

Computational Characterization of Dimerization and Ligand Binding in Biological Systems

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von
Lixian Zhang
aus Fujian (China)

Basel, 2013

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von:

Prof. Dr. Markus Meuwly

Prof. Dr. Thomas Pfohl

Basel, den 21. Mai 2013

Prof. Dr. Jörg Schibler
Dekan

Acknowledgements

First and foremost, I would like to thank my PhD supervisor Prof. Markus Meuwly for giving me the opportunity to work as a doctoral student in his group and providing me with this very interesting project. I am very grateful to him for his guidance, support, helpful discussions throughout my PhD study.

I would also like to thank Prof. Thomas Pfohl who kindly accepted to act as the co-examiner.

I would like to thank all the current and past members of the Meuwly group. It has been a pleasure to work in such a friendly group. Special thanks go to Dr. Franziska Schmid, Dr. Jing Huang, for their help when I started my PhD, Dr. Michael Devereux, Dr. Maurus Schmid, Dr. Tibor Nagy for english proofreading parts of this thesis, Dr. Pierre-Andre Cazade and Florent Hedin for sharing the nice time in the office, Franziska Hofmann for her kind help during my knee injury, Dr. Ana Patricia Gamiz-Hernandez, Dr. Tristan Bereau, Juvenal Yosa Reyes for the great time during lunch in Mensa, Dr. Myung Won Lee, Maksym Soloviov, Dr. Jaroslav Padevet, Prashant Gupta, Dr. Yonggang Yang, Dr. Stephan Lutz, Dr. Nuria Plattner, Dr. Jaroslaw Szymczak, Dr. Marek Orzechowski, Vijay Solomon Rajadurai, Dr. Christian Kramer, Dr. Tobias Schmidt.

I am very grateful to everyone who helps me during my PhD study in Basel, especially

Acknowledgements

secretaries in the Department of Physical Chemistry, Ms. Daniela Tischhauser, Ms. Maya Greuter and Ms. Esther Stalder.

Most importantly, I wish to thank my family, Maurus and Schmid family for their love and support.

Abstract

The self-association of proteins to form dimers or higher-order oligomers is a very common phenomenon in biology. Protein dimerization or oligomerization acts as a control tool for the execution of functions in many biological systems. Three systems were studied by computational methods in this thesis. Cyclic diguanylic acid (CDG) is a ubiquitous messenger involved in bacterial signaling networks. CDG can form an intercalated dimer and bind at the inhibition site of PleD. MD simulations were carried out for the CDG dimer as well as the analogue of CDG (endo-S-CDG) in solution and binding to the PleD protein. It was demonstrated, that dimeric CDG is only marginally stabilized even in high concentration. The results help the fundamental understanding of *c*-di-GMP and preventing biofilm formation. Insulin is a small protein that plays an eminent role in controlling glucose uptake in cells. Insulin can associate as a dimer which leads to diabetes. The key role of the B24 residue for insulin dimerization was identified. Our work provided an insight for designing analogues of human insulin and thus a therapy for diabetes.

Contents

Acknowledgements	iii
Abstract	v
1 Overview	1
2 Introduction	3
2.1 Proteins	3
2.1.1 Structure and Function	4
2.1.2 Allostery	9
2.2 Protein-Protein Interactions	10
2.2.1 Protein Dimerization	11
2.2.2 Thermodynamics and Protein Dimerization	17
2.2.3 Diseases Caused by Protein Dimerization	19
2.3 Computational Simulations Methods	22
2.3.1 Atomistic MD Simulations	22
2.3.2 Accuracy Improvement and Other Methods	25
3 Determination of Protein-Protein Binding Affinities	31
3.1 Experimental Approaches	32
3.1.1 Experimental Determination of Binding Affinities	32
3.1.2 High-throughput Screening	33

Contents

3.2	Computational Methods	34
3.2.1	Virtual Screening	34
3.2.2	Docking and Scoring	35
3.2.3	Molecular Mechanics Methods	37
4	Bacterial Second Messenger c-di-GMP	49
4.1	Stability and Dynamics of Cyclic Diguanylic Acid in Solution	52
4.2	Dynamics of Analogue endo-S-c-di-GMP in Solution	62
4.3	PleD Dimerization	93
5	Computational Characterization of the Insulin Dimerization	95
5.1	Influence of Mutations at Position B24 on the Stability of the Insulin Dimer	96
6	Conclusion and Outlook	151
	References	153
	Curriculum Vitae	169

1 Overview

Proteins are organic macromolecules which constitute one of the four major building blocks of molecular biology and are essential to all organisms. The majority of protein tasks in living cells are mediated by protein-protein interactions. Protein-protein interactions occur when two or more proteins bind together, often in order to carry out their biological function. Self-association of proteins to form dimers and higher-order oligomers is a very common phenomenon. It is evident that protein dimerization acts as a control tool for the execution of functions in many biological systems. The formation of a dimer can be responsible for enzyme activation. For example, the dimerization of the diguanylate cyclase PleD of *C. crescentus* will lead to synthesizing the bacterial second messenger c-di-GMP, which is involved in the transition of Caulobacter cells from the motile to the sessile form. On the other hand, two c-di-GMP monomers form an intercalated dimer which binds at the inhibition site of PleD and induces product inhibition (see Chapter 4). Conversely, dimerization can inhibit an active monomeric protein, e.g, insulin and lead to disease. Thus, identification of protein-protein interactions is at the heart of functional genomics and prediction of protein-protein interactions is also crucial for drug discovery (see Chapter 5).

The aim of this thesis is to characterize the dimerization and ligand-binding in important biological systems using computational methods. The free energies of the systems which can describe their tendencies to associate and react were investigated

1 Overview

using computer simulations. Parameters which were missing were parametrized before Molecular Dynamics Simulations were employed . Then numerous theoretical tools were used to illustrate the structural changes upon dimerization.

2 Introduction

“Nature has simplicity and therefore a great beauty.”

Richard P. Feynman

2.1 Proteins

Proteins are large polymeric organic compounds made of amino acids, which together with nucleic acids, lipids and saccharides constitute the four major types of biological molecules. Proteins play an essential role in biology. They are involved in nearly every aspect of physiology and biochemistry, carrying out the most important tasks in living organisms. Ultimately it is the three-dimensional structure of proteins that determines their function. Therefore, it is crucial to study the details of the three-dimensional structure of the protein. The basic cornerstone for the three dimensional structure is the amino acid, which consists of an α -carbon, an amine ($-\text{NH}_2$), a carboxyl acid ($-\text{COOH}$), a side chain and one more hydrogen. There are left-handed (L-amino acids) and right-handed (D-amino acids) isomers due to asymmetry of the “ α -carbon”. Only the side chain varies between different amino acids. Two amino acids are connected through a peptide bond, which is an amide bond formed by the reaction of an α -amino group ($-\text{NH}_2$) of one amino acid with the carboxyl group ($-\text{COOH}$) of another. In general, the basic twenty standard amino acids can be classified by the properties of their side

2 Introduction

chain into four groups: hydrophobic, hydrophilic, acidic, and basic. The first two refer to nonpolar and polar side chains, respectively, while the last two may be negatively or positively charged respectively according to the surrounding pH and their pK_a .

The main interactions, i.e., hydrogen bonding and the hydrophobic effect, that drive them to fold into intricate secondary, tertiary and quaternary functional structures (Figure 2.1) are weak compared to their own conformational entropy, which is one of the reasons that proteins often exist in large scale so they have to cooperate with each other to keep themselves in their native structure. Besides, many proteins are believed to be “disordered”, i.e., unstable in solution. The structures of disordered proteins are not “random”, but have a significant residual structure and differ from one another primarily due to the different sequence of amino acids, which results in folding of the protein into a characteristic three-dimensional shape that determines its activity. In the “disordered” state, a protein exists in an ensemble of conformers. It has been found that the main driving forces in folding globular proteins is to pack the hydrophobic side-chains in the interior of the molecule, the so called hydrophobic core; the hydrophilic backbones are hydrogen-bonded, forming secondary structure elements; the hydrophobic side-chains of the α -helices and of the β -sheets often interact to form higher-order secondary structure motifs.^{1,2} The ability to fold provides the basis for the many diverse functions that proteins are responsible for within living organisms, e.g., molecular motors, cell signaling, catalyzing reactions, transporting, transmitting information from DNA to RNA, traversing membranes to yield regulated channels, and forming the building blocks of viral capsids.³

2.1.1 Structure and Function

Most of the three-dimensional macromolecular structure data in the Protein Data Bank⁵ were obtained mainly by: X-ray crystallography ($> 80\%$), solution nuclear magnetic

resonance (NMR technique) ($\sim 16\%$) and theoretical modeling (2%). The first two techniques are experimental methods and provide a resolution at the level of distinguishing individual atoms. In X-ray crystallography, the first step is protein molecule crystallization which is as much an art as science. The X-ray beam is directed on a crystalline sample, the crystalline atoms cause it to diffract into many specific directions. This can produce a three-dimensional picture of the density of electrons within the crystal and provide the 3D model of the protein. However, this technique is limited to molecules that form regular crystals. Membrane proteins and flexible fibril-like proteins are very difficult to crystallize.

Nuclear magnetic resonance (NMR) spectroscopy, relies on the principle that the nuclear magnetic moment is aligned with the electromagnetic field, and elucidates the solution structure of small proteins using chemical shifts (describing the local structure) and nuclear Overhauser effect (NOE) data (measuring short atomic distances). While these two methods share certain similarities, they are two complementary techniques for probing a wide range of structural and dynamical properties of macromolecules. Combination of NMR and X-ray diffraction data is able to obtain more precise models.⁶⁻⁸

Table 2.1: X-ray crystallography and NMR are complementary techniques

X-ray crystallography	NMR
long time scale (s \sim hours), static structure	short time scale (ns \sim s), protein folding
single crystal, purity <100 kDa, domain, complex	solution, purity < 30 kDa, domain
all atomic properties by a Fourier transformation	chemically specific, local structural information by NOE, J-coupling/chemical shift
active or inactive	functional active site
electron density	atomic nuclei, chemical bonds
resolution limit 2-3.5 Å	resolution limit 2-3.5 Å

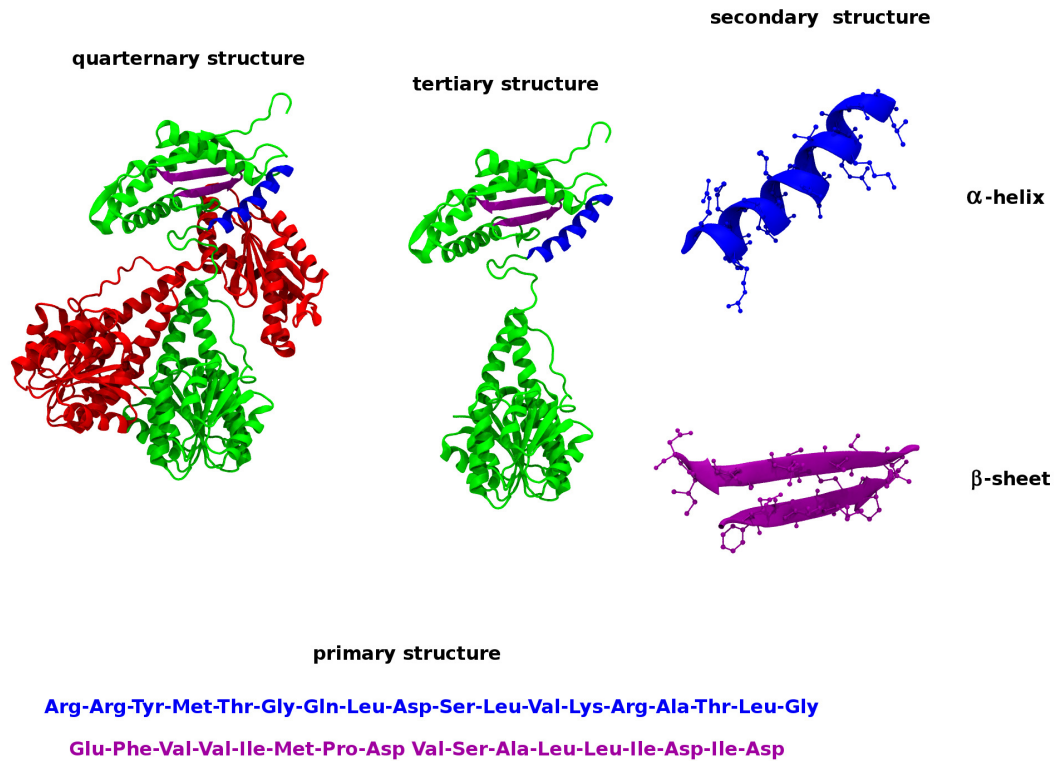


Figure 2.1: Biomolecular structure of the protein PleD. (PDB code:2V0N⁴) Proteins have complex shapes based on four levels of structure: primary structure; secondary structure; tertiary structure and quaternary structure. Primary structure—a protein’s unique linear sequence of amino acids; secondary structure—alpha-helix or beta-pleated sheets; tertiary structure—determined by the interaction of the amino acid’s side groups with their environment, generating the 3-dimensional shape of the protein molecule. The folded areas may be held together by disulfide linkages in some proteins.

A major goal of structural biology is to predict the three dimensional structures of proteins from the sequence,⁹ many proteins are simply too large for NMR analysis and can not be crystallized for X-ray diffraction. Therefore, there is a huge gap between available experimentally determined structures and residue sequences that have been determined. Protein modeling, such as homology modeling is an alternative strategy that is being applied to obtain structural information if experimental techniques fail. Homology modeling is a knowledge-based prediction of protein 3D structures. In homology modeling, a protein sequence with unknown structure (the target) is aligned with one or more protein sequences with known structures (the templates). The method is based on the principle that homologous proteins have similar structure and is a fast tool in drug discovery. Homology modeling obtains more reliable results than pure theory such as *ab initio* modeling. Figure 2.2 shows the common four processes for building homology models: template selection, target-template alignment, model construction, and model quality evaluation.

In the first step, the program/server compare the sequence of the target protein to the template protein in the protein data bank (PDB)¹⁰. The most popular servers such as BLAST¹¹, and FASTA¹² perform the searching and give a list of known protein structures that matches the sequence when the sequence identity between target and template sequences is above 30%. If it is below 30%, alternative strategies based on multiple sequence alignment^{13,14} have to be used. Once a suitable template has been selected, it should be aligned to the target using programs such as t-coffee¹⁵, Expresso¹⁵, PSI-BLAST¹² and PROBCONS¹⁶. The sequence alignment step is crucial for the accuracy of the homology model, as no modeling procedure can recover from incorrect alignment. Based on the template structure and the alignment, in the third step the model for the target protein is constructed (including backbone generation, loop modeling, sidechain modeling, and model optimization). Several methods can be employed such as those based on rigid-body assembly (SWISS-MODEL^{17,18}, Composer¹⁹, 3D-JIGSAW,

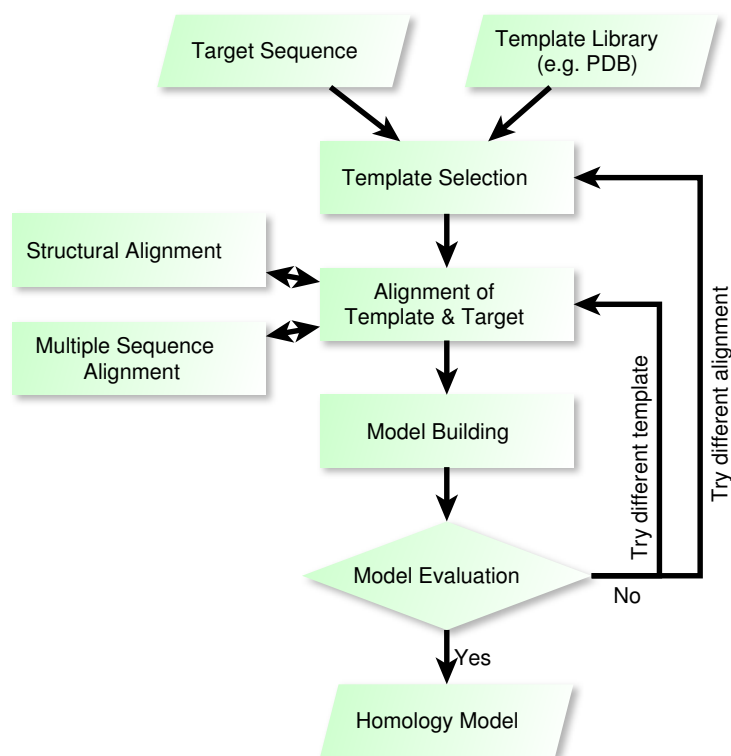


Figure 2.2: The key processes for homology modeling.

PrISM²⁰, CONGEN^{21,22}), or based on spatial restraint, (Modeller²³) whereas side-chains and loops can be modeled with other processes. If no suitable template structure is available for comparative modeling, *de novo* modeling methods also called *ab initio* modeling may be used. There are many factors that will lead homology models to fail, i.e., errors in side-chain prediction, alignment errors, errors in the regions without templates, misalignment, incorrect templates. Therefore, it is necessary to evaluate the homology model. Currently, there are many programs and web servers such as PROCHECK²⁴, WHATIF²⁵, VERIFY3D²⁶ and ANOLEA²⁷ available for assisting the evaluation. Typical applications of a homology model in drug discovery require a very high accuracy of the local side chain positions in the binding site. A large number of homology models have been built over the years, including antibodies²⁸ and proteins in human biology and

medicine^{29,30}.

There are some other techniques to study different aspects of structures of cellular components, such as Hydrogen-deuterium (H/D) exchange, a chemical reaction where a covalently bonded hydrogen atom is replaced by a deuterium atom upon changing the solvent from H₂O to D₂O, or vice versa. FRET detects distance changes in real-time. All of these techniques provide essential tools to determine protein structure and moreover the protein-protein interactions which are at the core of the entire interaction system of any living cell.

2.1.2 Allostery

Allostery is the phenomenon that a change such as binding an effector molecule at one site (allosteric site), affects the activity at another site which will control processes such as signal transmission, catalysis, receptor trafficking, turning genes on or off and apoptosis. Effectors which enhance the protein activity are referred to as allosteric activators, while those that decrease the protein activity are called allosteric inhibitors. The distance between an allosteric site and the functional part of the protein could be several tens of Ångströms. The classical allosteric views such as Monod-Wyman-Changeux (MWC)³¹, and Koshland-Némethy-Filmer (KNF)³² described allosteric regulation via a conformational change. The MWC model described the transition as a concerted action between two co-existing, discrete states (R and T); and the KNF model formulates it as a sequential, induced conformational change by the binding event at the first site which is responsible for the allosteric effect. The old views indicate that: first, there are only two states R and T which exist in a ratio governed by an equilibrium constant when the ligand is absent; second, allostery involved a shape change in the substrate binding site; third, the allosteric signal is transmitted via a single pathway.^{33,34} The new

2 Introduction

views indicate that: first, proteins exist in ensembles instead of just two conformational states; second, allostery is a thermodynamic phenomenon and can be driven by enthalpy, enthalpy and entropy, or entropy. Allostery can work without a change in shape. Third, the existence of multiple conformational and dynamic states implies multiple pathways through which the strain energy is released from the allosteric site. If the enthalpy change does not reverse the free-energy change due to the change in entropy, entropy may be the factor responsible for the ligand binding.³³⁻³⁹

2.2 Protein-Protein Interactions

The majority of protein tasks in living cells are mediated by protein-protein interactions. Protein-protein interactions occur when two or more proteins bind together, in order to carry out the biological function. Protein interactions have been studied from the perspectives of biochemistry, quantum chemistry, molecular dynamics, chemical biology, signal transduction and other metabolic or genetic networks. If one can identify the function of at least one of the components with which the protein interacts, its function pathway can be assigned³. Through the network of protein-protein interactions, it is possible to map cellular pathways and their intricate cross-connectivity. Identification of protein-protein interactions is at the heart of functional genomics. The types of protein-protein complexes can be classified as homodimeric proteins, heterodimeric proteins, enzyme-inhibitor complexes and antibody-protein complexes.⁴⁰ Prediction of protein-protein interactions is also crucial for drug discovery.^{3,41} In order to predict protein-protein interactions, it is necessary to figure out the chemical and physical features of the associations, including the shape complementary to the organization and physical/chemical contributions to their stability. In the next section, we will introduce protein dimerization, which is a subset of protein-protein interactions.

2.2.1 Protein Dimerization

In biochemistry, a dimer is a macromolecular complex formed by two, usually non-covalently bonded, macromolecules like proteins or nucleic acids, and can be classified as homodimer or heterodimer.⁴² Homodimers, which are present in abundance in the Protein Data Bank (PDB⁵), are the simplest case of non-covalent self-assembly in proteins. Dimerization is a subset of protein-protein interactions. The self-assembly of proteins to form dimers and higher oligomers is a common theme. Self-assembly can help to minimize genome size while maintaining the advantages of modular complex formation.⁴³ It was proposed that there are three pathways for the evolution of dimers: 1) formation of a functional dimer directly without going through an ancestor monomer. This kind of homodimeric proteins are permanent assemblies and their polypeptide chains assemble at the time they fold. 2) formation of an energetically stable monomer and then mutating its surface residues. This kind of complex involves proteins that fold separately and remain in monomeric forms until they meet and associate. 3) a domain swapping mechanism, replacing one segment of the monomer by an equivalent one from an identical chain in the dimer.^{44,45} Homomultimeric proteins are responsible for the diversity and specificity of many pathways, ion channels, activities of enzymes, mediation and regulation of gene expression, receptors, cell adhesion processes and so on.

It is evident that protein dimerization acts as a control tool for the execution of functions in many biological systems. The formation of a dimer can be responsible for enzyme activation. For example, PleD is a protein that is involved in the transition of *Caulobacter* cells from the motile to the sessile form. It has been suggested that the activation can be triggered via the dimerization of the D1/D2 domains in PleD monomers and bring two DGC domains close as a condition for the condensation reaction to occur.⁴ Conversely, dimerization can inhibit an active monomeric enzyme, e.g. insulin. Insulin, a protein that controls glucose uptake in cells, performs the biological function in its

2 Introduction

monomeric form. However, under physiological conditions, the monomers are ready to aggregate to dimers and lose their critical biological function.⁴⁶

Protein-Protein Interface Proteins interact through interfaces, which consist of residues or fragments that belong to two different subunits. Figure 2.3 is one of the examples of protein-protein interfaces. The subunits of a multimeric protein are identical in a homomultimeric protein, whereas in a heteromultimeric protein they are different. There are several fundamental properties people use to characterize protein-protein interfaces.⁴⁰

1) Compared to a transient interaction which is continuously forming and dissociating, the interface of obligated interaction is larger, more conserved and tends to have more hydrophobic residues, whereas transient interfaces consist of more polar residues.⁴⁷ 2) The interfaces of homodimers on average are more hydrophobic and bury twice as much protein surfaces as in complexes. 3) For “weak” transient homodimers, monomers and dimers can exist at physiological concentration with dissociation constants in the micro-molar scale. The weak homodimers have smaller contact areas between protomers and the interfaces are usually more planar and polar. In contrast, the “strong” transient dimers often experience large conformational changes upon association/dissociation and have larger, less planar and sometimes more hydrophobic interfaces.⁴⁸ 4) If the interface is larger than 1000 \AA^2 , the complex will undergo conformational changes upon complexation.^{49,50}

Besides, there are several criteria that are used to define interface residues:¹ a) the distance between two CA atoms belonging to each chain respectively is less than 9.0 \AA , b) the distance between any two atoms of two residues from the different chains is less than 5.0 \AA , c) the van der Waals energy between the residues is less than -0.5 kcal/mol , d) all atoms or amino acid residues in the monomer that lost more than 0.1 \AA^2 solvent accessible surface area in the dimer are regarded as interface atoms or residues.⁵¹ Criteria have been defined for these residues to be considered as belonging to the interface.

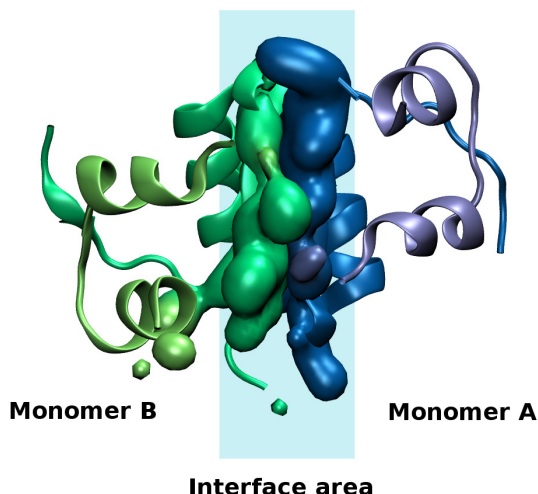


Figure 2.3: Illustration of protein – protein interface. The figure represents two interacting proteins (insulin monomer, PDB code: 4INS). In monomer A: chain A is colored in light blue and chain B is in dark blue. In monomer B: chain A is colored in light green and chain B is in dark green. Interacting residues from the two monomers are shown with surface representation while the rest of the proteins are illustrated with ribbon representations.

The chemical properties involved in these residues are key factors to understand and determine the architecture of the interfaces.⁵²⁻⁵⁵

Protein-Protein Interaction Forces Protein-protein interactions are mainly driven by :

Covalent Bonds (~ 60 kcal/mol) Chemical bonds due to the sharing of electrons pairs between atoms; short-ranged, directional and strong. For single, double and triple carbon-carbon bonds, one finds approximate energy values of 83, 142 and 196 kcal/mol respectively.

Electrostatic Interactions (~ 6 kcal/mol) Such as ion pairs and salt bridges. Attractive electrostatic interactions between oppositely charged residues, i.e. nitrogen

2 Introduction

atoms in the side-chains of His, Arg and Lys as positively charged groups and oxygens in the side-chains of aspartic and glutamic acids as negatively charged groups. A distance of less than or equal to 4 Å between the involved atoms is typical for these interactions.¹ “The energy penalty paid due to the desolvation of the charged residues may not be recovered by favorable interaction among the charged residues [...]. Calculating the electrostatic field in a protein molecule correctly is thus akin to hitting a moving target using a shotgun with a bent barrel while being in the middle of an earthquake.”⁵³

Hydrogen Bonds ($\sim 1.8\text{--}6$ kcal/mol) H-bonds are attractive, relatively weak, non-bonded interactions between a hydrogen atom (“donor”) and an highly electronegative atom (“acceptor”), e.g. nitrogen or oxygen. The electron belonging to the hydrogen is strongly pulled towards the oxygen; the hydrogen atom must be covalently bonded with an electronegative atom and can bond with another electronegative atom of another molecule.⁵⁶

van der Waals (vdW) Interactions ($\simeq 0.6$ kcal/mol) The van der Waals interactions are formed by fluctuations in the electric dipoles of two atoms. The charge in one atom will induce a dipole in the other atoms, which makes the two atoms attract each other. It is a weak interaction, the interaction energy is comparable to and often lower than the thermal vibrational energy (~ 0.6 kcal/mol). When two molecules are near each other in a liquid, the distance of closest approach (vdW radii) is generally greater than the sum of their covalent radii.⁵⁷

Hydrophobic Effects ($\simeq 0.6$ kcal/mol) A primarily entropic effect arising from the rearrangement of hydrogen bonds (loss of water entropy) between solvent molecules (i.e. water) around non-polar solute (demonstrated in Figure 2.4). This is the entropic driving force for self-association of non-polar groups in water.⁵⁸

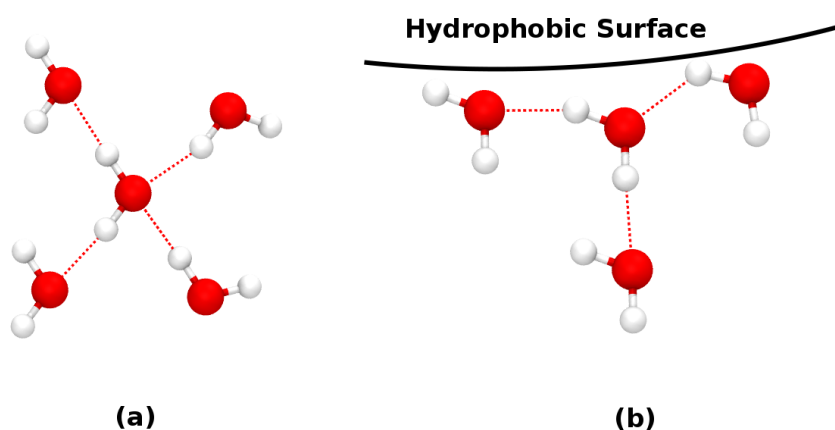


Figure 2.4: Hydrophobic effect. (a) Bulk water molecules without any constraint which are highly mobile and have a high state of entropy which is favorable. (b) Hydrophobic protein surface - surface waters are “frozen”. They reorient and place the hydrogen bonding sites facing the solvent. They are less mobile and the entropy is decreased, which costs energy. Adapted from⁵⁸.

S-S Bridge (~ 60 kcal/mol) Covalent bond between pairs of cysteins in many native structures of proteins. Prerequisite for proper folding and biological function; thermodynamic stabilization of the native structure by forming conformational constraints within the protein.^{58,59}

Compared to covalent bonds, all of the noncovalent interactions are known to be relatively weak interactions. However, small stabilizing interactions can add up and lead to significant contributions to the overall stability of a conformer. Reversible protein aggregation usually results from non-covalent protein interactions while covalent aggregation benefits from formation of a chemical bond, such as a disulfide bridge.⁶⁰

Binding Hot Spots Protein-protein interactions are critically reliant on just a few ‘hot spot’ residues at the interface. Hot spots make a dominant contribution to the binding free energy and can disrupt the interaction if they are mutated to alanine.⁶¹ Chothia and Janin⁶², found that the binding hot spot may be any H-bond or ion pairs in the interface. Mutation of one of the residues forming the ion pair on one side to alanine will cause its partner on the other side to remain unpaired and decrease the binding energy.⁶³ However, a study by Wells found some exceptions. The mutation of amino acids that make important hydrogen or ionic bonds across the interface to alanine caused only minimal effect on the binding energy. One possible explanation was that after removal of the hydrogen acceptor, the donor can shift and find another acceptor to form a new hydrogen bond. At present, two algorithms are considered to be fairly accurate in predicting the amino acids which will be a hot spot.^{64,65} Both are based on the computational calculation results of binding free energy between the interface by decomposing in terms of H-bonds, ionic interactions, vdW interactions and desolvation of part of the amino acids. In general, a hot spot amino acid is defined as the one which decreases ΔG by 2 kcal/mol when mutated to alanine.⁶¹

Point Mutations A point mutation, or substitution, is a type of mutation. It can be induced by chemicals or malfunction of DNA replication, by exchanging a single nucleotide for another.⁶⁶ As the difference between the original and the new amino acid is the side chain, for computational modeling techniques, it can essentially be yielded by removing or replacing the side chain but conserving the backbone atoms, (i.e. for a mutation to alanine, the side chains need to be removed and only the beta carbon kept). Proteins are marginal, mutation of some residues, especially on the hot spot, may change the thermodynamic equilibrium. This may make the protein either less stable, which

lowers the effective concentration of protein and affects its biochemical function^{67,68}, or more stable, which causes a loss of flexibility and increases the rigidity. The lower flexibility of a protein will affect ligand binding, allosteric effects and degradation.⁶⁹ Computational modeling can be used as a tool to understand and predict the effects of mutations. It is obvious that rearrangements of the protein structure are necessary to accommodate the changes of amino acid size owing to mutation. It has been repeatedly observed that proteins are surprisingly robust to site mutations, and can endure multiple substitutions with little change in structure, stability, or function.⁷⁰ Mutations of key amino acids or hot spot residues will induce rearrangement of the protein and affect the stability or the enzyme's specificity. The mutations will affect the protein structure by a) disruption of the hydrophobic core through over-packing using a large side chain instead of the original small chain, or through cavity formation by replacing a larger side chain with a smaller one, or putting a charged/polar residue in the core; b) removal of residues forming disulphide bridges, or salt bridges, or polar interactions, or hydrogen bonding partners; c) replacement of charged/polar residues on the surfaces with hydrophobic ones; d) mutations at ligand binding sites, catalytic sites, or allosteric sites, or other sites of specific function in proteins.

2.2.2 Thermodynamics and Protein Dimerization

The native protein conformation must be energetically stable. From a thermodynamic point of view, for protein folding, the major stabilizing contributions are the hydrophobic effect and hydrogen bonds, while the major destabilizing contributions to the stability of the folded state is the conformational entropy of the polypeptide chain. Therefore, the Gibbs free energy barrier for folding is determined by the unfavorable loss in configurational entropy upon folding and the gain in stabilizing native interactions. For example, the transformation of a long chain to a specific compact structure results in

2 Introduction

significant entropical loss due to the restricted motion of the backbone and the side chains. Under physiological conditions, proteins exist in their native structure because the favorable enthalpic term arising from the solvent and protein interactions exceeds in magnitude the unfavorable entropic term. The free-energy difference between the folded and the unfolded state, is marginal and on the order of 5–10 kcal/mol.⁵⁸ Since the pioneering work of Anfinsen,⁷¹ protein folding has become a popular subject in statistical physics. The energy landscape theory^{72–74} has opened a new research direction for protein folding. It suggests that the most realistic model of a protein is a minimally frustrated heteropolymer with a funnel-like landscape biased towards the native structure. Other recent reviews have been published.^{75,76}

Protein aggregation is affected by environmental conditions, e.g. temperature, pH and the concentration of components.⁷⁷ A dimer can be formed spontaneously in a thermodynamically stable state when the two proteins are confined in a small enough region, i.e. being surrounded by other macromolecules.⁷⁸ Compared to protein folding, electrostatics can enhance association rates while destabilizing the final complex.^{79–82} Hydrogen bonds and ion pairs contribute more to the stability of protein binding than to protein folding, while the contribution of the hydrophobic effect to protein-protein associations is not as strong as in protein folding.^{1,81,83–85}

It is worth mentioning that in protein dimerization, when two subunits come together and form the interface, the layer of “frozen” water (see Figure 2.4) will release from each surface and become mobile to form H-bonds in all directions, thus, increasing the entropy and having a lower (more favorable) free energy. The hydrophobic interaction here is not a positive attraction of the two hydrophobic surfaces, but the water that drives the subunits together. The interaction (van der Waals) between the hydrophobic surface and the water was approximately as favorable as that with the other hydrophobic surface,

and water molecules have a much stronger interaction with each other than with the subunit interface.

2.2.3 Diseases Caused by Protein Dimerization

Protein misfolding and aggregation is the reason for many protein conformational diseases, including neurodegenerative (e.g. Alzheimer's, Huntington's, and Parkinson's diseases, familial British and Danish dementias), systemic (e.g. type II diabetes, light chain amyloidosis) and other (e.g. cystic fibrosis) diseases.^{76,86,87}

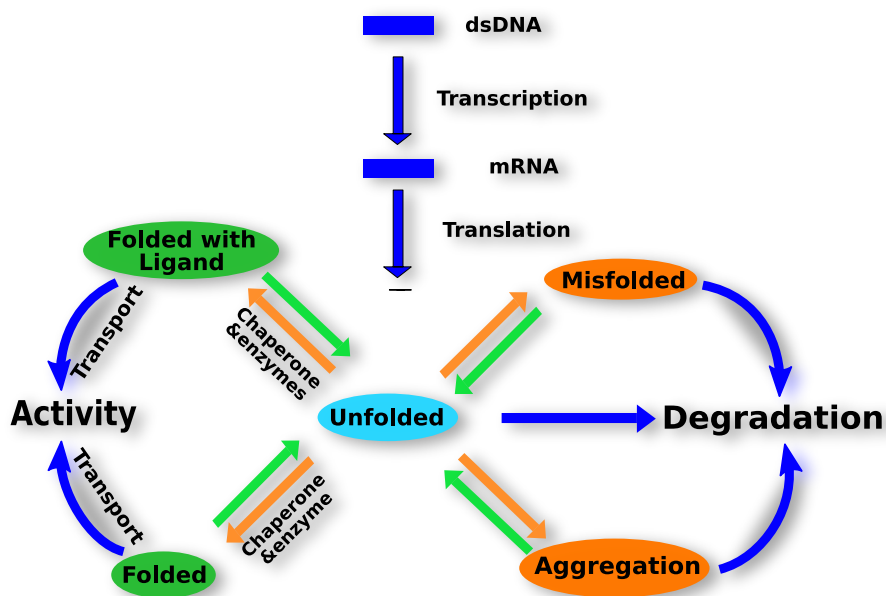


Figure 2.5: A proteostasis network comprising pathways represented by the arrows. Imbalances in proteostasis often lead to disease. Potential therapies of diseases of proteostasis can be yielded by shifting the equilibria toward active forms with small ligands, or replacing aberrant proteins, or modulating the pathways with agents that influence pathways such as transcription, translation, degradation and translocation using molecules like siRNAs to modulate concentrations of chaperones, disaggregates and signal pathways. Adapted from⁸⁸.

2 Introduction

Diabetes is a very common disease which can be very dangerous if not treated. There are two types of diabetes. Type I is an autoimmune disease in which the immune system attacks its own insulin-producing cells by mistake, so that insufficient amounts of insulin are produced. Type II is insulin resistance, which means that the cells do not react to insulin the way they are supposed to. Insulin is a hormone, which is produced in the pancreas as a hexamer and is a main regulator of the glucose levels in the blood. When we eat, glucose levels rise, and insulin is released into the bloodstream as a monomer. The insulin acts like a key, opening up cells so they can take in the sugar and use it as an energy source. Usually insulin dissociates from its hexameric storage form through an intermediate dimer state to the bioactive monomer before binding to its transmembrane insulin receptor. The interface which the insulin monomer uses to bind to the receptor, is the same one that forms dimer and hexamer. Thus, once the monomers form the dimers or hexamers, they can not bind to the receptors and subsequently, lose their biological function which leads to diabetes. Therefore, understanding the dynamics of insulin dissociation is critical for devising formulations for the treatment of insulin-dependent diabetes which will be reported in detail in a later chapter.

Bacterial biofilm formation is involved in life-threatening infectious diseases, such as cystic fibrosis, or the colonization of medical devices. Cyclic diguanosine-monophosphate (*c*-di-GMP) is a ubiquitous second messenger that regulates cell surface-associated traits in bacteria and thus is important for biofilm formation. It is produced from 2 molecules of GTP by the activity of diguanylate cyclases (DGCs) and is degraded by specific phosphodiesterases (PDE) into 5'-phosphoguananylyl-(3'-5')-guanosine (pGpG); pGpG is subsequently split into two GMP molecules. Through the interaction with different receptors, such as PilZ- containing proteins, *c*-di-GMP negatively modulates cell motility and traits associated with bacterial virulence and stimulates several biofilm-associated functions. *c*-di-GMP signaling has been important for the development of anti-biofilm or

anti-virulence drugs. The concentration of *c*-di-GMP influences the biofilm formation or dissolution and thus cause diseases or not. *c*-di-GMP is ready to form dimers, or higher aggregates in solution. Analogues of *c*-di-GMP have been designed, and it was found that they selectively target binding proteins.⁸⁹ This is helping to study *c*-di-GMP signaling in bacteria and may become lead compounds for the design of anti-biofilm agents. Chapter 4 will present the investigations in detail.

Development of new therapeutic solutions is an expensive and time-consuming process. Progress in computer power makes it possible to simulate systems involving protein-ligand and protein-protein interactions with millions of atoms. This opens the possibility to tackle more physiologically relevant biological problems, and computational simulations become an essential tool in modern drug design and development. In the next section, some of these simulation techniques available for protein-protein interaction are reviewed.

2.3 Computational Simulations Methods

“...all things are made of atoms, and that everything that living things do can be understood in terms of the jiggings and wiggings of atoms”

Richard P. Feynman

2.3.1 Atomistic MD Simulations

Molecular dynamics (MD), first developed in the late 1970s, is a computer simulation of physical movements of atoms and molecules over time, according to Newton’s second law of motion

$$F_i(t) = m_i a_i(t) \quad (2.1)$$

where F_i is the force exerted on particle i , m_i is the mass of particle i and a_i is the acceleration of particle i . The force can also be expressed as the gradient of the potential energy

$$F_i = -\nabla_i V \quad (2.2)$$

Combining the above two equations yields

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \quad (2.3)$$

where V is the potential energy of the system and describes how the particles in the simulation interact with each other and with the environment. Newton’s equation of motion can then relate the derivative of the potential energy to the changes in position as a function of time.

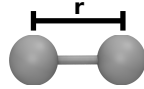
The forces between the particles and potential energy are defined by molecular mechanics force fields. Classical force fields consider the contributions of various atomic forces that govern molecular dynamics as the interactions between atoms, including

bonds, angles, dihedral angles, electrostatic and van der Waals terms

$$V_{ff} = \sum V_{bonds} + \sum V_{angles} + \sum V_{dihe} + \sum V_{elec} + \sum V_{vdW} \quad (2.4)$$

where each term can be expressed as a function of the atomic coordinates

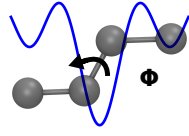
$$V_{bond} = \sum k_r (r - r_e)^2 \quad (2.5)$$



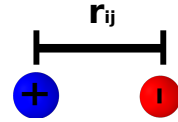
$$V_{angle} = \sum k_\theta (\theta - \theta_e)^2 \quad (2.6)$$



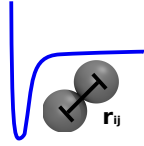
$$V_{dihe} = \sum k_\phi (1 + \cos(n\phi - \delta)) \quad (2.7)$$



$$V_{elec} = \frac{1}{4\pi\epsilon_0} \sum \frac{q_i q_j}{r_{ij}} \quad (2.8)$$



$$V_{vdW} = \sum \epsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^6 \right] \quad (2.9)$$



The first three terms represent the “bonded” interactions while the latter two describe the “non-bonded” interactions. Chemical bonds and bond angles are modeled using simple springs (Eq. 2.5 -2.6) and dihedral angles (Eq. 2.7) are modeled using a sinusoidal function. Non-bonded forces arise due to van der Waals interactions, modeled by the Lennard-Jones potential (Eq. 2.9), and electrostatic interactions, modeled by Coulomb’s law (Eq. 2.8). k_r , k_θ , k_ϕ is the bond, angle, and dihedral angle force constant, respectively. r_e and θ_e are equilibrium values, n is the periodicity of the dihedral and δ is the phase which governs the position of the maximum. q_i , q_j are the partial charges on atoms i and j and ϵ_0 is the vacuum dielectric constant. ϵ_{ij} is the Lennard-Jones well depth, $R_{min,ij}$

2 Introduction

is the finite distance at the Lennard-Jones minimum. r_{ij}^{-12} is an empirical repulsive term, describing Pauli repulsion at short ranges due to overlapping electron orbitals. The r_{ij}^{-6} term describes the attraction at long ranges and can be obtained from theory. The functional form of the attractive term has a clear physical meaning, which the repulsive term lacks. The exponent value of the former is chosen for computational convenience. In principle, the nonbonded energy terms between every pair of atoms which are more than two bonds apart or belonging to different molecules, should be evaluated. However, the details of nonbonded treatment depend on the system and the method used for the system. In general, there are two main ways to handle the nonbonded interaction. One method is to truncate the interactions at a pre-defined distance, which is faster but less accurate. Another one is using Ewald summation,⁹⁰ which is more accurate.

Eq. 2.4 shows a minimal model for force fields. There are several classical force fields commonly used in molecular dynamics simulations, including AMBER⁹¹, CHARMM^{92,93}, OPLS^{94,95}, and GROMOS⁹⁶, which are often used with a software that bears the same name. All of them are parameterized in a different way but generally give similar results. Force field parameters can be obtained from *ab initio* calculations or by fitting of calculated system properties to experimental data. Compared to Eq. 2.4, some force fields add more terms, such as improper dihedral terms and the Urey-Bradley (1 - 3 bond length) terms in CHARMM, to improve the description of vibrational spectra and out of plane motions. Some more terms were introduced, such as CMAP.^{97,98} A number of studies have demonstrated good agreement between computational and experimental measurements of macromolecular dynamics.

2.3.2 Accuracy Improvement and Other Methods

Although molecular dynamics simulations have achieved significant successes in studying biological systems, the utility is still limited mainly by two aspects: a) the force fields require further refinement to be more accurate^{99,100}; b) the high computational demands prohibit simulations longer than microsecond timescale,¹⁰¹ which may cause inadequate sampling of conformational states in many cases, e.g. binding pocket configuration in drug design.

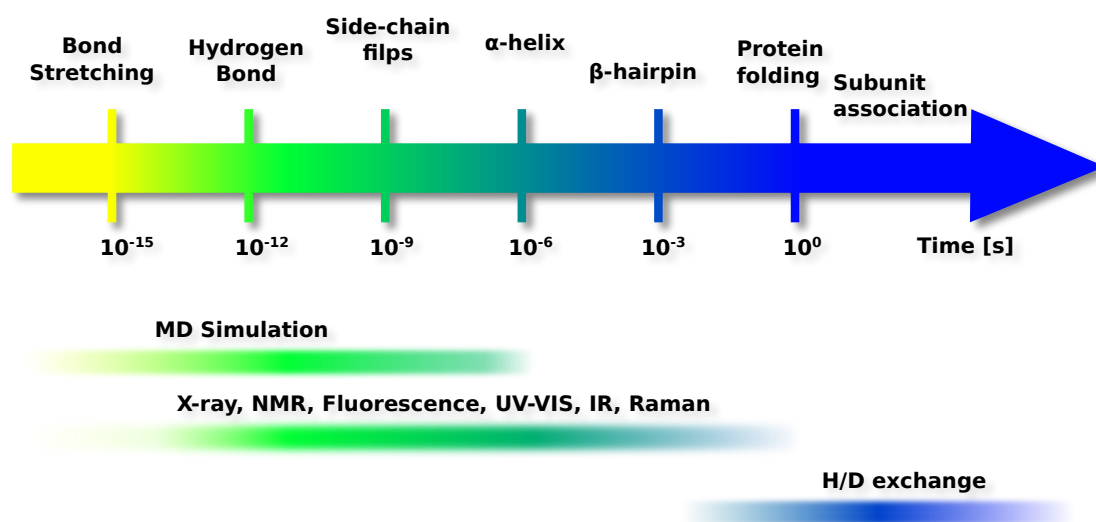


Figure 2.6: Several timescales involved in protein association. The color changes in the axis represent the different timescale and amplitude [\AA] of the protein motions. Yellow and light green: Local motion with the amplitude $0.001\text{--}0.1$ \AA . Light green: Medium-scale motion with the amplitude $0.1\text{--}10$ \AA . Dark green: Large-scale motion with the amplitude $1\text{--}100$ \AA . Light blue: Large-scale motion with the amplitude $10\text{--}100$ \AA . Dark Blue: Global motion with amplitude max 10 \AA .

To overcome the challenges in force field refinement (besides more accurate parameters, i.e. force constants and partial charges), better functional forms for describing the potential are essential. Using Eq. 2.8 - Eq. 2.9, the long-range intermolecular interactions are simply expressed by a coulomb potential (Eq. 2.8) instead of electrostatics, induction

2 Introduction

and dispersion. Although dispersion can be approximately covered by the Lennard-Jones potential attractive term Eq. 2.9, induction is completely neglected. To address these problems, for bonded terms, anharmonic functional potentials instead of the simple harmonic potentials were introduced. For the nonbonded terms, e.g., multipole moments can be used to replace the point charges in the electrostatic interaction and a Morse-potential can be used for describing covalent bonds. Additionally, new terms for missing types of interatomic interactions can be added, i.e. inter- and intramolecular electronic polarization. So far, a generally accepted polarizable force field has not been available but is under development,^{102,103} and future implementations will lead to improvement of accuracy.

For classical MD simulation, the most CPU intensive task is to evaluate the potential (force field) as a function of the particles' internal coordinates. The most expensive part in energy evaluation is the nonbonded term which is scaled by $O(n^2)$. Another factor for the cost of CPU time is the size of the integration timestep, the time interval between evaluations of the potential energy which needs to be chosen roughly 10 times smaller than the fastest vibrational frequency in the system (in the range 1 - 2 fs for atomistic resolution). The timestep value can be extended by using algorithms, i.e. SHAKE,¹⁰⁴ which fixes the vibration of the fastest atoms (hydrogens). Aside from algorithms, novel hardware has been used and made great progress towards overcoming the time-scale limitations of classical MD simulation, i.e. graphics-processing-units (GPUs) which can be developed to speed up the MD simulation by an order of magnitude¹⁰⁵⁻¹⁰⁷, or special purpose MD hardware like Anton¹⁰¹ and MDGRAPE¹⁰⁸.

ab initio MD simulation In *ab initio* molecular dynamics, quantum mechanical methods are used to calculate the potential energy of a system “on the fly” for conformations

in a trajectory. The Born-Oppenheimer approximation, as the basis of *ab initio* MD simulation, is used to solve Schrödinger equation by separating nuclear and electronic wavefunctions. In Born-Oppenheimer MD simulations, the motion of electrons (by solving Schrödinger equations) and motion of nuclei (by solving Newtonian equations) is solved at the same time

$$\hat{H}_e|\Psi_0\rangle = E_0|\Psi_0\rangle \quad (2.10)$$

$$m_i \frac{d^2 r_i}{dt^2} = -\frac{dE_0}{d\vec{r}_i} \quad (2.11)$$

where H_e is the electronic Hamiltonian, Ψ_0 is the wavefunction, and E_0 is the eigenenergy. In these simulations, the electronic Schrödinger equation is solved at each step and the nuclei are propagated classically on the potential energy surface. The most computationally time consuming part at each timestep is the calculation of the electronic structure which can be handled by different levels of methods, such as density functional theory (DFT),^{109,110} Møller-Plesset second order (MP2),¹¹¹ and semi-empirical methods with different levels of basis set leading to different accuracy of the results. It is obvious that *ab initio* simulation can yield more accurate results, however their use is severely limited to very small systems (i.e. ~ 100 atoms) due to the intensive computational cost.

QM/MM In quantum mechanics/molecular mechanics (QM/MM) methods, the “active site” is treated by QM methods and the rest of the system is handled by MM force fields. Therefore, within the QM region, the motions of atoms are described by *ab initio* methods as described above, while for the MM region, the motions of atoms are propagated by an empirical potential. All the methods are compromised between accuracy and speed.^{112–114}

2 Introduction

Coarse-graining As mentioned above, molecular modeling can provide insight in biological systems in all-atom detail within the limitations of simulation time and system sizes of less than 100 ns and 10 nm respectively. To address problems arising from larger time-scale and length-scale, or the time-scale and length-scale gap between computational and experimental methods of studying biological systems, alternative techniques are needed. A possible way to extend molecular modeling and bridge it with experimental techniques is using coarse-grained methods. Compared to an all-atom description, coarse-graining methods represent a system by a reduced number of degrees of freedom, such as using “pseudo-atoms” to represent groups of atoms instead of every atom explicitly. Computationally, coarse-graining has great advantages: a) with a smaller number of “pseudo-atoms” or beads, it decreases the computational requirements and accelerates the speed of molecular dynamics simulations. b) the coarse-graining potentials tend to be softer than the atomistic one, so that larger integration time steps can be used. c) coarse-graining reduces molecular friction and therefore smoothes the free energy landscape which makes the dynamics faster. The reduced representations, so called coarse-graining models, such as discontinuous molecular dynamics (CG-DMD)¹¹⁵ and Go-models^{116,117} can be used to tackle the problem. In coarse-graining methods, parameterization (potential of interaction, i.e. van der Waals interactions with other groups have the proper distance-dependence, and the same for the bonds, angles, dihedrals etc), mapping (between atoms and beads, i.e in the MARTINI^{118–120} force field, the beads can be used in normal size 4:1 mapping or small size 3:1 mapping) is not easy to yield. The results need to agree with experimental data or all-atom simulations. Compared to other molecular modeling methods, coarse-graining provides substantial savings in computer time, but at the cost of accuracy and lose the microscopic information. The higher the level of coarse-graining that is used, the lower the accuracy of the result can be. Overall, coarse-graining, as a complementary tool to experiments and atomic simulation, can identify the important degrees of freedom and provide a big picture of the main structural

2.3 Computational Simulations Methods

mechanisms in biological systems. The applications of coarse-graining methods include in protein folding, liquid crystals, packaging of DNA, RNA structure in the ribosome, etc.

3 Determination of Protein-Protein Binding Affinities

For the understanding of protein aggregation it is important to be able to predict protein-protein binding reliably, but it is a challenging task. Protein-protein binding can be obligate, meaning that the subunits are not observed on their own *in vivo*, or transient, meaning continuously forming and dissociating *in vivo*.^{3,49,121} From a physical chemical standpoint, any two proteins can interact. The question is under what conditions and at which strength. The Gibbs free energy upon complex formation — also called binding free energy — can be used to assess how stable the interactions are. Experimentally, the Gibbs free energy can be evaluated via the equilibrium constant of a reaction. In this chapter, first the experimental approaches for equilibrium constant determination are introduced, then high throughput screening which is used in drug discovery is briefly presented. Computational methods such as virtual screening (docking) are introduced and then the focus will be on the binding free energy calculation based on MD simulation. Protein-protein interactions can be treated similarly to protein-ligand interactions, where one protein takes the role of the ligand.

3.1 Experimental Approaches

3.1.1 Experimental Determination of Binding Affinities

Numerous binding assays have been designed to determine binding affinities experimentally. They can be classified into separation assays¹²² and direct assays¹²³ according to whether they require separation of the components for analysis. In most experiments, the binding affinity is given as a dissociation constant K_d . For the aggregation of two proteins P, one is considered as ligand L. K_d is thus defined as



$$K_{eq} = K_a = K_d^{-1} = \frac{[PL]}{[P][L]} \quad (3.2)$$

where K_{eq} is the equilibrium constant for the binding, K_a the association and K_d the dissociation constant. The range of K_d values is very wide in protein-protein interactions, ranging from Micromolar to Picomolar and resulting in free energy changes (ΔG) of -6 to -19 kcal/mol.³ $[PL]$ is the concentration of the protein-protein complex and $[P]$ respectively $[L]$ are the concentrations of the free proteins. Upon titration of a protein to an excess concentration of another (see Figure 3.1), the free concentration of one protein, the ligand, reaches the value of K_d when the receptor binding sites are half saturated with ligand (see Figure 3.1). The value of K_d is the maximal specific binding, which is equal to the half maximal inhibitory concentration (IC_{50}).

The binding free energy (ΔG) of the aggregation of two proteins can be calculated from the equilibrium constant K_{eq} .

$$\Delta G_{\text{bind}} = -RT \ln K_{eq} = \Delta H_{\text{bind}} - T \Delta S_{\text{bind}} \quad (3.3)$$

where R is the gas constant and T is the absolute temperature.

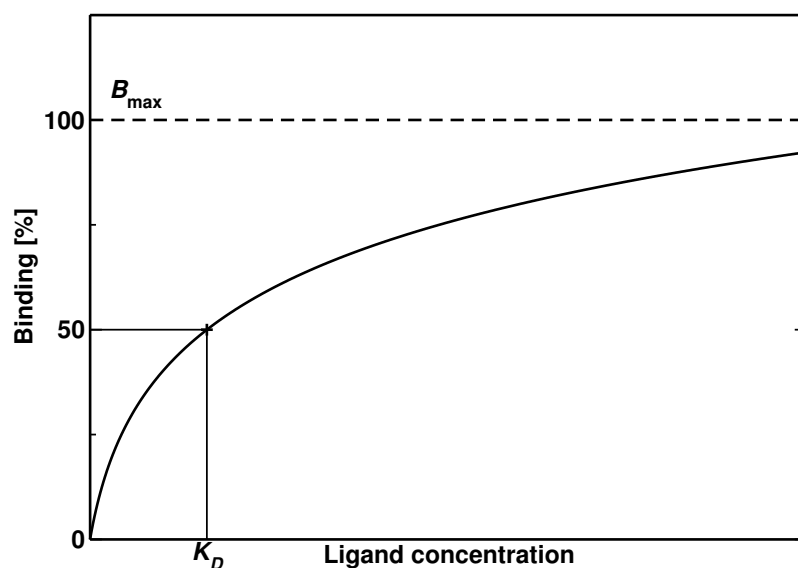


Figure 3.1: This ligand binding graph displays the ratio of bound protein vs. the total ligand concentration. The K_d is at the ligand concentration where 50% of the ligand is bound.

The thermodynamics of ligand binding can be measured directly by Isothermal titration calorimetry (ITC), one of the latest techniques to be used in characterizing binding affinity of ligands for proteins. ITC can directly determine the binding free energy ΔG_{bind} , the enthalpy changes ΔH_{bind} and the entropic component ΔS_{bind} .

3.1.2 High-throughput Screening

High-throughput screening¹²⁴ is a drug-discovery process widely used in the pharmaceutical industry. It utilizes robotics, data processing and control software, liquid handling devices, and sensitive detectors to quickly assay the biological or biochemical activity of large libraries of compounds, with easily thousands of molecules in a library. The goal is to identify compounds that interact with the target protein and provide a starting point for drug design. Usually, the results are organized in a list sorted according to activity level, and a threshold value above which the compounds are considered active will be

3 Determination of Protein-Protein Binding Affinities

chosen for further analysis. These compounds are called hits. After hit identification, the selected hits are tested in a new and more focused screening, i.e. testing the activity based on the concentration of the compound and calculating the maximal inhibitory concentration IC_{50} . Assay hits which have well-behaved titration curves and IC_{50} values of typically less than 10 μ M are subjected to the next step studies, such as validation assays and selectivity tests. This process is usually done manually with only a small number of molecules to test.¹²⁵

3.2 Computational Methods

3.2.1 Virtual Screening

To reduce the number of compounds that have to be synthesized and tested in HTS, computational methods such as virtual screening are used. In virtual screening, molecules from large libraries of available compounds are docked computationally into the binding site of the protein target and their binding is evaluated.

To identify compounds that are potential drug leads, the binding energy between the protein and compounds is calculated and those that have the most favorable interaction are selected for further analysis. The drawback of such virtual screening methods is that the algorithms currently used have to make a number of approximations in order to be able to screen large numbers of compounds in a reasonable time, which results in less accurate description of the binding energies. One way to overcome these limitations is to use more accurate force field based methods to determine the binding energies of the compounds and thereby more reliably reject those poses that don't have a favorable binding energy.

3.2.2 Docking and Scoring

In molecular modeling, docking is a method which aims to predict the preferred orientation of one molecule to a second when bound to each other to form a stable complex. Docking can be between protein/small ligand, protein/protein, protein/peptide, protein/nucleotide. In general, there are two aims of studies in high-throughput docking which is used as a hit identification tool.¹²⁶ The first one is to identify the ligands by virtual screening (docking), including finding possible poses, and orientation of a ligand to fit the active site of the macromolecular target. The second is to predict the binding affinities of the binding modes (scoring). Docking is generally devised as a multi-step process and begins with the application of docking algorithms that pose small molecules in the active site. Algorithms are complemented by scoring functions which evaluate the interactions between the compounds and the targets. The success of a docking program depends on both components: the search algorithm and the scoring function.¹²⁷

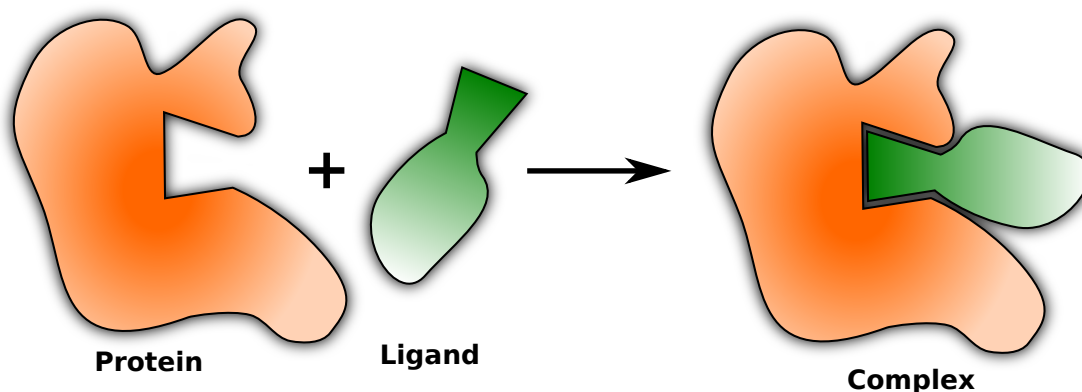


Figure 3.2: Protein, ligand docking.

There are three categories of docking programs which are divided as: Random/stochastic methods, systematic methods and simulation methods.^{126,128} Three types or classes of scoring functions are currently applied: force field based, empirical and knowledge-based

3 Determination of Protein-Protein Binding Affinities

scoring functions (see Table 3.1).

Table 3.1: Types of flexible ligand-search methods and scoring functions, adapted from Kitchen et.al.¹²⁶

Types of flexible ligand-search methods	Types of scoring functions
Random, stochastic	Force-field-based
AutoDock ¹²⁹	AutoDock
GOLD ¹³⁰	GOLD
MOE-Dock ¹³¹	D-Score
PRO_LEADS ¹³²	G-Score
	DOCK
Systematic	Empirical
DOCK ¹³³	LUDI
FlexX ¹³⁴	F-Score
Glide ¹³⁵	ChemScore
Hammerhead ¹³⁶	SCORE
FLOG ¹³⁷	Fresno
	X-SCORE
Simulation	Knowledge-based
DOCK	PMF
Glide	DrugScore
MOE-Dock	SMoG
AutoDock	
Hammerhead	

There is a need for fast, accurate and reliable methods to calculate the binding affinity of ligands to a protein.¹³⁸ The methods ideally should help drug discovery by pre-screening the potential drugs so that we can reduce the number of compounds that need to be synthesized for experimental screening. The aim of most docking programs is to screen huge number of structures within a relatively short time. To achieve this goal, numerous approximations are applied to obtain a reasonable ranking of compounds, rather than calculate absolute binding affinities accurately. On the other hand, more elaborate simulation methods such as Monte Carlo or Molecular Dynamics can get more accuracy with conformational sampling, but are more computationally intensive and not applicable to high-throughput virtual screening.¹³⁹ In the section below a number of

methods that are used to calculate binding affinities between protein and ligands/protein are described briefly.

3.2.3 Molecular Mechanics Methods

The free energies of molecular systems can describe their tendencies to associate and react. In rational drug design, to reach the required affinity and specificity, accurate estimates for both structure and binding energy are needed but unfortunately are still lacking at present.¹³⁹ Estimating binding free energies accurately is a very time-consuming process. Low-throughput computational approaches for the calculation of ligand binding free energies can be divided into “pathway” and “endpoint” methods. In pathway methods, the system is converted from one state (e.g., the complex) to the other (e.g., the unbound protein/ligand). This can be yielded by introducing a set of finite or infinitesimal “alchemical” changes to the energy function (the Hamiltonian) of the system through Free Energy Perturbation (FEP) or Thermodynamic Integration (TI), respectively. Currently, using methods such as FEP and TI combined with atomistic molecular dynamics (MD) or Monte Carlo (MC) simulations in explicit water solvent models, can obtain the most accurate results. Similar results with lower computational cost can be obtained with methods such as MM-GBSA/MM-PBSA or Linear Interaction Energy. All of the methods are still regarded as computationally too expensive to be broadly used in virtual screening.

Thermodynamic Integration

The aim of thermodynamic integration is to compute the difference in a thermodynamic property (usually the free energy) of the system between some reference state and the state of interest. This is done via sampling of state configurations in Molecular Dynamics or Metropolis Monte Carlo simulations. The free energy difference between two states

3 Determination of Protein-Protein Binding Affinities

cannot be calculated directly, because the free energy of a system is not simply a function of the phase space coordinates of the system, but instead, is related to the canonical partition function $Q(N,V,T)$. To measure the free energy change from initial to final state, thermodynamic parameters representative for the system are changed infinitesimally slowly in an effort to let the system equilibrate at each stage along the path, so that the path can be considered as reversible. The free energy difference is then calculated by defining a thermodynamic path between the states and integrating over enthalpy changes along the path. In addition to the thermodynamic variables such as volume, pressure, temperature, in a molecular simulation, one can change the interaction potential of the system and introduce suitable external potentials, to provide a larger variety of reversible paths and reference states.¹⁴⁰ For a reaction, $A \rightarrow B$, a variable λ is introduced, which is 0 for A (reactant) and 1 for B (product). Then, the potential energy of the system can be written as

$$U(r_1, \dots, r_N, \lambda) = f(\lambda)U_A(r_1, \dots, r_N) + g(\lambda)U_B(r_1, \dots, r_N) \quad (3.4)$$

For simplicity, $f(\lambda) = 1 - \lambda$ and $g(\lambda) = \lambda$ can be used. As we have

$$U(\lambda) = (1 - \lambda)U_A + \lambda U_B, \quad (3.5)$$

we obtain

$$\frac{\partial U(\lambda)}{\partial \lambda} = U_B - U_A. \quad (3.6)$$

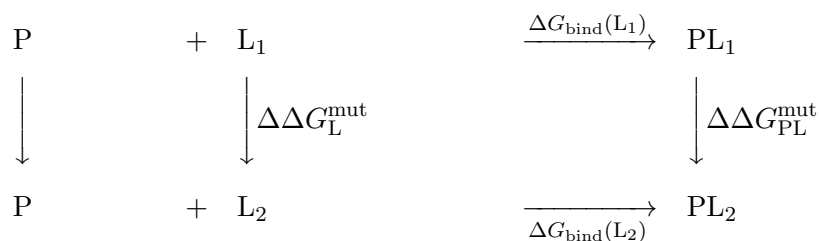
U is decomposed into solvent-solvent interactions (ww) and solute-solvent interactions (ow) in the case of solvation free energy and we get $U_A = U_{ww}$ and $U_B = U_{ww} + U_{ow}$. So we have

$$\frac{\partial U(\lambda)}{\partial \lambda} = U_{ow} = U_{ow}^{\text{vdW}} + U_{ow}^{\text{es}}, \quad (3.7)$$

where $U_{\text{ow}}^{\text{vdW}}$ is the solute-solvent van der Waals interaction and $U_{\text{ow}}^{\text{es}}$ is the solute-solvent electrostatic interaction. Therefore, we have to calculate $\langle U_{\text{ow}} \rangle_{\lambda}$ for different λ values between 0 and 1 to evaluate the thermodynamic integral, where $\langle \dots \rangle_{\lambda}$ is the average from the simulations performed at a particular value of λ .

Free Energy Perturbation

Free energy perturbation methods (FEP) and thermodynamic integration are often referred as computational alchemy. Compared to a number of approaches used for computing ligand-protein binding affinities with use of empirical models, and a model¹⁴¹ that treats part of the system as a continuum,^{142–144} molecular dynamics simulations in full atomic detail employed with the FEP methodology offer the prospect of a generally applicable rigorous “first principles”^a solution to the “binding problem”.¹⁴⁵ The FEP methodology has usually been used to compute $\Delta\Delta G_{\text{bind}}$, differences between the binding free energies of two similar ligands to one protein target, or of one ligand to a protein and its mutant.



These methods generally give very good results for the binding energy, with errors less than 1 kcal/mol.^{146,147}

$$\langle \Delta G \rangle (A \rightarrow B) = G_B - G_A = -k_B T \ln \left\langle \exp \left(-\frac{E_B - E_A}{k_B T} \right) \right\rangle_A \quad (3.8)$$

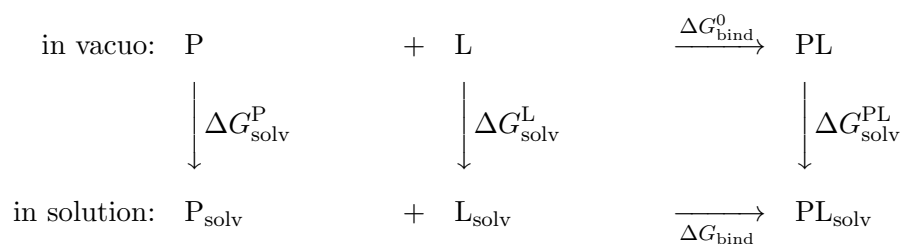
^aIn physics, when a calculation starts directly at the level of established laws of physics and does not make assumptions such as empirical model and fitting parameters, it is said to be from first principles.

3 Determination of Protein-Protein Binding Affinities

where T is the temperature, k_B is Boltzmann’s constant, the triangular brackets show an average over a simulation run for state A. In practice, one runs a normal simulation for state A, but the energy for state B is also computed. Free energy perturbation calculations only converge properly when the difference between the two states is small enough. So it is usually necessary to divide a perturbation into a series of smaller “windows” and compute independently.

MM-PBSA and MM-GBSA

The MM-PBSA/GBSA approach represents the post-processing end-state method to evaluate free energies of binding or to calculate absolute free energies of molecules in solution. The MM-PBSA/MM-GBSA approach employs molecular mechanics, the Poisson Boltzmann model respectively the Generalized Born model and solvent accessibility method to obtain energies from structural information circumventing the computational complexity of free energy simulations. In these approaches, a thermodynamic cycle is considered as follows:



The binding free energy is estimated as the sum of contributions from gas phase (“vacuo”) binding and solvation free energies that arise from the gas phase to water transition:

$$\Delta G_{\text{bind}} = \Delta G_{\text{bind}}^0 + \Delta G_{\text{solv}}^{\text{PL}} - \Delta G_{\text{solv}}^{\text{P}} - \Delta G_{\text{solv}}^{\text{L}} \quad (3.9)$$

The binding free energies are calculated for many snapshots from a molecular dynamics simulation and averaged to yield the final energies.

Gas Phase Contribution The gas-phase contribution to the binding free energy (ΔG_{bind}^0) is calculated according to

$$\Delta G_{\text{bind}}^0 = \Delta H_{\text{gas}} - T \Delta S_{\text{gas}} \quad (3.10)$$

These parts are calculated using molecular mechanics. $\Delta H_{\text{gas}} = \Delta E_{\text{bind}}^0 = \Delta E_{\text{MM}}$ is calculated from the sum of the internal, the van der Waals and the electrostatic interaction energies between the two monomers

$$\langle \Delta E_{\text{bind}}^0 \rangle = \langle \Delta E_{\text{intra}} \rangle + \langle E_{\text{vdW}} \rangle + \langle E_{\text{elec}} \rangle \quad (3.11)$$

where $\langle \Delta E_{\text{intra}} \rangle$ is the difference in the internal energy and

$$\langle \Delta E_{\text{intra}}^X \rangle = \langle \Delta E_{\text{intra,bond}}^X \rangle + \langle \Delta E_{\text{intra,vdW}}^X \rangle + \langle \Delta E_{\text{intra,elec}}^X \rangle \quad (3.12)$$

$\langle \Delta E_{\text{intra,bond}}^X \rangle$, $\langle \Delta E_{\text{intra,vdW}}^X \rangle$ and $\langle \Delta E_{\text{intra,elec}}^X \rangle$ are the energy of the bonded terms (bonds, angles, dihedral angles and improper angles) for a given molecule X, and the van der Waals and electrostatic interactions between the atoms of this molecule, respectively. $\langle \Delta E_{\text{intra}} \rangle$, $\langle E_{\text{vdW}} \rangle$, $\langle E_{\text{elec}} \rangle$ terms can be calculated according to the CHARMM molecular mechanics force field, with a dielectric constant of 1 and no cutoff for nonbonded interactions.

Entropy Contribution The entropy contribution can be decomposed into translational $\langle \Delta S_{\text{trans}} \rangle$, rotational $\langle \Delta S_{\text{rot}} \rangle$, and vibrational contributions $\langle \Delta S_{\text{vib}} \rangle$:

$$-T \langle \Delta S \rangle = -T \langle \Delta S_{\text{vib}} \rangle - T \langle \Delta S_{\text{trans}} \rangle - T \langle \Delta S_{\text{rot}} \rangle \quad (3.13)$$

3 Determination of Protein-Protein Binding Affinities

S_{trans} and S_{rot} are functions of the mass and moments of inertia, whereas calculation of S_{vib} is according to the quantum formula from a normal mode analysis¹⁴⁸. The force constant matrix which is used to determine the normal mode vectors and frequencies can be calculated and diagonalized by the VIBRAN normal mode module in the CHARMM program.

The equations for the translational, rotational and vibrational contributions to the free energy in the gas phase are as below, translations and rotations are treated semi-classically and vibrations are treated quantum-mechanically.

$$A(\text{trans}) = \frac{3}{2}k_B T - \left[\frac{5}{2} + \frac{3}{2} \ln \left(\frac{2\pi m k_B T}{h^2} \right) - \ln(\rho) \right] k_B T \quad (3.14)$$

$$A(\text{rot}) = \frac{3}{2}k_B T - \left[\frac{3}{2} + \frac{1}{2} \ln(\pi I_A I_B I_C) + \frac{3}{2} \ln \left(\frac{8\pi^2 k_B T}{h^2} \right) - \ln(\sigma) \right] k_B T \quad (3.15)$$

$$A(\text{vib}) = \sum_{i=1}^{3N-6} \left\{ \left[\frac{1}{2} h \nu_i + \frac{h \nu_i}{e^{\frac{h \nu_i}{k_B T}} - 1} \right] - \left[\frac{h \nu_i}{e^{\frac{h \nu_i}{k_B T}} - 1} - k_B T \ln \left(1 - e^{-\frac{h \nu_i}{k_B T}} \right) \right] \right\} \quad (3.16)$$

where m is the mass, ρ is the number density (corresponding to the standard state, which equal to 1M here), $I_A I_B I_C$ is the product of the 3 principal moments of inertia, σ is the symmetry factor (1 for non-symmetric molecules, 2 for the effectively symmetric dimer between 2 non-symmetric monomers), N is the number of atoms, h is the Planck constant, and ν_i is the frequency of the i th internal normal mode. The enthalpic contribution from translational and rotational degrees of freedom is a small constant ($1.5 k_B T$, which is 0.9 kcal/mol at 300K, for translations and an equal amount for rotations), the entropic contribution increases with the logarithm of the size of the molecule as they are dependent on mass and principal moments of inertia according to Eq. 3.14. It is accepted that using gas-phase equations to study the translational, rotational and vibrational properties of molecules in solution is a reasonable approximation.^{62,149–151} It has been previously

shown that the influence of S_{vib} on *relative* binding free energies $\Delta\Delta G$ is in general small and does not affect the ranking of ligands.^{46,152-154}

Solvation Contribution The solvation contributions $\langle G_{\text{solv}}^{\text{P}} \rangle$, $\langle G_{\text{solv}}^{\text{L}} \rangle$ and $\langle G_{\text{solv}}^{\text{PL}} \rangle$ consist of contributions from electrostatics and non-polar interactions

$$\langle G_{\text{solv}} \rangle = \langle G_{\text{elec}} \rangle + \langle G_{\text{nonpolar}} \rangle \quad (3.17)$$

In MM-PBSA, the electrostatic contribution to the solvation term is calculated by solving the Poisson-Boltzmann (PB) equation.^{155,156} The Poisson-Boltzmann equation (PB) describes the electrostatic environment of a solute in a solvent containing ions.

$$\vec{\nabla}[\epsilon(\vec{r})\vec{\nabla}\Psi(\vec{r})] = -\rho^f(\vec{r}) - \sum_i c_i^\infty z_i q \lambda(\vec{r}) \exp\left[\frac{-z_i q \Psi(\vec{r})}{k_B T}\right] \quad (3.18)$$

where $\vec{\nabla}$ is the divergence operator, $\epsilon(\vec{r})$ is the position-dependent dielectric, $\vec{\nabla}\Psi(\vec{r})$ is the gradient of the electrostatic potential, $\rho^f(\vec{r})$ represents the charge density of the solute, c_i^∞ represents the concentration of the ion i at a distance of infinity from the solute, z_i is the charge of the ion, q is the elementary charge, k_B is the Boltzmann constant, T is the temperature, and $\lambda(\vec{r})$ is a factor for the accessibility of position r to the ions in solution (often set to uniformly 1). The equation can be linearized to be solved more efficiently.¹⁵⁷

The Generalized Born (GB) model used in MM-GBSA is an approximation to the exact linearized Poisson-Boltzmann equation. In the GB model, the protein is treated as a set of spheres whose internal dielectric constant differs from the external solvent.¹⁵⁸

$$G_{\text{GB}} = \frac{1}{8\pi} \left(\frac{1}{\epsilon_0 - \frac{1}{\epsilon}} \right) \sum_{i,j}^N \frac{q_i q_j}{f_{\text{GB}}} \quad (3.19)$$

3 Determination of Protein-Protein Binding Affinities

where

$$f_{\text{GB}} = \sqrt{r_{ij}^2 + a_{ij}^2} e^{-D} \quad (3.20)$$

where $D = \left(\frac{r_{ij}}{2a_{ij}}\right)$, and $a_{ij} = \sqrt{a_i a_j}$. ϵ_0 is the permittivity of free space, ϵ is the dielectric constant of the solvent. q_i is the electrostatic charge of particle i , r_{ij} is the distance between particles i and j , and a_i is the effective Born radius which can be thought as the distance from the atom to the molecular surface. It is very important to estimate accurately the effective born radii for the GB model.

GBSA is one of the most commonly used implicit solvent model combinations which is a Generalized Born model with a hydrophobic solvent accessible surface area SA term. The nonpolar contribution is the sum of a solvent-solvent cavity term and a solute-solvent van der Waals term

$$G_{\text{cav}} + G_{\text{vdW}} = \sum \sigma_k SA_k \quad (3.21)$$

where SA_k is the total solvent-accessible surface area (SASA) of atoms of type k and σ_k is an empirical atomic solvation parameter, with a value of 0.0072 kcal/mol/ \AA^2 . The electrostatic contribution to the solvation energy in MM-GBSA is calculated within the generalized Born approximation

$$G_{\text{pol}} = -166 \left(1 - \frac{1}{\epsilon}\right) \sum \sum \frac{q_i q_j}{f_{\text{GB}}} \quad (3.22)$$

where ϵ is the dielectric constant of water ($\epsilon = 80$), q_i and q_j are the charges of atoms i and j . f_{GB} is an expression that depends on the Born atomic radii α_i and α_j and distances r_{ij} .¹⁵⁹

$$f_{\text{GB}} = r_{ij}^2 + \alpha_i \alpha_j \exp\left(\frac{-r_{ij}^2}{8\alpha_i \alpha_j}\right) \quad (3.23)$$

To sum up, $\langle \Delta G_{\text{bind}}^0 \rangle$ is the gas-phase term obtained by molecular mechanics calculations **MM**, $\langle G_{\text{elec}} \rangle$ can be determined using the Poisson-Boltzmann or Generalized

Born (**PB/GB**) implicit solvent model, $\langle G_{nonpolar} \rangle$ is assumed to be proportional to the solvent accessible surface area (**SASA**), the methods are called: MM-PB/GBSA. After averaging the energy terms, the binding free energy can be expressed as follows:

$$\langle \Delta G_{bind} \rangle = \langle \Delta E_{intra} \rangle + \langle \Delta E_{vdW} \rangle + \langle \Delta E_{elec} \rangle + \langle \Delta G_{elec,desolv} \rangle + \langle \Delta G_{np,desolv} \rangle - T \langle \Delta S \rangle \quad (3.24)$$

Decomposition of Binding Free Energy MM-GBSA allows to easily decompose the binding free energy at the atomic level (with the exception of the contribution of the entropy terms), which can be used to evaluate the contribution of each residue to the total binding free energy as well as the contribution of its side-chain and backbone. From Eq. 3.24, for the same trajectory method (STM), $\langle \Delta E_{intra} \rangle$ is equal to zero, as the internal energies of the complex and the individual one are calculated from the same trajectory. In this thesis, we used STM which has also been employed by others,^{160–162} and it has been shown to provide reasonable results. In the different trajectory method (DTM), the $\langle \Delta E_{intra} \rangle$ term is important, $\langle \Delta E_{intra,vdW}^X \rangle$ and $\langle \Delta E_{intra,bond}^X \rangle$ can be calculated in the same way as $\langle \Delta E_{vdW}^X \rangle$ and $\langle \Delta E_{bond}^X \rangle$. For the dimerization free energy, in principle, using separate trajectories for the monomers and dimer can capture the conformational changes upon binding, but the results will be very sensitive to the detailed structures. Furthermore, the DTM introduces additional errors compared to the STM.⁴⁶ For electrostatic interaction energy between the 2 monomers, the contribution of atom i is given by

$$\langle E_{elec}^i \rangle = \frac{1}{2} \sum_j \frac{q_i q_j}{r_{ij}} \quad (3.25)$$

where j belongs to monomer **I**, and i belongs to the other. $r_{i,j}$ is the distance between the two atoms with charge q_i and q_j , respectively. The SASA of each atom in the dimer is $SASA^i$, and the contribution to the non-polar term $\langle \Delta G_{np,solv}^i \rangle = \sigma \times [SASA^{i,d} - (SASA^{i,m_1} + SASA^{i,m_2})]$. The GB-MV2 model^{159,163} uses the following expression for

3 Determination of Protein-Protein Binding Affinities

the electrostatic solvation energy term

$$\langle \Delta G_{elec,solv} \rangle = -166.0(\epsilon_{solute}^{-1} - \epsilon_{solvent}^{-1}) \sum_{i,j} \frac{q_i q_j}{\sqrt{r_{ij}^2 + a_i a_j \exp\left(\frac{-r_{ij}^2}{K_s a_i a_j}\right)}} \quad (3.26)$$

where ϵ_{solute} and $\epsilon_{solvent}$ are the dielectric constant of the solute and the solvent, in this thesis we used the value of 1 and 80, respectively. a_i and a_j are the Born radii of atoms i and j . In the GB-MV2 model, the value of K_s is 8 instead of 4 as in the Still equation.¹⁵⁸ Therefore, the contribution of atom i to $\langle G_{elec,solv} \rangle$

$$\begin{aligned} \langle \Delta G_{elec,solv}^i \rangle = & -166.0(\epsilon_{solute}^{-1} - \epsilon_{solvent}^{-1}) \frac{q_i^2}{a_i} \\ & - \frac{1}{2} 166.0(\epsilon_{solute}^{-1} - \epsilon_{solvent}^{-1}) \sum_{j \neq i} \frac{q_i q_j}{\sqrt{r_{ij}^2 + a_i a_j \exp\left(\frac{-r_{ij}^2}{K_s a_i a_j}\right)}} \end{aligned} \quad (3.27)$$

Compared to the PB approach, GB is much faster and can allow one to decompose easily and rapidly the electrostatic solvation energy and thus, the binding free energy, into atomic contributions from only 1 calculation. The per-atom contributions can be summed up to atomic groups such as residues, backbones, and sidechains, to obtain their contributions to total binding free energy. A decomposition of the binding free energy in the PB calculations is also possible^{82,164} but requires separate and time-consuming calculations. The results obtained by using the GB-MV II model are very close to the solvation free energies calculated by solving the PB equation.¹⁵⁵ The MM-PB/GBSA approach has been applied to estimate the binding energy for protein-protein and protein-ligand systems.¹⁶⁵⁻¹⁶⁷ It has also been used to predict the effect of residue mutations on the binding energy of protein-protein systems with the ‘‘computational alanine scanning’’¹⁶⁰

Limitations of MM-GB/PBSA Calculations MM-GB/PBSA methods are valuable tools used in computer-aided drug design. However, as with any other method, they have limitations which have to be taken into consideration when they are employed. First, they are useful for ranking relative ligand binding affinities, but not accurate enough for absolute binding free energy predictions.¹⁶⁸ PB and GB results strongly rely on atomic charges and van der Waals radii, which are often optimized for MD simulations. The molecular structure of the solvent was not considered by continuum electrostatics models, which may affect the result, particularly when key receptor-ligand interactions are bridged by water molecules, as we found in this thesis. Furthermore, the inclusion of entropic contributions make the results closer to experimental absolute affinities¹³⁸, but it is time consuming to calculate the entropic terms and also contains a lot of uncertainties. Moreover, the value of the protein/ligand dielectric constant is chosen empirically, Hou et.al,¹⁵⁴ suggested $\epsilon=4$ for a lightly charged protein-ligand binding interface, $\epsilon=2$ for a moderately charged binding interface and $\epsilon=1$ for a hydrophobic binding interface may improve ligand ranking.

While molecular docking algorithms are computationally efficient methods for screening a large number of ligands against a targeted protein in a reasonable time, it is still not very practical to generate MD ensembles for more than certain number of receptor-ligand structures in a MM-GB/PBSA calculation. However, MM-GB/PBSA can serve well as a post-docking method in virtual screening experiments. Using MM-GB/PBSA algorithms on single docking poses is proven to improve correlations between predicted and experimental binding affinities.¹⁶⁹⁻¹⁷¹ Post-docking MM-GBSA is implemented in the Schrodinger software in the program Prime, and the entropy term is neglected by default. Manta et al. extended MM-PBSA by exploiting quantum mechanics methods in QM/MM-PBSA^{172,173}, where a hybrid gas phase energy term ($E_{QM/MM}$) replaces the pure molecular mechanics energy E_{MM} in Eq. 3.11. While this method can eliminate the problem arising from deficient ligand forcefield parameters, it is significantly more

3 Determination of Protein-Protein Binding Affinities

expensive than MM-GB/PBSA, and therefore of limited viability.

4 Bacterial Second Messenger c-di-GMP

Cyclic diguanosine-monophosphate (c-di-GMP, CDG) is a bacterial signaling molecule that triggers a switch from motile to sessile bacterial lifestyles.^{174,175} It has attracted considerable attention from different fields since it was discovered 20 years ago.^{89,174–183} Extensive research on the small signaling molecule cyclic diguanosine monophosphate (c-di-GMP) has highlighted its critical regulator role in bacterial metabolism. c-di-GMP is produced from two molecules of GTP by the activity of diguanylate cyclases (DGCs) which usually contain a conserved GGDEF (Gly-Gly-Asp-Glu-Phe) domain. c-di-GMP can also bind to some DGC proteins, repressing the activity of DGC via allosteric changes.^{4,184} On the other hand, c-di-GMP is degraded by specific phosphodiesterase (PDE) into 5'-phosphoguanylyl-(3'-5')-guanosine (pGpG); pGpG is subsequently split into two GMP molecules. The latter enzyme usually has a conserved EAL domain (a domain enriched in Glu-Ala-Leu). In some bacteria, the HD-GYP domain (His-Asp, Gly-Tyr-Pro) replaces the EAL domain.

Most GGDEF and EAL (or HY-GYP) domains are linked to a signal input (or sensory) domains such as the PAS (oxygen sensing), blue light sensing, red/far red light sensing (GAP-PHY domain), gas sensing (haemerythrin), REC (phospho receiver) and GAF domains. Afterwards, through the interaction with different receptors, such as transcription factors, a Riboswitch or a PilZ domain^{185–187}, c-di-GMP modulates diverse biological functions, such as biofilm formations, motility, virulence or the cell cycle. It

4 Bacterial Second Messenger c-di-GMP

has been characterized that c-di-GMP negatively modulates cell motility and traits associated with bacterial virulence and stimulates several biofilm-associated functions.¹⁸⁸ High concentrations of c-di-GMP promote sessile growth (biofilm formation), while low concentrations of c-di-GMP promote motile growth.

The discovery that most bacteria use c-di-GMP as a ubiquitous second messenger to orchestrate the switch between a planktonic and a sedentary, biofilm-related lifestyle established an entirely new field of research for studying the cellular, molecular and structural details of the components involved in this process.¹⁷⁴ The first example of this regulatory network was characterized in *G. xylinus* (see Figure 4.1).^{189,190} The proteins are involved in c-di-GMP synthesis and degradation through three domains, a PAS, a GGDEF and an EAL domain. Oxygen is the first messenger that binds to the PAS domain, which has a heme or flavin binding pocket and serves as an oxygen or redox sensing domain. c-di-GMP is the second messenger. The PDE activity (EAL domain) is tightly coupled to the binding status of heme in the PAS domain. Under aerobic conditions, O₂ binds to PAS-heme, PDE activity is diminished and c-di-GMP levels are high which promotes its binding to downstream cellulose synthase to invoke an allosteric change and produce cellulose. On the other hand, under anaerobic conditions (no oxygen binding to PAS-heme), PDE activity is restored and c-di-GMP concentration diminishes, resulting in the removal of the cellulose synthase activity.^{190,191}

Inhibition by Domain Immobilization The GGDEF domain was first identified in PleD in the aquatic bacterium *Caulobacter crescentus*. PleD^{4,184,192,192-197} contains three domains, with a REC1-REC2-GGDEF domain organization. REC1 and REC2 are phosphorylation receiver domains connected by a loop. Phosphorylation of REC1-REC2 triggers PleD to dimerize.¹⁹⁰ In non-phosphorylated PleD (pdb code: 1W25⁴) and phos-

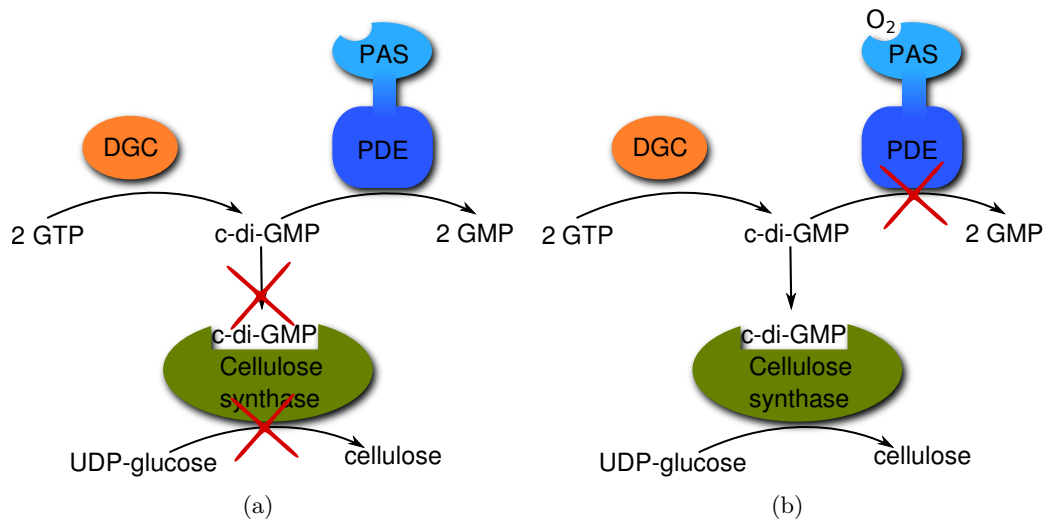


Figure 4.1: Regulation of cellulose synthesis in *G. xylinus* under anaerobic (a) and aerobic (b) conditions. Under anaerobic conditions, PDE is fully functional, leading to the degradation of c-di-GMP. Under aerobic conditions, PDE activity is repressed. Accumulated c-di-GMP leads to the activation of cellulose synthase. Adapted from¹⁹⁰

phorylated PleD (pdb code: 2WB4¹⁹⁸), structure determination identified an allosteric site for the mutually intercalated c-di-GMP dimer [c-di-GMP]₂. Binding of a c-di-GMP dimer to this allosteric site (inhibition site, I-site) via residues R539, D362 and R390 on the DGC domain (pdb code:1W25), leads to strong non-competitive inhibition of PleD.^{4,196} Moreover, c-di-GMP present in the I-site crosslinks to either the neighboring D2 domain via interaction to R148, R178, or to another neighboring DGC domain via interactions to R313.^{4,184} A mechanistic model by Jenal *et al.*^{4,184} suggested that binding c-di-GMP dimer in the I-site locks the DGC domain in a non-productive orientation which blocks the encounter of two GTP-loaded active site (A-site) and subsequent the generation of c-di-GMP monomer. Two different immobilization modes are found, named DGC-D2 and DGC-DGC crosslinking (Figure 4.2)

Subsequent studies has confirmed that product inhibition is a general feature of

DGCs (GGDEF domain)¹⁹⁶. When c-di-GMP concentration reaches a threshold, with a K_i of 0.5 μM , product inhibition, inactivates the DGC enzyme. However, the nature of communication between I- and A- site for self-regulation is still unknown. Several aspects of this signaling molecule remain far from being understood. When c-di-GMP binds to PleD I-site, does it bind as a monomer and then forms dimer or two c-di-GMP monomers form the dimer first, and bind to protein as a dimeric form? In this chapter, we present our results from the all-atom computational simulations to study the structure and energetic of the c-di-GMP dimer and analog in the solution and proteins.

4.1 Stability and Dynamics of Cyclic Diguanylic Acid in Solution

4.1 Stability and Dynamics of Cyclic Diguanlyic Acid in Solution

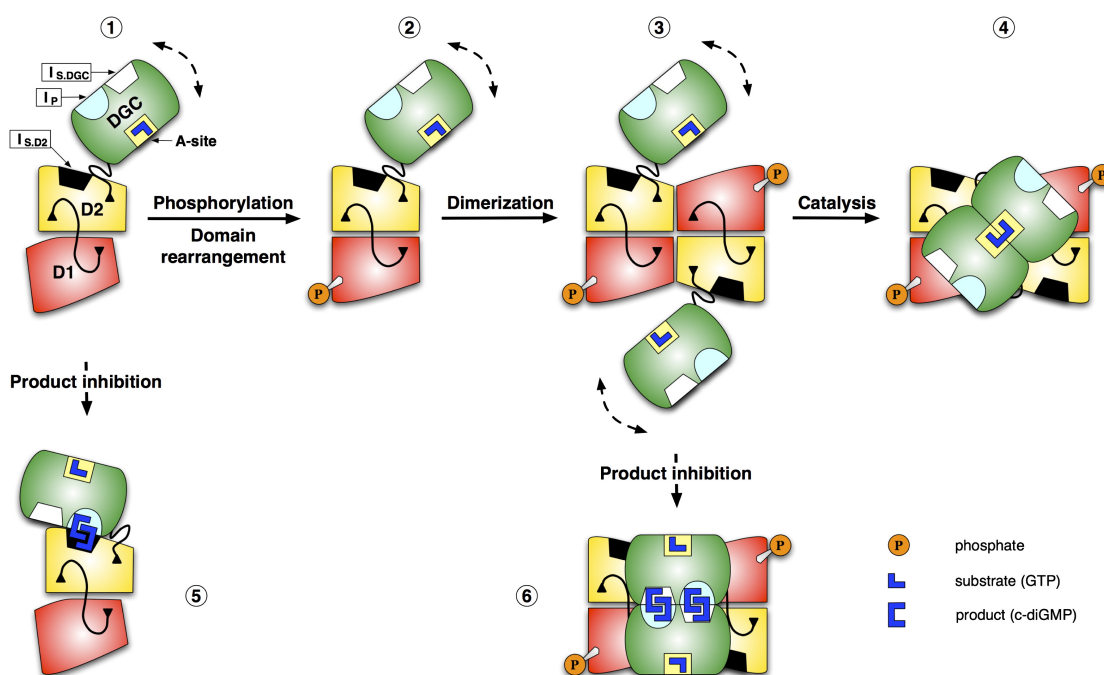


Figure 4.2: Mechanistic model of PleD regulation. The model is adapted from Chan et al.⁴. The DGC domain (green) is connected via a flexible linker to the stem (receiver domain D1 [red] and adaptor domain D2 [yellow]) and is supposed to be mobile relative to it. (Upper row) Activation. Phosphorylation of domain D1 leads to a rearrangement of the stem domains, which, in turn, allows for formation of a tight dimeric stem (3). The dimeric arrangement is a prerequisite for an efficient and productive encounter of the two substrate-loaded DGC domains to form the c-di-GMP product (4). (Lower row) Product inhibition. Dimeric product molecules, (c-di-GMP)₂, can crosslink the primary inhibition site on DGC, Ip, with a secondary binding site either on D2, Is,D2 (5) or on the adjacent DGC domain, Is,DGC (6). The former structure has been observed experimentally with nonactivated PleD, the latter structure is presented in this report. In both cases, the DGC domains become immobilized, and the active sites are hampered from a productive encounter. Note that a possible direct communication between Ip and A sites (Christen et al., 2006) is not depicted. Reprint¹⁸⁴ with permission of Elsevier.

DOI: 10.1002/cphc.201000692

Stability and Dynamics of Cyclic Diguanylic Acid in Solution

Lixian Zhang and Markus Meuwly*^[a]

Cyclic diguanylic acid (CDG) is a ubiquitous messenger involved in bacterial signaling networks. Despite its central role in motility, biofilm formation, virulence, and flagellum development, fundamental properties such as its aggregation state are still poorly understood. Here the dynamics and stability of metal-free and Mg²⁺-bound CDG are characterized. Atomistic simulations establish that the CDG dimer is slightly favored (by -5 kcal mol^{-1}) over its dissociated form (2CDG), while the Mg²⁺ ion coordinated in the X-ray structure readily dissociates

from (CDG)₂ in solution and prefers water coordination. As a ligand in a protein, CDG binds both as a U-shaped and a quasi-linear monomer. The current results indicate that the energy difference between these two conformations is only a few kilocalories per mole, which explains the facile adaptation to different protein environments. This, together with the slight preference of (CDG)₂ over 2CDG suggests that (CDG)₂ binding to a protein does probably not occur via sequential binding of two individual monomers.

1. Introduction

Cyclic diguanylic acid (cyclic diguanosine monophosphate, c-di-GMP, CDG) is an important second messenger in bacteria, primarily involved in signaling the switch between a motile (planctonic) and a sessile (biofilm-related) lifestyle.^[1–3] CDG-dependent signaling in bacteria is a complex process in which a multitude of diguanylate cyclases and phosphodiesterases are involved, even in a single bacterial species. The molecule consists of two guanosine monophosphate (GMP) moieties linked by two phosphodiester bonds. Cyclic dinucleotides in solution tend to form multimeric structures. The X-ray structure of CDG was found to consist of a Mg²⁺-bound dimer,^[4] but higher oligomeric states were proposed and found in solution depending on the type and concentration of the cation used and the local concentration of CDG.^[5] Other arrangements, such as a tetrameric form with a central cavity, were proposed but not observed in the crystal structure.^[6,7]

Of particular importance for the in vivo situation and in vitro characterization is the question whether CDG in solution occurs in monomeric or dimeric form. This is relevant to better understanding molecular mechanisms underlying interaction of CDG with proteins or RNA. Experiments with the protein PleD from *Caulobacter crescentus*, a protein that efficiently catalyzes the conversion of GTP to c-di-GMP,^[8] showed that in the protein inhibition site (I site) the second messenger binds as a metal-free dimer,^[9] as did a crystal structure of the diguanylate cyclase WspR.^[10] In addition to the metal-free dimer, the X-ray structure of PleD also contains CDG monomers in the active site (A site). Contrary to that, the crystal structure of *Vibrio cholera* VCA0042 protein, consisting of YcgR-N and PilZ domains, shows that CDG binds in its monomeric form,^[11] whereas in PP4397 of *Pseudomonas putida*, which also has YcgR-N and PilZ domains, the CDG dimer binds.^[12]

For understanding the interaction of PleD,^[9] WspR,^[10] and other proteins with CDG, the aggregation state of the messenger

in solution is of fundamental interest. Several scenarios can be imagined: 1) CDG is synthesized in its metal-bound form and released to the solvent, where it loses Mg²⁺ and binds as a metal-free dimer. 2) After synthesis the Mg²⁺-bound dimer dissociates into its monomeric form and binds sequentially as two separate monomers. 3) The metal-free dimer is also stable in solution and binds in this form. Experiments in solution found that CDG exists in five distinct but related complexes.^[5] The methods employed (UV and CD spectroscopy) could establish guanine stacking, which also occurs in the crystal structures, and NMR spectroscopy showed that different oligomerization states occur in solution with different propensity depending on the chemical identity of the cation and its concentration. However, no structural data in solution is yet available.

In the following we determine the stability and dynamics of solvated metal-free and metal-bound CDG from atomistic simulations and electronic structure calculations.

Methods Section

Electronic Structure and Molecular Dynamics Simulations: Density functional theory (DFT) calculations were carried out with Gaussian 03^[13] at the B3LYP/6-31G** level. The starting structures for the CDG monomer and dimer (see Figure 1a and b) were those found in the X-ray structure of PleD^[9] (code 1W25, Brookhaven Protein Data Bank (PDB)^[14]). For the Mg²⁺-bound CDG dimer (Figure 1c) coordinates from the Cambridge Structural Database^[4,15] were used. The overall charge is -4 for the mutually intercalated CDG dimer and -2 for the Mg²⁺-bound dimer.

[a] L. Zhang, Prof. Dr. M. Meuwly
Department of Chemistry, University of Basel
Basel (Switzerland)
Fax: (+41) 612673855
E-mail: m.meuwly@unibas.ch

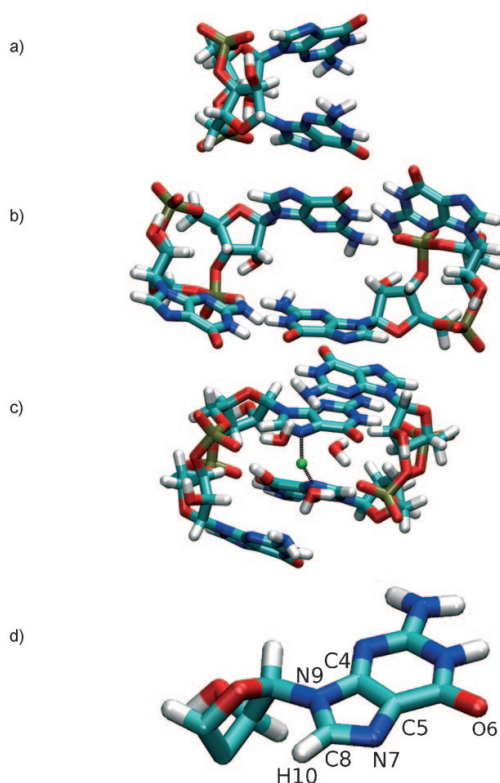


Figure 1. Optimized geometry of: a) CDG monomer, b) CDG dimer, c) CDG dimer with Mg²⁺ and four water molecules from ab initio calculation, and d) labeled atoms for which charges were increased in the central guanosine base part (electrostatic models I, II, and III).

All molecular dynamics (MD) simulations were carried out with the CHARMM program^[16] using the CHARMM27 force field^[17] with periodic boundary conditions (PBC). Force-field parameters for CDG were those determined and used previously.^[18,19] The structures of metal-free and Mg²⁺-bound CDG dimer were optimized with 5000 steps of steepest descent minimization followed by adopted Newton–Raphson minimization until a gradient of 10^{−7} was reached. Then the two complexes were solvated in a pre-equilibrated water box containing 1452 TIP3P^[20] water molecules of dimension 37.3 × 34.1 × 34.1 Å³. The solvent was optimized and equilibrated at 300 K for 30 ps in the presence of the fixed solute. Finally, the structures of the solutes were relaxed by means of 2000 steps of steepest descent minimization. The two systems, with a total of 4261 and 4283 atoms, respectively, were gradually heated to 300 K for 15 ps followed by 100 ps of equilibration. All MD simulations were carried out by using SHAKE^[21] to constrain all bonds involving hydrogen atoms. For metal-free and metal-bound CDG dimer five independent trajectories (100 ns in total) and two independent trajectories (20 ns in total) were run, respectively.

During the MD simulation, the CDG dimer was restrained to the center of the water box by applying a weak harmonic potential to the center of mass of the CDG dimer with a force constant of 0.1 kcal mol^{−1} Å^{−2} to avoid translation of the solute within the periodic box. To follow overall structural changes, root mean square

deviations (RMSDs) were calculated along the trajectories. For simulations of the Mg²⁺-bound CDG dimer it was observed that the metal has a finite probability to escape to the solvent on the time-scale of the simulations (10 ns). Therefore, additional simulations were carried out in which the position of the magnesium ion was constrained relative to the N7 nitrogen atoms of the central guanosine unit by using NOE constraints. The energy biases added to the system are $E = 0.5K_{\min}(R - R_{\min})^2$ for $R < R_{\min}$ and $E = 0.5K_{\max}(R_{\max} - R)^2$ for $R_{\max} < R < R_{\min}$. In all simulations, $R_{\min} = 2.0$, $R_{\text{lim}} = 4.9$, and $R_{\max} = 2.9$ Å were used. The force constants [K_{\min} , K_{\max}] included [100,200] and [500,1000] kcal mol^{−1} Å^{−2}, respectively. Alternatively, to avoid biasing the simulations, enhanced charges on the two N7 atoms were used. More precisely, the following charge combinations were used for N7, C8, C5, C4, and H10 (see Figure 1 d): model I (−0.60, 0.25, 0.00, 0.26, 0.16), model II (−0.80, 0.30, 0.10, 0.29, 0.18), and model III (−0.90, 0.37, 0.12, 0.29, 0.19), where model I is the original parameterization.

Free Energy of Dimerization: The binding free energy ΔG consists of an enthalpic (ΔE) and an entropic contribution (ΔS), which can be calculated according to a thermodynamic cycle.^[22,23] Here the molecular mechanics with generalized Born surface area (MM-GBSA) approach^[24] is used, which decomposes the binding free energy ΔG_{bind} into a sum of the gas-phase contribution, the desolvation energy of the system upon binding ΔG_{desolv} , and an entropic contribution $-T\Delta S$ [Eq. (1)]:

$$\Delta G_{\text{bind}} = \langle \Delta G_{\text{bind}}^0 \rangle + \langle \Delta G_{\text{desolv}} \rangle - T \langle \Delta S \rangle \quad (1)$$

The brackets indicate averages of the individual terms along an equilibrium MD simulation.^[25] The gas-phase contribution $\langle \Delta G_{\text{bind}}^0 \rangle$ consists of the difference in van der Waals (vdW) and electrostatic energies between the dimer (D) and the individual monomers (M), whereas the desolvation (desolv) term consists of a nonpolar (np) contribution, which is assumed to be proportional to the solvent-accessible surface area (SASA), and an electrostatic (ele) contribution calculated with the GB-MV2 model.^[26,27] A value of 0.0072 kcal mol^{−1} Å^{−2} was used to relate the buried SASA to the nonpolar desolvation free energy:^[28,29,25]

$$\begin{aligned} \Delta G_{\text{bind}} &= \langle \Delta G_{\text{bind}}^0 \rangle + \langle \Delta G_{\text{desolv}} \rangle - T \langle \Delta S \rangle \\ &= \langle \Delta E_{\text{intra}} \rangle + \langle \Delta E_{\text{vdW}} \rangle + \langle \Delta E_{\text{elec}} \rangle + \langle \Delta G_{\text{ele,desolv}} \rangle + \langle \Delta G_{\text{np,desolv}} \rangle - T \langle \Delta S \rangle \end{aligned} \quad (2)$$

The energies calculated for snapshots taken every 10 ps by using the “same trajectory method” (STM)^[25] were averaged over 10 ns. In the STM, the energy terms relative to the isolated monomers are calculated by using coordinates taken from the simulation of the dimer. The entropy term is partitioned into translational, rotational, and vibrational terms following standard equations of statistical mechanics.^[30] The translational contribution only depends on the mass of the system and the rotational part is related to the moments of inertia. The vibrational part can be calculated from the normal modes (NM) ν_i [Eq. (3)]:

$$-T\Delta S_i = \frac{h\nu_i}{e^{\frac{h\nu_i}{k_B T}} - 1} - k_B T \ln(1 - e^{-\frac{h\nu_i}{k_B T}}) \quad (3)$$

where k_B is the Boltzmann constant, h the Planck constant, and $T = 300$ K. The VIBRAN module of the CHARMM program was used to calculate and diagonalize the force constant matrix to determine the normal-mode vectors and frequencies. Normal modes were calculated both for the fully minimized structures of the dimer and the monomers in vacuo with a distance dependent dielectric ($\epsilon =$

4), and a cutoff of 12 Å for non-bonded interactions was applied. The snapshots were minimized by using the adopted basis Newton–Raphson (ABNR) method until a gradient of 10^{-7} was reached. All reported values are averaged over 1000 frames, equally distributed over the 10 ns trajectories.

2. Results and Discussion

2.1. Electronic-Structure Calculations

DFT calculations were carried out for the isolated monomer, for the CDG dimer, and for the CDG dimer with hydrated Mg^{2+} ($\text{Mg}^{2+}(\text{H}_2\text{O})_4$) bound to it. For all three systems the structure was fully optimized starting from the coordinates mentioned in the Methods Section. The final RMSD between the initial and optimized structures is 0.86, 3.21, 0.34 Å for CDG monomer, CDG dimer, and CDG dimer with hydrated Mg^{2+} , respectively. The rather large RMSD for the CDG dimer is related to the overall charge of -4 , which is not shielded in a vacuum calculation. The RMSD for the individual monomers of optimized CDG dimer are 2.10 and 2.17 Å, respectively, and a Mg^{2+} -free structure in solution is not available, which makes direct comparison difficult. On the other hand, the agreement for the monomer and the Mg^{2+} -bound dimer is encouraging.

The stabilization energy of the dimer with respect to the asymptotes (either the two monomers in the dimer configuration $E_{\text{M1}}^{\text{dimer}}$ and $E_{\text{M2}}^{\text{dimer}}$ or the optimized, isolated monomer $E_{\text{M}}^{\text{opt}}$) is $\Delta E_1 = E^{\text{dimer}} - (E_{\text{M1}}^{\text{dimer}} + E_{\text{M2}}^{\text{dimer}}) = 21.8 \text{ kcal mol}^{-1}$ or $\Delta E_2 = E^{\text{dimer}} - 2 \times E^{\text{opt.M}} = 18.2 \text{ kcal mol}^{-1}$, respectively (see Table 1). Thus, in vacuum the metal-free dimer is destabilized by about 20 kcal mol^{-1} . Solvation will, however, affect this due to shielding of the charges. Thus, the energies of all optimized structures were also calculated with the implicit polarizable continuum model (PCM)^[31] and the resulting binding energies are $\Delta E_1 = -26.4 \text{ kcal mol}^{-1}$ and $\Delta E_2 = -27.8 \text{ kcal mol}^{-1}$, respectively. For the Mg^{2+} -bound dimer in all cases a stable system was found. Additional calculations on the optimized B3LYP/6-31G** structures were carried out to assess effects of basis set superposition error (BSSE),^[32,33] larger basis sets [6-311++G(2d,2p)], and a different functional (PBE1PBE).^[34] Conventional (Boys–Bernardi) BSSE correction^[32] leads to slightly decreased stabilization, whereas a more recent counterpoise scheme^[33] reduces the stabilization of the metal-free dimer in vacuum considerably. With a much larger basis set, also a destabilization of 11 kcal mol^{-1} is found, whereas a calculation

Table 1. Total energies of minimized monomers (M1, M2) and dimers (D) with and without the solvated Mg^{2+} ion.^[a]

Structure	$E_{\text{vac}}/E_{\text{h}}$	$\Delta E_{\text{vac}} [\text{kcal mol}^{-1}]$	$E_{\text{solv}}/E_{\text{h}}$	$\Delta E_{\text{solv}} [\text{kcal mol}^{-1}]$
D	-6118.13425		-6118.74477	
M1	-3059.08454		-3059.35134	
M2	-3059.08452		-3059.35140	
B3LYP/6-31**		21.8		-26.4
B3LYP/6-31** (BSSE)		35.8		-12.4
B3LYP/6-31** (CP)		68.8		20.6
B3LYP/6-311++G(2d,2p)		32.7		-14.2
PBE1PBE/6-311++G(2d,2p)		26.9		-19.3
2M	-6118.16374		-6118.70038	
B3LYP/6-31**		18.2		-27.8
B3LYP/6-311++G(2d,2p)		27.0		-17.5
PBE1PBE/6-311++G(2d,2p)		21.4		-22.9
D($\text{Mg}^{2+}(\text{H}_2\text{O})_4$)	-6624.30192		-6624.55805	
M1	-3059.04533		-3059.33935	
M2	-3059.04542		-3059.33944	
($\text{Mg}^{2+}(\text{H}_2\text{O})_4$)	-505.30574		-505.69605	
B3LYP/6-31**		-568.2		-115.0
B3LYP/6-31** (BSSE)		-528.7		-77.5
B3LYP/6-31** (CP)		-282.5		-7.4
B3LYP/6-311++G(2d,2p)		-533.0		-77.3
PBE1PBE/6-311++G(2d,2p)		-551.9		-95.6

[a] Total energies (for B3LYP/6-31G**) in Hartree (E_{h}), relative stabilization energies in kcal mol^{-1} . The structures in vacuum are optimized starting from the X-ray structure and give the vacuum energy E_{vac} ; energies with implicit solvent E_{solv} (PCM) are single-point calculations on the optimized vacuum structure. Stabilization energies with B3LYP/6-31** (BSSE),^[32] (CP),^[33] B3LYP/6-311++G(2d,2p), and PBE1PBE/6-311++G(2d,2p) were calculated on the optimized B3LYP/6-31G** structures.

with the 6-311++G(2d,2p) basis and the more recent PBE1PBE functional yields a similar stabilization to B3LYP/6-31**. The PBE1PBE functional was recently found to perform best for describing nonbonded interactions, in particular H-bonding.^[35] Including solvation effects in all cases increases the interaction energies and, with the exception of one of the counterpoise schemes, leads to a stable dimer. The same considerations were applied to the metal-bound dimer, and in all cases increased stabilization compared to the metal-free dimer was found.

2.2. Molecular Dynamics Simulations

With molecular dynamics simulations it is possible to account for solvation effects beyond the approximation of implicit solvents. Furthermore, the influence of entropic effects can be estimated and assessed. Following the methods outlined above the dynamics of metal-free and metal-bound CDG dimer was investigated in separate simulations. First, simulations of the metal-free form are described.

Figure 2 shows the RMSD from the X-ray structure, calculated for the entire dimer (black), and for each of the monomers (red for monomer I, green for monomer II). Snapshots were taken every 10 ps. The RMSD was calculated after optimally superimposing the atoms of the instantaneous conformation of the dimer or each monomer on the X-ray structure.^[9] The RMSD compared to the X-ray structure remains below 2 Å for most of the time during all individual 20 ns simulations, except for occasional increases up to nearly 4 Å. Detailed analysis of

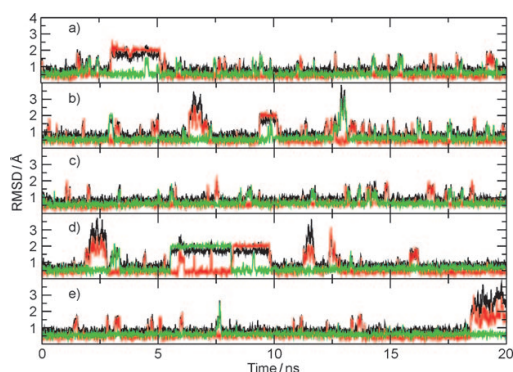


Figure 2. RMSD of the c-di-GMP dimer (black lines), as well as the first (red lines) and second (green lines) monomers from individual 20 ns MD simulations. Monomers I and II were fitted to the X-ray structure of monomers I and II, respectively. Despite some large RMSD excursions (RMSD > 2 Å), the dimer usually returns to a compact structure. Panels a) to e) are from five independent MD simulations.

the trajectories reveals that the guanosine part of monomer I opens its U-shaped structure during that time and then returns to its original conformation.

In the minimum-energy structure, the Mg^{2+} ion is coordinated to two N7 atoms from each CDG monomer, which leads to a wedge shape of the two central bases and induces a bend in the stacking of the intercalated dimer. Furthermore, the X-ray structure^[4] shows that the Mg^{2+} ion, which is located between the two monomers, is hydrated. This is partly because two of the water molecules of the hydrated Mg^{2+} complex interact with both oxygen atoms of the phosphate groups that are not involved in H-bonding to the base N1 atoms. They also form H-bonds to O6 atoms of the two central bases (see Figure 3).

In unbiased simulations of the metal-bound system the Mg^{2+} ion spontaneously leaves the CDG dimer to form more stable $Mg^{2+}(H_2O)_6$ on a timescale of several hundred picoseconds. This is explained by comparing the partial charges on the water oxygen atom (−0.83e) and the CDG N7 atoms (−0.60e). Mulliken charge analysis of the ab initio calculations

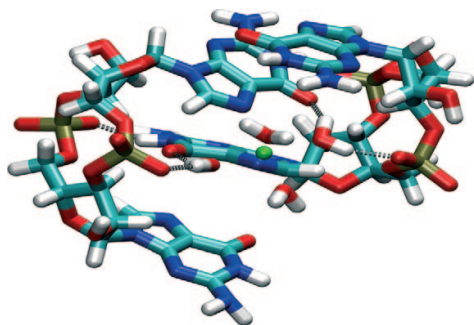


Figure 3. Water molecules form hydrogen bonds (dotted lines) with the O6 atoms of the central bases of c-di-GMP and two oxygen atoms of the phosphate groups.

(see above) on the metal-bound structure reveal that the partial charges on the water oxygen and CDG nitrogen atoms are −0.65e and −0.60e, respectively. Thus, the parameterization of the force field, especially for the nitrogen atoms of CDG are meaningful and the propensity for Mg^{2+} to diffuse out of the CDG dimer into solvent is qualitatively correct. Note that the charges from ab initio calculations also include charge transfer between Mg^{2+} and the water molecules in the first solvation shell. Such effects can not be reliably captured with empirical force fields. As discussed in the next paragraph, additional simulations with slightly increased charges on the CDG N7 atoms (charge models II and III) lead to stabilization of the metal ion.

To sufficiently sample and estimate binding free energies for the Mg^{2+} -bound system two different approaches were explored. In one, the distance between the metal ion and the two N7 atoms of the central guanosine base was biased with an NOE restraint. In the other, the negative partial charge on the N7 atoms was increased (models II and III in the Methods Section) to provide a stronger interaction with the metal ion. Using NOE restraints to maintain the Mg^{2+} ion at a position close to that from the X-ray experiments reveals that the Mg^{2+} -bound CDG dimer remains close to its starting structure (see Figure 4a), which is the X-ray structure (see Figure 1). Even

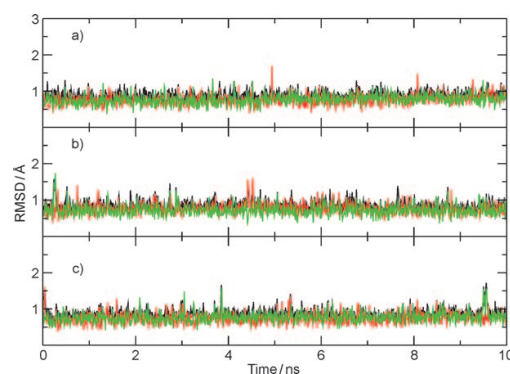


Figure 4. RMSD for CDG (with respect to the X-ray structure) of Mg^{2+} -bound c-di-GMP dimer during the 10 ns MD simulations. CDG dimer (black), first monomer (red), second monomer (green). a) RMSD of Mg^{2+} -bound c-di-GMP dimer with a NOE restraint [K_{min}, K_{max}] = [100, 200] kcal mol^{−1} Å^{−2}. b, c) RMSD for simulations with models II, III.

short openings of one of the monomers that were observed for the metal-free system (Figure 2) do not occur, although this would, in principle, be possible because the constraints only affect the two N7 atoms of the two central guanosine bases close to the Mg^{2+} ion and not the outer bases. To obtain more physical insight, additional simulations were carried out with charge models II and III. The increased negative charges on the two N7 atoms were counterbalanced by increased positive charges on neighboring atoms to maintain an integer overall charge of the system (see Methods Section). It is found that the metal ion remains coordinated to the CDG dimer and the structure of the Mg^{2+} -bound CDG dimer is stable for the entire

10 ns simulation (Figure 4b and c). Figure 5a shows the temporal variation of the N7–Mg distances $r(t)$ for a simulation with NOE constraints [K_{\min}, K_{\max}] of [100, 200] kcal mol⁻¹ Å⁻², and Figure 5b and c show the results of simulations with charge models II and III. The distance $r(t)$ typically fluctuates around 2–2.5 Å, which compares with a separation of 2.3 Å from the X-ray structure.

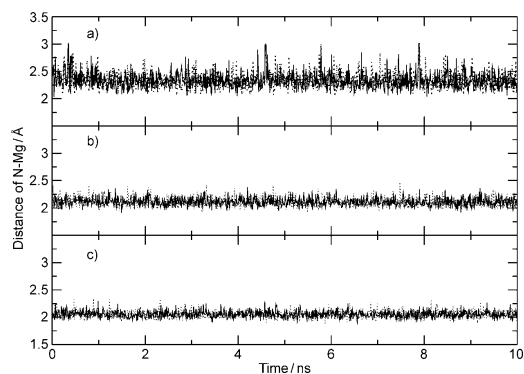


Figure 5. Distance r between Mg²⁺ and two N7 atoms in monomer I (.....) and monomer II (—). a) Analysis from the trajectory using NOE restraint [K_{\min}, K_{\max}] = [100, 200] kcal mol⁻¹ Å⁻² b, c) Analysis from trajectories with different charges redistribution of the guanosine base: b) model II; c) model III.

In summary, the CDG dimer with and without Mg²⁺ appears to be stable on the nanosecond timescale. The average RMSD(t) is lower for the metal-bound dimer compared to the system without metal as long as the metal ion remains coordinated to the CDG dimer. After dissociation, the Mg²⁺ ion is coordinated to six water molecules. Depending on the details of the atomic charges, the charge disparity between the N7 nitrogen atoms of CDG and the oxygen atoms of the surrounding solvent creates a driving force that leads to dissociation of the metal ion and its subsequent solvation.

2.3. Dimerization Free Energy

To more quantitatively assess the stabilization of the metal-free and metal-bound CDG dimer, the binding free energy ΔG_{bind} is calculated from the MD simulations by using MM-GBSA (see Methods Section). Table 2 reports the different contributions to the calculated dimerization free energy for the CDG dimer from three different runs (each 10 ns). The general finding is that the metal-free dimer is slightly favored over two separated monomers (by ≈ 5 kcal mol⁻¹). This is in line with the MD simulations, which do not find dissociation of the dimer. However, the binding is weak, and within typical error bars of MM-GBSA the dimer can be expected to be only marginally stable. The error bars for the individual terms are typical for MM-GBSA, as was already found in ligand-binding studies on HIV-1 protease.^[36]

By using MM-GBSA it is possible to decompose ΔG_{bind} into favorable and unfavorable contributions. The loss of three

Table 2. Dimerization free energies [kcal mol⁻¹] of *c*-di-GMP dimer from three independent 10 ns MD simulations in explicit solvent at $T = 300$ K.^[a]

Energy contribution	Run 1	Run 2	Run 3
$\langle \Delta E_{\text{vdw}} \rangle$	-30.06(4.15)	-30.01(3.87)	-30.76(3.37)
$\langle \Delta E_{\text{ele}} \rangle$	78.88(9.25)	78.56(9.26)	76.94(9.09)
$\langle \Delta G_{\text{ele, desolv}} \rangle$	-67.27(7.58)	-67.80(7.82)	-65.96(7.46)
$\Delta G_{\text{np, desolv}}$	-3.62(0.25)	-3.62(0.23)	-3.69(0.18)
$\langle \Delta G_{\text{bind}}^{\text{CDG}} \rangle + \langle \Delta G_{\text{desolv}} \rangle$	-22.70(5.17)	-22.30(5.33)	-23.47(4.36)
$-T \langle \Delta S_{\text{trans}} \rangle$	13.02	13.02	13.02
$-T \langle \Delta S_{\text{rot}} \rangle$	10.61	10.62	10.62
$-T \langle \Delta S_{\text{vib}} \rangle$	-6.09	-5.89	-5.59
$\langle \Delta G_{\text{bind}} \rangle$	-5.16	-4.55	-5.42

[a] The free energies were calculated using MM-GBSA as outlined in the Methods Section.

translational and three rotational degrees of freedom in the process is entropically unfavorable. The total entropy change ($T \langle \Delta S \rangle = T(\langle \Delta S_{\text{trans}} \rangle + \langle \Delta S_{\text{rot}} \rangle + \langle \Delta S_{\text{vib}} \rangle)$) is negative, because $T \langle \Delta S_{\text{vib}} \rangle$ is too small to compensate the contribution from $T \langle \Delta S_{\text{trans}} \rangle$. So the entropic contribution $-T \langle \Delta S \rangle$ is overall positive and thus destabilizing (see Table 2). It is also of interest to compare the RMSD(t) traces in Figure 2a and c in view of the computed ΔG values. Given the rather stable trajectory in Figure 2c, which shows almost no large-RMSD excursions, it is expected that this dimer is most strongly bound, whereas for the trajectory in Figure 2b multiple attempted separations, in particular between 6 and 7 ns, are observed, which should destabilize the system. This is indeed also reflected in the ΔG values, which differ by 1 kcal mol⁻¹. A distinction between run 1 and run 2 is less obvious because the RMSD(t) in Figure 2a is around 2 Å for 2 ns contrary to an increase of the RMSD(t) in Figure 2b to more than 3 Å for about 1 ns.

For estimating the stabilization of the metal-bound CDG dimer, the definition of the thermodynamic cycle to be considered is less obvious. Because in the X-ray structure the Mg²⁺ ion is coordinated to four water molecules and the CDG dimer, the following scheme was chosen. First, the stability of (CDG)₂Mg²⁺(H₂O)₄ with respect to separated (CDG)₂ + Mg²⁺(H₂O)₄ was determined. In a next step, the stability of (CDG)₂ compared to 2CDG was calculated (see above). This is not a unique decomposition, but given the observation that in unbiased simulations the Mg²⁺ ion spontaneously leaves the CDG dimer, such a decomposition scheme is useful.

In order to have a sufficient number of snapshots from which to obtain meaningful statistics, trajectories with NOE constraints [100, 200] kcal mol⁻¹ Å⁻² on the N7–Mg²⁺ separation were analyzed. Table 3 reports the different contributions to the dimerization free energy of the metal-bound system. The total $\langle \Delta G \rangle$ for the first step [(CDG)₂Mg²⁺(H₂O)₄ → (CDG)₂ + Mg²⁺(H₂O)₄] is 15.8 kcal mol⁻¹, which suggests that Mg²⁺(H₂O)₄ prefers to dissociate from the CDG dimer. This is in line with the observation that in unconstrained simulations of solvated (CDG)₂Mg²⁺(H₂O)₄, the Mg²⁺ ion prefers the aqueous environment (see above). On the other hand, the decomposition energy of (CDG)₂ into two separate monomers is 1.9 kcal mol⁻¹ which is again indicative of a marginally stable CDG dimer and,

Table 3. Different contributions [kcal mol⁻¹] to the dimerization free energy of c-di-GMP with Mg²⁺ from simulations at T = 300 K with NOE restraint [K_{min}, K_{max}] = [100,200] kcal mol⁻¹ Å⁻².^[a]

	a) (CDG) ₂ Mg ²⁺ (H ₂ O) ₄ → Mg ²⁺ (H ₂ O) ₄ + (CDG) ₂ ^[b]	b) (CDG) ₂ → 2CDG ^[b]
⟨ΔE _{vdw} ⟩	8.03(3.85)	-28.14(3.06)
⟨ΔE _{ele} ⟩	-555.58(16.74)	100.97(9.21)
⟨ΔG _{ele,desolv} ⟩	550.49(15.26)	-79.70(6.50)
ΔG _{np,desolv}	-1.94(0.04)	-3.53(0.12)
⟨ΔG _{bind} ^[c] ⟩ + ⟨ΔG _{desolv} ⟩	1.00(6.31)	-10.39(5.64)
⟨ΔH _(a+b) ⟩	-9.39	
-T⟨ΔS _{trans} ⟩	11.82	13.02
-T⟨ΔS _{rot} ⟩	8.09	10.63
-T⟨ΔS _{vib} ⟩	-5.09	-11.38
⟨ΔG _{bind(a/b)} ⟩ ^[c]	15.82	1.88
⟨ΔG _{bind(a+b)} ⟩ ^[c]	17.7	

[a] The total simulation time is 10 ns. [b] a) The Mg²⁺-bound dimer dissociates into the hydrated Mg²⁺ and metal-free dimer. b) The metal-free dimer dissociates into two separate monomers. [c] ⟨ΔG_{bind(a/b)}⟩ is the binding free energy for step a) or b), respectively, whereas ⟨ΔG_{bind(a+b)}⟩ is the free energy difference for the overall process (a and b).

3. Discussion and Conclusions

Despite its central role as a second messenger involved in cell signaling, surprisingly little is known about the structure, dynamics, and energetics of cyclic diguanosine monophosphate at the molecular and atomistic level. Here, atomistic simulations of metal-free and metal-bound CDG were used to better characterize the association state of CDG in solution. Estimated binding free energies from MM-GBSA including explicit solvation and

if error bars are included, in agreement with the results from Table 2. The difference between ΔG calculated from the (CDG)₂ and (CDG)₂Mg²⁺(H₂O)₄ trajectories is that the latter were constrained, which also influences the conformational space available to the individual CDG molecules and therefore affects the estimated ΔG. This can also be seen in comparing the RMSD(t) traces in Figures 2 and 4.

Simulations with different values of the NOE constraint gave slightly different dimerization free energies. Three trajectories (with restraints [0,0], [100,200], and [500,1000]) were analyzed. The binding free energies of Mg²⁺(H₂O)₄ with (CDG)₂ and the total binding energies of Mg²⁺(H₂O)₄ and two CDG monomers are [17.0, 15.8, 14.6] kcal mol⁻¹, and [18.6, 17.7, 15.7] kcal mol⁻¹, respectively, that is, the NOE restraint does not appreciably affect the total binding free energy of the system. The binding free energy for the simulation without restraints [0,0] was evaluated over a 3 ns portion of the [100,200] simulation for which the N7–Mg distance was between 2.0 and 2.9 Å which corresponds to distances for which the NOE constraint is not active (see Methods Section). In addition, the total binding free energy of (CDG)₂Mg²⁺(H₂O)₄ with respect to Mg²⁺(H₂O)₄ + CDG + CDG calculated from trajectories with increased negative partial charges on the two N7 atoms (model III, see Methods Section) is 10.7 kcal mol⁻¹ (see Table 4). Thus, also with increased charges for which the Mg²⁺ ion remains coordinated to (CDG)₂ and the binding free energy is -5.7 kcal mol⁻¹, the (CDG)₂ dimer is not stable.

Table 4. Different contributions to the dimerization free energy of c-di-GMP with Mg²⁺ from simulations at T = 300 K using charge model III and MM-GBSA (See Methods Section).^[a]

	a) (CDG) ₂ Mg ²⁺ (H ₂ O) ₄ → Mg ²⁺ (H ₂ O) ₄ + (CDG) ₂ ^[b]	b) (CDG) ₂ → 2CDG ^[b]
⟨ΔE _{vdw} ⟩	14.42(4.79)	-25.29(3.12)
⟨ΔE _{ele} ⟩	-613.41(15.75)	109.03(9.43)
⟨ΔG _{ele,desolv} ⟩	583.29(12.14)	-80.12(6.65)
ΔG _{np,desolv}	-1.95(0.04)	-3.44(0.13)
⟨ΔG _{bind} ^[c] ⟩ + ⟨ΔG _{desolv} ⟩	-17.65(5.90)	0.18(6.31)
⟨ΔH _(a+b) ⟩	-17.47	
-T⟨ΔS _{trans} ⟩	11.82	13.02
-T⟨ΔS _{rot} ⟩	8.10	10.64
-T⟨ΔS _{vib} ⟩	-7.94	-7.50
⟨ΔG _{bind(a/b)} ⟩ ^[c]	-5.67	16.34
⟨ΔG _{bind(a+b)} ⟩ ^[c]	10.67	

[a] In this simulation the Mg²⁺ ion remains coordinated to the CDG dimer (see text). The total simulation time is 10 ns. [b] a) The Mg²⁺-bound dimer dissociates into hydrated Mg²⁺ and metal-free dimer. b) The metal-free dimer dissociates into two separate monomers. [c] ⟨ΔG_{bind(a/b)}⟩ is the binding free energy for step a) or b), respectively, and ⟨ΔG_{bind(a+b)}⟩ the free energy difference for the overall process (a and b).

the nuclear dynamics suggest that the CDG dimer is only marginally stable with respect to its dissociated form (2CDG), whereas the Mg²⁺ ion readily dissociates from (CDG)₂ and prefers water coordination. The fact that Mg²⁺ dissociates from (CDG)₂ in solution is related to the partial charges on the atoms with which the metal ion interacts preferentially (N7 of CDG and O of water). Electronic structure calculations show that the charges and their relative magnitude used here are meaningful. Simulations with slightly modified charges (models II and III) show that Mg²⁺ indeed stabilizes in the form known from X-ray crystallography. However, it is quite likely that the interaction mode of Mg²⁺ with (CDG)₂ in the crystal and in solution differs and comparison with data from experiments in solution will be very interesting.

Previous work with MM-GBSA has shown that, depending on the system, typical error bars for free energies range between 3 and 6 kcal mol⁻¹.^[29,25] Not surprisingly, calculations excluding entropic effects tend to overestimate experimentally

determined binding free energies. For Ras-Raf the calculated binding free energy of $-47.1(\pm 5.4)$ kcal mol $^{-1}$ compares with -9.6 kcal mol $^{-1}$ from experiment, whereas for Ras-RalGDS and for insulin ΔG without entropic contributions is $-50.0(5.6)$ versus -8.4 and $-38.7(5.8)$ versus -7.2 kcal mol $^{-1}$, respectively.^[29,25] Including entropic contributions brings calculations and experiment into much better agreement. For Ras-Raf $\Delta G_{\text{MM-GBSA}} \approx -15.0(6.3)$ kcal mol $^{-1}$, whereas for Ras-RalGDS and insulin dimer they decrease to $-19.5(5.9)$ and $-11.9(6.7)$ kcal mol $^{-1}$, respectively. For (CDG) $_2$ excluding entropic effects, MM-GBSA yields $\langle \Delta G_{\text{bind}}^0 \rangle + \langle \Delta G_{\text{desolv}} \rangle \approx -22$ kcal mol $^{-1}$ (see Table 2). As for the other systems mentioned above, it is expected that this overestimates the real stabilization considerably. However, it is interesting to note that the solvation energies ΔE_{solv} from most of the quantum chemical calculations are in the range of -12 to -26 kcal mol $^{-1}$ and thus agree quite favorably with the MD results. Including entropic effects, accounting for the errors in MM-GBSA and the fact that MM-GBSA usually leads to overstabilization of a few kilocalories per mole, $\Delta G_{\text{MM-GBSA}} \approx -5$ kcal mol $^{-1}$ (see Table 1) indicates that (CDG) $_2$ in solution is only marginally stable, if at all.

These observations have implications for the molecular mechanisms involved in binding of CDG to its protein partners. First, the results indicate that CDG in the cell is present in its metal-free form. Second, the simulations in solution suggest that the metal-free CDG dimer (CDG) $_2$ is thermodynamically stable by at most 5 kcal mol $^{-1}$ (see Table 2) relative to two monomers. This is in agreement with experimental studies which suggest that, depending on the concentration of CDG present, dimers or higher oligomers form.^[5] The present simulations correspond to a concentration of about 80 mM, which is at the upper range of concentrations found in vivo. Where reported, typical concentrations in cocrystallization experiments range from 0.2 to 5 mM.^[11,37] Average physiological concentrations of CDG are in the submicromolar range.^[38] However, because in vitro experiments are usually carried out with whole-cell extracts and disregard cell structure, local concentrations may be considerably higher.^[5] It is therefore possible and likely that the CDG concentration, and thus the equilibrium between its monomeric and dimeric form, varies considerably under physiologically relevant conditions, which is also consistent with the view that control of c-di-GMP concentrations in bacteria is extremely complex and may also involve concentration gradients.^[39]

The experimental observation of a linear monomer^[11,37,40,41] lead us to consider whether the linear monomer is energetically favored over the U-shaped structure found in (CDG) $_2$. This was quantitatively analyzed here by carrying out ab initio calculations at the B3LYP/6-31G** level on the CDG monomer conformations in different crystal structures. PilZ (PDB code: 2RDE^[11]) and PleD (PDB code: 1W25^[9]) bind a U-shaped monomer whereas FimX, Ykul, and BlrP1 bind a stretched monomer (PDB codes: 2W27,^[41] 3V8,^[11] 3GFX^[37]). To compare these structures, the CDG monomers were all optimized to a local minimum, and total energies in vacuo and with implicit solvent (see above) were determined. In vacuo the linear structure from Ykul (PDB code: 2W27) is lowest in energy followed by

the optimized U-shaped structure from PleD (2.2 kcal mol $^{-1}$ higher). Other conformations are destabilized by between 2.8 and 11.4 kcal mol $^{-1}$. Including solvation effects makes the U-shaped structure from PleD lowest in energy, followed by the linear structure from 2W27 (1.1 kcal mol $^{-1}$), and the remaining local minima are destabilized by between 2.1 and 7.9 kcal mol $^{-1}$. Starting from a U-shaped structure from PleD, the CDG monomer adopts a more elongated conformation upon optimization (see Figure 6). Thus, the potential-energy

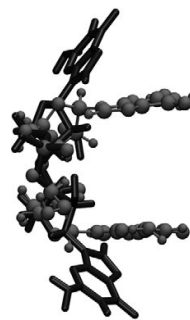


Figure 6. Structures of the CDG monomer from PleD (PDB code 1W25). Ball-and-stick structure: crystal structure of the U-shaped CDG monomer from the intercalated dimer. Line structure: optimization structure of the U-shaped CDG monomer from the intercalated dimer.

surface seems to be quite rugged, to support a larger number of local minima, and CDG appears to be sufficiently flexible to adapt to linear and U-shaped structures likewise with a relatively small energy penalty. This is also required given the different protein environments it must adapt to. The electronic structure calculations support the notion that CDG monomer is flexible and able to either adopt quasilinear (extended) or more U-shaped (compact) structures.

Recently, it was demonstrated that activated PleD specifically localizes at the developing pole of *Caulobacter crescentus* cells. Coupling of activation and subcellular sequestration could indicate local concentration maxima.^[8,42] The results presented here suggest that what binds to PleD in such high CDG concentration regions is actually (CDG) $_2$. Finally, experimentally, it is found that both, CDG monomers (PleD,^[9] BlrP1,^[37] Ykul,^[40] FimX,^[41] PilZ^[11]) and dimers (PleD,^[9] WspR,^[10] PP4397^[12]) bind to proteins. Therefore, the simultaneous presence of CDG and (CDG) $_2$ may be advantageous or even required in a concrete biological context. Furthermore, the weak preference of (CDG) $_2$ over 2CDG found in the present case for the high-concentration limit would provide the cell with a flexible mechanism to make both forms of the messenger available. Consequently, at any given instance, there will be a more or less significant amount of CDG or (CDG) $_2$, which is capable of binding as a monomer or as a dimer, respectively, depending on the requirements of the protein. Consequently, the proposal of protein-induced sequestration of (CDG) $_2$ in PleD, although still speculative, may be a very appropriate picture.^[43]

Acknowledgements

We wish to acknowledge the use of the Chemical Database Service at Daresbury and helpful comments by Prof. U. Jenal and Prof. T. Schirmer. The authors gratefully acknowledge financial support from the Swiss National Science Foundation through grant number 200021-117810.

Keywords: cyclic diguanylic acid · density functional calculations · electronic structure · magnesium · molecular dynamics

- [1] R. Hengge, *Nat. Rev. Microbiol.* **2009**, *7*, 263–273.
- [2] T. Schirmer, U. Jenal, *Nat. Rev. Microbiol.* **2009**, *7*, 724–735.
- [3] H. Yan, W. Chen, *Chem. Soc. Rev.* **2010**, *39*, 2914–2924.
- [4] M. Egli, R. V. Gessner, L. D. Williams, G. J. Quigley, G. A. van der Marel, J. H. van Boom, A. Rich, C. A. Frederick, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 3235–3239.
- [5] Z. Zhang, S. Kim, B. L. Gaffney, R. A. Jones, *J. Am. Chem. Soc.* **2006**, *128*, 7015–7024.
- [6] Y.-C. Liaw, Y.-G. Gao, H. Robinson, G. M. Sheldrick, L. A. J. M. Sliedregt, G. A. van der Marel, J. H. van Boom, A. H.-J. Wang, *FEBS Lett.* **1990**, *264*, 223–227.
- [7] Y. Guan, Y.-G. Gao, Y.-C. Liaw, H. Robinson, A. H.-J. Wang, *J. Biomol. Struct. Dyn.* **1993**, *11*, 253–276.
- [8] R. Paul, S. Weiser, N. C. Amiot, C. Chan, T. Schirmer, B. Giese, U. Jenal, *Genes Dev.* **2004**, *18*, 715–727.
- [9] C. Chan, R. Paul, D. Samoray, N. C. Amiot, B. Giese, U. Jenal, T. Schirmer, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 17084–17089.
- [10] N. De, M. Pirruccello, P. V. Krasteva, N. Bae, R. V. Raghavan, H. Sondermann, *PLoS Biol.* **2008**, *6*, 0601–0617.
- [11] J. Benach, S. S. Swaminathan, R. Tamayo, S. K. Handelman, E. Folta-Stogniew, J. E. Ramos, F. Forouhar, H. Neely, J. Seetharaman, A. Camilli, J. F. Hunt, *EMBO J.* **2007**, *26*, 5153–5166.
- [12] J. Ko, K.-S. Ryu, H. Kim, J. -S.-Shin, J.-O. Lee, C. Cheong, B.-S. Choi, *J. Mol. Biol.* **2010**, *398*, 97–110.
- [13] Gaussian 03 (Revision B.04), M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Rob, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian Inc., Wallingford, CT, **2004**.
- [14] H. M. Berman, T. Battistuz, T. N. Bhat, W. F. Bluhm, P. E. Bourne, K. Burkhardt, G. L. G. Z. Feng, L. Iype, S. Jain, J. M. P. Fagan, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J. D. Westbrook, C. Zardecki, *Acta Crystallogr. Sect. D* **2002**, *58*, 899–907.
- [15] D. A. Fletcher, R. F. McMeeking, D. Parkin, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 746–749.
- [16] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187–217.
- [17] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. E. Straub, M. Watanabe, J. Wiorcikiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- [18] B. Christen, M. Christen, R. Paul, F. Schmid, M. Folcher, P. Jenoe, M. Meuwly, U. Jenal, *J. Biol. Chem.* **2006**, *281*, 32015–32024.
- [19] F. F.-F. Schmid, M. Meuwly, *J. Mol. Biol.* **2007**, *374*, 1270–1285.
- [20] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.
- [21] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.
- [22] J. Srinivasan, T. E. Cheatham III, P. Cieplak, P. A. Kollman, D. A. Case, *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409.
- [23] P. A. Kollman, I. Massoval, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, T. E. Cheatham III, *Acc. Chem. Res.* **2000**, *33*, 889–897.
- [24] H. Gohlke, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 238–250.
- [25] V. Zoete, M. Meuwly, M. Karplus, *Proteins Struct. Funct. Bioinf.* **2005**, *61*, 79–93.
- [26] M. S. Lee, F. R. Salsbury Jr., C. L. Brooks III, *J. Chem. Phys.* **2002**, *116*, 10606–10614.
- [27] M. S. Lee, M. Feig, F. R. Salsbury, Jr., C. L. Brooks III, *J. Comput. Chem.* **2003**, *24*, 1348–1356.
- [28] W. C. Still, A. Tempczyk, R. C. Hawley, T. Hendrickson, *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- [29] H. Gohlke, C. Kiel, D. A. Case, *J. Mol. Biol.* **2003**, *330*, 891–913.
- [30] B. Tidor, M. Karplus, *J. Mol. Biol.* **1994**, *238*, 405–414.
- [31] B. Mennucci, J. Tomasi, *J. Chem. Phys.* **1997**, *106*, 5151–5158.
- [32] S. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553.
- [33] S. Simon, M. Duran, J. Dannenberg, *J. Chem. Phys.* **1996**, *105*, 11024–11031.
- [34] J. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- [35] Y. Zhao, D. Truhlar, *J. Chem. Theory Comput.* **2005**, *1*, 415–432.
- [36] H. B. Thorsteinsdottir, T. Schwede, V. Zoete, M. Meuwly, *Proteins Struct. Funct. Bioinf.* **2006**, *65*, 407–423.
- [37] T. R. M. Baerends, E. Hartmann, J. J. Griese, T. Beitlich, N. V. Kirienco, D. A. Ryjenkov, J. Reinstein, R. L. Shoeman, M. Gomelsky, I. Schlichting, *Nature* **2009**, *459*, 1015–1018.
- [38] R. Simm, M. Morr, U. Remminghorst, M. Andersson, U. Römling, *Anal. Biochem.* **2009**, *386*, 53–58.
- [39] U. Römling, *Cell. Mol. Life Sci.* **2005**, *62*, 1234–1246.
- [40] G. Minasov, S. Padavattan, L. Shuvalova, J. S. Brunzelle, D. J. Miller, A. Basle, C. Massa, F. R. Collart, T. Schirmer, W. F. Anderson, *J. Biol. Chem.* **2009**, *284*, 13174–13184.
- [41] M. V. A. S. Navarro, N. De, N. Bae, Q. Wang, H. Sondermann, *Structure* **2009**, *17*, 1104–1116.
- [42] R. Paul, S. Abel, P. Wassmann, A. Beck, H. Heerklotz, U. Jenal, *J. Biol. Chem.* **2007**, *282*, 29170–29177.
- [43] R. Paul, T. Jaeger, S. Abel, I. Wiederkehr, M. Folcher, E. G. Biondi, M. T. Laub, U. Jenal, *Cell* **2008**, *133*, 452–461.

Received: August 27, 2010

Revised: November 3, 2010

Published online on January 14, 2011

4.2 Dynamics of Analogue endo-S-c-di-GMP in Solution

In the previous section, the structure and association affinity of c-di-GMP in solution were investigated. Through interaction with different receptors, c-di-GMP negatively modulates cell motility and traits associated with bacterial virulence and stimulates several biofilm-associated functions. High concentrations of c-di-GMP promote sessile (biofilm formation) growth while low concentrations of c-di-GMP promote motile growth. The aggregation of c-di-GMP into higher aggregates will reduce its concentration in the cell. It will be of interest to determine which moieties of c-di-GMP promote aggregate formation.

In the next section, to further characterize c-di-GMP, investigations on a c-di-GMP analogue called endo-S-CDG are presented. endo-S-CDG is obtained by modification of one of the phosphate groups in c-di-GMP with a sulfur in the phosphodiester linkage. To compare with c-di-GMP, MD simulations with endo-S-CDG in solution were carried out. Moreover, endo-S-CDG binding to PleD protein was studied to reveal the role played by the phosphodiester linkage in c-di-GMP dimerization and binding to processing proteins. A paper is in preparation.

Dynamics of Analogue endo-S-c-di-GMP in Solution

Lixian Zhang and Markus Meuwly*

*University of Basel, Department of Chemistry,
Klingelbergstrasse 80, CH-4056, Basel, Switzerland.*

Abstract

Cyclic diguanosine-monophosphate (c-di-GMP, CDG) is a bacterial signaling molecule that triggers a switch from motile to sessile bacterial lifestyles^{1,2}. Analogues of c-di-GMP, which can selectively modulate the activities of c-di-GMP processing proteins, will be useful chemical tools for studying and altering bacterial behavior. Recent studies^{3,4} revealed that c-di-GMP is a monomeric state at low micromolar concentration ($1 \mu\text{M}$) in the absence of metal ions and dimerization may occur only on the proteins (i.e., diguanylate cyclase, PilZ). Higher oligomer formation occurs only in the presence of monovalent (particularly K^+) metal ions. Another report⁵ showed that a conservative modification of one of the phosphate groups in c-di-GMP with a bridging sulfur in the phosphodiester linkage affords an analogue called endo-S-c-di-GMP does not readily form higher aggregates. Local c-di-GMP pools with higher concentrations have been discussed¹, but so far remain hypothetical. In our work, we presented Molecular Dynamic Simulation on endo-S-c-di-GMP in aqueous solution with a concentration of 80 mMol. The result shows the dimerization free energy of endo-S-c-di-GMP is -0.43 kcal/mol compared to -5 kcal/mol for the c-di-GMP dimer, which indicates that the former is less favoured in dimeric form. This finding suggests that the phosphate in c-di-GMP is important in aggregate formation and may also in the binding of c-di-GMP to proteins.

I. INTRODUCTION

Cyclic diguanosine-monophosphate (c-di-GMP) is a ubiquitous second messenger that regulates cell surface-associated traits in bacteria. It has attracted considerable attention from different fields since it was discovered 20 years ago.¹⁻¹⁴ Cyclic di-GMP is a monocyclic RNA dinucleotide and has a conformation in which the 12-membered circular sugar-phosphate backbone provides a rigid framework to hold the two adenines 6.8 Å apart in parallel planes.⁷ It is produced from 2 molecules of GTP by the activity of diguanylate cyclases (DGCs) and is degraded by specific phosphodiesterase (PDE) into 5'-phosphoguananylyl-(3'-5')-guanosine (pGpG); pGpG is subsequently split into two GMP molecules. Through the interaction with different receptors, such as PilZ¹⁵ - containing proteins, c-di-GMP negatively modulates cell motility and traits associated with bacterial virulence and stimulates several biofilm-associated functions. High concentrations of c-di-GMP promote sessile (biofilm formation) growth while low concentrations of c-di-GMP promote motile growth.

Binding of c-di-GMP to a secondary site (I site) that is present in most DGCs results in feedback product inhibition of these enzymes and contributes to a physiological upper limit to cellular accumulation of c-di-GMP.¹⁶ Indeed, two c-di-GMP molecules can form a mutually intercalated dimer structure stabilised by multiple forces including hydrogen bonds from the base of one molecule to the sugar phosphate backbone of the other, and guanine guanine base-pairing interactions.^{3,6,7}

Chemical modifications of c-di-GMP have been recently studied because structural changes may affect the aggregation properties of the molecule and also modulate its affinity and interaction with the proteins to which it binds. One conservative chemical change to c-di-GMP is to modify one of the oxygen atoms in the phosphodiester linkage by a bridging sulfur which yields the endo-S-CDG analogue (see Figure 1). Experimentally, endo-S-CDG was studied by using NMR and CD spectroscopies which found a reduced aggregation propensity compared to c-di-GMP.⁵ Furthermore endo-S-CDG binds to RocR, a PDE that binds monomeric c-di-GMP, and inhibits hydrolysis while it did not bind to WspR (to which dimeric c-di-GMP binds) or the PilZ domain protein Alg44.¹⁷ This demonstrates that selective binding to different classes of c-di-GMP binding proteins is possible by chemically altering c-di-GMP.

4.2 Dynamics of Analogue *endo-S-c-di-GMP* in Solution

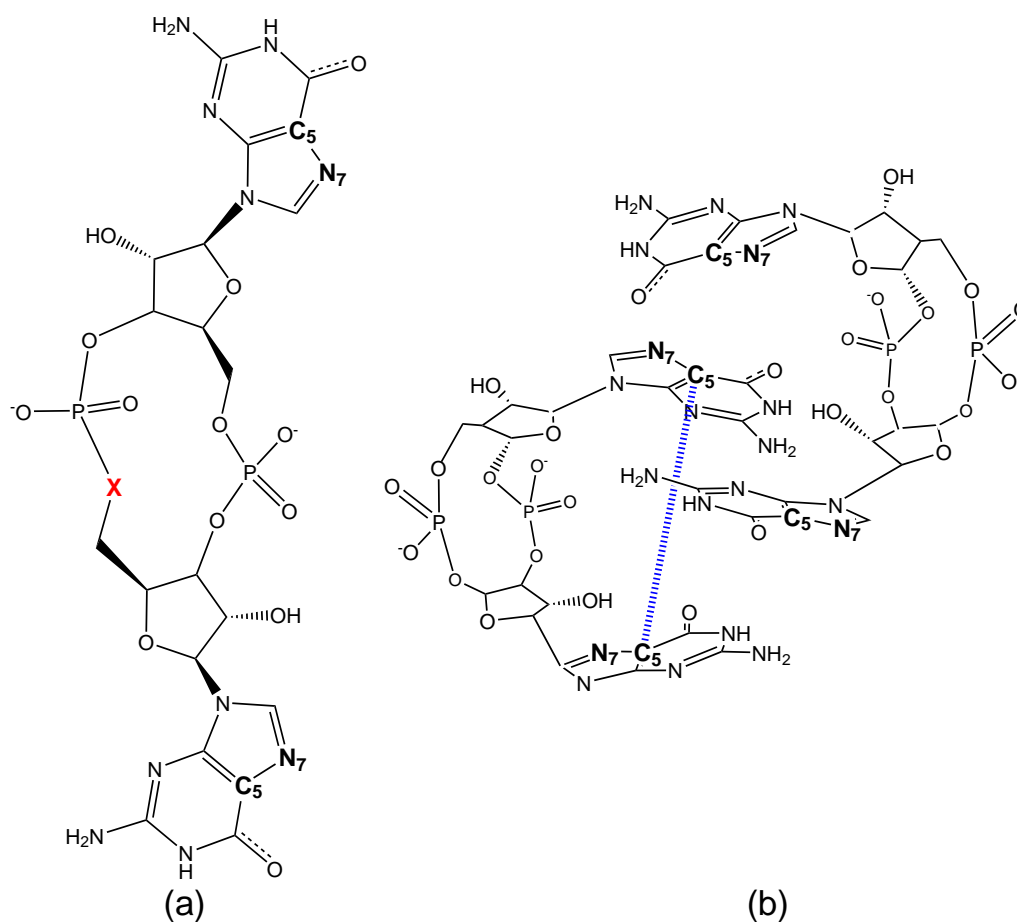


FIG. 1: (a) Structure of *c-di-GMP* monomer ($X = O$) and *endo-S-CDG* by substituting the bridging Oxygen atom by sulfur. Atoms C5, and N7 are labelled as well. (b) Structure of intercalated *c-di-GMP* dimer.

Currently, there is no available crystal structure in the PDB that contains *endo-S-c-di-GMP* cocrystallized in a host protein. As is shown in Table I, both closed and open *c-di-GMP* can be bound to the active site of a range of proteins, including Wspr^{18,19}, PleD^{16,20}, the PilZ domain^{15,21} Ykui²², Fimx,²³ or LapD²⁴. DGC and PilZ-domain-containing proteins bind closed *c-di-GMP* whereas EAL-domain-containing proteins bind open *c-di-GMP*.

In the present work we use computational techniques to study the aggregation behaviour of *endo-*

4 Bacterial Second Messenger *c*-di-GMP

TABLE I: Proteins, Domains containing *c*-di-GMP in the protein data bank

Domain	Protein	PDB code	Unit	Num. and form CDG
EAL	Ykui	2W27 ²⁵	Dimeric	2 open CDG
EAL	Fimx	3HV8 ²⁶	Monomeric	1 open CDG
BLUF-EAL	BlrP1	3GG1 ²⁷	Dimeric	2 open CDG
BLUF-EAL	BlrP1	3GG0 ²⁷	Dimeric	2 open CDG
BLUF-EAL	BlrP1	3GFZ ²⁷	Dimeric	2 open CDG
BLUF-EAL	BlrP1	3GFY ²⁷	Dimeric	2 open CDG
BLUF-EAL	BlrP1	3GFX ²⁷	Dimeric	2 open CDG
EAL	LapD	3PJU ²⁸	Dimeric	2 open CDG
EAL	LapD	3PJT ²⁸	Dimeric	2 open CDG
EAL-GGDEF	Q3SJE6	3N3T ²⁹	Dimeric	2 open CDG
REC-GGDEF	Wspr	3I5A ³⁰	Tetrameric	1 dimer
REC-GGDEF	Wspr	3BRE ³¹	Dimeric	2 dimer
REC-REC-GGDEF	PleD	2V0N ³²	Dimeric	2 dimer
REC-REC-GGDEF	PleD	1W25 ³³	Dimeric	2 dimer, 1 closed CDG
REC-REC-GGDEF	PleD	2WB4	Dimeric	2 dimer, 1 closed CDG
GGDEF	XCC4471	3QYY ³⁴	Dimeric	2 closed CDG
bzzp-1 GGDEF	Wspr/GCn4	3I5C ³⁵	Dimeric	2 closed CDG
GGDEF	MqR89a	3IGN	Monomeric	1 closed CDG
GGDEF- <i>GAF</i> ₂	Peld156-455	4DN0 ³⁶	Monomeric	1 closed CDG
PilZ	pa4608	2L74 ³⁷	Monomeric	1 dimer
PilZ	pp4397	3KYF ³⁸	Monomeric	1 dimer
PilZ	PlzD(Vca0042)	2RDE ³⁹	Dimeric	2 closed CDG
PilZ	XAC1133	3KYG ⁴⁰	Dimeric	2 dimer

S-CDG in gas phase and in solution. For this, electronic structure calculations and molecular dynamics simulations with explicit solvation are carried out. From the latter the binding free energy of the dimer in solution can be estimated which provides information about the aggregation state in solution. Finally, we also consider the dynamics of endo-S-CDG and c-di-GMP in their monomeric and dimeric forms in the typical c-di-GMP-binding protein PleD.

II. COMPUTATIONAL METHODS

A. Electronic Structure Calculations

Density functional theory (DFT) calculations were carried out with Gaussian 03⁴¹ at the B3LYP/6-31G(d,p) level.^{42,43} The starting structure for endo-S-CDG was obtained by substituting one of the oxygen atoms in the phosphodiester linkage in c-di-GMP ([Brookhaven Protein Data Bank (PDB)⁴⁴]. For the Mg²⁺-bound c-di-GMP dimer coordinates from the Cambridge Structural Database were used.^{6,45} The overall charge is -4 for the mutually intercalated endo-S-CDG dimer and -2 for the Mg²⁺-bound endo-S-CDG dimer.

B. Molecular Dynamics Simulations

Molecular dynamics (MD) simulations for the dinucleotides were carried out with the CHARMM program⁴⁶ using the CHARMM22 force field⁴⁷ with periodic boundary conditions (PBC). Force field parameters for c-di-GMP were those determined and used previously.^{3,48,49} Force field parameters for sulfur were fitted to reproduce the energy surface scans of the respective coordinates obtained from DFT calculations. Bonded force field parameters (see below) involving the sulfur atom were fitted to reproduce DFT energy scans of the respective internal coordinates which included the P-S and S-C distances, the CCS, CSP, HCS and SPO angles. The coordinates were scanned on a grid of 0.05 \AA , 1° around the minimum structure for 30 points. The fitting was performed using Chnolls⁵⁰, the CHARMM interface for Inolls.⁵¹ The overall correlation coefficient for the fit is $R = 0.90$ with a mean square deviation for the energies of 2.89 kcal/mol . This data is shown in figure 1 of the supplementary information. The Mulliken charges from DFT

4 Bacterial Second Messenger *c*-di-GMP

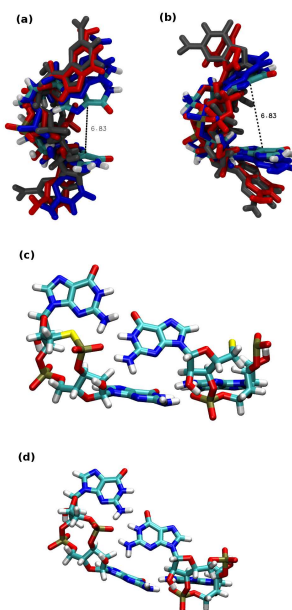


FIG. 2: Optimized geometry of (a) endo-S-CDG monomer, (b) CDG monomer, (c) endo-S-CDG dimer, (d) CDG dimer from DFT calculation at B3LYP/6-31G(d,p) level. Four structures are shown in (a) and (b) respectively, corresponding to constraint and fully optimized structures of U-shaped endo-S-CDG, CDG (PDB code: 1w25) and open-shaped endo-S-CDG, CDG (PDB code: 2W27²²). For U-Shaped endo-S-CDG, CDG, the distance of C5 in the guanine of CDG is 6.83 Å, while for open-shape endo-S-CDG, CDG, the distance of C5 is 13.5 Å. Fully optimized endo-S-CDG, CDG, the distance of C5 is between 6.8 –13.5 Å.

minimizations were used for the MD simulations.

The structures of the metal-free and Mg^{2+} -bound endo-S-CDG dimer were optimized with 5000 steps of steepest descent minimization. Then the complexes were individually solvated in a pre-equilibrated water box of dimension $37.3 \times 34.1 \times 34.1 \text{ \AA}^3$. The solvent was optimized and equilibrated at 300 K for 30 ps in the presence of the fixed solute. Finally, the structures of the solutes were relaxed with 2000 steps of steepest descent minimization and then gradually

4.2 Dynamics of Analogue endo-S-c-di-GMP in Solution

heated to 300 K for 200 ps followed by 1 ns of equilibration. All MD simulations were carried out by using SHAKE⁵² to constrain the bonds involving hydrogen atoms. For the metal-bound and metal-free endo-S-CDG dimer 3 independent trajectories (60 ns in total) and 12 independent trajectories (240 ns in total) were run, respectively.

For simulations of the Mg^{2+} -bound endo-S-CDG dimer the metal ion has a finite probability to escape to the solvent because it only interacts via electrostatic interactions with its environment. Therefore, the MD simulations used a NOE constraint to maintain the magnesium ion closer to the N7 nitrogen atoms of the central guanosine. The force constants $[K_{\min}, K_{\max}]$ were [100,200] kcal/(mol \AA^2).

For endo-S-CDG and CDG dimer bound to PleD, MD simulations were carried out with either CHARMM or NAMD⁵³ using the CHARMM22 force field. The starting structures are the A chain from the X-ray dimer (PDB code: 1W25²⁰) and the GGDEF domains of X-ray dimer (PDB code: 2WB4), respectively. Thus for both proteins the simulation systems consist of the protein monomer, an intercalated endo-S-CDG/CDG dimer in the inhibition (I-)site, and a endo-S-CDG/CDG monomer in the active (A-) site. The systems have an overall charge of $-9e$ and $-14e$ for inactive PleD (1W25) and the active PleD (2WB4), respectively. First, 500 steps of steepest descent minimization were run. followed by heating and equilibration dynamic in a pre-equilibrated waterbox of TIP3P⁵⁴ water molecules extending 10 \AA in all dimensions around the solute which leads to simulation systems consisting of ≈ 100000 atoms. Afterwards, the systems were neutralized with an adequate number of Na^+ counter-ions and PME⁵⁵ was employed for the calculation of the electrostatics beyond the cutoff of 12 \AA of the explicit calculation. All systems were heated for 30 ps to 300 K and equilibrated for 200 ps with the solute fixed and then totally relaxed for another 1 ns before production simulations started. The simulation length is 20 ns for each trajectory.

C. Dimerization Free Energy

The binding free energy ΔG consists of an enthalpic (ΔE) and an entropic contribution (ΔS), which can be calculated according to a thermodynamic cycle.^{56,57} Here, the molecular mechan-

4 Bacterial Second Messenger *c*-di-GMP

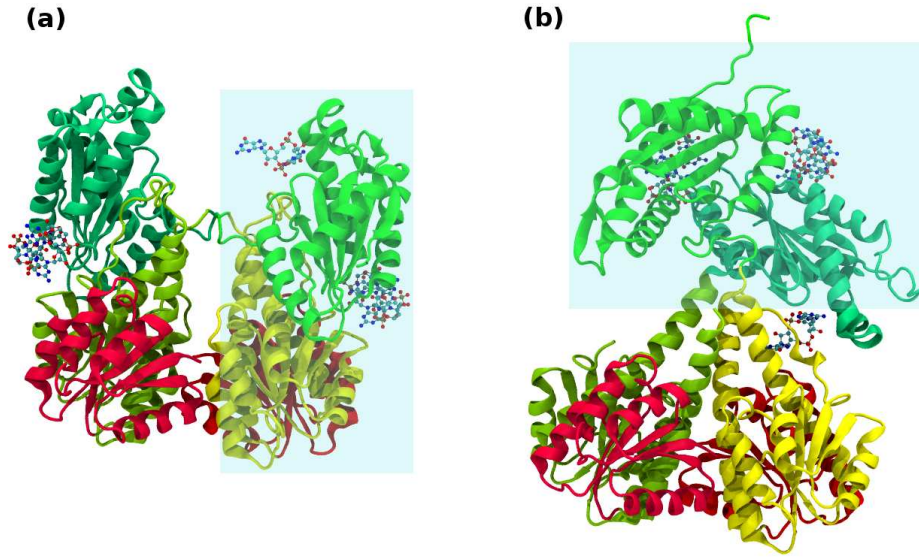


FIG. 3: Structure of inactive (panel a, PDB code: 1W25) and activated (panel b, PDB code: 2WB4) PleD. The part marked by cyan color is the system studied here. Green: DGC; red: D1 domain; yellow: D2 domain.

ics with generalized Born surface area (MM-GBSA) approach⁵⁸ is used which decomposes the binding free energy, ΔG_{bind} into a sum of the gas phase contribution, the desolvation energy of the system upon binding, ΔG_{desolv} , and an entropic contribution, $-T\Delta S$:

$$\Delta G_{\text{bind}} = \langle \Delta G_{\text{bind}}^0 \rangle + \langle \Delta G_{\text{desolv}} \rangle - T \langle \Delta S \rangle \quad (1)$$

The brackets indicate averages of the individual terms along an equilibrium MD simulation⁵⁹. The gas phase contribution, $\langle \Delta G_{\text{bind}}^0 \rangle$, consists of the difference in van der Waals and electrostatic energies between the dimer (D) and the individual monomers (M) whereas the desolvation term consists of a nonpolar contribution which is assumed to be proportional to the solvent-accessible surface area (SASA) and an electrostatic contribution calculated with the GB-MV2 model.^{60,61} A value of $0.0072 \text{ kcal}/(\text{mol } \text{\AA}^2)$ was used to relate the buried SASA to the non-polar desolvation

4.2 Dynamics of Analogue endo-S-c-di-GMP in Solution

free energy^{59,62,63}.

$$\begin{aligned}\Delta G_{\text{bind}} &= \langle \Delta G_{\text{bind}}^0 \rangle + \langle \Delta G_{\text{desolv}} \rangle - T \langle \Delta S \rangle \\ &= \langle \Delta E_{\text{intra}} \rangle + \langle \Delta E_{\text{vdw}} \rangle + \langle \Delta E_{\text{elec}} \rangle + \langle \Delta G_{\text{ele,desolv}} \rangle + \langle \Delta G_{\text{np,desolv}} \rangle - T \langle \Delta S \rangle\end{aligned}\quad (2)$$

The energies calculated for snapshots taken every 10 ps using the "same trajectory method" (STM)⁵⁹ were averaged over 20 ns. In the STM, the energy terms relative to the isolated monomers are calculated using coordinates taken from the simulation of the dimer. The entropy term is partitioned into translational, rotational and vibrational terms following standard equations of statistical mechanics.⁶⁴ The translational contribution only depends on the mass of the system and the rotational part is related to the moments of inertia. The vibrational part can be calculated from the normal modes (NM) ν_i :

$$-T\Delta S_i = \frac{h\nu_i}{e^{\frac{h\nu_i}{k_B T}} - 1} - k_B T \ln(1 - e^{-\frac{h\nu_i}{k_B T}})\quad (3)$$

where k_B is the Boltzmann constant h is the Planck constant and $T = 300$ K. The VIBRAN module of the CHARMM program was used to calculate and diagonalize the force constant matrix to determine the normal mode vectors and frequencies. Normal modes were calculated both for the fully minimized structures of the dimer and the monomers in vacuo with a distance dependent dielectric ($\epsilon = 4$), and a cutoff of 12 Å for non-bonded interactions was applied. The snapshots were minimized using the adopted basis Newton Raphson (ABNR) method until a gradient of 10^{-7} was reached. All reported values are averaged over 1000 frames, equally distributed over the 20 ns trajectories.

III. RESULTS AND DISCUSSION

A. Electronic Structure Calculations

DFT calculations for the minimum energy conformation of the endo-S-CDG monomer were started from the closed and the open conformation. The two states are characterized by a C5–C5 distance of 6.8 Å and 13.5 Å, respectively, see Figure 1. For the endo-S-CDG dimers the optimized structures of the metal-free and hydrated Mg^{2+} ($\text{Mg}^{2+}(\text{H}_2\text{O})_4$) were determined. For all systems the structures were fully optimized starting from the X-ray coordinates. Additionally, constrained optimizations, whereby the C5–C5 distance was fixed at 6.8 Å (U-shaped, "closed") and 13.5 Å (fully extended, "open"), respectively, were carried out. These distances correspond to

4 Bacterial Second Messenger *c*-di-GMP

the distance of the monomer in the X-ray structure of the *c*-di-GMP dimer⁶ and that in the crystal together with Protein PleD²⁰. For analyzing the structures of monomeric and dimeric endo-S-CDG the structures of *c*-di-GMP in the crystal structures of 1W25²⁰, 2WB4⁴⁴, 2W27²² were used whereas for the Mg²⁺-bound dimer the reference was the X-ray structure of Mg²⁺-bound *c*-di-GMP.⁶

The RMSD of the closed endo-S-CDG monomer with and without C5 constraints, and the open endo-S-CDG monomer with and without C5 constraint are 0.62, 2.03, 1.01, and 1.11 Å, respectively. For the constrained systems, which are less relevant, similar RMSDs were found for conventional *c*-di-GMP (CDG) - they are 0.47 and 1.14, compared to 0.62 and 1.01 Å for endo-S-CDG - whereas the unconstrained optimizations remain closer to the X-ray structure in the case of conventional CDG (they are 0.80 and 0.60, compared to 2.03 and 1.11 Å).³ Hence, replacing an oxygen atom by a sulfur atom leads to observable differences in the minimum energy structure, in particular for the monomer. For the dimeric structures, the RMSD to the X-ray structure for CDG and endo-S-CDG are similar. For the endo-S-CDG dimers with and without hydrated Mg²⁺ bound to it the RMSDs are 0.32 and 3.21 Å, respectively.

The energies of endo-S-CDG dimer and endo-S-CDG dimer with Mg²⁺ from DFT calculations are reported in Table II. The destabilization energy of the endo-S-CDG dimer in the gas phase relative to two separated monomers is $\Delta E_1 = E^{\text{dimer}} - (E_{\text{M1}}^{\text{dimer}} + E_{\text{M2}}^{\text{dimer}}) = 27.6$ kcal/mol. Including solvation effects this leads to a stabilization by -19.4 kcal/mol. As a comparison, the CDG dimer has been found to be stabilized by -27.8 kcal/mol relative to the separated monomers. This suggests that both, endo-S-CDG and CDG are stable in their dimeric forms although CDG is more strongly bound. For the Mg²⁺-bound endo-S-CDG dimer the binding energies ΔE are -555.8 and -111.9 kcal/mol in gas phase and solution, respectively.

Optimized structures, with and without constraint of the C5 distance for both, the endo-S-CDG and the CDG monomer, are shown in Figures 4a,b. The total electronic energies of both monomers in solution are reported in Figure 4 and Table II relative to the energy of the constrained optimization of the closed monomers. Starting from the closed conformer, the minimum energy structure for endo-S-CDG is stabilized by -3.50 kcal/mol whereas for CDG it is almost isoenergetic which suggests that the potential energy surface is very flat. Starting from the constrained optimization

4.2 Dynamics of Analogue endo-S-c-di-GMP in Solution

TABLE II: Total energies of minimized endo-S-CDG monomers, M1,M2 and dimers (D) with and without the solvated Mg^{2+} ion. Total energies in Hartrees (E_h), relative stabilization energies in kcal/mol. Structures in vacuum are optimized starting from the X-ray structure and give the vacuum energy E_{vac} ; energies with implicit solvent E_{solv} (PCM) are single-point calculations on the optimized vacuum structure.

Structure	E_{vac}/E_h	$\Delta E_{\text{vac}}/(\text{kcal/mol})$	E_{solv}/E_h	$\Delta E_{\text{solv}}/(\text{kcal/mol})$
D	-6764.05667003		-6764.66080314	
M1	-3382.04943669		-3382.31214372	
M2	-3382.03477879		-3382.30356901	
		17.3		-28.3
$2 \times \text{M}$	-6764.10064174		-6764.62985988	
		27.6		-19.4
$2 \times \text{M-close}$	-6764.09456578	23.8	-6764.62013426	-25.5
$\text{D}(\text{Mg}^{2+}(\text{H}_2\text{O})_4)$	-7270.18691562		-7270.44294429	
M1	-3381.99433357		-3382.28445380	
M2	-3381.99445385		-3382.28476163	
$(\text{Mg}^{2+}(\text{H}_2\text{O})_4)$	-505.312501382		-505.695405589	
		-555.8		-111.9

of the extended structures, the two fully optimized structures lead to a slight contraction with concomitant energy reductions by 0.2 and 0.4 kcal/mol. Again, the PES is very flat. Overall, both extended structures are energetically favoured over the U-shaped structure: for endo-S-CDG the stabilization is 1.0 kcal/mol whereas for CDG it is 2.5 kcal/mol. The small energy differences suggest that in solution both systems exist in a conformational equilibrium consisting of extended and partially closed structures. However, for endo-S-CDG the energy differences between the two states is somewhat larger (3.5 kcal/mol) than for CDG (2.5 kcal/mol). Thus, it is expected that the CDG-dimer is more stable than the endo-S-CDG-dimer.

4 Bacterial Second Messenger *c*-di-GMP

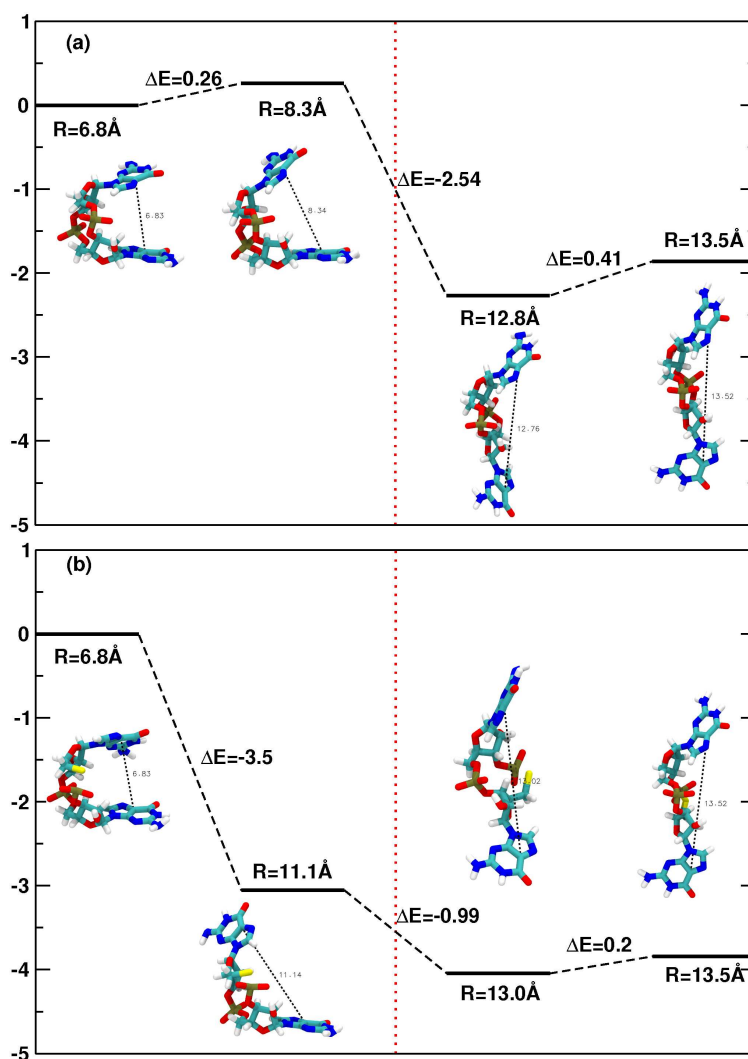


FIG. 4: The energy of CDG and endo-S-CDG in solution from electronic structure calculation. R is the distance of C5 in the guanine of CDG. The left panel shows the results of close CDG structures—partially optimized CDG (constraint the C5 distance at 6.8\AA) and fully optimized ($R = 8.3\text{\AA}$ for CDG and $R = 11.1\text{\AA}$ for endo-S-CDG); the right panel is for open CDG structures—constraint the distance of C5 at 13.5\AA and without constraint just fully optimized ($R = 12.8\text{\AA}$ and 13.5\AA).

B. Dimer Dynamics in Solution

In order to characterize the dynamics of the *endo-S*-CDG dimer, atomistic simulations in full solvation were carried out. Figure 5a shows the root mean square deviation (RMSD) from the X-ray structure, calculated for the entire *endo-S*-CDG dimer. Snapshots are taken every 10 ps. The RMSD was calculated after optimally superimposing the atoms of the instantaneous conformation of the dimer or each monomer onto the X-ray structure (see above). The RMSD fluctuates between 2 and 3 Å with eventual excursions up to 4 Å. This is somewhat larger than for the CDG dimer for which we previously found typical RMSDs of around 1 Å. The distance between the C5-atoms of the guanine bases, which can be used to distinguish between an open and a closed monomer, is around 7 Å for monomer I (see Figure 5b), with infrequent elongations up to 9 Å. This suggests that monomer I is still in its closed conformation. The distance between the C5 atoms of the guanine bases on monomer II (Figure 5c) oscillates considerably since monomer II fluctuates between closed and open conformations during the entire simulation. Compared to this, the chemically unmodified CDG dimer shows less structural variability.³

Representative snapshots of *endo-S*-CDG dimer structures along the 20 ns MD simulation are shown in Figure 5 of the supplementary information. In the simulations we find all possible combinations of monomeric structures, including closed-closed, open-closed and open-open, see Figure 5.

It is also of interest to consider the interaction energies of specific parts of *endo-S*-CDG within the dimer. Figure 6(A) reports the total interaction energy contributed by individual energy contributions: the inner guanine base of monomer II and monomer I; the outer, flexible guanine base of monomer II and monomer I; the phosphodiester linkage of monomer II and monomer I. The interaction energies of two monomers along the trajectory is in the range of 25 to 75 kcal/mol whereas the interaction energies of monomer I and the inner and outer guanine of monomer II are between -50 and -25 kcal/mol and between -20 and -10 kcal/mol. The interaction between monomer I and the phosphodiester linkage of monomer II is strongly destabilizing and fluctuates around 100 kcal/mol. Hence, we find overall enthalpic destabilization based on pure interaction energies and stabilization of the dimer is entirely solvent-driven.

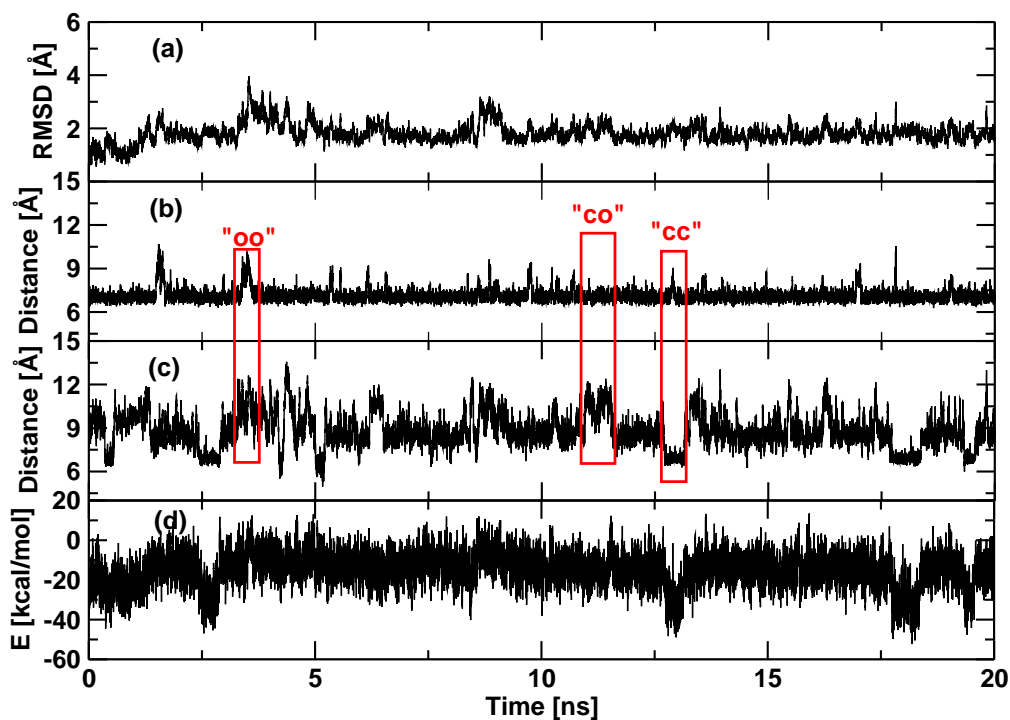


FIG. 5: (a) RMSD of endo-s-CDG dimer from 20 ns MD simulations. Monomers I and II were fitted to the X-ray structure of monomers I and II in 1W25, respectively. (b) The distance of C5-C5 of the guanine in monomer I. Along the 20 ns simulation, the distance is around 7 Å, which shows that this monomer is still in its closed state. (c) The distance of C5-C5 of the guanine in monomer II. The distance fluctuates since monomer II oscillates between open and closed structures. (d) Enthalpic part ΔH of the dimerization free energy for endo-S-CDG dimer along the 20 ns trajectory. When the distance is large in (c), the dimerization energy is small.

C. Dimerization Free Energy

A more quantitative measure for the stability of endo-S-CDG than interaction energies (either from electronic structure calculations or force field computations) is provided by the dimerization free energy ΔG_{bind} as calculated from the MD simulations. This is done according to MM-GBSA (see Methods section). Figure 5d already shows that the enthalpic part of the total energy (including gas phase contribution and solvation part to the enthalpic) is closely related

4.2 Dynamics of Analogue *endo-S-c-di-GMP* in Solution

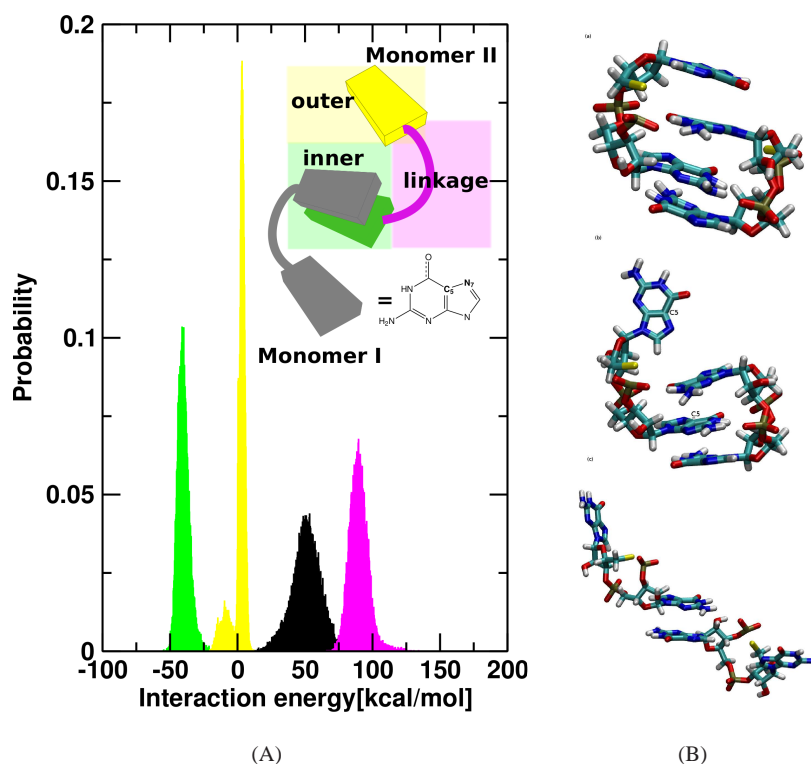


FIG. 6: (A) The interaction energy along a 20 ns trajectory for *endo-S-CDG*. Black: total interaction energy between *endo-S-CDG* monomer I and monomer II. Green: interaction energy between monomer I and inner guanine base of monomer II. Yellow: interaction energy between monomer I and outer guanine base (swing and turn the conformer to open shape) of monomer II. Pink: interaction energy between monomer I and the phosphodiester linkage motif of monomer II. The inner guanine base, outer guanine base, and the phosphodiester linkage motif is marked by green, yellow, pink respectively. (B) The snapshots of *endo-S-CDG* dimer structures from the 13 individual 20ns simulation. (PDB code:1W25): (a) The intercalated *endo-S-CDG* dimer; (b) One *endo-S-CDG* monomer stay as close shape, while another one turns to open shape; (c) Two *endo-S-CDG* monomers are in open shape with two guanine base-pair interaction.

to the conformations of the two *endo-S-CDG* monomers. Whenever the two monomers are in a closed form the two monomers lock in an enthalpically favoured state whereas opening of one monomer leads to destabilization by several 10 kcal/mol. The overall enthalpic stabilization excluding entropic contributions is -18.8 kcal/mol, about 4 kcal/mol smaller than that for

4 Bacterial Second Messenger *c*-di-GMP

TABLE III: Contributions from the Side-Chains and the Backbone Atoms in both endo-S-CDG Monomers to total Dimerization Free Energies (in kcal/mol) at $T = 300$ K. Free energies were calculated using MM-GBSA outlined in the Methods Section.

	Monomer I		Monomer II		D-(M1+M2)		
	linkage bases	M1	linkage bases	M2			
$\langle \Delta E_{\text{vdw}} \rangle$	-2.48	-10.29	-12.77	-2.15	-10.62	-12.77	-25.54(3.57)
$\langle \Delta E_{\text{ele}} \rangle$	47.38	-8.93	38.44	42.87	-4.43	38.44	76.89(10.17)
$\langle \Delta G_{\text{ele,desolv}} \rangle$	-47.55	14.07	-33.48	-46.88	13.47	-33.41	-66.89(8.25)
$\langle \Delta G_{\text{np,desolv}} \rangle$	-0.31	-1.36	-1.66	-0.17	-1.43	-1.59	-3.26(0.25)
$\langle \Delta G_{\text{bind}}^0 \rangle + \langle \Delta G_{\text{desolv}} \rangle$	-2.96	-6.51	-9.47	-6.32	-3.01	-9.33	-18.80(4.89)
$\langle -T\Delta S_{\text{tran}} \rangle$							17.58
$\langle -T\Delta S_{\text{rot}} \rangle$							10.69
$\langle -T\Delta S_{\text{vib}} \rangle$							-6.89
$\langle -T\Delta S_{\text{tot}} \rangle$							16.98
$\langle \Delta G_{\text{bind}} \rangle$							-1.82

c-di-GMP.³ Table III reports the different contributions to the calculated dimerization free energies for the endo-S-CDG dimer from a total of 0.25 μ s simulations whereas analyses of selected individual runs are presented in Table 1 of the SI. Including translational, rotational and vibrational contributions to ΔS leads to $\Delta G_{\text{bind}} \approx 0$ and suggests that the metal-free endo-S-CDG dimer is only marginally stable in solution. This compares with a binding free energy of ≈ -5 kcal/mol for the *c*-di-GMP dimer which has also been found to exist in dimeric form from NMR experiments.^{3,4} Hence, the present simulations find that chemical modification at the phosphodiester linkage destabilizes the endo-S-CDG and yields predominantly monomeric forms in solution.

The dimerization free energy for Mg^{2+} -bound endo-S-CDG was also determined molecular dynamics simulations in explicit solvent. As mentioned in the methods section, the position of the magnesium ion was constrained with an NOE constraint. Following an earlier study on *c*-di-GMP, the thermodynamic cycle considered for MM-GBSA considered the dissociation of the solvated $\text{Mg}^{2+}(\text{H}_2\text{O})_4$ from the solvated endo-S-CDG dimer.³ This leads to an enthalpic stabilization of the magnesium-bound endo-S-CDG dimer by $\Delta H = -3.24$ kcal/mol, compared to -9.39 kcal/mol for

4.2 Dynamics of Analogue *endo-S-c-di-GMP* in Solution

the CDG dimer. Details are provided in Table 2 in the SI. Including entropic contributions similar to the Mg^{2+} -free dimer above leads to an overall destabilization which was, however, not determined explicitly here.

D. *endo-S*-CDG and CDG bound to inactive and active PleD

Physiologically it is of considerable interest to characterize the dynamics of *endo-S*-CDG and *c-di-GMP* bound to proteins with which they interact. To provide more insight on this, MD simulations for both molecules in the inactive (1W25) and activated (2WB4) structure of PleD have been carried out. In both structures a mutually intercalated CDG dimer is bound in the allosteric inhibition site (I-site) while a CDG monomer is present in the active site (A-site). The structures used in these simulations are reported in Figure 3.

Figure 7 shows that both, *endo-S*-CDG and CDG dimers are stable in the I-site. The fluctuations are larger in the inactive form (panels (a) and (b)) compared to the active form. Both monomers show similar fluctuations which suggests that *endo-S*-CDG should also be stable in the inactive and activated PleD. This is different for the dimers in solution where one of the monomers can occasionally make a transition to an open form, specifically for *endo-S*-CDG. This is not observed for any of the dimers bound to the protein. Binding monomeric *endo-S*-CDG and CDG into 1W25 and 2WB4 leads to a slightly different picture. While CDG fluctuates appreciably in the active site of inactive PleD, *endo-S*-CDG appears to better stabilize in this site, as is shown in Figure 2 of the SI. The C5-C5 distance (see Figure 8), indicative of a “open” or a “closed” monomer suggests that CDG in inactive PleD is in an equilibrium between the two conformations (on the 10 ns time scale) whereas in the active form of the protein CDG is rather in an “open” conformation. Conversely, *endo-S*-CDG is predominantly in a “closed” form in inactive PleD whereas it also prefers an open conformation in activated PleD. It is worthwhile to note that all four systems have RMSD around or below 2 Å during the 20 ns dynamics and hence the proteins are overall stable (see Figure 3 in the SI). Hence, the activated form of the protein seems to favour open *endo-S*-CDG/CDG which is less likely to dimerize whereas inactive PleD shifts this to more closed forms reminiscent of a dimer structure.

4 Bacterial Second Messenger *c*-di-GMP

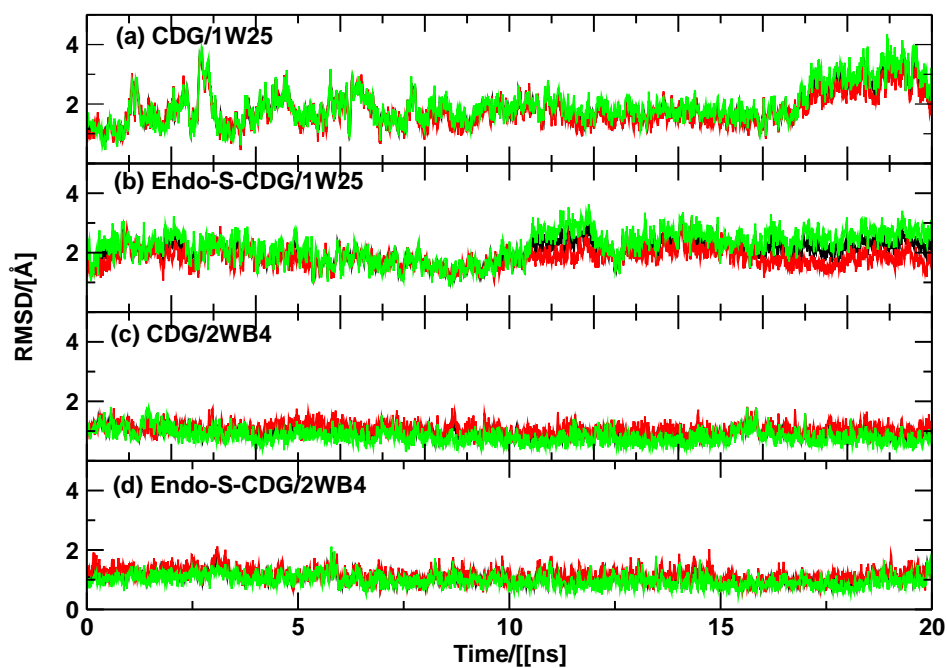


FIG. 7: RMSD of heavy atoms of CDG and endo-S-CDG which is bound to the inactive PleD (PDB code:1W25) and active PleD (PDB code:2WB4). (a) CDG in inactive PleD (PDB code:1W25); (b) endo-S-CDG in inactive PleD (PDB code:1W25); (c) CDG in active PleD (PDB code:2WB4); (d) endo-S-CDG in active PleD (PDB code:2WB4). Black: CDG/endo-S-CDG dimer; Red: CDG/endo-S-CDG monomer1; Green: CDG/endo-S-CDG monomer2.

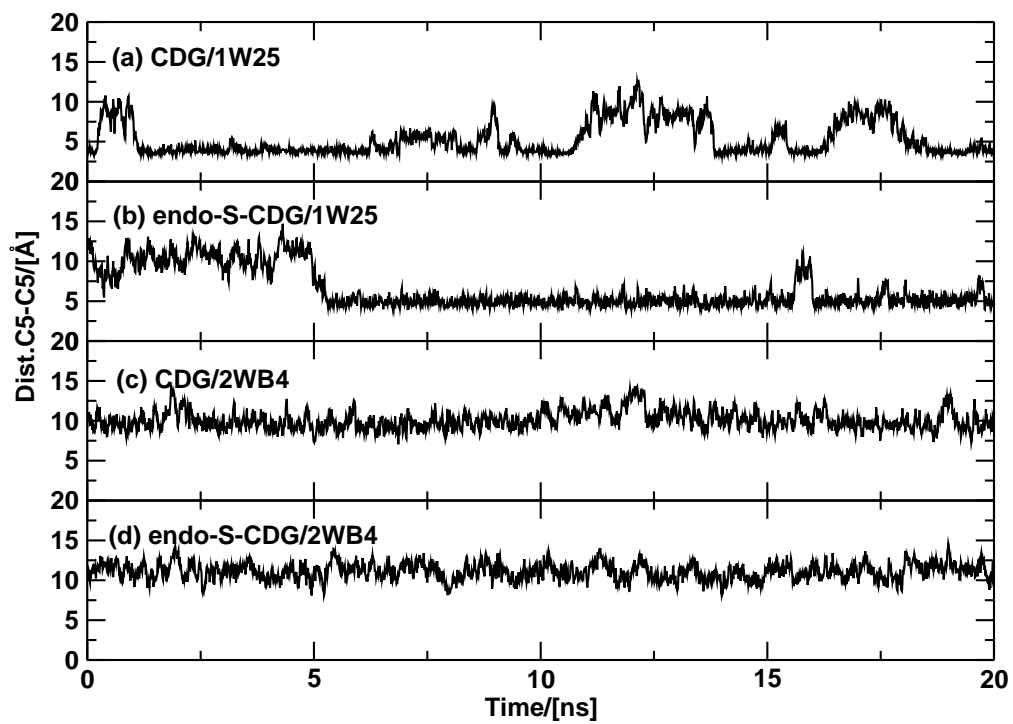


FIG. 8: Distance of C5-C5 of CDG/endo-S-CDG monomer in protein PleD

IV. CONCLUSIONS

The present work establishes that chemically modified CDG - endo-S-CDG - is less probable to form dimers in solution than CDG for which marginal stabilization has been found from computations and experiment.^{3,4} Including entropic contributions in estimating the dimerization free energy of endo-S-CDG, 250 ns of atomistic simulations give a slight stabilization of ≈ -2 kcal/mol. The metal-bound dimer is slightly more stable but comparison with Mg^{2+} -bound $(\text{CDG})_2$ suggests that the endo-S-CDG dimer is also only marginally stable in the presence of magnesium. Both messengers are found to favourably interact with the active and inactive form of PleD, a CDG-binding protein. In the inhibition site, both dimers are stable and in a predominantly “closed” form whereas in the active site the behaviour of endo-S-CDG and CDG differ to some extent. It is interesting to note that the activated protein appears to favour “open” forms of the monomeric messenger whereas the inactive protein supports both, “open” and “closed” forms. These findings may provide essential insights into better characterizing the dynamics and function of natural and chemically modified CDG.

Acknowledgments

Financial support from the Swiss National Science Foundation through funding of projects Nr. 200021-117810 is gratefully acknowledged.

¹ Schirmer, T.; Jenal, U. *Nat. Rev. Microbiol.* **2009**, *7*, 724–735.

² Hengge, R. *Nat. Rev. Microbiol.* **2009**, *7*, 263–273.

³ Zhang, L.; Meuwly, M. *ChemPhysChem* **2011**, *12*, 295–302.

⁴ Gentner, M.; Allan, M.; Zaehring, F.; Schirmer, T.; Grzesiek, S. *J. Am. Chem. Soc.* **2012**, *134*, 1019–1029.

⁵ Wang, J.; Zhou, J.; Donaldson, G.; Nakayama, S.; Yan, L.; Lam, Y.; Lee, V.; Sintim, H. *J. Am. Chem. Soc.* **2011**, *133*, 9320–9330.

⁶ Egli, M.; Gessner, R. V.; Williams, L. D.; Quigley, G. J.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Frederick, C. A. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 3235–3239.

4.2 Dynamics of Analogue endo-S-c-di-GMP in Solution

- ⁷ Liaw, Y.-C.; Gao, Y.-G.; Robinson, H.; Sheldrick, G. M.; Sliedregt, L. A. J. M.; van der Marel, G. A.; van Boom, J. H.; Wang, A. H.-J. *FEBS Lett.* **1990**, *264*, 223–227.
- ⁸ Smith, K.; Lipchock, S.; Livingston, A.; Shanahan, C.; Strobel, S. *Biochem* **2010**, *49*, 7351–7359.
- ⁹ Yan, H.; Chen, W. *Chem.Soc.Rev.* **2010**, *39*, 2914–2924.
- ¹⁰ Sondermann, H.; Shikuma, N.; Yildiz, F. *Curr.Opin. Microbiol.* **2012**, *15*, 140–146.
- ¹¹ Römling, U.; Gomelsky, M.; Galperin, M. Y. *Mol. Microbiol.* **2005**, *57*, 629–639.
- ¹² Römling, U.; Amikam, D. *Curr. Op. Microbiol.* **2006**, *9*, 218–228.
- ¹³ Zhang, Z.; Gaffney, B. L.; Jones, R. A. *J. Am. Chem. Soc.* **2004**, *126*, 16700–16701.
- ¹⁴ Zhang, Z.; Kim, S.; Gaffney, B. L.; Jones, R. A. *J. Am. Chem. Soc.* **2006**, *128*, 7015–7024.
- ¹⁵ Ko, J.; Ryu, K.-S.; Kim, H.; J.-S-Shin.; Lee, J.-O.; Cheong, C.; Choi, B.-S. *J. Mol. Biol.* **2010**, *398*, 97–110.
- ¹⁶ Wassmann, P.; Chan, C.; Paul, R.; Beck, A.; Heerklotz, H.; Jenal, U.; Schirmer, T. *Structure* **2007**, *15*, 915–927.
- ¹⁷ Amikam, D.; Galperin, M. Y. *Bioinformatics* **2006**, *22*, 3–6.
- ¹⁸ De, N.; Navarro, M.; Raghavan, R.; Sondermann, H. *J. Mol. Biol.* **2009**, *393*, 619–633.
- ¹⁹ De, N.; Pirruccello, M.; Krasteva, P. V.; Bae, N.; Raghavan, R. V.; Sondermann, H. *PLoS Biol.* **2008**, *6*, 0601–0617.
- ²⁰ Chan, C.; Paul, R.; Samoray, D.; Amiot, N. C.; Giese, B.; Jenal, U.; Schirmer, T. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 17084–17089.
- ²¹ Benach, J.; Swaminathan, S.; Tamayo, R.; Handelman, S.; Folta-Stogniew, E.; Ramos, J.; Forouhar, F.; Neely, H.; Seetharaman, J.; Camilli, A.; Hunt, J. *EMBO J.* **2007**, *26*, 5153–5166.
- ²² Minasov, G.; Padavattan, S.; Shuvalova, L.; Brunzelle, J. S.; Miller, D. J.; Basle, A.; Massa, C.; Collart, F. R.; Schirmer, T.; Anderson, W. F. *J. Biol. Chem.* **2009**, *284*, 13174–13184.
- ²³ Navarro, M. V. A. S.; De, N.; Bae, N.; Wang, Q.; Sondermann, H. *Structure* **2009**, *17*, 1104–1116.
- ²⁴ Navarro, M.; Newell, P.; Raghavan, R.; Chatterjee, D.; Madden, D.; Toole, G. O.; Sondermann, H. *PLoS Biol.* **2011**, *9*, e1000588–e1000588.
- ²⁵ Minasov, G.; Padavattan, S.; Shuvalova, L.; Brunzelle, J. S.; Miller, D. J.; Baslé, A.; Massa, C.; Collart, F. R.; Schirmer, T.; Anderson, W. F. *J. Biol. Chem.* **2009**, *284*, 13174–84.
- ²⁶ Navarro, M. V. a. S.; De, N.; Bae, N.; Wang, Q.; Sondermann, H. *Structure (London, England : 1993)* **2009**, *17*, 1104–16.
- ²⁷ Barends, T. R. M.; Hartmann, E.; Griese, J. J.; Beitlich, T.; Kirienko, N. V.; Ryjenkov, D. a.; Reinstein, J.;

4 Bacterial Second Messenger c-di-GMP

- Shoeman, R. L.; Gomelsky, M.; Schlichting, I. *Nature* **2009**, *459*, 1015–8.
- ²⁸ Navarro, M. V. a. S.; Newell, P. D.; Krasteva, P. V.; Chatterjee, D.; Madden, D. R.; O'Toole, G. a.; Sondermann, H. *PLoS Biol.* **2011**, *9*, e1000588.
- ²⁹ Tchigvintsev, A.; Xu, X.; Singer, A.; Chang, C.; Brown, G.; Proudfoot, M.; Cui, H.; Flick, R.; Anderson, W. F.; Joachimiak, A.; Galperin, M. Y.; Savchenko, A.; Yakunin, A. F. *J. Mol. Biol.* **2010**, *402*, 524–38.
- ³⁰ De, N.; Navarro, M. V. A. S.; Raghavan, R. V.; Sondermann, H. *J. Mol. Biol.* **2009**, *393*, 619–33.
- ³¹ De, N.; Pirruccello, M.; Krasteva, P. V.; Bae, N.; Raghavan, R. V.; Sondermann, H. *PLoS Biol.* **2008**, *6*, e67.
- ³² Wassmann, P.; Chan, C.; Paul, R.; Beck, A.; Heerklotz, H.; Jenal, U.; Schirmer, T. *Structure (London, England : 1993)* **2007**, *15*, 915–27.
- ³³ Chan, C.; Paul, R.; Samoray, D.; Amiot, N. C.; Giese, B.; Jenal, U.; Schirmer, T. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 17084–9.
- ³⁴ Yang, C.-Y.; Chin, K.-H.; Chuah, M. L.-C.; Liang, Z.-X.; Wang, A. H.-J.; Chou, S.-H. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **2011**, *67*, 997–1008.
- ³⁵ De, N.; Navarro, M. V. A. S.; Raghavan, R. V.; Sondermann, H. *J. Mol. Biol.* **2009**, *393*, 619–33.
- ³⁶ Whitney, J. C.; Colvin, K. M.; Marmont, L. S.; Robinson, H.; Parsek, M. R.; Howell, P. L. *J. Biol. Chem.* **2012**, *287*, 23582–93.
- ³⁷ Habazettl, J.; Allan, M. G.; Jenal, U.; Grzesiek, S. *J. Biol. Chem.* **2011**, *286*, 14304–14.
- ³⁸ Ko, J.; Ryu, K.-S.; Kim, H.; Shin, J.-S.; Lee, J.-O.; Cheong, C.; Choi, B.-S. *J. Mol. Biol.* **2010**, *398*, 97–110.
- ³⁹ Benach, J.; Swaminathan, S. S.; Tamayo, R.; Handelman, S. K.; Folta-Stogniew, E.; Ramos, J. E.; Forouhar, F.; Neely, H.; Seetharaman, J.; Camilli, A.; Hunt, J. F. *The EMBO journal* **2007**, *26*, 5153–66.
- ⁴⁰ Ko, J.; Ryu, K.-S.; Kim, H.; Shin, J.-S.; Lee, J.-O.; Cheong, C.; Choi, B.-S. *J. Mol. Biol.* **2010**, *398*, 97–110.
- ⁴¹ Frisch, M. J. et al. Gaussian 03, Revision B.01; Gaussian, Inc.: Wallingford, CT, 2004.
- ⁴² Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- ⁴³ Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- ⁴⁴ Berman, H. M. et al. *Acta Cryst.* **2002**, *D58*, 899–907.
- ⁴⁵ Fletcher, D. A.; McMeeking, R. F.; Parkin, D. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 746–749.
- ⁴⁶ Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput.*

4.2 Dynamics of Analogue endo-S-c-di-GMP in Solution

- Chem.* **1983**, *4*, 187–217.
- ⁴⁷ MacKerell, Jr., A. D. et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- ⁴⁸ Christen, B.; Christen, M.; Paul, R.; Schmid, F.; Folcher, M.; Jenoe, P.; Meuwly, M.; Jenal, U. *J. Biol. Chem.* **2006**, *281*, 32015–32024.
- ⁴⁹ Schmid, F. F.-F.; Meuwly, M. *J. Mol. Biol.* **2007**, *374*, 1270–1285.
- ⁵⁰ Devereux, M.; Meuwly, M. *J. Chem. Inf. Model.* **2010**, *50*, 349–357.
- ⁵¹ Law, M.; Hutson, J. *Comput. Phys. Commun.* **1997**, *102*, 252–268.
- ⁵² Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Chem. Phys.* **1977**, *23*, 327–341.
- ⁵³ Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- ⁵⁴ Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- ⁵⁵ Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10093.
- ⁵⁶ Srinivasan, J.; Cheatham III, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409.
- ⁵⁷ Kollman, P. A.; Massoval, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham III, T. E. *Acc. Chem. Res.* **2000**, *33*, 889–897.
- ⁵⁸ Gohlke, H.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 238–250.
- ⁵⁹ Zoete, V.; Meuwly, M.; Karplus, M. *Proteins: Struct. Funct. Bioinf.* **2005**, *61*, 79–93.
- ⁶⁰ Lee, M. S.; Salsbury Jr., F. R.; Brooks III, C. L. *J. Chem. Phys.* **2002**, *116*, 10606–10614.
- ⁶¹ Lee, M. S.; Feig, M.; Salsbury Jr., F. R.; Brooks III, C. L. *J. Comput. Chem.* **2003**, *24*, 1348–1356.
- ⁶² Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- ⁶³ Gohlke, H.; Kiel, C.; Case, D. A. *J. Mol. Biol.* **2003**, *330*, 891–913.
- ⁶⁴ Tidor, B.; Karplus, M. *J. Mol. Biol.* **1994**, *238*, 405–414.

Dynamics of Analogue endo-S-c-di-GMP in Solution

Lixian Zhang, and Markus Meuwly*

Department of Chemistry, University of Basel, Basel, Switzerland

E-mail: m.meuwly@unibas.ch

*To whom correspondence should be addressed

4.2 Dynamics of Analogue endo-S-c-di-GMP in Solution

Table 1: Dimerization Free Energies (in kcal/mol) of endo-S-cdg dimer from three independent 20 ns MD simulations in explicit solvent at $T = 300$ K. Free energies were calculated using MM-GBSA outlined in the Methods Section

	$\langle \Delta E_{\text{vdw}} \rangle$	$\langle \Delta E_{\text{ele}} \rangle$	$\langle \Delta G_{\text{ele,desolv}} \rangle$	$\langle \Delta G_{\text{np,desolv}} \rangle$	$\langle \Delta E_{\text{bind}}^0 \rangle + \langle \Delta G_{\text{desolv}} \rangle$	$\langle -T\Delta S_{\text{tran}} \rangle$	$\langle -T\Delta S_{\text{rot}} \rangle$	$\langle -T\Delta S_{\text{vib}} \rangle$	$-T\langle \Delta S \rangle$	$\langle \Delta G_{\text{bind}} \rangle$
Run1	-27.36	77.94	-64.43	-3.37	-17.23	13.18	10.71	-7.09	16.80	-0.43
Run2	-28.66	77.20	-65.21	-3.46	-20.13	13.18	10.57	-7.61	16.29	-3.84
Run3	-25.23	70.37	-62.61	-3.31	-20.78	13.18	10.57	-6.89	16.86	-3.92
Run4	-26.39	72.91	-62.57	-3.38	-19.44	13.18	10.61	-6.90	16.89	-2.55
Run5	-25.24	76.39	-62.93	-3.26	-15.05	13.17	10.69	-6.90	16.96	1.91

4 Bacterial Second Messenger *c*-di-GMP

Table 2: The different contributions to the dimerization free energy of endo-S-cdg with Mg^{2+} from simulations at $T = 300$ K with NOE restraint [K_{\min}, K_{\max}] = [100, 200] kcal/(mol \AA^2). The total simulation time is 10 ns. (a) Mg^{2+} - bound dimer dissociates into the hydrated Mg^{2+} and metal free dimer. (b) The metal-free dimer dissociated into two separate monomers. $\langle \Delta G_{\text{bind}(a/b)} \rangle$ is the binding free energy for step (a) or (b), respectively, whereas $\langle \Delta G_{\text{bind}(a+b)} \rangle$ is the free energy difference for the overall process (a and b).

	a)(endo-S-cdg) ₂ Mg ²⁺ (H ₂ O) ₄ → Mg ²⁺ (H ₂ O) ₄ + (endo-S-cdg) ₂	b)(endo-S-cdg) ₂ → 2endo-S-cdg
$\langle \Delta E_{\text{vdw}} \rangle$	15.23 (4.67)	-23.52 (4.11)
$\langle \Delta E_{\text{ele}} \rangle$	-579.09(14.82)	117.78(7.54)
$\langle \Delta G_{\text{ele,desolv}} \rangle$	555.91(13.20)	-84.42(6.44)
$\Delta G_{\text{np,desolv}}$	-1.80(0.09)	-3.33(0.26)
$\langle \Delta G_{\text{bind}}^0 \rangle + \langle \Delta G_{\text{desolv}} \rangle$	-9.75(7.16)	6.51(5.66)
$\langle \Delta H_{(a+b)} \rangle$		-3.24

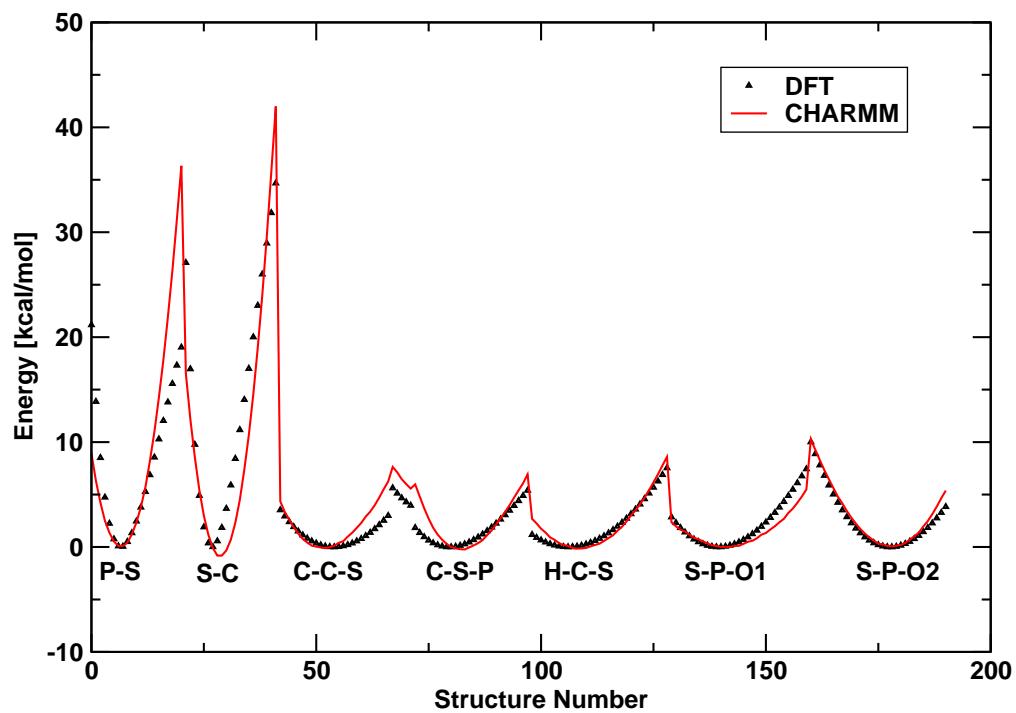


Figure 1: Fit of the distance and angle parameters in the force field. The reference energies (from DFT calculations) and the fitted values are shown as black symbols and red lines, respectively.

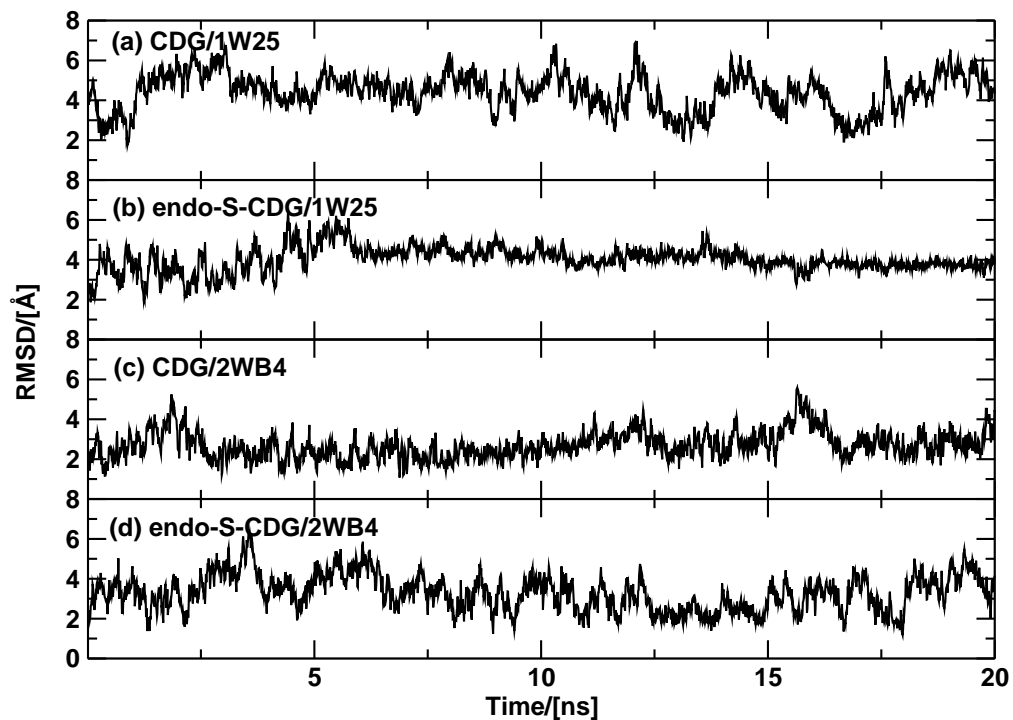


Figure 2: RMSD of heavy atoms of CDG and endo-S-CDG monomer which bound to A-site of inactive PleD (PDB code: 1W25) and active PleD (PDB code: 2WB4). (a) CDG in inactive PleD (PDB code: 1W25); (b) endo-S-CDG in inactive PleD (PDB code: 1W25); (c) CDG in active PleD (PDB code: 2WB4); (d) endo-S-CDG in active PleD (PDB code: 2WB4).

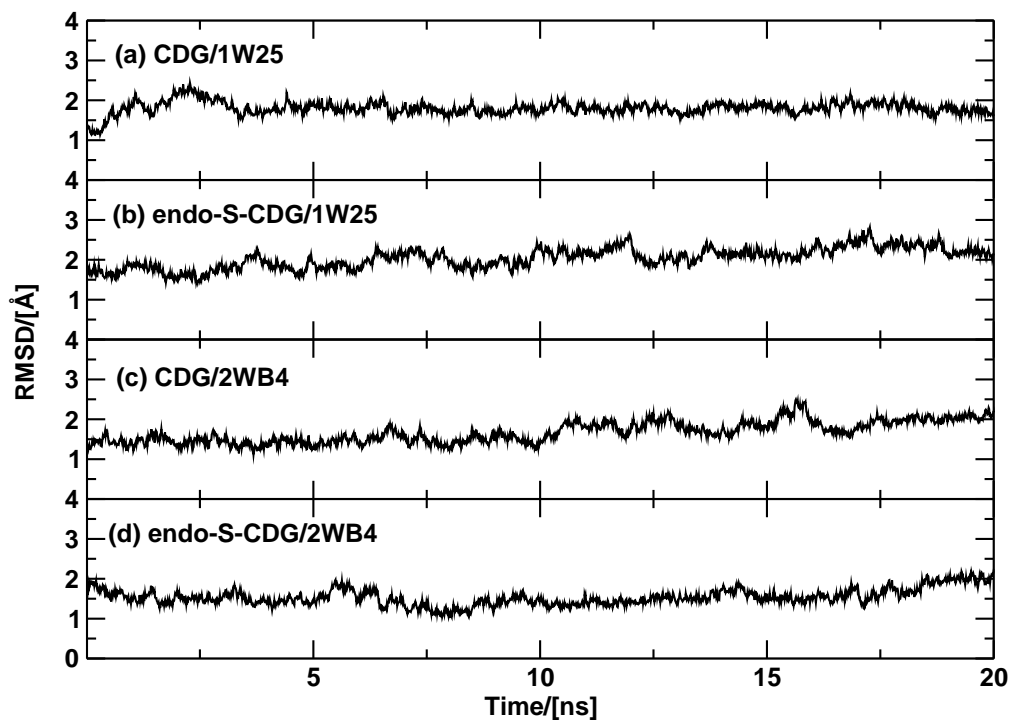


Figure 3: RMSD of backbone of inactive PleD (PDB code: 1W25) protein and active PleD (PDB code: 2WB4) protein with CDG or endo-S-CDG binding. (a) system: CDG/inactive PleD (PDB code: 1W25); (b) endo-S-CDG/inactive PleD (PDB code: 1W25); (c) CDG/active PleD (PDB code: 2WB4); (d) endo-S-CDG/active PleD (PDB code: 2WB4).

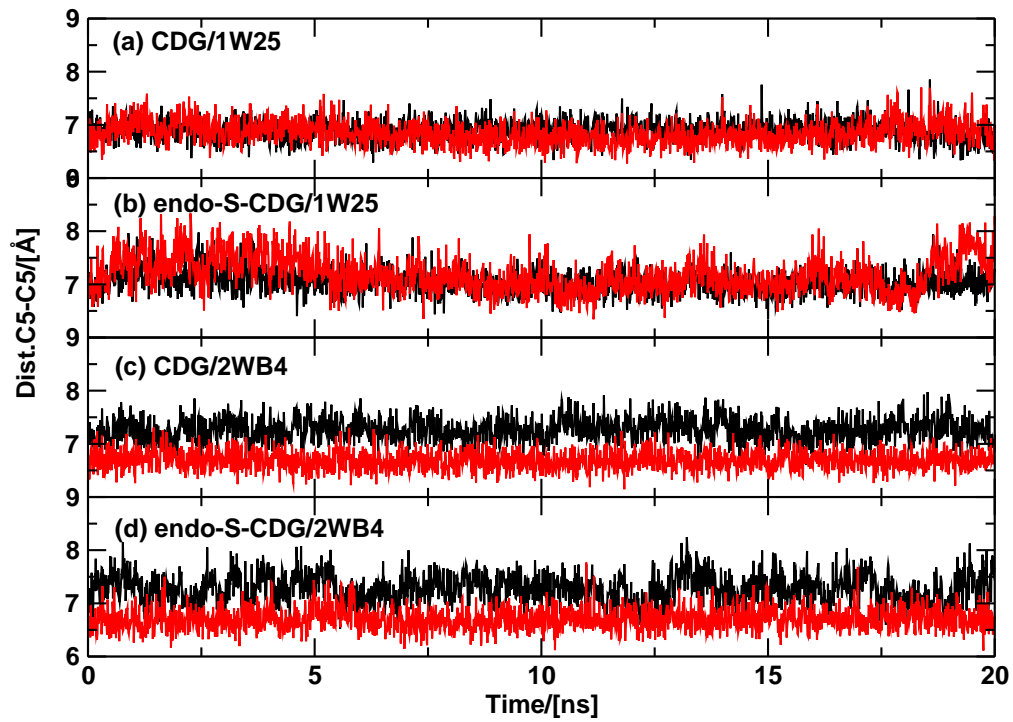


Figure 4: Distance of C5-C5 of CDG/endo-S-CDG dimer in protein PleD. Black and red lines represent the C5-C5 distance from the two monomers of the dimer.

4.3 PleD Dimerization

Atomic Simulations of PleD with/without c-di-GMP Bound As we mentioned above, a central question for understanding the function of PleD is its primary function (production of c-di-GMP upon PleD dimerization) with self-inhibition by c-di-GMP. Both I- and A-sites are located in the DGC domain. So far, still little is known about how DGCs work at a molecular level. For this, one of the works in this thesis is to study inhibited monomeric PleD and dimeric PleD with and without intercalated c-di-GMP present in the inhibition site to investigate a possible biological mechanism for allosteric regulation. All MD simulations were carried out with NAMD using the CHARMM 22 force field. The starting structures were taken from the X-ray structure of the PleD dimer (pdb code: 1W25, 2V0N/2WB4). The MD simulations were carried out for 2 ns each system (20 ns in the case of c-di-GMP/endo-S-CDG binding PleD monomer). It has been proposed that binding of c-di-GMP to the PleD monomer and dimer will eliminate catalytic activity by domain immobilization.⁴ Moreover, inhibition (I-site) and the active site (A-site) of PleD are dynamically coupled through an network of atomic motions that involves the $\beta 2$ strand (direct coupling with DGC, pathway I) and the structural elements $\beta 3$, $\alpha 2$ and $\alpha 3$ of DGC (network coupling, pathway II).^{196,199} To obtain information on the proposed pathway and to discriminate these two pathways, longer simulation and further analysis are required.

Atomic Simulations of c-di-GMP/endo-S-CDG Bound to DGC Domain To verify that DGC domain is sufficient for feedback inhibition and to exclude path II, DGC domain of PleD individually (excluding D1/D2) with c-di-GMP and endo-S-CDG bound were investigated. All MD simulations were carried out with NAMD using the CHARMM 22 force field for 20 ns each trajectory.

Atomic Simulations of Phosphorylation of D1/D2 Domain The DGC activity of PleD is activated upon phosphorylation of D1/D2 domain. Phosphorylation of D1/D2 domains invokes repacking of the D1/D2 loop region, and promotes PleD dimerization.^{4,184,193} A lot of study has been performed for the mechanism of CheY-like domain (PleD) activity upon phosphorylation:^{35,200} a major repositioning of the $\beta 4$ - $\alpha 4$ loop and the reorientation of the sidechain of a conserved Tyr/Phe residue from solvent exposed to a buried position; repositioning of the Thr/Ser due to hydrogen-bond formation leaves more space for the Tyr/Phe to adopt the buried rotameric state. However, still, detail structural changes associated with the $\beta 4$ - $\alpha 4$ loop relaxation from the active to its inactive form, interface changes and interaction between monomers when dimerization happens are still unknown. Four systems were investigated: D1/D2 with and without beryll fluoride binding in inactive PleD (pdb code: 1W25 and 2V0N). All MD simulations were carried out with NAMD using the CHARMM 22 force field for 2 ns each trajectory.

5 Computational Characterization of the Insulin Dimerization

Chapter 4 presented the work of the ubiquitous second messenger c-di-GMP which regulates cell surface-associated traits and biofilm formation. Biofilm formation is involved in life-threatening infectious diseases such as cystic fibrosis and the colonization of medical devices.

Another case of protein dimerization which leads to a common disease (diabetes) is insulin. Insulin is a hormone, which is produced in the pancreas as a hexamer and is a main regulator of the glucose levels in the blood. When we eat, glucose levels rise, and insulin is released into the bloodstream as a monomer. The insulin acts like a key, opening up cells so they can take in the sugar and use it as an energy source. Usually insulin dissociates from its hexameric storage form through an intermediate dimer state to the bioactive monomer before binding to its transmembrane insulin receptor. The interface which the insulin monomer uses to bind to the receptor, is the same one that forms the dimer and hexamer. Thus, once the monomers form dimers or hexamers, they can not bind to the receptors and subsequently, lose their biological function which lead to diabetes.

Many detailed questions concerning the stability of aggregated insulin, the conformation of inactive and active native and mutated insulin monomers, designing insulin mutants which keep the active conformation but have lower aggregation ability still need to be resolved. Although extensive research on insulin has been conducted, more efforts

5 Computational Characterization of the Insulin Dimerization

are still needed and provide new progress in insulin treatment. The following manuscript is in preparation.

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

Lixian Zhang, Manuela Koch, and Markus Meuwly*

Department of Chemistry, University of Basel, Basel, Switzerland

E-mail: m.meuwly@unibas.ch

Abstract

*To whom correspondence should be addressed

Introduction

Insulin is a small protein that plays an eminent role in controlling glucose uptake in cells.¹ The crystallized, native hormone is a hexamer and consists of two trimers with either two or four zinc atoms bound to it. Each dimer contains two monomers (chain A with 21 amino acids and chain B with 30 amino acids) which are connected by two interchain (CysA7–CysB7 and the other from CysA20–CysB19) and one intrachain (CysA6–CysA11) disulphide bond. Under physiological conditions, insulin monomers readily aggregate to dimers. The interactions stabilizing the native dimer are predominantly nonpolar^{2–4} with the β -sheet hydrogen bonds mainly replacing water-hydrogen bonds and contributing to orienting the two monomers. One possibility to prevent aggregation is to modify the complexation interface.³ It is known that the sequence and structure of insulin is intimately linked to its biological function. However, since the structure of the complex formed by insulin and its receptor is unknown, the study of the binding process at an atomistic level, even by theoretical means, can not be addressed. When insulin is mutated, the resulting change in its activity (increase or decrease) may be due to modification of the insulin/receptor interactions, to a change of the insulin fold required for binding to the receptor⁵ or to altered pharmacokinetics.

Insulin shows an important propensity to aggregation. Although it circulates in the serum and binds to its receptor in its monomeric form, insulin self-associates to dimers at micromolar concentrations, and in the presence of zinc ions it further assembles to hexamers.⁶ This strongly affects the physiological function and pharmacokinetics of the hormone. When associating to dimers, the extended C-terminal ends of the two B-chains are brought together, forming a two-stranded antiparallel β -sheet. In the hexameric form, the form in which insulin is stored in the β cells, the two central zinc ions are coordinated by the B10 histidine residues. Structural studies of native insulin in the monomeric state are made difficult by self-association. All published high resolution X-ray structures of insulin are aggregated species except for a partial report of results for monomeric insulin⁷ for which no coordinates have been released so far. Thus, the

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

current view of insulin structure-function relationship is derived primarily from insulin hexamer and dimer crystal structures and from studies of the structures and activity of chemically modified and/or naturally occurring mutant insulins in solution.^{5,8,9}

Many studies suggest that the conformation of insulin found in the X-ray structures (dimer/hexamer) is inactive and that the separation of the N-terminus of chain A and the C-terminus of chain B is required for interaction with the insulin receptor.⁷⁻¹⁵ This separation exposes a hydrophobic surface including the nonpolar residues Leu(A16), Tyr(A19), Leu(B11) and Leu(B15) and, in particular, Gly(A1), Ile(A2), Val(A3), which is crucial for insulin binding to its receptor. The C-terminus of chain B is situated near residues Gly(A1), Ile(A2) and Val(A3) and shields them from the solvent.

Several NMR studies of active insulin mutants show a rearrangement of the C-terminus end of chain B.^{8,14} Also, activity studies of semi-synthetic analogs have found that the C-terminal pentapeptide of the chain B (B26-B30) can be deleted without a decrease in biological potency.¹⁵ A recently published preliminary crystallographic investigation of a low pH native porcine insulin monomer⁷ shows that most of the insulin monomer is well ordered and similar in conformation to other insulin structures. However, residues B21-B25 appear to have multiple conformations, whereas the C-terminus of chain B is highly flexible when not involved in dimer formation (electron density for B25-B30 is essentially absent). A Raman spectroscopy and microscopy study of insulin in different aggregation states (monomer, dimer, hexamer and fibril) shows that dimerization damps fluctuations at an intermolecular β -sheet.¹⁶ Experimental alanine scanning finds that substitution of alanine at various positions reduces insulin affinity for the receptor by more than 20-fold.¹⁷ The residues that are most likely to be directly involved in binding are A1, A2, A3, A19, B12, B23, and B24. Any substitution of residues A1-A3 has been shown to impair function.¹⁵

5 Computational Characterization of the Insulin Dimerization

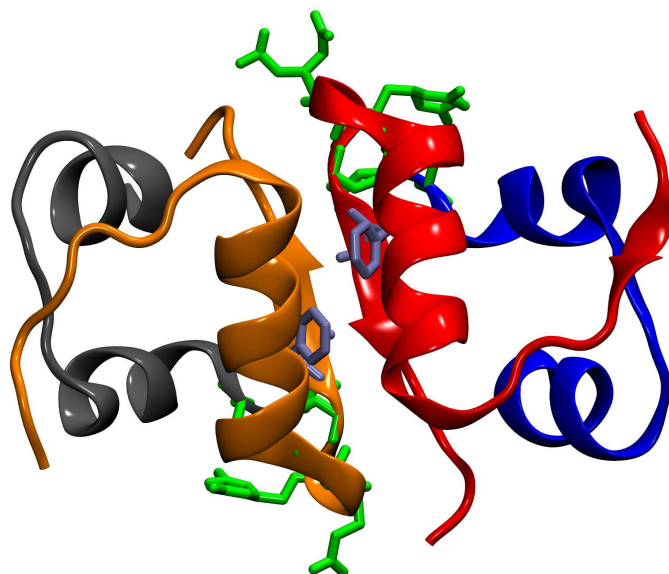


Figure 1: Structure of insulin dimer (PDB Code: 4INS). Chain A (blue), Chain B (red), Chain C (grey), Chain D (orange). β -turn (residue 20-23) and residue 24 are shown as stick representation in green and iceblue respectively.

Phenylalanine at position B24 is invariant among insulin sequences and is located at the dimerization interface and maintains the orientation of the B-chain of the monomer.^{8,18} These observations together with studies of low-potency B24 analogues suggest that PheB24 plays an important role in the activity of insulin. SerB24 and LeuB24 analogues show reduced binding potency.¹⁹ However, it was also found that certain B24 substitutions, such as glycine²⁰ and D-Ala,¹¹ are well tolerated in view of insulin affinity to its receptor. Their bioactivity has also been referred to as “anomalous” as it can not be readily explained by crystal models. The role of PheB24 in stabilizing the insulin dimer has also been studied to some extent. It was found that unlike native insulin, the GlyB24 mutant does not dimerize in aqueous solution at pH 1.9.⁸ Furthermore, alanine scanning of the dimerization interface revealed that the AlaB24 mutant is a monomeric insulin and does not readily aggregate. This suggests that the AlaB24 and GlyB24

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

mutant dimers are less stable than the native dimer.

In the present work the dynamics and stability of the native and four B24 mutant insulin dimers (AlaB24, GlyB24, D-AlaB24, Des-B25) were investigated by using computer simulations. Previous studies have shown that computer simulations provide meaningful and complementary information to experimental characterizations.^{4,21,22} For the native insulin dimer experimental data on the dimerization energy (-7.2 kcal/mol in favour of the dimer²³) is available which provides a validation for the simulations. The work is structured as follows. First, the computational methods are presented. Then results on the dimerization energies and internal dynamics of insulin monomers and dimers are reported. Finally, the results are discussed in light of experimental data.

Computational Methods

Molecular dynamics simulations

Molecular dynamics (MD) simulations were carried out using CHARMM²⁴ (version 35b1) and the "all-atom" CHARMM22²⁵ force field with periodic boundary conditions (PBC).²⁶

Native and Mutant Insulin Monomer:

The starting coordinates for the MD simulations were the X-ray structure of the native porcine insulin dimer resolved at 1.5 Å (Protein Data Bank (PDB^{27,28}), Code: 4INS.²⁹ The structure contains the coordinates of the insulin dimer and two aggregated zinc atoms. As the zinc atoms play an important role only in the hexamer formation they were moved for this investigation. Hydrogen atoms were added to the X-ray structure. The crystal structure was also used to generate mutants computationally. The Phenylalanine at position B24 was mutated into Glycine (Gly), Alanine (Ala), and D-Alanine (D-Ala) which yields mutants B24Gly, B24Ala,

5 Computational Characterization of the Insulin Dimerization

and B24D-Ala, respectively. Besides, wildtype insulin dimer without the Phe(B25) amino acid on both monomers, Des-B25 mutant is also studied here.

Native and Mutant Insulin Dimer:

For the simulation of the insulin dimers the same X-ray structure (Protein Data Bank (PDB^{27,28}), entry 4INS²⁹) was used. The wildtype dimer and mutants were solvated in a $77.6 \times 62.8 \times 55.8$ Å box of TIP3P water molecules. Water molecules overlapping the protein were removed which leads to a system with approximately 1550 protein atoms and 8495 water molecules. The solvent was equilibrated at 300 K during 30 ps with the insulin kept fix. Then 2000 steps of steepest descent (SD) minimization were carried out. The entire system was heated to 300 K during 15 ps using harmonic constraints with a force constant of $5 \text{ kcal/mol}\text{\AA}^2$ on the position of the backbone atoms. The system was further equilibrated for 120 ps with gradually decreasing harmonic constraints (from $5 \text{ kcal/mol}\text{\AA}^2$ to a fully unconstrained system) on the backbone atoms. For all simulations the Verlet leapfrog integrator was used for time propagation with a time step of 1 fs. A 12 Å cut off was applied to the shifted electrostatic and switched Van der Waals interactions and images for periodic boundary conditions were updated every 10 time steps. All distances to hydrogen atoms were constrained by using SHAKE.³⁰ For the native system simulations of 15 to 20 ns in length were carried out to verify earlier results.²¹ For wildtype dimer and mutants (B24Gly, B24Ala, B24D-Ala, Des-B25) nine individual trajectories were run starting from different structures taken from the equilibration run. In the following the trajectories are labeled by letters to indicate the amino acid at position B24 (F for Phenylalanine, A for Alanine and G for Glycine) and numerals indicate the starting structure. The trajectories are labeled by double letters for the mutation and a numeral for the run.

Calculation of the Binding Free Energy

The binding free energy ΔG_{bind} for a dimer consisting of two monomers can be calculated by considering a thermodynamic cycle and using the molecular mechanics-generalized Born solvation model (MM-GBSA).³¹ For details the reader is referred to Ref.⁴ Here only a short description of the main steps involved in calculating ΔG_{bind} is given. ΔG_{bind} can be separated in two terms, the entropic and the enthalpic contribution. In this approach the binding free energy ΔG_{bind} is calculated as the sum of the average (indicated with brackets) gas phase energy ΔG_{bind}^0 , the desolvation energy of the system ΔG_{desolv} and an entropic part $-T\Delta S$:

$$\Delta G_{\text{bind}} = \langle \Delta G_{\text{bind}}^0 \rangle + \langle \Delta G_{\text{desolv}} \rangle - \langle T\Delta S \rangle \quad (1)$$

Enthalpic Contribution: The enthalpic contribution $\Delta G_{\text{enthalpic}}$ to the binding free energy and was calculated using equation 2.

$$\Delta G_{\text{enthalpic}} = \langle \Delta G_{\text{bind}}^0 \rangle + \langle \Delta G_{\text{desolv}} \rangle \quad (2)$$

where ΔG_{bind}^0 is the sum of the intramolecular contribution ΔE_{intra} , the Van der Waals term ΔE_{vdW} and the electrostatic contribution ΔE_{elec} . The Van der Waals and electrostatic interaction energy between the two monomers were calculated together with the difference in internal energy between the dimer and the two isolated monomers. Because the calculations were carried out with the "same trajectory method" (STM) where the monomeric and dimeric contributions were taken from the same trajectory, the difference of the internal energy cancels.

$$\Delta G_{\text{bind}} = \Delta E_{\text{vdW}} + \Delta E_{\text{elec}} \quad (3)$$

This method is expected to give more stable results than analysis of different trajectories for the monomers and the dimer.⁴ The contribution ΔG_{desolv} is the difference between the solvation energy of the dimer and the solvation energies of the isolated monomers: $\Delta G_{\text{desolv}} = \Delta G_{\text{solv}}^{\text{dimer}} -$

5 Computational Characterization of the Insulin Dimerization

($\Delta G_{\text{solv}}^{\text{monomer1}} + \Delta G_{\text{solv}}^{\text{monomer2}}$). Each term can be separated into an electrostatic part $\Delta G_{\text{solv,elec}}$ and a non polar contribution $\Delta G_{\text{solv,np}}$.

$$\Delta G_{\text{solv}} = \Delta G_{\text{solv,elec}} + \Delta G_{\text{solv,np}} \quad (4)$$

The nonpolar contribution to the solvation free energy is assumed to be proportional to the solvent accessible surface area (SASA). This approximation is often used and is justified by the fact that the solvation energy of saturated non polar hydrogen bonds is related to the SASA through $\Delta G_{\text{solv,np}} = 0.0072 \times \text{SASA}$.³¹⁻³³ The solvent accessible surface areas were calculated analytically with CHARMM. The $\Delta G_{\text{solv,elec}}$ was calculated with the generalized Born GB-MV2 model as implemented in CHARMM which is much faster than solving the Poisson equation and makes it applicable to a large set of structures. All enthalpic contributions were calculated using a distant dependent dielectric with $\epsilon = 1$, a 12 Å cutoff for Van der Waals and electrostatic interactions. In the GB-MV2 model the expression of Still et al.³³ is used to calculate $\Delta G_{\text{solv,elec}}$.

$$\Delta G_{\text{solv,elec}} = k \sum_{ij} (q_i q_j / \sqrt{r_{ij}^2 + \alpha_i \alpha_j e^{-r_{ij}^2 / K_s \alpha_i \alpha_j}}) \quad (5)$$

with $k = -166.0(\epsilon_{\text{solute}}^{-1} - \epsilon_{\text{solvent}}^{-1})$, ϵ_{solute} and $\epsilon_{\text{solvent}}$ are the dielectric constants of the solvent ($\epsilon_{\text{solvent}} = 80$ for water) and the solute ($\epsilon_{\text{solute}} = 1$ for the protein). α_i , α_j are the Born radii of the atoms i and j and $K_s = 8$ which is different from the original Still expression, where $K_s = 4$. All enthalpic terms were calculated for frames taken every 4 ps of the trajectory.

Entropic contribution: The entropic contribution to the binding free energy was separated in translational, rotational and vibrational parts as seen in equation 6. The total entropic contribution is the sum of three parts, which were calculated separately.

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

$$-T\Delta S_{\text{total}} = -T\Delta S_{\text{trans}} - T\Delta S_{\text{rot}} - T\Delta S_{\text{vib}} \quad (6)$$

Each of the terms is calculated using standard formulae from statistical mechanics.³⁴ For the vibrational contribution to the free energy, snapshots were taken every 100 ps along the trajectory. These structures were minimized by 10000 steps of SD minimization and followed by ABNR minimization until the root-mean-square of the energy gradient reached 10^{-5} kcal/mol/Å. All calculations were carried out with a distant depending dielectric with $\epsilon = 4$. From these minimized structures the normal mode frequencies were calculated with the VIBRAN module implemented in CHARMM. A cutoff of 12 Å for the non bonded interactions and the distance dependent dielectric constant was set to $\epsilon = 4$ as for the structure minimization. The dimerization energy was then calculated by taking the difference between the energy of the dimer and two times the value calculated for the monomer.

Dynamical Cross Correlation Maps

Dynamical cross correlation maps quantify the coupling between different residues during an MD simulation. The cross correlation coefficient C_{ij} of residues i and j is given by

$$C_{ij} = \frac{\langle \Delta r_i \Delta r_j \rangle}{\sqrt{\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle}} \quad (7)$$

where Δr_i and Δr_j are the displacements of the backbone atoms from the reference position. C_{ij} varies between 1 and -1 , where 1 corresponds to movement in the same direction (correlated) and -1 movement in the opposite direction (anti-correlated). In a dynamical cross-correlation map (DCCM), motion in the same direction (correlated) can be found in the upper left triangle, while motion in opposite directions (anti-correlated) is in the lower right panel. Typically, correlated motions are much more pronounced than anti-correlated motions. α -helices are manifest as a broadening of the diagonal while β -sheets appear off-diagonal (ascending for parallel, and

descending for anti-parallel β -sheets). It should be noted that DCCM do not give information about the magnitude or the direction of the motion. The calculated cross correlations were averaged over windows of different length. Windows of 50 ps were chosen to evaluate along the calculated timescale for each trajectory separately.

Results and Discussion

First, the structures of the native and mutant insulin monomers and dimers are characterized. Then, the dimerization energies of the native and mutant proteins are compared and discussed at an atomistic level. Finally, the results are discussed in the context of experimental and previous simulation results.

Dynamics of the native and mutant insulin monomers

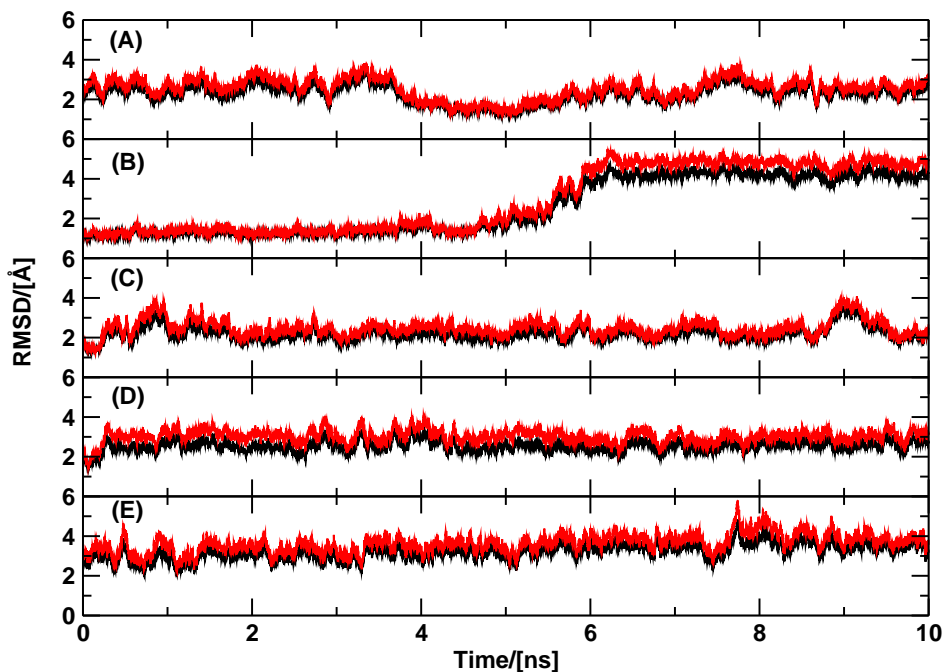


Figure 2: RMSD along the trajectories for the backbone atoms of the native and mutant insulin monomers. (A) Wildtype (B) B24Gly (C) B24Ala. Alanine substitutions (alanine scanning mutagenesis) are one of the standard procedures (most conservative mutation) to interrogate protein stability^{35,36} or protein function⁵ upon changing the amino acid sequence of a protein. (D) B24D-Ala (E) B24Des-B25. Black: RMSD of the entire protein superimposed by the backbone of the protein. Red: RMSD of the entire protein superimposed by the backbone atoms of helices (residue 1 to 8 of chain A, residue 12 to 18 of chain A, and residue 9 to 19 of chain B).

5 Computational Characterization of the Insulin Dimerization

The Native monomer: The insulin monomer shows a RMSD relative to the initial X-ray structure of around 3 Å. This is indicative of a high degree of flexibility and was already observed in previous simulations.²¹ The transition in the present work already occurs during the equilibration phase and is therefore not visible in the RMSD(*t*) curves shown in Figure 2a). The black trace corresponds to superimposing the backbone atoms whereas the red trace measures the RMSD for the helices only (residues 1 to 8 and 12 to 18 of chain A; residues 9 to 19 of chain B). Both RMSD are quite close to each other, which means that the RMSD changes of the entire monomer is caused by the C-terminal of chain B or/and N-terminal of chain A.

The GlyB24 monomer: The RMSD fluctuations for the GlyB24 mutant are considerably different from those of the native dimer. 2-dimensional NMR experiments have suggested that residues B19-B30 of the C-terminal part have a distinctly different structure than that of native insulin.⁸ GlyB24 is an insulin analog which maintains considerable biological activity (22 % and 78 % potency relative to native insulin, respectively^{11,37}) compared to the native hormone even though NMR studies suggest complete loss of the characteristic B-chain β -turn⁸ which in native insulin allows the extended B-chain C-terminal region to fold against the central B-chain helix. Fluorescence anisotropy studies of the GlyB24 mutant (with an additional TrpB25 mutation) found partial maintenance of the B-chain β -turn under near-physiological conditions.³⁸ It is interesting to note that the B-chain β -turn has been found to be maintained in all crystal structures except for one structure⁸ where this β -turn was absent. It is possible that the experimental conditions (pH 1.9), under which the original NMR studies were carried out, contribute to the observation of this unusual structure.^{8,38}

The AlaB24 monomer: For the Alanine mutant also a departure from the monomer structure in the native dimer is observed. Typical RMSDs are around 3 Å with occasional excursions to up to 4 Å. This is between the behaviour found for the native and the GlyB24 monomer. The structure of the AlaB24 mutant monomer has not been characterized experimentally. However,

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

this mutant shows a lower propensity for dimerization³⁹ (*vide infra*) which makes it interesting for the discussion in the present context. Also, insulin analogs with PheB24 replaced by biphenylalanine have recently been synthesized and characterized.⁴⁰

The D-AlaB24 monomer: It was found that substitution with D-Alanine at position B24 leads to a very active insulin (affinity of 150 % relative to native insulin).¹¹ The structural changes found in the present simulations indicate that the monomer deviates from the reference native monomer structure quite rapidly but does not show much different behaviour than the native or AlaB24 monomer.

The Des-B25 monomer: Finally, the RMSD from the starting conformation of the native insulin monomer during the 10 ns MD simulation, are calculated for the entire protein superimposed by the backbone of the protein (black line) or superimposed by the helices. As expected, the Des-B25 monomer is considerably flexible with RMSD value around 4 Å, and sometimes even up to 6 Å during the trajectory (see Figure 2e). The absence of B25 residue may cause the β -sheet to unfold, which is part of dimer interface. This agrees with the smaller dimerization free energy of Des-B25 dimer (*vide infra*) compared to native insulin. In previous work by Jørgensen et al.,² they found that the C-terminal end of the B-chain packs closer against the rest of molecular in des-B25 monomer. Compared to native insulin monomer, deletion of the B25 residue shifts the hydrophobic ProB28 residue to position B27, subsequently turns the hydrophobic side-chain of this residue away from the surface and to position B chain α -helix of the monomer and eliminate the hydrophobic surface at the C-terminal end of the B-chain.

Dynamics of the native and mutant dimers

To assess the stability of the simulations for the dimers, the RMSD for backbone atoms of entire protein and backbone atoms of the helical regions (A2 to A9, A14 to A20 and B9 to B19) along

5 Computational Characterization of the Insulin Dimerization

all trajectories was calculated.

The RMSDs for backbone atoms of wildtype dimers and insulin mutants (SI-Figure 1) were calculated vs. X-ray native insulin structure show that for WT insulin, the RMSD of the entire protein is well within 2 Å. Hence, the WT insulin is stable in its dimeric form. The RMSD of the B24Ala mutant are indicative of more flexible structures whereas the RMSD of B24Gly and B24D-Ala are ≈ 3 Å and the RMSD of Des-B25 insulin dimer is large since the structures change dramatically during the simulation.

Dynamical cross-correlation maps (DCCM) allow to analyze persistent, (anti-)correlated motions of residue pairs from the MD simulations and serve as a quantitative measure for the collective motion in a protein. The helices extend from residues A2-A8, A13-A19, and B9-B19 and the disulfide bonds are located between residues (A6,A11), (A7,B7), and (A20,B19). Correlated motions between the A- and B-chain are quite similar for the native, AlaB24, and GlyB24 whereas they are considerably reduced for the D-AlaB24 and Des-B25 mutant. In particular, the correlation of the residues on the dimer interaction interface B24-D26, B25-D25 and B26-D24 of the Des-B25 mutant are much lower, which will affect the dimerization free energy of Des-B25 mutant dimer, as shown below.

The Native dimer: In Figure 3, the DCCMs for the native and mutant insulin dimers are shown. Most of the strong correlations have their origin in structural features such as disulfide (A7-B7, A20-B19, and A6-A11) or hydrogen bonds. In all DCCMs it is evident that the two monomers are not equivalent because the C-termini of the two B-chains have different conformations. For monomer II, correlated motions between A1 and B30 are found which are weaker or absent for monomer I, as the conformation is different and does not allow for interaction between these residues. The only appreciable correlation between the two insulin monomers are between the C-termini of chains B1 and B2 (residues B41 to B51) which constitutes the dimerization inter-

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

face. In particular, residues B24 to B26 are correlated and it has been found that these residues contribute considerably to the stability of the insulin dimer.^{4,22} Superposition of selected snapshots reveals a compact structure of the native dimer. This is also reflected in the RMSD along the trajectory. It is interesting to note that - except for the fluctuations at the end of the B-chain - the RMSDs for monomer I in the dimer and in the monomer are comparable in magnitude. However, the RMSD-pattern differs and fluctuations around the average are larger in monomer II than in monomer I.

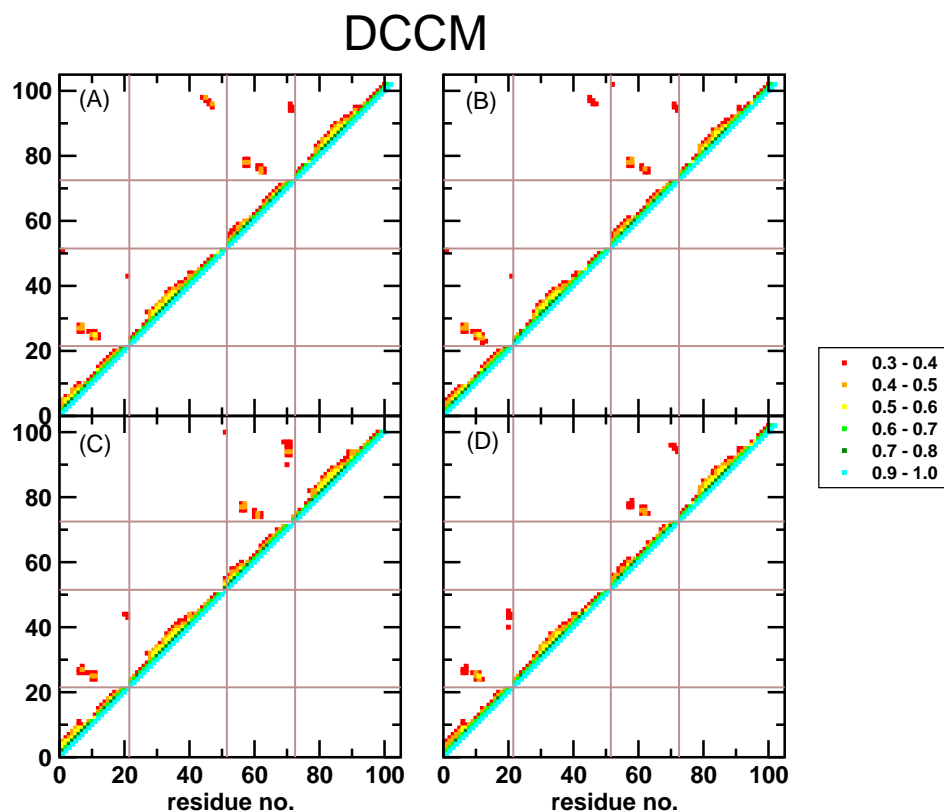


Figure 3: Dynamical cross-correlation maps for wildtype (A), B24Gly-SI (B), B24Gly-WI (C) and B24-Des-B25 (D) insulin dimer. The correlation coefficient are averaged over windows of 50 ps. Positive cross-correlated coefficients are collected in the upper-left triangle. Only cross-coefficients larger than 0.30 are shown. The color code is: red: $C_{ij} = 0.30 - 0.40$; orange: $C_{ij} = 0.40 - 0.5$; yellow: $C_{ij} = 0.5 - 0.6$; green: $C_{ij} = 0.60 - 0.70$; deep green: $C_{ij} = 0.7 - 0.8$; cyan: $C_{ij} = 0.8 - 0.9$

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

The GlyB24 dimer: The DCCM for this mutant is comparable to the one for the AlaB24 mutant dimer. However, the two correlations (TyrB26(I)–AsnA21(II) and CysA20(I)–ProB28(II)) are stronger, whereas the GlyA1(II)–AlaB(30) correlation is weaker. The superposition of selected structures shows more variability compared to the native dimer than the AlaB24 mutant whereas the RMSD are qualitatively similar. The fluctuations in the RMSD of monomer II are larger whereas RMSD and its fluctuations in monomer I are comparable to those of the native and the AlaB24 mutant. The Ramachandran plot of the β -turn in GlyB24 (see Figure 4b) shows that compared to wildtype insulin, the β -turn (residues 20 to 23) of the GlyB24 undergoes more structural changes especially for residue 23 (black trace). Moreover, the Ramachandran plot of the β -sheet (residues 24 to 28) in GlyB24 also experience larger structural changes. Residue Pro28 switches between two conformations and the structure of residue 24 fluctuates more which may affect the structure of the β -turn. To further study the structural changes upon the mutation by Glycine, we investigated the pseudodihedral angle (four successive CA atoms) of the β -turn and the β -sheet. (see Figure 6 in the SI). Compared to the native insulin dimer (SI-Figure 6a), the β -turn (black peak) samples a wider range of the pseudo-dihedral in the B24Gly mutant.

The AlaB24 dimer: The correlated motions include all the ones already found for the native dimer, although with slight differences in their intensity. Correlations around the disulfide bonds are somewhat larger as is the correlation between the C-termini between chains B1 and B2. More significantly, an additional correlation between TyrB26(I) and AsnA21(II) was found which is only weak in the native dimer. The superimposed structures are less compact than the native dimer although the RMSD fluctuations are quite similar. Differences are found in the C-terminal regions of monomers I and II. In particular, the C-terminus of monomer II is pushed towards chain A. Again, the two monomers are found to behave differently and the fluctuations around the mean are larger for monomer II than for monomer I.

5 Computational Characterization of the Insulin Dimerization

The D-AlaB24 dimer: Compared with the native insulin dimer, the DCCM of the D-AlaB24 calculated from all eight independent trajectories shows that there are weaker interactions between residue 24 of chain B (B24) and residue 26 of chain D (D26). This hydrogen bond interaction between B24 and D26 is one of important interactions at the dimerization interface and contributes to the dimerization free energy of the dimer. The RMSD of backbone atoms of the D-AlaB24 dimer is stable with slightly larger amplitudes compared to native insulin dimer (see Figure 1d in the SI). In the Ramachandran plot of the C-terminal of chain B, Figure 4d), it is found that the Φ value of the D-Alanine residue (black trace) covers a wider conformational range compared to AlaB24 and the wildtype insulin dimer. Moreover, the residue 26 (blue trace) suggests that two conformations are present and the conformation of the C-terminal of chain B is changing accordingly. For the wildtype insulin, the PheB24 aromatic ring stabilizes the β -turn (residue B20 - B23) structure by packing against residue 19. When residue 24 is substituted by D-Ala the structure of the β -turn is more flexible during the trajectory with the larger fluctuation area of $[\Phi, \Psi]$ for all 4 residues in the β -turn.

The Des-B25 dimer: The cross-correlation map of the Des-B25 insulin mutant (see Figure 3d) shows that there is no or little correlated motion between B24 and D26. This is related to the loss of H-bonding interaction between B24 and D26 as the deletion of B25 changes the monomer-monomer interface and leads to two weakly interacting monomers. The RMSD of Des-B25 dimer is larger than all other four insulin mutants, around 4 Å as is shown in SI-Figure 1. As already discussed for the monomers, in the Des-B25 mutant the hydrophobic side-chain of ProB28 can interact with the hydrophobic surface formed by residues ValB12, GluB13, and LeuB15. The structure reveals that the lost ability of Des-B25 insulin to self-associate is caused by a conformational change of the C-terminal region of the B-chain. Des-B25 insulin does not form the anti-parallel β -sheet between two monomers that characterizes the insulin dimer. This is primarily found for the C-terminal end of the B-chain which packs closer against the rest of the molecule in the Des-B25 monomer than in the monomers of the dimeric porcine insulins.

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

Deletion of the PheB25 residue not only moves the hydrophobic ProB28 residue to position B27, it also results in an intra-molecular hydrophobic interaction between Pro(B28) and the hydrophobic region LeuB11-LeuB15 of the B-chain α -helix. This interaction interferes with the inter-molecular hydrophobic interactions responsible for the dimerization of native insulin, depriving the mutant of the ability to dimerize. It results in an intra-molecular hydrophobic interaction between Pro(B28) and the hydrophobic region LeuB11-LeuB15 of the B-chain α -helix. This interaction interferes with the inter-molecular hydrophobic interactions responsible for the dimerization of native insulin, depriving the mutant of the ability to dimerize.²

5 Computational Characterization of the Insulin Dimerization

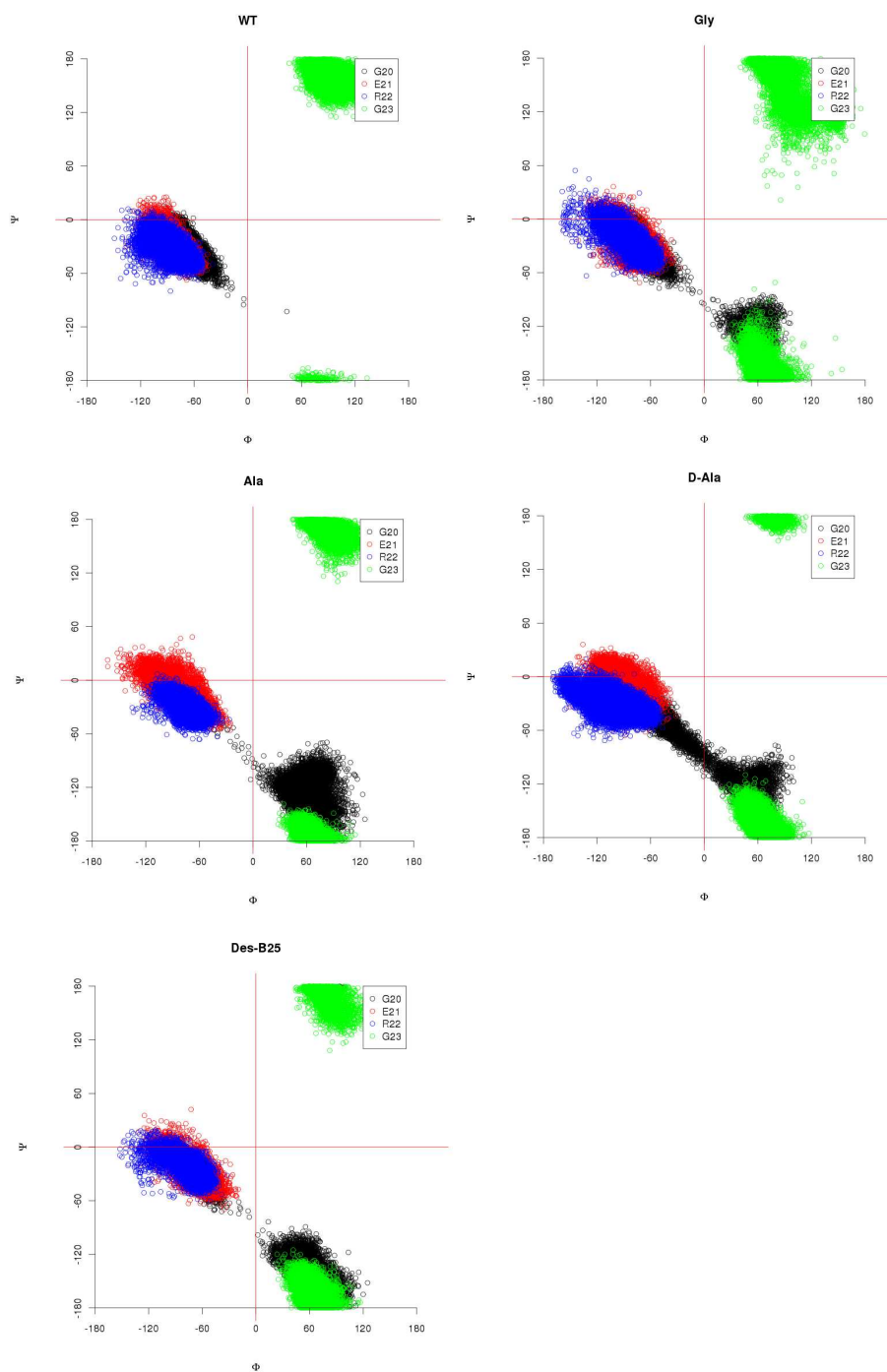


Figure 4: Ramachandran plot of β -turn (residue 20 to 23) of (a) wildtype insulin dimer (b) GlyB24 insulin dimer (c) AlaB24 insulin dimer (d) FB24D-Ala insulin dimer (e) Des-B25 insulin dimer during 10 ns trajectory length. One dot per ps, 10000 dots for each residue. Black is for B20Gly, red for B21Glu, blue for B22Arg, and green for B23Gly.

Dimerization Energy Calculation

A central aspect of the present work is the calculation of the relative stabilities of mutated insulin dimers relative to that of the native dimer. Experimentally, a value of -7.2 kcal/mol in favor of the dimer was determined⁴¹ for native insulin which compares with $\Delta G_{\text{bind}} = -11.9$ kcal/mol (absolute binding free energy) and $\langle \Delta G_{\text{bind}}^0 \rangle + \langle \Delta G_{\text{desolv}} \rangle = -38.7$ kcal/mol (enthalpic contribution to ΔG_{bind}) from molecular dynamics simulations,⁴ where $\Delta G_{\text{bind}} = \langle \Delta G_{\text{bind}}^0 \rangle + \langle \Delta G_{\text{desolv}} \rangle - T\Delta S$.

Table 1: Binding Free Energy for the Dimerization of insulin and Contributions of Solvation, van der Waals, and Electrostatic Interactions and Entropic Terms, using the Same Trajectory Method (kcal/mol)

	WT	Gly	Ala	D-Ala	Des-B25
$\langle \Delta E_{\text{vdw}} \rangle$	-66.59(4.36)	-53.39(4.35)	-56.85(4.19)	-60.39(5.06)	-44.47(5.21)
$\langle \Delta E_{\text{ele}} \rangle$	-114.38(34.28)	-107.84(40.75)	-91.79(35.53)	-104.85(33.88)	-56.15(32.19)
$\langle \Delta G_{\text{ele,desolv}} \rangle$	146.23(32.58)	132.83(38.36)	120.17(34.27)	142.77(30.85)	92.43(29.57)
$\Delta G_{\text{np,desolv}}$	-10.28(0.61)	-9.76(0.49)	-9.89(0.42)	-10.54(0.48)	7.99(0.61)
$\langle \Delta G_{\text{bind}}^0 \rangle + \langle \Delta G_{\text{desolv}} \rangle$	-45.02(6.46)	-38.16(7.27)	-38.36(7.49)	-33.00(9.47)	-16.19(6.47)
$\langle \Delta S_{\text{trans}} \rangle$	14.98	14.98	14.98	14.99	14.98
$\langle \Delta S_{\text{rot}} \rangle$	13.94	13.82	13.85	13.82	13.82
$\langle \Delta S_{\text{vib}} \rangle$	3.82	1.48	-0.58	-1.14	-1.70
$-T\langle \Delta S \rangle$	32.74	30.28	28.26	27.67	27.11
$\langle \Delta G_{\text{bind}} \rangle$	-12.3	-7.9	-10.1	-5.3	10.9

The enthalpic contribution $G_{\text{enthalpic}} = E_{\text{vdw}} + E_{\text{elec}} + G_{\text{solv,elec}} + G_{\text{solv,nb}}$ is a first quantity which allows to qualitatively compare the stabilities of the different dimers relevant in the present study. For the native structure $\langle \Delta G_{\text{bind}}^0 \rangle + \langle \Delta G_{\text{desolv}} \rangle$ (ΔG_{bind}^0 is the gas phase contribution to the binding free energy $\Delta G_{\text{bind}}^0 = \Delta H_{\text{gas}} - T\Delta S_{\text{gas}}$) from the two trajectories is $\approx -45 \pm 5$ kcal/mol which is reduced to -38 ± 6 kcal/mol for the AlaB24 mutant and $\approx -40 \pm 7$ kcal/mol for the GlyB24 mutant. From all three systems investigated the native protein forms the most stable dimer. The Gly mutant dimer is less stable than the native protein by 5 kcal/mol and the Alanine mutant is destabilized by 7 kcal/mol with respect to the native structure. These destabilizations compare with a stabilizing enthalpic contribution of the two phenylalanine residues at position

5 Computational Characterization of the Insulin Dimerization

B24 which has been calculated to be -3.92 and -2.68 kcal/mol, respectively.⁴

The Native dimer: Wildtype insulin dimer is enthalpically stabilized by ≈ -45 kcal/mol as calculated from all the 15 trajectories with 10 ns for each trajectory, see Table 1. In Table 1, all the mutants have lower binding free energy compared to the wildtype insulin dimer which suggests that these dimers are less stable. As the absolute binding free energy of wild type insulin is -7.2 kcal/mol and as we do not expect that differences in going from relative to absolute binding free energies are greatly affected, one already can conclude that all mutant dimers are only marginally stable.

The B24Gly dimer: As shown in Table 1, the absolute dimerization free energy is around -8 kcal/mol with the entropy contribution around 30 kcal/mol. The loss of translation and rotation and vibration entropy due to dimerization process indicate that the process is entropically highly unfavorable. However, the more negative enthalpy value propose that the dimerization is driven by the enthalpy contribution. For the GlyB24 dimer, two types of situations are found in the simulations. In one of them the GlyB24 monomers interact with each other with a similar strength as the WT dimer, i.e. $\Delta \approx -43$ kcal/mol. In the other case, the GlyB24 dimer is destabilized by 20 kcal/mol compared to the WT dimer. (Detail discussion will be listed later.)

The B24Ala dimer: The absolute dimerization free energy of B24Ala mutant dimer is about -10 kcal/mol averaged from all the trajectories. Compared to B24Gly, the dimerization free energy of B24Ala are relative more negative (favoring association), which may due to the side-chain of alanine is larger in size and more hydrophobic.

The B24D-Ala dimer: The average absolute dimerization free energy of B24D-Ala mutant dimer from 10 trajectories is around -5 kcal/mol. (See Table 1) This is in line with the conclusion we got from the structure that D-Ala substitution has more flexible structure and a bit more

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

favorable in monomer form compared to native insulin.

The Des-B25 dimer: Des-B25 mutant dimer has a positive dimerization free energy of around 11 kcal/mol with the contributions of enthalpy (-16 kcal/mol) and entropy (27 kcal/mol). Table 1 the positive dimerization free energy value indicates that Des-B25 mutant dimer has dramatic structure changed, and are monomeric form favored. This is in line with what we observed in the structure characters. Compared to native and other mutant insulin dimer, the vibration entropic part is relative larger but still can not compensate the lost translational and rotational motion. Meanwhile the enthalpic is relative unfavorable, which may due to the loss of two hydrophobic surfaces as we discussed before for dimerization compared to native and other mutants.

To better characterize the structural changes along a trajectory, the secondary structural elements analyzed for each snapshot in a 10 ns simulation are reported in Figure 5. For both, wildtype and mutant insulin dimer, the structure component of monomer I and monomer II along the trajectory are not identical. This is particularly evident for the AlaB24 and D-AlaB24 mutants, see Figure 5c and d. In the AlaB24 mutant, residues 12 to 18 form a 3-10 helix or a turn in monomer I instead of a α -helix as in monomer II. In the D-AlaB24 mutant, residues 20 to 23 of chain D in monomer II form a 3-10 helix instead of a β -turn in monomer I. Secondly, there are different structural components between wildtype and mutant insulin dimers. This is mainly found by considering the helix (residues 13 to 19) of chain B, which indicates that residue 24 interacts closely with residues on helix of chain B, whereas substitution at position B24 will affect the structure around the β -turn (residues 20-23). Moreover, in D-AlaB24 there is a turn formed by the C-terminal of β -sheet of chain B in monomer I and the N-terminal of the α -helix of chain A in monomer II. This demonstrates that there is a major structural change for the D-AlaB24 mutant dimer, and the turn between chain B of monomer I and chain A of monomer II causes the structure changed accordingly but not contribute to dimer binding affinity, which maybe another evidence that the dimerization free energy is contributed mainly by

5 *Computational Characterization of the Insulin Dimerization*

residues on the monomer-monomer interface. In the Des-B25 mutant the N-terminal β -strand in chain B of monomers I and II are not extended configuration anymore but more disordered compared to other mutants or the wildtype insulin dimer.

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

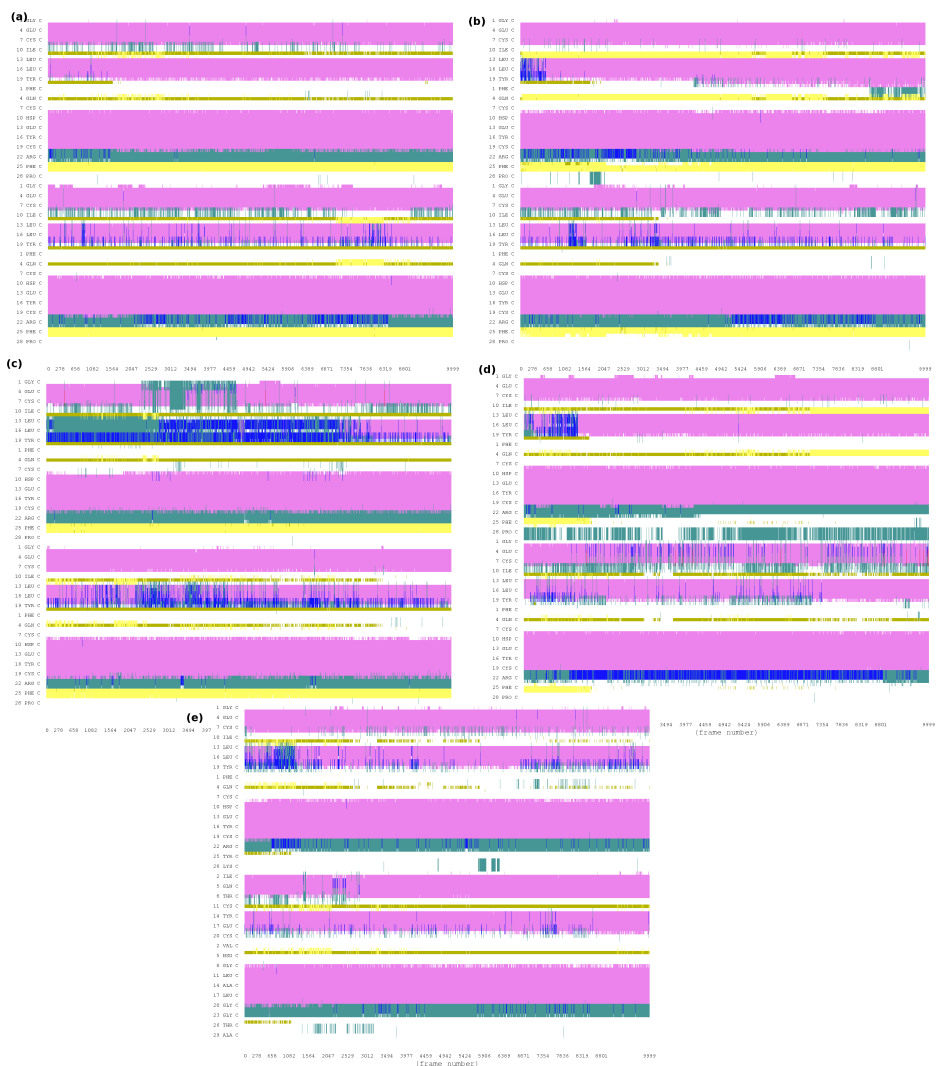


Figure 5: Time vs. structural component of (a) wildtype, (b) FB24G, (c) FB24A, (d) FB24D-Ala, (e) Des-B25 insulin dimer. Green represents Turn, yellow represents extended configuration, light green represents isolated bridge, pink represents alpha helix, blue represents 3-10 helix, and red represents pi-helix.

The B24Gly Mutant

For the GlyB24 dimer, two types of situations are found in the simulations. As these observations may also be relevant for understanding the stabilization found in the wild type and other mutant insulins, these aspects are discussed in more detail in the following. The simulations analyzed so far show that either the GlyB24 dimer interacts almost as strongly as the WT dimer ($\Delta G \approx 43$ kcal/mol) or that the dimer is considerably destabilized by almost 20 kcal/mol relative to the WT dimer. These two situations are labelled “weakly” and “strongly” interacting (WI and SI) dimers, respectively.

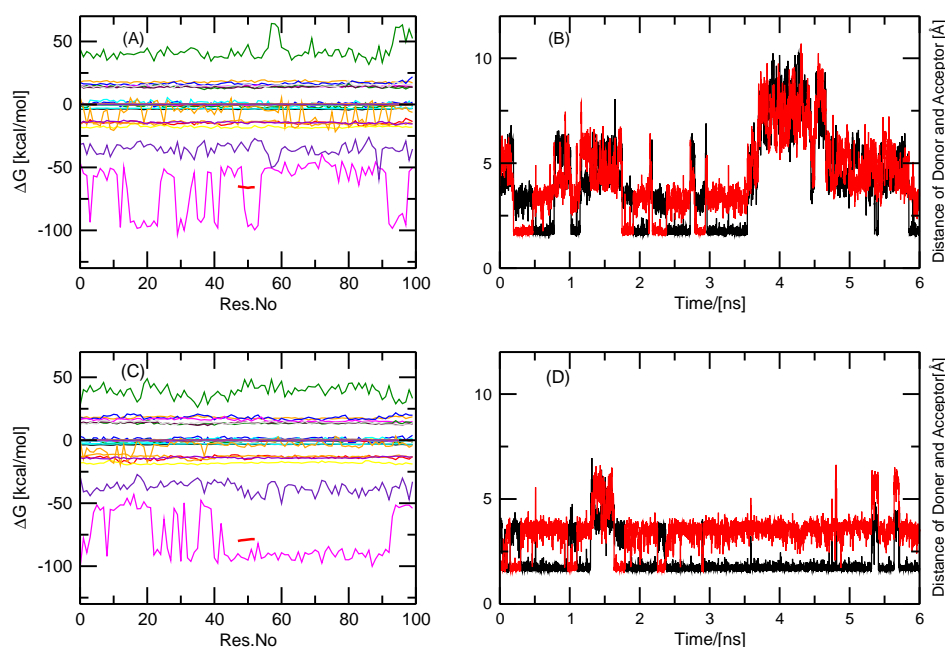


Figure 6: The Electrostatic interaction energy between B13Glu and all other residues on the B24Gly mutant insulin dimer. (A): B24Gly-SI while (C): B24Gly-WI; (B)-(D) are the distances between the hydrogen bond donor and acceptor on residue B13Glu and residue B10Hip. red line: the distance between OE1-B13Glu and N1-B10Hip; black line: the distance between OE2-B13Glu and N1-B10Hip along the trajectories.

Detailed analysis of the per-residue contributions and the average interaction energies of sidechains

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

with their neighboring sidechains shows that these differences mainly arise from contributions of the electrostatic interaction $\langle E_{\text{elec}} \rangle$ and solvation energy $\langle G_{\text{elec,desolv}} \rangle$. We found that for wildtype insulin and the SI-GlyB24 mutant the per-residue binding free energies shared more similarity while for the WI-GlyB24 they per-residue contributions differ. One residue that contributes much to the difference between SI and WI is GluB13 (see Figure 7), specifically in terms of the electrostatic interaction. To further explore this, we studied the electrostatic interaction energy between GluB13 and the surrounding residues for WI and SI-GlyB24 (see Figure 6). Figure 6a,c reports the electrostatic interaction energy between GluB13 and all other residues for two different trajectories. In one trajectory (pink, SI) $E_{\text{elec}} \approx -50$ kcal/mol while in the trajectory corresponding to the WI dimer $E_{\text{elec}} \approx -80$ kcal/mol. Structurally, this can be related to a hydrogen bond between GluB13_O and HisB10_H, as is shown in Figure 6b,d. This H-bond will lead to a monomer-self-interaction which eliminates it from the dimerization energy and reduces the solvation energy and partially explains the difference between the two types of behaviors of the GlyB24 mutant.

5 Computational Characterization of the Insulin Dimerization

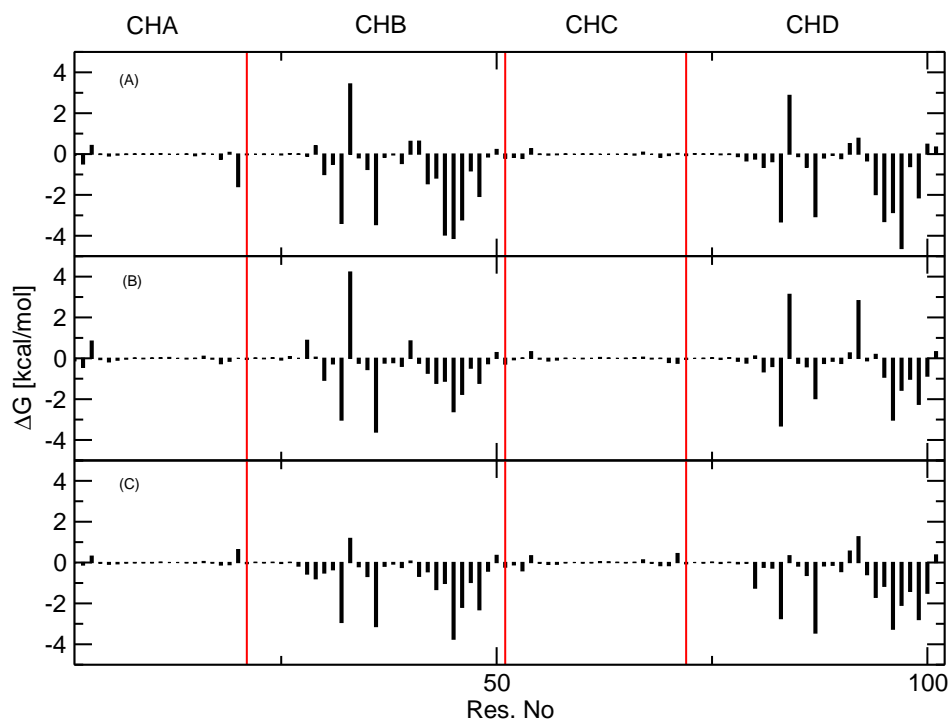


Figure 7: Per-residue contribution to the total dimerization free energy of wildtype insulin(A), B24Gly-mutant-WI (B) and B24Gly-mutant-SI (C). X axis is the residue number, 102 in total for insulin dimer; Y axis is the dimerization free energy in kcal/mol.

To further explore this effect, additional analysis of the per-residue contributions (see Figure 8) is carried out. The dimerization free energy is dominated by contributions from chains B and D. Particularly large differences between WI and SI are found for residues 22 to 26 of chain B and residues 19 to 26 for chain D. A comparison of WT and a WI-GlyB24 trajectory is shown in Figure 8. The left panel reports the difference in per-residue contribution to the total ΔG for chain B whereas the right panel corresponds to chain D. For both chains, 20 % of the difference is for the first 20 residues whereas the remaining 10 residues contribute 80 % to the difference. Further analysis shows that the backbone of chain B contributes more to $\Delta\Delta G_{WT/GlyB24}$ than the side-chain by ≈ 3 kcal/mol, whereas for chain D the contributions are about equal.

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

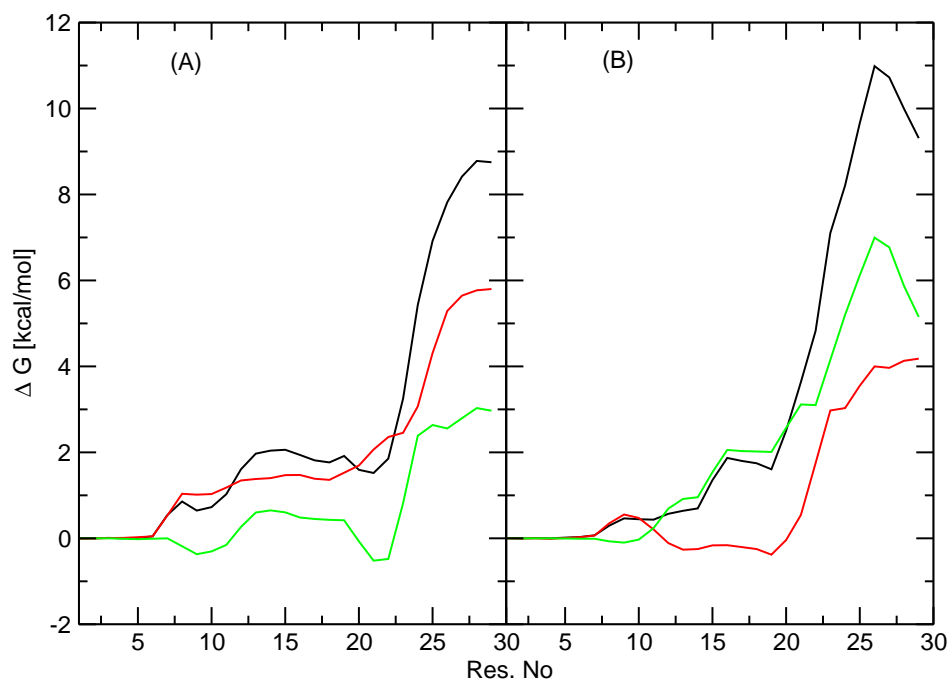


Figure 8: The difference of per-residue contribution to the total dimerization free energy from wildtype insulin and B24Gly-mutant-WI insulin. (A) The difference between chain B of WT and chain B of B24Gly-mutant-WI. (B) The difference between chain D of WT and chain D of B24Gly-mutant-WI. Black line: all atoms on chain B; red line: backbone atoms; green line: sidechain atoms. X axis is the residue number, 30 in total for chain B and chain D respectively; y axis is the dimerization free energy in kcal/mol.

The RMSD for the backbone of residues 22-26 of chain B is shown in Figure 9 for the wildtype and glycine mutant, respectively, for a 10ns trajectories. The RMSD was calculated for residue 22-26 chain B and 20-28 chain D after superimposing to itself separately. The black trace is the RMSD of chain B for the wildtype, the red trace corresponds to the GlyB24 mutant and the green curve is the RMSD difference between these two trajectories while the blue trace is the average of the RMSD difference which is around 1 Å. This suggests that there is no pronounced structural change for residues 22-26. Thus, the difference in binding free energy between wildtype and glycine mutants may be caused by the covalent bonds between residues and finally affect the dimerization free energy.

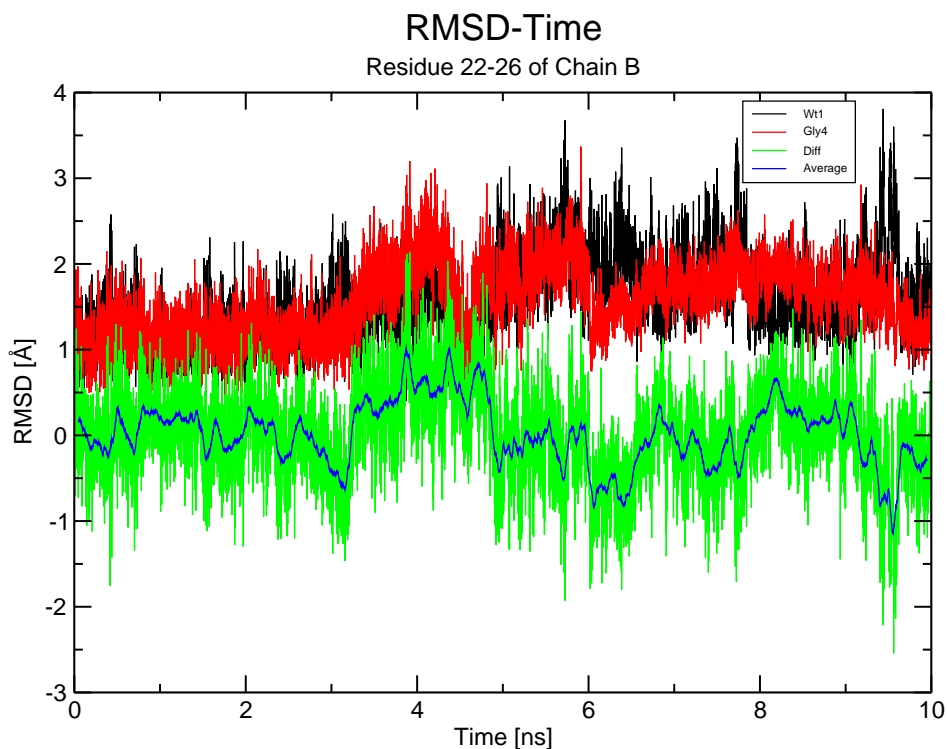


Figure 9: RMSD of backbone of residue 22-26 chain B during 10ns trajectories.

The following 6 H-bonds (donor-acceptor) between chains B and D are analyzed in more detail: Y26B-G24D, Y26B-G20D, G24B-Y26D, Y26B-F25D/G24B-T27D, G24D-Y26B, and Y26D-G24B (see Table 2 - Table 2). The total number of H-bonds as a function of simulation time between residues 22-26 (CHB) and residues 20-28 (CHD) are shown in Figure 10 for WI-GlyB24. It can be clearly seen that there are fewer hydrogen bonds in this case. Therefore, the contribution of H-bonds to the dimerization energy is reduced for the WI-GlyB24 mutant compared to the WT. The histograms in Figure 10 underline this finding. The probability distributions for the WT and SI-GlyB24 mutant are quite similar and show an average of 2.5 H-bonds compared to only 1.2 H-bonds for the WI-GlyB24 mutant.

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

Table 2: The hbonds between residue22-26 chain B and residue 20-28 chain D from B24Gly-SI ($\Delta G \approx -43$ kcal/mol) and B24Gly-WI ($\Delta G \approx -25$ kcal/mol)

Donor	Acceptor	Occupancy
B24Gly-SI		
CHB-Tyr26	CHD-GLy24	31%
CHB-Tyr26	CHD-Gly20	3 %
CHB-Gly24	CHD-Tyr26	7 %
CHB-Tyr26	CHD-Phe25	0.08%
CHD-Gly24	CHB-Tyr26	28 %
CHD-Tyr26	CHB-Gly24	36 %
B24Gly-WI		
CHB-Gly24	CHD-Tyr26	5.8 %
CHB-Tyr26	CHD-Gly24	6.6 %
CHB-Gly24	CHD-Tyr27	0.04 %
CHB-Tyr26	CHD-Gly20	0.02 %
CHD-Tyr26	CHB-Gly24	10.3 %
CHD-Gly24	CHB-Tyr26	3.12 %

5 Computational Characterization of the Insulin Dimerization

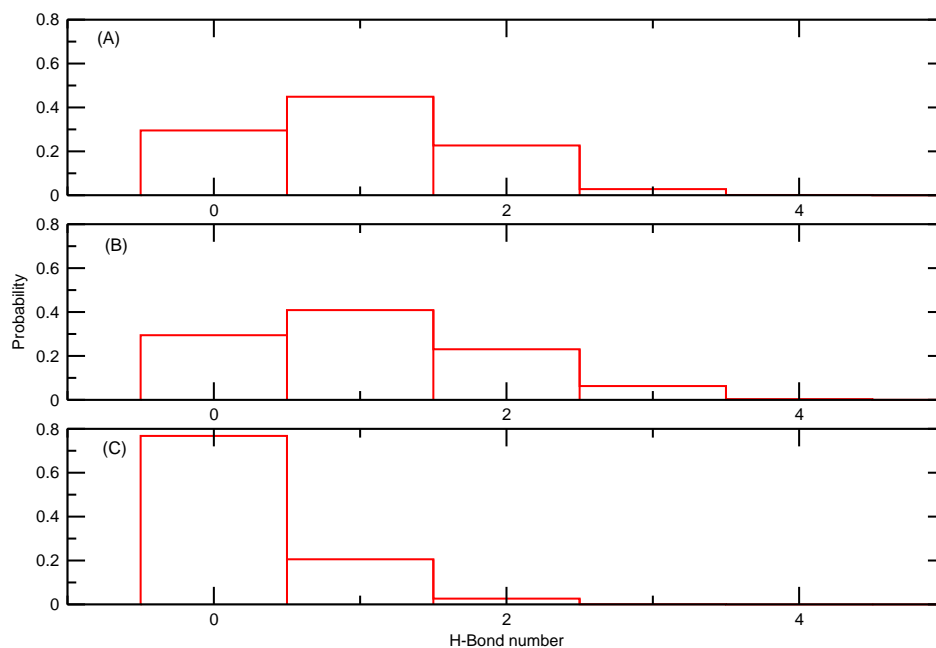


Figure 10: The number of hydrogen bonds between residue 22-26 chain B and residue 20-28 chain D. (A) wildtype insulin dimer; (B) SI-B24Gly insulin mutant; (C) WI-B24Gly insulin mutant.

The 6 H-bond distances for SI and WI are reported in *Figure 11*. For SI-GlyB24 (*Figure 11(A)–(F)*), the distances between the donors (N, O) and acceptors (O, N) is $\approx 3 \text{ \AA}$ for H-bonds Y26B-G24D (*Figure 11 (A)*, occupancy 31 %), G24B-Y26D (*Figure 11 (B)*, occupancy 7 %), G24D-Y26B (*Figure 11 (C)*, occupancy 28 %), Y26D-G24B (*Figure 11 (D)*, occupancy 36 %); $\approx 5 \text{ \AA}$ for Y26B-G20D (*Figure 11 (E)*, occupancy 3 %), and Y26B-F25D (*Figure 11 (F)*, occupancy 0.08 %). Conversely, *Figure 11 (G)–(L)* report the same H-bonds for the WI-GlyB24 dimer. The donor-acceptor distances are considerably larger. In some cases they increase to more than 6 \AA , which means that no H-bond between chains B and D are formed and therefore reduce the binding free energy accordingly.

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

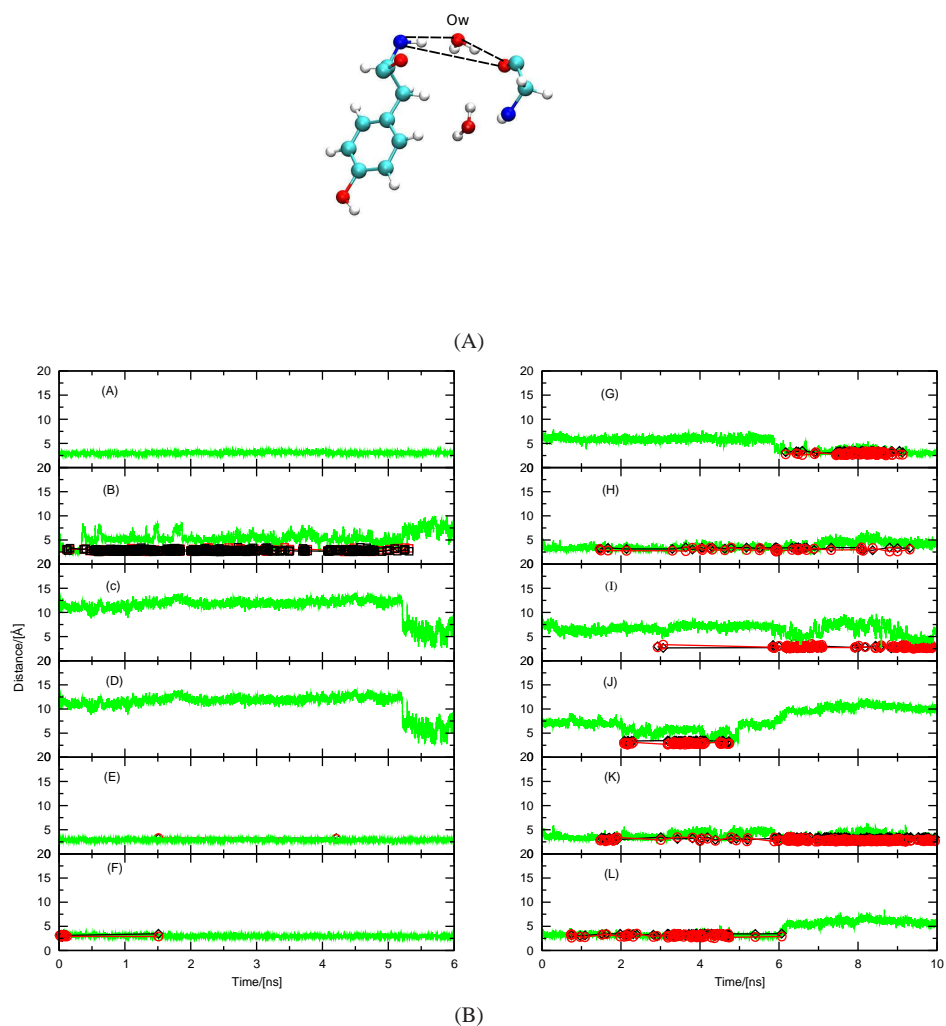


Figure 11: H-Bonds of donor and acceptor and water of Gly mutant. The distances (in Å) of H-Bond donors and acceptors on residue 22-26 chain B and residue 20-28 chain D. (A)-(F): SI-B24Gly mutant; (G)-(L): WI-B24Gly mutant. Green line : the donor-acceptor distance; black line: the distance of the water-oxygen atom (Ow) closest to the protein-donor atom; red line: the distance of the water-oxygen atom (Ow) closest to the protein-acceptor atom. Only hydrogen bonds (both Ow-donor and Ow-acceptor) distances within 3.5 Å at the same time are showed.

Because all simulations were carried out in explicit solvent, it is even possible to analyze whether the H-bonds are replaced by interactions with the surrounding solvent. This is indeed the case. In Figure 11 distances between the water-oxygen atoms and the protein H-bond donors are shown in black whereas the water-hydrogen atoms and the protein H-acceptors are displayed in red. Only distances shorter than 3.5 Å are reported. This analysis shows in which H-bonds a possible B↔D H-bond is replaced by a protein–water–protein H-bond. As can be seen, the WI-GlyB24 dimer has many such water-mediated H-bonds which, however, are not included in a typical MM-GBSA analysis.

Structural Rearrangement of the B20-B23 turn

It was supposed by Kaarsholm et al. that mutation of B24 by glycine may rearrange the C-terminal decapeptide involving a perturbation of the B20-23 turn. This may allow the PheB25 residue occupy the position normally taken up by PheB24 in native insulin.¹⁴ It will be of interest to verify this hypothesis with MD simulations. When phenylalanine in residue 24 position is mutated by glycine, the pocket which is occupied by phenyl ring become empty. We compared the structure of wild-type insulin dimer and GlyB24 insulin dimer (SI Figure 5 -Figure 6). During the 10ns simulation, for wildtype insulin, the dihedral of aromatic side-chain of residue 24 and residue 26 is quite stable while dihedral of aromatic side-chain of residue 25 which is outward from the dimer interface, is more flexible. For Gly24 mutant, for position 24 there is no sidechain, the dihedral of sidechain of residue 26 is as stable as wildtype. However the dihedral of residue 25 sidechain is much more stable compared to wildtype. From the structure, for the wildtype insulin the Phe25 is much easily accommodated in the vicinity of the A19-A21 region, while for the Gly24 mutant, the Phe25 of Chain B is far away from the A19-21 region. As the sidechain of PheB25 is outward of the dimer interface while the sidechain of PheB24 is inward of the interface, if the sidechain of PheB25 would occupy the pocket of PheB24, the backbone of residue 24 to 26 will change subsequently. From SI-Figure7, the dihedral of back-

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

bone of residue 24 to 26 of Gly24 mutant doesn't dramatically changed compared to wildtype insulin, which indicate 10ns trajectory is not long enough to see the changes for the Kaarsholm hypothesis if any. Further more, dimer may be not the best system to study this motion, because the interaction between chain B and chain D in the dimer interface.

Additionally, an alternative way were used to study whether residue 25 will move into the position occupied by PheB24, via shifting the residue B24 up to the β -turn, then putting the sidechain and backbone of residue 25 to the place where B24 were, and finally shifting residue 26-30 to compensate the space of residue 25. We observed that β -turn (Figure 12b, orange) and β -sheet of chain B (Figure 12b, cyan) have more fluctuation compared to those of wildtype insulin dimer (Figure 12a). It means that the structures of B23 to B28 undergo rearrangement and inhibit this region from forming an antiparallel β -sheet with another monomer which is necessary for dimer formation. The enthalpic stabilization free energy of this analog dimer is around -24 kcal/mol, Table 3 compared to -45 kcal/mol for wildtype insulin, along 10 ns simulation. This indicates that this analog is not dimeric favored.

5 Computational Characterization of the Insulin Dimerization

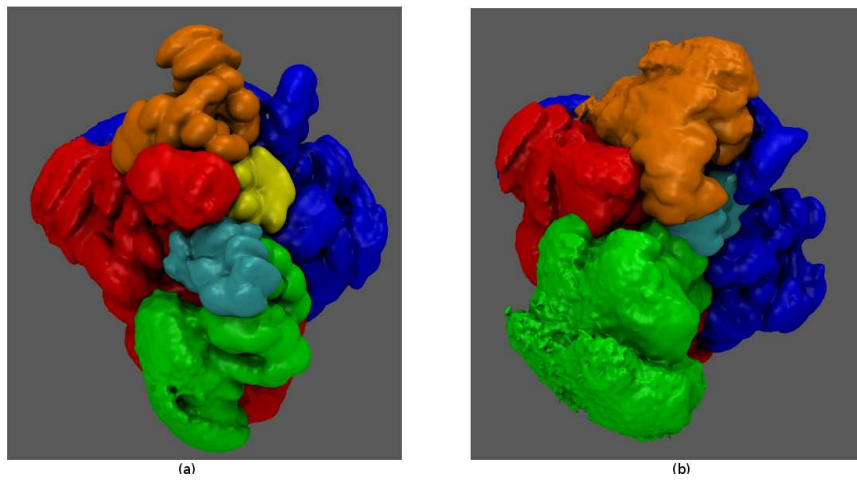


Figure 12: Density map surfaces of Chain A and Chain B of insulin dimers. (a) Wildtype (b) FB24Gly with PheB25 occupies the pocket of GlyB24. Blue: chain A; Red: residue 1 to 19 of chain B; Orange: β -turn (residue 9 to 19); Yellow: residue 24; Cyan: residue 25; Green: β -sheet (residue 26 to 30).

Conclusion

Molecular dynamics (MD) simulations have been used to study the dynamical behavior of the native and mutant insulin monomers and dimers. The native insulin dimer is very stable during the MD simulation in solution, stays close to the starting X-ray structure and has dimerization free energy around -12 kcal/mol absolute dimerization free energy (dimer favored). In native insulin monomer, chain A and helix of chain B are stable, but the C-terminal of chain B is more flexible than in dimer. GlyB24 mutant and AlaB24 mutant insulin have lower propensity for dimerization. AlaB24 dimer is slightly more stable than GlyB24 dimer, which may be due to hydrophobic side chain. The aromatic ring of B24 plays a critical role in maintaining the β -turn which tether the C-terminal of B-chain domain to hydrophobic core. D-Ala mutant monomer has more flexible structure and less stable dimerization energy. Des-B25 has monomeric character with very unstable dimerization energy. Des-B25 mutant experiences large structural changes, which leads to the disappearance of hydrophobic surface at the C-terminal of the B-chain. The large difference (≈ 20 kcal/mol) between the WI and SI-GlyB24 dimers most likely originates from H-bonds. They are formed directly H-bonds between chain B and chain D in the case of SI dimer, whereas they are water-mediated in the case of WI dimer. Our work provides an insight for designing analogues of human insulin and thus a therapy for diabetes.

REFERENCES

- (1) Voet, D.; Voet, J. 2nd ed.; John Wiley and Sons, Inc.: New York, NY, 1995.
- (2) Jorgensen, A. M. M.; Olsen, H. B.; Balschmidt, P.; Led, J. J. *J. Mol. Biol.* **1996**, *257*, 684–699.
- (3) Brems, D.; Alter, L.; Beckage, M.; Chance, R.; DiMarchi, R.; Green, L.; Long, H.; Pekar, A.; Shields, J.; Frank, B. *Protein Eng.* **1992b**, *5*, 527–533.
- (4) Zoete, V.; Meuwly, M.; Karplus, M. *Proteins: Struct. Funct. Bioinf.* **2005**, *61*, 79–93.
- (5) Kristensen, C.; Kjeldsen, T.; Wiberg, F.; SchÄd'ffer, L.; Hach, M.; Havelund, S.; Bass, J.; Steiner, D.; Andersen, A. *J. Biol. Chem.* **1997**, *272*, 12978–12983.
- (6) Dodson, G.; Steiner, D. *Curr. Opin. Struct. Biol.* **1998**, *8*, 189–194.
- (7) Zhang, Y.; Whittingham, J. L.; Turkenburg, J. P.; Dodson, E. J.; Brange, J.; Dodson, G. G. *Acta Crystallogr D Biol Crystallogr.* **2002**, *58*, 186–187.
- (8) Hua, Q.; Shoelson, S.; Kochoyan, M.; Weiss, M. *Nature* **1991**, *354*, 238–241.
- (9) Derewenda, U.; Derewenda, Z.; Dodson, E.; Dodson, G.; Binga, X.; Markussen, J. *J. Mol. Biol.* **1991**, *220*, 425–433.
- (10) Baker, E.; Blundell, T.; Cutfield, J.; Cutfield, S.; Dodson, E.; Dodson, G.; Hodgkin, D.; Hubbard, R.; Isaacs, N.; Reynolds, C. *Philos Trans R Soc Lond B Biol Sci.* **1991**, *319*, 369–456.
- (11) Mirmira, R.; Tager, H. *J. Biol. Chem.* **1989**, *264*, 6349–6354.
- (12) Nakagawa, S.; Tager, H. *Biochem.* **1992**, *31*, 3204–3214.
- (13) Cutfield, J.; Cutfield, S.; Dodson, E.; Dodson, G.; Hodgkin, D.; Reynolds, C. *Hoppe-Seyler's Z. Physiol. Chem.* **1981**, *62*, 755–761.

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

- (14) Ludvigsen, S.; Olsen, H.; Kaarsholm, N. *J. Mol. Biol.* **1998**, *279*, 1–7.
- (15) Nakagawa, S.; Tager, H. *J. Biol. Chem.* **1986**, *261*, 7332–7341.
- (16) Dong, J.; Wan, Z.; Popov, M.; Carey, P.; Weiss, M. *J. Mol. Biol.* **2003**, *330*, 431–432.
- (17) Chen, H.; Shi, M.; Guo, Z. Y.; Tang, Y. H.; Qiao, Z. S.; Liang, Z. H.; Feng, Y. M. *Protein Eng.* **2000**, *13*, 779–782.
- (18) Xu, B.; Hu, S.-Q.; Chu, Y.-C.; Huang, K.; Nakagawa, S.; Whittaker, J.; Katsoyannis, P.; Weiss, M. *Biochem.* **2004**, *43*, 8356–8372.
- (19) Bao, S.-J.; Xie, D.; Zhang, J.; Chang, W.; Liang, D. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2975–2980.
- (20) Shoelson, S.; Lu, Z.; Parlautan, L.; Lynch, C.; Weiss, M. *Biochem.* **1992**, *31*, 1757–1767.
- (21) Zoete, V.; Meuwly, M.; Karplus, M. *J. Mol. Biol.* **2004**, *342*, 913–929.
- (22) Zoete, V.; Meuwly, M. *J. Comp. Chem.* **2006**, *27*, 1843–1857.
- (23) Strazza, S.; Hunter, R.; Walker, E.; Darnall, D. *Arch. Biochem. Biophys.* **1985**, *238*, 30–42.
- (24) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (25) MacKerell, Jr., A. D. et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (26) Brooks, C.; Karplus, M. *J. Chem. Phys.* **1983**, *79*, 6312–6325.
- (27) Bernstein, F.; Koetzle, T.; Williams, G. J.; Jr., E. F. M.; Brice, M.; Rodgers, J.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535–542.
- (28) Berman, H. M. et al. *Acta Cryst. D* **58** 899–907.

5 Computational Characterization of the Insulin Dimerization

- (29) Baker, E. N.; Blundell, T. L.; Cutfield, J. F.; Cutfield, S. M.; Dodson, E. J.; Dodson, G. G.; Hodgkin, D. M. C.; Hubbard, R. E.; Isaacs, N. W.; Reynolds, C. D.; Sakabe, K.; Sakabe, N.; Vijayan, N. *Phil. Trans. R. Soc. Lond. B* **1988**, *319*, 369–456.
- (30) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (31) Gohlke, H.; Kiel, C.; Case, D. A. *J. Mol. Biol.* **2003**, *330*, 891–913.
- (32) Hasel, W.; Hendrickson, T.; Still, W. *Tetrahedron Comput. Methodol.* **1988**, *1*, 103.
- (33) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (34) McQuarrie, D. *Statistical mechanics*; Harper and Row: New York, 1976.
- (35) Kim, D.; Fisher, C.; Baker, D. *J. Mol. Biol.* **2001**, *298*, 971–984.
- (36) L.Serrano.; Kellis, J.; Cann, P.; Matouschek, A.; Fersht, A. *J. Mol. Biol.* **1992**, *224*, 783–804.
- (37) Shoelson, S.; Lu, Z.; Parlautan, L.; Lynch, C. S.; Weiss, M. A. *Biochem.* **1992**, *31*, 1757–1767.
- (38) IV, I.; Nakagawa, S.; Tager, H.; Steiner, D. *Biochem.* **1997**, *36*, 3430–3437.
- (39) Chen, H.; Shi, M.; Guo, Z. Y.; Tang, Y. H.; Qiao, Z. S.; Liang, Z. H.; Feng, Y. *Protein Eng.* **2000**, *13*, 779–782.
- (40) Du, H.; Shi, J.; Cui, D.; Zhang, Y. *Acta Biochim Biophys Sin* **2008**, *40*, 133–139.
- (41) Tidor, B.; Karplus, M. *J. Mol. Biol.* **1994**, *238*, 405–414.

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

Table 3: The contributions to the dimerization free energy at $T = 300$ K.

	$\langle \Delta E_{\text{vdw}} \rangle$	$\langle \Delta E_{\text{ele}} \rangle$	$\langle \Delta G_{\text{ele,desolv}} \rangle$	$\Delta G_{\text{np,desolv}}$	$\langle \Delta G_{\text{bind}}^0 \rangle + \langle \Delta G_{\text{desolv}} \rangle$
wt1	-62.87(4.49)	-102.32(38.74)	125.71(36.99)	-9.48(0.65)	-48.96(6.32)
wt2	-61.83(4.33)	-87.89(39.17)	116.35(42.73)	-9.65(0.78)	-43.01(5.75)
wt3	-58.34(5.03)	-124.27(46.08)	143.45(38.66)	-10.47(0.57)	-49.63(9.37)
wt4	-66.59(4.36)	-114.38(34.28)	146.23(32.58)	-10.28(0.61)	-45.02(6.46)
wt5	-59.45(3.68)	-121.51(29.98)	145.68(28.10)	-9.30(0.46)	-44.58(5.82)
wt6	-61.36(3.91)	-90.87(26.06)	119.87(24.76)	-9.55(0.55)	-41.91(5.62)
wt7	-62.85(3.40)	-125.20(42.81)	149.39(44.43)	-9.25(0.46)	-47.90(5.02)
wt8	-60.93(4.02)	-138.28(48.70)	161.65(47.78)	-9.62(0.49)	-47.17(6.23)
gly1	-62.36(4.08)	-104.76(25.39)	130.29(24.99)	-10.71(0.47)	-47.54(6.16)
gly2	-53.32(4.12)	-125.17(27.38)	151.12(22.78)	-9.64(0.38)	-37.01(7.38)
gly3	-58.09(4.93)	-117.11(31.58)	141.39(31.27)	-10.05(0.43)	-43.86(6.53)
gly4	-51.11(6.17)	-115.90(27.53)	147.56(22.56)	-9.89(0.47)	-29.34(11.69)
gly5	-56.11(5.94)	-115.43(32.93)	140.98(31.20)	-9.90(0.59)	-40.47(7.88)
gly6	-53.39(4.35)	-107.84(40.75)	132.83(38.36)	-9.76(0.49)	-38.16(7.27)
gly7	-64.08(6.42)	-143.78(37.88)	170.47(34.54)	-10.65(0.47)	-48.04(8.34)
gly8	-55.28(4.50)	-136.76(37.17)	176.94(35.75)	-9.80(0.50)	-24.90(6.81)
gly9	-57.52(3.92)	-98.19(31.22)	145.35(27.56)	-9.48(0.48)	-19.84(10.31)
ala1	-54.16(3.82)	-87.25(37.41)	116.56(36.76)	-9.65(0.53)	-34.50(6.59)
ala2	-61.34(5.91)	-114.52(31.64)	142.50(31.27)	-10.29(0.77)	-43.64(7.36)
ala3	-58.00(3.96)	-26.04(23.07)	53.51(20.69)	-9.92(0.37)	-40.45(6.11)
ala4	-53.76(4.66)	-146.85(30.00)	167.65(28.81)	-10.15(0.40)	-43.10(7.69)
ala5	-56.85(4.19)	-91.79(35.53)	120.17(34.27)	-9.89(0.42)	-38.36(7.49)
ala6	-58.52(5.49)	-121.13(34.93)	150.53(31.66)	-10.37(0.49)	-39.49(7.97)
ala7	-56.20(3.99)	-128.37(26.39)	149.69(26.18)	-9.74(0.55)	-44.62(6.51)
ala8	-55.04(3.96)	-112.98(32.67)	141.54(30.39)	-9.86(0.39)	-36.34(6.29)
d-ala1	-60.39(5.06)	-104.85(33.88)	142.77(30.85)	-10.54(0.48)	-33.00(9.47)
d-ala2	-59.20(4.41)	-98.69(33.85)	128.59(31.52)	-10.87(0.48)	-40.18(7.07)
d-ala3	-42.83(5.26)	-136.84(45.45)	167.94(41.11)	-8.21(0.74)	-19.95(8.68)
d-ala4	-58.63(4.25)	-90.45(27.05)	135.90(24.50)	-10.47(0.47)	-23.64(8.10)
d-ala5	-45.51(5.82)	-91.32(31.90)	127.91(31.78)	-8.54(0.74)	-17.46(10.07)
des-b251	-42.19(7.40)	-35.89(23.36)	73.77(26.39)	-7.37(1.17)	-11.69(8.09)
des-b252	-48.57(4.45)	-47.82(31.74)	84.51(29.39)	-8.28(0.59)	-20.16(7.11)
des-b253	-40.76(5.03)	-26.01(18.79)	51.12(18.38)	-7.39(0.45)	-23.03(5.50)
des-b254	-29.94(5.42)	-67.34(29.45)	88.82(28.31)	-6.96(0.76)	-15.43(6.76)
des-b255	-44.47(5.21)	-56.15(32.19)	92.43(29.57)	-7.99(0.61)	-16.19(6.47)
des-b256	-56.27(4.32)	-46.75(49.52)	82.96(49.44)	-9.27(0.58)	-29.33(6.26)
des-b257	-44.52(5.82)	-35.70(21.52)	68.84(23.67)	-7.85(0.80)	-19.24(8.19)
des-b258	-44.43(3.68)	-30.09(18.94)	66.42(19.40)	-7.48(0.45)	-15.58(6.65)
b25turn24	-59.21(4.79)	-63.19(26.05)	107.31(27.20)	-8.94(0.57)	-24.03(7.60)

41

Influence of Mutations at Position B24 on the Stability of Insulin Dimer

Lixian Zhang, and Markus Meuwly*

Department of Chemistry, University of Basel, Basel, Switzerland

E-mail: m.meuwly@unibas.ch

*To whom correspondence should be addressed

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

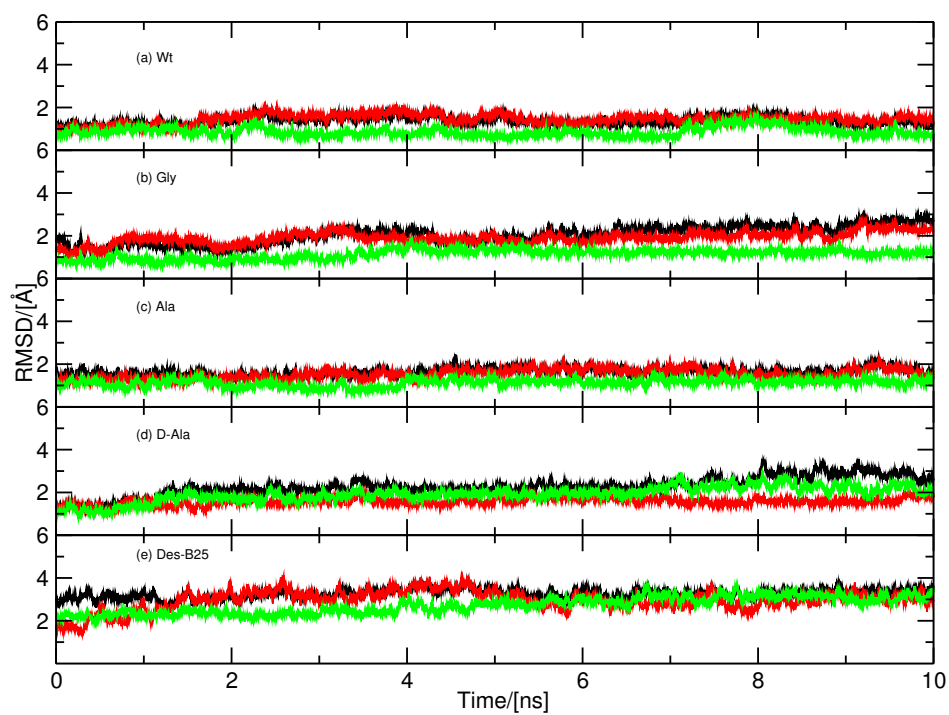


Figure 1: RMSD along the trajectories for the backbone atoms of the entire protein. (a) Wild-type (b) B24Gly (c) B24Ala (d) B24D-Ala (e) Des-B25.

5 Computational Characterization of the Insulin Dimerization

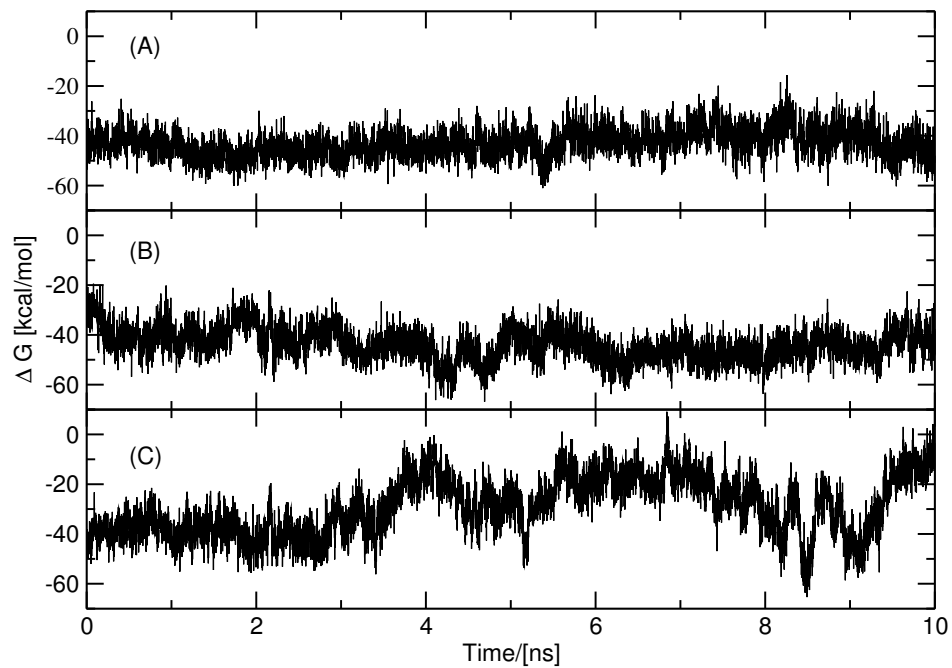


Figure 2: Dimerization Free Energy of wildtype insulin dimer -panel (A) and B24Gly mutant insulin dimer -panel (C) and (D).

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

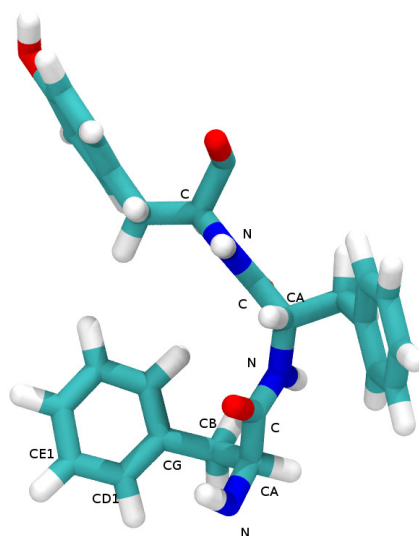


Figure 3: Structure of residue 24 to 26 of chain B in wild-type insulin dimer.

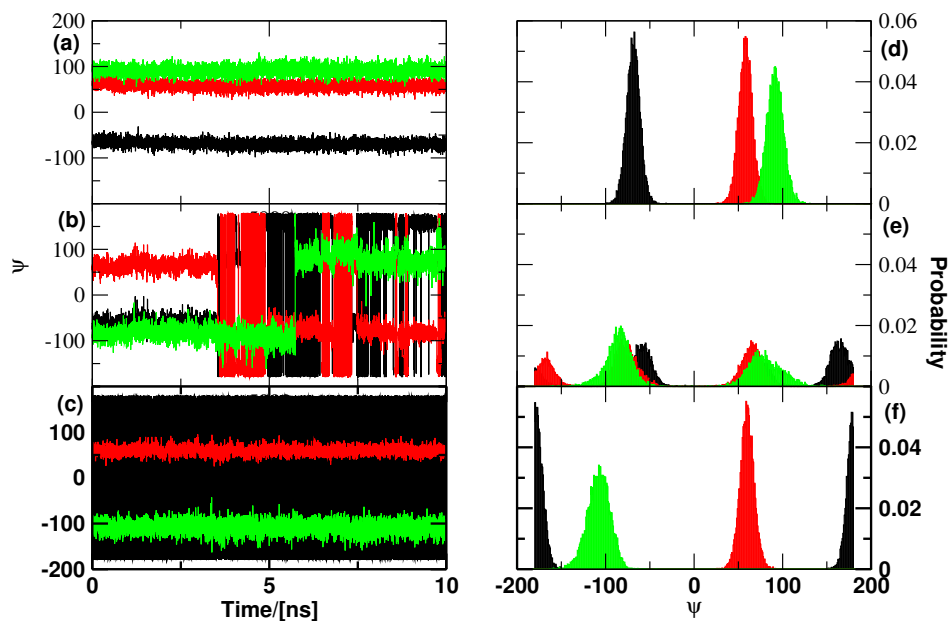


Figure 4: Dihedral Angel of sidechain of residue 24 to 26 of wild-type insulin dimer. (a) -(c) The dihedral angle of side-chain of residue 24 to residue 26 in wild-type insulin dimer. Black is the dihedral of N-CA-CB-CG, Red is the dihedral of C-CA-CB-CG, Green is the dihedral of CA-CB-CG-CD1. (a) residue 24; (b) residue 25; (c) residue 26. (d) - (f) the histograms of the dihedral angles of side-chain of residue 24 to residue 26 in wild-type insulin-dimer: (d) residue 24; (e) residue 25; (f) residue 26.

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

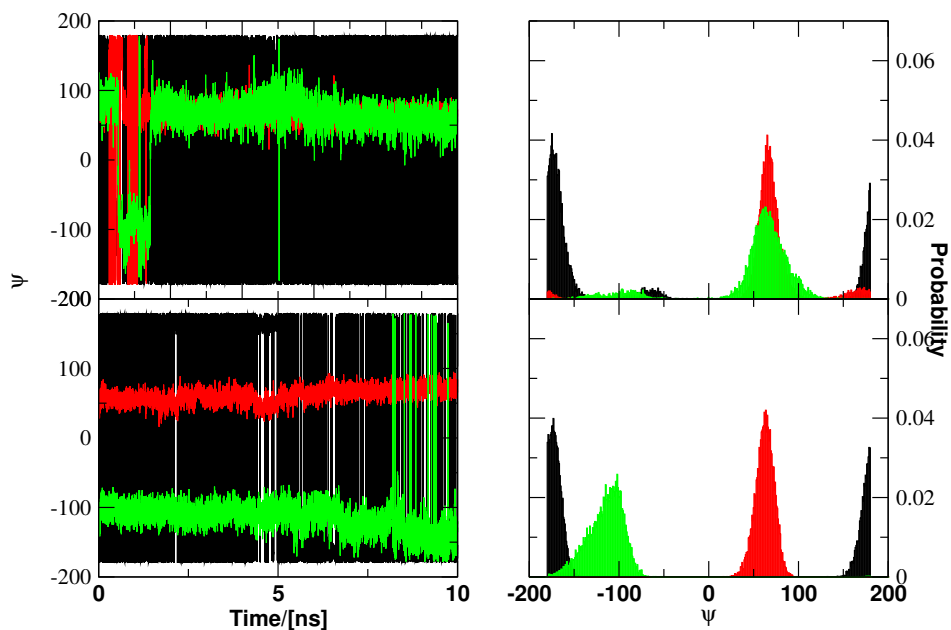


Figure 5: (a) -(b) The dihedral angle of side-chain of residue 25 to residue 26 in FB24Gly mutant insulin dimer. Black is the dihedral of N-CA-CB-CG, Red is the dihedral of C-CA-CB-CG, Green is the dihedral of CA-CB-CG-CD1. (a) residue 25; (b) residue 26. (c) - (d) the histograms of the dihedral angles of side-chain of residue 25 to residue 26 in FB24Gly insulin dimer: (c) residue 25; (d) residue 26.

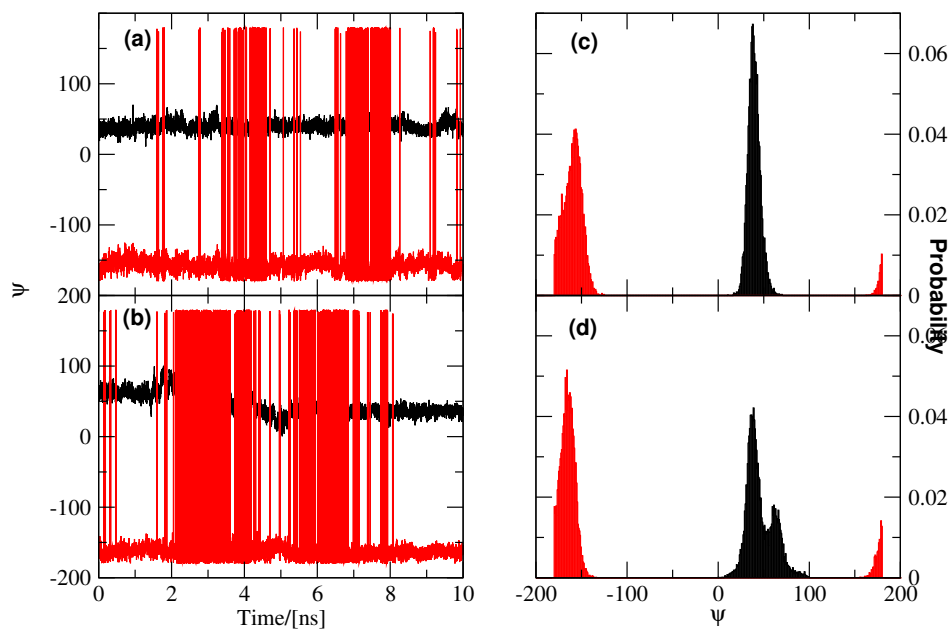


Figure 6: Pseudodihedral angle (four successive CA atoms). black is the pseudodihedral for β turn-(residue 20 to 23) in chain B; red is the for residue 24 to 26. (a) wildtype insulin; (b) FB24Gly insulin; (c) Histogram for wildtype insulin; (d) Histogram for FB24Gly insulin.

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

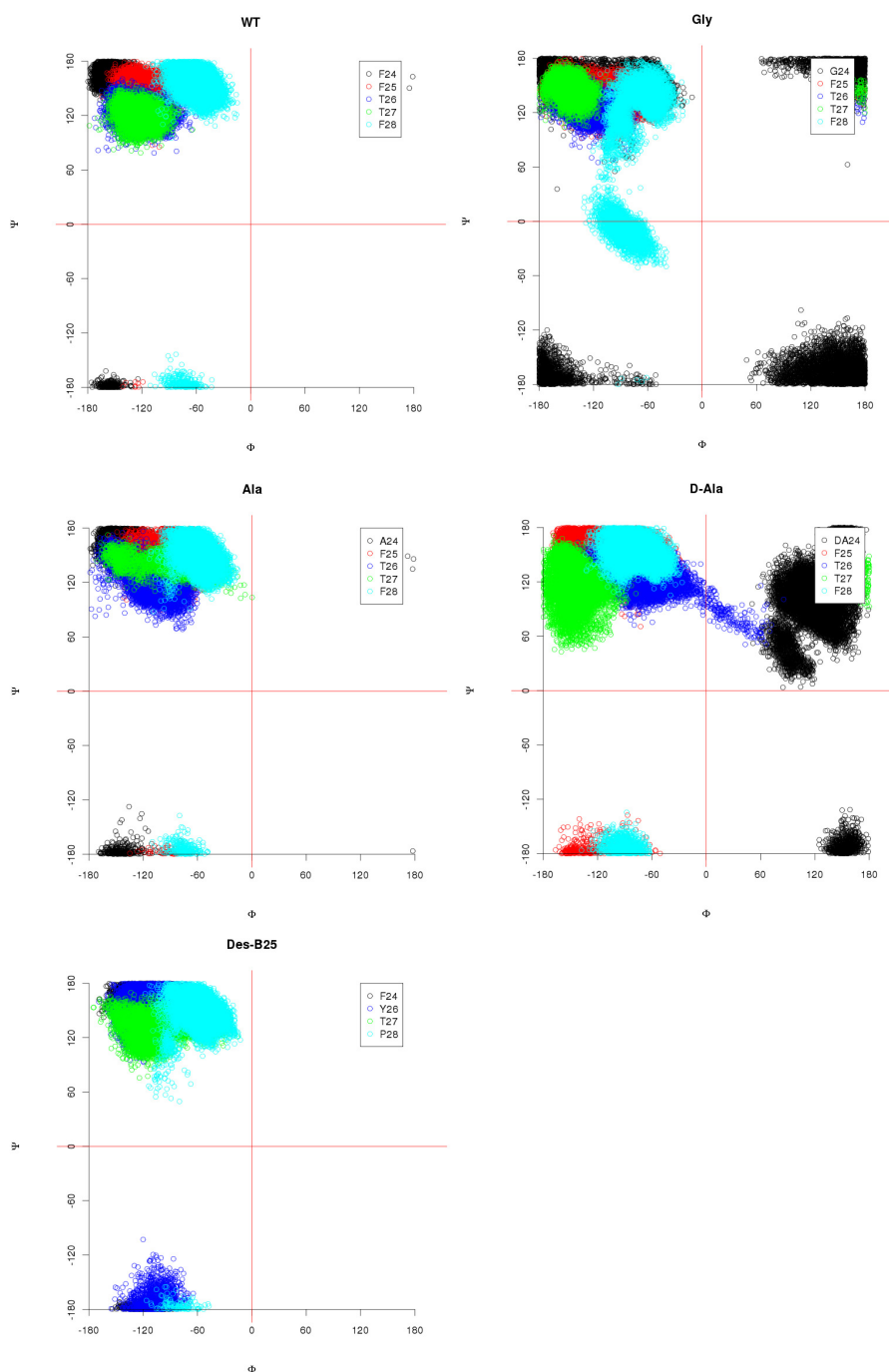


Figure 7: Ramachandran plot of β -sheet (residue 24 to 28) of chain B of (a) wildtype insulin dimer (b) FB24Gly (c) FB24Ala (d) FB24D-Ala (e) Des-B25 insulin dimer during 10 ns trajectory length. One dot per ps, 10000 dots for each residue. Black is for B24, red for B25Phe, blue for B26Tyr, green for B27Tyr, and cyan for B28Pro.

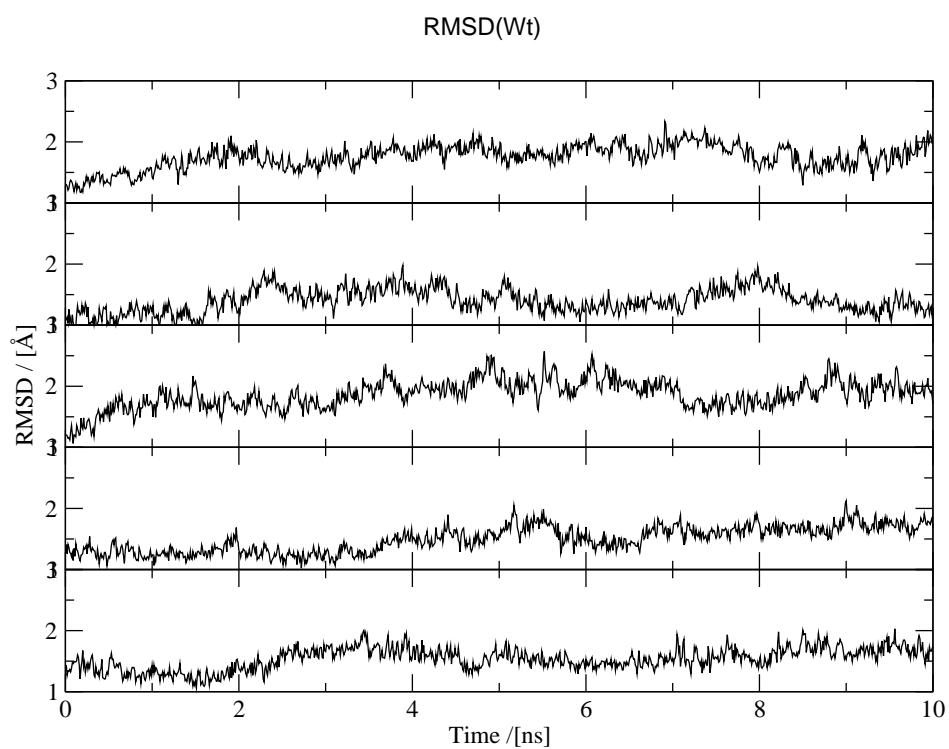


Figure 8: RMSD of Wt insulin

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

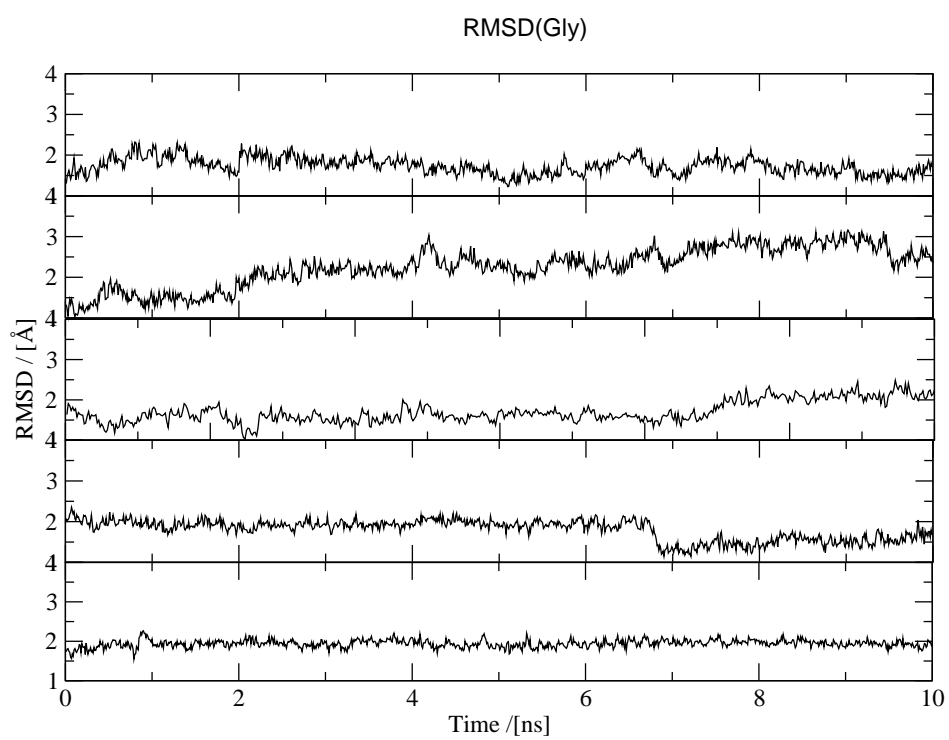


Figure 9: RMSD of Gly insulin

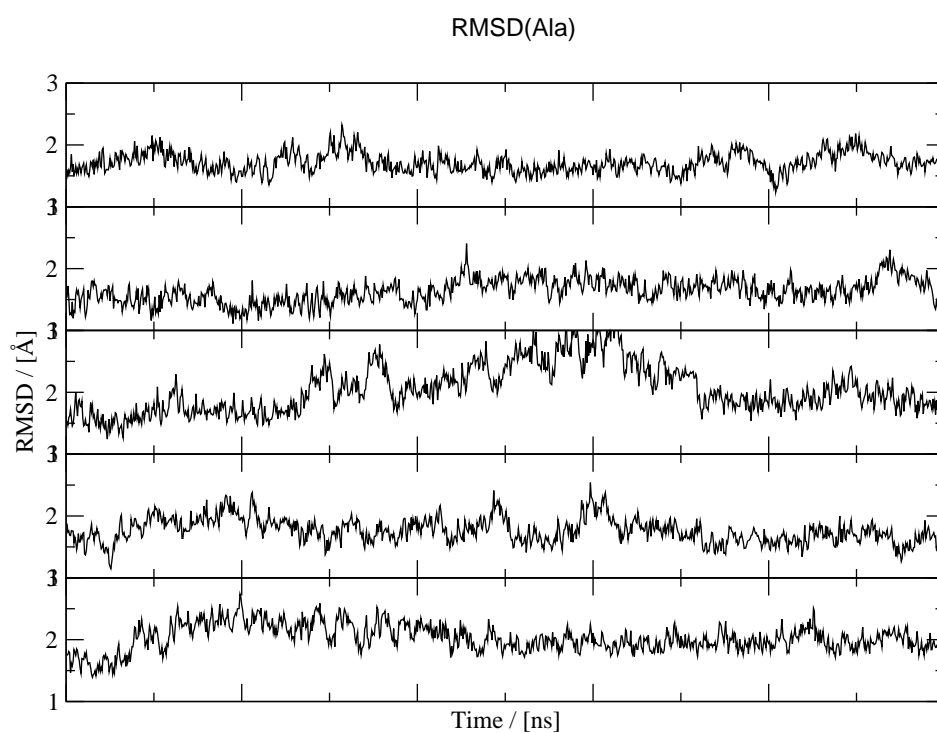


Figure 10: RMSD of Ala insulin

5.1 Influence of Mutations at Position B24 on the Stability of the Insulin Dimer

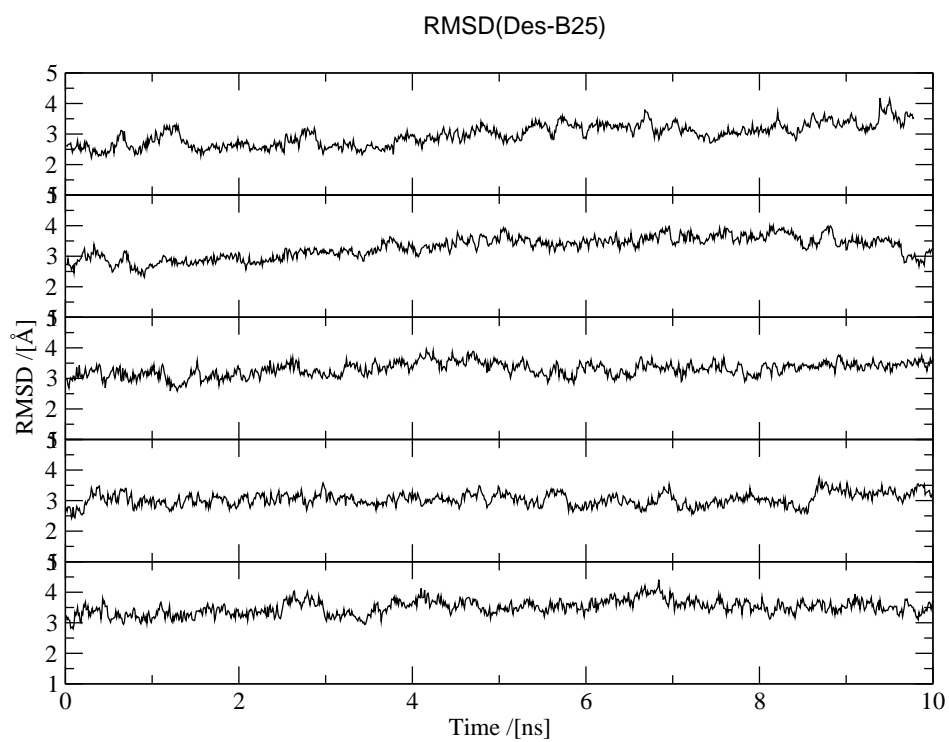


Figure 11: RMSD of Des-b25 insulin

6 Conclusion and Outlook

In this thesis some important dimerization processes in biological systems were addressed. Quantum mechanical modelling methods (Density Functional Theory), together with atomic molecular mechanism MD simulations were employed to characterize the dimerization processes for *c*-di-GMP, endo-S-*c*-di-GMP, PleD, and insulin systems. Structural changes and binding free energies were analyzed and very good correlations to experimental data were obtained.

The studies of *c*-di-GMP and its analogue endo-S-CDG in solution and binding with proteins in this thesis (Chapter 4) improved our fundamental understanding of this bacterial signaling molecule. It was demonstrated, that dimeric *c*-di-GMP is only marginally stabilized even in high concentration. We revealed that changing the phosphodiester backbone of *c*-di-GMP decreases the aggregation ability to support the designing of non-aggregation-prone *c*-di-GMP like molecules. More investigation in this field is still required for better understanding of the mechanism and for the development of anti-film or anti-virulence drugs. And several aspects of this signalling molecule remain far from being understood, such as the enzymatic process of PleD (biosynthesis of *c*-di-GMP upon PleD dimerization), and the adaptor proteins that bind to *c*-di-GMP and transmit the binding event into a processes leading to biofilm formation.

In Chapter 5 we investigated the hormone insulin, which upon dimerization leads to

6 Conclusion and Outlook

diabetes. The structural changes between monomers and dimers of the wildtype insulin as well as several mutants (mutating the dimerization key residue phenylalanine B24 to alanine, glycine, D-alanine, and Des-(Phe25)) were successfully characterized. Moreover, the dimerization free energies of insulin and other analogues were obtained and indicated the importance of residue phenylalanine B24 for insulin dimerization. Our work provided an insight for designing analogues of human insulin and thus a therapy for diabetes.

In this thesis MM-GBSA was employed for calculating the dimerization free energy of systems. MM-PB/GBSA methods are valuable tools used in computer-aided drug design. But as with any other method, they have some limitations that need to be considered when they are employed, such as they ignore protein-ligand/protein interactions bridged by water molecules, as we found in this thesis. More research is required to ultimately develop a fast, accurate and reliable method to calculate the binding affinity of ligand/protein to protein for utilisation in drug discovery.

Bibliography

- [1] C. J. Tsai, S. L. Lin, H. J. Wolfson and R. Nussinov, *Crit. Rev. Biochem. Mol. Biol.*, 1996, **31**, 127–152.
- [2] C. Pace, B. Shirley, M. McNutt and K. Gajiwala, *FASEB J.*, 1996, 75–83.
- [3] O. Keskin, N. Tuncbag and A. GURSOY, *Curr. Pharm. Biotechnol.*, 2008, **9**, 67–76.
- [4] C. Chan, R. Paul, D. Samoray, N. C. Amiot, B. Giese, U. Jenal and T. Schirmer, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 17084–17089.
- [5] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, *Arch. Biochem. Biophys.*, 1978, **185**, 584–591.
- [6] B. Shaanan, a. M. Gronenborn, G. H. Cohen, G. L. Gilliland, B. Veerapandian, D. R. Davies and G. M. Clore, *Science*, 1992, **257**, 961–964.
- [7] D. W. Hoffman, C. S. Cameron, C. Davies, S. W. White and V. Ramakrishnan, *J. Mol. Biol.*, 1996, **264**, 1058–1071.
- [8] S. W. Muchmore, M. Sattler, H. Liang, R. P. Meadows, J. E. Harlan, H. S. Yoon, D. Nettlesheim, B. S. Chang, C. B. Thompson, S. L. Wong, S. L. Ng and S. W. Fesik, *Nature*, 1996, **381**, 335–341.
- [9] H.-X. Zhou and M. K. Gilson, *Chem. Rev.*, 2009, **109**, 4092–4107.

Bibliography

- [10] M. Guharoy, A. Pal, M. Dasgupta and P. Chakrabarti, *J. Struct. Funct. Genomics*, 2011, **12**, 33–41.
- [11] W. R. Pearson and D. J. Lipman, *Proc. Natl. Acad. Sci. U. S. A.*, 1988, **85**, 2444–2448.
- [12] S. Altschul, *Nucleic Acids Res.*, 1997, **25**, 3389–3402.
- [13] G. J. Barton, *Methods Enzymol.*, 1990, **183**, 403–428.
- [14] M. Suyama, Y. Matsuo and K. Nishikawa, *J. Mol. Evol.*, 1997, **44**, S163–S173.
- [15] C. Notredame, D. G. Higgins and J. Heringa, *J. Mol. Biol.*, 2000, **302**, 205–217.
- [16] C. B. Do, M. S. P. Mahabhashyam, M. Brudno and S. Batzoglou, *Genome Res.*, 2005, **15**, 330–340.
- [17] K. Arnold, L. Bordoli, J. Kopp and T. Schwede, *Bioinformatics*, 2006, **22**, 195–201.
- [18] T. Schwede, J. Kopp, N. Guex and M. C. Peitsch, *Nucleic Acids Res.*, 2003, **31**, 3381–3385.
- [19] M. S. Johnson, N. Srinivasan, R. Sowdhamini and T. L. Blundell, *Crit. Rev. Biochem. Mol. Biol.*, 1994, **29**, 1–68.
- [20] A.-S. Yang and B. Honig, *Proteins: Struct., Funct., Bioinf.*, 1999, **37**, 66–72.
- [21] H. Li, R. Tejero, D. Monleon, D. Bassolino-Klimas, C. Abate-Shen, R. E. Brucoleri and G. T. Montelione, *Protein Sci.*, 1997, **6**, 956–970.
- [22] M. J. Bower, F. E. Cohen and R. L. Dunbrack, *J. Mol. Biol.*, 1997, **267**, 1268–1282.
- [23] A. Sali and T. L. Blundell, *J. Mol. Biol.*, 1993, **234**, 779–815.
- [24] R. A. Laskowski, M. W. MacArthur, D. S. Moss and J. M. Thornton, *J. Appl. Crystallogr.*, 1993, **26**, 283–291.

- [25] G. Vriend, *J. Mol. Graphics*, 1990, **8**, 52–56.
- [26] D. Eisenberg, R. Lüthy and J. U. Bowie, *Methods Enzymol.*, 1997, **277**, 396–404.
- [27] F. Melo and E. Feytmans, *J. Mol. Biol.*, 1998, **277**, 1141–1152.
- [28] M. T. Mas, K. C. Smith, D. L. Yarmush, K. Aisaka and R. M. Fine, *Proteins: Struct., Funct., Bioinf.*, 1992, **14**, 483–498.
- [29] I. T. Weber, *Proteins: Struct., Funct., Bioinf.*, 1990, **7**, 172–184.
- [30] P. V. Sahasrabudhe, R. Tejero, S. Kitao, Y. Furuichi and G. T. Montelione, *Proteins: Struct., Funct., Bioinf.*, 1998, **33**, 558–566.
- [31] J. Monod, J. Wyman and J.-P. Changeux, *J. Mol. Biol.*, 1965, **12**, 88–118.
- [32] D. E. Koshland, G. Némethy and D. Filmer, *Biochemistry*, 1966, **5**, 365–85.
- [33] C.-J. Tsai, A. del Sol and R. Nussinov, *J. Mol. Biol.*, 2008, **378**, 1–11.
- [34] C.-J. Tsai, A. Del Sol and R. Nussinov, *Mol. BioSyst.*, 2009, **5**, 207–216.
- [35] Q. Cui and M. Karplus, *Protein Sci.*, 2008, **17**, 1295–1307.
- [36] F. F.-F. Schmid, *Ph.D. thesis*, Universität Basel, 2009.
- [37] K. Gunasekaran, B. Ma and R. Nussinov, *Proteins: Struct., Funct., Bioinf.*, 2004, **57**, 433–443.
- [38] M. S. Formanek, L. Ma and Q. Cui, *Proteins: Struct., Funct., Bioinf.*, 2006, **63**, 846–867.
- [39] N. M. Goodey and S. J. Benkovic, *Nat. Chem. Biol.*, 2008, **4**, 474–482.
- [40] S. Jones and J. M. Thornton, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 13–20.
- [41] P. M. Kim, L. J. Lu, Y. Xia and M. B. Gerstein, *Science*, 2006, **314**, 1938–1941.

Bibliography

- [42] J. D. Klemm, S. L. Schreiber and G. R. Crabtree, *Annu. Rev. Immunol.*, 1998, **16**, 569–592.
- [43] N. J. Marianayagam, M. Sunde and J. M. Matthews, *Trends Biochem. Sci.*, 2004, **29**, 618–625.
- [44] D. Xu, C. J. Tsai and R. Nussinov, *Protein Sci.*, 1998, **7**, 533–544.
- [45] C. J. Tsai and R. Nussinov, *Protein Sci.*, 1997, **6**, 1426–1437.
- [46] V. Zoete, M. Meuwly and M. Karplus, *Proteins: Struct., Funct., Bioinf.*, 2005, **61**, 79–93.
- [47] D. La, M. Kong, W. Hoffman, Y. I. Choi and D. Kihara, *Proteins*, 2013, **81**, 805–518.
- [48] I. M. Nooren and J. M. Thornton, *J. Mol. Biol.*, 2003, **325**, 991–1018.
- [49] L. Zhao, S. C. H. Hoi, L. Wong, T. Hamp and J. Li, *PloS one*, 2012, **7**, e50821.
- [50] J. a. Wells and C. L. McClendon, *Nature*, 2007, **450**, 1001–1009.
- [51] R. P. Bahadur, P. Chakrabarti, F. Rodier and J. Janin, *Proteins*, 2003, **53**, 708–719.
- [52] B. Honig and A. Nicholls, *Science*, 1995, **268**, 1144–1149.
- [53] M. Neves-Petersen and S. Petersen, *Biotech. Ann. Rev.*, 2003, **9**, 315–395.
- [54] P. Kukić and J. E. Nielsen, *Future Med.Chem.*, 2010, **2**, 647–666.
- [55] C. L. Vizcarra and S. L. Mayo, *Curr. Opin. Chem. Biol.*, 2005, **9**, 622–626.
- [56] Y. Marechal, *The Hydrogen Bond and the Water Molecule: The Physics and Chemistry of Water, Aqueous and Bio-Media*, Elsevier B.V., 1st edn., 2007.
- [57] C. L. Tien, A. Majumdar and F. M. Gerner, *Microscale Energy Transfer*, CRC Press, 1997.

- [58] A. Szilagyi, J. Kardos, S. Osvath, L. Barna and P. Zavodszky, in *Handbook of Neurochemistry and Molecular Neurobiology: Neural Protein Metabolism and Function*, ed. A. Lajtha and N. Banik, Springer, 2007, ch. Protein folding, pp. 303–344.
- [59] W. J. Wedemeyer, E. Welker, M. Narayan and H. A. Scheraga, *Biochemistry*, 2000, **39**, 4207–4216.
- [60] V. Zoete, M. Meuwly and M. Karplus, *J. Mol. Biol.*, 2004, **342**, 913–929.
- [61] S. Lise, D. Buchan, M. Pontil and D. T. Jones, *PloS one*, 2011, **6**, e16774.
- [62] C. Chothia and J. Janin, *Nature*, 1975, **256**, 705–708.
- [63] W. L. DeLano, *Science*, 2000, **287**, 1279–1283.
- [64] T. Kortemme and D. Baker, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 14116–14121.
- [65] R. Guerois, J. E. Nielsen and L. Serrano, *J. Mol. Biol.*, 2002, **320**, 369–387.
- [66] E. Freese, *Proc. Natl. Acad. Sci. U. S. A.*, 1959, **45**, 622–633.
- [67] A. A. Pakula, V. B. Young and R. T. Sauer, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 8829–8833.
- [68] H. Erlandsen and R. C. Stevens, *Mol. Genet. Metab.*, 1999, **68**, 103–125.
- [69] M. Hochstrasser, *Annu. Rev. Genet.*, 1996, **30**, 405–439.
- [70] D. M. Taverna and R. A. Goldstein, *J. Mol. Biol.*, 2002, **315**, 479–484.
- [71] C. B. Anfinsen, *Science*, 1973, **181**, 223–230.
- [72] J. N. Onuchic, Z. Luthey-Schulten and P. G. Wolynes, *Annu. Rev. Phys. Chem.*, 1997, **48**, 545–600.
- [73] M. S. Cheung, L. L. Chavez and J. N. Onuchic, *Polymer*, 2004, **45**, 547–555.

Bibliography

- [74] J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten and N. D. Socci, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 3626–3360.
- [75] G. R. Bowman, V. A. Voelz and V. S. Pande, *Curr. Opin. Struct. Biol.*, 2011, **21**, 4–11.
- [76] D. J. Selkoe, *Nature*, 2003, **426**, 900–904.
- [77] J. L. Yap, J. Chauhan, K.-y. Jung, L. Chen, E. V. Prochownik and S. Fletcher, *Med. Chem. Commun.*, 2012, **3**, 541–551.
- [78] K. Nakanishi and M. Kikuchi, *J. Phys. Soc. Jpn.*, 2006, **75**, 064803.
- [79] F. B. Sheinerman, R. Norel and B. Honig, *Curr. Opin. Struct. Biol.*, 2000, **10**, 153–159.
- [80] A. H. Elcock, R. R. Gabdouliline, R. C. Wade and J. a. McCammon, *J. Mol. Biol.*, 1999, **291**, 149–462.
- [81] C. J. Camacho, Z. Weng, S. Vajda and C. DeLisi, *Biophys. J.*, 1999, **76**, 1166–1178.
- [82] Z. S. Hendsch and B. Tidor, *Protein Sci.*, 1999, **8**, 1381–1392.
- [83] C. J. Tsai, S. L. Lin, H. J. Wolfson and R. Nussinov, *Protein Sci.*, 1997, **6**, 53–64.
- [84] T. A. Larsen, A. J. Olson and D. S. Goodsell, *Structure*, 1998, **6**, 421–427.
- [85] S. Jones, *Adv. Exp. Med. Biol.*, 2012, **747**, 42–54.
- [86] A. Quist, I. Doudevski, H. Lin, R. Azimova, D. Ng, B. Frangione, B. Kagan, J. Ghiso and R. Lal, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 10427–10432.
- [87] C. M. Dobson, *Nature*, 2003, **426**, 884–890.
- [88] W. E. Balch, R. I. Morimoto, A. Dillin and J. W. Kelly, *Science*, 2008, **319**, 916–919.

- [89] J. Wang, J. Zhou, G. P. Donaldson, S. Nakayama, L. Yan, Y.-f. Lam, V. T. Lee and H. O. Sintim, *J. Am. Chem. Soc.*, 2011, **133**, 9320–9330.
- [90] T. Darden, L. Perera, L. Li and L. Pedersen, *Structure*, 1999, **7**, R55–R60.
- [91] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta and P. Weiner, *J. Am. Chem. Soc.*, 1984, **106**, 765–784.
- [92] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217.
- [93] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J. Comput. Chem.*, 2009, **30**, 1545–614.
- [94] W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1988, **110**, 1657–1666.
- [95] J. Hermans, H. J. C. Berendsen, W. F. Van Gunsteren and J. P. M. Postma, *Biopolymers*, 1984, **23**, 1513–1518.
- [96] M. Christen, P. H. Hünenberger, D. Bakowies, R. Baron, R. Bürgi, D. P. Geerke, T. N. Heinz, M. A. Kastenholtz, V. Kräutler, C. Oostenbrink, C. Peter, D. Trzesniak and W. F. van Gunsteren, *J. Comput. Chem.*, 2005, **26**, 1719–1751.
- [97] M. Buck, S. Bouguet-Bonnet, R. W. Pastor and A. D. MacKerell, *Biophys. J.*, 2006, **90**, L36–L38.
- [98] A. D. Mackerell, M. Feig and C. L. Brooks, *J. Comput. Chem.*, 2004, **25**, 1400–1415.

Bibliography

- [99] P. Maragakis, K. Lindorff-Larsen, M. P. Eastwood, R. O. Dror, J. L. Klepeis, I. T. Arkin, M. O. Jensen, H. Xu, N. Trbovic, R. A. Friesner, A. G. P. Iii and D. E. Shaw, *J. Phys. Chem. B*, 2008, **112**, 6155–6158.
- [100] M. W. van der Kamp, K. E. Shaw, C. J. Woods and A. J. Mulholland, *J. R. Soc. Interface*, 2008, **5 Suppl 3**, S173–S190.
- [101] D. E. Shaw, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. Grossman, C. R. Ho, D. J. Lerardi, I. Kolossváry, J. L. Klepeis, T. Layman, C. McLeavey, M. M. Deneroff, M. A. Moraes, R. Mueller, E. C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, S. C. Wang, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson and K. J. Bowers, *Comm. ACM*, 2008, **51**, 91.
- [102] W. L. Jorgensen, *J. Chem. Theory Comput.*, 2007, **3**, 1877–1877.
- [103] P. Cieplak, F.-Y. Dupradeau, Y. Duan and J. Wang, *J. Phys. Condens. Matter*, 2009, **21**, 333102.
- [104] J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
- [105] J. Yang, Y. Wang and Y. Chen, *J. Comput. Phys.*, 2007, **221**, 799–804.
- [106] W. Liu, B. Schmidt, G. Voss and W. Müller-Wittig, *Comput. Phys. Commun.*, 2008, **179**, 634–641.
- [107] M. S. Friedrichs, P. Eastman, V. Vaidyanathan, M. Houston, S. Legrand, A. L. Beberg, D. L. Ensign, C. M. Bruns and V. S. Pande, *J. Comput. Chem.*, 2009, **30**, 864–872.
- [108] P. Hut and J. Makino, *Science*, 1999, **283**, 501–505.
- [109] P. Hohenberg, *Phys. Rev.*, 1964, **136**, B864–B871.

- [110] W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133–A1138.
- [111] C. Møller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618–622.
- [112] J. D. Durrant and J. A. McCammon, *BMC biology*, 2011, **9**, 71.
- [113] M. Karplus and J. A. McCammon, *Nat. Struct. Biol.*, 2002, **9**, 646–652.
- [114] W. F. van Gunsteren and H. J. C. Berendsen, *Angew. Chem. Int. Ed.*, 1990, **29**, 992–1023.
- [115] L. Hernández de la Peña, R. van Zon, J. Schofield and S. B. Opps, *J. Chem. Phys.*, 2007, **126**, 074106.
- [116] F. Ding, J. M. Borreguero, S. V. Buldyrey, H. E. Stanley and N. V. Dokholyan, *Proteins: Struct., Funct., Bioinf.*, 2003, **53**, 220–228.
- [117] A. Voegler Smith and C. K. Hall, *Proteins: Struct., Funct., Bioinf.*, 2001, **44**, 344–360.
- [118] X. Periole, A. M. Knepp, T. P. Sakmar, S. J. Marrink and T. Huber, *J. Am. Chem. Soc.*, 2012, **134**, 10959–10965.
- [119] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries, *J. Phys. Chem. B*, 2007, **111**, 7812–7824.
- [120] S. J. Marrink, A. H. de Vries and A. E. Mark, *J. Phys. Chem. B*, 2004, **108**, 750–760.
- [121] S. J. Maerkl and S. R. Quake, *Science*, 2007, **315**, 233–237.
- [122] R. T. Strachan, G. Ferrara and B. L. Roth, *Drug discovery today*, 2006, **11**, 708–716.
- [123] M. Allen, D. Hall, B. Collins and K. Moore, *J. Biomol. Screening*, 2002, **7**, 35–44.

Bibliography

- [124] K. H. Bleicher, H.-J. Böhm, K. Müller and A. I. Alanine, *Nat. Rev. Drug Discovery*, 2003, **2**, 369–378.
- [125] L. Salwinski and D. Eisenberg, *Curr. Opin. Struct. Biol.*, 2003, **13**, 377–382.
- [126] D. B. Kitchen, H. Decornez, J. R. Furr and J. Bajorath, *Nat. Rev. Drug Discovery*, 2004, **3**, 935–949.
- [127] B. Mukesh and K. Rakesh, *Int. J. Res. Ayur. Pharm.*, 2011, **2**, 1746–1751.
- [128] N. Brooijmans and I. D. Kuntz, *Annu. Rev. Biophys. Biomol. Struct.*, 2003, **32**, 335–373.
- [129] D. S. Goodsell, H. Lauble, C. D. Stout and A. J. Olson, *Proteins: Struct., Funct., Bioinf.*, 1993, **17**, 1–10.
- [130] G. Jones, P. Willett and R. C. Glen, *J. Mol. Biol.*, 1995, **245**, 43–53.
- [131] Chemical Computing Group, *MOE Dock*, 2003.
- [132] D. R. Westhead, D. E. Clark and C. W. Murray, *J. Comput.-Aided Mol. Des.*, 1997, **11**, 209–228.
- [133] T. J. Ewing, S. Makino, A. G. Skillman and I. D. Kuntz, *J. Comput.-Aided Mol. Des.*, 2001, **15**, 411–428.
- [134] M. Rarey, B. Kramer, T. Lengauer and G. Klebe, *J. Mol. Biol.*, 1996, **261**, 470–489.
- [135] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *J. Med. Chem.*, 2004, **47**, 1739–1749.
- [136] W. Welch, J. Ruppert and A. N. Jain, *Chemistry & Biology*, 1996, **3**, 449–462.
- [137] S. K. Kearsley, D. J. Underwood, R. P. Sheridan and M. D. Miller, *J. Comput.-Aided Mol. Des.*, 1994, **8**, 565–582.

- [138] M. K. Gilson and H.-X. Zhou, *Annu. Rev. Biophys. Biomol. Struct.*, 2007, **36**, 21–42.
- [139] J. Wang and G. Verkhivker, *Phys. Rev. Lett.*, 2003, **90**, 188101.
- [140] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*, Wiley, 2nd edn., 1985, p. 493.
- [141] N. Froloff, A. Windemuth and B. Honig, *Protein Sci.*, 1997, **6**, 1293–301.
- [142] J. Aqvist, C. Medina and J. E. Samuelsson, *Protein Eng.*, 1994, **7**, 385–391.
- [143] R. D. Head, M. L. Smythe, T. I. Oprea, C. L. Waller, S. M. Green and G. R. Marshall, *J. Am. Chem. Soc.*, 1996, **118**, 3959–3969.
- [144] M. D. Paulsen and R. L. Ornstein, *Protein Eng.*, 1996, **9**, 567–571.
- [145] V. Helms and R. C. Wade, *J. Am. Chem. Soc.*, 1998, **120**, 2710–2713.
- [146] C. Oostenbrink and W. F. van Gunsteren, *Proteins: Struct., Funct., Bioinf.*, 2004, **54**, 237–246.
- [147] B. G. Rao, E. E. Kim and M. A. Murcko, *J. Comput.-Aided Mol. Des.*, 1996, **10**, 23–30.
- [148] B. Tidor and M. Karplus, *J. Mol. Biol.*, 1994, **238**, 405–414.
- [149] H. P. Erickson, *J. Mol. Biol.*, 1989, **206**, 465–474.
- [150] M. I. Page and W. P. Jencks, *Proc Natl Acad Sci U S A.*, 1971, **68**, 1678–1683.
- [151] I. Z. Steinberg and H. A. Scheraga, *J. Biol. Chem.*, 1963, **238**, 172–181.
- [152] H. Gohlke, C. Kiel and D. a. Case, *J. Mol. Biol.*, 2003, **330**, 891–913.
- [153] H. B. Thorsteinsdottir, T. Schwede, V. Zoete and M. Meuwly, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 407–423.

Bibliography

- [154] T. Hou, J. Wang, Y. Li and W. Wang, *J. Chem. Inf. Model.*, 2011, **51**, 69–82.
- [155] J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman and D. A. Case, *J. Am. Chem. Soc.*, 1998, **120**, 9401–9409.
- [156] M. K. Gilson and B. H. Honig, *Proteins: Struct., Funct., Bioinf.*, 1988, **3**, 32–52.
- [157] F. Fogolari, A. Brigo and H. Molinari, *J. Mol. Recognit.*, 2002, **15**, 377–392.
- [158] W. C. Still, A. Tempczyk, R. C. Hawley and T. Hendrickson, *J. Am. Chem. Soc.*, 1990, **112**, 6127–6129.
- [159] M. S. Lee, M. Feig, F. R. Salsbury and C. L. Brooks, *J. Comput. Chem.*, 2003, **24**, 1348–1356.
- [160] I. Massova and P. a. Kollman, *J. Am. Chem. Soc.*, 1999, **121**, 8133–8143.
- [161] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. a. Case and T. E. Cheatham, *Acc. Chem. Res.*, 2000, **33**, 889–897.
- [162] B. Kuhn and P. A. Kollman, *J. Med. Chem.*, 2000, **43**, 3786–3791.
- [163] M. S. Lee, F. R. Salsbury and C. L. Brooks, *J. Chem. Phys.*, 2002, **116**, 10606.
- [164] G. Archontis, T. Simonson and M. Karplus, *J. Mol. Biol.*, 2001, **306**, 307–327.
- [165] M. K. Gilson and B. Honig, *Proteins: Struct., Funct., Bioinf.*, 1988, **4**, 7–18.
- [166] J. Warwicker and H. C. Watson, *J. Mol. Biol.*, 1982, **157**, 671–679.
- [167] N. K. Rogers, *Prog. Biophys. Mol. Biol.*, 1986, **48**, 37–66.
- [168] N. Singh and A. Warshel, *Proteins: Struct., Funct., Bioinf.*, 2010, **78**, 1705–1723.
- [169] J. Du, H. Sun, L. Xi, J. Li, Y. Yang, H. Liu and X. Yao, *J. Comput. Chem.*, 2011, **32**, 2800–2809.

- [170] T. Hou, J. Wang, Y. Li and W. Wang, *J. Comput. Chem.*, 2011, **32**, 866–877.
- [171] P. D. Lyne, M. L. Lamb and J. C. Saeh, *J. Med. Chem.*, 2006, **49**, 4805–4808.
- [172] F. Gräter, S. M. Schwarzl, A. Dejaegere, S. Fischer and J. C. Smith, *J. Phys. Chem. B*, 2005, **109**, 10474–10483.
- [173] S. Manta, A. Xipnitou, C. Kiritsis, A. L. Kantsadi, J. M. Hayes, V. T. Skamnaki, C. Lamprakis, M. Kontou, P. Zoumpoulakis, S. E. Zographos, D. D. Leonidas and D. Komiotis, *Chemical biology & drug design*, 2012, **79**, 663–673.
- [174] T. Schirmer and U. Jenal, *Nat. Rev. Microbiol.*, 2009, **7**, 724–735.
- [175] R. Hengge, *Nature reviews. Microbiology*, 2009, **7**, 263–73.
- [176] M. Egli, R. V. Gessner, L. D. Williams, G. J. Quigley, G. a. van der Marel, J. H. van Boom, A. Rich and C. a. Frederick, *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**, 3235–3239.
- [177] Y. C. Liaw, Y. G. Gao, H. Robinson, G. M. Sheldrick, L. a. Sliedregt, G. a. van der Marel, J. H. van Boom and a. H. Wang, *FEBS Lett.*, 1990, **264**, 223–227.
- [178] M. Gentner, M. G. Allan, F. Zaehring, T. Schirmer and S. Grzesiek, *J. Am. Chem. Soc.*, 2012, **134**, 1019–1029.
- [179] K. D. Smith, S. V. Lipchock, A. L. Livingston, C. a. Shanahan and S. a. Strobel, *Biochemistry*, 2010, **49**, 7351–7359.
- [180] U. Römling, M. Gomelsky and M. Y. Galperin, *Mol. Microbiol.*, 2005, **57**, 629–639.
- [181] L. Zhang and M. Meuwly, *ChemPhysChem*, 2011, **12**, 295–302.
- [182] Z. Zhang, B. L. Gaffney and R. A. Jones, *J. Am. Chem. Soc.*, 2004, **126**, 16700–16701.

Bibliography

- [183] Z. Zhang, S. Kim, B. L. Gaffney and R. a. Jones, *J. Am. Chem. Soc.*, 2006, **128**, 7015–7024.
- [184] P. Wassmann, C. Chan, R. Paul, A. Beck, H. Heerklotz, U. Jenal and T. Schirmer, *Structure*, 2007, **15**, 915–927.
- [185] J. Benach, S. S. Swaminathan, R. Tamayo, S. K. Handelman, E. Folta-Stogniew, J. E. Ramos, F. Forouhar, H. Neely, J. Seetharaman, A. Camilli and J. F. Hunt, *EMBO J.*, 2007, **26**, 5153–5166.
- [186] R. P. Ryan, T. Tolker-Nielsen and J. M. Dow, *Trends Microbiol.*, 2012, **20**, 235–242.
- [187] C. A. Shanahan, B. L. Gaffney, R. A. Jones and S. A. Strobel, *J. Am. Chem. Soc.*, 2011, **133**, 15578–15592.
- [188] H. Sondermann, N. J. Shikuma and F. H. Yildiz, *Curr. Opin. Microbiol.*, 2012, **15**, 140–146.
- [189] P. Ross, H. Weinhouse, Y. Aloni, D. Michaeli, P. Weinberger-Ohana, R. Mayer, S. Braun, E. de Vroom, G. A. van der Marel, J. H. van Boom and M. Benziman, *Nature*, 1987, **325**, 279–281.
- [190] H. Yan and W. Chen, *Chem. Soc. Rev.*, 2010, **39**, 2914–2924.
- [191] A. L. Chang, J. R. Tuckerman, G. Gonzalez, R. Mayer, H. Weinhouse, G. Volman, D. Amikam, M. Benziman and M.-A. Gilles-Gonzalez, *Biochemistry*, 2001, **40**, 3420–3426.
- [192] P. Aldridge, R. Paul, P. Goymer, P. Rainey and U. Jenal, *Mol. Microbiol.*, 2003, **47**, 1695–1708.
- [193] M. Christen, B. Christen, M. Folcher, A. Schauerte and U. Jenal, *J. Biol. Chem.*, 2005, **280**, 30829–33087.

- [194] G. B. Hecht and A. Newton, *J. Bacteriol.*, 1995, **177**, 6223–6229.
- [195] R. Paul, S. Weiser, N. C. Amiot, C. Chan, T. Schirmer, B. Giese and U. Jenal, *Genes & development*, 2004, **18**, 715–727.
- [196] B. Christen, M. Christen, R. Paul, F. Schmid, M. Folcher, P. Jenoe, M. Meuwly and U. Jenal, *J. Biol. Chem.*, 2006, **281**, 32015–32024.
- [197] R. Paul, S. Abel, P. Wassmann, A. Beck, H. Heerklotz and U. Jenal, *J. Biol. Chem.*, 2007, **282**, 29170–29177.
- [198] H. M. Berman, T. Battistuz, T. N. Bhat, W. F. Bluhm, P. E. Bourne, K. Burkhardt, Z. Feng, G. L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J. D. Westbrook and C. Zardecki, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2002, **58**, 899–907.
- [199] F. F.-F. Schmid and M. Meuwly, *J. Mol. Biol.*, 2007, **374**, 1270–1285.
- [200] L. Ma and Q. Cui, *J. Am. Chem. Soc.*, 2007, **129**, 10261–10268.

Curriculum Vitae

Personal Information

Lixian Zhang
Mittlere Strasse 33
4056 Basel
E-mail: lixian.zhang@unibas.ch

Date of birth: 13.04.1982
Nationality: China

Education

University of Basel, Basel, Switzerland

PhD Studies Chemistry **January 2009 - May 2013**

- Dissertation Topic: “Computational Characterization of Dimerization and Ligand Binding in Biological Systems”
- Advisors: Prof. Dr. Markus Meuwly

MSc Chemistry **September 2005 - June 2008**

- Master Thesis Topic: “Topology of Electron Charge Density: A Prospect from Valence Bond Theory; Cytochrome P450 Oxidation of Sulfur-containing Substrate”
- Advisor: Prof. Dr. Wei Wu

BSc Chemistry **September 2001 - June 2005**

Publications

Stability and Dynamics of Cyclic Diguanylic Acid in Solution.
Zhang, L.; Meuwly, M. *ChemPhysChem* 2011, 12, 295-302.

Topology of Electron Charge Density. A Prospect from Valence Bond Theory.

Curriculum Vitae

Zhang, L.; Ying, F.; Wu, W.; Hiberty, P. C.; Shaik, S. *Chem. Eur. J.* 2009, 15, 2979-2989.

Which Oxidant Is Really Responsible for Sulfur Oxidation by Cytochrome P450?

Li, C.; Zhang, L.; Zhang, C.; Hirao, H.; Wu, W.; Shaik, S. *Angew. Chem. Int. Ed.* 2007, 46, 8168-8169.

Publications in preparation

Dimerization and Protein Binding of the Analogues of c-di-GMP.
Zhang, L.; Meuwly, M. (Submitted)

Influence of Mutations at Position B24 on the Stability of Insulin Dimer.
Zhang, L.; Koch, M.; Meuwly, M. (Manuscript in preparation)