	L89A	2.60	2.3422	±0.3008	-0.6101	0.3164	1.6439	1.5142	-0.5223
1STN	L89G	3.20	3.4725	±0.3136	-0.1929	0.0483	1.9104	1.9862	-0.2794
1STN	M26A	1.50	1.4533	±0.3024	-0.8873	0.8831	1.1896	0.8904	-0.6226
1STN	M26G	2.20	2.3474	±0.3152	-0.0559	0.4944	1.2939	1.1072	-0.4922
1STN	M32A	1.70	1.7683	±0.2960	-0.4270	0.3334	1.2081	1.1518	-0.4981
1STN	M32G	2.40	3.3217	±0.3147	-0.5955	0.7052	1.8378	1.9457	-0.5716
1STN	M65A	2.00	2.6022	±0.2954	0.0213	-0.0808	1.6714	1.8419	-0.8516
1STN	M65G	4.60	3.6078	±0.3163	0.2228	0.0577	1.9296	2.0932	-0.6955
1STN	M98A	4.60	2.8607	±0.2929	-0.7370	0.3928	1.8699	1.9025	-0.5675
1STN	M98G	4.50	3.8393	±0.3157	-0.0894	-0.1329	2.1441	2.3150	-0.3976
1STN	N100A	5.20	1.6906	±0.2994	-1.2930	1.6802	0.6031	1.0681	-0.3678
1STN	N100G	5.10	2.2059	±0.3159	-1.4714	1.5724	0.8428	1.5644	-0.3023
1STN	N118A	2.10	1.1494	±0.2969	-2.6708	2.0465	1.0168	1.1116	-0.3547
1STN	N118D	2.40	4.5357	±0.2813	20.1100	-16.5478	0.0926	0.7617	0.1193
1STN	N118G	1.90	2.0195	±0.3243	-2.5338	2.0129	1.2716	1.6184	-0.3496
1STN	N119A	1.30	0.5361	±0.2959	-0.6713	0.6852	0.2511	0.6583	-0.3873
1STN	N119G	1.30	0.7852	±0.3110	-0.5619	0.0515	0.4405	1.1155	-0.2603
1STN	N138A	1.10	1.0841	±0.3001	-0.8758	0.9472	0.5930	0.7833	-0.3636
1STN	N138G	-0.10	1.2543	±0.3255	-1.4432	1.3099	0.7527	0.9442	-0.3094
1STN	N68A	0.50	0.4247	±0.2980	-2.0262	2.2055	0.2718	0.3128	-0.3392
1STN	N68G	0.50	0.5282	±0.3111	-1.4423	1.6138	0.2797	0.2767	-0.1996
1STN	P117A	-0.80	0.1780	±0.3020	-0.5441	0.4732	0.1084	0.1558	-0.0153
1STN	P117G	-0.90	0.5060	±0.3165	-0.4360	0.5834	-0.0317	0.3707	0.0196
1STN	P11A	0.40	0.1457	±0.2966	-0.8895	0.8226	0.3027	0.0183	-0.1084
1STN	P11G	1.00	0.3442	±0.3057	-0.9281	1.0373	0.2401	-0.0352	0.0301
1STN	P31A	0.50	-0.0964	±0.2943	-0.9344	0.5756	0.3506	0.0494	-0.1376
1STN	P31G	1.60	0.5901	±0.3201	-0.2303	0.3608	0.2715	0.1593	0.0288
1STN	P42A	-0.10	0.7782	±0.2995	-0.8941	0.4941	0.6388	0.6679	-0.1284
1STN	P42G	0.40	1.4920	±0.3073	-1.1770	1.0732	0.8308	0.8432	-0.0782
1STN	P47A	0.60	0.1017	±0.2976	-0.3234	0.3290	0.2232	-0.0299	-0.0974
1STN	P47G	0.10	0.4926	±0.3108	-1.3305	0.9439	0.5641	0.2642	0.0510
1STN	P56A	0.00	0.6074	±0.2979	-0.7857	0.7148	0.3622	0.4410	-0.1249
1STN	P56G	1.00	0.7761	±0.3192	-1.0196	0.8039	0.4894	0.5565	-0.0541
1STN	Q106A	-0.10	1.0579	±0.3023	-0.4712	0.6932	0.5704	0.7402	-0.4748
1STN	Q106G	1.50	2.2114	±0.3103	-0.1596	0.6656	0.9998	1.0962	-0.3905
1STN	Q123A	0.40	0.0845	±0.3024	0.0989	-0.0649	0.3001	0.2140	-0.4636
1STN	Q123G	0.60	0.3652	±0.3187	-0.1096	0.2658	0.3599	0.3015	-0.4523
1STN	Q131A	0.20	-0.0975	±0.2984	-1.3330	1.0159	0.2495	0.4542	-0.4840
1STN	Q131G	2.40	1.6414	±0.3167	-1.7049	1.7006	0.7698	1.2532	-0.3773
1STN	Q30A	0.30	0.5882	±0.3105	-0.5161	0.5549	0.4552	0.5790	-0.4848
1STN	Q30G	0.90	0.7660	±0.3075	0.3815	-0.1588	0.5221	0.4212	-0.4000
1STN	Q80A	0.10	1.3076	±0.2981	-2.3253	2.3466	0.8506	1.0038	-0.5682
1STN	Q80G	1.40	1.5938	±0.3064	-2.2310	2.1762	1.0028	1.0377	-0.3917
1STN	R105A	1.40	1.5608	±0.2943	3.0525	-3.4123	1.1891	1.7097	-0.9782
1STN	R105G	2.40	2.3779	±0.3182	2.7362	-3.2029	1.5798	2.1384	-0.8735
1STN	R126A	1.70	1.7153	±0.3062	6.3429	-6.8060	1.3391	1.6163	-0.7769
1STN	R126G	2.90	2.5957	±0.3185	6.8775	-6.9289	1.4507	1.9196	-0.7232
1STN	R35A	1.40	1.6823	±0.3020	3.7206	-4.9258	1.6401	2.0226	-0.7752
1STN	R35G	2.20	2.0984	±0.2993	4.3915	-5.4154	1.7101	2.1511	-0.7389
1STN	R81A	1.10	0.7332	±0.2855	9.4355	-10.6499	1.3853	1.0747	-0.5124
1STN	R81G	2.20	1.9497	±0.3109	9.0455	-10.0041	1.6605	1.6176	-0.3698
1STN	R87A	0.90	2.0129	±0.3020	6.9026	-7.7137	1.5213	2.0346	-0.7318
1STN	R87G	2.60	3.1915	±0.3187	7.6571	-8.0079	1.7236	2.4967	-0.6780
1STN	S128A	-0.70	0.2127	±0.3041	-0.5739	1.1922	-0.2584	0.0379	-0.1852

PDB	Mutation	$\Delta\Delta G_{\text{exp}}$	$\Delta\Delta G_{\text{calc}}$		$\Delta\Delta G_{\text{RF}}$	$\Delta\Delta G_{\text{Coul}}$	$\Delta\Delta G_{\text{LJ}}$	$\Delta\Delta G_{\text{SA}}$	$-T\Delta\Delta S$
1STN	S128G	1.60	1.2335	±0.3036	-0.4693	1.0666	0.1954	0.4988	-0.0580
1STN	S141A	0.40	0.0506	±0.3006	0.4031	-0.0915	-0.0939	-0.0799	-0.0871
1STN	S141G	0.90	1.0126	±0.3140	-0.7829	0.7959	0.4091	0.6856	-0.0951
1STN	S59A	-0.40	0.2343	±0.2847	-0.8333	0.4591	0.3817	0.3826	-0.1558
1STN	S59G	1.10	0.7055	±0.3133	-1.0764	0.8460	0.4831	0.4587	-0.0057
1STN	T120A	1.20	0.7220	±0.3179	-1.7079	2.2783	-0.0079	0.5212	-0.3617
1STN	T120G	2.10	1.3556	±0.3301	-1.0999	1.4080	0.2229	0.9131	-0.0886
1STN	T120V	1.80	0.4789	±0.2932	-1.3242	1.9636	-0.2243	0.0726	-0.0088
1STN	T13A	0.70	-0.0193	±0.2965	-0.3414	0.0519	0.2762	0.1281	-0.1340
1STN	T13G	1.10	0.6306	±0.3119	-0.4097	0.4186	0.3879	0.3033	-0.0695
1STN	T13V	0.40	-0.4753	±0.2924	0.4051	-0.6676	0.0172	-0.1651	-0.0648
1STN	T22A	1.60	0.7769	±0.3150	-1.5995	1.3371	0.5201	0.8057	-0.2866
1STN	T22G	2.40	1.9623	±0.3223	-2.0068	1.6027	1.0039	1.5153	-0.1527
1STN	T22V	0.90	-0.4324	±0.3005	-0.7659	1.0221	-0.2802	-0.2277	-0.1807
1STN	T33A	1.40	0.7675	±0.2944	-1.0931	0.8396	0.4831	0.6468	-0.1090
1STN	T33G	2.50	1.5041	±0.3121	-0.8219	0.7991	0.6030	0.8859	0.0380
1STN	T33V	-0.40	-0.4242	±0.2985	-0.5079	0.5018	-0.0945	-0.1170	-0.2066
1STN	T41A	0.00	0.0042	±0.3028	-0.1129	-0.4487	0.1020	0.5622	-0.0985
1STN	T41G	2.00	1.2875	±0.3117	-0.3168	-0.5304	0.8198	1.3464	-0.0315
1STN	T41V	-0.80	-0.4324	±0.3019	0.0431	-0.8062	0.2369	0.2210	-0.1273
1STN	T44A	0.40	0.4870	±0.2992	-1.0430	0.5822	0.6124	0.5727	-0.2373
1STN	T44G	0.60	0.9638	±0.3197	-1.7226	1.0613	0.8528	1.0000	-0.2277
1STN	T44V	-0.10	0.2750	±0.2943	-0.7305	1.1512	-0.0469	-0.0297	-0.0692
1STN	T62A	2.40	2.0847	±0.3037	-0.5277	0.6334	0.9482	1.1755	-0.1447
1STN	T62G	3.50	3.1250	±0.3240	-0.0614	0.2112	1.4482	1.8121	-0.2851
1STN	T62V	0.20	-0.2184	±0.2861	-0.5408	0.1562	0.1190	0.1535	-0.1063
1STN	T82A	0.90	0.5994	±0.3033	-0.0079	0.2023	0.3275	0.2519	-0.1745
1STN	T82G	2.00	1.1567	±0.3125	0.3046	0.3297	0.3246	0.3215	-0.1238
1STN	T82V	-0.20	-0.0653	±0.2985	-0.5064	0.3060	0.0849	0.0910	-0.0409
1STN	V104A	2.90	1.1620	±0.3002	2.8348	-1.6831	-0.2568	0.1091	0.1581
1STN	V104T	2.50	1.0310	±0.2891	0.6801	0.3303	-0.0441	0.1562	-0.0915
1STN	V111A	4.20	1.2623	±0.3033	0.1295	-0.2885	0.8479	0.6710	-0.0975
1STN	V111G	4.90	3.1587	±0.3149	-0.1399	0.1131	1.4898	1.7837	-0.0880
1STN	V111T	2.30	0.8025	±0.2886	0.5967	-0.6186	0.4348	0.3749	0.0147
1STN	V114A	0.00	0.9366	±0.2928	0.4369	-0.3679	0.4638	0.5261	-0.1222
1STN	V114G	0.20	1.7955	±0.3046	1.1858	-0.9023	0.5295	1.0957	-0.1131
1STN	V114T	0.30	-0.0885	±0.2946	1.0764	-1.0950	0.1201	-0.0587	-0.1313
1STN	V23A	2.90	2.6043	±0.2986	-0.2558	0.2570	1.1513	1.5194	-0.0676
1STN	V23G	5.60	3.9772	±0.3181	-1.0338	1.0276	1.8498	2.2213	-0.0876
1STN	V23T	3.20	0.7316	±0.2935	1.0175	-0.7237	0.2628	0.1195	0.0555
1STN	V39A	2.20	1.4265	±0.3031	0.2196	-0.1129	0.6831	0.8308	-0.1941
1STN	V39G	4.70	3.6131	±0.3214	0.6215	-0.0163	1.2058	1.9180	-0.1159
1STN	V39T	1.30	0.7639	±0.2887	1.1493	-0.2978	-0.1190	0.0429	-0.0116
1STN	V51A	0.30	0.8479	±0.3064	-0.3336	0.3558	0.4950	0.3389	-0.0082
1STN	V51G	0.40	1.2378	±0.3127	-0.2073	0.5432	0.4415	0.4761	-0.0157
1STN	V51T	-0.20	0.2434	±0.2919	-0.4014	0.4969	0.0950	0.0433	0.0096
1STN	V66A	2.20	2.4276	±0.3058	-0.2868	0.4504	1.1309	1.4027	-0.2696
1STN	V66G	4.40	3.7745	±0.3171	-0.4106	0.6801	1.5049	2.0545	-0.0543
1STN	V66T	1.40	0.7255	±0.2922	0.7047	-0.0745	0.0198	0.1200	-0.0445
1STN	V74A	3.10	1.5303	±0.3071	1.2013	-0.6315	0.3032	0.7195	-0.0621
1STN	V74T	3.80	0.3745	±0.2913	0.6778	0.4095	-0.4247	-0.3038	0.0158
1STN	V99A	3.20	3.0577	±0.3011	-0.7166	0.8963	1.3821	1.7902	-0.2943
1STN	V99G	5.00	4.1522	±0.3221	-0.0660	0.3771	1.7616	2.3291	-0.2495
1STN	V99T	3.30	1.1498	±0.2944	0.0626	0.5585	0.2900	0.3520	-0.1133
1STN	Y113A	0.00	1.4595	±0.3404	-0.6002	1.2327	0.6656	0.6494	-0.4880

PDB	Mutation	$\Delta\Delta G_{\text{exp}}$	$\Delta\Delta G_{\text{calc}}$		$\Delta\Delta G_{\text{RF}}$	$\Delta\Delta G_{\text{Coul}}$	$\Delta\Delta G_{\text{LJ}}$	$\Delta\Delta G_{\text{SA}}$	$-T\Delta\Delta S$
1STN	Y113F	0.00	0.4047	±0.3344	0.0743	-0.1009	0.3863	0.3351	-0.2901
1STN	Y113G	0.30	1.9153	±0.3516	-0.3495	1.1760	0.6836	0.7742	-0.3689
1STN	Y113L	-0.20	0.5078	±0.3306	0.4658	0.2162	0.0569	-0.0795	-0.1517
1STN	Y115A	0.30	1.3440	±0.3349	-0.0595	0.6423	0.7562	0.4657	-0.4607
1STN	Y115F	0.10	0.5968	±0.3339	-0.5913	0.5812	0.4671	0.2778	-0.1380
1STN	Y115G	0.70	2.2419	±0.3518	-0.2223	0.6956	1.1457	0.9909	-0.3681
1STN	Y115L	0.30	0.6395	±0.3378	0.1439	0.3586	0.1070	0.1493	-0.1193
1STN	Y27A	2.80	3.2876	±0.3351	-1.7758	1.5343	1.9953	2.2307	-0.6970
1STN	Y27F	0.60	-0.1096	±0.3381	-1.2456	0.4417	0.3749	0.4825	-0.1630
1STN	Y27G	5.10	5.0879	±0.3573	-1.9411	1.9953	2.5184	3.1789	-0.6636
1STN	Y27L	1.50	0.9493	±0.3342	-1.0637	0.4801	0.8472	0.7952	-0.1095
1STN	Y54A	2.20	2.5804	±0.3351	-2.0564	1.6834	1.7646	1.8960	-0.7073
1STN	Y54F	0.50	-0.2829	±0.3402	-0.7245	0.4812	-0.0245	0.0186	-0.0338
1STN	Y54G	1.90	3.2809	±0.3479	-1.1359	0.5751	1.8423	2.5577	-0.5582
1STN	Y54L	3.40	1.0415	±0.3311	-2.0855	1.7220	0.8535	0.7831	-0.2316
1STN	Y85A	0.40	1.3363	±0.3445	-1.9932	1.7530	0.8058	1.1640	-0.3933
1STN	Y85F	0.00	0.1793	±0.3369	-2.7739	1.9688	0.4054	0.7786	-0.1995
1STN	Y85G	1.00	1.9862	±0.3498	-1.8889	1.7673	1.0503	1.3716	-0.3140
1STN	Y85L	0.10	0.5694	±0.3294	-1.3162	1.5246	0.1146	0.2842	-0.0378
1STN	Y91A	5.30	5.8830	±0.3396	-2.2727	2.4729	2.9380	3.3826	-0.6379
1STN	Y91F	2.40	1.3753	±0.3427	-0.3629	1.2265	0.3848	0.3115	-0.1845
1STN	Y91L	3.90	3.6663	±0.3309	-2.5424	2.9937	1.6424	1.8027	-0.2301
1STN	Y93F	2.00	-0.4976	±0.3363	-0.0346	-0.5540	0.1737	0.0020	-0.0848
1STN	Y93G	7.50	5.2864	±0.3461	-1.7187	0.7603	3.2401	3.3975	-0.3929
1STN	Y93L	4.50	2.0697	±0.3339	-1.4929	1.0409	1.3813	1.3289	-0.1885
1YPC	A16G	1.09	1.2403	±0.2307	0.4285	-0.1703	0.3401	0.6109	0.0311
1YPC	A58G	1.88	1.1118	±0.2284	-0.0263	0.1706	0.3199	0.5145	0.1330
1YPC	D23A	0.96	0.5965	±0.2060	-8.4956	8.0910	0.6877	0.6922	-0.3789
1YPC	D45A	0.80	0.8935	±0.2061	-13.7761	13.6185	0.6428	0.8960	-0.4878
1YPC	D52A	3.41	0.3808	±0.2061	-6.2983	5.9803	0.6156	0.4618	-0.3786
1YPC	E14D	0.52	0.1497	±0.2010	-2.9044	2.7208	0.1746	0.2593	-0.1007
1YPC	E14N	0.70	-0.2674	±0.1981	-5.7170	5.3267	0.0396	0.1669	-0.0836
1YPC	E14Q	0.29	-0.3253	±0.1972	-6.5794	6.4307	-0.1593	-0.1303	0.1129
1YPC	E15D	0.74	0.5800	±0.1968	-2.0380	1.9967	0.3157	0.3465	-0.0408
1YPC	E15N	1.07	0.8665	±0.1932	-7.4612	6.9863	0.7368	0.6341	-0.0296
1YPC	E15Q	0.47	0.3592	±0.1968	-7.8167	7.7602	0.1109	0.3345	-0.0298
1YPC	E26A	0.32	0.6901	±0.1913	-2.0388	3.2741	-0.3737	-0.3606	0.1891
1YPC	E41A	0.70	0.1743	±0.2025	-14.8276	14.0037	0.5808	0.8596	-0.4423
1YPC	E7A	0.47	0.6013	±0.2029	-10.7589	9.4409	1.2150	1.1155	-0.4113
1YPC	E7Q	0.62	0.3861	±0.1980	-9.2459	8.0435	0.8874	0.9216	-0.2205
1YPC	F50A	3.84	3.5671	±0.2018	-0.5165	0.6075	1.8865	2.0846	-0.4950
1YPC	F50L	2.11	1.1937	±0.1922	0.4516	-0.2017	0.2613	0.5586	0.1240
1YPC	F50V	2.39	2.3602	±0.1854	-0.5586	0.7860	1.0670	1.4496	-0.3837
1YPC	I20V	1.30	1.5113	±0.1959	0.4794	-0.3292	0.7306	0.7910	-0.1606
1YPC	I29A	3.90	2.4516	±0.2000	0.5641	-0.2568	0.9039	1.3512	-0.1108
1YPC	I29V	1.11	0.7030	±0.2005	0.3455	-0.3459	0.4463	0.3399	-0.0828
1YPC	I30A	2.12	1.1280	±0.2101	0.3445	-0.3446	0.4946	0.7653	-0.1318
1YPC	I30G	3.52	2.3319	±0.2197	0.3770	-0.1084	0.7924	1.3204	-0.0495
1YPC	I30T	1.34	0.7181	±0.2010	0.1000	0.4595	0.0732	0.1697	-0.0843
1YPC	I30V	-0.08	0.0380	±0.2002	0.0511	-0.1106	0.0896	0.1567	-0.1487
1YPC	I37A	0.03	0.1403	±0.2039	0.3827	-0.2235	0.1221	-0.2365	0.0955
1YPC	I57A	4.29	4.2241	±0.2125	-0.1962	0.5183	1.7200	2.3444	-0.1624
1YPC	I57V	-0.19	1.0475	±0.2017	-0.1721	0.1328	0.5009	0.5554	0.0304
1YPC	K11A	-0.42	1.1565	±0.2029	-11.8812	12.0504	0.7534	0.7190	-0.4851
1YPC	K17A	0.49	2.1406	±0.2071	-12.6573	12.6137	1.2494	1.5653	-0.6304

PDB	Mutation	$\Delta\Delta G_{\text{exp}}$	$\Delta\Delta G_{\text{calc}}$		$\Delta\Delta G_{\text{RF}}$	$\Delta\Delta G_{\text{Coul}}$	$\Delta\Delta G_{\text{LJ}}$	$\Delta\Delta G_{\text{SA}}$	$-T\Delta\Delta S$
1YPC	K17G	2.32	3.1426	±0.2238	-12.6028	12.7046	1.5615	2.0722	-0.5929
1YPC	K18A	-0.21	0.2072	±0.1977	-5.6350	5.9773	0.1686	-0.0369	-0.2668
1YPC	K18G	0.99	0.7773	±0.2183	-5.4867	6.0104	0.3268	0.0795	-0.1527
1YPC	K24A	0.65	2.3588	±0.2060	-9.8790	9.1569	1.5920	2.1131	-0.6242
1YPC	K24G	3.19	3.3129	±0.2197	-9.5684	9.0081	1.9114	2.4876	-0.5258
1YPC	K2A	0.55	2.0867	±0.2054	-15.3363	14.9584	1.3445	1.6399	-0.5200
1YPC	K2M	0.67	0.9960	±0.1997	-14.9989	15.5176	-0.0534	0.5195	0.0113
1YPC	L21A	1.33	1.7644	±0.2109	-0.1406	0.2497	1.1047	1.0333	-0.4827
1YPC	L21G	1.38	2.3293	±0.2228	0.2172	0.1816	1.1622	1.2102	-0.4419
1YPC	L32A	2.37	2.4447	±0.2007	-0.4825	0.6756	1.1480	1.5456	-0.4421
1YPC	L32I	0.26	-0.1857	±0.1903	0.8183	-0.3815	-0.1932	-0.3274	-0.1019
1YPC	L32V	0.50	0.3627	±0.1854	0.2133	0.0635	0.2486	0.1383	-0.3010
1YPC	L49A	3.84	2.8431	±0.2061	0.2499	-0.0615	1.1252	1.8472	-0.3176
1YPC	L8A	2.68	1.9316	±0.2014	0.5701	-0.3564	0.8745	1.0538	-0.2104
1YPC	N56A	0.83	0.9080	±0.2063	-0.8584	0.6149	0.6948	0.6923	-0.2356
1YPC	N56D	1.21	1.1854	±0.1950	-7.2129	8.4501	-0.2180	0.0010	0.1652
1YPC	P25A	1.76	0.2720	±0.2017	-0.5476	0.5815	0.2359	-0.0198	0.0220
1YPC	P33A	0.17	0.7477	±0.2078	-0.6362	0.7952	0.3940	0.1995	-0.0048
1YPC	P61A	3.34	2.0440	±0.2043	0.0723	0.2993	0.8860	0.7380	0.0484
1YPC	P6A	1.57	0.4341	±0.1939	0.6977	-0.4116	0.1798	-0.0318	-0.0000
1YPC	Q22A	0.02	0.0171	±0.2090	0.0638	0.1173	0.0737	0.0815	-0.3192
1YPC	Q22G	0.60	0.6081	±0.2272	0.0624	0.2004	0.3366	0.2791	-0.2704
1YPC	R43A	0.58	-0.0192	±0.2043	-8.2402	7.6463	0.5273	0.6464	-0.5989
1YPC	S12A	0.89	0.7659	±0.2003	-1.6847	1.7977	0.3527	0.4208	-0.1206
1YPC	S12G	0.80	1.1538	±0.2205	-1.9383	1.9975	0.5072	0.6465	-0.0591
1YPC	T36A	-0.23	0.6401	±0.2007	-0.3233	0.2284	0.1882	0.3603	0.1866
1YPC	T36S	0.02	0.3450	±0.1913	0.0703	0.2032	-0.1481	-0.1263	0.3460
1YPC	T36V	0.76	-0.4830	±0.1945	-0.0448	-0.3456	-0.0805	-0.0250	0.0131
1YPC	T39A	0.72	0.7860	±0.2042	-0.9800	1.1623	0.2246	0.3885	-0.0094
1YPC	T39D	-0.02	0.2278	±0.1950	1.6136	-1.8685	0.1504	0.1430	0.1893
1YPC	T3A	0.85	0.9336	±0.2022	-0.1627	0.4289	0.3458	0.3676	-0.0460
1YPC	T3G	1.16	1.3482	±0.2255	-0.1011	0.5805	0.3329	0.4382	0.0977
1YPC	T3V	0.32	0.1331	±0.1908	-0.1343	0.4544	-0.1380	-0.1574	0.1083
1YPC	V19A	0.49	0.6258	±0.2049	-0.4082	-0.0386	0.6639	0.6313	-0.2228
1YPC	V34A	0.64	1.1104	±0.2019	0.2060	-0.0258	0.4172	0.6010	-0.0880
1YPC	V34G	2.43	1.8569	±0.2218	0.7133	-0.3379	0.5466	0.8930	0.0419
1YPC	V34T	1.03	0.2301	±0.1862	1.0971	-0.2066	-0.5772	-0.3410	0.2578
1YPC	V38A	1.47	1.5283	±0.2001	-0.6213	0.7162	0.5695	0.8870	-0.0230
1YPC	V47A	4.93	3.4354	±0.2055	-0.3759	0.7360	1.4114	1.7651	-0.1013
1YPC	V51A	1.98	1.4299	±0.2081	0.2914	-0.0967	0.5699	0.7676	-0.1022
1YPC	V60A	1.51	1.4186	±0.2023	0.1746	-0.1213	0.6402	0.7277	-0.0026
1YPC	V60G	3.24	2.4654	±0.2197	-0.2603	0.3708	0.9501	1.3596	0.0451
1YPC	V60T	0.38	0.2593	±0.1912	0.1901	0.0795	-0.0932	-0.1435	0.2265
1YPC	V63A	1.45	1.3707	±0.2082	0.2495	-0.1267	0.6051	0.6549	-0.0122
1YPC	V63G	3.50	2.6512	±0.2198	-0.4164	0.1437	1.1537	1.6921	0.0780
1YPC	V63T	1.15	0.8996	±0.1982	0.4821	-0.0127	0.1408	0.1986	0.0908
2LZM	E11A	-1.10	1.7152	±0.3584	-18.0360	17.7298	0.8522	1.4931	-0.3240
2LZM	E128A	0.16	0.7588	±0.3686	-22.4258	22.5745	0.5121	0.5065	-0.4086
2LZM	I3A	0.70	2.8336	±0.3714	-0.1031	0.3016	1.2255	1.5548	-0.1453
2LZM	I3G	2.10	3.7282	±0.3709	0.3379	0.2060	1.3812	1.8447	-0.0415
2LZM	I3T	2.30	1.0609	±0.3567	0.8662	-0.2569	0.2863	0.1999	-0.0347
2LZM	I3V	0.40	1.7426	±0.3636	-0.3124	0.6587	0.7314	0.7130	-0.0481
2LZM	K124G	0.10	-0.8264	±0.3742	4.0245	-6.2053	0.8276	0.9636	-0.4368
2LZM	L133A	3.60	3.5470	±0.3754	0.0838	0.0432	1.8469	2.0013	-0.4282
2LZM	N116D	-0.60	-0.2713	±0.3459	21.4055	-21.7141	0.0261	-0.2210	0.2322

PDB	Mutation	$\Delta\Delta G_{\mathrm{exp}}$	$\Delta\Delta G_{\mathrm{calc}}$		$\Delta\Delta G_{\mathrm{RF}}$	$\Delta\Delta G_{\mathrm{Coul}}$	$\Delta\Delta G_{\mathrm{LJ}}$	$\Delta\Delta G_{\mathrm{SA}}$	$-T\Delta\Delta S$
2LZM	N144D	-0.50	-1.0599	±0.3638	18.3443	-18.7243	-0.3021	-0.4406	0.0628
2LZM	N55G	0.60	0.7410	±0.3723	-1.7963	2.2431	0.0651	0.3880	-0.1589
2LZM	P37A	0.00	0.3662	±0.3637	-0.5165	0.7338	0.1004	0.1134	-0.0649
2LZM_pW	D20A	0.30	0.0143	±0.3760	-18.3642	18.5995	-0.0150	0.0669	-0.2729
2LZM_pW	D47A	0.95	0.2427	±0.3695	-21.1741	20.1132	0.7186	0.8644	-0.2793
2LZM_pW	D92N	1.40	-0.6302	±0.3863	-26.1846	26.0719	-0.1333	-0.2149	-0.1694
2LZM_pW	E45A	-0.01	0.7393	±0.3913	-18.7688	19.0294	0.3347	0.4995	-0.3555
2LZM_pW	F153A	3.80	4.7497	±0.3851	-0.4937	0.6420	2.4228	2.5375	-0.3589
2LZM_pW	F153L	-0.30	2.0754	±0.3751	-0.2152	0.3984	1.0017	0.8944	-0.0038
2LZM_pW	F67A	1.90	4.1671	±0.4031	-0.5669	0.7109	2.0105	2.3950	-0.3825
2LZM_pW	I100A	3.40	2.7283	±0.3770	-0.1956	0.3929	1.0574	1.5052	-0.0316
2LZM_pW	I17A	2.70	2.8361	±0.3788	0.5077	-0.2110	1.2202	1.3648	-0.0456
2LZM_pW	I27A	3.10	1.5650	±0.3959	0.8058	-0.9398	0.7858	1.0365	-0.1232
2LZM_pW	I29A	2.60	1.9056	±0.3805	-0.0554	-0.1791	1.0096	1.3035	-0.1730
2LZM_pW	I50A	2.00	1.8815	±0.3768	-0.2804	0.4837	0.7113	1.0482	-0.0813
2LZM_pW	I58A	3.20	2.5577	±0.3814	0.3032	-0.2265	1.2653	1.3401	-0.1244
2LZM_pW	I78A	1.60	2.1050	±0.3915	0.5966	-0.3818	1.0841	0.9553	-0.1492
2LZM_pW	K43A	1.03	0.8240	±0.3782	4.9887	-5.9596	0.9357	1.4177	-0.5586
2LZM_pW	K48A	0.56	-1.0669	±0.3805	-3.1758	2.5862	-0.0682	0.0277	-0.4369
2LZM_pW	L118A	3.50	2.5925	±0.3804	-0.2382	0.6490	1.1195	1.3018	-0.2397
2LZM_pW	L121A	2.30	3.1791	±0.3820	-0.2176	0.2766	1.7228	1.6908	-0.2935
2LZM_pW	L33A	3.60	3.7743	±0.3923	0.1607	0.0642	1.8963	1.9727	-0.3196
2LZM_pW	L39A	0.90	0.6282	±0.3857	0.0951	0.1327	0.4421	0.2351	-0.2768
2LZM_pW	L46A	1.86	2.2193	±0.3973	-0.1106	0.1248	1.1284	1.5875	-0.5107
2LZM_pW	L66A	3.90	2.0730	±0.3830	0.0201	-0.0270	1.1684	1.3024	-0.3910
2LZM_pW	L7A	2.60	2.2460	±0.3810	0.1949	-0.6337	1.4595	1.6089	-0.3836
2LZM_pW	L84A	3.90	3.3378	±0.3917	-0.3001	0.5352	1.5394	1.8805	-0.3173
2LZM_pW	L91A	3.10	2.8153	±0.3902	-0.2833	0.3113	1.3672	1.6616	-0.2415
2LZM_pW	L99A	4.50	3.7784	±0.3827	-0.0125	0.2190	1.7623	2.1142	-0.3046
2LZM_pW	L99G	6.30	5.3065	±0.3786	-0.0914	0.3768	2.3473	2.9269	-0.2531
2LZM_pW	M106A	2.30	2.1443	±0.3920	-0.4873	0.4350	1.2023	1.3531	-0.3588
2LZM_pW	M120A	0.20	1.2386	±0.3847	-0.1547	-0.5951	1.4820	1.0753	-0.5688
2LZM_pW	M6A	1.90	2.5890	±0.3969	-1.2993	0.4073	2.0546	1.9564	-0.5299
2LZM_pW	N116A	-0.17	-0.6469	±0.3838	-1.0628	1.3542	-0.2914	-0.4281	-0.2187
2LZM_pW	N163D	0.21	-1.0151	±0.3908	11.2315	-11.7342	-0.3249	-0.3586	0.1711
2LZM_pW	N40A	-0.32	-0.5934	±0.3856	-0.8664	1.2323	-0.3858	-0.3979	-0.1755
2LZM_pW	N40D	-0.44	-0.5018	±0.3881	16.5730	-16.6910	-0.2385	-0.3540	0.2087
2LZM_pW	N68A	-0.05	0.4464	±0.3868	-0.5166	0.2869	0.4475	0.4300	-0.2014
2LZM_pW	Q122A	0.24	0.8014	±0.3921	-0.0972	-0.3749	0.7602	0.8079	-0.2945
2LZM_pW	Q123A	0.22	-0.3968	±0.3844	-0.4491	-0.3304	0.2565	0.5231	-0.3970
2LZM_pW	R119A	0.18	-0.0615	±0.3885	11.5935	-12.3886	0.6251	0.6813	-0.5728
2LZM_pW	S117A	-1.27	-0.4807	±0.3760	-0.9039	0.2057	0.1947	0.0550	-0.0323
2LZM_pW	S44A	-0.34	-0.1581	±0.3836	0.4219	-0.1990	-0.1245	-0.1658	-0.0906
2LZM_pW	S44G	0.53	0.1874	±0.4000	0.4750	-0.2384	0.0612	-0.0789	-0.0315
2LZM_pW	T115A	0.14	-0.5169	±0.3862	0.2747	-0.3520	-0.0893	-0.3871	0.0368
2LZM_pW	T151S	-0.39	-0.1861	±0.3714	-0.0303	-0.3146	-0.0512	0.1323	0.0776
2LZM_pW	T152S	2.60	0.2912	±0.3709	0.2094	-0.4149	0.1998	0.2426	0.0542
2LZM_pW	T26S	-0.57	0.0512	±0.3817	0.1894	-0.7130	0.3203	0.2077	0.0467
2LZM_pW	T59A	1.50	0.1698	±0.3888	-0.4709	0.3013	0.2453	0.1770	-0.0829
2LZM_pW	T59G	1.60	0.7761	±0.3897	-0.7091	0.3616	0.5319	0.4808	0.1108
2LZM_pW	T59S	0.20	-0.3686	±0.3877	0.1901	-0.5007	-0.0752	-0.0445	0.0617
2LZM_pW	T59V	1.50	-0.3798	±0.3844	-0.3257	0.3572	-0.1460	-0.2101	-0.0551
2LZM_pW	V111A	1.10	0.7774	±0.3902	0.1273	-0.3041	0.5018	0.5634	-0.1110
2LZM_pW	V149T	3.00	1.2467	±0.3777	0.8021	0.4159	0.0400	-0.0596	0.0482
2LZM_pW	V71A	1.50	1.8297	±0.3920	0.2675	-0.0877	0.7717	0.9641	-0.0860

PDB	Mutation	$\Delta\Delta G_{\text{exp}}$	$\Delta\Delta G_{\text{calc}}$		$\Delta\Delta G_{\text{RF}}$	$\Delta\Delta G_{\text{Coul}}$	$\Delta\Delta G_{\text{LJ}}$	$\Delta\Delta G_{\text{SA}}$	$-T\Delta\Delta S$
2LZM_pW	V75T	1.30	0.3072	±0.3806	0.5199	0.1310	-0.1879	-0.1678	0.0121
2LZM_pW	V87A	1.70	1.5537	±0.3931	0.4393	-0.2999	0.6740	0.6789	0.0613
2LZM_pW	V87T	1.60	0.2560	±0.3817	0.6656	-0.2236	-0.0592	-0.1333	0.0064
2LZM_pW	V94A	1.80	0.7833	±0.3934	0.0581	-0.1377	0.4197	0.3698	0.0735
2LZM	Q105A	0.60	2.1824	±0.3635	-2.3017	2.0274	1.1825	1.5670	-0.2928
2LZM	Q105E	1.10	0.6802	±0.3536	21.0233	-19.3743	-0.8489	-0.4436	0.3237
2LZM	Q105G	3.11	2.5340	±0.3698	-2.9951	2.1051	1.5786	2.0454	-0.2000
2LZM	Q123E	-0.40	-0.7031	±0.3644	20.7350	-21.1032	-0.3720	-0.0424	0.0795
2LZM	T157A	0.50	0.7614	±0.3705	-1.0694	1.5675	0.0794	0.2635	-0.0795
2LZM	T157G	1.10	1.2508	±0.3800	-1.0551	1.3893	0.2936	0.6761	-0.0530
2LZM	T157S	0.66	0.4842	±0.3674	0.2009	-0.1649	0.1019	0.3221	0.0242
2LZM	T157V	1.20	0.4266	±0.3581	-1.2121	1.8563	-0.2488	-0.0149	0.0462
2LZM	V103A	1.91	2.0695	±0.3753	0.0659	0.2962	0.7185	0.9742	0.0147
2LZM	V131A	-0.39	0.8371	±0.3754	-0.0861	0.2540	0.2871	0.4229	-0.0408
2LZM	V131G	0.68	1.3444	±0.3840	-0.1656	0.3664	0.5514	0.5246	0.0676
2LZM	V131T	0.12	0.6522	±0.3621	-0.3257	0.5288	0.1030	0.1973	0.1488
2LZM	V149A	2.87	2.9766	±0.3683	-0.3222	0.5285	1.3631	1.3667	0.0406
2LZM	Y25G	4.55	6.4469	±0.4236	-1.7923	2.1667	2.9194	3.6985	-0.5453
3CHY	A101G	1.00	0.9925	±0.3293	0.4378	-0.2242	0.1691	0.3703	0.2395
3CHY	A113G	1.30	1.0997	±0.3231	1.5854	-1.2318	0.0561	0.4022	0.2878
3CHY	A114G	0.80	0.6208	±0.3255	1.2532	-0.9232	-0.0141	0.0417	0.2633
3CHY	A74G	0.30	0.3431	±0.3349	0.7524	-0.2867	-0.1284	-0.1873	0.1930
3CHY	A99G	0.50	0.6634	±0.3203	2.6903	-2.2020	-0.0336	-0.1444	0.3531
3CHY	D12A	-2.50	-2.6546	±0.3093	0.1073	-4.4995	1.1990	1.0226	-0.4841
3CHY	D13A	-2.70	-2.1046	±0.3115	3.7135	-6.8440	0.7855	0.6535	-0.4131
3CHY	D57A	-3.40	-2.3101	±0.2971	-1.2608	-2.8374	1.0669	1.1166	-0.3954
3CHY	F14A	-0.80	-1.2703	±0.3101	0.6169	-1.5190	0.1721	-0.3906	-0.1497
3CHY	F14N	-2.90	-1.8279	±0.3093	-0.6377	-0.9711	0.1121	-0.3111	-0.0202
3CHY	P61G	0.60	1.4092	±0.3328	-0.6386	0.7873	0.4974	0.5954	0.1677
3CHY_pW	A103G	1.70	1.5571	±0.3332	-0.8915	0.8438	0.6183	1.1206	-0.1341
3CHY_pW	A36G	3.10	1.3395	±0.3251	-0.5105	0.4658	0.7001	0.8031	-0.1190
3CHY_pW	A42G	2.30	0.9322	±0.3287	-0.0696	0.1725	0.3712	0.5168	-0.0587
3CHY_pW	A97G	1.40	0.9794	±0.3200	-0.3129	0.3044	0.4343	0.6198	-0.0662
3CHY_pW	A98G	1.30	1.9445	±0.3283	-1.2940	1.1883	0.9109	1.3317	-0.1924
3CHY_pW	D38A	1.90	0.5499	±0.3128	4.9449	-5.3753	0.6839	1.0041	-0.7076
3CHY_pW	D38G	1.00	2.0989	±0.3215	4.5780	-5.2099	1.2416	2.0911	-0.6018
3CHY_pW	D64A	1.00	0.6460	±0.3018	4.5537	-4.7136	0.7155	0.6625	-0.5722
3CHY_pW	G39A	1.00	0.3149	±0.3160	-0.4938	0.1599	0.5814	0.3093	-0.2418
3CHY_pW	G76A	-0.50	-0.1015	±0.3188	-0.5814	0.2516	0.1928	0.2956	-0.2601
3CHY_pW	I123V	0.80	1.3167	±0.2962	0.1489	-0.0263	0.6339	0.7360	-0.1757
3CHY_pW	I55V	1.50	1.2283	±0.2955	-0.6284	0.4589	0.8134	0.8941	-0.3097
3CHY_pW	I72V	1.50	1.3813	±0.2918	-1.1092	0.9833	0.7856	0.9928	-0.2712
3CHY_pW	N23G	0.00	0.7547	±0.3213	0.0728	-0.1918	0.5884	0.6015	-0.3162
3CHY_pW	T112A	1.50	0.6972	±0.3077	-0.9192	0.8823	0.4440	0.5664	-0.2764
3CHY_pW	T112G	1.00	0.4311	±0.3135	-0.0477	0.0745	0.2233	0.1772	0.0038
3CHY_pW	V108T	1.00	0.5349	±0.3021	-1.3685	0.8094	0.5757	0.6063	-0.0881
3CHY_pW	V10T	5.70	1.4345	±0.3020	-0.2716	0.6643	0.5221	0.6388	-0.1192
3CHY_pW	V11T	3.20	1.2605	±0.2942	-0.2689	0.7141	0.6241	0.4388	-0.2476
3CHY_pW	V21T	0.20	0.6043	±0.2911	-0.2153	0.3600	0.2697	0.3044	-0.1144
3CHY_pW	V33T	1.50	1.2890	±0.2995	-0.1063	0.4512	0.4952	0.5751	-0.1262
3CHY_pW	V40T	0.70	0.4737	±0.2968	-0.7862	0.7578	0.2468	0.3784	-0.1231
3CHY_pW	V54T	4.80	1.0595	±0.3042	0.1663	0.4776	0.1593	0.3825	-0.1262
3CHY_pW	V83T	3.50	1.6970	±0.2947	0.0722	0.8459	0.2970	0.5514	-0.0696

Table A.2: Reproducibility of 1YPC A16G with an experimental stability of 1.09 kcal/mol

number	$\Delta\Delta G_{\mathrm{CC/PBSA}}$	$\Delta\Delta G_{\mathrm{es}}$	$\Delta\Delta G_{\mathrm{LJ}}$	$\Delta\Delta G_{\mathrm{SA}}$	$-T\Delta\Delta S$
test set	1.2403	0.3401	0.2582	0.6109	0.0311
1	1.12000	-0.0527000	0.586000	0.773000	-0.188000
2	1.47000	-0.0353000	0.664000	0.885000	-0.0480000
3	1.20000	0.133000	0.412000	0.627000	0.0295000
4	1.42000	0.0971000	0.587000	0.847000	-0.106000
5	1.15000	0.137000	0.432000	0.637000	-0.0576000
6	0.968000	0.127000	0.310000	0.489000	0.0419000
7	0.748000	0.0823000	0.307000	0.417000	-0.0581000
8	0.669000	0.188000	0.254000	0.366000	-0.139000
9	1.19000	0.00585000	0.587000	0.698000	-0.0964000
10	1.27000	0.0791000	0.528000	0.688000	-0.0239000
mean	1.12050	0.0761350	0.466700	0.642700	-0.0645600
σ	0.260787	0.0792196	0.143511	0.174465	0.0712651

Table A.3: Reproducibility of 1YPC D45A with an experimental stability of 0.80 kcal/mol

number	$\Delta\Delta G_{\mathrm{CC/PBSA}}$	$\Delta\Delta G_{\mathrm{es}}$	$\Delta\Delta G_{\mathrm{LJ}}$	$\Delta\Delta G_{\mathrm{SA}}$	$-T\Delta\Delta S$
test set	0.8935	-0.157600	0.6428	0.8960	-0.4878
1	0.646000	-0.575000	0.800000	0.958000	-0.536000
2	1.07000	-0.400000	0.897000	1.11000	-0.531000
3	0.283000	-0.374000	0.492000	0.644000	-0.479000
4	0.760000	-0.482000	0.749000	0.985000	-0.493000
5	0.739000	-0.300000	0.652000	0.923000	-0.536000
6	1.17000	-0.205000	0.732000	1.10000	-0.455000
7	0.680000	-0.310000	0.629000	0.871000	-0.510000
8	0.568000	-0.331000	0.678000	0.805000	-0.584000
9	0.650000	-0.396000	0.729000	0.891000	-0.574000
10	0.562000	-0.446000	0.665000	0.804000	-0.460000
mean	0.712800	-0.381900	0.702300	0.909100	-0.515800
σ	0.253302	0.104226	0.108187	0.141063	0.0444767

Table A.4: Reproducibility of 1YPC E15Q with an experimental stability of 0.47 kcal/mol

number	$\Delta\Delta G_{\mathrm{CC/PBSA}}$	$\Delta\Delta G_{\mathrm{es}}$	$\Delta\Delta G_{\mathrm{LJ}}$	$\Delta\Delta G_{\mathrm{SA}}$	$-T\Delta\Delta S$
test set	0.3592	-0.0565	0.1109	0.3345	-0.0298
1	0.217000	-0.259000	0.182000	0.466000	-0.172000
2	0.600000	-0.140000	0.263000	0.559000	-0.0819000
3	0.431000	-0.0209000	0.101000	0.465000	-0.113000
4	0.506000	0.0205000	0.116000	0.432000	-0.0633000
5	0.355000	0.0174000	0.0613000	0.286000	-0.0105000
6	0.187000	-0.118000	1.23000e-05	0.288000	0.0174000
7	-0.204000	-0.0806000	-0.150000	0.0402000	-0.0136000
8	0.214000	0.0291000	0.0713000	0.344000	-0.230000
9	0.459000	-0.145000	0.196000	0.476000	-0.0676000
10	0.542000	-0.288000	0.296000	0.567000	-0.0329000
mean	0.330700	-0.0984500	0.113661	0.392320	-0.0767400
σ	0.237607	0.113604	0.130867	0.158498	0.0769757

Table A.5: Reproducibility of 1YPC F50A with an experimental stability of 3.84 kcal/mol

number	$\Delta\Delta G_{\mathrm{CC/PBSA}}$	$\Delta\Delta G_{\mathrm{es}}$	$\Delta\Delta G_{\mathrm{LJ}}$	$\Delta\Delta G_{\mathrm{SA}}$	$-T\Delta\Delta S$
test set	3.5671	0.0910	1.8865	2.0846	-0.4950
1	2.75000	-0.765000	2.03000	2.06000	-0.581000
2	3.36000	-0.589000	2.20000	2.35000	-0.600000
3	2.75000	-0.513000	1.92000	2.03000	-0.687000
4	3.09000	-0.354000	1.99000	2.12000	-0.668000
5	2.90000	-0.297000	1.85000	1.97000	-0.625000
6	2.69000	-0.708000	1.91000	2.01000	-0.513000
7	2.89000	-0.389000	1.87000	2.00000	-0.593000
8	3.25000	-0.586000	2.19000	2.39000	-0.747000
9	2.84000	-0.664000	2.07000	2.05000	-0.622000
10	2.94000	-0.637000	2.05000	2.17000	-0.646000
mean	2.94600	-0.550200	2.00800	2.11500	-0.628200
σ	0.221971	0.157863	0.123720	0.146686	0.0641644

Table A.6: Reproducibility of 1YPC N56D with an experimental stability of 1.21 kcal/mol

number	$\Delta\Delta G_{\mathrm{CC/PBSA}}$	$\Delta\Delta G_{\mathrm{es}}$	$\Delta\Delta G_{\mathrm{LJ}}$	$\Delta\Delta G_{\mathrm{SA}}$	$-T\Delta\Delta S$
test set	1.1854	1.23720	-0.2180	0.0010	0.1652
1	0.979000	1.08000	-0.126000	0.0482000	-0.0264000
2	1.50000	1.23000	-0.0437000	0.220000	0.0875000
3	1.22000	1.37000	-0.224000	-0.00793000	0.0735000
4	1.36000	1.24000	-0.0574000	0.226000	-0.0564000
5	1.17000	1.30000	-0.222000	0.0124000	0.0808000
6	1.30000	1.30000	-0.205000	0.0984000	0.110000
7	1.02000	1.13000	-0.190000	0.0262000	0.0541000
8	1.18000	1.31000	-0.185000	0.0560000	-0.000268000
9	1.35000	0.968000	0.0719000	0.302000	0.00371000
10	1.52000	1.03000	0.0768000	0.335000	0.0743000
mean	1.25990	1.19580	-0.110440	0.131627	0.0400842
σ	0.181830	0.135543	0.116423	0.127233	0.0557362

Table A.7: Reproducibility of 1YPC S12A with an experimental stability of 0.89 kcal/mol

number	$\Delta\Delta G_{\mathrm{CC/PBSA}}$	$\Delta\Delta G_{\mathrm{es}}$	$\Delta\Delta G_{\mathrm{LJ}}$	$\Delta\Delta G_{\mathrm{SA}}$	$-T\Delta\Delta S$
test set	0.7659	0.1130	0.3527	0.4208	-0.1206
1	0.672000	-0.106000	0.512000	0.532000	-0.266000
2	1.27000	0.0951000	0.570000	0.719000	-0.110000
3	0.294000	0.204000	0.127000	0.127000	-0.164000
4	0.692000	-0.0412000	0.394000	0.446000	-0.106000
5	0.414000	-0.0452000	0.284000	0.329000	-0.155000
6	0.752000	-0.0970000	0.407000	0.508000	-0.0661000
7	0.277000	0.104000	0.208000	0.181000	-0.215000
8	0.512000	0.113000	0.328000	0.351000	-0.280000
9	0.593000	-0.131000	0.503000	0.469000	-0.248000
10	0.765000	-0.132000	0.535000	0.581000	-0.219000
mean	0.624100	-0.00363000	0.386800	0.424300	-0.182910
σ	0.288179	0.121602	0.148569	0.180897	0.0737198

Table A.8: Reproducibility of 1YPC T39D with an experimental stability of −0.02 kcal/mol

number	$\Delta\Delta G_{\mathrm{CC/PBSA}}$	$\Delta\Delta G_{\mathrm{es}}$	$\Delta\Delta G_{\mathrm{LJ}}$	$\Delta\Delta G_{\mathrm{SA}}$	$-T\Delta\Delta S$
test set	0.2278	-0.25490	0.1504	0.1430	0.1893
1	-0.289000	-0.439000	0.0433000	0.166000	-0.0592000
2	0.405000	-0.257000	0.251000	0.431000	-0.0198000
3	-0.259000	-0.169000	-0.118000	-0.0476000	0.0758000
4	0.176000	-0.143000	0.0647000	0.250000	0.00497000
5	-0.461000	-0.254000	-0.153000	-0.0573000	0.00312000
6	-0.0139000	-0.221000	-0.0425000	0.180000	0.0696000
7	-0.393000	-0.266000	-0.0904000	-0.0174000	-0.0201000
8	0.0326000	-0.186000	0.0645000	0.246000	-0.0911000
9	-0.0862000	-0.368000	0.0775000	0.159000	0.0454000
10	-0.318000	-0.419000	0.00609000	0.111000	-0.0152000
mean	-0.120650	-0.272200	0.0103190	0.142070	-0.000650999
σ	0.274230	0.103573	0.118092	0.152879	0.0532417

Table A.9: Reproducibility of 1YPC V63T with an experimental stability of 1.15 kcal/mol

number	$\Delta\Delta G_{\mathrm{CC/PBSA}}$	$\Delta\Delta G_{\mathrm{es}}$	$\Delta\Delta G_{\mathrm{LJ}}$	$\Delta\Delta G_{\mathrm{SA}}$	$-T\Delta\Delta S$
test set	0.8996	0.46940	0.1408	0.1986	0.0908
1	0.560000	0.0840000	0.334000	0.214000	-0.0713000
2	1.18000	0.443000	0.329000	0.347000	0.0593000
3	0.717000	0.456000	0.133000	0.0708000	0.0568000
4	1.15000	0.615000	0.278000	0.306000	-0.0497000
5	0.683000	0.437000	0.147000	0.115000	-0.0166000
6	0.853000	0.500000	0.179000	0.145000	0.0288000
7	0.905000	0.511000	0.215000	0.172000	0.00723000
8	1.20000	0.630000	0.311000	0.326000	-0.0660000
9	0.687000	0.237000	0.289000	0.166000	-0.00469000
10	1.09000	0.408000	0.381000	0.382000	-0.0775000
mean	0.902500	0.432100	0.259600	0.224380	-0.0133660
σ	0.238215	0.164649	0.0857168	0.108002	0.0516952

Appendix B

GYF-binding results

Table B.1: Concoord/PBSA results for GYF calculations using the Concoord/PBSA web interface. Binding free energies are shown relative to the wild type. All energies are in kcal/mol. Also the pK_a calculations for the two histidines in the peptide chain in bound and unbound conformation are presented. The input structures for concoord are used in pK_a calculations.

					pK_a			
					bound		free	
Mutation	$\Delta\Delta G_{bind}$	$\Delta\Delta G_{LJ}$	$\Delta\Delta G_{es}$	$\Delta\Delta G_{PPIS}$	H64	H71	H64	H71
S63A	1.82	0.305	-0.368	1.88	6.463	6.964	6.388	6.702
S63C	1.74	0.154	-0.289	1.88	6.679	7.167	6.621	6.863
S63D	2.06	0.368	-0.192	1.88	7.207	7.164	7.152	6.859
S63E	2.01	0.226	-0.095	1.88	6.792	7.108	6.967	6.805
S63F	1.82	0.33	-0.388	1.88	6.336	6.883	6.372	6.647
S63G	1.77	0.157	-0.265	1.88	6.671	7.157	6.548	6.847
S63H	1.77	0.193	-0.302	1.88	6.437	7.125	6.311	6.809
S63I	1.76	0.194	-0.316	1.88	6.229	6.901	6.274	6.652
S63K	1.62	0.198	-0.461	1.88	6.226	7.190	6.006	6.860
S63L	1.82	0.196	-0.26	1.88	6.448	6.852	6.387	6.614
S63M	1.51	-0.354	-0.0109	1.88	6.274	7.066	6.267	6.817
S63N	1.87	0.209	-0.218	1.88	6.606	7.103	6.476	6.773
S63P	1.91	0.0418	-0.0133	1.88	6.710	7.056	6.252	6.804
S63Q	1.8	0.24	-0.315	1.88	6.738	7.155	6.481	6.842
S63R	1.46	-0.145	-0.278	1.88	6.399	7.186	6.057	6.870
S63T	1.57	-0.273	-0.0351	1.88	6.406	6.945	6.397	6.677
S63V	1.72	0.148	-0.305	1.88	6.598	7.155	6.512	6.844
S63W	1.89	-0.00855	0.0217	1.88	6.497	6.934	6.360	6.684
S63Y	1.78	0.247	-0.349	1.88	6.244	7.168	6.364	6.715
H64A	1.89	0.367	-0.358	1.88		7.029		7.300
H64C	1.81	0.222	-0.289	1.88		6.694		6.923
H64D	1.84	0.214	-0.25	1.88		6.718		7.004
H64E	2.02	0.316	-0.176	1.88		6.862		7.165
H64F	1.59	-0.191	-0.104	1.88		6.606		6.876
H64G	1.86	0.318	-0.334	1.88		6.783		7.075
H64I	1.82	0.327	-0.389	1.88		6.792		7.068
H64K	1.77	0.238	-0.344	1.88		6.593		6.828
H64L	1.77	0.2	-0.306	1.88		6.637		6.944
H64M	1.87	0.262	-0.275	1.88		6.557		6.773
H64N	1.77	0.207	-0.316	1.88		6.750		6.998
H64P	1.77	0.266	-0.377	1.88		6.591		6.846
H64Q	1.84	0.215	-0.255	1.88		6.755		6.989

Mutation	$\Delta\Delta G_{bind}$	$\Delta\Delta G_{LJ}$	$\Delta\Delta G_{es}$	$\Delta\Delta G_{PPIS}$	pK_a bound H64	pK_a bound H71	pK_a free H64	pK_a free H71
H64R	1.86	0.184	-0.201	1.88		6.963		7.211
H64S	2	0.206	-0.0814	1.88		6.727		7.035
H64T	1.92	0.291	-0.255	1.88		6.726		7.016
H64V	1.87	0.32	-0.334	1.88		6.735		6.953
H64W	1.9	0.0535	-0.0337	1.88		6.604		6.848
H64Y	1.72	0.239	-0.401	1.88		6.606		6.900
R65A	1.9	0.882	-0.858	1.88	6.910	7.041	6.716	6.746
R65C	1.86	0.598	-0.62	1.88	6.997	7.249	6.758	6.900
R65D	2.02	0.525	-0.387	1.88	7.107	7.142	6.902	6.821
R65E	2.97	-0.259	1.35	1.88	7.076	7.247	6.658	6.887
R65F	1.35	0.0821	-0.61	1.88	6.888	7.128	6.645	6.805
R65G	2.26	1.3	-0.925	1.88	6.802	7.067	6.767	6.762
R65H	1.45	0.212	-0.644	1.88	6.656	7.079	6.571	6.769
R65I	1.29	0.191	-0.782	1.88	6.937	7.208	6.571	6.910
R65K	2.11	0.419	-0.187	1.88	6.816	7.211	6.684	6.843
R65L	1.36	0.127	-0.651	1.88	6.847	7.282	6.733	6.750
R65M	1.47	-0.135	-0.279	1.88	7.009	6.983	6.738	6.687
R65N	2.06	0.453	-0.268	1.88	6.924	7.263	6.742	6.904
R65P	1.67	0.457	-0.672	1.88	6.889	7.223	6.683	6.899
R65Q	1.52	-0.115	-0.25	1.88	7.018	7.273	6.746	6.911
R65S	2.09	0.895	-0.684	1.88	6.941	7.201	6.519	6.870
R65T	1.87	0.702	-0.71	1.88	6.810	7.027	6.730	6.719
R65V	1.7	0.591	-0.772	1.88	6.919	7.251	6.757	6.904
R65W	1.36	-0.275	-0.245	1.88	6.795	7.123	6.610	6.784
R65Y	1.13	-0.0706	-0.675	1.88	6.962	6.954	6.760	6.667
P66A	2.11	0.597	-0.367	1.88	6.747	7.175	6.659	6.854
P66C	1.83	0.34	-0.39	1.88	6.607	6.887	6.470	6.596
P66D	2.16	0.586	-0.303	1.88	6.810	7.068	6.705	6.799
P66E	1.91	-0.0474	0.0795	1.88	6.836	7.285	6.812	6.954
P66F	2.12	0.41	-0.169	1.88	6.601	7.035	6.504	6.788
P66G	2.18	0.629	-0.332	1.88	6.677	6.888	6.547	6.632
P66H	1.43	-0.339	-0.111	1.88	6.566	6.928	6.433	6.663
P66I	1.8	0.0698	-0.151	1.88	6.655	6.885	6.573	6.639
P66K	2.03	0.539	-0.394	1.88	6.449	6.927	6.367	6.726
P66L	2.04	0.493	-0.331	1.88	6.683	7.054	6.493	6.767
P66M	2.07	0.526	-0.331	1.88	6.706	7.148	6.597	6.858
P66N	2.12	0.472	-0.236	1.88	6.539	7.139	6.455	6.868
P66Q	1.96	-0.0843	0.165	1.88	6.712	7.098	6.632	6.834
P66R	1.77	-0.109	-0.00245	1.88	6.486	7.239	6.313	7.049
P66S	2.13	0.643	-0.393	1.88	6.690	7.109	6.478	6.817
P66T	2.03	0.393	-0.239	1.88	6.730	7.123	6.642	6.841
P66V	2.1	0.486	-0.262	1.88	6.658	6.877	6.543	6.638
P66W	2.04	0.443	-0.287	1.88	6.665	7.168	6.584	6.848
P66Y	1.1	-0.706	-0.075	1.88	6.555	7.041	6.559	6.807
P67A	1.73	0.299	-0.445	1.88	6.655	7.112	6.523	6.807
P67C	1.74	0.181	-0.324	1.88	6.536	7.153	6.493	6.932
P67D	1.88	0.211	-0.206	1.88	6.623	7.214	6.508	6.935
P67E	1.81	0.207	-0.282	1.88	6.469	7.322	6.541	6.906
P67F	1.74	0.172	-0.312	1.88	6.780	7.156	6.590	6.865
P67G	1.9	0.359	-0.341	1.88	6.362	6.923	6.499	6.670
P67H	1.59	-0.00707	-0.284	1.88	6.666	6.950	6.564	6.639
P67I	1.53	0.0914	-0.446	1.88	6.429	6.925	6.492	6.625
P67K	1.45	0.196	-0.626	1.88	6.611	6.809	6.448	6.579
P67L	1.75	0.251	-0.377	1.88	6.697	7.129	6.529	6.813
P67M	1.49	0.0817	-0.473	1.88	6.621	6.976	6.504	6.647

Mutation	$\Delta\Delta G_{bind}$	$\Delta\Delta G_{LJ}$	$\Delta\Delta G_{es}$	$\Delta\Delta G_{PPIS}$	pK_a bound H64	pK_a bound H71	pK_a free H64	pK_a free H71
P67N	1.67	0.118	-0.332	1.88	6.758	7.138	6.590	6.794
P67Q	1.75	0.145	-0.277	1.88	6.717	7.009	6.507	6.753
P67R	1.14	-0.0953	-0.642	1.88	6.470	6.425	6.339	6.211
P67S	2.01	0.252	-0.125	1.88	6.698	7.205	6.613	6.903
P67T	1.69	0.159	-0.351	1.88	6.674	7.071	6.578	6.752
P67V	1.55	0.0303	-0.358	1.88	6.676	7.031	6.524	6.723
P67W	1.66	0.117	-0.342	1.88	6.228	7.116	6.190	6.657
P67Y	1.67	0.13	-0.339	1.88	6.706	7.116	6.608	6.829
P68A	2.03	0.393	-0.243	1.88	6.628	6.992	6.485	6.731
P68C	1.65	-0.161	-0.0644	1.88	6.441	6.972	6.471	6.785
P68D	5.02	-0.907	4.05	1.88	6.547	7.522	6.620	7.094
P68E	5.51	-1.09	4.72	1.88	6.323	7.068	6.605	6.883
P68F	0.316	-1.84	0.281	1.88	6.674	7.213	6.568	6.909
P68G	2.52	0.853	-0.209	1.88	6.774	7.173	6.604	6.864
P68H	1.43	-0.8	0.351	1.88	6.571	7.028	6.413	6.733
P68I	1.56	0.0184	-0.343	1.88	6.694	7.190	6.573	6.877
P68K	3.47	-1.15	2.75	1.88	6.444	6.890	6.502	6.515
P68L	1.29	-0.399	-0.191	1.88	6.647	7.248	6.484	6.911
P68M	1.1	-0.951	0.168	1.88	6.685	7.144	6.550	6.859
P68N	1.83	-0.487	0.44	1.88	6.272	6.756	6.572	6.511
P68Q	1.69	-1.14	0.953	1.88	6.580	7.019	6.574	6.733
P68R	1.34	-1.96	1.42	1.88	6.490	6.646	6.514	6.498
P68S	2.11	0.191	0.0381	1.88	6.761	7.154	6.631	6.876
P68T	2.03	0.139	0.00701	1.88	6.688	7.147	6.559	6.840
P68V	1.83	0.0344	-0.0824	1.88	6.498	7.001	6.372	6.929
P68W	0.992	-2.02	1.14	1.88	6.427	7.163	6.579	6.642
P68Y	0.384	-2.38	0.888	1.88	6.553	6.625	6.512	6.448
P69A	2.41	0.564	-0.0322	1.88	6.505	6.951	6.443	6.778
P69C	2.29	0.692	-0.282	1.88	6.710	7.267	6.579	6.954
P69D	2.59	0.396	0.31	1.88	6.730	7.549	6.513	7.216
P69E	2.42	-0.269	0.807	1.88	6.712	7.490	6.493	7.085
P69F	2.05	0.224	-0.0568	1.88	6.588	7.339	6.439	6.941
P69G	2.71	1.2	-0.362	1.88	6.700	7.212	6.587	6.964
P69H	1.97	-0.168	0.253	1.88	6.342	7.123	6.500	6.801
P69I	2.15	0.36	-0.0883	1.88	6.478	6.956	6.548	6.670
P69K	1.87	0.217	-0.224	1.88	6.208	6.952	6.437	6.818
P69L	2.03	0.284	-0.13	1.88	6.692	7.215	6.561	6.913
P69M	2.14	0.432	-0.172	1.88	6.642	7.136	6.522	6.869
P69N	2.12	0.323	-0.0877	1.88	6.601	7.276	6.665	7.036
P69Q	1.9	0.102	-0.0829	1.88	6.586	7.242	6.459	7.071
P69R	1.94	-0.0503	0.115	1.88	6.404	6.695	6.505	6.674
P69S	2.41	0.796	-0.269	1.88	6.366	7.116	6.445	6.846
P69T	2.26	0.389	-0.00578	1.88	6.296	7.043	6.492	6.666
P69V	2.26	0.492	-0.114	1.88	6.522	6.872	6.459	6.718
P69W	2.23	0.0168	0.334	1.88	6.540	7.309	6.373	7.031
P69Y	2	0.142	-0.0264	1.88	6.511	7.187	6.352	6.940
G70A	2.08	0.266	-0.0635	1.88	6.666	6.975	6.533	6.706
G70C	2.34	0.108	0.35	1.88	6.365	7.181	6.492	6.926
G70D	4.47	-0.278	2.86	1.88	6.494	7.492	6.488	7.209
G70E	4.06	-1.67	3.86	1.88	6.614	7.303	6.539	7.062
G70F	0.504	-1.48	0.105	1.88	6.580	7.299	6.411	7.092
G70H	1.54	-1.02	0.678	1.88	6.545	7.360	6.433	7.037
G70I	1.75	-0.352	0.222	1.88	6.439	7.116	6.491	6.879
G70K	3.86	-0.77	2.75	1.88	6.645	6.902	6.456	6.687
G70L	1.82	-0.45	0.386	1.88	6.554	7.144	6.388	6.956

Mutation	$\Delta\Delta G_{bind}$	$\Delta\Delta G_{LJ}$	$\Delta\Delta G_{es}$	$\Delta\Delta G_{PPIS}$	pK_a bound H64	pK_a bound H71	pK_a free H64	pK_a free H71
G70M	1.23	-1.46	0.801	1.88	6.537	7.096	6.411	6.856
G70N	2.6	-0.0214	0.744	1.88	6.712	7.141	6.595	6.863
G70P	1.94	-0.316	0.372	1.88	6.459	7.247	6.439	7.017
G70Q	1.81	-1.16	1.08	1.88	6.724	7.188	6.638	6.821
G70R	1.73	-1.74	1.59	1.88	6.719	6.831	6.583	6.537
G70S	2.55	0.483	0.185	1.88	6.630	7.107	6.508	6.815
G70T	2.34	0.438	0.0218	1.88	6.647	7.240	6.461	6.962
G70V	2.2	0.449	-0.125	1.88	6.723	7.204	6.635	6.921
G70W	0.328	-2.46	0.913	1.88	6.498	7.255	6.492	7.121
G70Y	0.683	-1.54	0.346	1.88	6.626	7.372	6.482	7.125
H71A	1.89	0.419	-0.414	1.88	6.468		6.600	
H71C	1.85	0.312	-0.338	1.88	6.587		6.574	
H71D	2.01	0.0739	0.0597	1.88	6.571		6.736	
H71E	1.98	0.0725	0.031	1.88	6.645		6.783	
H71F	1.74	0.179	-0.323	1.88	6.509		6.564	
H71G	1.92	0.297	-0.258	1.88	6.566		6.324	
H71I	1.67	0.076	-0.289	1.88	6.535		6.528	
H71K	1.75	0.241	-0.375	1.88	6.465		6.458	
H71L	1.71	0.217	-0.383	1.88	6.466		6.498	
H71M	1.71	0.22	-0.388	1.88	6.575		6.697	
H71N	1.94	0.367	-0.311	1.88	6.555		6.683	
H71P	1.56	-0.142	-0.179	1.88	6.545		6.639	
H71Q	1.87	0.375	-0.388	1.88	6.484		6.628	
H71R	1.68	0.305	-0.502	1.88	6.563		6.362	
H71S	1.95	0.368	-0.299	1.88	6.605		6.737	
H71T	1.74	0.236	-0.373	1.88	6.618		6.737	
H71V	1.73	0.151	-0.296	1.88	6.510		6.627	
H71W	1.82	0.114	-0.17	1.88	6.489		6.460	
H71Y	1.89	0.288	-0.275	1.88	6.559		6.672	
R72A	2.12	0.549	-0.312	1.88	6.699	7.222	6.566	6.902
R72C	2.12	0.646	-0.408	1.88	6.720	7.300	6.601	6.997
R72D	2.59	0.651	0.0551	1.88	6.354	7.889	6.352	7.563
R72E	2.48	0.554	0.047	1.88	6.711	7.433	6.582	7.120
R72F	1.8	0.259	-0.339	1.88	6.712	7.258	6.451	6.953
R72G	2.16	0.734	-0.451	1.88	6.437	6.816	6.421	6.538
R72H	2.04	0.478	-0.313	1.88	6.392	6.909	6.481	6.646
R72I	2.13	0.692	-0.445	1.88	6.435	7.422	6.464	7.088
R72K	1.97	0.316	-0.229	1.88	6.718	7.180	6.572	6.847
R72L	2.01	0.556	-0.426	1.88	6.688	7.312	6.490	7.008
R72M	2.1	0.425	-0.203	1.88	6.717	7.292	6.532	6.984
R72N	2.22	0.581	-0.245	1.88	6.718	7.236	6.527	6.901
R72P	2.05	0.578	-0.406	1.88	6.549	7.390	6.564	7.117
R72Q	2.17	0.374	-0.0851	1.88	6.563	7.016	6.500	6.726
R72S	2.25	0.753	-0.386	1.88	6.469	6.907	6.479	6.617
R72T	2.14	0.678	-0.414	1.88	6.353	7.109	6.467	6.834
R72V	2.15	0.719	-0.451	1.88	6.511	7.491	6.502	7.221
R72W	1.99	0.367	-0.26	1.88	6.663	7.194	6.523	6.875
R72Y	1.8	0.0368	-0.121	1.88	6.711	7.176	6.475	6.865
V73A	1.72	0.133	-0.294	1.88	6.462	7.070	6.450	6.776
V73C	1.8	0.15	-0.232	1.88	6.766	7.276	6.601	6.950
V73D	1.91	0.126	-0.0997	1.88	6.749	7.748	6.653	7.446
V73E	1.99	0.258	-0.153	1.88	6.791	7.528	6.584	7.258
V73F	1.84	0.255	-0.296	1.88	6.642	7.134	6.521	6.796
V73G	1.92	0.33	-0.289	1.88	6.359	6.979	6.429	6.678
V73H	1.89	0.24	-0.234	1.88	6.659	7.049	6.582	6.701

Mutation	$\Delta\Delta G_{\text{bind}}$	$\Delta\Delta G_{\text{LJ}}$	$\Delta\Delta G_{\text{es}}$	$\Delta\Delta G_{\text{PPIS}}$	pK_a bound H64	bound H71	free H64	free H71
V73I	1.8	0.143	-0.221	1.88	6.340	6.940	6.499	6.705
V73K	1.74	0.33	-0.467	1.88	6.552	6.821	6.423	6.455
V73L	1.79	0.238	-0.324	1.88	6.386	7.288	6.502	6.998
V73M	1.7	0.154	-0.33	1.88	6.701	7.279	6.619	6.937
V73N	1.73	0.164	-0.312	1.88	6.371	7.227	6.496	6.962
V73P	1.83	0.261	-0.312	1.88	6.480	7.064	6.423	6.841
V73Q	1.72	0.0309	-0.196	1.88	6.726	7.100	6.575	6.848
V73R	1.77	0.355	-0.46	1.88	6.672	6.833	6.477	6.550
V73S	1.9	0.293	-0.269	1.88	6.649	7.109	6.511	6.809
V73T	1.91	0.345	-0.315	1.88	6.702	7.198	6.605	6.842
V73W	1.9	0.0858	-0.0678	1.88	6.306	6.918	6.460	6.614
V73Y	2	0.278	-0.161	1.88	6.480	7.161	6.532	6.825

Bibliography

[1] A. Benedix, C. M. Becker, B. L. de Groot, A. Caflisch, and Rainer A. Böckmann. Predicting free energy changes using structural ensembles. *Nat. Methods*, 6(1):3–4, 2009.

[2] L. Stryer. *Biochemistry*. W.H. Freeman and Company, New York, fourth edition, 1995.

[3] B. Alberts, D. Bray, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, New York, fourth edition, 2002.

[4] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. E. Darnell. *Molecular Cell Biology*. W.H. Freeman and Company, New York, fourth edition, 2000.

[5] T. Gallagher, P. Alexander, P. Bryan, and G. L. Gilliland. Two Crystal Structures of the B1 Immunoglobulin-Binding Domain of Streptococcal Protein G and Comparison with NMR. *Biochemistry*, 33(15):4721–4729, 1994.

[6] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The Protein Data Bank. *Nucleic Acids Res.*, 28(1):235–242, 2000.

[7] A. V. Finkelstein and O. V. Galzitskaya. Physics of protein folding. *Phys. Life Rev.*, 1(1):23–56, 2004.

[8] D. Narzi, I. Daidone, A. Amadei, and A. Di Nola. Protein folding pathways revealed by essential dynamics sampling. *J. Chem. Theor. Comp.*, 4 (11):1940–1948, 2008.

[9] D. Neri, M. Billeter, G. Wider, and K. Wüthrich. NMR Determination of Residual Structure in a Urea–Denatured Protein, the 434–Repressor. *Science*, 257(5076):1559–1563, 1992.

[10] H. J. Dyson and P. E. Wright. Equilibrium NMR studies of unfolded and partially folded proteins. *Nat. Struct. Biol.*, 5:499–503, 1998.

[11] C. B. Anfinsen. Principles that Govern the Folding of Protein Chains. *Science*, 181(4096):223–230, 1973.

[12] C. N. Pace, B. A. Shirley, M. McNutt, and K. Gajiwala. Forces contributing to the conformational stability of proteins. *Faseb J.*, 10(1):75–83, 1996.

[13] M. Ohgushi and A. Wada. Molten–globule state: A compact form of globular proteins with mobile side–chains. *FEBS Lett.*, 164(1):21–24, 1983.

[14] C. Levinthal. Are there pathways for protein folding. *J. Chim. Phys.-Chim. Biol.*, 65(1):44–45, 1968.

[15] B. van den Berg, R. J. Ellis, and C. M. Dobson. Effects of macromolecular crowding on protein folding and aggregation. *Embo J.*, 18(24):6927–6933, 1999.

[16] R. J. Ellis and S. M. van der Vies. Molecular chaperones. *Annu. Rev. Biochem.*, 60:321–347, 1991.

[17] O. Crescenzi, S. Tomaselli, R. Guerrini, S. Salvadori, A.M. D'Ursi, P.A. Temussi, and D. Picone. Solution structure of the Alzheimer amyloid β–peptide (1-42) in an apolar microenvironment. Similarity with a virus fusion domain. *Europ. J. Biochem.*, 269:5642–5648, 2002.

[18] T. Luhrs, C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Dobeli, D. Schubert, and R. Riek. 3D structure of Alzheimer's amyloid–β (1-42) fibrils. *Proc. Natl. Acad. Sci. U. S. A.*, 102: 17342–17347, 2005.

[19] F. Chiti and C. M. Dobson. Protein Misfolding, Functional Amyloid, and Human Disease. *Ann. Rev. Biochem.*, 75:333–366, 2006.

[20] C. Chakraborty, S. Nandi, and S. Jana. Prion disease: A deadly disease for protein misfolding. *Curr. Pharm. Biotechnol.*, 6(2):167–177, 2005.

[21] C. Holmes, D. Boche, D. Wilkinson, G. Yadegarfar, V. Hopkins, A. Bayer, R. W. Jones, R. Bullock, S. Love, J. W. Neal, E. Zotova, and J. A. R. Nicoll. Long–term effects of Aβ_{42} immunisation in Alzheimer's disease: follow–up of a randomised, placebo–controlled phase I trial. *Lancet*, 372(9634):216–223, 2008.

[22] F. Lottspeich and H. Zorbas, editors. *Bioanalytik*. Spektrum, Akademischer Verlag, Heidelberg, Berlin, 1998.

[23] J. Breckow and R.Greinert. *Biophysik: eine Einführung*. de Gryuter, Berlin, New York, 1994.

[24] A. R. Fersht and S. Sato. ϕ-Value analysis and the nature of protein-folding transition states. *Proc. Natl. Acad. Sci. USA*, 101(21):7976–7981, 2004.

[25] R. A. Böckmann and H. Grubmüller. Conformational dynamics of the F_1–ATPase β-subunit: A molecular dynamics study. *Biophys. J.*, 85(3): 1482–1491, 2003.

[26] W. E. Stites. Protein–Protein Interactions: Interface Structure, Binding Thermodynamics, and Mutational Analysis. *Chem. Rev.*, 97(5):1233–1250, 1997.

[27] J. L. Morrison, R. Breitling, D. J. Higham, and D. R. Gilbert. A lock-and-key model for protein–protein interactions. *Bioinformatics*, 22(16): 2012–2019, 2006.

[28] D. E. Koshland. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. U. S. A.*, 44(2):98–104, 1958.

[29] O. F. Lange, N. A. Lakomek, C. Fares, G. F. Schröder, K. F. A. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger, and B. L. de Groot. Recognition dynamics up to microseconds revealed from an rdc-derived ubiquitin ensemble in solution. *Science*, 320(5882):1471–1475, 2008.

[30] C. Becker. Prediction of Protein-Protein Binding Affinities. Master's thesis, Universität des Saarlandes, Center for Bioinformatics, 2007.

[31] B. C. Cunningham and J. A. Wells. High-resolution epitope mapping of hgh-receptor interactions by alanine-scanning mutagenesis. *Science*, 244(4908):1081–1085, 1989.

[32] R. Frank. Spot-synthesis - an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron*, 48(42):9217–9232, 1992.

[33] W. Raith, editor. *Bergmann Schaefer - Lehrbuch der Experimentalphysik*, volume 6 - Festkörper. de Gryuter, Berlin, New York, 1992.

[34] K. Wüthrich. Protein structure determination in solution by NMR spectroscopy. *J. Biol. Chem.*, 265(36):22059–22062, 1990.

[35] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten. Molecular Dynamics Simulations of the Complete Satellite Tobacco Mosaic Virus. *Structure*, 14(3):437–449, 2006.

[36] Y. Duan and P. A. Kollman. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, 282(5389):740–744, 1998.

[37] Martin B. Ulmschneider and Jakob P. Ulmschneider. Folding Peptides into Lipid Bilayer Membranes. *J. Chem. Theor. Comp.*, 4(11):1807–1809, 2008.

[38] A. Warshel and M. Karplus. Calculation of Ground and Excited State Potential Surfaces of Conjugated Molecules. I. Formulation and Parametrization. *J. Am. Chem. Soc.*, 94(16): 5612–5625, 1972.

[39] Y. Sugita and Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314(1-2):141–151, 1999.

[40] M. Shirts and V. S. Pande. Computing: Screen Savers of the World Unite! *Science*, 290(5498): 1903–1904, 2000.

[41] W. F. van Gunsteren and H. J. C. Berendsen. Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. *Angew. Chem.-Int. Edit.*, 29(9):992–1023, 1990.

[42] D. L. Beveridge and F. M. DiCapua. Free Energy via Molecular Simulation: Applications to Chemical and Biomolecular Systems. *Ann. Rev. Biophys. Biophys. Chem.*, 18:431–492, 1989.

[43] T. P. Straatsma and J. A. McCammon. Computational Alchemy. *Ann. Rev. Phys. Chem.*, 43:407–435, 1992.

[44] W. F. van Gunsteren, X. Daura, and A. E. Mark. Computation of Free Energy. *Helv. Chim. Acta*, 85(10):3113–3129, 2002.

[45] W. L. Jorgensen. Free Energy Calculations: A Breakthrough for Modeling Organic Chemistry in Solution. *Accounts Chem. Res.*, 22(5):184–189, 1989.

[46] P. Kollman. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.*, 93(7):2395–2417, 1993.

[47] W. Wang, O. Donini, C. M. Reyes, and P. A. Kollman. Biomolecular Simulations: Recent Developments in Force Fields, Simulations of Enzyme Catalysis, Protein–Ligand, Protein–Protein, and Protein–Nucleic Acid Noncovalent Interactions. *Annu. Rev. Biophys. Biomolec. Struct.*, 30:211–243, 2001.

[48] R. W. Zwanzig. High–Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.*, 22(8):1420–1426, 1954.

[49] H. Y. Liu, A. E. Mark, and W. F. van Gunsteren. Estimating the Relative Free Energy of Different Molecular States with Respect to a Single Reference State. *J. Phys. Chem.*, 100(22):9485–9494, 1996.

[50] J.G. Kirkwood. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.*, 3:300, 1935.

[51] J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman, and D. A. Case. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. *J. Am. Chem. Soc.*, 120(37):9401–9409, 1998.

[52] J. Åqvist, C. Medina, and J. E. Samuelsson. A new method for predicting binding-affinity in computer-aided drug design. *Protein Eng.*, 7(3): 385–391, 1994.

[53] A. Beck, C. Klinguer-Hamour, M. C. Bussat, T. Champion, J. F. Haeuw, L. Goetsch, T. Wurch, M. Sugawara, A. Milon, A. van Dorsselaer, T. Nguyen, and N. Corvaia. Peptides as tools and drugs for immunotherapies. *J. Pept. Sci.*, 13(9): 588–602, 2007.

[54] A. M. Slovic, H. Kono, J. D. Lear, J. G. Saven, and W. F. DeGrado. Computational design of water–soluble analogues of the potassium channel KcsA. *Proc. Natl. Acad. Sci. USA*, 101(7): 1828–1833, 2004.

[55] I. Massova and P. A. Kollman. Computational Alanine Scanning To Probe Protein-Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *J. Am. Chem. Soc.*, 121(36): 8133–8143, 1999.

[56] W. Wang and P. A. Kollman. Free Energy Calculations on Dimer Stability of the HIV Protease using Molecular Dynamics and a Continuum Solvent Model. *J. Mol. Biol.*, 303(4):567–582, 2000.

[57] W. L. DeLano. Unraveling hot spots in binding interfaces: progress and challenges. *Curr. Opin. Struct. Biol.*, 12(1):14–20, 2002.

[58] T. J. Hou, K. Chen, W. A. McLaughlin, B. Z. Lu, and W. Wang. Computational Analysis and Prediction of the Binding Motif and Protein Interacting Partners of the Abl SH3 Domain. *PLoS Comput. Biol.*, 2(1):46–55, 2006.

[59] R. Guerois, J. E. Nielsen, and L. Serrano. Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations. *J. Mol. Biol.*, 320(2):369–387, 2002.

[60] N. Pokala and T. M. Handel. Energy functions for protein design I: Efficient and accurate continuum electrostatics and solvation. *Protein Sci.*, 13(4):925–936, 2004.

[61] N. Pokala and T. M. Handel. Energy Functions for Protein Design: Adjustment with Protein-Protein Complex Affinities, Models for the Unfolded State, and Negative Design of Solubility and Specificity. *J. Mol. Biol.*, 347(1):203–227, 2005.

[62] E. Capriotti, P. Fariselli, and R. Casadio. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.*, 33:W306–W310, 2005.

[63] J. L. Cheng, A. Randall, and P. Baldi. Prediction of Protein Stability Changes for Single-Site Mutations Using Support Vector Machines. *Proteins*, 62(4):1125–1132, 2006.

[64] B. L. de Groot, D. M. F. van Aalten, R. M. Scheek, A. Amadei, G. Vriend, and H. J. C. Berendsen. Prediction of Protein Conformational Freedom From Distance Constraints. *Proteins*, 29(2):240–251, 1997.

[65] E. Rebhan. *Theoretische Physik*, volume 2 - Quantenmechanik, Quantenfeldtheorie, Elementarteilchentheorie, Thermodynamik und Statistik. Spektrum, Akademischer Verlag, Heidelberg, Berlin, 2004.

[66] A.M. Buckle, G. Schreiber, and A.R. Fersht. Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0 Å resolution. *Biochemistry*, 33(30):8878–8889, 1994.

[67] Shirley W.I. Siu and Rainer A. Böckmann. Low Free Energy Barrier for Ion Permeation Through Double-Helical Gramicidin. *J. Phys. Chem. B*, 2009.

[68] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.*, 118(45):11225–11236, 1996.

[69] J. P. Ryckaert and A. Bellemans. Molecular Dynamics of Liquid Alkanes. *Faraday Discuss.*, 66 (66):95–106, 1978.

[70] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.*, 106(3):765–784, 1984.

[71] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comp. Chem.*, 25(13): 1656–1676, 2004.

[72] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comp. Chem.*, 4(2):187–217, 1983.

[73] W. L. Jorgensen and J. Tirado-Rives. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.*, 110(6):1657–1666, 1988.

[74] H. C. Andersen. Molecular-dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72(4):2384–2393, 1980.

[75] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Di Nola, and J. R. Haak. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690, 1984.

[76] S. Nose. A unified formulation of the constant temperature molecular-dynamics methods. *J. Chem. Phys.*, 81(1):511–519, 1984.

[77] W. G. Hoover. Canonical dynamics: Equilibrium phase–space distributions. *Phys. Rev. A*, 31(3): 1695–1697, 1985.

[78] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes . *J. Comp. Phys.*, 23(3):327–341, 1977.

[79] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comp. Chem.*, 18(12):1463–1472, 1997.

[80] V. Krautler, W. F. van Gunsteren, and P. H. Hünenberger. A Fast SHAKE Algorithm to Solve Distance Constraint Equations for Small Molecules in Molecular Dynamics Simulations. *J. Comp. Chem.*, 22(5):501–508, 2001.

[81] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. GROMACS: a message-passing parallel molecular-dynamics implementation. *Comput. Phys. Commun.*, 91(1-3):43–56, 1995.

[82] E. Lindahl, B. Hess, and D. van der Spoel. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.*, 7(8):306–317, 2001.

[83] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: Fast, Flexible, and Free. *J. Comp. Chem.*, 26(16):1701–1718, 2005.

[84] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. Scalable Molecular Dynamics with NAMD. *J. Comp. Chem.*, 26(16): 1781–1802, 2005.

[85] J. W. Ponder. *TINKER: Software Tools for Molecular Design, 4.2 ed.* Washington University School of Medicine, Saint Louis, MO, 2001.

[86] W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Kruger, and W. F. van Gunsteren. The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A*, 103(19): 3596–3607, 1999.

[87] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvary, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, and P. A. Kollman. *AMBER 10*. University of California, San Francisco, 2008.

[88] A.D. MacKerel Jr., C.L. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus. *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*, volume 1 of *The Encyclopedia of Computational Chemistry*, pages 271–277. John Wiley & Sons: Chichester, 1998.

[89] W. H. Press and et al. *Numerical Recipes in C*. Cambridge University Press, second edition edition, 2002.

[90] J. Nocedal. Updating Quasi-Newton Matrices with Limited Storage. *Math. Comput.*, 35(151): 773–782, 1980.

[91] D. C. Liu and J. Nocedal. On the limited memory BFGS method for large-scale optimization. *Math. Program.*, 45(3):503–528, 1989.

[92] P. A. Bash, U. C. Singh, R. Langridge, and P. A. Kollman. Free Energy Calculations by Computer Simulation. *Science*, 236(4801):564–568, 1987.

[93] U. C. Singh, F. K. Brown, P. A. Bash, and P. A. Kollman. An Approach to the Application of Free Energy Perturbation Methods Using Molecular Dynamics: Applications to the Transformations of $CH_3OH \rightarrow CH_3CH_3$, $H_3O^+ \rightarrow NH_4^+$, Glycine→Alanine, and Alanine→Phenylalanine in Aqueous Solution and to $H_3O^+(H_2O)_3 \rightarrow NH_4^+(H_2O)_3$ in the Gas Phase. *J. Am. Chem. Soc.*, 109(6):1607–1614, 1987.

[94] Y. Rodriguez, M. Mezei, and R. Osman. Association free energy of dipalmitoylphosphatidylserines in a mixed dipalmitoylphosphatidylcholine membrane. *Biophys. J.*, 92(9):3071–3080, 2007.

[95] F. Schwab, W. F. van Gunsteren, and B. Zagrovic. Computational study of the mechanism and the relative free energies of binding of anticholesteremic inhibitors to squalene-hopene cyclase. *Biochemistry*, 47(9):2945–2951, 2008.

[96] M. Prevost, S. J. Wodak, B. Tidor, and M. Karplus. Contribution of the hydrophobic effect to protein stability: Analysis based on simulations of the Ile–96→Ala mutation in barnase. *Proc. Natl. Acad. Sci. USA*, 88(23):10880–10884, 1991.

[97] B. Tidor and M. Karplus. Simulation Analysis of the Stability Mutant R96H of T4 Lysozyme. *Biochemistry*, 30(13):3217–3228, 1991.

[98] S. Boresch, F. Tettinger, M. Leitgeb, and M. Karplus. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B*, 107(35):9535–9551, 2003.

[99] J. Y. Wang, Y. Q. Deng, and B. Roux. Absolute Binding Free Energy Calculations Using Molecular Dynamics Simulations with Restraining Potentials. *Biophys. J.*, 91(8):2798–2814, 2006.

[100] L. X. Dang, K. M. Merz, and P. A. Kollman. Free Energy Calculations on Protein Stability: Thr–157→Val–157 Mutation of T4 Lysozyme. *J. Am. Chem. Soc.*, 111(22):8505–8508, 1989.

[101] B. Roux. The calculation of the potential of mean force using computer-simulations. *Comput. Phys. Commun.*, 91(1-3):275–282, 1995.

[102] C. Horejs, D. Pum, U. B. Sleytr, and R. Tscheliessnig. Structure prediction of an S-layer protein by the mean force method. *J. Chem. Phys.*, 128(6):065106, 2008.

[103] G. M. Torrie and J. P. Valleau. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comp. Phys.*, 23(2):187–199, 1977.

[104] T. B. Woolf and B. Roux. Conformational Flexibility of o-Phosphorylcholine and o-Phosphorylethanolamine: A Molecular Dynamics Study of Solvation Effects. *J. Am. Chem. Soc.*, 116(13):5916–5926, 1994.

[105] D. Trzesniak, A. P. E. Kunz, and W. F. van Gunsteren. A Comparison of Methods to Compute the Potential of Mean Force. *ChemPhysChem*, 8 (1):162–169, 2007.

[106] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comp. Chem.*, 13(8):1011–1021, 1992.

[107] J. Shen and J. A. McCammon. Molecular-dynamics simulation of superoxide interacting with superoxide-dismutase. *Chem. Phys.*, 158(2-3):191–198, 1991.

[108] C. Jarzynski. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.*, 78(14): 2690–2693, 1997.

[109] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E*, 56(5): 5018–5035, 1997.

[110] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante. Equilibrium Information from Nonequilibrium Measurements in an Experimental Test of Jarzynski's Equality. *Science*, 296(5574):1832–1835, 2002.

[111] J.E. Nielsen and G. Vriend. Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pK_a calculations. *Proteins*, 43: 403–412, 2001.

[112] E. Rebhan. *Theoretische Physik*, volume 1 - Mechanik, Elektrodynamik, spezielle und allgemeine Relativitätstheorie, Kosmologie. Spektrum, Akademischer Verlag, Heidelberg, Berlin, 1999.

[113] M. K. Gilson and B. H. Honig. The Dielectric Constant of a Folded Protein. *Biopolymers*, 25 (11):2097–2119, 1986.

[114] C. N. Schutz and A. Warshel. What Are the Dielectric "Constants" of Proteins and How To Validate Electrostatic Models? *Proteins*, 44(4):400–417, 2001.

[115] D. Voges and A. Karshikoff. A model of a local dielectric constant in proteins. *J. Chem. Phys.*, 108 (5):2219–2227, 1998.

[116] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.*, 112(16):6127–6129, 1990.

[117] R. Abagyan and M. Totrov. Biased Probability Monte Carlo Conformational Searches and Electrostatic Calculations for Peptides and Proteins. *J. Mol. Biol.*, 235(3):983–1002, 1994.

[118] J. J. Havranek and P. B. Harbury. Tanford-Kirkwood electrostatics for protein modeling. *Proc. Natl. Acad. Sci. USA*, 96(20):11145–11150, 1999.

[119] P. Debye and E. Hückel. Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen. *Phys. Z.*, 24(9):185–206, 1923.

[120] J. Warwicker and H. C. Watson. Calculation of the Electric Potential in the Active Site Cleft due to α–Helix Dipoles. *J. Mol. Biol.*, 157(4):671–679, 1982.

[121] I. Klapper, Hagstrom R., R. Fine, K. Sharp, and B. Honig. Focusing of Electric Fields in the Active Site of Cu–Zn Superoxide Dismutase: Effects of Ionic Strength and Amino-Acid Modification. *Proteins*, 1:47–59, 1986.

[122] W. Rocchia, E. Alexov, and B. Honig. Extending the Applicability of the Nonlinear Poisson-Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. *J. Phys. Chem. B*, 105(28):6507–6514, 2001.

[123] W. Rocchia, S. Sridharan, A. Nicholls, E. Alexov, A. Chiabrera, and B. Honig. Rapid Grid–Based Construction of the Molecular Surface and the Use of Induced Surface Charge to Calculate Reaction Field Energies: Applications to the Molecular Systems and Geometric Objects. *J. Comp. Chem.*, 23(1):128–137, 2002.

[124] M. Totrov and R. Abagyan. Rapid Boundary Element Solvation Electrostatics Calculations in Folding Simulations: Successful Folding of a 23–Residue Peptide. *Biopolymers*, 60(2):124–133, 2001.

[125] M. L. Connolly. Analytical Molecular Surface Calculation. *J. Appl. Crystallog.*, 16(5):548–558, 1983.

[126] B. Lee and F. M. Richards. The Interpretation of Protein Structures: Estimation of Static Accessibility. *J. Mol. Biol.*, 55(3):379–400, 1971.

[127] F. M. Richards. Areas, Volumes, Packing, and Protein Structure. *Ann. Rev. Biophys. Bioeng.*, 6 (1):151–176, 1977.

[128] M. E. Davis, J. D. Madura, B. A. Luty, and J. A. McCammon. Electrostatics and diffusion of molecules in solution - simulations with the university-of-houston-brownian dynamics program. *Comput. Phys. Commun.*, 62(2-3):187–197, 1991.

[129] J. D. Madura, J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. R. Scott, and J. A. McCammon. Electrostatics and diffusion of molecules in solution - simulations with the university-of-houston brownian dynamics program. *Comput. Phys. Commun.*, 91(1-3):57–95, 1995.

[130] O. Kohlbacher and H. P. Lenhof. BALL – rapid software prototyping in computational molecular biology. *Bioinformatics*, 16(9):815–824, 2000.

[131] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA*, 98(18):10037–10041, 2001.

[132] M. Born. Volumen und Hydrationswärme der Ionen. *Z. f. Physik*, 1(1):45–48, 1920.

[133] M. Schaefer and M. Karplus. A Comprehensive Analytical Treatment of Continuum Electrostatics. *J. Phys. Chem.*, 100(5):1578–1599, 1996.

[134] D. Qiu, P. S. Shenkin, F. P. Hollinger, and W. C. Still. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem. A*, 101(16):3005–3014, 1997.

[135] G. D. Hawkins, C. J. Cramer, and D. G. Truhlar. Parametrized Model for Aqueous Free Energies of Solvation Using Geometry–Dependent Atomic Surface Tensions with Implicit Electrostatics. *J. Phys. Chem. B*, 101(36):7147–7157, 1997.

[136] M. S. Lee, M. Feig, F. R. Salsbury, and C. L. Brooks. New Analytic Approximation to the Standard Molecular Volume Definition and Its Application to Generalized Born Calculations. *J. Comp. Chem.*, 24(11):1348–1356, 2003.

[137] E. Gallicchio and R. M. Levy. AGBNP: An Analytic Implicit Solvent Model Suitable for Molecular Dynamics Simulations and High–Resolution Modeling. *J. Comp. Chem.*, 25(4):479–499, 2004.

[138] J. A. Grant, B. T. Pickup, M. J. Sykes, C. A. Kitchen, and A. Nicholls. The Gaussian Generalized Born model: application to small molecules. *Phys. Chem. Chem. Phys.*, 9(35):4913–4922, 2007.

[139] U. Haberthür and A. Caflisch. FACTS: Fast Analytical Continuum Treatment of Solvation. *J. Comp. Chem.*, 29(5):701–715, 2008.

[140] D. Bashford and D. A. Case. Generalized Born Models of Macromolecular Solvation Effects. *Ann. Rev. Phys. Chem.*, 51:129–152, 2000.

[141] C. Chothia. Hydrophobic bonding and accessible surface-area in proteins. *Nature*, 248(5446):338–339, 1974.

[142] D. Sitkoff, K. A. Sharp, and B. Honig. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.*, 98(7):1978–1988, 1994.

[143] M. Zacharias. Continuum Solvent Modeling of Nonpolar Solvation: Improvement by Separating Surface Area Dependent Cavity and Dispersion Contributions. *J. Phys. Chem. A*, 107(16):3000–3004, 2003.

[144] H. S. Ashbaugh, E. W. Kaler, and M. E. Paulaitis. A "Universal" Surface Area Correlation for Molecular Hydrophobic Phenomena. *J. Am. Chem. Soc.*, 121(39):9243–9244, 1999.

[145] R. M. Levy, L. Y. Zhang, E. Gallicchio, and A. K. Felts. On the nonpolar hydration free energy of proteins: Surface area and continuum solvent models for the solute–solvent interaction energy. *J. Am. Chem. Soc.*, 125(31):9523–9530, 2003.

[146] T. J. Richmond. Solvent Accessible Surface Area and Excluded Volume in Proteins — Analytical Equations for Overlapping Spheres and Implications for the Hydrophobic Effect. *J. Mol. Biol.*, 178(1):63–89, 1984.

[147] R. Fraczkiewicz and W. Braun. Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules. *J. Comp. Chem.*, 19(3):319–333, 1998.

[148] J. Busa, J. Dzurina, E. Hayryan, S. Hayryan, C. K. Hu, J. Plavka, I. Pokorny, J. Skrivanek, and M. C. Wu. ARVO: A Fortran package for computing the solvent accessible surface area and the excluded volume of overlapping spheres via analytic equations. *Comput. Phys. Commun.*, 165(1): 59–96, 2005.

[149] S. Hayryan, C. K. Hu, J. Skrivanek, E. Hayryan, and I. Pokorny. A New Analytical Method for Computing Solvent–Accessible Surface Area of Macromolecules and its Gradients. *J. Comp. Chem.*, 26(4):334–343, 2005.

[150] W. Hasel, T. F. Hendrickson, and W. C. Still. A Rapid Approximation to the Solvent Accessible Surface Areas of Atoms. *Tetrahedron Computer Methodology*, 1(2):103–116, 1988.

[151] A. Shrake and J. A. Rupley. Environment and Exposure to Solvent of Protein Atoms - Lysozyme and Insulin. *J. Mol. Biol.*, 79(2):351–371, 1973.

[152] J. A. Wagoner and N. A. Baker. Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms. *Proc. Natl. Acad. Sci. USA*, 103(22):8331–8336, 2006.

[153] J. Numata, M. Wan, and Knapp. E.-W. Conformational Entropy of Biomolecules: Beyond the Quasi–Harmonic Approximation. *Genome Informatics*, 18:192–205, 2007.

[154] J. Schlitter. Estimation of absolute and relative entropies of macromolecules using the covariance-matrix. *Chem. Phys. Lett.*, 215(6): 617–621, 1993.

[155] B. Tidor and M. Karplus. The Contribution of Vibrational Entropy to Molecular Association — The Dimerization of Insulin. *J. Mol. Biol.*, 238 (3):405–414, 1994.

[156] M. Karplus and J. N. Kushick. Method for Estimating the Configurational Entropy of Macromolecules. *Macromolecules*, 14(2):325–332, 1981.

[157] B. Brooks and M. Karplus. Harmonic dynamics of proteins: Normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. US – Biol. Sci.*, 80(21):6571–6575, 1983.

[158] M. Levitt, C. Sander, and P. S. Stern. Protein Normal-mode Dynamics: Trypsin Inhibitor, Crambin, Ribonuclease and Lysozyme. *J. Mol. Biol.*, 181(3):423–447, 1985.

[159] B. Tidor and M. Karplus. The Contribution of Cross–Links to Protein Stability: A Normal Mode Analysis of the Configurational Entropy of the Native State. *Proteins*, 15(1):71–79, 1993.

[160] S. Hayward and N. Go. Collective variable description of native protein dynamics. *Ann. Rev. Phys. Chem.*, 46:223–250, 1995.

[161] F. Reinhard and H. Grubmüller. Estimation of absolute solvent and solvation shell entropies via permutation reduction. *J. Chem. Phys.*, 126(1): 014102, 2007.

[162] B. Roux, H. A. Yu, and M. Karplus. Molecular basis for the Born model of ion solvation. *J. Phys. Chem.*, 94(11):4683–4688, 1990.

[163] J. Carlsson, M. Ander, M. Nervall, and J. Aqvist. Continuum Solvation Models in the Linear Interaction Energy Method. *J. Phys. Chem. B*, 110 (24):12034–12041, 2006.

[164] V. Zoete, O. Michielin, and M. Karplus. Protein-ligand binding free energy estimation using molecular mechanics and continuum electrostatics. Application to HIV-1 protease inhibitors. *J. Comput.-Aided Mol. Des.*, 17(12):861–880, 2003.

[165] D. Huang and A. Caflisch. Efficient Evaluation of Binding Free Energy Using Continuum Electrostatics Solvation. *J. Med. Chem.*, 47(23):5791–5797, 2004.

[166] I. Massova and P. A. Kollman. Computational Alanine Scanning To Probe Protein–Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *J. Am. Chem. Soc.*, 121(36): 8133–8143, 1999.

[167] I. Massova and P. A. Kollman. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discov. Design*, 18:113–135, 2000.

[168] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Accounts Chem. Res.*, 33(12): 889–897, 2000.

[169] V. Zoete, M. Meuwly, and M. Karplus. Study of the Insulin Dimerization: Binding Free Energy Calculations and Per–Residue Free Energy Decomposition. *Proteins*, 61(1):79–93, 2005.

[170] P. Liberali, E. Kakkonen, G. Turacchio, C. Valente, A. Spaar, G. Perinetti, R. A. Böckmann, D. Corda, A. Colanzi, V. Marjomaki, and A. Luini. The closure of Pak1–dependent macropinosomes requires the phosphorylation of CtBP1/BARS. *Embo J.*, 27(7):970–981, 2008.

[171] H. Gohlke, C. Kiel, and D. A. Case. Insights into Protein–Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras–Raf and Ras–RalGDS Complexes. *J. Mol. Biol.*, 330(4):891–913, 2003.

[172] B. Kuhn and P. A. Kollman. Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models. *J. Med. Chem.*, 43(20):3786–3791, 2000.

[173] V. Zoete and M. Meuwly. Importance of Individual Side Chains for the Stability of a Protein Fold: Computational Alanine Scanning of the Insulin Monomer. *J. Comp. Chem.*, 27(15):1843–1857, 2006.

[174] A. J. Bordner and R. A. Abagyan. Large–Scale Prediction of Protein Geometry and Stability Changes for Arbitrary Single Point Mutations. *Proteins*, 57(2):400–413, 2004.

[175] F. Ding and N. V. Dokholyan. Emergence of Protein Fold Families through Rational Design. *PLoS Comput. Biol.*, 2(7):725–733, 2006.

[176] S. Y. Yin, F. Ding, and N. V. Dokholyan. Eris: an automated estimator of protein stability. *Nat. Methods*, 4(6):466–467, 2007.

[177] E. Capriotti, P. Fariselli, R. Calabrese, and R. Casadio. Predicting protein stability changes from sequences using support vector machines. *Bioinformatics*, 21:54–58, 2005.

[178] W. Raith, editor. *Bergmann Schaefer - Lehrbuch der Experimentalphysik*, volume 4 - Teilchen. de Gryuter, Berlin, New York, 1992.

[179] G.W.H. Höhne, W. Hemminger, and H.-J. Flammersheim. *Differential scanning calorimetry : an introduction for practitioners*. Springer-Verlag, Berlin, New York, 1996.

[180] P. Connelly, L. Ghosaini, C. Q. Hu, S. Kitamura, A. Tanaka, and J. M. Sturtevant. A Differential Scanning Calorimetric Study of the Thermal Unfolding of Seven Mutant Forms of Phage T4 Lysozyme. *Biochemistry*, 30(7):1887–1891, 1991.

[181] B. S. McCrary, S. P. Edmondson, and J. W. Shriver. Hyperthermophile Protein Folding Thermodynamics: Differential Scanning Calorimetry and Chemical Denaturation of Sac7d. *J. Mol. Biol.*, 264(4):784–805, 1996.

[182] A. Tanaka, J. Flanagan, and J. M. Sturtevant. Thermal unfolding of staphylococcal nuclease and several mutant forms thereof studied by differential scanning calorimetry. *Protein Sci.*, 2(4): 567–576, 1993.

[183] R.B.M. Schasfoort and A.J. Tudos, editors. *Handbook of Surface Plasmon Resonance*. RSC Publishing, 2008.

[184] M. M. Pierce, C. S. Raman, and B. T. Nall. Isothermal Titration Calorimetry of Protein–Protein Interactions. *Methods*, 19(2):213–221, 1999.

[185] M. M. Gromiha, J. An, H. Kono, M. Oobatake, H. Uedaira, and A. Sarai. ProTherm: Thermodynamic Database for Proteins and Mutants. *Nucl. Acids Res.*, 27(1):286–288, 1999.

[186] K. Saraboji, M. M. Gromiha, and M. N. Ponnuswamy. Average Assignment Method for Predicting the Stability of Protein Mutants. *Biopolymers*, 82(1):80–92, 2006.

[187] C. A. Dennis, H. Videler, R. A. Pauptit, R. Wallis, R. James, G. R. Moore, and C. Kleanthous. A structural comparison of the colicin immunity proteins Im7 and Im9 gives new insights into the molecular determinants of immunity–protein specificity. *Biochem. J.*, 333:183–191, 1998.

[188] A. P. Capaldi, C. Kleanthous, and S. E. Radford. Im7 folding mechanism: misfolding on a path to the native state. *Nat. Struct. Biol.*, 9(3):209–216, 2002.

[189] J. W. O'Neill, D. E. Kim, D. Baker, and K. Y. J. Zhang. Structures of the B1 domain of protein L from Peptostreptococcus magnus with a tyrosine to tryptophan substitution. *Acta Crystallogr. Sect. D-Biol. Crystallogr.*, 57:480–487, 2001.

[190] D. E. Kim, C. Fisher, and D. Baker. A Breakdown of Symmetry in the Folding Transition State of Protein L. *J. Mol. Biol.*, 298(5):971–984, 2000.

[191] E. L. McCallister, E. Alm, and D. Baker. Critical role of β–hairpin formation in protein G folding. *Nat. Struct. Biol.*, 7(8):669–673, 2000.

[192] T. R. Hynes and R. O. Fox. The Crystal Structure of Staphylococcal Nuclease Refined at 1.7 Å Resolution. *Proteins*, 10(2):92–105, 1991.

[193] D. Shortle, W. E. Stites, and A. K. Meeker. Contributions of the Large Hydrophobic Amino Acids to the Stability of Staphylococcal Nuclease. *Biochemistry*, 29(35):8033–8041, 1990.

[194] S. M. Green, A. K. Meeker, and D. Shortle. Contributions of the Polar, Uncharged Amino Acids to the Stability of Staphylococcal Nuclease: Evidence for Mutational Effects on the Free Energy of the Denatured State. *Biochemistry*, 31(25): 5717–5728, 1992.

[195] T. Nakano, L. C. Antonio, R. O. Fox, and A. L. Fink. Effect of Proline Mutations on the Stability and Kinetics of Folding of Staphylococcal Nuclease. *Biochemistry*, 32(10):2534–2541, 1993.

[196] M. R. Eftink, C. A. Ghiron, R. A. Kautz, and R. O. Fox. Fluorescence and Conformational Stability Studies of Staphylococcus Nuclease and Its Mutants, Including the Less Stable Nuclease–Concanavalin A Hybrids. *Biochemistry*, 30(5): 1193–1199, 1991.

[197] Y. Harpaz, N. Elmasry, A. R. Fersht, and K. Henrick. Direct observation of better hydration at the N terminus of an α–helix with glycine rather than alanine as the N–cap residue. *Proc. Natl. Acad. Sci. U. S. A.*, 91(1):311–315, 1994.

[198] L. S. Itzhaki, D. E. Otzen, and A. R. Fersht. The Structure of the Transition State for Folding of Chymotrypsin Inhibitor 2 Analysed by Protein Engineering Methods: Evidence for a Nucleation–condensation Mechanism for Protein Folding. *J. Mol. Biol.*, 254(2):260–288, 1995.

[199] L. H. Weaver and B. W. Matthews. Structure of Bacteriophage T4 Lysozyme Refined at 1.7 Å Resolution. *J. Mol. Biol.*, 193(1):189–199, 1987.

[200] T. Alber, D. P. Sun, K. Wilson, J. A. Wozniak, S. P. Cook, and B. W. Matthews. Contributions of hydrogen bonds of Thr 157 to the thermodynamic stability of phage T4 lysozyme. *Nature*, 330(6143):41–46, 1987.

[201] M. Matsumura, W. J. Becktel, and B. W. Matthews. Hydrophobic stabilization in T4 lysozyme determined directly by multiple substitutions of Ile 3. *Nature*, 334(6181):406–410, 1988.

[202] H. Nicholson, E. Soderlind, D. E. Tronrud, and B. W. Matthews. Contributions of Left-handed Helical Residues to the Structure and Stability of Bacteriophage T4 Lysozyme. *J. Mol. Biol.*, 210 (1):181–193, 1989.

[203] M. Blaber, J. D. Lindstrom, N. Gassner, J. Xu, W. H. Dirk, and B. W. Matthews. Energetic Cost and Structural Consequences of Burying a Hydroxyl Group within the Core of a Protein Determined from Ala→Ser and Val→Thr Substitutions in T4 Lysozyme. *Biochemistry*, 32(42): 11363–11373, 1993.

[204] A. E. Eriksson, W. A. Baase, X. J. Zhang, D. W. Heinz, M. Blaber, E. P. Baldwin, and B. W. Matthews. Response of a Protein Structure to Cavity-Creating Mutations and Its Relation to the Hydrophobic Effect. *Science*, 255(5041):178–183, 1992.

[205] B. K. Shoichet, W. A. Baase, R. Kuroki, and B. W. Matthews. A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci. USA*, 92(2):452–456, 1995.

[206] X. J. Zhang, W. A. Baase, and B. W. Matthews. Multiple alanine replacements within α–helix 126-134 of T4 lysozyme have independent, additive effects on both structure and stability. *Protein Sci.*, 1(6):761–776, 1992.

[207] A. E. Eriksson, W. A. Baase, and B. W. Matthews. Similar Hydrophobic Replacements of Leu99 and Phe153 within the Core of T4 Lysozyme Have Different Structural and Thermodynamic Consequences. *J. Mol. Biol.*, 229(3):747–769, 1993.

[208] X. J. Zhang, W. A. Baase, B. K. Shoichet, K. P. Wilson, and B. W. Matthews. Enhancement of protein stability by the combination of point mutations in T4 lysozyme is additive. *Protein Eng.*, 8(10):1017–1022, 1995.

[209] J. W. Wray, W. A. Baase, J. D. Lindstrom, L. H. Weaver, A. R. Poteete, and B. W. Matthews. Structural Analysis of a Non–contiguous Second–site Revertant in T4 Lysozyme Shows that Increasing the Rigidity of a Protein can Enhance its Stability. *J. Mol. Biol.*, 292(5): 1111–1120, 1999.

[210] D. W. Heinz, W. A. Baase, and B. W. Matthews. Folding and Function of a T4 Lysozyme Containing 10 Consecutive Alanines Illustrate the Redundancy of Information in an Amino Acid Sequence. *Proc. Natl. Acad. Sci. USA*, 89(9):3751–3755, 1992.

[211] H. Nicholson, D. E. Anderson, S. Daopin, and B. W. Matthews. Analysis of the Interaction between Charged Side Chains and the α–Helix Dipole Using Designed Thermostable Mutants of Phage T4 Lysozyme. *Biochemistry*, 30(41):9816–9828, 1991.

[212] D. W. Heinz, W. A. Baase, X. J. Zhang, M. Blaber, F. W. Dahlquist, and B. W. Matthews. Accommodation of Amino Acid Insertions in an α–Helix of T4 Lysozyme — Structural and Thermodynamic Analysis. *J. Mol. Biol.*, 236(3):869–886, 1994.

[213] P. Pjura, M. Matsumura, W. A. Baase, and B. W. Matthews. Development of an in-vivo method to identify mutants of phage T4-lysozyme of enhanced thermostability. *Protein Sci.*, 2(12):2217–2225, 1993.

[214] J. Xu, W. A. Baase, E. Baldwin, and B. W. Matthews. The response of T4 lysozyme to large–to–small substitutions within the core and its relation to the hydrophobic effect. *Protein Sci.*, 7 (1):158–177, 1998.

[215] J. Xu, W. A. Baase, M. L. Quillin, E. P. Baldwin, and B. W. Matthews. Structural and thermodynamic analysis of the binding of solvent at internal sites in T4 lysozyme. *Protein Sci.*, 10(5): 1067–1078, 2001.

[216] E. Baldwin, J. Xu, O. Hajiseyedjavadi, W. A. Baase, and B. W. Matthews. Thermodynamic and Structural Compensation in "Size–switch" Core Repacking Variants of Bacteriophage T4 Lysozyme. *J. Mol. Biol.*, 259(3):542–559, 1996.

[217] J. A. Bell, W. J. Becktel, U. Sauer, W. A. Baase, and B. W. Matthews. Dissection of Helix Capping in T4 Lysozyme by Structural and Thermodynamic Analysis of Six Amino Acid Substitutions at Thr 59. *Biochemistry*, 31(14):3590–3596, 1992.

[218] J. D. Klemm, J. A. Wozniak, T. Alber, and D. P. Goldenberg. Correlation between Mutational Destabilization of Phage T4 Lysozyme and Increased Unfolding Rates. *Biochemistry*, 30(2): 589–594, 1991.

[219] B. H. M. Mooers, D. Datta, W. A. Baase, E. S. Zollars, S. L. Mayo, and B. W. Matthews. Repacking the Core of T4 Lysozyme by Automated Design. *J. Mol. Biol.*, 332(3):741–756, 2003.

[220] P. Pjura, L. P. McIntosh, J. A. Wozniak, and B. W. Matthews. Perturbation of Trp 138 in T4 Lysozyme by Mutations at Gln 105 Used to Correlate Changes in Structure, Stability, Solvation, and Spectroscopic Properties. *Proteins*, 15(4): 401–412, 1993.

[221] S. Daopin, U. Sauer, H. Nicholson, and B. W. Matthews. Contributions of Engineered Surface Salt Bridges to the Stability of T4 Lysozyme Determined by Directed Mutagenesis. *Biochemistry*, 30(29):7142–7153, 1991.

[222] M. Blaber, X. J. Zhang, J. D. Lindstrom, S. D. Pepiot, W. A. Baase, and B. W. Matthews. Determination of α–Helix Propensity within the Context of a Folded Protein — Sites 44 and 131 in Bacteriophage T4 Lysozyme. *J. Mol. Biol.*, 235(2): 600–624, 1994.

[223] M. Blaber, W. A. Baase, N. Gassner, and B. W. Matthews. Alanine Scanning Mutagenesis of the α–Helix 115-123 of Phage T4 Lysozyme: Effects on Structure, Stability and the Binding of Solvent. *J. Mol. Biol.*, 246(2):317–330, 1995.

[224] K. Volz and P. Matsumura. Crystal Structure of Escherichia coli CheY Refined at 1.7 Å Resolution. *J. Biol. Chem.*, 266(23):15511–15519, 1991.

[225] E. López-Hernández and L. Serrano. Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2. *Fold. Des.*, 1(1):43–55, 1996.

[226] G. M. Crippen. A Novel Approach to Calculation of Conformation: Distance Geometry. *J. Comp. Phys.*, 24(1):96–107, 1977.

[227] S. Saitoh, T. Nakai, and K. Nishikawa. A Geometrical Constraint Approach for Reproducing the Native Backbone Conformation of a Protein. *Proteins*, 15(2):191–204, 1993.

[228] D. Seeliger and B. L. de Groot. Prediction of Protein Flexibility from Geometrical Constraints. *Biotech International*, 18:20–22, 2006.

[229] K. A. Dill and D. Shortle. Denatured states of proteins. *Ann. Rev. Biochem.*, 60:795–825, 1991.

[230] U. Börjesson and P. H. Hünenberger. Effect of mutations involving charged residues on the stability of staphylococcal nuclease: a continuum electrostatics study. *Protein Eng.*, 16(11):831–840, 2003.

[231] H. X. Zhou. A Gaussian–chain model for treating residual charge–charge interactions in the unfolded state of proteins. *Proc. Natl. Acad. Sci. USA*, 99(6):3569–3574, 2002.

[232] P. J. Kundrotas and A. Karshikoff. Effects of charge–charge interactions on dimensions of unfolded proteins: A Monte Carlo study. *J. Chem. Phys.*, 119(6):3574–3581, 2003.

[233] T. P. Creamer, R. Srinivasan, and G. D. Rose. Modeling Unfolded States of Peptides and Proteins. *Biochemistry*, 34(50):16245–16250, 1995.

[234] T. P. Creamer, R. Srinivasan, and G. D. Rose. Modeling Unfolded States of Proteins and Peptides. II. Backbone Solvent Accessibility. *Biochemistry*, 36(10):2832–2835, 1997.

[235] P. Bernadó, M. Blackledge, and J. Sancho. Sequence-Specific Solvent Accessibilities of Protein Residues in Unfolded Protein Ensembles. *Biophys. J.*, 91(12):4536–4543, 2006.

[236] P. Koehl and M. Delarue. Application of a Self–consistent Mean Field Theory to Predict Protein Side–chains Conformation and Estimate Their Conformational Entropy. *J. Mol. Biol.*, 239(2): 249–275, 1994.

[237] A. J. Doig and M. J. E. Sternberg. Side–chain conformational entropy in protein-folding. *Protein Sci.*, 4(11):2247–2251, 1995.

[238] B. W. Chellgren and T. P. Creamer. Side–Chain Entropy Effects on Protein Secondary Structure Formation. *Proteins*, 62(2):411–420, 2006.

[239] A. H. Elcock. Realistic modeling of the denatured states of proteins allows accurate calculations of the ph dependence of protein stability. *J. Mol. Biol.*, 294(4):1051–1062, 1999.

[240] D. P. Goldenberg. Computational Simulation of the Statistical Properties of Unfolded Proteins. *J. Mol. Biol.*, 326(5):1615–1633, 2003.

[241] A. K. Jha, A. Colubri, K. F. Freed, and T. R. Sosnick. Statistical coil model of the unfolded state: Resolving the reconciliation problem. *Proc. Natl. Acad. Sci. USA*, 102(37):13099–13104, 2005.

[242] G. Vriend. WHAT IF — A Molecular Modeling And Drug Design Program. *J. Mol. Graph.*, 8(1): 52–, 1990.

[243] G. Chinea, G. Padron, R. W. W. Hooft, C. Sander, and G. Vriend. The Use of Position-Specific Rotamers in Model Building by Homology. *Proteins*, 23(3):415–421, 1995.

[244] D. Bashford and M. Karplus. pK_a's of Ionizable Groups in Proteins: Atomic Detail from a Continuum Electrostatic Model. *Biochemistry*, 29(44): 10219–10225, 1990.

[245] M. Stone. Cross–Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc. Ser. B-Methodol.*, 36(2):111–147, 1974.

[246] A. R. Leach. *Molecular Modelling — Principles and Applications*. Pearson Education Limited, 2001.

[247] J. Gasteiger and Th. Engel, editors. *Chemoinformatics*. Wiley-VCH, 2003.

[248] T. Wang, S. Tomic, R. R. Gabdoulline, and R. C. Wade. How Optimal Are the Binding Energetics of Barnase and Barstar? *Biophys. J.*, 87(3): 1618–1630, 2004.

[249] R. F. Greene and C. N. Pace. Urea and Guanidine Hydrochloride Denaturation of Ribonuclease, Lysozyme, α–Chymotrypsin, and β–Lactoglobulin. *J. Biol. Chem.*, 249(17):5388–5393, 1974.

[250] J. K. Myers, C. N. Pace, and J. M. Scholtz. Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci.*, 4(10):2138–2148, 1995.

[251] G. L. Wang and R. L. Dunbrack. PISCES: a protein sequence culling server. *Bioinformatics*, 19 (12):1589–1591, 2003.

[252] R. A. Jarvis and E. A. Patrick. Clustering Using a Similarity Measure Based on Shared Near Neighbors. *IEEE Transactions on Computers*, 11: 1025–1034, 1973.

[253] L. Kaufman and P. J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley-Interscience, 1990.

[254] B. Wallner and A. Elofsson. Can correct protein models be identified? *Protein Sci.*, 12(5):1073–1086, 2003.

[255] S. Hiller, G. Wider, L. L. Imbach, and K. Wuthrich. Interactions with hydrophobic clusters in the urea–unfolded membrane protein OmpX. *Angew. Chem.-Int. Edit.*, 47(5):977–981, 2008.

[256] H. Tafer, S. Hiller, C. Hilty, C. Fernandez, and K. Wuthrich. Nonrandom structure in the urea–unfolded Escherichia coli outer membrane protein X (OmpX). *Biochemistry*, 43(4):860–869, 2004.

[257] Z. Zhou, M. Bates, and J. D. Madura. Structure Modeling, Ligand Binding, and Binding Affinity Calculation (LR-MM-PBSA) of Human Heparanase for Inhibition and Drug Design. *Proteins*, 65(3):580–592, 2006.

[258] J. Marelius, M. Graffner-Nordberg, T. Hansson, A. Hallberg, and J. Åqvist. Computation of affinity and selectivity: binding of 2,4-diaminopteridine and 2,4-diaminoquinazoline inhibitors to dihydrofolate reductases. *J. Comput.-Aided Mol. Des.*, 12:119–131, 1998.

[259] D. Narzi, K. Winkler, J. Saidowski, R. Misselwitz, A. Ziegler, R. A. Böckmann, and U. Alexiev. Molecular recognition of MHC class I complex stability: shaping antigenic features through short– and long–range electrostatic interactions. *J. Biol. Chem.*, 283:1837–1850, 2008.

[260] A. S. Yang, M. R. Gunner, R. Sampogna, K. Sharp, and B. Honig. On the Calculation of pK_as in Proteins. *Proteins*, 15(3):252–265, 1993.

[261] J.E. Nielsen and J.A. McCammon. On the evaluation and optimisation of protein X–ray structures for pK_a calculations. *Protein Sci.*, 12:313–326, 2003.

[262] J.E. Nielsen and J.A. McCammon. Calculating pK_a values in enzyme active sites. *Protein Sci.*, 12:1894–1901, 2003.

[263] R. E. Georgescu, E. G. Alexov, and M. R. Gunner. Combining Conformational Flexibility and Continuum Electrostatics for Calculating pK_as in Proteins. *Biophys. J.*, 83(4):1731–1748, 2002.

[264] G. W. Rudgers and T. Palzkill. Identification of Residues in β–Lactamase Critical for Binding β–Lactamase Inhibitory Protein. *J. Biol. Chem.*, 274(11):6963–6971, 1999.

[265] S. Albeck, R. Unger, and G. Schreiber. Evaluation of direct and cooperative contributions towards the strength of buried hydrogen bonds and salt bridges. *J. Mol. Biol.*, 298(3):503–520, 2000.

[266] D. Reichmann, O. Rahat, S. Albeck, R. Meged, O. Dym, and G. Schreiber. The modular architecture of protein–protein binding interfaces. *Proc. Natl. Acad. Sci. USA*, 102(1):57–62, 2005.

[267] T. Selzer, S. Albeck, and G. Schreiber. Rational design of faster associating and tighter binding protein complexes. *Nat. Struct. Biol.*, 7(7):537–541, 2000.

[268] T. Kortemme and D. Baker. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. U. S. A.*, 99(22):14116–14121, 2002.

[269] D. Lim, H. U. Park, L. De Castro, S. G. Kang, H. S. Lee, S. Jensen, K. J. Lee, and N. C. J. Strynadka. Crystal structure and kinetic analysis of β–lactamase inhibitor protein–II in complex with TEM–1 β–lactamase. *Nat. Struct. Biol.*, 8 (10):848–852, 2001.

[270] S. Shangary and S. Wang. Small–Molecule Inhibitors of the MDM2–p53 Protein–Protein Interaction to Reactivate p53 Function: A Novel Approach for Cancer Therapy. *Annu. Rev. Pharmacol. Toxicol.*, 49:223–41, 2009. *in press*, doi:10.1146/annurev.pharmtox.48.113006.094723.

[271] M. G. Mateu, M. M. Sánchez Del Pino, and A. R. Fersht. Mechanism of folding and assembly of a small tetrameric protein domain from tumor suppressor p53. *Nat. Struct. Biol.*, 6(2):191–198, 1999.

[272] M. G. Mateu and A. R. Fersht. Mutually compensatory mutations during evolution of the tetramerization domain of tumor suppressor p53 lead to impaired hetero–oligomerization. *Proc. Natl. Acad. Sci. USA*, 96(7):3595–3599, 1999.

[273] P. H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A. J. Levine, and N. P. Pavletich. Structure of the MDM2 Oncoprotein Bound to the p53 Tumor Suppressor Transactivation Domain. *Science*, 274(5289):948–953, 1996.

[274] W. Gu and V. Helms. Dynamical binding of proline–rich peptides to their recognition domains. *BBA-Proteins Proteomics*, 1754(1-2):232–238, 2005.

[275] M. M. Kofler and C. Freund. The GYF domain. *FEBS J.*, 273(2):245–256, 2006.

[276] M. Kofler, K. Heuer, T. Zech, and C. Freund. Recognition Sequences for the GYF Domain Reveal a Possible Spliceosomal Function of CD2BP2. *J. Biol. Chem.*, 279(27):28292–28297, 2004.

[277] A. Kramer, U. Reineke, L. Dong, B. Hoffmann, U. Hoffmuller, D. Winkler, R. Volkmer-Engert, and J. Schneider-Mergener. Spot synthesis: observations and optimizations. *J. Pept. Res.*, 54(4): 319–327, 1999.

[278] A. A. Weiser, M. Or-Guil, V. Tapia, A. Leichsenring, J. Schuchhardt, C. Frommel, and R. Volkmer-Engert. SPOT synthesis: Reliability of array–based measurement of peptide binding affnity. *Anal. Biochem.*, 342(2):300–311, 2005.

[279] C. Freund, R. Kuhne, H. L. Yang, S. Park, E. L. Reinherz, and G. Wagner. Dynamic interaction of CD2 with the GYF and the SH3 domain of compartmentalized effector molecules. *Embo J.*, 21 (22):5985–5995, 2002.

[280] W. Gu, M. Kofler, I. Antes, C. Freund, and V. Helms. Alternative Binding Modes of Proline–Rich Peptides Binding to the GYF Domain. *Biochemistry*, 44(17):6404–6415, 2005.

[281] M. Ahmad, W. Gu, and V. Helms. Mechanism of Fast Peptide Recognition by SH3 Domains. *Angew. Chem. Int. Ed.*, 47(40):7626–7630, 2008.

[282] D. V. Goeddel, D. G. Kleid, F. Bolivar, H. L. Heyneker, D. G. Yansura, R. Crea, T. Hirose, A. Kraszewski, K. Itakura, and A. D. Riggs. Expression in escherichia coli of chemically synthesized genes for human insulin. *Proc. Natl. Acad. Sci. USA*, 76:106–110, 1979.

[283] J. Brange, U. Ribel, J. F. Hansen, G. Dodson, M. T. Hansen, S. Havelund, S. G. Melberg, F. Norris, K. Norris, L. Snel, A. R. Sørensen, and H. O. Voigt. Monomeric insulins obtained by protein engineering and their medical implications. *Nature*, 333:679–682, 1988.

[284] J. Brange and A. Vølund. Insulin analogs with improved pharmacokinetic profiles. *Advanced Drug Delivery Reviews*, 35:307–335, 1999.

[285] I. Rakatzi, S. Ramrath, D. Ledwig, O. Dransfeld, T. Bartels, G. Seipke, and I. Eckel. A Novel Insulin Analog With Unique Properties Lys^{B3}, Glu^{B29} Insulin Induces Prominent Activation of Insulin Receptor Substrate 2, but Marginal Phosphorylation of Insulin Receptor Substrate 1. *Diabetes*, 52(9):2227–2238, 2003.

[286] *Rote Liste*. Rote Liste Service GmbH, 2008.

[287] C. Kristensen, T. Kjeldsen, F. C. Wiberg, L. Schaffer, M. Hach, S. Havelund, J. Bass, D. F. Steiner, and A. S. Andersen. Alanine Scanning Mutagenesis of Insulin. *J. Biol. Chem.*, 272(20):12978–12983, 1997.

[288] C. C. Yip and P. Ottensmeyer. Three–dimensional Structural Interactions of Insulin and Its Receptor. *J. Biol. Chem.*, 278(30):27329–27332, 2003.

[289] M. Z. Lou, T. P. J. Garrett, N. M. McKern, P. A. Hoyne, V. C. Epa, J. D. Bentley, G. O. Lovrecz, L. J. Cosgrove, M. J. Frenkel, and C. W. Ward. The first three domains of the insulin receptor differ structurally from the insulin–like growth factor 1 receptor in the regions governing ligand specificity. *Proc. Natl. Acad. Sci. USA*, 103(33): 12429–12434, 2006.

[290] V. Zoete, M. Meuwly, and M. Karplus. A Comparison of the Dynamic Behavior of Monomeric and Dimeric Insulin Shows Structural Rearrangements in the Active Monomer. *J. Mol. Biol.*, 342 (3):913–929, 2004.

[291] E. N. Baker, T. L. Blundell, J. F. Cutfield, S. M. Cutfield, E. J. Dodson, G. G. Dodson, D. M. C. Hodgkin, R. E. Hubbard, N. W. Isaacs, C. D. Reynolds, K. Sakabe, N. Sakabe, and N. M. Vijayan. The structure of 2ZN pig insulin crystals at 1.5 Å resolution. *Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci.*, 319(1195):369–&, 1988.

[292] H. Chen, M. Shi, Z. Y. Guo, Y. H. Tang, Z. S. Qiao, Z. H. Liang, and Y. M. Feng. Four new monomeric insulins obtained by alanine scanning the dimer–forming surface of the insulin molecule. *Protein Eng.*, 13(11):779–782, 2000.

[293] N. C. Kaarsholm and S. Ludvigsen. The high–resolution solution structure of the insulin monomer determined by NMR. *Receptor*, 5(1):1–8, 1995.

[294] D. Seeliger, J. Haas, and B. L. de Groot. Geometry–Based Sampling of Conformational Transitions in Proteins. *Structure*, 15(11):1482–1492, 2007.

[295] J. Seibert. MHC:peptide structure prediction. Bachelor's Thesis, Universität des Saarlandes, Center for Bioinformatics, 2008.

[296] S. Eyrisch and V. Helms. Transient pockets on protein surfaces involved in protein-protein interaction. *J. Med. Chem.*, 50(15):3457–3464, 2007.

[297] S. Eyrisch and V. Helms. What induces pocket openings on protein surface patches involved in protein–protein interactions? *J. Comput. Aided. Mol. Des.*, 23(2):73–86, 2009.