In Silico Facets of Biochemical Research: Accounts from Protein Folding and Protein-Ligand Interaction Studies

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Ajay Abisheck Paul George

aus

Neyoor, Indien

Bonn, November 2019

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Diana Imhof
2. Gutachter: PD Dr. Arijit Biswas
Tag der Promotion: 12.03.2020
Erscheinungsjahr: 2020

Dedicated to Swathi, for her unconditional love, support, patience and sacrifice.

Zusammenfassung

Die exponentiellen Fortschritte in der Computertechnologie in den letzten fünf Jahrzehnten haben der Wissenschaft und der Menschheit als Ganzes allgegenwärtig geholfen. Speziell Teilbereiche der Wissenschaft, die unter das Forschungsgebiet „Biochemie“ fallen, profitieren immens vom Anstieg der Rechenleistung. Insbesondere Proteine, die zu den vielseitigsten aller biologischen Makromoleküle zählen, sind aufgrund ihrer Zielfunktion für die Mehrheit der pharmazeutischen Arzneimittel schon immer Gegenstand umfangreicher experimenteller Untersuchungen gewesen. *In silico* Methoden unterstützen die experimentelle Forschung von Proteinen auf vielfältige Weise. Dies reicht von relativ einfachen Aufgaben wie der Einordnung von Sequenzen und Strukturen in biologische Datenbanken bis hin zu atomar aufgelösten Erkenntnissen der Proteinstruktur und der Dynamik, die die Grundlage für die biologische Funktion der Proteine bilden. Angesichts der unbestreitbaren Gewissheit, dass die dreidimensionale Struktur die Proteinfunktion bestimmt, wirft das Verständnis der Strukturbildung aus der Sequenz, d.h. die Proteinfaltung, trotz massiver Erkenntnisse der letzten 50 Jahre, weiterhin einige Fragen auf.

In dieser Arbeit werden Computermethoden vorgestellt, die mittels molekularer Modellierung und biomolekularen Simulationen wichtige biochemische Phänomene, wie Proteinfaltung und Protein-Ligand-Interaktionen, untersuchen. Im Hinblick auf die Proteinfaltung wird speziell die oxidative Selbstfaltung behandelt. Die oxidative Faltung von Proteinen beschreibt die kovalente Verknüpfung von Cysteinresten zu Disulfidbrücken. Disulfidbrücken in Peptiden und Proteinen weisen vielfältige Funktionen, wie die Stabilisierung der Struktur oder die Funktionskontrolle als allosterische Schalter, auf. Das aktuelle Wissen über die oxidative Faltung beschränkt sich auf Untersuchungen der Faltung verschiedener Modellpeptide und Proteine. Vor allem die Faltungswege der Proteine Rinder-Pankreas-Trypsin-Inhibitor (BPTI) und Hirudin wurden bisher als Extremmodelle verwendet. Allerdings sind nur wenige Informationen über die Faltung von disulfidreichen Peptiden und Proteinen, die ein Faltungsmuster zwischen diesen beiden Extremfällen aufweisen, bekannt. In der vorliegenden Studie werden Conotoxine aus marinen Kegelschnecken der Gattung *Conus* auf ihre oxidative Faltung untersucht. Auf Grundlage des Faltungsverlaufs werden die Conotoxine zwischen den von BPTI und Hirudin definierten Extremen eingeordnet. Neben der Untersuchung der oxidativen Faltung wecken Conotoxine mit ihrer Fähigkeit, potent und selektiv spannungsgesteuerte Natriumkanäle zu blockieren, ein breites pharmazeutisches Interesse. Zusätzlich zu den Conotoxinen werden Disulfidisomere von Tridegin, einem 66mer Peptid des Amazonas-Riesenblutegels *Haementeria ghilianii*, auf die Rolle von Disulfidbrücken hinsichtlich Faltung, Stabilität und Funktion untersucht. Tridegin ist pharmazeutisch interessant, da es der bisher einzig bekannte potente Peptidinhibitor des Blutgerinnungsfaktors XIIIA ist und ein großes Potenzial besitzt, als Leitsubstanz in der Antikoagulationstherapie zu dienen.

Häm wirkt als Effektormolekül regulatorisch auf bestimmte Proteine und deren physiologische Funktionen. Protein-Ligand-Wechselwirkungen in Form der transienten Bindung von Häm an Proteine werden mit Hilfe von molekularem Docking und molekulardynamischen Simulationen untersucht. Zudem wird das existente experimentelle Wissen über transiente Häm-Protein-Interaktionen als algorithmische Implementierung, die transiente Häm-bindende Motive in Proteinsequenzen vorhersagt und somit die Identifikation neuer Häm-regulierter Proteine ermöglicht, präsentiert. Insgesamt dient diese Arbeit als Beweis für die wachsende Bedeutung von *in silico* Methoden zur Unterstützung der experimentellen biochemischen Forschung.

Abstract

Exponential advancements in computer technology over the last five decades have ubiquitously benefited science and humanity as a whole. Consequent beneficiaries of this surge of computational power include all subfields of science that fall under the umbrella of “biochemical research”. Specifically, proteins, possibly the most versatile of all biological macromolecules, have always been the subject of extensive experimental investigation and more so from a pharmaceutical perspective, since the majority of pharmaceutical drugs target proteins. *In silico* methods assist experimental research on proteins in multiple ways ranging from relatively simple tasks such as organizing sequences and structures in biological databases to providing atomistic level insights into the structure and dynamics, that form the basis of the biological function of the protein. Given the unquestionable certitude that the three-dimensional structure of the protein determines its function, understanding the formation of structure from sequence, i.e. protein folding, is a central theme of investigation. Despite massive improvements in the understanding of protein folding over the last 50 years, it still remains an unsolved problem.

Herein, computational approaches involving a combination of molecular modeling and biomolecular simulations are pursued to study important biochemical phenomena, namely, protein folding and protein-ligand interactions. In terms of protein folding, a specific problem, i.e. oxidative self-folding is investigated. Oxidative folding refers to folding that involves the covalent linkage of cysteine residues in proteins to form disulfide bonds that stabilize the folded structure. Disulfide bonds play multifaceted roles in the peptides and proteins that they occur in, from providing structural integrity to acting as allosteric switches that regulate function. Current knowledge on oxidative folding has been restricted to consensuses derived from observing the folding of various model peptides and proteins. Most notably, the oxidative folding pathways of the proteins bovine pancreatic trypsin inhibitor (BPTI) and hirudin have been used as extreme models based on the manner in which the disulfide bond formations occur. However, the folding of a large group of such disulfide-rich peptides and proteins has been vaguely described to fall in between these extreme models. In this study, conotoxins, a class of venom peptides derived from marine cone snails of the genus *Conus* are used as candidates to study their oxidative folding, and to determine their place between the aforementioned extremes defined by BPTI and hirudin. With their ability to potently and selectively block voltage-gated sodium channels, conotoxins invoke broader a pharmaceutical interest than being mere model peptides to study oxidative folding. Furthermore, disulfide isomers of tridegin, a 66mer peptide produced by the giant Amazon leech *Haementeria ghilianii* are investigated for the role of disulfide bonds concerning folding, stability and function. The pharmaceutical significance of tridegin is that, it is the only known peptide inhibitor of the blood coagulation factor XIIIA, and shows great promise as a lead substance in anti-coagulation therapy.

Heme being an effector molecule, conveys a regulatory effect on the proteins it binds to, affecting their physiological functions. Protein-ligand interactions in the form of the transient binding of heme to proteins is investigated herein using molecular docking and molecular dynamic simulations. Finally, a decade’s worth of experimental knowledge obtained on transient heme-protein interactions is presented as an algorithmic implementation to predict transient heme-binding motifs in protein sequences, enabling the identification of novel heme-regulated proteins. Overall, this work serves as a testament to the growing significance of *in silico* methods in aiding experimental biochemical research.

Table of Contents

1 Introduction

The conception and execution of any form of scientific research endeavor without a significant *in silico* component is a virtually non-existent notion in the current day and age. This unprecedented advancement of computer technology over the last few decades, both in terms of hardware and algorithmic novelty, has rendered computational methods a mainstay in biochemical, biological and biomedical research. The completion of the human genome project in 2003 proved to be a watershed moment, as large volumes of biological data in the form of DNA sequences and whole-genomes became available for analysis.[1] A consistent drop in sequencing costs owing to advancement in next generation sequencing (NGS) technologies paved the way for 'bioinformatics' and 'computational biology' to gain mainstream attention.[2–7] Christian Anfinsen's experiments with ribonuclease in the late 1950s provided the foundations of protein folding, linking protein sequence to structure and function.[8,9] Rapid developments in structural biology,[10,11] molecular modeling and computational chemistry[12–14] have enabled close to accurate means of predicting protein structure,[15] protein-protein/protein-ligand interactions, and to observe macromolecular dynamics which is key to function.[16] Methods such as homology modeling[17–19], threading,[20–23] molecular docking[24–27] and molecular dynamics (MD)[16,28,29] simulations serve as tools to achieve these goals. It therefore goes without saying that computational methods have gained their rightful place in all areas of biochemical research. The current study leverages on these methods to augment existing knowledge on critically significant biochemical phenomena, namely protein folding and protein-ligand interactions. Understanding protein folding[30] despite its apparent conceptual simplicity,[31] has been one of the core objectives of the scientific community over the last five decades.[9] Besides being a non-deterministic polynomial complete (NP complete) problem[32] that piques the interests of mathematicians and computer scientists alike, the pivotal motivation to address protein folding lies in the fact that improper protein folding[33] (misfolding) is a significant contributor and a basis of several human diseases.[34–36] A bite-sized means to approach this daunting challenge is via bioactive peptides as model candidates of study.[37,38] The novelty of peptides is not limited to their relatively small size (<100 amino acids), but extends to their vast therapeutic potential that is only being commercially explored in the last decade.[39–45] Herein, from a large variety of available animal toxins,[46] a group of small, disulfide-rich peptides,[47–49] called conotoxins from cone snails of the genus *Conus* are studied.[50] This group of toxins exceeds over 700 species, and is classified according to their pharmacological targets.[51,52] In this study, selected members of the family of μ-conotoxins, known to selectively target and block voltage-gated sodium channels (VGSCs) with high specificity, are investigated.[50] Another animal-derived disulfide-rich peptide, i.e. tridegin, produced by the blood-sucking giant Amazon leech *Haementeria ghilianii*, which acts as a peptide inhibitor of the activated coagulation factor XIIIA (FXIIIA) is also investigated.[53] An improved understanding of the oxidative folding of disulfide-rich peptides which are till date defined only via limited models (BPTI and hirudin)[54–56], and the role of disulfide bonds in folding, stability and function are the main outcomes of this part of the study. [57–60]

On the front of protein-ligand interactions, the role of labile heme as a regulatory molecule[61] is explored by studying its transient binding to heme-regulatory motifs (HRMs) on the surfaces of proteins. Finally, an algorithmic implementation to detect transient heme-binding motifs (HBMs) is introduced as a means to rapidly screen large datasets of protein sequences for potentially novel heme-regulated proteins. [62]

2 Theoretical background

2.1 Peptide therapeutics

The term ‘drug’ is usually associated with small molecules (molecular weight < 500 Da) that act on their biological target (e.g., a protein) to evoke a physiological response in the organism. This association to small molecules is partly driven by the fact that most pharmaceutical drugs that exist are small molecules (Figure. 1).

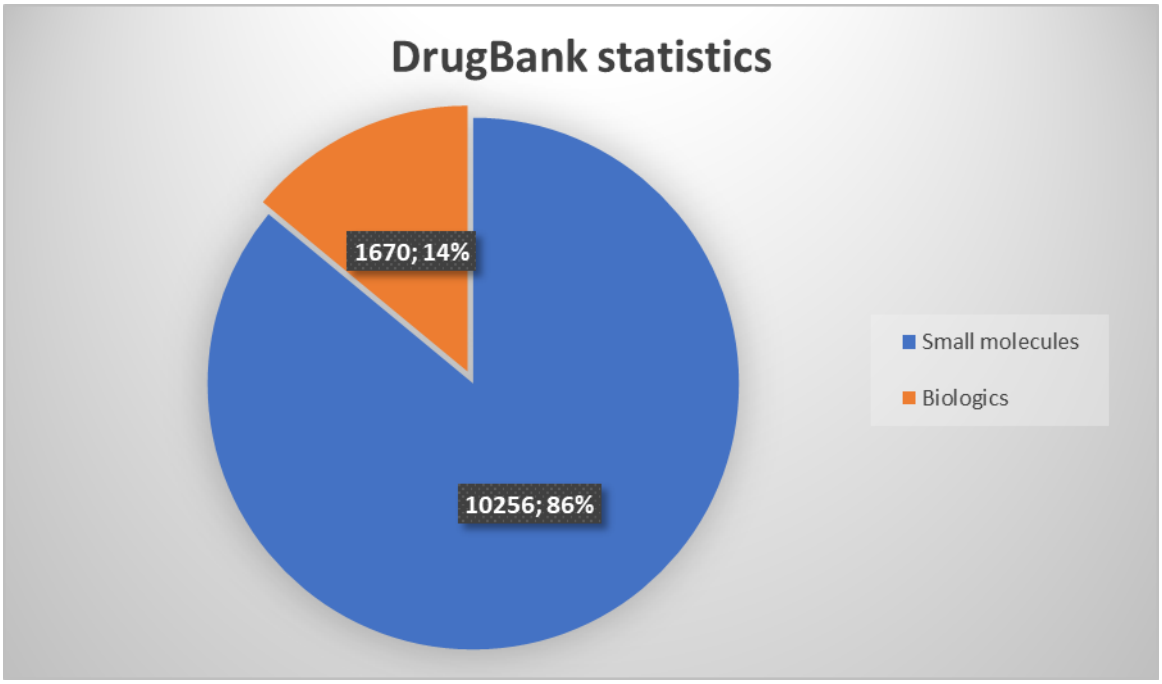


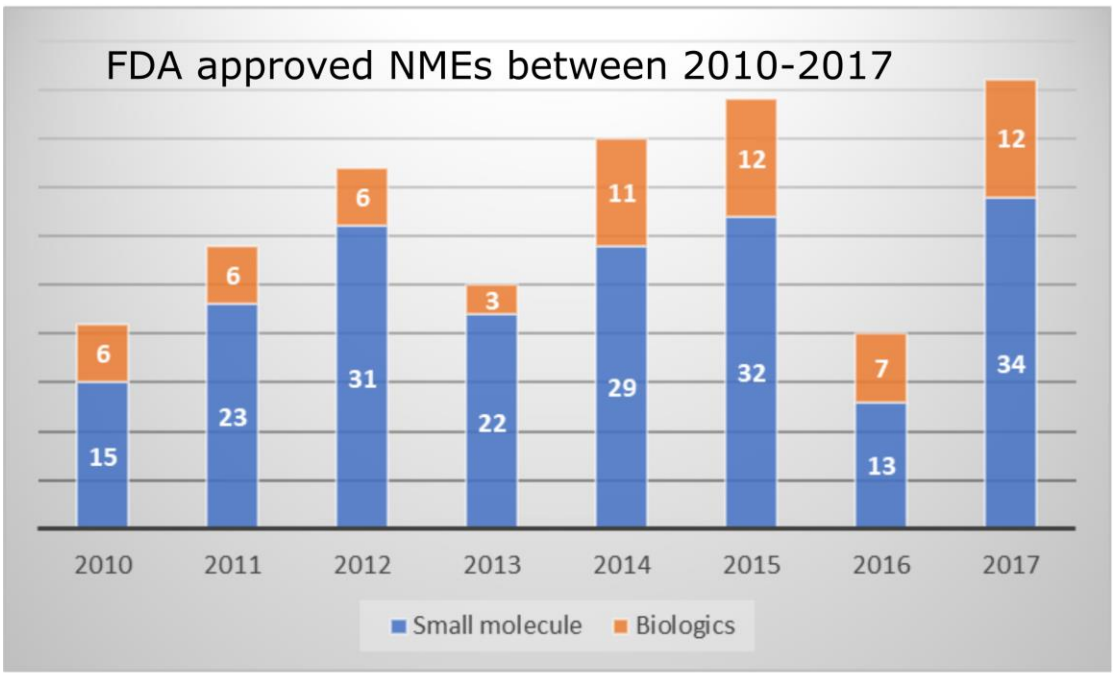
Figure 1. Comparison of the distribution of small molecule vs. biologics based on DrugBank data. *Data presented as a pie chart from the DrugBank database*[63–65] *(version 5.1.4, August 2019) shows the distribution of small molecule drugs and biologics currently in the market.*

Pharmaceutical companies also eagerly tend to pursue research on small molecules since they have significantly outweighed other molecule classes such as peptides (molecular weight 500 – 5000 Da) and biologics (molecular weight > 5000 Da) in terms of obtaining approvals from regulatory agencies such as the Food and Drug Administration (FDA) of the United States. In 2018 alone, small molecules accounted for 71% of all new molecular entities (NMEs) approved by the FDA that constituted a record-high of 42 small molecule approvals.[66] Small molecules have simple, well established synthetic routes, low production cost, high conformational restriction, good membrane permeability, good metabolic stability and oral bioavailability.[47] It is therefore clear that small molecules have an upper hand in terms of driving profits for mainstream pharmaceutical companies.

On the other end of the molecular size spectrum feature biologics (molecular weight > 5000 Da), whose market potential is growing with time. Prominent classes of biologics include proteins, like monoclonal antibodies, vaccines as well as gene and cell therapy products.[67] Triggered by the approval of recombinant human insulin in 1982, biologics have established their own expansive presence in the pharmaceutical market.[68] Biologics have contributed towards the development of a number of life-altering therapies against diseases like cancer (e.g. Trastuzumab and Campath), rare genetic diseases (e.g. Cerezyme, Fabrazyme,

Aldurazyme, Myozyme, and Mipomersan), and immune disorders (e.g. Mdalimumab and Rituximab).[68]

A



B

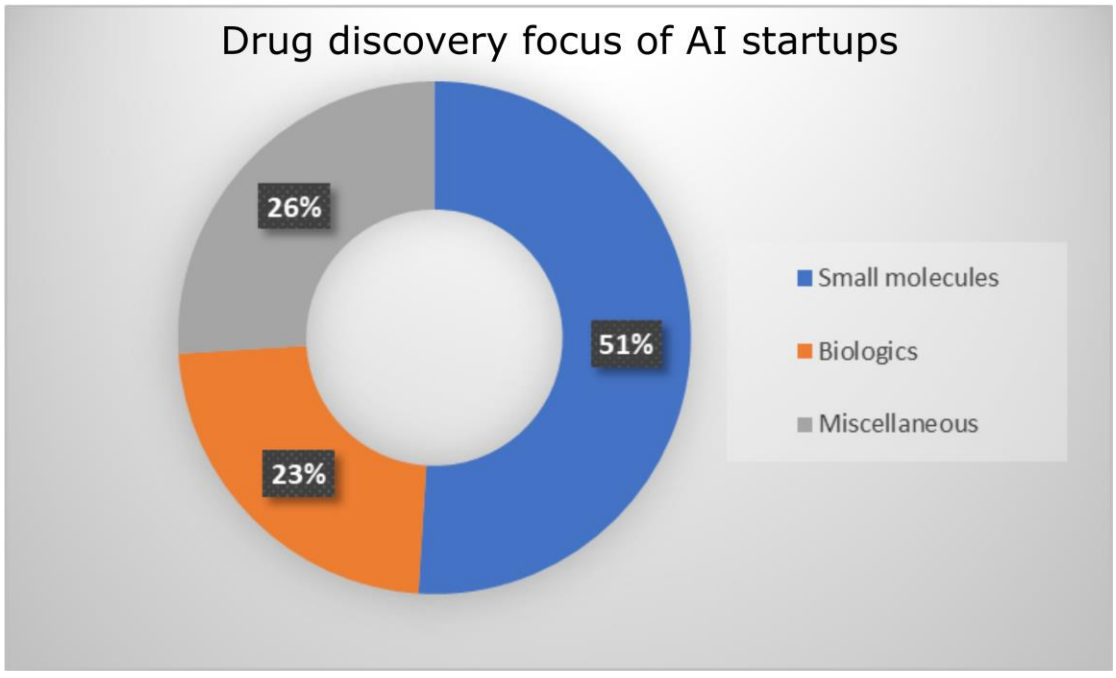


***Figure 2. Comparative statistics of small molecules and biologics in the pharmaceutical industry**. **A)** Stacked bar charts comparing the number of FDA-approved small molecules vs biologics over a 8 year period between 2010 and 2017. The data reveals that the approval for biologics per year ranges between ~20% and ~55%. **B)** Doughnut plot showing the proportion of the different molecule classes that drug discovery based AI startups focus on. More than half of these companies focus on small molecules while biologics take up about 23% of the focus. Data obtained from online source.*[69]

Chronic diseases such as asthma have also been addressed by biologics therapies. The first and only approved biologic against asthma is the immunoglobulin E-targeted monoclonal antibody Omalizumab.[70] Omalizumab has been found to reduce disease exacerbations, emergency room visits and hospitalizations in patients with difficult-to-treat allergic asthma.[70] Biologics have an advantage over small molecules in terms of attrition rates of new drug candidates in clinical studies.[67] Clinical success data from 2016 show that biologics have a success rate of 11.5% which is twice as much as that for small molecules (6.5%).[67] This higher success rate is attributed to a combination of better selectivity and lower off-target effects (of biologics) over small molecules.[67] Data from the past decade shows that the interest towards biologics is steadily increasing. FDA approval data collected between 2010 and 2017 show that biologics make up approximately 25% of all NMEs (Figure 2A).[69] The growing adoption of artificial intelligence (AI) in the drug discovery process[71–74] has also helped the development of biologics. 23% of all AI based startups that are involved with drug discovery work on biologics (Figure 2B) and this proportion is only bound to increase in the future.

That being said, the last decade has also seen peptides, that are poised in a sweet spot between small molecules and larger biologics gain pharmaceutical interest (Figure. 3).[47,75] With over 7000 naturally occurring peptides identified and approximately 140 candidates being evaluated in clinical trials, the peptide-drug market is generating commercial gravity.[75] This class of biomolecules play crucial roles as hormones, neurotransmitters, growth factors, ion-channel ligands and anti-infectives.[75] The therapeutic potential of peptides presents a case of ‘two sides of the same coin’. While on the one hand poor membrane penetration, inadequate absorption through mucosa and propensity to be degraded by proteolytic enzymes present significant challenges, their high selectivity, efficacy and activity at low concentrations resulting in relatively non-toxic amino acid metabolites (in comparison to small molecules), present huge opportunities.[43]

The very initial efforts into the industrial production of peptide therapeutics in the 1980s proved to have limited success as it was revealed that most natural peptides lacked drug-like properties.[47] Added to this, inefficient manufacturing methodologies resulting in increased production costs drove companies and investors away from researching the potential of peptides as drug candidates and hampering further developments in the field.[43,47] This dire situation has taken a turn for the better more recently as Roche demonstrated a cost effective synthetic means to produce the peptide drug enfuvirtide (Fuzeon®, Roche), a human immunodeficiency virus (HIV) membrane fusion inhibitor on a ton-scale per annum.[76] The current state of peptides as therapeutic agents span over a broad spectrum of biomedical applications[47] ranging from tissue engineering,[77–80] vaccines,[81–86] drug delivery systems[87–91] to medical imaging technologies.[92,93]

As illustrated in Figure 3, constrained peptides can be considered as biomolecules that combine the desirable properties of therapeutic proteins (biologics) and small molecules and, naturally occurring such peptides are considered as good starting points for peptide-based pharmaceutical product design and development.[94–96] This means, that obtaining a pharmaceutically relevant bioactive molecule from this starting point requires modifications and fine-tuning in order to achieve (and retain) drug-like properties.[97,98] Obviously, for peptides, attaining such a status can be reasonably more complex than for small molecules whose drug-likeness has been defined by principles such as the Lipinski’s rule of five established in 1997.[99] Recent reports have, however, suggested that this somewhat archaic rule-of-thumb does not stand the test of time.[97,100–103] While nature, in the form of evolution has been involved in fine-tuning peptide conformations, there are still several intrinsically undesirable pharmacokinetic properties in native peptides that need the expertise and

intervention of chemists.[47] A prime shortcoming of therapeutic peptide candidates is the inherent conformational instability (in comparison to a small molecule) that in turn confers energetic and entropic ramifications on target recognition and binding. Furthermore, conformational instability renders the peptide bond exposed and venerable to attack from proteolytic enzymes.[47,75,104–106] Several methods are in place, and in constant development to address these inadequacies of peptides as therapeutics that focus on improved properties such as selectivity and potency.[107–109]

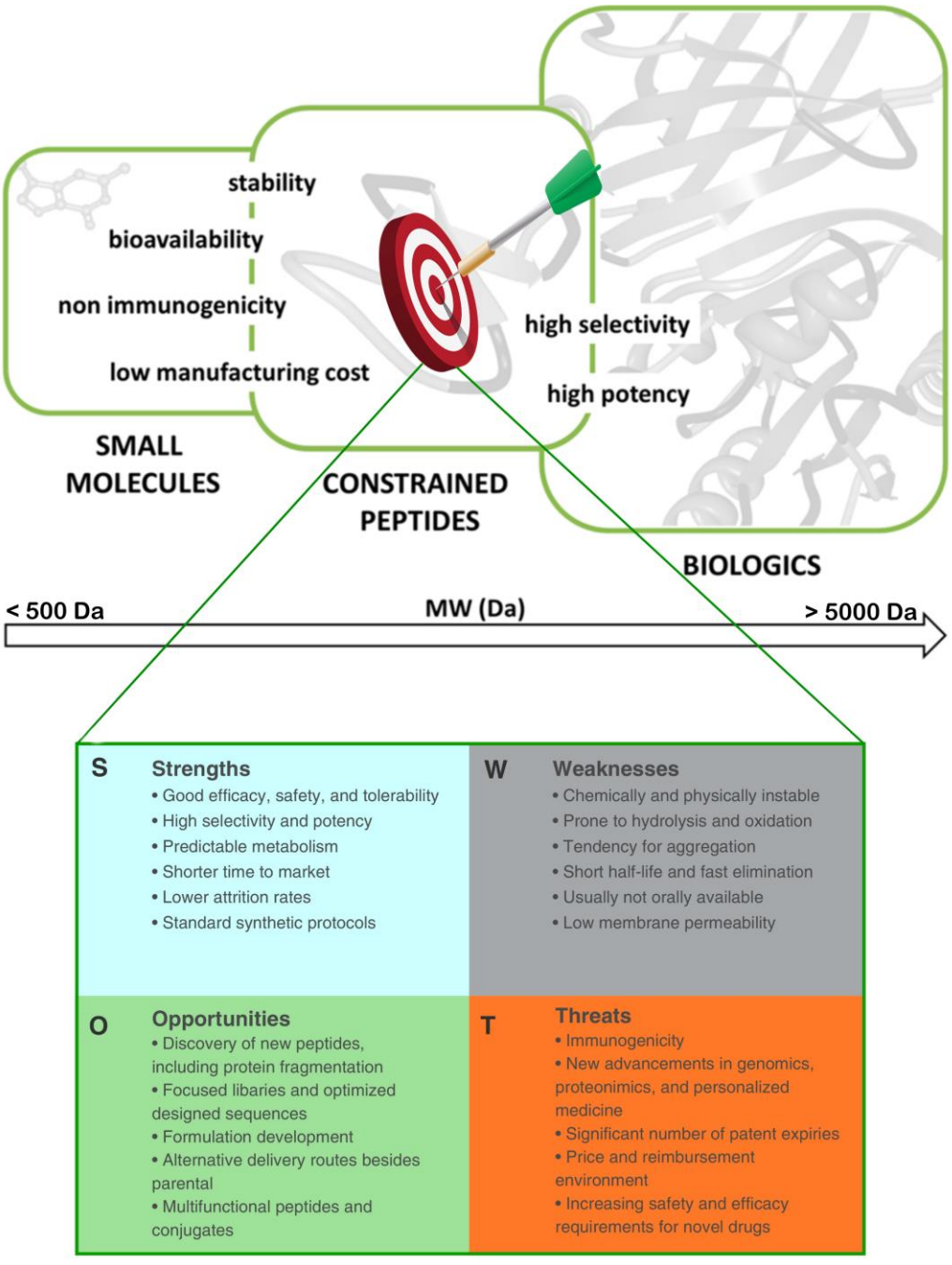


Figure 3. Constrained peptides as therapeutic agents. Top) *Peptides fill the gap between small molecules and biologics as they share favorable properties between small molecules and biologics.* ***Bottom)*** *A strength, weakness, opportunity and threat (SWOT) analysis of peptides as drugs is presented that serves as a guideline for research in the direction of peptide therapeutics. Adapted from literature.*[47,75]

As a means to overcome the lack of oral bioavailability that many peptide drugs suffer from, other routes of administration have been explored.[43] Commercial liposome-based delivery systems have found success in transdermal and pulmonary routes of administration.[110,111] Parenteral administration, despite its efficiency, does not address the oral bioavailability problem of peptide drugs and more prudent and innovative solutions are sought after. Approaches to constrain backbone conformational flexibility provide straightforward solutions for the design and synthesis of bioactive peptides. Constraining a peptide to explore its biologically relevant conformations by imposing structural restrictions on backbone mobility is usually the first step before rational chemical modifications that enhance affinity are attempted.[47] Backbone constraining methods are adopted to additionally induce structure-stabilizing secondary structural elements that add to the peptide's stability.[112] Other methods that have witnessed substantial success in restricting conformational flexibility of peptides include cyclization and stapling approaches.[113] Further enhancements to the conformational restrictions imposed by cyclized peptides are achieved by exchanging natural amino acids by non-proteinogenic ones that induce local conformational restrictions.[114] N- and C- termini modifications, PEGylation, glycosylation and fatty acid acylation are few of the other methods employed to achieve improved peptide stability and/or pharmacological properties.[43,108,115]

Disulfide bonds that occur in many constrained peptides of natural origin are an exceedingly important source of conformational stability. Multiple disulfide bonds present in disulfide-rich peptides and mini-proteins confer to these molecules privileged statuses.[47] The multifaceted roles of disulfide bonds in folding, stability, and function are one of the core concepts investigated in this work using computational methods. The role of disulfide bonds in enhancing pharmacological properties, enhancing peptide selectivity and elucidating active conformations are discussed in detail in the upcoming sections.[47]

2.2 Disulfide-bonded peptides and proteins

2.2.1 Role of disulfide bonds in protein folding, stability and function

The physical process by which protein sequences transform from a linear chain of covalently linked amino acids into distinct three-dimensional (3D) structures usually resulting in a biologically relevant state or conformation is called protein folding. Despite the apparent triviality with which the question is often formulated i.e., determining the final 3D arrangement of a linear string of beads, 50 years later, protein folding is still one of the biggest challenges in biological sciences.[8,9] Disulfide bonds are covalent links formed between the sulfur atoms of the thiol groups of two cysteine residues in the protein chain (or between cysteines from different protein chains) during the process of protein folding.[116,117] They are considered as one of the most commonly occurring post-translational modifications seen in secreted peptides and proteins.[118] Disulfide bonds occur in both prokaryotic (bacterial) as well as in eukaryotic systems but secreted extracellular proteins are the ones that most frequently contain disulfide bonds in their folded conformation.[116] The redox environment within the confines of the cytosolic space does not favor the formation of disulfide bonds as this environment maintains the cysteine sulfhydryls in a reduced state. Hence, rapid disulfide formation is observed in the extracellular space under oxidizing conditions.[116] In eukaryotic cells, disulfide bonds are formed in the endoplasmic reticulum (ER) owing to the availability of oxidizing conditions, chaperones and disulfide isomerases.[116,119,120]

In proteins, the disulfide bonded architecture can be decomposed into a six-atom configuration ($C_{\alpha1}$-$C_{\beta1}$-$S_{\gamma1}$-$S_{\gamma2}$-$C_{\beta2}$-$C_{\alpha2}$) that links the two cysteine residues (Figure 4).[116] In 1981, Janet Thornton studied the atomic coordinates from 55 structures of globular proteins available at that time to define the characteristics and bond geometries of disulfides in proteins.[121] Almost two decades later a set of 131 non-homologous protein structures containing a total of 351 disulfide bonds were studied by Petersen *et al*. providing further insights into the bond angle and distance distributions of naturally occurring disulfides, serving as the foundation for future work on disulfide engineering.[116,122] Further work by Dombkowski *et al*. using a much larger dataset of 1505 disulfide bonds from 331 non-homologous proteins determined updated values of the χ_3 torsion angles, that enables the rotation of the C_β atoms about the S-S bond (Figure 4).[123] This χ_3 torsion angle, touted to be critical for the stability of a disulfide bond, peaks at -87° and +97° (Figure 4).[116,124,125] The role of disulfide bonds on protein folding and structure discussed to be of foremost importance is that they impart structural and thermodynamic stability to the folded structure.[47,49,116,126,127] The formation of disulfide bonds affects the thermodynamics of protein folding by decreasing the entropy of the unfolded state, thereby favoring the folded state over the unfolded.[49,57] A consequence of the stability conferred to proteins by their disulfide bonds, these proteins often have higher resistance to damage and increased half-lives.[128] The quantitative contribution of disulfide bonds to the stability was investigated by observing and computing the change in stability upon systematic disruption of the native disulfide bonds in the protein.[116,129] The first of these efforts involved methodical rupture of the two native disulfide bonds of the protein ribonuclease T_1.[116] It was inferred from this study that each disulfide bond in ribonuclease T_1 had an energetic contribution of ~3.5 kcal/mol towards the thermodynamic stability of this protein.[116,130] Other similar experiments reported a range between 2.3 – 5.2 kcal/mol to the contribution of native disulfide bonds to thermodynamic stability.[116,131] Based on experimental consensus, the change in conformational entropy caused by a disulfide bond was mathematically modelled as:

$$\Delta S = -2.1 - \left(\frac{3}{2}\right) R ln(n)$$

where n is the number of residues enclosed by the loop formed by the disulfide bond and R is the universal gas constant.[116,131]

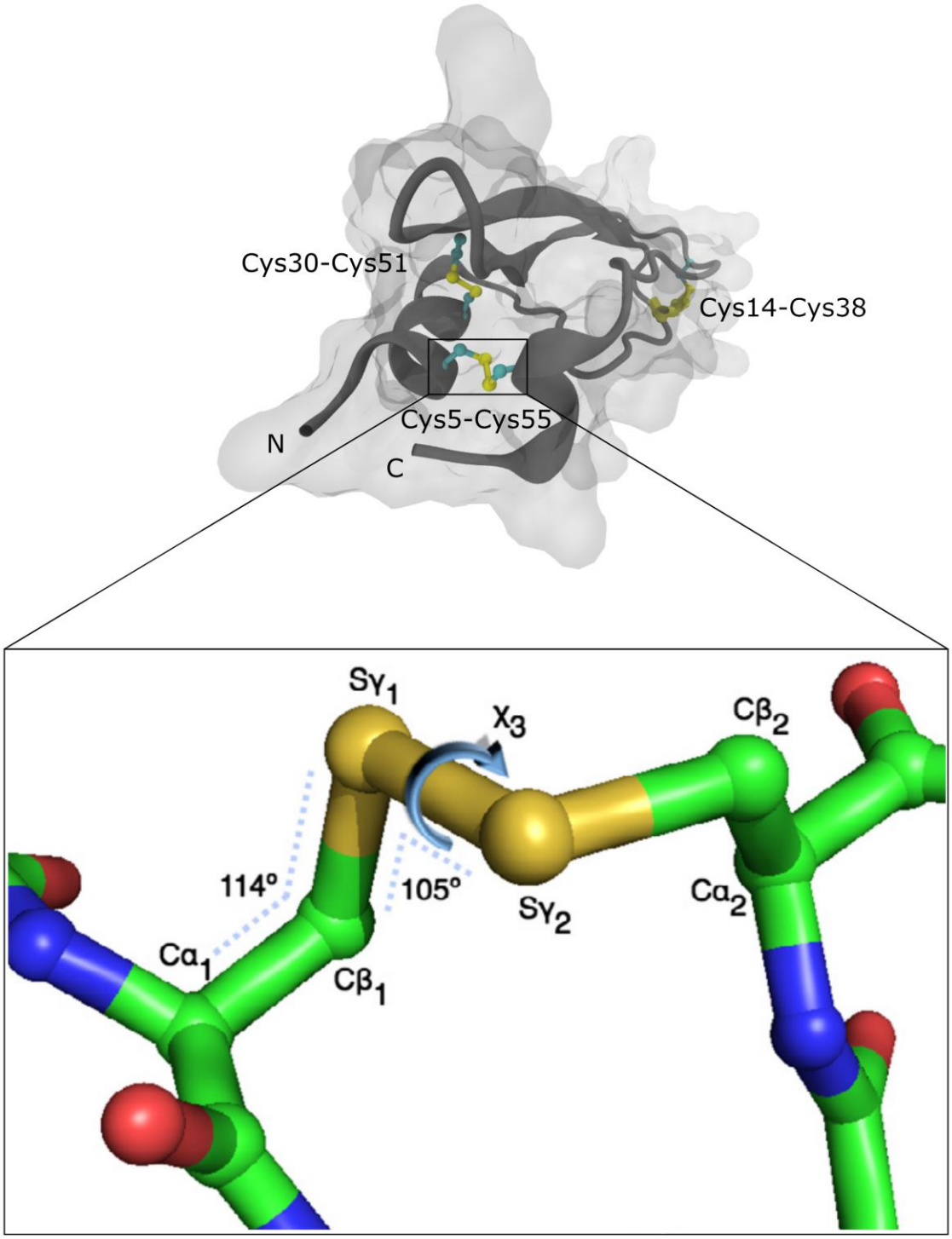


Figure 4. The architecture of a disulfide bond. *The three native disulfide bonds of protein bovine trypsin pancreatic inhibitor (BPTI) (PDB ID: 1QLQ[132]) are marked and labelled. BPTI is shown in the cartoon representation (gray) engulfed in a translucent molecular surface representation (gray). A zoomed in version of the Cys5-Cys55 disulfide bond is shown at the bottom with the important atoms, bonds and bond angles labelled.*

BPTI is one of the earliest and most important disulfide-bonded protein to be studied.[133] Modified from literature.[116]

For proteins defined by a simple two-state unfolding i.e., those exhibiting reversible denaturation, the above model produces results consistent with experimental observations.[116,130] Later studies using the protein lysozyme from the Enterobacteria phage T4, using the "loop perturbation analysis" method revealed that the energy contribution of a disulfide bond might also depend on the amino acid composition of the disulfide loop.[116,134] Further evidence from literature indicates that the introduction of disulfide bonds to the protein structure alters the free energy via an enthalpic change.[57,116,128,135] The idea of adding or engineering novel disulfide bonds into the protein structure to improve thermodynamic properties such as stability and half-life is a well-founded one and accordingly several such attempts have been explored already. T4 lysozyme, subtilisin, dihydrofolate reductase, and λ repressor were some of the proteins used as candidates to investigate the stabilizing effects of disulfide bonds.[136–138] In the case of T4 lysozyme, a 164 amino acid long protein, introducing a single novel disulfide bond between one of its native free cysteines (Cys97) and a mutated cysteine (Ile3Cys) was relatively easy.[139] Alternative disulfide bonded configurations within T4 lysozyme such as the Cys21-Cys142 engineered disulfide, proved beneficial by increasing the melting temperature by 11 °C.[140] At this point it might seem like a good idea to simply introduce disulfide bonds into sequentially distant parts of the protein structure and obtain a protein variant with increased stability. This is unfortunately not the case. Different cysteine pairings namely Cys90-Cys122 and Cys127-Cys154 in T4 lysozyme resulted in protein variants with lower stabilities and melting temperatures lowered by -0.5 and -2.4 °C, respectively.[139] Beyond their contribution to structural and thermodynamic stability, disulfide bonds also play alternative functional roles, as redox-switches for example.[141] Glutathionylation and nitrosylation of cysteine residues are further examples of disulfides serving as thiol-based switches.[142] These observations are based out of more recent research, establishing that the disulfide proteome can be broadly classified into two subclasses based on function namely, "structural disulfides" (discussed so far) and "functional disulfides".[143] Functional disulfides are often dubbed "forbidden disulfides" since they disobey standard stereochemistry rules as the ones established by Richardson[144] and Thornton.[121] Some of these violations include, the occurrence of disulfide bonds on adjacent β-strands, in a single helix or sheet, on non-adjacent strands of the same β-sheet and between adjacent cysteines in the sequence.[141] These disulfides were also found to result in strained, high-energy conformations thereby having little to no positive contribution towards structural and thermodynamic stability of the proteins they occur in.[143,145,146] Whether the disulfide serves more of a structural role or not can be imputed from calculating the redox potential.[143] While structural disulfides are found to have potentials as low as -470 mV,[147] the redox potentials for functional disulfides vary between -90 mV and -330 mV.[143,148–151] Thus, it was identified that some disulfide bonds can also have a destabilizing effect and despite experimental and computational advances, discerning whether an engineered disulfide bond might increase or decrease the stability of a protein is not straightforward. Existing studies provide some guidelines in this regard while computational modeling and simulation studies have made the process somewhat easier.[116] However, the experimental synthesis of multiple disulfide-bonded peptides and proteins is still a daunting challenge that requires chemical acumen and expertise in a broad range of experimental techniques.[152] The obscurity in understanding the folding of disulfide-bonded peptides and proteins directly relates to complexity of the underlying process itself, namely, oxidative folding. The classification of the oxidative folding process into generic yet diverse pathways open the possibility to understand the folding of such peptides and proteins better as discussed in the upcoming section.

2.2.2 Oxidative folding pathways of disulfide-bonded peptides and proteins

Oxidative folding can be defined as a complex process by which an unfolded (reduced) protein or peptide attains its folded functional conformation (the so-called 'native state') by the formation of its native disulfide bonds along its folding pathway.[50] The oxidative folding process is said to encompass two distinct processes that occur in complementary fashion namely, "disulfide bond regeneration" and "conformational folding".[126] Experimental studies pioneered by Creighton *et al*.[153] towards understanding oxidative folding usually followed the process of unfolding the protein using conditions that favor unfolding of the protein (by disruption of the native disulfide bonds), followed by refolding the protein to be able to trap partially formed disulfide intermediates. Experimentally, the unfolding component is achieved using a strong reducing agent (e.g., dithiothreitol) in the presence of a denaturant (e.g., 6 M GdmCl), extreme pH coupled with elevated temperatures while the opposite, i.e., removal of the denaturant, a pH jump and temperature restoration is carried out in order to induce spontaneous refolding of the protein.[54] The protein in its fully reduced state is allowed to refold in buffer solution containing redox agents and the intermediates are quenched (either by sample acidification or chemical reaction) and further analyzed by techniques such as HPLC.[54] The analysis of folding intermediates is intended to characterize the disulfide isomers that form along their refolding pathway. The main goal of studying oxidative folding is to understand the folding of multiple-disulfide-bonded peptides and proteins with regards to disulfide bonding patterns and the associated kinetics and stabilities. This motive in itself is non-trivial for several reasons. Firstly and most importantly, the fact that different disulfide bonded conformations of the same primary sequence resulted in proteins with varying biological activities, stresses on the importance of knowing the underlying mechanism involved in the folding and disulfide isomer formation of these biomolecules.[154–157] In biopharmaceutical production, the clear characterization of the disulfide structure is regarded as a critical quality attribute (CQA) as it determines the stabilities and efficacies of products such as antibody drugs and antibody-drug conjugates (ADCs).[44,158–161]

Individual research groups have used specific disulfide bonded proteins as candidates to study oxidative folding. The protein bovine pancreatic ribonuclease A (RNase A) with its four native disulfide bonds has been one such model protein.[60,162,163] Studies on RNase A resulted in the definition of four generic pathways that governed the oxidative folding of the protein to its folded state.[126] Scheraga and co-workers in 2000, formulated these four generic pathways based on three observed characteristics using RNase A.[126] These are : i) Are there folded disulfide intermediates observed?. ii) In the case of folded disulfide intermediates, are the intermediates metastable?, and iii) Among the folded intermediates, is preferential oxidation or reshuffling of disulfide bonds observed? Though this study established some ideas and a framework as to how the oxidative folding of such multiple-disulfide-bonded proteins, it focuses primarily on RNase A and stops short of applying such analysis to other proteins in-depth. This gap was however sufficiently filled in further reviews by Arolas *et al*.[49] in 2006 and and Chang in 2011.[54] Both Arolas *et al*. and Chang discuss oxidative folding pathways of disulfide-bonded proteins using the proteins bovine pancreatic trypsin inhibitor (BPTI) and hirudin, described to fold using opposite and extreme pathways as illustrated in Figure 5.[54] The simplest way of interpreting the differences between the two extreme pathways is by answering the following question. Does the folding of the multiple-disulfide-bonded peptide result in the formation of a single major disulfide isomer or does more than one alternative disulfide isomer accumulate besides the 'native' disulfide isomer? This bottom-up approach is valid since the aforementioned question can be answered by observing the HPLC elution profiles of the protein or peptide of concern.[54,129]

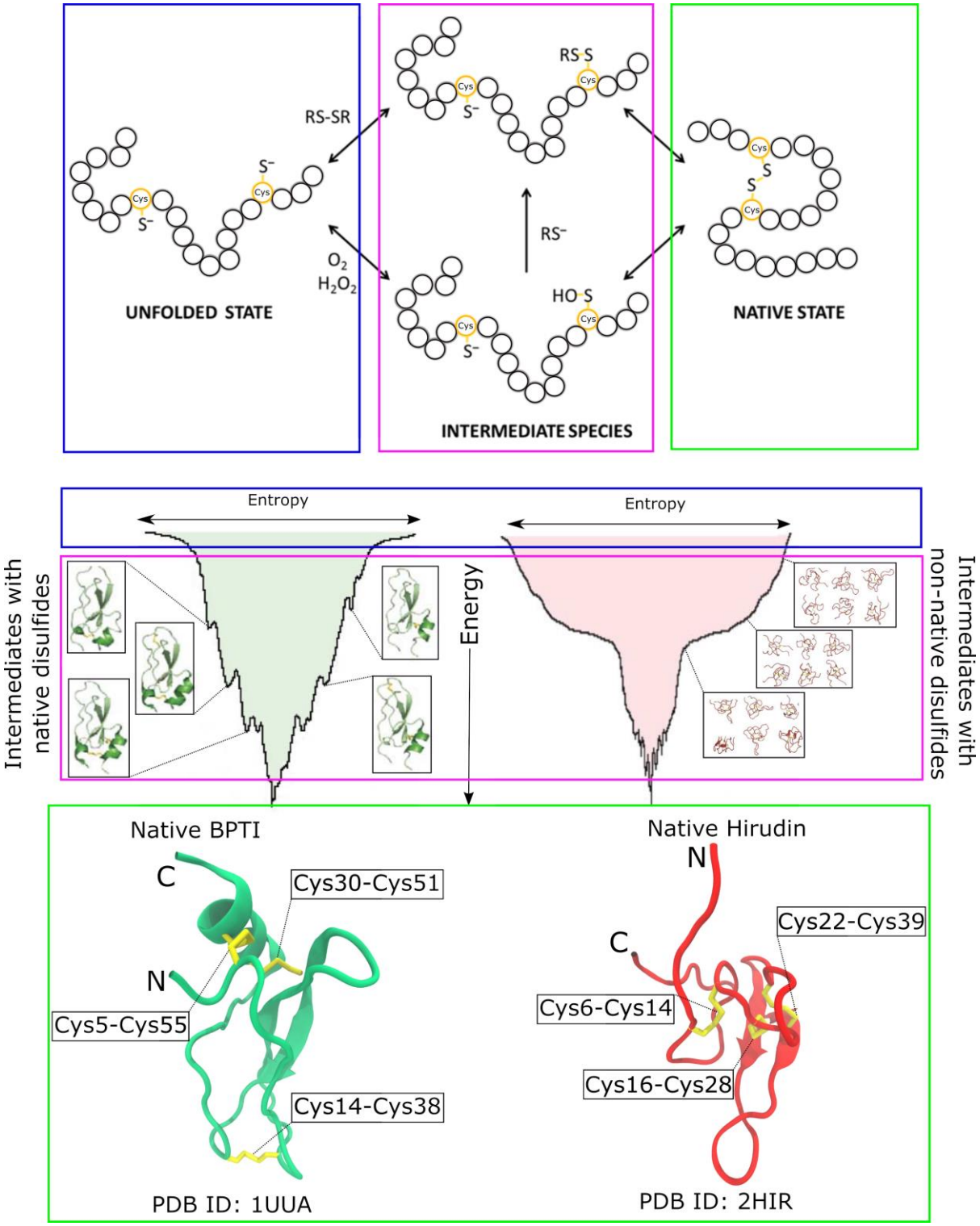


***Figure 5. Extreme and opposite pathways of oxidative folding of BPTI and Hirudin**. **Top)** Generic illustration of the key states encountered during the folding of a disulfide-bonded peptide i.e., the unfolded state (blue rectangle) where the none of the disulfide bonds have formed, the intermediate species (pink rectangle) wherein the peptide is partially folded and the native state (green rectangle) in which the disulfide bonds are correctly formed. **Center)** Folding energy funnels of BPTI (green) and hirudin (red). In BPTI, at each local minimum of the energy funnel, intermediates represented by partially folded species with native disulfide bonds are formed. In the case of hirudin (red), as the folding energy gradually drops, several partially folded intermediates containing non-native disulfide bonds are formed, leading to the existence of more than one isomer with alternative disulfide bonding patterns folded into low energy conformations, besides the native isomer. **Bottom)** The completely folded 3-disulfide-bonded isomers of BPTI (green ribbon) and hirudin (red ribbon) with their respective native disulfide bonds (yellow sticks) labelled. Figure adapted from literature sources.* [47,49]

Going one level higher using this bottom-up approach, we arrive at attempting to characterize the nature of the intermediates formed during the oxidative folding process. As seen from Figure 5 (center), here again a clear distinction can be drawn between the folding pathways of BPTI and hirudin. In the case of BPTI, as the folding proceeds by the gradual reduction in potential energy, partially formed disulfide isomers with one or more of the native disulfide bonds accumulate as intermediates. The formation of native disulfide pairings is accompanied with the appearance of native-like secondary structures. This in turn accelerates and guides the folding towards the correct disulfide isomer, resulting in the formation of a single 3-disulfide-bonded isomer i.e., the native isomer.[49] Arolas *et al*. and Chang relate the folding of BPTI to the 'framework model' of protein folding wherein local interactions contribute significantly towards the formation of native-like secondary structures from the early stages, that efficiently funnel the folding towards the native state.[49,54] In the case of hirudin, the opposite is observed (Figure 5, center). The intermediates formed along the folding pathway in this case, consists of heterogenous mixtures of partially folded isomers with non-native disulfide pairings.[49,54] As a consequence, at the bottom of the folding energy funnel, the native 3-disulfide-bonded isomeric species forms along with other 3-disulfide-bonded isomers with non-native disulfide pairings. Another means of viewing protein folding of this kind, is by considering the native state (or disulfide isomer in this case) as a uniquely selected metastable state among other possible states with configurational energies being local minima at the bottom of the folding energy funnel.[164] This description was proposed by one of the pioneers of protein folding studies Cyrus Levinthal way back in 1968.[164]

Besides describing extreme and opposite models of oxidative folding using BPTI and hirudin, Chang and Arolas *et al.* also described the folding pathways of several other proteins relative to the extreme models.[54] Accordingly, disulfide-bonded proteins that fold via the BPTI-like pathway include leech derived tryptase inhibitor (LDTI),[165] insulin-like growth factor-I (IGF-I)[166] and bovine α-interferon (IFN-α).[167] Similarly, the hirudin-like folding pathway is adopted by proteins such as potato carboxypeptidase inhibitor (PCI),[168] leech carboxypeptidase inhibitor (LCI),[169] ascaris carboxypeptidase inhibitor (ACI),[170] and tick carboxypeptidase inhibitor (TCI).[171] Interestingly, a few proteins have been shown to exhibit characteristics from both the extreme models forming the hybrid 'BPTI-hirudin-like' adopted by proteins such as tick anticoagulant peptide (TAP),[172] human epidermal growth factor (EGF)[173] and secretory leucocyte protease inhibitor (SLPI).[174]

A family of ω-conotoxins have been described to fold using the hirudin-like pathway by Chang.[54,175] Though this classification is subjectively true (for candidates with the one specific family of ω-conotoxins investigated[175]), the fact is that, conotoxins comprise a vast group of peptides, classified under a wide range hierarchies, and current folding studies do not adequately span the breadth of their diversity.[50] The following section discusses conotoxins, which are within the scope of this thesis, used as key candidates for computationally probing oxidative folding.[175]

2.2.3 Conotoxins

A diverse range of animals have evolved to be able to produce venoms for either predatory or self-defense motives.[176] A majority of these venoms comprise disulfide-rich, protease resistant peptides or proteins usually found as complex chemical cocktails.[176] The fact that most of these venom peptides target the cardiovascular and nervous systems have been exploited as therapeutic agents against pain, diabetes and hypertension to name a few.[176] Barring the usual suspects of venom producing animals such as snakes, scorpions and spiders, cone snails with over 700 species producing 100000 unique peptides present an interesting case for investigation.[50,176–179] Marine cone snails of the genus *Conus* are carnivorous gastropods that use their harpoon-like appendages to attack and inject powerful neurotoxins into their prey.[50,180] As most cone snail peptides (conopeptides) target membrane proteins such as ion-channels, receptors and transporters, they have been the subject of rigorous investigation to exploit their pharmaceutical potential.[50,51,177,181,182] That being said, the pharmaceutical success obtained from these peptides so far is limited. Till date there is only one FDA approved drug in the market of conotoxin origin. This is the conopeptide ω-MVIIA (from the species C. *magnus*) commercially sold as Ziconotide, used in the treatment of chronic pain for patients suffering serious conditions AIDS or cancer.[183,184] Consequently, conotoxin's therapeutic potential in pain-relief has gained public recognition and has been referred to as the "venom that's good for you".[185] One of the most beneficial tools in the study of conotoxins is the Conoserver database that maintains high quality, up to date information on sequence, structure, classification, post-translational modifications and statistics.[186,187] The webserver (http://www.conoserver.org/) is hosted by the Institute of Molecular Bioscience, Brisbane, Australia.

Conotoxins are classified under various levels of hierarchy (Figure 6). The highest in this level of hierarchy is the 'gene superfamily' classification which is based on sequence clustering of the precursor protein processed in the endoplasmic reticulum.[188] This classification describes the evolutionary relationship among the different conotoxin families and 28 published gene superfamilies are currently available in the latest update (13-09-2019) of the Conoserver. The next level of classification is based on the 'cysteine framework' defined by the pattern of neighboring and non-neighboring cysteines found in the sequence.[177,189] 27 of these unique cysteine frameworks have been identified till now.[177] The further general levels of classifications include the 'family' classification and most importantly the 'pharmacological target' based classification which segregates peptides based on their specific protein targets.[50] Additionally, structure based classifications on conotoxins take into consideration the disulfide connectivity and the nature of the 3D fold that results from the disulfide connectivity. Some of the important fold classes in this regard include, 'globular', 'cysteine knot', 'ribbon', 'beaded' and 'kunitz fold'.[50,177]

Several conotoxins have been rigorously investigated by both experimental and computational methods alike. Experimental investigations have primarily been in the direction of efforts such as chemical modifications and mutations to improve peptide stabilities, to tap into the underlying pharmaceutical potential of these toxins.[190] Ironically though, conotoxins have been used very rarely to study oxidative folding, and as mentioned earlier, a group of ω-conotoxins have been found to fold via the hirudin-like pathway.[175] On the other hand, members within the same conotoxin family, sharing identical cysteine framework but different primary sequences have been shown to display differing behavior during oxidative folding. Fuller *et al.* in 2005 studied the oxidative folding behavior of four μ-conotoxins (μ-GIIIA, μ-PIIIA, μ-SmIIIA and μ-RIIIK) belonging to the M-superfamily to infer that there existed noticeable variations in the folding properties and behavior among the four

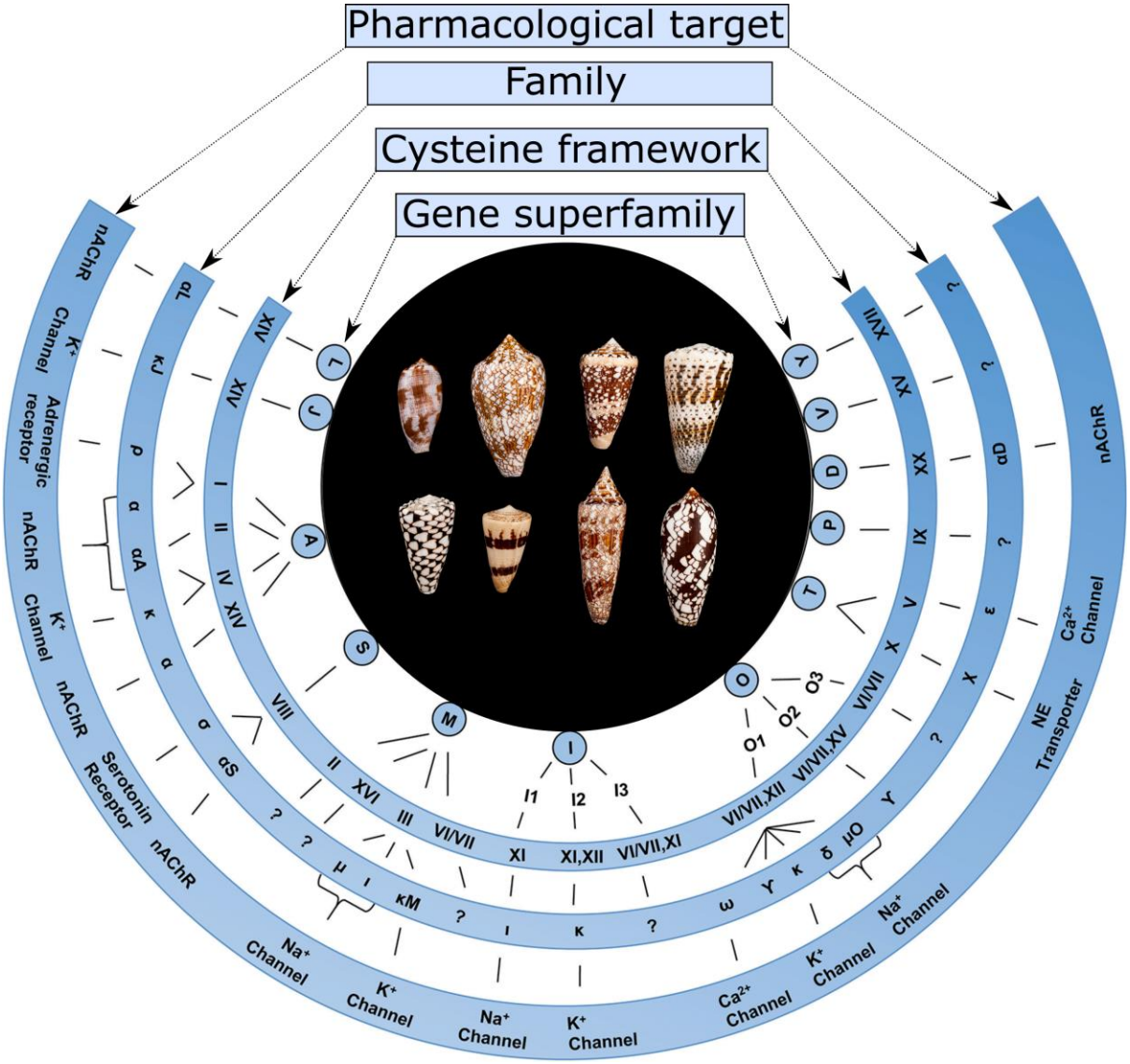


Figure 6. Schematic representation of the conotoxin classification hierarchy. *The major levels of hierarchy in conotoxin classification are based on the gene superfamily, cysteine framework, family and the pharmacological target. At the 'pharmacological target' level, some targets which do not have a specific receptor defined till now are denoted by NE (norepinephrine) and nAChR (nicotinic acetyl choline receptor). The black circle in the center contains the image of a representative collection of cone snails found in display at the Florida museum of natural history. Figure adapted from literature*[50] *and online*[191] *sources.*

investigated µ-conotoxins.[178] The main outcome of this study was the conclusion that, the folding behavior is determined to a great extent by sequence features even though the peptides shared comparable globular 3D structures induced by the identical disulfide bonding patterns present.[178] A similar study (featuring later in this thesis) uses experimental data and a novel computational approach to elucidate the nature of the underlying folding pathway of five µ-conotoxins namely µ-GIIIA, µ-KIIIA, µ-PIIIA, µ-SIIIA and µ-SmIIIA, assigning their relative positions between the BPTI-like and hirudin-like models.[129] Tietze *et al.* in 2012 questioned the validity of the popular concept that, for a protein only a single native, bioactive form exists, based on observations from the oxidative folding of the µ-conotoxin PIIIA in buffer solution.[192] The observations of this study found that isomers with alternative disulfide bonding patters in comparison to the native (I-IV, II-V, III-VI) pattern typical for this family, displayed comparable bioactivity at its target VGSC Nav1.4.[192] As an extension to this work, Heimer *et al.* in 2018 published the synthesis and NMR structure elucidation of twelve (of the fifteen) theoretically possible disulfide isomers of µ-PIIIA.[152] This

study described the impact of cysteine pairing on structure elucidation. More importantly, this study highlighted the need for different analytical methods and standards for the characterization of disulfide-rich peptides and proteins. To the best of current knowledge, this is the first study to have produced NMR structural ensembles for such a number of disulfide isomers of the same sequence.

μ-conotoxins have also been used as candidates to study disulfide engineering. Possessing mostly globular 3D conformations and a huge positive net charge, μ-conotoxins function by selectively and preferentially blocking voltage-gated sodium channels with impressive potencies.[50] To alleviate the complexity involved in the synthesis of multiple-disulfide-bonded peptides, approaches that essentially involve synthesis of the peptide with fewer disulfide bonds than what makes up the native isomeric fold are used. The toxin μ-KIIIA was shown to retain comparable bioactivity at its target channel with only two of its three disulfide bonds present.[193] However this mode of optimization does not prove successful in all cases as other studies using μ-GIIIA[194] and μ-PIIIA[195] proved that the peptides needed all three of their native disulfide bonds to be bioactive. Finally, several conotoxins have been investigated by *in silico* methods especially methods that combine molecular docking and MD simulations and additionally the usage of machine learning methods to study conotoxins is also on the rise.[177] This thesis is yet another example of *in silico* methods used in an effort to improve existing understanding on protein folding, using conotoxins as candidates of investigation.

2.2.4 Tridegin

Another source of animal-derived peptides are leeches, whose usage in therapy has a long trail in the annals of history.[196–198] The study of leech-derived peptides is therefore something that has been mainstay for quite a long while, both for the therapeutic aspects as well as towards understanding of biochemical processes such as protein folding in general. This is also obvious from the previous sections that introduce the leech-derived (*Hirudo medicinalis*) peptide hirudin, as a model to study oxidative folding.[54] In fact, the folding of hirudin itself is one extreme model of oxidative folding, and the 'hirudin-like' folding pathway is used to describe the oxidative folding of similar peptides. Hirudin is easily one of the most extensively studied leech-derived peptide, for his inhibitory activity on the hemostatic enzyme thrombin.[199] Leech-derived peptides have also been used in disulfide engineering applications to check stabilities and activities of the resultant peptide upon deletion of a disulfide bond.[200] Herein, we discuss tridegin, a 66mer peptide derived from the giant amazon leech *Haementeria ghilianii*.[196,201] Tridegin was first identified in 1997 and described by T. Sawyer and coworkers as a potent inhibitor of the coagulation factor XIIIa (FXIIIa).[202]

Factor XIIIa, a transglutaminase, is known to catalyze the final step of the blood coagulation cascade.[203] By far, tridegin is the only known naturally-derived peptide inhibitor to effectively inhibit the enzymatic activity of FXIIIa.[53] Recent experimental studies on tridegin have revealed that, under oxidative self-folding conditions, tridegin forms three different, 3-disulfide-bonded isomers in solution.[203] The disulfide connectivities of the so far identified disulfide isomers are Cys5-Cys17/Cys19-Cys25/Cys31-Cys37 for the first, Cys5-Cys37/Cys17-Cys31/Cys19-Cys25 for the second and Cys5-Cys31/Cys17-Cys37/Cys19-Cys25 for the third disulfide bonds respectively.[203] It is obvious from these disulfide connectivities that all three 3-disulfide-bonded isomers of tridegin identified thus far have a common disulfide bond between the residues Cys19 and Cys25. Unlike conotoxins, several of whose structures have been resolved by NMR[204] or X-ray crystallography, and

systematically maintained in databases,[186,187] tridegin has till date, no experimentally resolved structure. This has opened the possibility for computational studies, especially in the area of computational structure prediction to be conducted with tridegin. Currently, structures predicted by fold-recognition and threading methods implemented by the program I-Tasser,[23,205] are available for five of the fifteen disulfide isomers of tridegin.[203] Molecular docking studies conducted using these isomer models revealed that, two out of the three isomers bound over the active site of FXIIIa defined by the catalytic triad Cys314, His373 and Asp396, while the third isomer bound to a region distant from the catalytic site. It is proposed that, the residue Gln52 in the C-terminal region of tridegin plays a significant role towards the catalytic activity.[203] It was further proposed that the N-terminal disulfide constrained region of the peptide might play a crucial role for the binding of the peptide while the C-terminal segment between residues 38 and 66 could be significant for the inhibitory activity.[203] It is therefore obvious that the scope for in-depth research into the various aspects of tridegin, especially on how the individual disulfide bonds contribute to the folding, stability and function is very high. This work attempts to improve on the current understanding of tridegin based on computer-aided analysis of the folding and stability of its different disulfide isomers.

2.3 Heme as an effector molecule regulating protein function

The role of heme as an effector molecule that regulates protein function is a notion that is often understated and understudied. The abundance and ubiquitous nature of heme can simply be exemplified by just considering the amount of hemoglobin in the human body. Among the ~2.5 x 10^8 million hemoglobin molecules found in ~25 trillion erythrocytes, an estimated 5 x 10^9 molecules of heme are bound.[206,207] Discovered first as a prosthetic group in hemoglobin by Fritz Ludwig Hünefeld in the 1840s,[208] heme-related research won Nobel prizes for researchers involved in the field. In 1930, Hans Fischer won two Nobel prizes for the synthesis of hemin.[209] About three decades later, in 1962 Max Petruz and John Kendrew were awarded the Nobel prize for their pioneering work in the exploration of the structures of hemoglobin and myoglobin.[62] It was however, only in the early 1990s that an alternative role of heme as an effector or signaling molecule was recognized.[210] The identification of transient heme-protein interactions in human Aminolevulinic acid synthase (ALAS) by Lathrop and Timko in 1993 proved to be a crucial transition in the field, as they also defined the term heme regulatory motif (HRM).[211] HRMs are in general defined as a short stretches of amino acids found in the surface of a protein carrying a heme coordination site.[212] Heme binding to these HRMs may regulate protein function by either altering the stability or by forming a catalytic complex.[213] Over the years, several such motifs have been identified. The cysteine-proline dipeptide motif called 'CP motif' is one of the most prominent among the known motifs.[214–219] With over two decades of research in the search of such motifs that regulate protein function having passed, a more general definition is sought. Particularly, it has become clear that heme binding can occur via the residues histidine, cysteine, tyrosine and in some cases methionine. However, whether the binding has a functional impact on the protein or not is a different question all together. Hence the term heme binding motifs (HBMs) was introduced to describe heme binding with no functional consequence at the protein level or if the functional consequence was yet to be known.[62] In essence, all HRMs are HBMs but not all HBMs may be HRMs.

Heme-protein interactions can occur at various molecular levels. The key defining factor of a transient heme-protein interaction is the fast dissociation of heme post the binding event.[62] Figure 7 illustrates the differences between permanent and transient heme interactions to proteins. For regulatory heme, it is an absolute necessity that a coordinative bond between the central iron ion in the heme molecule and the heteroatom of the coordinating residue's side chain be formed (Figure 7C).[62] The most commonly occurring residues that enable such binding are cysteine (via sulfur), histidine (via oxygen) and tyrosine (via oxygen) while coordination via lysine and methionine are rarely observed.[215] Additionally, hydrophobic interactions and pi-pi stacking between the porphyrin ring as well as typical electrostatic interactions such as hydrogen bonds contribute to the stabilization of heme binding (Figure 7C). It is therefore clear that, for transient heme-protein interactions to be established, several other factors must be in favor besides just the coordination between the iron ion and the heteroatom of the binding residue's side chain.[215,220]

A substantial amount of knowledge is currently available on how heme regulates the function of various proteins. The regulatory impact span from processes related to transcription and translation,[221,222] ion channel modulation,[223]coagulation cascade modulation,[224] to its interaction with the Aβ peptide that is related to Alzheimer's disease.[225] With the knowledge of mechanisms of how heme and its target proteins interact, a need to define sequence characteristics desirable to produce transient heme-protein interactions arose. Based on evidence that short protein-derived sequence stretches of ~9 residues were able to identify heme-binding behavior,[214,217,226,227] a combinatorial peptide library approach was pursued.[228]

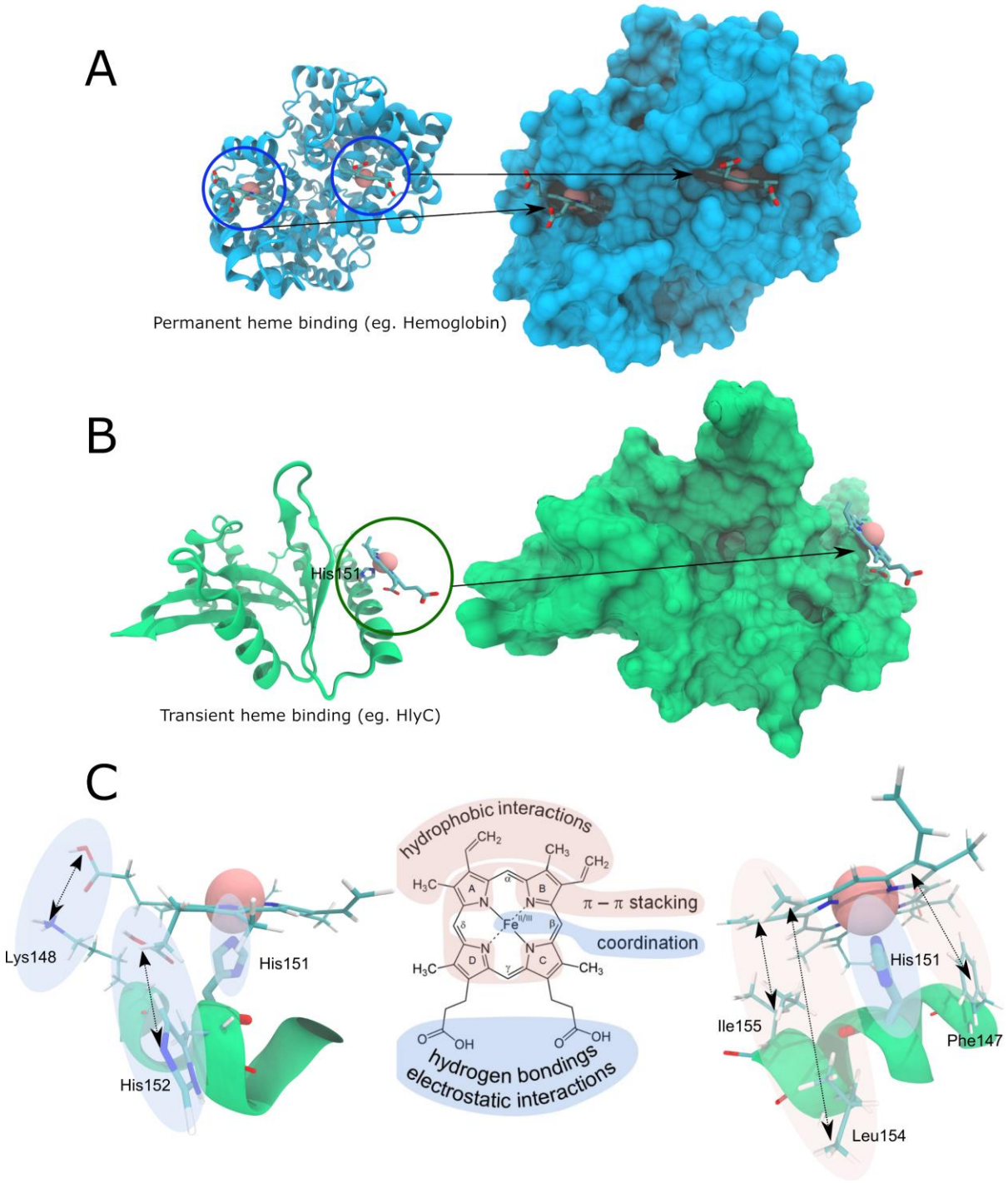


Figure 7. Permanent vs. transient heme binding. ***A)*** *A single subunit of the deoxy human hemoglobin (PDB ID: 1A3N*[229]*) with 2 heme molecules bound in the structure is shown as an example of permanent heme binding. The 2 heme molecules are circled in the structure (blue ribbon) and the surface representation on the right shows a clearer picture of how the heme molecules are buried into deep hydrophobic grooves/pockets in the protein.* ***B)*** *A homology modeled ribbon structure (green) of the protein hemolysin-activating lysine acyltransferase with heme bound to the surface exposed His151 residue is shown as an example of transient heme binding. The surface representation on the right clearly depicts that the heme molecule interacts with the protein on its surface and not in a hydrophobic pocket like in hemoglobin.* ***C)*** *The scheme in the center shows the different ways in which heme can interact with a protein. On the left, the structure of the HBM nonapeptide from HlyC is shown with the heme coordinating His151, as well as the residues that form hydrogen bonds with heme, namely, Lys148 and His152 shown as sticks. The interactions are highlighted by translucent ellipses matching the color coding in the central scheme. Similarly, the image on the right shows the hydrophobic interactions between heme and HlyC via Ile155 and Leu154, and π-π stacking interactions between Phe147 and heme. The STRIDE*[230] *and SURF*[231] *algorithms implemented in the program VMD (version 1.9.3),*[232] *are used to draw the ribbon and surface representations respectively. Figure components were adapted from literature.*[233,234]

This effort produced results that indicated an abundance of histidine and tyrosine based interactions (~40%) over cysteine based interactions (20%).[62] The findings thus far enabled the construction of a generic set of rules that a sequence must satisfy in order for it to qualify as a HBM. As a basic requirement, the sequence must contain at least one of the heme coordinating amino acids namely, Cys, His and Tyr. Next, the sequence must contain basic residues in its neighborhood that confer an overall net positive charge to the motif. Presence of a (Cys-Pro) CP motif is considered as a privilege and in this case the other conditions carry a lower weightage. This rule-based step-by-step validation is something that can easily be converted to an algorithm to predict transient heme binding motifs from large protein sequence datasets in an automated manner.[62] Current computational solutions such as TargetS[235] and HemeBind+[236] strictly rely on the availability of 3D structures to base their predictions on. Moreover all of these programs focus primarily on predicting permanent heme binding usually within a deep hydrophobic groove or pocket in the protein, and hence the current sequence characteristics validation approach seems to be an intriguing opportunity for the development of a new computational method, dedicated to the prediction of transient heme binding. Chapter V of this work showcases the successful implementation of such a solution.

2.4 *In silico* approaches to study protein folding and protein-ligand interactions

2.4.1 Molecular docking

Molecular docking, often just called docking is one of the most commonly used computational methods in structure-based drug discovery (SBDD).[237] The core idea of docking is to model and predict the interaction between a ligand (usually a small molecule but can also be peptides or proteins) and its receptor (the target protein).[238–240] Since the interaction of a ligand to its target largely defines a consequent underlying biochemical process, the docking process is considered as a critical one. Docking was first developed on the 'lock-and-key' theory proposed by Emil Fischer,[241] which defines the interaction between a ligand (the key) and its receptor (the lock) akin to that of a key fitting into its corresponding lock. However this approach considers both ligands and receptors as rigid bodies which is unfortunately not the case with biological molecules. Koshland refined this lock-and-key theory, reasoning that, the active site of the target protein is usually continually reshaped in the process of ligand binding and came up with the 'induced fit' theory.[242,243] The theory essentially proposes that, both the ligand and receptor should be considered as flexible entities in order to accurately characterize binding events. In practice though, considering both the ligand and receptor as fully flexible entities would be computationally inefficient and hence most docking experiments are done with a rigid receptor and a flexible ligand.[244–246]

Conceptually, the docking process involves two distinct steps. First it involves the prediction of the ligand conformations relative to the receptor by sampling different conformations of the ligand, resulting in a receptor-bound ligand conformation referred to as the bound 'pose' of the ligand.[247] The second step is called 'scoring' where multiple such obtained poses are ranked (using various methods) in order of predicted binding affinities.[247] Molecular docking experiments are conducted using two major approaches. First, in cases where the binding site of the ligand on the target protein is not known *a priori*, an approach called 'blind docking' is followed wherein the ligand is allowed to span the entire surface of the protein to detect possible binding sites.[248] This in turn is usually followed by closer examination of the best ranked binding sites and poses obtained. The next case, which is also the most common case in pharmaceutical applications, is when the binding site for a particular ligand on the receptor has been determined. In this case, docking is carried out with a restricted search space around the binding site to obtain finer details of the mode and nature of the binding.

During docking, sampling all six degrees of rotational and translational freedom and additionally the conformational degrees of freedom for both the ligand and protein could require astronomical number of computational steps to be performed, clearly deeming it infeasible for practice.[247] Various sampling algorithms have been developed and implemented in molecular docking software to overcome this issue. Matching algorithms using molecular shape as a basis, try to place the ligand at the active site of the protein by matching compatible shape features and chemical information, representing both the protein and ligand as pharmacophores.[249,250] The programs DOCK,[251] FLOG,[252] LibDock,[253] and SANDOCK[254] use this approach.

Docking packages such as DOCK 4.0,[255] FlexX,[245] Hammerhead,[256] SLIDE,[257] and eHiTS[258] use what is called as an 'incremental construction' method that first fragments the ligand and puts them at the receptor in an incremental fashion.

Finally, conformational sampling of ligand poses can also be carried out by stochastic methods. Two of the most important of these methods are the Monte Carlo method and genetic algorithms. Monte Carlo methods used bond rotation and rigid-body rotation or translation to generate ligand conformations.[247,259,260] The Monte Carlo sampling method is used in programs such as ICM,[261] QXP,[262] and Affinity (Accelrys Inc., San Diego, CA, USA).

Yet another class of stochastic method often used are genetic algorithms stemming from Darwin's theory of evolution, wherein, the degrees of conformational freedom of the ligand are encoded as 'genes' (in the form of binary strings) that are subjected to 'mutation' and 'crossover' operations, and the 'survival' of a gene is determined by a scoring function.[263,264] This method is used in programs such as AutoDock,[263] GOLD,[265] DIVALI,[266] and DARWIN.[267]

The second main component of docking, i.e., the scoring of various generated poses is done via a variety of different approaches. These range from classical force field based functions,[268,269] empirical scoring functions,[270,271] to knowledge-based scoring functions.[272,273]

It is therefore clear that molecular docking is a very powerful method that has gained a household name for itself in the field. Yet, results from docking simulations are often looked upon with skepticism, owing to its inherent flaws.[274] One of the main concerns with docking is the fact that, despite using methods that account for some amount of receptor flexibility,[275] docking does not fully represent the proteins flexibility which is detrimental to accuracy of the predicted results. The next section that discusses the method molecular dynamics (MD), is an efficient means of addressing this concern.

2.4.2 Molecular dynamics (MD) simulations

Molecular dynamics (MD) simulations are one of the most powerful and routinely used computational methods in chemical, biochemical, pharmaceutical and, materials research. In the pharmaceutical context, especially for structure-based drug discovery (SBDD), MD simulations have become an integral part of the process.[28,276] The simple yet, the most compelling reason as to why this is the case is since these simulations provide at an atomistic level, details on the dynamics and flexibility of the system that they simulate, which in most cases are protein-ligand systems.[277] Proteins like most biomolecules are not static structures and the function of most proteins is a direct consequence of their dynamics or the way they move. Structure determination methods such as X-ray crystallography that produce single static structures as their output cannot account for the underlying dynamics that dictate the function of the protein. Structures derived from NMR spectroscopic evaluation contain further detail on the protein's flexibility since NMR structures are usually published as an ensemble of an arbitrary number of low energy conformations (usually ~20). Here again, only the top 20 odd lowest energy conformations are made available. Hence in most studies that require a clear understanding of the protein dynamics, so that its functionality can be deduced, MD simulations play a crucial role.

MD simulations in practice start with the 3D structure of a protein or a protein-ligand complex. The protein is placed in a simulation cell, and a solvent, usually a well-tested model of water is added. Counter ions (usually Na^+/K^+ and Cl^-) to bring the net charge of the simulated system to zero, and optionally excess salt to achieve physiological conditions, are added. Protein folding and the stability of the folded protein arise from a combination of forces and interactions acting on the polypeptide chain namely, van der Waals interactions, salt bridges, hydrogen bonds and the hydrophobic effect.[120] In MD simulations, these forces

and their impact on the protein system are modeled by a mathematical energy function called the force field.[278,279] Several force fields exist for both proteins and nucleic acids, and the choice of a particular force field is a critical one which must be carefully done based on the system studied and the general aims of the investigation.[280–284] Though different force fields vary in the way they are derived and specifically what they model, i.e., whether they model an all-atom representation of the system or a coarse-grained one, all MD force fields have a common basis for modeling molecular motion. Every MD force field takes into account bonded interactions computing the energies from bond stretching, angle bending, as well as non-bonded interactions as the two main components of molecular motion and interaction. Non-bonded interactions are further divided into short range and long range interactions. Short range interactions caused by van der Waal's (VdW) forces are usually modeled by the Leonard-Jones potential while long range interactions are computed by a columbic potential. The initial configuration of the system to be simulated in accordance to the requirements of the force field is called the process of force field parameterization. This is one of the most important steps in the simulation protocol since any mistakes at this stage would result in either artifacts or in the worst case simulations with widely misleading results. Fortunately, most modern MD packages would fail to run simulations if discrepancies in the initial parametrization and topologies are encountered. Additional care must be taken when the simulation system involves non-standard amino acids, small molecule ligands or special molecules such as heme. The initial simulation setup and force field parameterization is followed by a process called energy minimization. Since, the states at which x-ray crystal structures or NMR ensembles are resolved might include high energy conformations, it is advisable to first energy minimize the system. Energy minimization essentially optimizes the potential energy of the system and brings the system to a local minimum. At the atomic level, the energy minimization process removes possible bumps and steric clashes in the starting structure. Two of the most commonly used algorithms to perform energy minimizations are the steepest descents and conjugate gradient algorithms.[285]

Once the system has achieved a low energy conformation, the solvent must be equilibrated to the target temperature and pressure values. Algorithms that simulate an external temperature bath are applied to maintain simulation temperature at the desired values while the pressure of the system is maintained by a scheme of rescaling atomic velocities and regular time intervals.[286–288] During the equilibration phase, the protein is usually held restrained while the solvent is allowed to 'soak' the protein sufficiently. Once the temperature and pressure of the system have been verified to fluctuate around the target value, the system is considered ready for the actual simulation usually referred to as the production phase of the simulation. The MD simulation algorithm in itself is a very simple one. It essentially entails iterating Newton's equation of motion over millions or billions of time steps with usually a 2 fs increment since atomic motion takes place in this range. Over the course of the simulation which is usually run for tens to hundreds of nanoseconds, the trail of atomic positions, forces and velocities are recorded and written to disc. This trail is referred to as the trajectory. The trajectory is what needs to be analyzed to obtain meaningful results from a simulation and is in a way the most important step of the process, since the entire simulation endeavor fails if the analysis of the simulated trajectory is not done correctly.[289–292]

Figure 8 shows a typical MD simulation workflow, starting from an initial structure, the MD algorithm, the generated conformational ensemble and two common measures of analysis of simulation data.

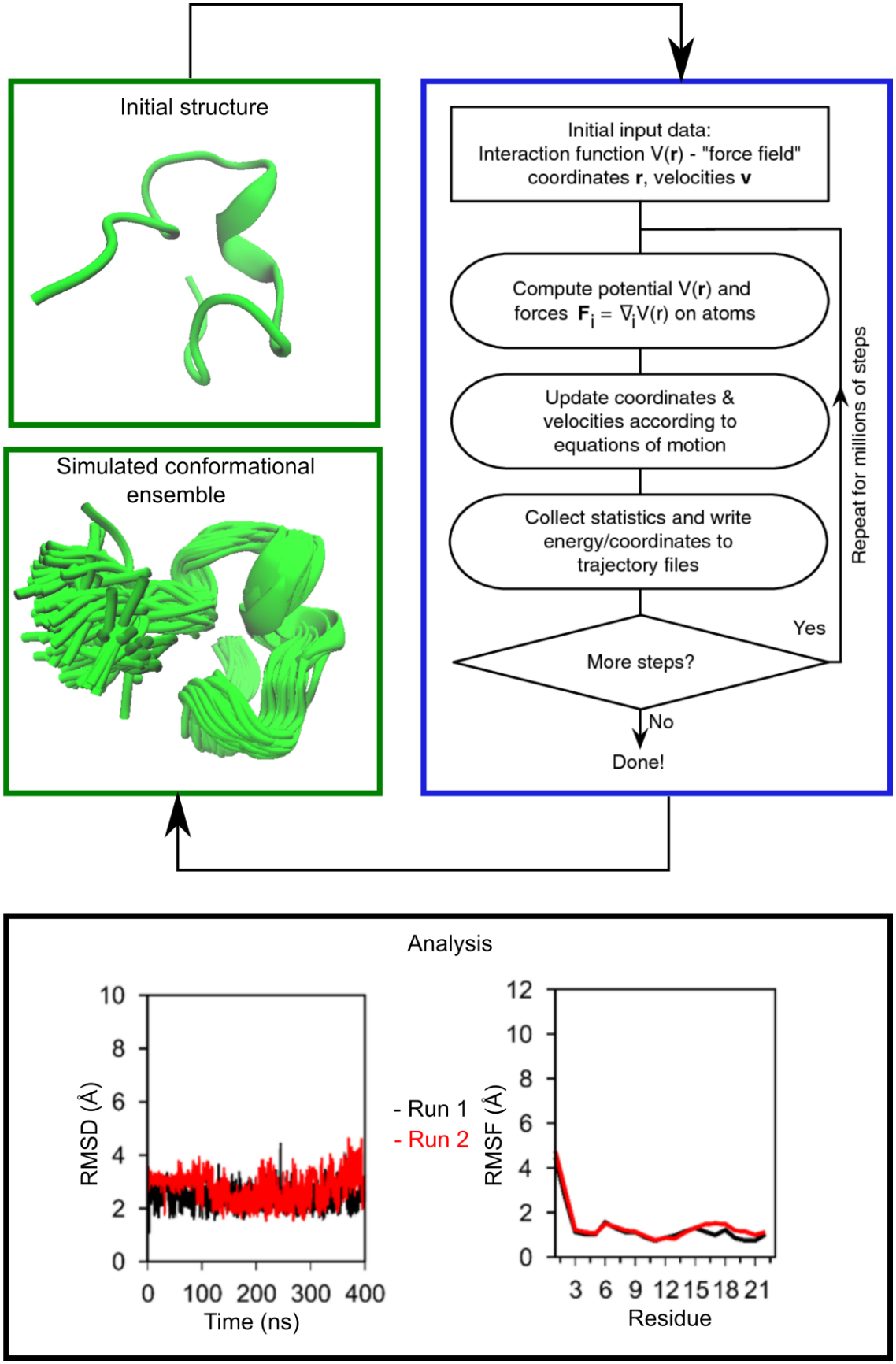


Figure 8. Molecular dynamics simulation workflow. *The figure represents the key components of a typical molecular dynamics simulation. The process starts with an initial peptide structure whose conformational dynamics is to be studied. This structure is fed as input the MD simulation program that runs a simulation for an user-defined amount of time. Upon completion of the simulation, a conformational ensemble can be visualized as*

well measures of protein structural change of the initial structure through the course of the simulation such as RMSD and RMSF and be analyzed and plotted. The example plots show RMSD and RMSF plotted from 2 independent MD simulations of the same peptide. Figure components adapted from literature.[293,294]

MD derived measures such as root mean square deviation of atomic positions (RMSD), root mean square fluctuations (RMSF) and radius of gyration (Rg) are some of the commonly computed quantities as they contain a wealth of information about the protein that has been simulated. Often times, the simulation trajectory can be visualized via a molecular visualization program to obtain visual insights into the simulated system. MD simulations have become ubiquitous in biochemical research and are an essential component of any industrial drug discovery pipeline. The computational chemistry community in recent years, has taken a great interest in teaching best practices in conducting MD simulations and related analyses.[295–297] In the current work MD simulations have been used to simulate disulfide-bonded peptides to observe folding/refolding in partially unfolded states induced by the rupture of a certain disulfide bond in the starting structure of the simulation. Though, it is not possible to dynamically simulate the rupture and formation of disulfide bonds via conventional MD simulations, underlying folding mechanisms can be deduced. Finally, MD simulations are also applied to assess the quality of heme-protein complexes that result from molecular docking simulations. As docking simulations treat the receptor protein as a rigid structure, MD-based refinement of the docked complex is almost always necessary to obtain reliable results.

3 Thesis outline

The aim of this thesis is to showcase how computational methods encompassing the fields of molecular modeling, computational chemistry and bioinformatics can be successfully employed in conjunction with experimental methods to produce meaningful insights in biochemical research. From a biochemical perspective, this thesis focusses on addressing protein folding, more precisely on the oxidative folding of multiple-disulfide-bonded peptides and proteins that often form the basis of peptide-based drug development.[42,47] Oxidative folding of such multiple-disulfide-bonded peptides and proteins have till date been defined by two extreme and opposite folding frameworks developed from studying the folding of two proteins namely bovine pancreatic trypsin inhibitor (BPTI) and hirudin.[54,298] In short, BPTI is said to fold in a manner wherein, its folding intermediates comprise of partially folded species with native disulfide linkages and native-like secondary structures that eventually efficiently drive the folding toward the correctly folded native isomer.[54] On the other hand, hirudin experiences an initial collapse into a molten globule with several non-native disulfide bonded species appearing at various stages of its folding. This in turn, results in a higher probability of alternative disulfide-bonded isomers being formed along with the native isomer defined by the lowest potential energy in the folded state.[54] The oxidative folding problem is studied in this work using conotoxins, a large group of small multiple-disulfide-bonded peptides produced by cone snails of the genus *Conus* that target various ion channels and membrane proteins with impressive potency and specificity.[50,180] The properties of conotoxins including their small, constrained structures and their potency, specificity and selectivity.[46,50,52] Conotoxins are yet to be extensively investigated to specifically determine the oxidative folding behavior relative to the standard BPTI and hirudin framework. In this study, a group of μ-conotoxins from the M superfamily are considered for this purpose. They are investigated using molecular dynamics (MD) simulations to understand the underlying folding mechanism, since earlier studies suggest that even within a set of conotoxins with identical disulfide-bonding patterns, differing oxidative folding behaviors were observed.[56] Additionally, using the fifteen theoretically possible disulfide isomers of the three-disulfide-bonded μ-conotoxin PIIIA made available as 12 NMR structures and 3 computer-generated models,[152] MD and docking based studies are conducted to investigate how the 3D structure induced by each of the distinct disulfide pairings influence the structure and consequently the bioactivity of these peptides. Finally. similar studies using tridegin,[53,203,299] a leech derived peptide inhibitor of the blood coagulation factor FXIIIa, are conducted to elucidate the impact of its disulfide bonds. MD and docking based studies done on three identified disulfide isomers[203] attempt to explain how disulfide bonds and the lack of one (of the three) of them affect the structure, dynamics and consequently the function.

In the second part of this thesis, specific ligand-protein interactions, namely the transient interaction of heme with proteins which is suggested to impart a regulatory effect on the target protein, is studied.[209] Transient heme binding is distinguished from the more commonly known insertion of heme as a prosthetic group in hemoproteins, such as hemoglobin and myoglobin, where heme is bound to a deep hydrophobic groove acting as a cofactor. In the case of transient binding, heme associates with the surface of proteins via sequence motifs called heme-binding motifs (HBMs)[212] and exerts a regulatory function on the protein. Herein, an example of such an interaction between heme and the protein hemolysin-activating lysine acyltransferase (HlyC), the inhibition of whose enzymatic activity is desired to curb hemolysis induced by uropathogenic bacteria, is investigated. Homology modeling, molecular docking and molecular dynamics simulations are pursued to achieve this aim. Finally, the growing need for a comprehensive assessment and summarization of

the current knowledge on transient heme binding is addressed with a computational solution that enables the rapid and efficient identification of transient heme-binding motifs from large sequence datasets is programmatically implemented and presented.

The thesis is divided into five separate chapters three of which (part 1) deal with the application of biomolecular simulation approaches such as molecular docking and MD simulation to study different multiple-disulfide-bonded peptides in each case, while the last two chapters (part 2) deal with the computational elucidation of transient heme-protein interactions using both simulation and programmatic approaches. A short summary of each chapter is given below.

Part 1

Chapter I studies five µ-conotoxins namely µ-GIIIA, µ-KIIIA, µ-PIIIA, µ-SIIIA and µ-SmIIIA sharing identical disulfide bonding patterns. Using MD simulations applied via a novel approach of *in silico* disulfide bond removal, the peptides are analyzed to provide an explanation to the difference in their experimentally observed folding behavior. The outcome of this study is a ranking of these 3-disulfide-bonded peptides based on 2-disulfide bonded structural stability as well as an evaluation of their individual places between the established extreme and opposite BPTI and hirudin oxidative folding models.

Chapter II studies the fifteen theoretically possible disulfide isomers of the conotoxin µ-PIIIA using MD simulations and molecular docking studies to explain at a molecular and structural basis, the differences in the bioactivity observed when the fifteen isomers were tested for the blocking activity at their target VGSC Nav1.4. Finer details of the interaction between selected µ-PIIIA isomers and the VGSC Nav1.4 are obtained from extensive all-atom MD simulations conducted on the isomer-channel docked complexes.

Chapter III yet again uses MD simulations and molecular docking based analysis to study three disulfide isomers of the leech derived, 3-disulfide-bonded inhibitor peptide, tridegin. MD simulations and docking studies attempt to explain as to how 2-disulfide-bonded variants of the three isomers of tridegin still were able to inhibit their target protein, the human coagulation factor FXIIIa at levels comparable to their 3-disulfide-bonded counterparts.

Part 2

Chapter IV uses molecular docking followed by MD simulations to prove the existence of transient binding of heme to a predicted nonapeptide HBM on the C-terminal helix of the protein hemolysin-activating lysine acyltransferase (HlyC) as predicted by the accompanying experimental studies.

Chapter V is a review on the current knowledge on the understated side of heme-related research namely, research into the role of heme as a regulatory/effector molecule. Based on knowledge gathered from over the years, a consensus on sequence requirements for transient heme binding is first established. This consensus is transformed into an algorithm that is implemented in the python programming language resulting in a program SeqD-HBM for the sequence-based detection of transient heme-binding motifs from sequence data.

The final section of the thesis is a summary of conclusions from the individual chapters and a holistic take home message from the work as a whole.

4 Manuscripts

Part 1

4.1 Chapter I

Insights into the Folding of Disulfide-Rich μ-Conotoxins

Original research article

Authors*

Ajay Abisheck Paul George, Pascal Heimer, Astrid Maaß, Jan Hamaekers, Martin Hofmann-Apitius, Arijit Biswas,and Diana Imhof

This peer-reviewed research article was published in *ACS Omega*.

Citation

ACS Omega 2018, 3, 12330−12340

DOI: 10.1021/acsomega.8b01465

4.1.1 Introduction

μ-Conotoxins are known to be peptides of large pharmaceutical interest owing to their ability to specifically and potently bind to their VGSC targets. However, not much is known about the folding mechanism and the underlying pathways of oxidative folding of several of these peptides. To address this concern, in this study five μ-conotoxins namely μ-GIIIA, μ-KIIIA, μ-PIIIA, μ-SIIIA and μ-SmIIIA sharing identical disulfide bonding patterns (C1-C4/C2-C5/C3-C6) are subjected to molecular dynamics based investigation. A novel approach is taken wherein, selected native disulfide bonds in these peptides are systematically removed computationally and the partially unfolded versions are observed under simulation. The folding and refolding behavior of the simulated, partially folded peptides, is compared to experimental results to assign each of the peptides in positions relative to the opposite extreme models of oxidative folding defined by the folding of proteins BPTI and hirudin.

***Own contribution**

The computational study including MD simulations and analysis of the results was performed by myself. The design and the content of the manuscript has been compiled by Diana Imhof and me. The layout and preparation of the figures has been planned and carried out by me in consultation and agreement with Diana Imhof.

ACS OMEGA

Article

Cite This: ACS Omega 2018, 3, 12330−12340

http://pubs.acs.org/journal/acsodf

Insights into the Folding of Disulfide-Rich μ-Conotoxins

Ajay Abisheck Paul George,[†] Pascal Heimer,[†,#] Astrid Maaß,[‡] Jan Hamaekers,[‡] Martin Hofmann-Apitius,[§,∥] Arijit Biswas,[⊥] and Diana Imhof*[,†]

[†]Pharmaceutical Biochemistry and Bioanalytics, Pharmaceutical Institute, University of Bonn, An der Immenburg 4, D-53121 Bonn, Germany

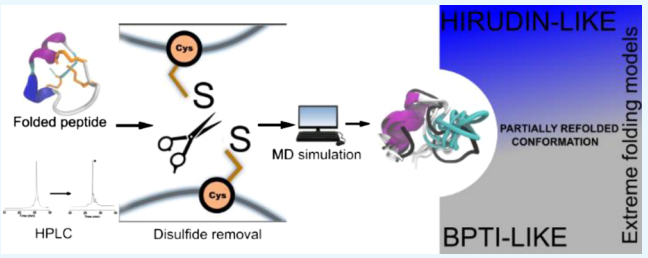
[‡]Department of Virtual Material Design and [§]Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing, Schloss Birlinghoven, D-53754 Sankt Augustin, Germany

[∥]Bonn-Aachen International Center for Information Technology, University of Bonn, Endenicher Allee 19 C, D-53115 Bonn, Germany

[⊥]Institute for Experimental Hematology, University Hospital Bonn, Sigmund-Freud-Straße 25, D-53127 Bonn, Germany

Supporting Information

ABSTRACT: The study of protein conformations using molecular dynamics (MD) simulations has been in place for decades. A major contribution to the structural stability and native conformation of a protein is made by the primary sequence and disulfide bonds formed during the folding process. Here, we investigated μ-conotoxins GIIIA, KIIIA, PIIIA, SIIIA, and SmIIIA as model peptides possessing three disulfide bonds. Their NMR structures were used for MD simulations in a novel approach studying the conformations between the folded and the unfolded states by systematically breaking the distinct disulfide bonds and monitoring the conformational stability of the peptides. As an outcome, the use of a combination of the existing knowledge and results from the simulations to classify the studied peptides within the extreme models of disulfide folding pathways, namely the bovine pancreatic trypsin inhibitor pathway and the hirudin pathway, is demonstrated. Recommendations for the design and synthesis of cysteine-rich peptides with a reduced number of disulfide bonds conclude the study.



INTRODUCTION

Conotoxins are neuropeptides from the venom of marine cone snails, which interact with a wide range of biological targets (e.g., ion channels, transmembrane receptors, and transporters) and hence are of pharmaceutical interest and of great potential as molecular probes to study the specific subtypes of ion channels and receptors.[1,2] Conotoxins consist of approximately 10−50 amino acid residues and are classified according to their cysteine patterns.[3,4] The typical CC−C−C−CC pattern defines the framework for conotoxins of the M-superfamily comprised by ψ-, μ-, and κM-conotoxins.[3,5,6] The family of the 27 currently known μ-conotoxins[4,7] selectively binds to the ion channel pore of the voltage-gated sodium channels, thus blocking the influx of sodium ions into the cell. Therefore, these peptides also gained interest as useful tools for research studies in electrophysiology.[8−10] μ-Conotoxins are cysteine-rich peptides consisting of 6 cysteines which can give rise to 15 conformational isomers of different disulfide connectivities in various combinations of disulfide bonds.[10] However, the dominant isomer among them bears the disulfide linkage of C1−C4, C2−C5, and C3−C8, often referred to as the "native fold".[10] The molecular principles underlying the folding bias contributing to one particular dominant isomer is as yet unclear. This subject is of great importance because an accumulation of cysteines may also occur in distinct regions of larger peptides and proteins such as that observed in defensins,[11,12] resistins,[13] Kunitz serine protease inhibitors,[14] and various growth factors.[13,15,16] In such cases, any information on their preferred disulfide connectivity would be relevant especially in the absence of actual chemical or structural data.

Conotoxins represent a promising tool for studying the impact of disulfide bonds on the folding process owing to their small to medium size which makes them intermediates between peptides and proteins and also their high disulfide bond content.[3,5,17] We studied the μ-conotoxins GIIIA, KIIIA, PIIIA, SIIIA, and SmIIIA (Figure 1) using unbiased all-atom molecular dynamics (MD) simulations performed on their NMR structures. The simulation output was correlated with the in vitro data from the oxidation reaction for all the μ-conotoxins that had been described earlier.[8−10] The individual μ-conotoxins could be grouped based on the reaction product yield and the side product formation,[18] which allowed for a functional comparison with the computational study. For the

Received: June 27, 2018
Accepted: September 12, 2018
Published: October 1, 2018

DOI: 10.1021/acsomega.8b01465
ACS Omega 2018, 3, 12330−12340

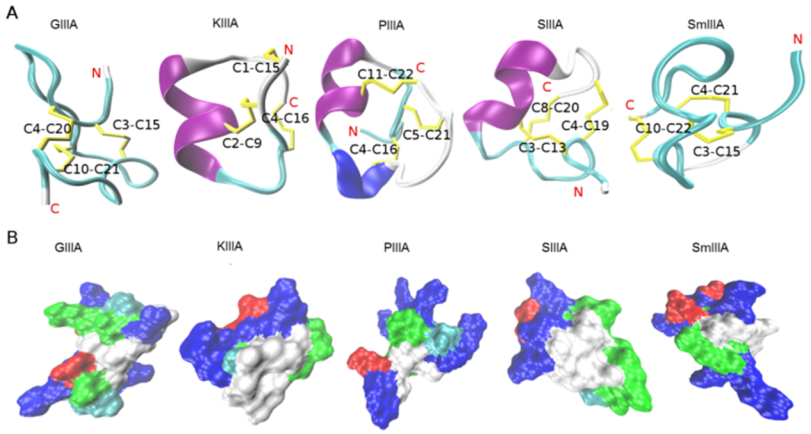


Figure 1. Structure and surface representations of the investigated μ-conotoxins. (A) NMR structures of the five μ-conotoxins (GIIIA, KIIIA, PIIIA, SIIIA, and SmIIIA) used in this study represented as a cartoon. The secondary structure elements, α-helix (purple), 3−10 helix (blue), turn (cyan), and coil (white) were generated by STRIDE[20] in visual molecular dynamics (VMD). The cysteine residues forming the disulfide bonds (yellow) were labeled. (B) Molecular surface was generated by SURF[21] in VMD, indicating the hydrophobic (white), basic (blue), acidic (red), and hydrophilic (green) regions. All structures were taken from the ConoServer database.[7,22]

Table 1. Comparison of Sequence Characteristics of the μ-Conotoxins Investigated in This Study[a]

μ-Cono-toxin	Sequence	Net Charge	Structure Code (pdb)/ Structure Card[4]	Bioactivity on $Na_V1.4$[#]
GIIIA	-RDCCTOOKKCKDRQCKOQ-RCCA*	+5	S00082 (1TCG) [23]	19 +/- 1 nM[25]
KIIIA	---CCN----CSSKWCRDHSRCC-*	+3	S00129 (2LXG) [26]	90 +/- 17 nM[27]
PIIIA	ZRLCCGFOKSCRSRQCKOH-RCC-*	+6	S00159 [28]	36 +/- 8 nM[25]
SIIIA	-ZNCCNG--GCSSKWCRDHARCC-*	+2	S00125 (BMRB 20025) [29]	130 +/- 15 nM[29]
SmIIIA	-ZRCCNGRRGCSSRWCRDHSRCC-*	+5	S00077 (1Q2J) [30]	0.22 +/- 1 nM[25]

[a]Residues are highlighted according to their character: basic (blue), acidic (red), polar uncharged (green), and cysteine (yellow) (Z: pyroglutamate, O: 4-hydroxyproline). *All peptides were used as amides. Native μ-SmIIIA occurs as the C-terminal acid, however, is usually used as an amide.[8,9] We used μ-SmIIIA as an amide for reasons of comparison because both structures were found to be identical.[31] [#]In general, the IC_{50} values determined for the toxins ion channel blocking activity at the skeletal muscle $Na_V1.4$ expressed in *Xenopus* oocytes are given, with the exception of SmIIIA where only K_D is available.

MD studies, a rather unconventional approach was pursued by analyzing the process of refolding; that is, the behavior and stability of an individual peptide was observed upon successive opening of disulfide bonds in the folded peptide NMR structures. The analyses conducted on the resulting MD trajectories gave rise to inferences on characteristic factors contributing to the conformational stability of the folded conopeptides. On the other hand, we used the distances sampled by the sulfur atoms between the reduced cysteine residues to observe and determine if the peptide with a disulfide bond removed in silico tries to refold or not. This enabled drawing a distinction between the influence of the disulfide bond and the properties exerted by the rest of the sequence to the maintenance of close-to-native backbone conformations in peptides with computationally reduced disulfide bonds. This consequently simulates the state at which the peptide is at its final stage of folding and in its near-native conformation.

The pathways for disulfide folding have been classified into two extreme models despite exhibiting a high degree of diversity.[19] One model is represented by the bovine pancreatic trypsin inhibitor (BPTI)-like folding, where there is a predominance of native intermediates at various steps down the folding funnel, and the other extreme, the hirudin pathway, is defined by highly heterogeneous non-native intermediates. Interestingly, conotoxins are so far placed in between these two extreme models in a hybrid BPTI−hirudin model.[19] In the present study, we thus try to improve the clarity regarding the classification of the peptides used herein considering the existing models. Finally, an attempt to relate the simulation inferences from these peptides with open disulfide bonds to their propensity (or) favorability to retain close-to-native conformations is presented via a qualitative grouping.

RESULTS AND DISCUSSION

Studies of conotoxin folding include, in general, experimental approaches such as regioselective oxidation strategies, spontaneous oxidative self-folding, and optimization of the folding methods tested, recombinant expression, and so far available information from the biosynthesis of conotoxins in the venom duct of cone snails. However, it is still unresolved how exactly the cone snails produce properly folded peptide and protein toxins. In order to approach the mechanisms behind conotoxin folding, several studies focused on using computational strategies to provide valuable insights into the structural features important for the folding process, but often a combination of both experimental and theoretical work is

DOI: 10.1021/acsomega.8b01465
ACS Omega 2018, 3, 12330−12340

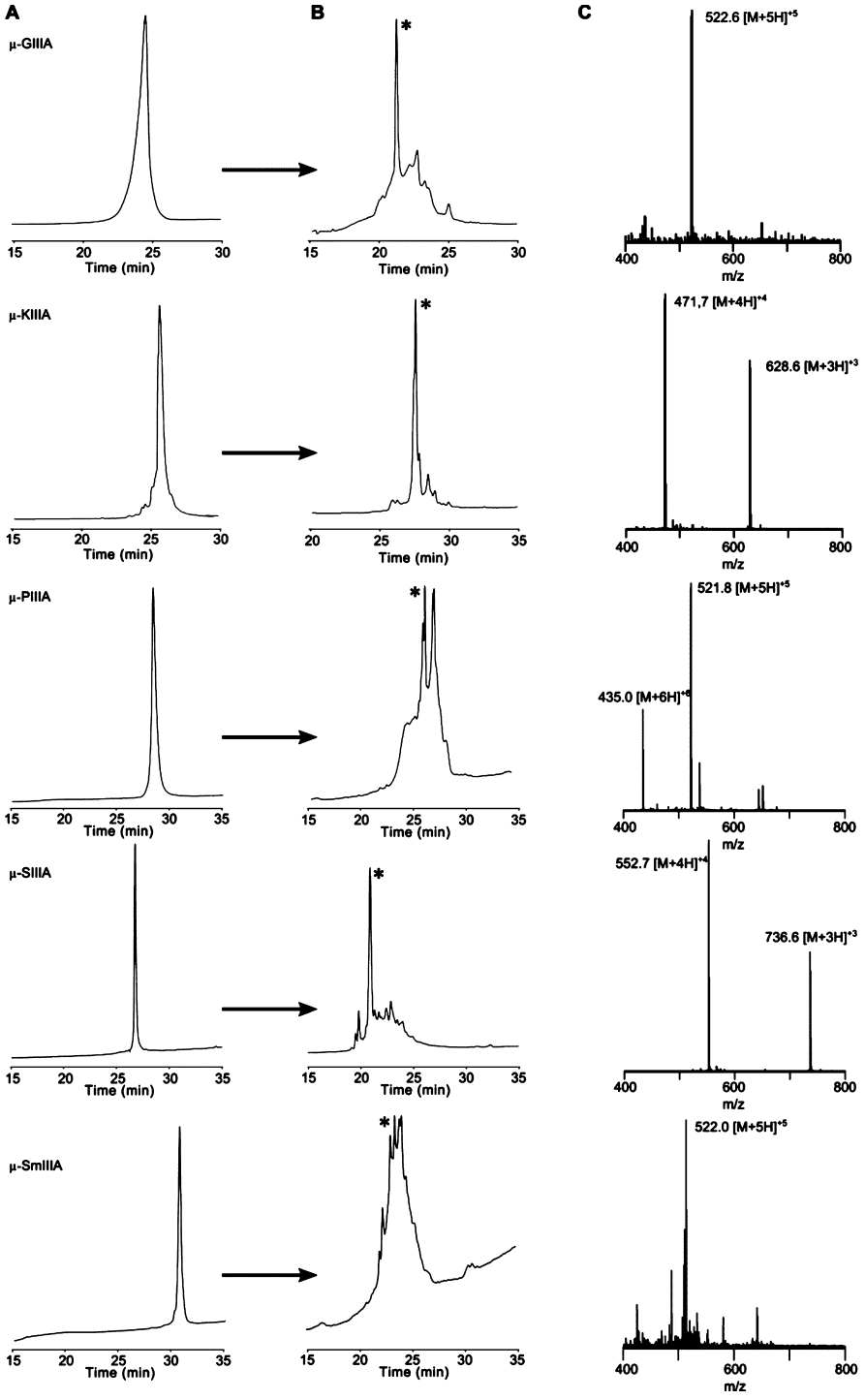


Figure 2. High-performance liquid chromatography (HPLC) elution profiles of μ-conotoxins studied. Linear, reduced precursors of μ-conotoxins (A) and folded crude mixtures after 1 h (B), with the main product marked with an asterisk. Respective electrospray ionization (ESI) mass spectra of the oxidized peptides are shown in (C).

missing or only one conotoxin was in the focus of most of these studies.

Oxidative Self-Folding of μ-Conotoxins. In order to provide the in-house experimental data for comparison with the computational results, equal amounts of each of the selected five μ-conotoxins (Figure 1, Table 1) were used for the oxidation reaction according to a protocol earlier described yielding an undirected and spontaneous formation of disulfide bonds.[8−10] Under the conditions applied, oxidation of μ-SmIIIA and μ-PIIIA resulted in several peaks of fully oxidized product(s) as confirmed by mass spectrometry; that is, the disulfide connectivity of the individual fractions is different from the native (major) fold.[10] In contrast, oxidation of all other peptides, that is, μ-GIIIA, μ-KIIIA, and μ-SIIIA, resulted in one major product confirming earlier reports (Figure 2).[18,23,24]

The formation of one major product for μ-GIIIA and μ-SIIIA can be explained by a rapid collapse into the favored

DOI: 10.1021/acsomega.8b01465
ACS Omega 2018, 3, 12330−12340

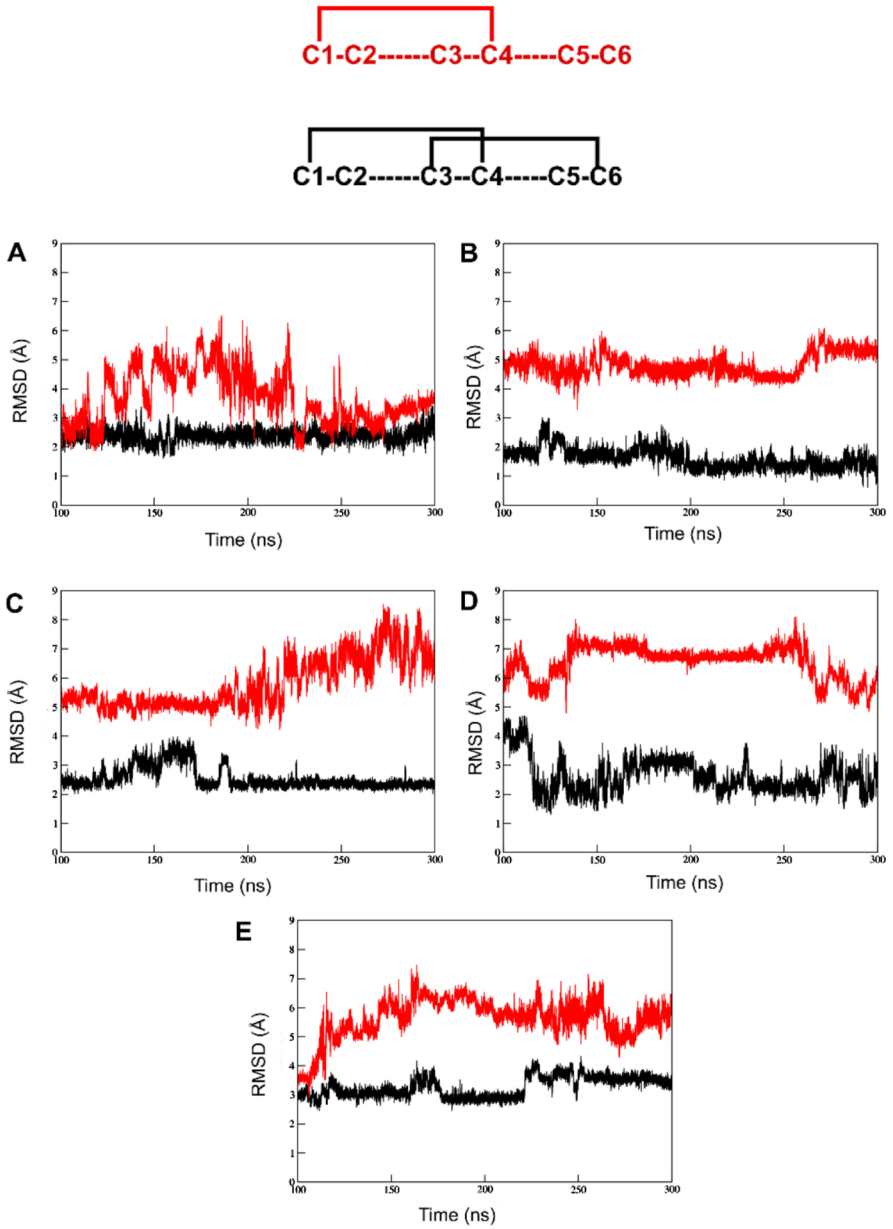


Figure 3. RMSD plots of disulfide bond opened versions of the five μ-conotoxins. The comparison of backbone stability between the peptides with the C2−C5 disulfide bond removed (black) and both the C2−C5 and C3−C6 bridges removed (red) between 100 and 300 ns of simulation time: (A) μ-GIIIA, (B) μ-KIIIA, (C) μ-PIIIA, (D) μ-SIIIA, and (E) μ-SmIIIA. Above the plots is a representation of two cases of disulfide connectivity discussed. Here, red represents the version with the single C1−C4 disulfide bond and black represents the C1−C4/C3−C6 disulfide connectivity.

native fold with the disulfide connectivity C1−C4/C2−C5/C3−C6 as introduced earlier.[18] For μ-KIIIA, such a rapid collapse also results in one main product; however, the connectivity C1−C5/C2−C4/C3−C6 is preferred here, and the pattern C1−C4/C2−C5/C3−C6 is only present as a minor fraction. Small differences in the elution profiles of buffer-oxidized μ-KIIIA (the crude product), that is, product formation, might result from the differences in batch size and composition of the oxidation buffer compared to the results of Khoo et al.[26] In the case of μ-PIIIA and μ-SmIIIA, a different folding mechanism indicative of a slower rearrangement results in the formation of several isomers. This can be attributed to more diverse noncovalent interactions and electrostatic forces compared to μ-GIIIA and μ-SIIIA. Here, it was suggested that the native isomer accumulates via reshuffling of disulfide bonds during the folding process and is dependent on the thermodynamic stability of the isomer formed.[18,32] Although μ-SIIIA and μ-SmIIIA have a high sequence similarity (Table 1), the higher number of basic residues in μ-SmIIIA (six Arg) might cause the formation of multiple isomers, whereas in case of μ-SIIIA (two Arg and one Lys), only one isomer is preferred.[18,33] μ-PIIIA forms multiple isomers possibly for the same reason.[10] In contrast, the structure of μ-GIIIA tolerates a high number of basic residues and forms only one major product compared to that of μ-PIIIA and μ-SmIIIA.[18] Several aspects influencing the folding, such as the number of hydroxyproline residues, the loop size between the two linked cysteines, or amidation of the C-terminus, are discussed controversially in the literature without a clear preference indicating the uniqueness of each sequence and the respective biological activity.[18,29,34,35]

DOI: 10.1021/acsomega.8b01465
ACS Omega 2018, 3, 12330−12340

Conformational Analysis Using Molecular Dynamics. The solution NMR structures used herein as initial conformations for starting MD simulations were derived from the ConoServer database (Figure 1, Table 1).[4,7] Comparison, alignment, and structural differences of the μ-conotoxin structures were discussed previously by Yao et al. and Tietze et al.[10,29] All μ-conotoxins, except μ-KIIIA, possess conserved structures that align significantly better in the C-terminal part compared to the N-terminal region.[29] A high similarity was seen for the backbone conformations between the loop 2 and loop 3 regions of μ-KIIIA, μ-SIIIA, and μ-SmIIIA, which have a higher selectivity for blocking $Na_V1.2$ over $Na_V1.4$ channels (Table 1). On the other hand, μ-GIIIA and μ-PIIIA, which prefer $Na_V1.4$ over $Na_V1.2$, superimpose well in the second loop between C2 and C5.[10] Besides contributing to the structural rigidity, the disulfide bridges cause cysteine residues to form a hydrophobic core, enveloped by other charged and hydrophilic residues (Figure 1). This hydrophobic effect plays a key role in the stability of the native fold.[36]

With three disulfide bridges present, there are six ways to open them one by one. The order of disulfide bond formation during the synthesis (and also in vivo) is not known, but as mentioned before, the process appears to be guided by thermodynamic aspects. Regarding the simulation strategy, it was decided to open the longest disulfide bridge first (bridging the longest sequence in between, see Table S1) as it is expected to instantly introduce the highest level of flexibility into the peptide backbone. The intention for this opening strategy was to increase the conformational entropy of the reduced version. In the case of μ-KIIIA (non-native connectivity), the shortest disulfide bridge was opened first. This would serve as a means to gauge the effect of loop size in retaining a stable structure. The loops in between the three disulfide bonds differ in size (Table S1) and thus needed to be considered in the analysis process.

It was observed during the simulations at room temperature that all peptides retained their initial conformation as demonstrated by the root-mean-square deviation (RMSD) of Cα atoms compared to the chosen starting NMR structure of each peptide (Table 1) with μ-KIIIA possessing the lowest RMSD of 1.1 Å and μ-SmIIIA the highest of 2.7 Å (Figure S1, Table S2). The RMSD of the Cα atoms, the root mean square of fluctuation (RMSF) of all atoms of each residue, and the radius of gyration (Rg) of the whole protein were computed from each simulation for all of the five peptides (Figure S1, Table S2).

The native structure of μ-GIIIA displayed a hydrophobic core formed by the cysteines and a salt bridge between R1 and D12. As the first disulfide bond C2−C5 was removed, the distance between the now unbound cysteines increased moderately. However, this did not affect the overall three-dimensional conformation of the backbone proved by only a 0.3 Å increase in the backbone RMSD, but it could be observed that its RMSD progression is inconsistent in the simulation especially between 30 and 50 ns. The 300 ns simulation showed that the peptide had a stable backbone indicated by a relatively unwavering RMSD curve between 100 and 300 ns of simulation time as shown in Figure 3.

Importantly, the residues significant for bioactivity experienced none to a very minimal increase in fluctuations (Figure 3, Table S4). Upon subsequent removal of the second disulfide bridge between C3−C6, an obvious stretch in the overall shape of the peptide was observed (Figure S3). The peptide had an excessive flexibility and adopted close to completely unfolded conformations during different intervals in the simulation (Figure S3). Interestingly, the cysteines forming the C3−C6 bond moved much closer to each other than the C2−C5 cysteine residues in the two disulfide bond opened structure. This is shown by the decrease in the RMSD of the C2−C5 and C3−C6 removed peptide between 60 and 70 ns (Figure S3). In longer time scales (300 ns), it was observed that the RMSD values dropped close to the ones with just one disulfide bond (C2−C5) removed, indicating that μ-GIIIA tends to fold back to retain its preferred native conformation (Figure 3). The Rg of μ-GIIIA followed an almost identical pattern of progression to the RMSD, peaking between 40 and 65 ns before falling back toward its initial values, representing an unfolding-(re)folding event. The fluctuations of the residues K8, K11, R13, K16, and R19, which were reported to be responsible for ion channel binding,[3] did not show a significant change compared to the structure with all disulfide bonds intact (Table S3). With two disulfide bridges removed, the C-terminus including R19 and K16 showed a significant increase in fluctuations. It has been reported that K16 has a low priority for biological activity. Moreover, the exchange of this residue increased the binding affinity compared to the native toxin.[37] Our findings suggest that if the C2−C5 disulfide bond alone was removed, the structure may adopt a conformation still representing a structure close to the native fold.

μ-KIIIA has the lowest RMSD of 1.1 Å on average with respect to the chosen starting structure among the five peptides in their natively folded form (Table S2). In 100 ns of simulation time, the removal of the C2−C4 disulfide bond, thereby altering the C4 residue, results in the loss of its helix and increases the RMSD by 0.7 Å. However, at the 300 ns time scale, conformations sampled by μ-KIIIA showed the reappearance of its native helix (Figure 5). The remaining C1−C5 and C3−C6 disulfide bonds were sufficient to retain the backbone stability and conformation, respectively, of μ-KIIIA. The progression of RMSD and Rg for μ-KIIIA followed the same scheme as observed for μ-GIIIA, that is, with the structure possessing two reduced disulfides revealing the largest variation in the conformational flexibility. It was observed from the 300 ns simulations that unlike μ-GIIIA, the two disulfide-deficient versions of μ-KIIIA did not regain the backbone stability of its one disulfide bond removed counterpart as shown in Figure 3.

Although the average RMSD of native μ-PIIIA was a decent 2.3 Å compared to the selected starting structure over the course of the simulation, the backbone was constantly subjected to changes as can be seen from the RMSD plot (Figure S1). The removal of the C2−C5 disulfide bond resulted in a lesser fluctuating RMSD progression, although it came at the expense of a 1.5 Å increase in RMSD in the first 100 ns of simulation time. Conformations sampled by this peptide between 100 and 300 ns of simulation time showed the reappearance of its native helices (α-helix between O8 and S14 and 3_10 helix between L3 and C5). The structure of μ-PIIIA with both the C2−C5 and C3−C6 disulfide bonds removed showed the largest extent of structural variation among the five conopeptides with an average RMSD of 4.9 Å (Figure S5). The peptide did not tend to refold within 300 ns simulation time. With one disulfide bond removed, none of its functionally significant residues showed a pronounced increase in fluctuations, the highest of which was a 1.3 Å increase for

DOI: 10.1021/acsomega.8b01465
ACS Omega 2018, 3, 12330−12340

R14 compared to the natively folded peptides. Unlike μ-GIIIA, the two disulfide-deficient versions of μ-PIIIA did not regain the backbone stability of its one disulfide bond removed counterpart. Meanwhile, the single disulfide-deficient version of μ-PIIIA adopted a very stable conformation, with the RMSD curve almost flat lining between 200 and 300 ns of the 300 ns simulation (Figure S3).

In contrast, almost all residues of μ-SIIIA displayed marginally higher RMSF values for the one and two bond-removed structures in comparison to the native structure containing all the three disulfide bonds. In the one disulfide-deficient version of μ-SIIIA, although the marginal increase in residue fluctuations was observed, the functionally significant restudies W12, R14, and H16 revealed only a minimal increase in the overall residue fluctuations. Another key residue R18, however, showed a larger mobility between the native and one disulfide-deficient versions. More importantly, the one disulfide-deficient version of μ-SIIIA retained its α-helical motif between K11 and H16, which is a significant aspect in terms of targeting sodium channels.[38]

We focused further on μ-GIIIA that forms a single oxidation product and μ-SmIIIA that forms multiple oxidation products during the synthesis to illustrate the phenomenon described above (Figure S3). The overall RMSD between the native connectivity and the structure with one disulfide bond removed was observed to be low in μ-GIIIA (2.1 Å) and high in μ-SmIIIA (2.9 Å) among the five peptides. Figure S3 shows average conformations for all the three 100 ns simulations of μ-GIIIA and μ-SmIIIA compared with their corresponding RMSD plots. An interesting observation from both peptides with an opened disulfide bond was the formation of new secondary structure elements that were not present in the native state. The structure of C2−C5-deficient μ-GIIIA achieved a reasonable equilibration between 200 and 300 ns of the simulation, and the inspection of the trajectory revealed that the conformations sampled by this peptide had a 3_10 helix between K16 and Q18 (Figure 3). In μ-SmIIIA with one opened disulfide bond, an α-helix was formed between residues R13 and H18 (Figure 2).

The stability of this helix through the entire course of the simulation can be accounted for by a combination of hydrogen bond formation and the presence of the bonded C15 in the center of the helix. In comparison with the native fold, the structure with the C2−C5 bond opened appeared well ordered. The distribution of the hydrogen bonds around the helix can be seen in Figure 4.

The μ-SmIIIA structure with two disulfide bonds reduced formed a less stable 3_10 helix between residues R16 and H18, increasing the flexibility of the conformation. However, the functionality of a conotoxin is dependent on a stable backbone structure coupled with the favorable orientation of basic side chain residues for binding to their target and not solely on the flexibility of a distinct part of the peptide.[3] In μ-SmIIIA, though the 3_10 helix formation reduced the Rg of the peptide, the overall peptide structure drifted significantly from that of the native state, and the orientations adopted by the side chains of its basic residues varied largely when compared to either the native or the structure with one disulfide bond opened. Even with a single disulfide bond removed, the functionally significant R7 showed a moderate increase in fluctuations (Figure 4, Table S4).

The 300 ns simulations for the disulfide-deficient species of both μ-SIIIA and μ-SmIIIA revealed similar trends based on

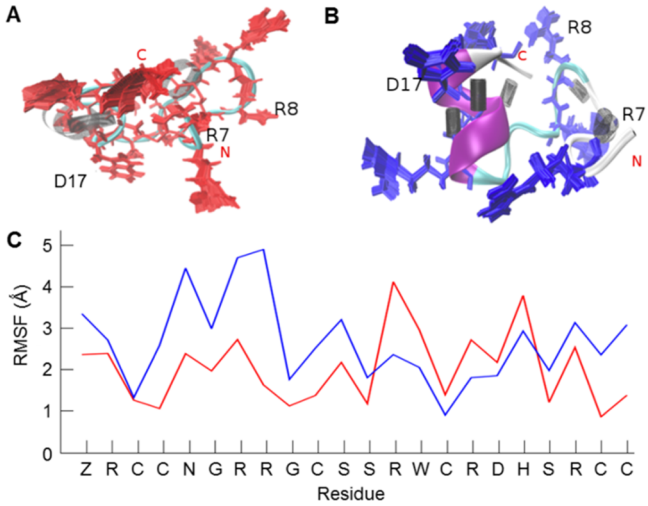


Figure 4. Comparison of μ-SmIIIA native fold and the structure with one disulfide bond opened. (A) Structure of the native peptide (completely oxidized, three disulfide bonds) with 100 conformations of its basic residues (red) superimposed with three residues marked as important for binding activity. The distribution of hydrogen bonds shows a sparse black area which indicates that the region surrounding it is relatively flexible. (B) Structure of the peptide containing one opened disulfide bond (C2−C5) showing (blue) its well-formed α-helix and the dense well-ordered hydrogen bonding illustrated as black cylinders. Higher rigidity inducing an improved structural stability of the peptide in (B) in comparison to the fully oxidized peptide in (A) is apparent from the reference RMSF plot (C). Despite the rigidity of the backbone, the orientations of the basic residues differ from the native structure.

the observed RMSD progression (Figure 3). In the C2−C5 disulfide removed version, between the two, μ-SmIIIA had a marginally more stable backbone than μ-SIIIA. However, with the C3−C6 disulfide bond also removed, both peptides had equally unstable backbones that did not show signs of refolding. Herewith, it is demonstrated how the individual disulfide bridges and the residues that occur within the loops affect the conformational stability on the basis of observed fluctuations in the backbone and the side chain residues.

A closer means of observing the folding behavior of the peptides in this study is to track the movement of the opened cysteine residues in the simulation. By tracking the distances spanned by the Sγ atoms of the cysteines, the tendency of the peptide to fold back to its original conformation or to explore completely new conformations can be identified. This behavior reflects on the underlying folding model that the peptide prefers to adopt. On the basis of this idea, we were able to find a clear correlation between the observations from the synthesis (Figure 2) and simulation (Figure 6). The peptides that preferred forming distinct main products (μ-GIIIA and μ-KIIIA) in the synthesis (Figure 2) also exhibited their preference to fold back to their original conformations as seen from the C2−C5 Sγ distance profiles in Figure 6. This is an indicator to the preference of the BPTI-like folding pathway. On the other hand, μ-PIIIA and μ-SmIIIA that formed a mixture of products in the synthesis (Figure 2) preferred moving away from their original conformations and exploring new conformations as indicated by the C2−C5 Sγ distance profiles in Figure 6. This reflects on the Hirudin-like folding pathway preferred by these peptides. The behavior of μ-SIIIA could be placed between these two extreme cases. The

DOI: 10.1021/acsomega.8b01465
ACS Omega 2018, 3, 12330−12340

videos from the MD simulations where the distances between the opened CYS residues were tracked are provided for the cases representing the opposite extremes, namely, μ-GIIIA and μ-PIIIA are provided as a part of the Supporting Information.

CONCLUSIONS

The folding of smaller disulfide-rich peptides and oligopeptides is a less well-understood folding event because of a much higher degree of flexibility and often a lower extent of structure-forming elements. MD simulations are increasingly used to assist the experimental work for understanding and predicting the folding process. Additionally, they have been routinely used in structure−activity relationship studies, drug discovery, and design pipelines.[39−43] Previous studies using the five μ-conotoxins investigated herein gave insights into the folding and binding modes adopted by these peptides.[10,26,29,33,37,44] Simulating the complete oxidative folding pathway following the formation of non-native disulfide intermediates until the native disulfide bonds are formed as reported by previous studies[45,46] using coarse-grained models is not within the scope of this study. Our work, however, aimed at determining how a particular disulfide bond contributes to the stability of the peptide. Consequently, this approach reviews the validity of the logic that the removal of a disulfide bridge, that is, herein C2−C5, represents a reduction of the backbone stability when considering RMSD and RMSF values.

With the C2−C5 disulfide bond removed, only μ-SmIIIA revealed a noticeable increase in the average fluctuations of its functionally significant residues. In addition, the fact that secondary structural elements such as α-helices were formed in some peptides containing only two disulfide bonds suggests that in distinct cases (e.g., μ-GIIIA), a greater structural rigidity of the backbone may be achieved if one disulfide bridge is removed. This helix-induced stability while strengthening the backbone might reduce the extent of overall fluctuations of the basic residues responsible for binding activity. The biological activity and selectivity of disulfide-deficient mutants might differ from the native conformation as shown for μ-GIIIA recently. Fifteen different disulfide isomers are possible for a peptide containing six cysteines, and still three different isomers (ribbon, bead, and globular) might occur in case of four cysteines. It has, however, not been mentioned in the report by Han et al.[47] which isomer of the μ-GIIIA analogs has been tested because the structural characterization of the respective products was not performed. Apart from the reports on μ-GIIIA regarding the disulfide-deficient variants, another study by Khoo et al.[48] provides an insight into the removal of disulfide bridge C1−C9 in μ-KIIIA, resulting in only a minimal change in the biological activity against $Na_V1.2$ and $Na_V1.4$. In contrast, there are no experimental data for the disulfide-deficient species of μ-conotoxins PIIIA, SIIIA, and SmIIIA available so far.

With respect to drug design and synthesis, the simplification to two disulfide bonds would be a clear benefit for disulfide-rich peptides and proteins. A similar study by Yu et al. on α-conotoxin cVc1.1 complements our idea of the reduction of the number of disulfide bonds.[49] In this respect, we can conclude from our MD simulations that two disulfide bridges could be sufficient to maintain a stable backbone for the majority of the μ-conotoxins studied. However, it is important that the deficient structure is sufficiently supported by at least one pair of cross-linked disulfide bridges that span to almost either ends of the sequence. From the results obtained, a rank order of the five peptides can be provided: μ-GIIIA and μ-KIIIA fall in the highly favorable category, μ-SIIIA falls in the moderately favorable group, and μ-PIIIA and μ-SmIIIA fall into the least favorable group (Figure 5). We also conclude that the C3−C6 disulfide bridge plays the greatest role in retaining the backbone stability for the current strategy of disulfide bond removal employed.

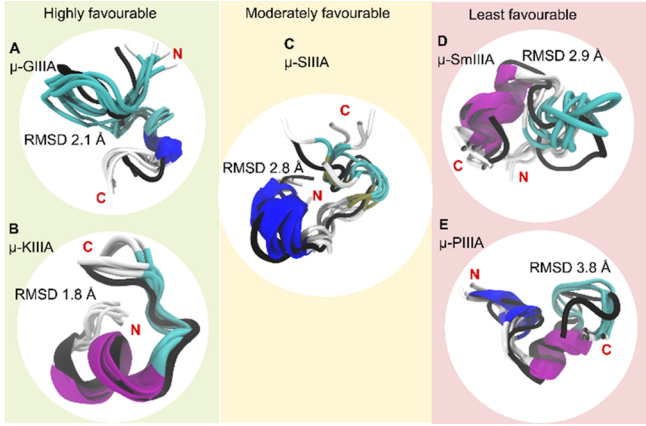


Figure 5. Grouping of the peptides based on the favorability of two disulfide bond stability. The C2−C5 disulfide-deficient conformations (cartoon representations colored to distinguish secondary structural elements) of the studied μ-conotoxins (A) GIIIA, (B) KIIIA, (C) SIIIA, (D) SmIIIA, and (E) PIIIA superimposed on their energy-minimized native structure (black cartoon representation). From the 300 ns trajectory, five conformations (one every 60 ns) have been used. The average RMSD of these conformations in comparison to the reference native structure is shown in Å. Besides displaying the regions of similarity and dissimilarity between the native and the C2−C5-deficient versions, the figure also provides a grouping for the five peptides in terms of favorability of the disulfide-deficient version retaining structural characteristics of the native peptide (based on RMSD, RMSF, and Rg).

As expected, the removal of two disulfide bridges led to an increased backbone flexibility, the formation of a series of intermediate conformations, and a less stable peptide. The increased fluctuations of basic side chain residues responsible for the interactions with the sodium ion channels are in this case unlikely to stay in favorable orientations for binding. Furthermore, the formation of different disulfide isomers for μ-KIIIA, μ-PIIIA, and μ-SmIIIA in the experimental self-folding approach indicates a difference in their folding behavior,[10,18,26,30] which cannot be explained unambiguously by simulations and disulfide removal. Finally, the observations of the structural stability of the backbone observed and the extent to which the peptide tries to fold back gives us clues on which disulfide folding model a particular peptide tends to likely prefer. It must be said that this distinction is still not completely in black and white but can be ranked or ordered relative to each other between the extreme models (BPTI and hirudin). Our attempt to classify the peptides between the BPTI and hirudin folding models is illustrated in Figure 6. The correlation of observations from the experiment and simulation as discussed earlier in the results adds further validity to this classification proposed in Figures 5 and 6. Thus, we have been able to demonstrate the usefulness of molecular simulations in applications beyond the observation of the structural behavior of a peptide in solution to being used as a tool for the

DOI: 10.1021/acsomega.8b01465
ACS Omega 2018, 3, 12330−12340

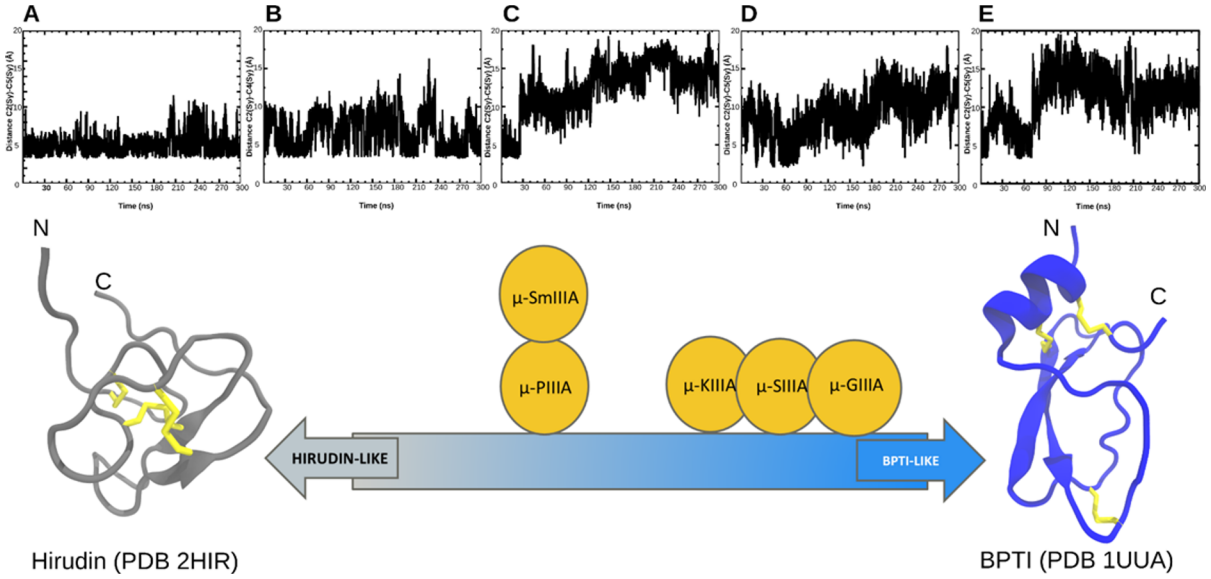


Figure 6. Relation between the unbound cysteine distances and the underlying folding pathway. Top: the distances between the Sγ atoms of the cysteine residues from the peptides with a single disulfide opened plotted over the 300 ns simulation for (A) μ-GIIIA, (B) μ-KIIIA, (C) μ-PIIIA, (D) μ-SIIIA, and (E) μ-SmIIIA. Bottom: a schematic representation of the placement of the five μ-conotoxins within the established BPTI-like and hirudin-like models.[19] The NMR structures of hirudin (the gray cartoon) and BPTI (the blue cartoon) with their three disulfide bridges shown as yellow sticks. This classification is based on the observations from this study.

generalized assignment of peptides to established folding models.

MATERIALS AND METHODS

MD Simulations. Disulfide bonds were systematically removed to yield a partially folded conformation as the starting structure for the simulation. From an atomic perspective, this translates to changing the bonded cysteines to nonbonded cysteines by protonating the sulfur atoms. This step was done by pdb2gmx program within the GROMACS 5.1.4 package.[50,51] This approach induces the least changes in the coordinate file to create a disulfide-deficient species as opposed to the usually performed mutation studies where the Cys residues are replaced by the Ala or Ser residues. It further eliminates the errors arising from the manual manipulation of the coordinate file. Four out of five μ-conotoxins in this study possess the native disulfide connectivity (C1−C4/C2−C5/C3−C6), whereas μ-KIIIA adopts a C1−C5/C2−C4/C3−C6 connection as the stable conformation.[26] The following structures were used herein: PDB ID 1TCG (μ-GIIIA),[23] PDB ID 2LXG (μ-KIIIA),[26] S00159 (μ-PIIIA),[10] BMRB 20025 (μ-SIIIA),[29] and PDB ID 1Q2J (μ-SmIIIA).[30] The NMR ensembles of μ-GIIIA, μ-KIIIA, μ-SIIIA, and μ-SmIIIA had 20 structures and μ-PIIIA had 15 structures in the respective PDB file. The first model was selected as the best representative structure except for SmIIIA, whereas model 13 was chosen as mentioned in the PDB file as the best model for this peptide.

GROMACS 5.1.4[50,51] was used for all the MD simulations in this study. An individual peptide was placed in the center of a cubic box that evolved to a final volume of 2.5 × 2.5 × 2.5 nm^3. TIP3P[52] water model was used as the solvent to fill the box. Appropriate amounts of Cl^- ions were added to balance the positive charge of the μ-conotoxins. Simulations were run using the AMBER99SB-ILDN[53] force field, which was chosen based on its better agreement with the NMR data and an accurate modeling of helical proteins in comparative studies.[54−56] In the process of preparing the peptide for the production MD simulation, energy minimization simulations were with 10 000 steps of the steepest descents minimization protocol and convergence reached when the maximum force on any atom is no greater than 100 kJ/mol/nm. A thermal equilibration in the NVT ensemble at 300 K using the velocity-rescaling modified Berendsen thermostat[57] and a constant pressure equilibration in the NPT ensemble using the Parrinello−Rahman barostat[58,59] at 1 atm were carried out for 20 ns each, prior to production MD. During both the temperature and the pressure ensemble simulations, positional restraints on the peptides were applied using the LINCS[60] algorithm. Each peptide was subjected to three production runs, and on each run, preprocessing and equilibration were performed independently. First, the conformation with all the three disulfide bonds was considered as the control simulation. In the second simulation, the link between C2−C5 (C2−C4 for KIIIA) was removed leaving the two other disulfide bonds intact. In the third simulation, both the C2−C5 (C2−C4 for KIIIA) and the C3−C6 disulfide bonds were removed, leaving the peptide constrained originally by three disulfide bonds now supported by a single disulfide bonds. The production MD was done for 100 ns for all the peptides, and for instances with the opening of a single or two disulfide bonds, a total of 300 ns of simulations were carried out by running an extended 200 ns simulation from the final checkpoint of the 100 ns trajectory. The extended 200 ns simulations were conducted to allow sufficient sampling for the observation of events from the refolding process such as the possibility of the unbound cysteines coming close to each other. All simulations were conducted with a 2 fs time step and data written to the logs and trajectory at every 5 ps. Periodic boundary conditions were applied to the system. Long-range electrostatics were accounted by the particle mesh Ewald method.[61,62] For every 100 ns of simulation, 20 000 frames were written to the trajectory. The effect of periodic boundary conditions was adjusted by the suppression of center of mass movement from

DOI: 10.1021/acsomega.8b01465
ACS Omega 2018, 3, 12330−12340

the trajectory prior to analysis. Visualizations of conformations for the analysis and creation of images were performed using VMD.[63] The RMSD, RMSF, and Rg plots were created using the program Grace (version 5.1.25), whereas the distances between the unbound cysteines were plotted using the tools within VMD. The RMSD curves were plotted for every 10 ps (10 000 frames), whereas the distance curves were plotted for every 5 ps (20 000 frames).

Chemical Synthesis and Purification of μ-Conotoxins. Peptides were produced by an automated solid-phase peptide synthesis using a standard Fmoc (*N*-(9-fluorenyl) methoxycarbonyl)-protocol and an EPS 221 peptide synthesizer (Intavis Bioanalytical Instruments AG, Cologne, Germany) as described earlier and purified by preparative reversed-phase (RP) HPLC (Shimadzu LC-8A system, Duisburg, Germany). The gradient used was 0−50% eluent B in 120 min with 0.1% trifluoroacetic acid (TFA) in water (eluent A) and 0.1% TFA in acetonitrile/water (9:1) (eluent B) on a C18 column (Knauer Eurosphere 100, Berlin, Germany) with the dimensions 50 mm × 300 mm (5 mm particle size, 100 Å pore size). Reduced and oxidized peptides were analyzed on a LC−ESI micrOTOF-Q III mass spectrometer (Bruker Daltonics GmbH, Bremen, Germany) coupled with Dionex Ultimate 3000 (Thermo Scientific, Dreieich, Germany) equipped with a EC100/2 Nucleoshell RP18 Gravity 2.7 μm column (Macherey-Nagel, Düren, Germany). Analysis of the MS data was performed using Bruker Compass Data Analysis 4.1. The LC conditions used were as follows: eluent A was water with 0.1% acetic acid, whereas eluent B was acetonitrile containing 0.1% acetic acid. A gradient of 0−60% of eluent B in 12 min was used, and detection was at 220 nm.

Oxidation of Reduced μ-Conotoxin Precursors. Oxidative folding of the linear μ-conotoxins GIIIA, KIIIA, PIIIA, SIIIA, and SmIIIA in a buffer system containing redox agents was performed as described earlier. Each μ-conotoxin (1 mg) was subjected to oxidation, and fractions of the reaction mixture were monitored over time by RP HPLC using a Shimadzu LC-10AT system (Duisburg, Germany) equipped with a C18 column (Vydac 218TP54, Worms, Germany, 4.6 mm × 25 mm, 5 mm particle size, 300 Å pore size) and the gradient 0−60% eluent B in 60 min with 0.1% TFA in water (eluent A) and 0.1% TFA in acetonitrile (eluent B). Reaction control was performed over a time period of 24 h, and oxidation was stopped by adding 1% TFA in water. Monitoring revealed that the oxidation reactions were completed within the first 60 min of the reaction time. The fractions were collected for each peptide and subjected to LC−ESI mass spectrometry for the confirmation of the molar mass corresponding to the oxidized products.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.8b01465.

RMSD, RMSF, and Rg of the five μ-conotoxins; backbone stability of the five μ-conotoxins with three, two, and one disulfide bridge; backbone conformation of μ-GIIIA and μ-SmIIIA and helix formation; backbone conformation of μ-PIIIA and loss of helix by disulfide bond removal; comparison of disulfide loop length of the five μ-conotoxins; assessment of the stability using RMSD and Rg for 100 ns MD simulation ; and impact of the C2−C5 and C3−C6 disulfide bridge on the backbone RMSF and the functionally stable residues (PDF)

Dynamics of the peptide μ-GIIIA with the C2−C5 disulfide bond removed (AVI)

Dynamics of the peptide μ-PIIIA with the C2−C5 disulfide bond removed (AVI)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: dimhof@uni-bonn.de (D.I.).

ORCID

Diana Imhof: 0000-0003-4163-7334

Present Address

#Bachem AG, Hauptstraße 144, 4416 Bubendorf (Switzerland).

Author Contributions

A.A.P.G., P.H., and D.I. conceived the presented idea and designed and planned the experimental studies. P.H. performed the synthesis, purification, and analytical characterization of the peptides. A.A.P.G., A.M., J.H., M.H.-A., and A.B. developed the theory. A.A.P.G. performed the computations. All authors discussed the results and contributed to the final manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank Veda Thota for technical assistance. Financial support by the University of Bonn (to D.I.) is gratefully acknowledged.

■ REFERENCES

(1) Hille, B. The Superfamily of Voltage-gated Channels. *Ion Channel Excitable Membranes*, 3rd ed.; Sinauer Associates, Inc.: Sunderland, Massachusetts, 2001; pp 1−92.

(2) Cruz, L. J.; Gray, W. R.; Yoshikami, D.; Olivera, B. M. Conus Venoms: A Rich Source of Neuroactive Peptides. *J. Toxicol., Toxin Rev.* **1985**, *4*, 107−132.

(3) Akondi, K. B.; Muttenthaler, M.; Dutertre, S.; Kaas, Q.; Craik, D. J.; Lewis, R. J.; Alewood, P. F. Discovery , Synthesis, and Structure − Activity Relationships of Conotoxins. *Chem. Rev.* **2014**, *114*, 5815−5847.

(4) Kaas, Q.; Yu, R.; Jin, A.-H.; Dutertre, S.; Craik, D. J. ConoServer: Updated Content, Knowledge, and Discovery Tools in the Conopeptide Database. *Nucleic Acids Res.* **2012**, *40*, D325−D330.

(5) Leipold, E.; DeBie, H.; Zorn, S.; Adolfo, B.; Olivera, B. M.; Terlau, H.; Heinemann, S. H. μO-Conotoxins Inhibit NaVChannels by Interfering with their Voltage Sensors in Domain-2. *Channels* **2007**, *1*, 253−262.

(6) Jacob, R. B.; McDougal, O. M. The M-Superfamily of Conotoxins: A Review. *Cell. Mol. Life Sci.* **2010**, *67*, 17−27.

(7) Kaas, Q.; Westermann, J.-C.; Halai, R.; Wang, C. K. L.; Craik, D. J. ConoServer, a Database for Conopeptide Sequences and Structures. *Bioinformatics* **2008**, *24*, 445−446.

(8) Heimer, P.; Tietze, A. A.; Böhm, M.; Giernoth, R.; Kuchenbuch, A.; Stark, A.; Leipold, E.; Heinemann, S. H.; Kandt, C.; Imhof, D. Application of Room Temperature Aprotic and Protic Ionic Liquids for Oxidative Folding of Cysteine-Rich Peptides. *ChemBioChem* **2014**, *15* (18), 2754−2765.

(9) Tietze, D.; Leipold, E.; Heimer, P.; Böhm, M.; Winschel, W.; Imhof, D.; Heinemann, S. H.; Tietze, A. A. Molecular interaction of δ-

DOI: 10.1021/acsomega.8b01465
ACS Omega 2018, 3, 12330−12340

conopeptide EVIA with voltage-gated Na+ channels. *Biochim. Biophys. Acta, Gen. Subj.* **2016**, *1860*, 2053−2063.
(10) Tietze, A. A.; Tietze, D.; Ohlenschläger, O.; Leipold, E.; Ullrich, F.; Kühl, T.; Mischo, A.; Buntkowsky, G.; Görlach, M.; Heinemann, S. H.; et al. Structurally Diverse μ-Conotoxin PIIIA Isomers Block Sodium Channel NaV1.4. *Angew. Chem., Int. Ed.* **2012**, *51*, 4058−4061.
(11) Winter, J.; Wenghoefer, M. Human Defensins: Potential Tools for Clinical Applications. *Polymers* **2012**, *4*, 691−709.
(12) Szyk, A.; Wu, Z.; Tucker, K.; Yang, D.; Lu, W.; Lubkowski, J. Crystal structures of human α-defensins HNP4, HD5, and HD6. *Protein Sci.* **2006**, *15*, 2749−2760.
(13) Lyons, M. S.; Bell, B.; Stainier, D.; Peters, K. G. Isolation of the zebrafish homologues for the tie1 and tie2 endothelium-specific receptor tyrosine kinases. *Dev. Dyn.* **1998**, *212*, 133−140.
(14) Ranasinghe, S.; McManus, D. P. Structure and Function of Invertebrate Kunitz Serine Protease Inhibitors. *Dev. Comp. Immunol.* **2013**, *39*, 219−227.
(15) Barnham, K. J.; Torres, A. M.; Norton, R. S.; Alewood, D.; Alewood, P. F.; Domagala, T.; Nice, E. C. Role of the 6-20 Disulfide Bridge in the Structure and Activity of Epidermal Growth Factor. *Protein Sci.* **1998**, *7*, 1738−1749.
(16) Christinger, H. W.; Fuh, G.; de Vos, A. M.; Wiesmann, C. The Crystal Structure of Placental Growth Factor in Complex with Domain 2 of Vascular Endothelial Growth Factor Receptor-1. *J. Biol. Chem.* **2004**, *279*, 10382−10388.
(17) Góngora-Benítez, M.; Tulla-Puche, J.; Albericio, F. Multifaceted Roles of Disulfide Bonds. Peptides as Therapeutics. *Chem. Rev.* **2014**, *114*, 901−926.
(18) Fuller, E.; Green, B. R.; Catlin, P.; Buczek, O.; Nielsen, J. S.; Olivera, B. M.; Bulaj, G. Oxidative Folding of Conotoxins Sharing an Identical Disulfide Bridging Framework. *FEBS J.* **2005**, *272*, 1727−1738.
(19) Chang, J.-Y. Diverse Pathways of Oxidative Folding of Disulfide Proteins: Underlying Causes and Folding Models. *Biochemistry* **2011**, *50*, 3414−3431.
(20) Frishman, D.; Argos, P. Knowledge-Based Protein Secondary Structure Assignment. *Proteins* **1995**, *23*, 566−579.
(21) Varshney, A.; Brooks, F. P.; Wright, W. V. Computing Smooth Molecular Surfaces. *IEEE Comput. Graph. Appl. Mag.* **1994**, *14*, 19−25.
(22) Kaas, Q.; Yu, R.; Jin, A.-H.; Dutertre, S.; Craik, D. J. ConoServer: Updated Content, Knowledge, and Discovery Tools in the Conopeptide Database. *Nucleic Acids Res.* **2012**, *40*, D325−D330.
(23) Wakamatsu, K.; Kohda, D.; Hatanaka, H.; Lancelin, J. M.; Ishida, Y.; Oya, M.; Nakamura, H.; Inagaki, F.; Sato, K. Structure-activity relationships of .mu.-conotoxin GIIIA: structure determination of active and inactive sodium channel blocker peptides by NMR and simulated annealing calculations. *Biochemistry* **1992**, *31*, 12577−12584.
(24) Patel, D.; Mahdavi, S.; Kuyucak, S. Computational Study of Binding of μ-Conotoxin GIIIA to Bacterial Sodium Channels NaVAb and NaVRh. *Biochemistry* **2016**, *55*, 1929−1938.
(25) Wilson, M. J.; Yoshikami, D.; Azam, L.; Gajewiak, J.; Olivera, B. M.; Bulaj, G.; Zhang, M.-M. -Conotoxins that differentially block sodium channels NaV1.1 through 1.8 identify those responsible for action potentials in sciatic nerve. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10302−10307.
(26) Khoo, K. K.; Gupta, K.; Green, B. R.; Zhang, M.-M.; Watkins, M.; Olivera, B. M.; Balaram, P.; Yoshikami, D.; Bulaj, G.; Norton, R. S. Distinct Disulfide Isomers of μ-Conotoxins KIIIA and KIIIB Block Voltage-Gated Sodium Channels. *Biochemistry* **2012**, *51*, 9826−9835.
(27) Zhang, M.-M.; Green, B. R.; Catlin, P.; Fiedler, B.; Azam, L.; Chadwick, A.; Terlau, H.; McArthur, J. R.; French, R. J.; Gulyas, J.; et al. Structure/Function Characterization of μ-Conotoxin KIIIA, an Analgesic, Nearly Irreversible Blocker of Mammalian Neuronal Sodium Channels. *J. Biol. Chem.* **2007**, *282*, 30699−30706.
(28) Tietze, A. A.; Tietze, D.; Ohlenschläger, O.; Leipold, E.; Ullrich, F.; Kühl, T.; Mischo, A.; Buntkowsky, G.; Görlach, M.; Heinemann, S. H.; et al. Structurally Diverse μ-Conotoxin PIIIA Isomers Block Sodium Channel NaV1.4. *Angew. Chem., Int. Ed.* **2012**, *51*, 4058−4061.
(29) Yao, S.; Zhang, M.-M.; Yoshikami, D.; Azam, L.; Olivera, B. M.; Bulaj, G.; Norton, R. S. Structure, Dynamics, and Selectivity of the Sodium Channel Blocker μ-Conotoxin SIIIA. *Biochemistry* **2008**, *47*, 10940−10949.
(30) Keizer, D. W.; West, P. J.; Lee, E. F.; Yoshikami, D.; Olivera, B. M.; Bulaj, G.; Norton, R. S. Structural Basis for Tetrodotoxin-resistant Sodium Channel Binding by μ-Conotoxin SmIIIA. *J. Biol. Chem.* **2003**, *278*, 46805−46813.
(31) Fuller, E.; Green, B. R.; Catlin, P.; Buczek, O.; Nielsen, J. S.; Olivera, B. M.; Bulaj, G. Oxidative Folding of Conotoxins Sharing an Identical Disulfide Bridging Framework. *FEBS J.* **2005**, *272*, 1727−1738.
(32) Walewska, A.; Skalicky, J. J.; Davis, D. R.; Zhang, M.-M.; Lopez-Vera, E.; Watkins, M.; Han, T. S.; Yoshikami, D.; Olivera, B. M.; Bulaj, G. NMR-Based Mapping of Disulfide Bridges in Cysteine-Rich Peptides: Application to the μ-Conotoxin SxIIIA. *J. Am. Chem. Soc.* **2008**, *130*, 14280−14286.
(33) Lopez-Vera, E.; Walewska, A.; Skalicky, J. J.; Olivera, B. M.; Bulaj, G. Role of Hydroxyprolines in the in Vitro Oxidative Folding and Biological Activity of Conotoxins. *Biochemistry* **2008**, *47*, 1741−1751.
(34) Bulaj, G.; Olivera, B. M. Folding of Conotoxins: Formation of the Native Disulfide Bridges during Chemical Synthesis and Biosynthesis of Conus Peptides. *Antioxid. Redox Signaling* **2008**, *10*, 141−156.
(35) Kang, T. S.; Kini, R. M. Structural Determinants of Protein Folding. *Cell. Mol. Life Sci.* **2009**, *66*, 2341−2361.
(36) Pace, C. N.; Shirley, B. A.; McNutt, M.; Gajiwala, K. Forces Contributing to the Conformational Stability of Proteins. *FASEB J.* **1996**, *10*, 75−83.
(37) Choudhary, G.; Aliste, M. P.; Tieleman, D. P.; French, R. J.; Dudley, S. C., Jr. Docking of μ-Conotoxin GIIIA in the Sodium Channel Outer Vestibule. *Channels* **2007**, *1*, 344−352.
(38) Schroeder, C. I.; Ekberg, J.; Nielsen, K. J.; Adams, D.; Loughnan, M. L.; Thomas, L.; Adams, D. J.; Alewood, P. F.; Lewis, R. J. Neuronally Selective μ-Conotoxins from Conus striatus Utilize an α-Helical Motif to Target Mammalian Sodium Channels. *J. Biol. Chem.* **2008**, *283*, 21621−21628.
(39) Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biol.* **2011**, *9*, 71.
(40) Zhao, H.; Caflisch, A. Molecular Dynamics in Drug Design. *Eur. J. Med. Chem.* **2015**, *91*, 4−14.
(41) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59*, 4035−4061.
(42) Nair, P. C.; Miners, J. O. Molecular Dynamics Simulations: From Structure Function Relationships to Drug Discovery. *In Silico Pharmacol.* **2014**, *2*, 1−4.
(43) Salsbury, F. R. Molecular Dynamics Simulations of Protein Dynamics and Their Relevance to Drug Discovery. *Curr. Opin. Pharmacol.* **2010**, *10*, 738−744.
(44) Leipold, E.; Markgraf, R.; Miloslavina, A.; Kijas, M.; Schirmeyer, J.; Imhof, D.; Heinemann, S. H. Molecular determinants for the subtype specificity of μ-conotoxin SIIIA targeting neuronal voltage-gated sodium channels. *Neuropharmacology* **2011**, *61*, 105−111.
(45) Qin, M.; Wang, W.; Thirumalai, D. Protein Folding Guides Disulfide Bond Formation. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 11241−11246.
(46) Chinchio, M.; Czaplewski, C.; Liwo, A.; Ołdziej, S.; Scheraga, H. A. Dynamic Formation and Breaking of Disulfide Bonds in Molecular Dynamics Simulations with the UNRES Force Field. *J. Chem. Theory Comput.* **2007**, *3*, 1236−1248.
(47) Han, P.; Wang, K.; Dai, X.; Cao, Y.; Liu, S.; Jiang, H.; Fan, C.; Wu, W.; Chen, J. The Role of Individual Disulfide Bonds of μ-

DOI: 10.1021/acsomega.8b01465
ACS Omega 2018, 3, 12330−12340

Conotoxin GIIIA in the Inhibition of Nav1.4. *Mar. Drugs* **2016**, *14*, 213.
(48) Khoo, K. K.; Feng, Z.-P.; Smith, B. J.; Zhang, M.-M.; Yoshikami, D.; Olivera, B. M.; Bulaj, G.; Norton, R. S. Structure of the Analgesic μ-Conotoxin KIIIA and Effects on the Structure and Function of Disulfide Deletion. *Biochemistry* **2009**, *48*, 1210−1219.
(49) Yu, R.; Seymour, V. A. L.; Berecki, G.; Jia, X.; Akcan, M.; Adams, D. J.; Kaas, Q.; Craik, D. J. Less Is More: Design of a Highly Stable Disulfide-Deleted Mutant of Analgesic Cyclic α-Conotoxin Vc1.1. *Sci. Rep.* **2015**, *5*, 13264.
(50) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1*−*2*, 19−25.
(51) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(52) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(53) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950−1958.
(54) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic Validation of Protein Force Fields against Experimental Data. *PLoS One* **2012**, *7*, No. e32131.
(55) Serafeim, A.-P.; Salamanos, G.; Patapati, K. K.; Glykos, N. M. Sensitivity of Folding Molecular Dynamics Simulations to Even Minor Force Field Changes. *J. Chem. Inf. Model.* **2016**, *56*, 2035−2041.
(56) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100*, L47−L49.
(57) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(58) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(59) Nosé, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **1983**, *50*, 1055−1076.
(60) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(61) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(62) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(63) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.

■ NOTE ADDED AFTER ASAP PUBLICATION

This paper was published ASAP on October 1, 2018, with errors in the references. The corrected paper was reposted on October 4, 2018.

DOI: 10.1021/acsomega.8b01465
ACS Omega 2018, 3, 12330−12340

4.1.2 Summary

MD simulation based conformational analysis was applied to the available NMR structural ensemble of five μ-conotoxins namely μ-GIIIA, μ-KIIIA, μ-PIIIA, μ-SIIIA and μ-SmIIIA sharing identical disulfide bonding patterns (C1-C4/C2-C5/C3-C6). Their folding behavior was investigated by creating partially folded forms of these peptides via a systematic *in silico* removal of disulfide bonds. First, the disulfide bond which covered the largest disulfide loop length in these peptides was removed. In 4 out of five cases this happened to be the C2-C5 disulfide bond (except for μ-KIIIA). Following this, the C3-C6 disulfide bond was removed in an independent round of simulations, leaving the peptide's structural integrity to be left to a single disulfide linkage. The observations from the structural evolution obtained by changes in the backbone root mean square deviations (RMSD), per residue root mean square fluctuation (RMSF) and radius of gyration (Rg) along with structural characterization based on the observed secondary structural elements sampled by the peptides in simulations formed the basis of the findings of this study. The first key finding of this study was that all of the five conotoxins needed at least two of their three native disulfide bonds to retain structural and thermodynamic stability. It was also observed that the conotoxins μ-GIIIA and μ-KIIIA were able to maintain native-like structures despite the disruption of their C2-C5 disulfide bond due to new stabilizing secondary structural formations upon disulfide removal. While μ-SIIIA displayed moderate stability upon disulfide removal, the peptides μ-PIIIA and μ-SmIIIA displayed poor conformational stabilities in their disulfide-deficient forms. This lead to the assignment of the five peptides between the opposite and extreme oxidative folding models represented by the proteins BPTI and hirudin. μ-GIIIA displayed the most BPTI-like behavior while μ-PIIIA and μ-SmIIIA exhibited folding behaviors resembling that of hirudin. μ-SIIIA's observed folding behavior indicated that it could follow either BPTI-like or hirudin-like folding pathways.

4.2 Chapter II

Effect of Conformational Diversity on the Bioactivity of μ-Conotoxin PIIIA Disulfide Isomers

Original research article

Authors*

Ajay Abisheck Paul George, Pascal Heimer, Enrico Leipold, Thomas Schmitz , Desirée Kaufmann, Daniel Tietze, Stefan H. Heinemann and Diana Imhof

This peer-reviewed research article was published in *Marine drugs.*

Citation

Mar. Drugs 2019, 17, 390

DOI: 10.3390/md17070390

4.2.1 Introduction

In the previous chapter five μ-conotoxins sharing identical disulfide connectivity but different primary sequences were investigated. In the current study the opposite is pursued, i.e., the fifteen theoretically possible disulfide isomers of μ-conotoxin PIIIA are subjected to a combination of electrophysiological and computational studies. These isomers synthesized from the study in an earlier study[152], were tested for their blocking activity at the VGSC Nav1.4 using electrophysiological experiments. Additionally three 2-disulfide-bonded variants of native μ-PIIIA were synthesized by replacing the cysteines that formed the native disulfide bonds by serines at each instance. Furthermore twelve NMR structures and three computationally derived structures of the disulfide isomers made available from the previous study[152] were subjected to rigorous MD based studies. To replicate the experimental setup computer models of disulfide-deficient mutants of native μ-PIIIA were created and simulated as well. Three isomers selected (μ-PIIIA 2, μ-PIIIA 7 and μ-PIIIA 15) based on the observed blocking activity at Nav1.4 were docked on the cryo-electron microscopic structure of the VGSC Nav1.4 and the best docked complexes were subject to further MD simulation based refinement. The outcome of this study was a clear understanding of how different disulfide bonding patterns on the same primary sequence influence the structure, dynamics and ultimately the function of the peptide μ-PIIIA.

***Own contribution**

The computational study including MD simulations of peptide in solution, MD simulations of the docked complexes and analysis of the results was performed by myself. Desirée Kaufmann and Daniel Tietze performed the docking simulations. The design and the content of the manuscript has been compiled by Diana Imhof and me. The layout and preparation of the figures has been planned and carried out by me in consultation and agreement with Diana Imhof, Enrico Leipold and Stefan H. Heinemann. All authors contributed to the writing of the manuscript.

marine drugs

MDPI

Article

Effect of Conformational Diversity on the Bioactivity of μ-Conotoxin PIIIA Disulfide Isomers

Ajay Abisheck Paul George [1,†], Pascal Heimer [1,†], Enrico Leipold [2,†], Thomas Schmitz [1], Desiree Kaufmann [3], Daniel Tietze [3], Stefan H. Heinemann [4] and Diana Imhof [1,*]

1 Pharmaceutical Biochemistry and Bioanalytics, Pharmaceutical Institute, University of Bonn, An der Immenburg 4, D-53121 Bonn, Germany
2 Department of Anesthesiology and Intensive Care, University of Lübeck, Ratzeburger Allee 160, D-23562 Lübeck, Germany
3 Eduard Zintl Institute of Inorganic and Physical Chemistry, Darmstadt University of Technology, Alarich-Weiss-Str. 4, D-64287 Darmstadt, Germany
4 Center for Molecular Biomedicine, Department of Biophysics, Friedrich Schiller University Jena and Jena University Hospital, Hans-Knöll-Str. 2, D-07745 Jena, Germany
* Correspondence: dimhof@uni-bonn.de; Tel.: +49-228-73-5254
† These authors contributed equally to this work.

Received: 5 June 2019; Accepted: 25 June 2019; Published: 2 July 2019

Abstract: Cyclic μ-conotoxin PIIIA, a potent blocker of skeletal muscle voltage-gated sodium channel $Na_V1.4$, is a 22mer peptide stabilized by three disulfide bonds. Combining electrophysiological measurements with molecular docking and dynamic simulations based on NMR solution structures, we investigated the 15 possible 3-disulfide-bonded isomers of μ-PIIIA to relate their blocking activity at $Na_V1.4$ to their disulfide connectivity. In addition, three μ-PIIIA mutants derived from the native disulfide isomer, in which one of the disulfide bonds was omitted (C4-16, C5-C21, C11-C22), were generated using a targeted protecting group strategy and tested using the aforementioned methods. The 3-disulfide-bonded isomers had a range of different conformational stabilities, with highly unstructured, flexible conformations with low or no channel-blocking activity, while more constrained molecules preserved 30% to 50% of the native isomer's activity. This emphasizes the importance and direct link between correct fold and function. The elimination of one disulfide bond resulted in a significant loss of blocking activity at $Na_V1.4$, highlighting the importance of the 3-disulfide-bonded architecture for μ-PIIIA. μ-PIIIA bioactivity is governed by a subtle interplay between an optimally folded structure resulting from a specific disulfide connectivity and the electrostatic potential of the conformational ensemble.

Keywords: μ-conotoxin; PIIIA; voltage-gated sodium channel; disulfide connectivity; peptide folding; electrophysiology; molecular docking; molecular dynamics

1. Introduction

Covalent linkage of cysteine residues by disulfide bonds is fundamental for the folding, stability, and function of many peptides and proteins [1–4]. The formation of correct disulfide linkages is particularly important for peptides and proteins with higher-order disulfide networks (>2 disulfide bonds) because incorrectly assembled disulfide bonds may directly interfere with the biological functions of these molecules.

The venom of marine cone snails is a rich source of small- to medium-sized disulfide-stabilized peptides, so-called conotoxins, most of which specifically target ion channels and receptors in the membranes of excitable cells [5]. Conotoxins are recognized for their pharmacological and therapeutic

potential and they are models to study the impact of disulfide bonds on the structural stability of disulfide-rich peptides and proteins [4,6–8]. For example, it was shown that the biological activities and receptor specificities of some μ-, μO-, and α-conotoxins changed significantly if the native disulfide patterns of the peptides were experimentally modified [9–12].

The 3-disulfide-bonded μ-conotoxins specifically antagonize voltage-gated sodium channels (Na_V channels) making them attractive research tools as well as potential pharmacological lead structures [13]. μ-Conotoxins are so-called pore blockers because they bind to the extracellular pore vestibule and, hence, obstruct the permeation of sodium ions (Na^+) through the channels [14]. While some μ-conotoxins, such as μ-GIIIA are very specific for skeletal muscle sodium channels ($Na_V1.4$), μ-PIIIA—originally isolated from *Conus purpurascens*—also affects neuronal Na_V channels [10] and even some voltage-gated K^+ channels [15]. Recently, we synthesized all 15 possible disulfide isomers (denoted in the current study as **1**–**15**) of the conotoxin μ-PIIIA (sequence: ZRLCCGFOKSCRSRQCKOHRCC-NH_2), determined their structure and verified their disulfide bond patterns by applying a combination of MS/MS analysis and 2D NMR spectroscopy [16]. In the present study, these isomers and further analogs (Table 1) were analyzed in electrophysiological experiments for their potency to block the human skeletal muscle voltage-gated sodium channel $Na_V1.4$. Interactions between the isomers and the channel protein were examined via molecular docking simulations. All the 21 PIIIA analogs (Table 1) were subjected to all-atom molecular dynamics (MD) simulations in solution while 3 selected isomer-channel complexes obtained from the docking simulations were subjected to further MD-based refinement for additional details of how the isomer-channel interaction stabilizes over time. The analysis of the 21 peptides (Table 1) revealed that the channel-blocking ability is determined by a complex interplay between conformational stability, orientations of the functionally significant side chains [17], and the electrostatic properties of the molecules. This study provides a conceptual framework to understand how different disulfide connectivities affect the structure of μ-PIIIA and ultimately its interaction with the pore of $Na_V1.4$.

Table 1. Nomenclature, disulfide connectivity, type of structure, and source references for the peptides used in this study.

Isomer Nomenclature	Number of Disulfides	Disulfide Connectivity	Type of Structure	Source Reference(s)
1	3	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	NMR	Heimer et al. [16], Tietze et al. [10].
2	3	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	NMR	Heimer et al. [16], Tietze et al. [10].
3	3	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	NMR	Heimer et al. [16], Tietze et al. [10].
4	3	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	NMR	Heimer et al. [16].
5	3	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	NMR	Heimer et al. [16].
6	3	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	NMR	Heimer et al. [16].
7	3	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	*In silico* model	Heimer et al. [16].
8	3	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	NMR	Heimer et al. [16].
9	3	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	NMR	Heimer et al. [16].
10	3	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	NMR	Heimer et al. [16].

Table 1. *Cont.*

Isomer Nomenclature	Number of Disulfides	Disulfide Connectivity	Type of Structure	Source Reference(s)
11	3	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	NMR	Heimer et al. [16].
12	3	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	*In silico* model	Heimer et al. [16].
13	3	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	*In silico* model	Heimer et al. [16].
14	3	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	NMR	Heimer et al. [16].
15	3	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	NMR	Heimer et al. [16].
16*	2	-C_4-C_5---S_{11}---C_{16}---C_{21}-S_{22}-	*In silico* model	Current study
17*	2	-S_4-C_5---C_{11}---S_{16}---C_{21}-C_{22}-	*In silico* model	Current study
18*	2	-C_4-S_5---C_{11}---C_{16}---S_{21}-C_{22}-	*In silico* model	Current study
Δ(C5-C21)**2**#	2	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	*In silico* model	Current study
Δ(C11-C21)**4**#	2	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	*In silico* model	Current study
Δ(C5-C22)**10**#	2	-C_4-C_5---C_{11}---C_{16}---C_{21}-C_{22}-	*In silico* model	Current study

* Mutated disulfide bonds replaced by serine residues. # Isomers with in silico disulfide bond opening yielding two SH-groups.

2. Results

2.1. Bioactivity of 3- and 2-Disulfide-bonded μ-PIIIA Analogs at $Na_v1.4$

Previously, we synthesized all 15 possible 3-disulfide-bonded isomers of the conopeptide μ-PIIIA (here termed **1**–**15**, Table 1) and determined their 3-dimensional structure [16]. Based on the native disulfide connectivity of isomer **2**, we now produced three additional 2-disulfide-bonded mutants of μ-PIIIA (termed **16**–**18**), each lacking one of the disulfide bonds of **2** (Table 1, Figure S1, Tables S1 and S2). Previous reports have shown that μ-PIIIA has the highest potency to interfere with skeletal muscle channel $Na_V1.4$ although also being active on neuronal Na_V channels, such as $Na_V1.6$ and $Na_V1.8$ [18], and even on select voltage-gated K^+ (K_V) channels [15,19]. We therefore applied all 3-disulfide-bonded as well as the 2-disulfide-bonded isomers to HEK293 cells expressing human $Na_V1.4$ channels and measured the impact of the toxin isomers on depolarization-elicited Na^+ inward currents, expressed as the time constant for current block and the maximal block after long toxin exposure (Figure 1, Figure S2). Figure 1 summarizes the inhibitory activity of the most active isomers **1**, **2**, **7**, **11**, **12**, **14**, and **15**. As demonstrated in Figure 1A, currents mediated by human $Na_V1.4$ were diminished by all isomers, albeit to different degrees. At a concentration of 10 μM, isomer **2**, which has the native disulfide connectivity of μ-PIIIA (C4-C16/C5-C21/C11-C22), was most effective in inhibiting $Na_V1.4$ channels, followed by isomers **15**, **11**, **14**, **1**, **7**, and **12**. Analysis of the concentration dependence revealed that isomer **2** blocked human $Na_V1.4$ with an apparent IC_{50} of 105.3 ± 29.9 nM, which is comparable to the value for the paralog rat $Na_V1.4$ channels (103.2 ± 9.9 nM) [10]. As observed for some μ-conotoxins, even at a saturating concentration of 100 μM a small current component of 7.2 ± 2.2% remained, indicating that channel occupancy by isomer **2** does not eliminate Na^+ conduction (Figure 1A,B). The remaining

isomers (**15**, **11**, **14**, **1**, **7**, **12**) blocked $Na_V1.4$ channels less potently than isomer **2**. In addition, the onset of block as estimated with single-exponential functions was substantially slower than for isomer **2**, thus precluding faithful assessment of channel block at lower concentrations than 10 µM and to determine the associated IC_{50} values (Figure 1C). Assessment of higher concentrations revealed that particularly isomers **15** and **7** are interesting: total current block at 10 and 100 µM was virtually identical (**15**: 48.2 ± 5.9% and 51.1 ± 2.2%, respectively; **7**: 32.3 ± 4.3% and 35.1 ± 4.6%, respectively), while the time constant characterizing the kinetics of onset of block, τ_{block}, roughly scaled linearly with the concentration (**15**: 10 µM: 160.0 ± 34.9 ms; 100 µM: 4.7 ± 1 ms; **7**: 10 µM: 853.0 ± 195.0 ms; 100 µM: 59.0 ± 18.1 ms; Figure 1C). This result indicates saturated association of isomers **15** and **7** and the channel, with imperfect (about 50% and 35%, respectively) occlusion of the Na^+ permeation pathway. In contrast, for isomers **11**, **14**, **1**, and **12** saturation of channel block was not apparent as increasing the concentration of the peptides from 10 µM to 100 µM also increased channel block (Figure 1C). The channel block of isomers with even lower potency (**3**–**6**, **8**, **10**, **13** and **16**–**18**) were analyzed at 10 µM. As shown in Figure S2, sample **13** was most active among this group, as it inhibited $Na_V1.4$ by 24.6 ± 3.0%, followed by samples **4** (21.7 ± 3.5%), **9** (17.7 ± 1.7%), **3** (17.6 ± 3.6%), **17** (16.5 ± 0.5%), and **18** (16.5 ± 0.3%). Isomer samples **6**, **8**, **16**, **5,** and **10** diminished $Na_V1.4$-mediated currents by less than 15% (**6**: 12.6 ± 1.3%, **8**: 11.4 ± 2.2%, **16**: 10.5 ± 1.3%, **5**: 6.7 ± 2.1%, **10**: 1.0 ± 3.0%), and thus were considered inactive under these conditions.

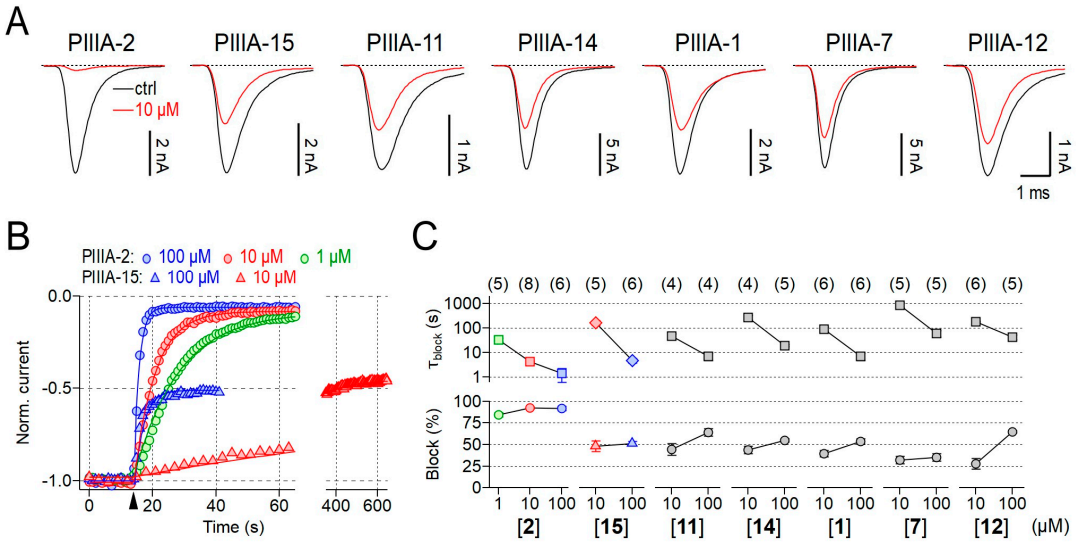


Figure 1. Disulfide isomers of µ-PIIIA differentially inhibit $Na_V1.4$-mediated currents. (**A**) Representative current traces of transiently expressed $Na_V1.4$ channels evoked at a test potential of −20 mV before (black, ctrl) and after (red) application of 10 µM of the indicated µ-PIIIA isomers. (**B**) Normalized peak current amplitudes obtained from repetitively evoked current responses were plotted as a function of time to follow the time course of current block mediated by various concentrations of µ-PIIIA isomers **2** (circles) and **15** (triangles). Continuous lines are single-exponential data fits used to characterize the onset of channel block. The arrowhead marks the start of peptide application. The time axis was split to illustrate that channel block by isomer **15** saturates at about 50% suggesting that isomer **15** seals the channel pore only partially. In contrast, isomer 2-mediated channel inhibition saturated at about 95%. (**C**) Steady-state block estimated from single-exponential fits of the time course of current inhibition (bottom) as well as the associated single-exponential time constant (top), describing the onset of channel block, for the indicated isomers and concentrations. Lines connect data points for clarity. Symbols and color-coding of data obtained with isomers **2** and **15** are as in (**B**). Numbers of individual experiments (n) are provided in parentheses.

2.2. *In Silico Toxin Binding Studies of the 3-Disulfide-Bonded μ-PIIIA Analogs at Nav1.4*

Interactions of the μ-PIIIA isomers with the $Na_V1.4$ channel were further investigated by docking isomers **1**, **2**, **4**, **7**, **11**–**15** (Table 1) to the channel (pdb ID 6AGF [20]) using the HADDOCK easy web interface (Figure 2, Figures S15 and S16). The remaining isomers **3**, **5**, **6**, **8**, **9**, **10**, and **16**–**18** were not further analyzed as they were considered to be inactive with respect to their poor pore blockage (Figure 1).

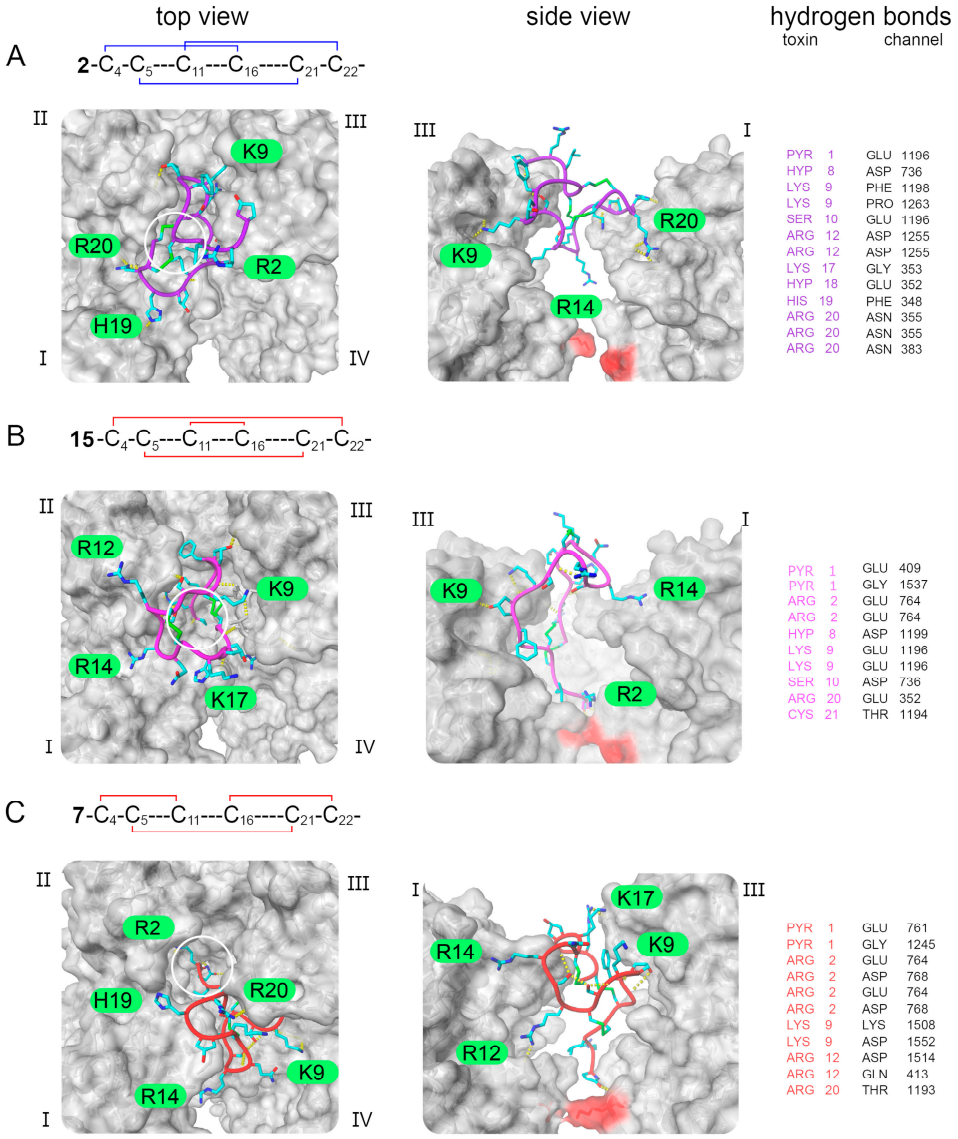


Figure 2. Visualization of μ-PIIIA–$Na_V1.4$ complex conformations obtained from docking experiments for (**A**) the native isomer **2**, (**B**) isomer **15**, and (**C**) isomer **7**. Left panel – top view of the toxin-channel complex. Middle panel – side view of the toxin-channel complex. The four $Na_V1.4$ domains are indicated and hydrogen bonds between the toxin and the channel are given as yellow dashed lines, specified in the right panel. The $Na_V1.4$ channel surface (molecular surface) is given in gray, the selectivity filter motif (DEKA) is highlighted in red. The toxin is shown with side chain atoms present (coloring scheme: carbon – cyan, nitrogen – blue, oxygen – red, sulfur – green, backbone – isomer 2: purple, 15: pink, 7: red).

The docking results were clustered according to the interface-ligand root mean square deviations (RMSDs) and scored according to the HADDOCK scoring function (Table S7). Multiple poses of the toxin isomers bound to the channel were observed and described in each case. The observations from

the docking experiments were in line with the experimental observations indicating that the native and most active isomer **2** (Figure 1, Figure 2A) preferred a binding conformation that covered the central pore, while the moderately active ones chose conformations that left large portions of the pore exposed (isomers **11**, **12**, **14**, **1** and **7**) and/or did not insert as deeply into the pore (isomers **13**, **14** and **15**) as the native isomer **2** (Figure 1, Figure 2B,C, Figures S15 and S16). Isomer **2** was observed to reside closest to the selectivity filter residues located at the center of the pore in a similar position as recently found for the structurally related μ-conotoxin KIIIA occluding the pore with its central R14, while tightly binding to three out of the four channel subunits (Figure 2A). The moderately active isomers, were found to bind away from the center of the pore at the interface of the subunits I, III, and IV leaving only a single toxin residue close to the pore (except for isomers **15** and **4**, Figure 1, Figure 2B,C, Supplementary Figures S14 and S15). Another general observation from our docking experiments was that the C- or N-terminus of the moderately active isomers was located closest to the selectivity filter region, again only allowing for an insufficient pore blockage by a single toxin residue.

However, at this point it shall be noted that the degree of pore blockage, which corresponds to the remaining current as revealed by the electrophysiological experiments, can only partly be rationalized from the docked pose of the isomers since the absolute degree of pore block could only be determined for isomers **2**, **15** and **7** (Figures 1 and 2). Nevertheless, the binding poses of isomers **2**, **15** and **7** might explain their decreasing ability to block the pore (Figure 2) unveiling a smaller number of hydrogen bonds with residues close to the selectivity filter for **15** and **7** compared to the native isomer **2** (Figure 2). Moreover, **15** and **7** were found to only block the pore by their flexible *N*-terminal residues Z1 and R2 (Figure 2). In contrast to isomer **2** and **15**, the center of isomer **7** is not located above the pore leaving enough space for sodium ions to get close to the selectivity filter region, which might explain the even lower degree of pore block compared to **2** and **15** (Figure 2.).

2.3. MD- and Molecular Electrostatic Potential (MEP)-Based Analysis

2.3.1. Analysis of 3-Disulfide-bonded μ-PIIIA Isomers **1**–**15**

With the biological activity of the 3-disulfide-bonded μ-PIIIA isomers **1**–**15** and the 2-disulfide-bonded analogs **16**–**18** determined via electrophysiological assessment, and their probable binding modes estimated by molecular docking, we conducted MD simulations to determine the impact of the distinct disulfide bonding on conformational stability and dynamics. From the electrophysiological investigation, it was evident that the native isomer **2** (C4-C16/C5-C21/C11-C22) of μ-PIIIA showed the highest activity by blocking >90% of channel current (Figure 1). The biological activity dropped to ~50% block for isomer **15** (C4-C22/C5-C21/C11-C16) and even below for all other active analogs **11**, **14**, **1**, **7**, and **12**. Based on this premise, two independent 400-ns MD simulations were conducted on each investigated disulfide isomer. The most obvious finding from the MD simulations was that isomer **2** was the most thermodynamically stable among the 3-disulfide-bonded analogs studied herein. All the structural and energetic properties computed from the two independent simulations (Figures S3–S14) were almost identical for the trajectories of isomer **2**, while the other 14 analogs, despite converging to similar conformations and measured average properties, took diverging paths to equilibration. The ΔG_{solv} estimated from atomic hydrophobicity and surface solvent accessibility [21,22] was the lowest for **2** among all the 3-disulfide isomers (Figures S3 and S4). This is an indicator that the folded conformation of **2** is energetically slightly preferred over e.g., isomers **15**, **11**, and **14**. From Figure 3, Figures S5 and S6 it is evident that the structure of isomer **2** exhibits good backbone stability with the average root mean square deviation (RMSD) value between the two trajectories of 2.55 ± 0.47 Å (Table S3). From the structural superimpositions (Figure 3A) and root mean square fluctuation (RMSF) plot (Figure 3E, Figure S7) it is evident that the leading three *N*-terminal residues possess greater flexibility than the rest of the structure as this part of the sequence is not constrained by disulfide bonds.

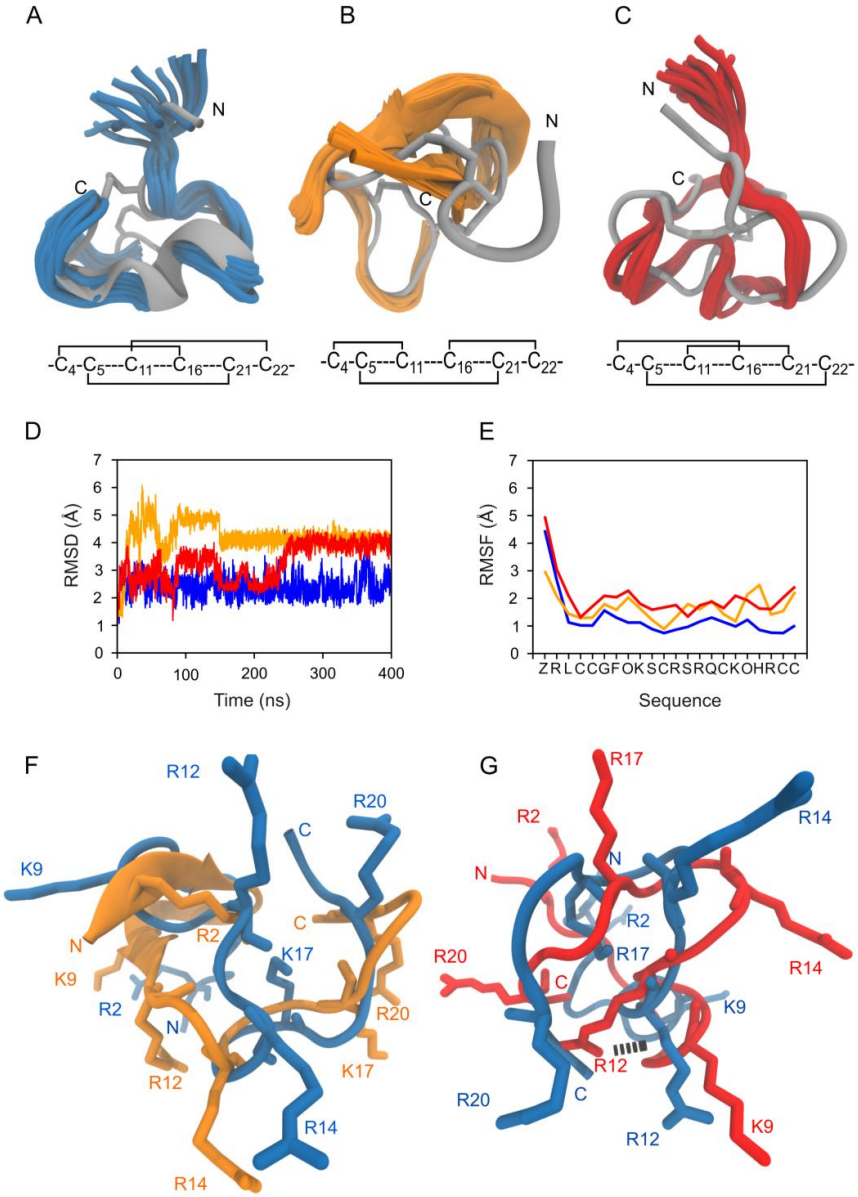


Figure 3. MD-based structural comparison of isomers **2**, **7**, and **10**. (**A**–**C**) 20 cartoon styled simulation snapshots of isomer **2** (blue), **7** (orange), and **10** (red) superimposed on their respective starting NMR structures (gray). The disulfide connectivity of each isomer is given below the individual isomer images. (**D**) Overlay of backbone RMSD progression plots for isomers **2** (blue), **7** (orange), and **10** (red) from a 400-ns MD simulation. (**E**) Overlay of per-residue root mean square fluctuation (RMSF) plots for isomers **2** (blue), **7** (orange), and **10** (red) from a 400-ns MD simulation. (**F**) Cartoon style structural superimposition of the representative structure of isomer **7** (orange) from MD simulation over the representative structure of isomer **2** (blue). The lysine and arginine residues significant for channel blocking are shown as sticks and labelled for both isomers. (**G**) Cartoon style structural superimposition of the representative structure of isomer **10** (orange) from MD simulation over the same of isomer **2** (blue). The lysine and arginine residues significant for channel blocking are shown as sticks and labelled for both isomers.

Comparing the structure and dynamics of **7** with **2** reveals that in terms of backbone mobility, the key difference again lies in the flexibility of the *N*-terminal residues (Figure 3). From the superimposed structures in Figure 3B, we noticed that the C4-C11 disulfide bond renders its bracketed residues to adopt an anti-parallel β-sheet conformation (Figure 3B). This secondary structural adoption reorients

the R2 residue towards the core of the peptide where it forms intramolecular H-bonds with C11 at the backbone in 40% of the trajectory, thus keeping it away from possible contacts with the channel. Added to this, **7** has the largest Molecular Electrostatic Potential (MEP) among all the studied isomers at 11.49 kJ/mol keeping the peptide bound to the channel by electrostatic attraction. The reduced channel-blocking ability of **7** primarily results from the orientations of its basic side chain residues, which are significantly different from the native **2** (Figure 2F). Moreover, the key residue R12 was involved in intramolecular H-bonds with Q15 for ~25% of the simulation taking it out of play for making contacts with the channel. **10** (C4-C16/C5-C22/C11-C21), whose disulfide architecture is the closest to that of **2** (Figure 2C), is surprisingly the isomer with the lowest measured bioactivity (Suppl. Figure 1). The isomer's backbone undergoes an average change of 3.3 ± 0.6 Å RMSD to reach an equilibrated state (Figure 3D) with the average fluctuations of all its residues being much higher than those of **2** (Figure 3E). Structural superimpositions of **10** on **2** reveals clear misalignments between their key arginine and lysine residues. Furthermore, a series of intramolecular H-bonds by these key residues R2, R12, R14 and K17 for up to ~60% of total simulation time render them unavailable for interacting with the channel residues. In addition, **10** has the lowest MEP value at 6.76 kJ/mol leaving it at a disadvantage in terms of favorable electrostatics to retain it near the channel pore. Despite the stable backbone conformations observed for **8**, it still falls in the inactive set for the same reason as **10**, the formation of intramolecular H-bonds between the side chains of key residues that take them out of play from channel binding. In **8**, H-bond interactions between R12 and Z1 for ~65% of the simulation, K9 and Q15 for ~8% of the simulation, R20 and K17 for ~10% of the simulation, K17 and C21 for ~24% of the simulation and the side chain of R2 to its own main chain for ~5% of the simulation ensure that these residues are almost never available to make contacts with the channel. Structures from the MD trajectories were clustered based on a backbone RMSD threshold of 1 Å using the "single-linkage" algorithm to facilitate the selection of representatives out of isomers that sample diverse conformations. From Table S4 it is evident that some of the less active isomers (e.g., **3**, **4**, **5**, and **9**) result in many clusters and their largest clusters represent only a small percentage of the trajectory. The corresponding RMSD plots (Figures S5 and S6) also convey the same message. It can be deduced that the extreme conformational instability in isomers **3** (C4-C5/C11-C16/C21-C22), **4** (C4-C5/C11-C21/C16-C22), **5** (C4-C5/C11-C22/C16-C21), **6** (C4-C11/C5-C16/C21-C22), and **9** (C4-C16/C5-C11/C21-C22), arise from their disulfide bonding architecture of pairing between adjacent cysteines with no crosslinking disulfides to stabilize the segment. Hence they all fall into the inactive category. Visual inspection of the MD trajectory of these isomers revealed excessive rotational and translational motion which can be quantified by their RMSDs averaging at ~ 6 Å from their initial conformations (Table S3). In terms of channel binding this means that these isomers are never found in a conformation to form stable contacts with the channel residues and leave the pore area constantly exposed. For the isomers in the active category (**1**, **7**, **11**, **12**, **14** and **15**), a combination of moderate to good backbone stability and favorable orientations of at least some of the key residues result in conformations that could partially but stably block the channel pore.

Additionally, 400-ns long refinement MD simulations of the isomer-channel complexes were conducted with a membrane system embedding the channel for the isomers **2**, **7**, and 15. All of the simulation parameters between the these isomer-channel complex MD simulations and the isomer in solution simulations were kept identical to maintain uniformity of the systems and comparability of results. All the 3 isomer-channel systems had attained equilibration providing molecular level insights into the toxin-channel interactions and how this related to the extent of pore block obtained from the electrophysiological studies. Differences in the dynamic behavior of these three isomers between their unbound and channel-bound forms were measured in terms of average RMSD, radius of gyration (Rg), and solvent accessible surface area (SASA) values (Table S8). As expected, the mean SASA values of the channel-bound forms of the isomer were comparatively lower than their unbound counterparts (Table S8) owing to the fact that some part of the peptide's solvent accessible surface is not available to the solvent but to the channel to interact with. An estimate of the binding energy (E_{bind}) was computed from the simulation and the values (Table S9) showed a clear correlation with the binding potency

quantified by the electrophysiological results. Isomer **2**, the best binder among all the isomers had an E_{bind} of -99.56 kJ/mol, while isomer **15** had -144.71 kJ/mol and finally isomer **7** with an E_{bind} of -332.83 kJ/mol (Table S9), preserving the trend observed from the electrophysiological experiments. As observed in the docking simulations, isomer **2** was found to occupy the central region of the channel with its R14 residue occluding the pore facilitating channel block at the beginning of the simulation. Over the course of the 400-ns simulation, it was observed that the isomer underwent a conformational change accounting to an average backbone RMSD of 3.2 Å. ~95% of conformations sampled by the isomer had the R14 residue placed directly above the channel pore persistently forming H-bonding interactions with the channel residue E184 making sure that the pore was covered. Additional isomer-channel H-bonding interactions namely Z1-D472, R12-L475, and R12-N474, stabilized the residence of the isomer **2** over the channel pore. For ~5% of the simulation, the isomer adopted a conformation where the sidechain of R14 reoriented away from its original position covering the pore and formed intramolecular H-bonds with the S13 residue on its own backbone. During this event it was observed that the channel pore was clearly exposed in a way that allows the passage of ions. This could be a reason as to why the channel was not blocked to 100% by this isomer (Figure 1). The docked conformation of isomer **15** had its *N*-terminal oriented towards the interior with Z1 and R2 residues occupying the area over the pore. A H-bond between the residues R2 of the isomer and D500 of the channel added to the initial stability of the configuration. As the simulation progressed, the isomer adopted an altered conformation in which the channel pore was only partly covered, and no single residue occluded the pore. The isomer underwent an average conformational change of 4.36 Å backbone RMSD during the simulation. For isomer **7**, the starting structure (docked complex conformation) in simulation had the isomer oriented over the channel pore with the Z1 residue found occluding the pore forming a H-bond with the channel residue G773. This conformation was different from what was observed in the MD simulation of this isomer in solution wherein the *N*-terminal residues formed an anti-parallel ß-sheet (Figure 3B). During the simulation it was observed that isomer 7 quickly (under 5 ns) adopted a similar conformation to that in solution forming anti-parallel ß-sheets between residues C5-O8 and K9-C11, pulling the Z1 residue away from the channel pore. Hence the channel pore remained exposed for large parts of the simulation. A small fraction of the conformations (~15%) sampled by the isomer involved the R2 residue stretching out over the pore providing the partial block. For the rest of the simulation this residue was found H-bonded to T772 which is a part of the loop structure of segment II of the channel, leaving the pore constantly exposed. The isomer had an overall RMSD of 1.69 Å during the simulation indicating that it was stable in its bound state although it did not block the pore sufficiently. It was also observed that this isomer's most preferred conformation, where the pore remains exposed, is different by 5.13 Å from its less preferred conformation where pore block was observed. This could further relate to the large time constant of block observed for this isomer because a large conformational change was required for channel block to be initiated. Finally, the per-residue contribution to the solvent accessible surface area ($SASA_{res}$) was computed for the three isomers and compared against their unbound forms (Table S10). Residues found close to the channel pore (e.g., R14 for isomer **2**, Z1 and R2 for isomer **15** and Z1 for isomer **7**) had lower $SASA_{res}$ values compared to their unbound versions, serving as a passive indicator to their involvement in and the extent of pore blockage (Table S10). The observations from these simulations serve as a guide to understanding the molecular details of how the difference in disulfide connectivity between these three selected isomers influence the way they bind to the channel, thereby defining the function.

2.3.2. Analysis of 2-Disulfide-bonded μ-PIIIA Isomers **16**–**18**

We studied the structural and functional contribution of the individual disulfide bonds to the native isomer **2** of μ-PIIIA. From the electrophysiological experiments, significant loss of channel-blocking activity among all the three 2-disulfide analogs (**16**, **17**, and **18**) was demonstrated. MD analysis revealed that all the three 2-disulfide analogs underwent a reasonable amount of backbone conformational change (Figure 4). Among the three, the backbone of isomer **18** lacking the C5-C21 disulfide bond evolved to be the most stable as observed from the RMSD profiles in Figure 4. **18** also had the least

increase of per-residue fluctuations as shown in the RMSF comparison plots in Figure 4. This results in the marginally higher activity seen for **18** over **16** in Figure S2. In all the three 2-disulfide-bonded analogs, clear misalignment of the key basic residues was observed in comparison to the native 3-disulfide-bonded isomer **2** and hence the loss of bioactivity (Figure 4J, K, L). A recent report using a modified version of μ-PIIIA produced similar results with respect to assessing the role of individual disulfide bonds on the structure and bioactivity of μ-PIIIA [23], although in this study the *N*-terminal pyroglutamate residue was replaced by proline and respective cysteines were substituted by alanine residues.

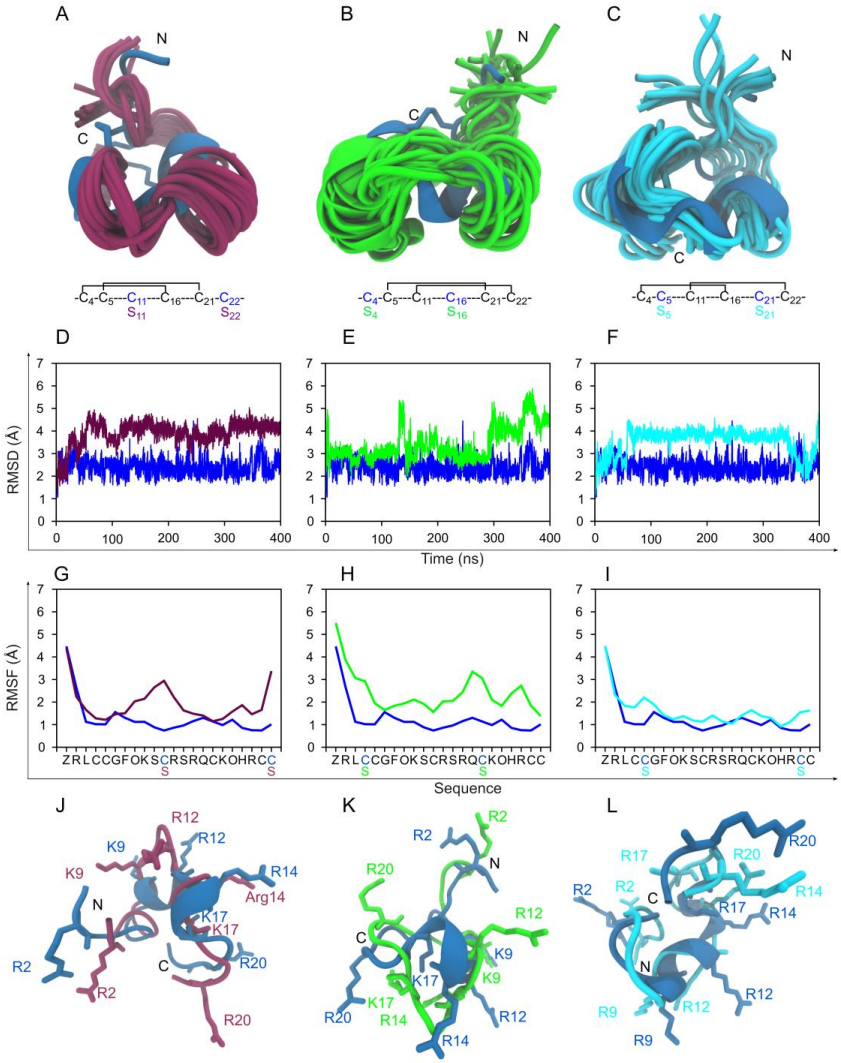


Figure 4. MD-based structural comparison of the 2-disulfide-bonded μ-PIIIA analogs **16**, **17**, and **18**. (**A**–**C**) 20 cartoon styled simulation snapshots of isomer **16** (maroon), 17 (green), and **18** (cyan) superimposed on the starting NMR structure of μ-PIIIA isomer **2** (blue). The disulfide connectivity and the C-S mutations of each isomer are given below the individual isomer images. (**D**) Overlay of backbone RMSD progression plots of isomer **16** (maroon) on the native isomer **2** (blue). (**E**) Overlay of backbone RMSD progression plots of isomers **17** (green) on the native isomer **2** (blue). (**F**) Overlay of backbone RMSD progression plots of isomers **18** (cyan) on the native isomer **2** (blue). (**G**) Overlay of per-residue RMSF plots of isomer **16** (maroon) on the native isomer **2** (blue). (**E**) Overlay of per-residue RMSF plots of isomer **17** (green) on the native isomer **2** (blue). (**F**) Overlay of per-residue RMSF plots of isomers **18** (cyan) on the native isomer **2** (blue). (**J**–**L**) Cartoon styled representative structures from MD for isomers **16** (maroon), **17** (green) and **18** (cyan) superimposed on the structure of the native isomer **2** (blue). The functionally significant lysine and arginine residues are shown as sticks and labelled.

2.3.3. Analysis of 2-Disulfide-bonded μ-PIIIA Isomers Δ(C5-C21)2, Δ(C11-C21)4, and Δ(C5-C22)10

The concept of disulfide bonds holding a "structurally frustrated" peptide in its bioactive form is known from the literature [24]. A means to measure this inherent frustration induced by the disulfide bonding pattern can be studied by observing the dynamics of partially folded states of the peptide. Previous studies have also shown that this observation could give hints into the folding pathways that the peptide prefers to take as it folds [8]. Isomer **2**, **4**, and **10** have been selected as representatives of structurally well-defined (**2**), moderately flexible (**10**), and highly flexible (**4**) classes among the 15 μ-PIIIA isomers [16]. We created partially folded analogs of these isomers by arbitrarily removing the second disulfide bond in these structures creating the Δ(C5-C21)**2**, Δ(C11-C21)**4,** and Δ(C5-C22)**10** to observe the refolding preference of these partially folded states (Figure 5). Isomer Δ(C5-C21)**2** did not prefer to refold maintaining an average distance of 15 Å between the Sγ atoms of the opened C5-C21 disulfide bond, while in analog Δ(C5-C22)**10** at several time points the Sγ atoms of the opened C5-C22 disulfide bond came up to 5 Å close but eventually sampled unfolded conformations towards the end of the simulation (Figure 5). The analog Δ(C11-C21)**4** displayed an equal mix of both unfolded and folded states with respect to its parent isomer **4** structure with the distances between the unbound Sγ atoms of the opened C11-C21 disulfide bond fluctuating heavily between 3 Å to 30 Å (Figure 5). This analysis confirms that the folding preferences between different disulfide isomers of the same sequence can be highly diverse, further supporting that μ-PIIIA does require all three of its disulfide bonds to remain structurally stable and bioactive.

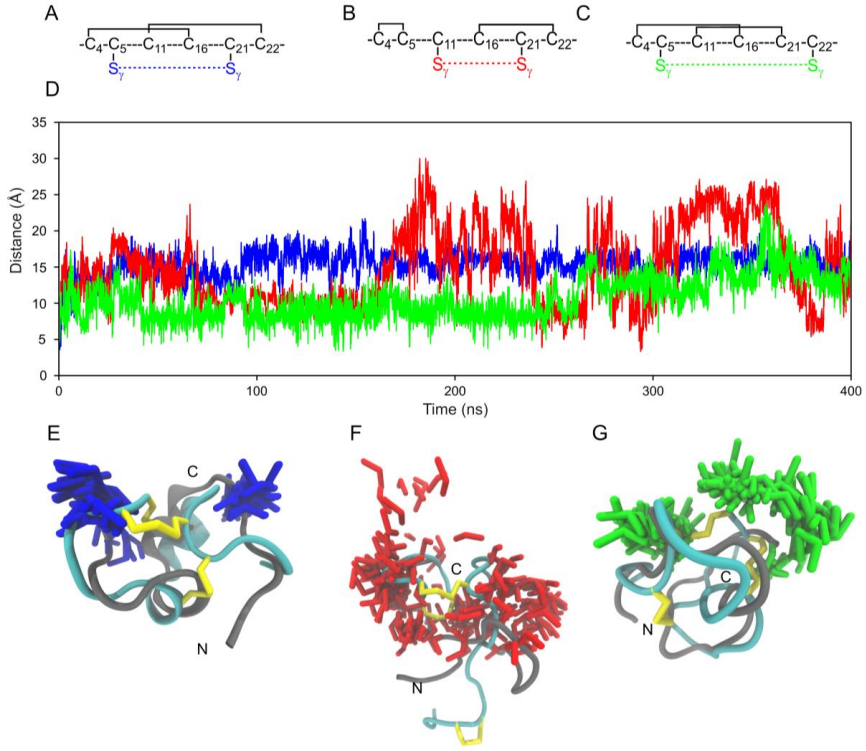


Figure 5. MD-based disulfide distance analysis of partially folded μ-PIIIA analogs Δ(C5-C21)**2**, Δ(C11-C21)**4** and Δ(C5-C22)**10**. (**A**–**C**) Disulfide connectivity (black solid lines) of the isomers. The colored dotted lines (blue – Δ(C5-C21)**2**, red – Δ(C11-C21)**4**, and green – Δ(C5-C22)**10**) represent the disulfide bonds opened *in silico*. (**D**) Overlay of distance plots between the *in silico* opened disulfide bonds (blue – Δ(C5-C21)**2**, red – Δ(C11-C21)**4**, and green – Δ(C5-C22)**10**). (**E**–**G**) The cyan cartoon structures represent the NMR structures of the isomers **2**, **4**, and **10** with their disulfide bonds shown as yellow sticks. The overlaid gray cartoon is a single MD snapshot of the partially folded 2-disulfide-bonded versions of the isomers with 100 conformations of the opened disulfide bond versions (blue – Δ(C5-C21)**2**, red – Δ(C11-C21)**4**, and green – Δ(C5-C22)**10**) to represent the distances sampled between them during simulation.

3. Discussion

The significance of disulfide bonds in the stability and folding of peptides and proteins is well-established and accepted [25]. Beyond their involvement in proteins for thermodynamic and kinetic stability [26], disulfide bonds also play multifaceted roles in peptides with therapeutic potential [27]. While the so-called "native fold" is often what is studied with much rigor, we have in recent studies ventured into exploring all alternative folds using the 3-disulfide-bonded μ-PIIIA as target [17]. In the present study, we have clarified using a combination of electrophysiological and computational approaches that despite the native fold (isomer **2**) still carrying the highest bioactivity, 6 of the remaining 14 isomers are still bioactive in a range between 30%–50% of maximum current block compared to **2**. Due to its globular conformation, homogenous distribution of functionally significant basic arginine and lysine residues and their stable orientations [17,28,29] indicated by the lowest RMSF values among the 15 isomers studied, isomer **2** is both structurally and energetically the most favorable to optimally block the channel (Figure 3). It is also the largest populated isomer and the most easily formed with ΔG_{solv} computation [23] serving as an indicator of stability.

Our current findings help to clarify questions about ambiguities regarding the potencies of the different isomers. In an earlier study (Tietze et al., 2012) [10], isomer **3** was active (~50% as active as the native isomer), while it was inactive in the present study. However, taken into account the recent study by Heimer et al., 2018 [16], we are now able to state that **3**, when obtained from the oxidative self-folding approach, obviously contained other bioactive isomers, such as **2** or **15,** as confirmed by corresponding HPLC elution profiles of the individual mixtures [16]. In addition, these impurities explain why NMR analysis could not be obtained for this isomer in our first study (Tietze et al., 2012) [10] but could be elucidated upon targeted synthesis and purification of isomer **3** by Heimer et al. [16]. In the same study by Tietze et al., 2012 [10], isomer **1** was more active than the natively folded isomer **2**. Considering our recent report [16], this discrepancy can also be explained by the HPLC analysis of the isomer mixtures. In this recent report, it was demonstrated that isomer **1** cannot unambiguously be separated from **2** [16], which was also confirmed by impurities (unassigned) found in the NMR spectra, although the analysis of the structure of isomer **1** was still feasible at the time.

The negligible to no activity of **8** and **10** represents an interesting case of how structural stability does not always relate to bioactivity and adds evidence to the notion that folding proceeds in the direction that preserves function [29], and that even slight alterations of the disulfide architecture can at times be functionally inefficient. The current study also provides experimental proof for an earlier in silico study which suggested that native μ-PIIIA does certainly need all its 3 disulfide bonds to remain functional [8] compared to other conotoxins such as μ-KIIIA [7] and α-conotoxin Vc1.1 [30]. In summary, using electrophysiology, molecular docking, MD simulations, and MEP analysis, we have established a molecular, structural and electrostatic basis for the impact of disulfide connectivity on the bioactivity of μ-PIIIA isomers. We have additionally used simulations to explain the loss of bioactivity in the 2-disulfide-bonded μ-PIIIA analogs.

4. Materials and Methods

4.1. Peptide Synthesis and Purification

Synthesis, purification, and chemical and structural analysis of the 15 3-disulfide-bonded PIIIA isomers **1**–**15** (sequence: ZRLCCGFOKSCRSRQCKOHRCC-NH_2, Table S1) has been described previously [16]. Freeze-dried aliquots of known concentrations for each isomer, as determined by amino acid analysis and HPLC, were submitted to electrophysiological experiments. In addition, three 2-disulfide-bonded isomers **16**–**18** were prepared and subjected to biological activity testing as well. In brief, these μ-PIIIA-mutants were automatically assembled using 9-fluorenyl-methyloxycarbonyl (Fmoc) chemistry and couplings employing 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) and *N*-methylmorpholine/DMF (1:1) on a Rink-amide MBHA resin (0.53 mmol/g) and an EPS221 peptide synthesizer (Intavis Bioanalytical Instruments AG, Cologne,

Germany). The different cysteine pairs were protected with Trt- and Acm-groups, or exchanged with Ser according to the desired disulfide connectivity (Table S2). Peptide cleavage and concomitant side chain deprotection was accomplished with reagent K (phenol/thioanisole/ethanedithiol 1:1:0.5, 150 µl/100 mg resin) and TFA (95%, 1 mL/100 mg resin) for 3 h. The crude peptides were precipitated with diethyl ether, centrifuged, and washed several times with diethyl ether. Linear precursor peptides were purified as previously described [16] and chemically characterized (Table S3). The oxidation of the 2-disulfide bonded peptides **16–18** was performed in a mixture of isopropanol/water/1 M HCl (31:62.5:7.5) with a final peptide concentration of 0.05 mM. Then, 1.1 equiv. iodine (0.1 M in MeOH) was added to the peptide solutions, and the reaction was stirred for 1 h at room temperature. After the first disulfide bond was closed, further 13.9 equiv. iodine (0.1 M in MeOH) was added to deprotect the Acm-group and to form the second disulfide bond. After the reaction was stirred for up to 1 h, the oxidation was stopped by the addition of an excess of ascorbic acid (1 M in water). Reaction progress was monitored via RP-HPLC and LC-ESI-MS. The products were freeze-dried and stored at −20 °C. The purification of the crude peptides was performed by semi-preparative RP-HPLC as reported [16]. Peptide purity and constitution was confirmed by analytical RP-HPLC, MALDI-TOF mass spectrometry and amino acid analysis (Figure S1, Table S3) using methods and instruments as described previously [16]. Amino acid analysis was used for the determination of peptide concentrations in solution prior to the further experiments to ensure equal concentrations in analysis and biological testing. In general, µ-PIIIA derivatives were dissolved in double distilled water at a stock concentration of 10 µM and further diluted as described in the electrophysiological experiments.

4.2. Electrophysiological Experiments

Human SCN4A (encoding the $Na_V1.4$ channel α subunit, UniProt ID P35499) on a plasmid with *Cytomegalovirus* (CMV) promoter was transiently expressed in HEK 293 cells as shown previously [10]. Co-transfection of a plasmid encoding the CD8 antigen ensured visual detection of transfected cells with CD8-specific Dynabeads (Deutsche Dynal, Hamburg, Germany). Currents were measured with the whole-cell patch clamp method 24–48 h after transfection [10]. The patch pipettes contained (in mM): 35 NaCl, 105 CsF, 10 EGTA (ethylene glycol bis(2-amino-ethylether)tetraacetic acid), 10 HEPES (pH 7.4 with CsOH). The bath solution contained (in mM): 150 NaCl, 2 KCl, 1.5 $CaCl_2$, 1 $MgCl_2$, 10 HEPES (pH 7.4 with NaOH). Holding potential was –120 mV, Na^+ currents were elicited with depolarizing steps to −20 mV. Series resistance was corrected electronically up to 80%. Peptides, diluted in the bath solution, were applied focally to cells under consideration with a fine-tipped glass capillary. Time course of peak current decrease after peptide application was described with single-exponential functions.

4.3. Molecular Modeling and Docking Simulations

The toxin-channel binding was predicted by docking the lowest energy conformer of the NMR structures of the µ-PIIIA isomers **1**, **2**, **4**, **7**, **11-15** [16] to the recently solved structure of the human $Na_V1.4$ voltage-gated sodium channel (pdb ID 6AGF [20]) using the Easy Interface of the HADDOCK online platform [31,32] (https://haddock.science.uu.nl/services/HADDOCK2.2/haddockserver-easy.html), a web service known to be suitable for handling more complex peptide ligand structures [32]. The receptor structure was energy-minimized in explicit solvent (TIP3 water) before the docking runs.

For the docking process, all toxin residues and the channel residues that are part of the channel's upper surface were defined as "active", because they were assumed to be able to form contacts with the toxin. As "passive" channel residues we defined all residues within a radius of 6.5 Å from the "active" residues. From the docking results, the best scoring structure from the highest scoring complex cluster was selected for further analysis (Table S7). As the HADDOCK web interface cannot handle γ-pyroglutamic acid, glutamic acid was used instead, and was re-converted to γ-pyroglutamic acid, followed by a subsequent energy minimization step for further analysis.

Analysis and energy minimizations were performed using the YASARA molecular modeling software (YASARA structure, Vers. 18.3.23, YASARA Biosciences GmbH, Vienna, Austria) [33,34].

The energy minimizations of the toxin-channel complex were performed in explicit water (TIP3P) and the Particle Mesh Ewald (PME) method [35] in order to describe long-range electrostatics at a cut-off distance of 8 Å in physiological conditions (0.9% NaCl, pH 7.4), at a constant temperature (298 K) using a Berendsen thermostat, and with constant pressure (1 bar). The charged amino acids were assigned according to the predicted pKa of the amino acid side chains from the Ewald summation, and were neutralized by adding counter ions (NaCl). The YASARA2 force field was used for energy minimization by simulated annealing, including the optimization of the hydrogen bond network and the equilibration of the water shell, until system convergence was achieved.

The molecular graphics were created using YASARA (YASARA structure, Vers. 18.3.23, YASARA Biosciences GmbH, Vienna, Austria, www.yasara.org) and POVRay (Persistence of Vision Raytracer Pty. Ltd., Williamstown, Australia www.povray.org).

4.4. MD Simulations of the μ-PIIIA Isomers and Analogs

All-atom unbiased MD simulations of the 15 3-disulfide-bonded isomers (**1**–**15**), three 2-disulfide bonded analogs (**16**–**18**) and three partially folded 2-disulfide bonded models **2-2S**, **4-2S** and **10-2S** were conducted with the GROMACS 2018 program [36,37]. The 3D coordinates for isomers **1**–**15** were obtained from the NMR structural ensembles reported [16]. The first out of the 20 structures of the ensemble was taken as the starting conformation for MD simulations. Analogs **16**–**18** were modeled in YASARA [34] by replacing the respective cysteine residues by serine. These models were energy-minimized using the AMBER99SB-ILDN [38] force field (Amber 11 in YASARA) before being used for simulations. The partially folded conformations were created in GROMACS via the *pdb2gmx* module as earlier described [8]. All steps and parameters involved in preparing the MD system were taken from the previous work [8] that involved the simulation of the native isomer of μ-PIIIA, except that the simulation run times increased herein. Solvent equilibration for 50 ns in constant temperature and constant-volume NVT ensemble, and 50 ns in the constant temperature and constant-pressure NPT ensemble conditions were carried out with position restraints on the energy-minimized peptide prior to the production MD run. Production runs at 300 K were conducted for 400 ns with a 2 fs time step. Two independent simulations were conducted per peptide using the random seed method that assigns different initial velocities (from the Boltzmann distribution) to the system thereby ensuring adequate sampling of the conformational space. 10,000 frames from each trajectory were used for analysis. The RMSD of the protein backbone atoms, per-residue RMSF of all peptide atoms, the radius of gyration (Rg) of all peptide atoms, and the SASA were computed using analysis tools within GROMACS. The double cubic lattice method [39] with the default probe radius of 1.4 Å was used to compute the SASA as well as the per-residue contribution to the solvent accessible surface ($SASA_{res}$). An estimate of solvation free energies (ΔG_{solv}) from the atomic solvation energies per exposed surface area as implemented by Eisenberg et al. [21] were computed and plotted as a function of simulation time. Single-linkage clustering analysis was done on each trajectory using the backbone RMSD cut-off of 1 Å as the metric for the clustering. The centroid of the largest cluster from each trajectory was used as a representative for further analysis. Visualizations of the MD trajectories, molecular graphics and movies from the simulations were created in Visual Molecular Dynamics (VMD) (version 1.9.3) [40]. Plots were created using the Grace program (version 5.1.25).

4.5. MD Simulations of the μ-PIIIA Isomer-Channel Complexes

Owing to the ease of setting up a membrane simulation, YASARA was used in conducting 400-ns long MD simulations of isomers **2**, **7**, and **15** bound to the recently resolved cryo-electron microscopy structure (PDB 6AGF [20]) of the human voltage-gated sodium channel $Na_V1.4$. The 3 isomers were selected based on a consensus obtained from the electrophysiological experiments. The preparation of the membrane embedded system was achieved by providing the docked

isomer-channel complex as input to the *md_runmembrane.mcr* macro in YASARA. The program first identified the transmembrane helices by scanning the channel structure for secondary structure elements with hydrophobic surface residues to orient and embed a membrane composed of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) fatty acid residues. The protein was oriented such that the axis through the transmembrane helices is perpendicular to the membrane. A cubic simulation cell large enough to enclose the entire membrane was then drawn. The cell was extended by 15 Å on either side of the membrane so that the membrane was 30 Å larger than the protein ensuring that during simulation, the protein never sees its periodic image. The simulation cell was filled with the water molecules including enough number of Na^+ and Cl^- ions to achieve a physiological concentration of 0.9% as a mass fraction. Next, successive steps of steepest descents energy minimizations followed by simulated annealing minimizations were done to remove bumps between the lipid molecules by deleting those lipids resulting in unfavorable geometries. The minimization steps were done iteratively until the potential energy of the system was optimized. Any solvent molecule found within the lipids was removed and the membrane was slowly packed closer by iteratively reducing the size of the simulation cell. As the final step, the membrane was artificially stabilized by position restraints and a pre-production simulation for 250 ps was done to equilibrate the solvent around the membrane and protein. The equilibrated system was composed of a total of ~163000 atoms including ~300 POPC molecules making up the membrane, within a cubic simulation cell with periodic boundaries of side ~120 Å was ready for the production MD simulation. A 2 fs time step was used to run the 400-ns long production MD simulations. Temperature was maintained at 300 K by a velocity rescaling algorithm and the pressure maintained at 1 bar by rescaling the volume. The cut-off for non-bonded forces was set at 8 Å. Analysis of the resulting trajectories were done via the standard macros within YASARA (*md_analzye.mcr* and *md_analyzeres.mcr*). An estimate of the binding energy E_{bind} was obtained using the Poisson-Boltzmann method [41,42] via the *md_analyzebindenergy.mcr* macro in YASARA by analyzing each frame in the 400-ns MD simulation. The E_{bind} was given by $E_{bind} = E_{pot}Recept + E_{solv}Recept + E_{pot}Ligand + E_{solv}Ligand - E_{pot}Complex - E_{solv}Complex$ where the subscripted "pot" and "solv" represent potential energy and solvation energy, respectively. Positive values from this result indicate better association between the ligand and receptor although negative values do not indicate no binding. The values obtained from the 3 simulations were compared against each other to estimate and compare the nature of the toxin-channel interactions for the 3 isomers (**2**, **7**, and **15**) in this study.

4.6. MEP Calculations

MEP calculations were made for multiple conformations for all peptides used in this study to help quantify the differences in the surface electrostatic properties. MEP calculations were done using the Poisson-Boltzmann method [41,42] via the Adaptive Poisson-Boltzmann Solver (ABPS) program [43] built into YASARA. The method uses an implicit solvation model wherein the solvent is treated as a high dielectric ($\varepsilon = 80$) continuum. To maintain consistency with the MD simulations, the AMBER99SB-ILDN force field was used. Frames from the last 300 ns of the MD trajectory were used as inputs for the MEP calculation.

Supplementary Materials: The following are available online at http://www.mdpi.com/1660-3397/17/7/390/s1, Figures S1–S16 and Tables S1–S10.

Author Contributions: P.H., T.S., and D.I. planned and performed synthesis and characterization of the peptides. E.L. and S.H.H. planned and conducted electrophysiological experiments. Molecular modeling, docking and simulation studies have been planned and carried out by A.A.P.G., D.K., D.T., and D.I. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding: Financial support by the University of Bonn (to D.I.) is gratefully acknowledged.

Acknowledgments: We are thankful to Tim Wermund (University of Bonn) for technical assistance in peptide synthesis and purification.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Weissman, J.S.; Kim, P.S. Response. *Science* **2006**, *256*, 112–114. [CrossRef] [PubMed]
2. Wedemeyer, W.J.; Welker, E.; Narayan, M.; Scheraga, H.A. Disulfide bonds and protein folding. *Biochemistry* **2000**, *39*, 4207–4216. [CrossRef] [PubMed]
3. Cemazar, M.; Zahariev, S.; Lopez, J.J.; Carugo, O.; Jones, J.A.; Hore, P.J.; Pongor, S. Oxidative folding intermediates with nonnative disulfide bridges between adjacent cysteine residues. *Proc. Natl. Acad. Sci.* **2003**, *100*, 5754–5759. [CrossRef] [PubMed]
4. Garrett, J.E.; Jimenez, E.C.; Kranski, J.; Olivera, B.M.; Bulaj, G.; Buczek, O.; Goodsell, I.; Nielsen, J.S. Efficient oxidative folding of conotoxins and the radiation of venomous cone snails. *Proc. Natl. Acad. Sci.* **2003**, *100*, 14562–14568.
5. Gao, B.; Peng, C.; Yang, J.; Yi, Y.; Zhang, J.; Shi, Q. Cone snails: A big store of conotoxins for novel drug discovery. *Toxins (Basel).* **2017**, *9*, 397. [CrossRef] [PubMed]
6. Chang, J.Y. Diverse pathways of oxidative folding of disulfide proteins: Underlying causes and folding models. *Biochemistry* **2011**, *50*, 3414–3431. [CrossRef] [PubMed]
7. Watkins, M.; Yoshikami, D.; Balaram, P.; Khoo, K.K.; Norton, R.S.; Olivera, B.M.; Zhang, M.-M.; Gupta, K.; Green, B.R.; Bulaj, G. Distinct Disulfide Isomers of μ-Conotoxins KIIIA and KIIIB Block Voltage-Gated Sodium Channels. *Biochemistry* **2012**, *51*, 9826–9835.
8. Paul George, A.A.; Heimer, P.; Maaß, A.; Hamaekers, J.; Hofmann-Apitius, M.; Biswas, A.; Imhof, D. Insights into the Folding of Disulfide-Rich μ-Conotoxins. *ACS Omega* **2018**, *3*, 12330–12340. [CrossRef] [PubMed]
9. Dutton, J.L.; Hogg, R.C.; Bansal, P.S.; Craik, D.J.; Adams, D.J.; Alewood, P.F. A New Level of Conotoxin Diversity, a Non-native Disulfide Bond Connectivity in α-Conotoxin AuIB Reduces Structural Definition but Increases Biological Activity. *J. Biol. Chem.* **2002**, *277*, 48849–48857. [CrossRef]
10. Tietze, A.A.; Tietze, D.; Ohlenschläger, O.; Leipold, E.; Ullrich, F.; Kühl, T.; Mischo, A.; Buntkowsky, G.; Görlach, M.; Heinemann, S.H.; et al. Structurally diverse μ-conotoxin PIIIA isomers block sodium channel Na V1.4. *Angew. Chemie Int. Ed.* **2012**, *51*, 4058–4061. [CrossRef]
11. Lee, H.S.; Carstens, B.B.; O'Donnell, T.; Craik, D.J.; Adams, D.J.; Berecki, G.; Tae, H.-S.; Sadeghi, M.; Daniel, J.T.; Deiteren, A.; et al. Structure-Activity Studies of Cysteine-Rich α-Conotoxins that Inhibit High-Voltage-Activated Calcium Channels via GABA B Receptor Activation Reveal a Minimal Functional Motif. *Angew. Chemie Int. Ed.* **2016**, *55*, 4692–4696.
12. Corpuz, G.P.; Jacobsen, R.B.; Jimenez, E.C.; Watkins, M.; Walker, C.; Colledge, C.; Garrett, J.E.; McDougal, O.; Li, W.; Gray, W.R.; et al. Definition of the M-conotoxin superfamily: Characterization of novel peptides from molluscivorous Conus venoms. *Biochemistry* **2005**, *44*, 8176–8186. [CrossRef] [PubMed]
13. Munasinghe, N.R.; Christie, M.J. Conotoxins that could provide analgesia through voltage gated sodium channel inhibition. *Toxins (Basel).* **2015**, *7*, 5386–5407. [CrossRef] [PubMed]
14. Tosti, E.; Boni, R.; Gallo, A. μ-Conotoxins modulating sodium currents in pain perception and transmission: A therapeutic potential. *Mar. Drugs* **2017**, *15*, 1–16. [CrossRef] [PubMed]
15. Leipold, E.; Ullrich, F.; Thiele, M.; Tietze, A.A.; Terlau, H.; Imhof, D.; Heinemann, S.H. Subtype-specific block of voltage-gated K+ channels by μ-conopeptides. *Biochem. Biophys. Res. Commun.* **2017**, *482*, 1135–1140. [CrossRef] [PubMed]
16. Heimer, P.; Tietze, A.A.; Bäuml, C.A.; Resemann, A.; Mayer, F.J.; Suckau, D.; Ohlenschläger, O.; Tietze, D.; Imhof, D. Conformational μ-Conotoxin PIIIA Isomers Revisited: Impact of Cysteine Pairing on Disulfide-Bond Assignment and Structure Elucidation. *Anal. Chem.* **2018**, *90*, 3321–3327. [CrossRef] [PubMed]
17. Akondi, K.B.; Muttenthaler, M.; Dutertre, S.; Kaas, Q.; Craik, D.J.; Lewis, R.J.; Alewood, P.F. Discovery, Synthesis, and Structure–Activity Relationships of Conotoxins. *Chem. Rev.* **2014**, *114*, 5815–5847. [CrossRef] [PubMed]
18. Chen, F.; Huang, W.; Jiang, T.; Yu, R. Determination of the μ-conotoxin PIIIA specificity against voltage-gated sodium channels from binding energy calculations. *Mar. Drugs* **2018**, *16*, 153. [CrossRef]

19. Ostroumov, V.; McMaster, D.; Al-Sabi, A.; French, R.J.; McArthur, J.R. Multiple, Distributed Interactions of μ-Conotoxin PIIIA Associated with Broad Targeting among Voltage-Gated Sodium Channels. *Biochemistry* **2010**, *50*, 116–124.
20. Pan, X.; Li, Z.; Zhou, Q.; Shen, H.; Wu, K.; Huang, X.; Chen, J.; Zhang, J.; Zhu, X.; Lei, J.; et al. Structure of the human voltage-gated sodium channel Na v 1.4 in complex with β1. *Science* **2018**, *362*. [CrossRef]
21. Eisenberg, D.; Mclachlan, A.D. Solvation energy in protein folding and binding. *Nature* **1986**, *319*, 199–203. [CrossRef] [PubMed]
22. Novotný, J.; Bruccoleri, R.; Karplus, M. An analysis of incorrectly folded protein models. *J. Mol. Biol.* **2004**, *177*, 787–818. [CrossRef]
23. Xu, X.; Xu, Q.; Chen, F.; Shi, J.; Liu, Y.; Chu, Y.; Wan, S.; Jiang, T.; Yu, R. Role of the disulfide bond on the structure and activity of μ-conotoxin PIIIA in the inhibition of Na $_{V}$ 1.4. *RSC Adv.* **2019**, *9*, 668–674. [CrossRef]
24. Zhang, Y.; Schulten, K.; Gruebele, M.; Bansal, P.S.; Wilson, D.; Daly, N.L. Disulfide bridges: Bringing together frustrated structure in a bioactive peptide. *Biophys. J.* **2016**, *110*, 1744–1752. [CrossRef] [PubMed]
25. Qin, M.; Wang, W.; Thirumalai, D. Protein folding guides disulfide bond formation. *Proc. Natl. Acad. Sci.* **2015**, *112*, 11241–11246. [CrossRef] [PubMed]
26. Dombkowski, A.A.; Sultana, K.Z.; Craig, D.B. Protein disulfide engineering. *FEBS Lett.* **2014**, *588*, 206–212. [CrossRef] [PubMed]
27. Góngora-Benítez, M.; Tulla-Puche, J.; Albericio, F. Multifaceted roles of disulfide bonds. peptides as therapeutics. *Chem. Rev.* **2014**, *114*, 901–926. [CrossRef] [PubMed]
28. Chen, R.; Robinson, A.; Chung, S.H. Mechanism of μ-conotoxin PIIIA binding to the voltage-gated Na+channel NaV1.4. *PLoS ONE* **2014**, *9*, e93267. [CrossRef]
29. Hogg, P.J. Disulfide bonds as switches for protein function. *Trends Biochem. Sci.* **2003**, *28*, 210–214. [CrossRef]
30. Craik, D.J.; Yu, R.; Kaas, Q.; Adams, D.J.; Berecki, G.; Jia, X.; Seymour, V.A.L.; Akcan, M. Less is More: Design of a Highly Stable Disulfide-Deleted Mutant of Analgesic Cyclic α-Conotoxin Vc1.1. *Sci. Rep.* **2015**, *5*, 13264.
31. Dominguez, C.; Boelens, R.; Bonvin, A.M.J.J. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737. [CrossRef] [PubMed]
32. Van Zundert, G.C.P.; Rodrigues, J.P.G.L.M.; Trellet, M.; Schmitz, C.; Kastritis, P.L.; Karaca, E.; Melquiond, A.S.J.; Van Dijk, M.; De Vries, S.J.; Bonvin, A.M.J.J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428*, 720–725. [CrossRef] [PubMed]
33. Krieger, E.; Vriend, G. YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics* **2014**, *30*, 2981–2982. [CrossRef] [PubMed]
34. Krieger, E.; Vriend, G. New ways to boost molecular dynamics simulations. *J. Comput. Chem.* **2015**, *36*, 996–1007. [CrossRef] [PubMed]
35. Lee, H.; Pedersen, L.G.; Essmann, U.; Darden, T.; Perera, L.; Berkowitz, M.L. A smooth particle mesh Ewald method. *J. Chem. Phys.* **2002**, *103*, 8577–8593.
36. Berendsen, H.J.C.; Hess, B.; Lindahl, E.; Van Der Spoel, D.; Mark, A.E.; Groenhof, G. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
37. Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory. Comput.* **2008**, *4*, 435–447. [CrossRef] [PubMed]
38. Klepeis, J.L.; Lindorff-Larsen, K.; Shaw, D.E.; Palmo, K.; Dror, R.O.; Maragakis, P.; Piana, S. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–1958.
39. Argos, P.; Eisenhaber, F.; Lijnzaad, P.; Scharf, M.; Sander, C. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.* **2005**, *16*, 273–284.
40. William Humphrey; Andrew Dalke; Klaus Schulten VMD – Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.
41. Fogolari, F.; Brigo, A.; Molinari, H. The Poisson-Boltzmann equation for biomolecular electrostatics: A tool for structural biology. *J. Mol. Recognit.* **2002**, *15*, 377–392. [CrossRef] [PubMed]

42. Baker, N.A. Poisson-Boltzmann Methods for Biomolecular Electrostatics. *Methods Enzymol.* **2004**, *383*, 94–118. [PubMed]
43. Holst, M.J.; McCammon, J.A.; Sept, D.; Baker, N.A.; Joseph, S. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci.* **2002**, *98*, 10037–10041.

4.2.2 Summary

The native isoform of the conotoxin μ-PIIIA with the C1-C4/C2-C5/C3-C6 disulfide connectivity is known to be a potent blocker of the VGSC Nav1.4. In this study we tried to investigate how this activity changes upon different altering the disulfide pairing. Accordingly, the fifteen theoretically possible disulfide isomers previously synthesized, were tested for their blocking activity on Nav1.4 using electrophysiological studies. It was clear that the isomer with the native disulfide pairing (μ-PIIIA 2) was the best active isomer with ~90% block at Nav1.4. The second most active isomer (μ-PIIIA 15) had upto ~50% loss of bioactivity and all of the two-disulfide mutants of native μ-PIIIA displayed poor activity. The MD and docking based computational investigations provided an improved understanding of the underlying causes to the experimentally observed behavior. It was clear that isomers with good blocking activity also showed reasonable conformational stability in the MD simulations. Isomers such as μ-PIIIA 3, μ-PIIIA 4 and μ-PIIIA 5 whose disulfide connectivity involved pairings with adjacent cysteine restudies were found to possess poor conformational stability in the MD simulations and correspondingly had poor blocking activity measured by the electrophysiological experiments. A similar correlation between bioactivity and conformational stability was observed for the serine-mutated two-disulfide versions μ-PIIIA 2. Finally, binding energies computed from the refinement MD simulations of the docked peptide-channel complexes were in correlation with the experimentally observed gradients in blocking activity. Overall, this study provides a clear structural, conformational and energetic basis for the impact of conformational diversity on the bioactivity of μ-PIIIA disulfide isomers.

4.3 Chapter III

Coagulation Factor XIIIa Inhibitor Tridegin: On the Role of Disulfide Bonds for Folding, Stability, and Function

Original research article

Authors*

Charlotte A. Bäuml, Thomas Schmitz, Ajay Abisheck Paul George, Monica Sudarsanam, Kornelia Hardes, Torsten Steinmetzer, Lori A. Holle, Alisa S. Wolberg, Bernd Pötzsch, Johannes Oldenburg, Arijit Biswas, and Diana Imhof

This peer-reviewed research article was published in the *Journal of Medicinal Chemistry*.

Citation

J. Med. Chem., 62(7), 3513−3523

DOI: 10.1021/acs.jmedchem.8b01982

4.3.1 Introduction

Tridegin is a 66mer peptide with six cysteines forming three disulfide bonds derived from the giant amazon leech *Haementeria ghilianii*. Its functional relevance lies in its ability to potently block the human coagulation factor FXIIIa. Previous studies[203] identified the formation of three of the fifteen possible disulfide isomers with all three isomers sharing a common disulfide bond between the cysteines in positions 19 and 25 in the sequence. In the current study, three two-disulfide variants of the isomers lacking the C19 and C25 disulfide bonds were synthesized and their activity on FXIIIa was tested. In parallel, computational models of the three three-disulfide bonded isomers generated from previous studies were used for *in silico* investigations. Accordingly, MD simulations were conducted on the three-disulfide bonded isomers, the two-disulfide bonded variants lacking the C19-C25 disulfide bond and the two-disulfide bonded variants created in this study to compare the conformational dynamics to obtain insights that aid to improved understanding of the experimental investigation. Furthermore, multiple MD-generated structures of all the isomer variants were docked on the crystal structure of FXIIIa to deduce a possible mechanism and mode these isomers might use to inhibit the activity of the protein. The prime motive of this study is to find out whether an easier synthetic route in the form of one less disulfide bond is available for the synthesis of the isomers of this biologically relevant peptide using both experimental and computational modes of investigation.

***Own contribution**

The computational study including MD simulations and analysis of the docked complexes was performed by me. Monica Sudarsanam provided technical assistance to the MD simulations. Arijit Biswas performed the docking simulations. The design and the content of the manuscript has been compiled by Diana Imhof, Arijit Biswas, and Alisa Wolberg. The layout and preparation of the figures related to the computational study has been planned and carried out by me in consultation and agreement with Diana Imhof and Arijit Biswas. The manuscript was written through the contribution of all authors involved.

Journal of Medicinal Chemistry

Article

Cite This: J. Med. Chem. 2019, 62, 3513−3523

pubs.acs.org/jmc

Coagulation Factor XIIIa Inhibitor Tridegin: On the Role of Disulfide Bonds for Folding, Stability, and Function

Charlotte A. Bäuml,[†] Thomas Schmitz,[†] Ajay Abisheck Paul George,[†] Monica Sudarsanam,[†] Kornelia Hardes,[‡] Torsten Steinmetzer,[‡] Lori A. Holle,[§] Alisa S. Wolberg,[§] Bernd Pötzsch,[∥] Johannes Oldenburg,[∥] Arijit Biswas,[∥] and Diana Imhof*[,†]

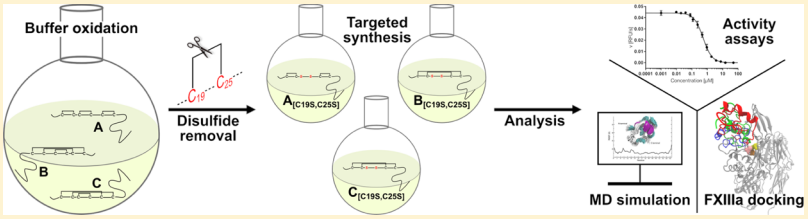
[†]Pharmaceutical Biochemistry and Bioanalytics, Pharmaceutical Institute, University of Bonn, An der Immenburg 4, 53121 Bonn, Germany

[‡]Institute of Pharmaceutical Chemistry, Philipps University of Marburg, Marbacher Weg 6, 35032 Marburg, Germany

[§]Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, 819 Brinkhous-Bullitt Building, Chapel Hill, North Carolina 27599, United States

[∥]Institute of Experimental Hematology and Transfusion Medicine, University Hospital Bonn, Sigmund-Freud-Str. 25, 53127 Bonn, Germany

Supporting Information



ABSTRACT: Tridegin is a potent and specific 66mer peptide inhibitor of coagulation factor XIIIa with six cysteines involved in three disulfide bonds. Three of the 15 possible 3-disulfide-bonded isomers have been identified, which share a bridge between cysteines 19 and 25. We synthesized the three possible 2-disulfide-bonded analogues using a targeted protecting group strategy to investigate the impact of the $C_{19}-C_{25}$ bond on tridegin's folding, stability, and function. The FXIIIa inhibitory activity of the analogues was retained, which was shown by in vitro fluorogenic activity and whole blood clotting assays. Molecular dynamics simulations of wild-type tridegin and the analogues as well as molecular docking studies with FXIIIa were performed to elucidate the impact of the $C_{19}-C_{25}$ bond on conformational stability and binding mode. The strategy of selectively reducing disulfide bonds to facilitate large-scale synthesis, while retaining the functionality of disulfide-bonded peptides, has been demonstrated with our present study.

INTRODUCTION

Cysteine-rich peptides and proteins are found in different organisms and possess a wide variety of biological functions.[1] Representatives, such as hirudin,[2] conotoxins,[3,4] and snake[5,6] and scorpion[7] toxins produced by different vertebrates and invertebrates, harbor high therapeutic potential. Recent research focused on the blood coagulation factor XIIIa (FXIIIa) inhibitor tridegin, a 66mer 3-disulfide-bonded peptide produced by the giant amazon leech *Haementeria ghilianii*.[8] The peptide represents a therapeutically promising yet synthetically challenging compound. It was first isolated and described in 1997.[8] Nonetheless, its folding pathway remains to be explored.[9,10] Tridegin inhibits FXIIIa with significant potency and high specificity.[8] Several efforts have been undertaken to produce this inhibitor by either synthetic or recombinant strategies.[9,10] These earlier reports also propose a major role for the three disulfide bonds located in the N-terminal region in the inhibitory activity of the peptide. In addition, tridegin seems to be composed of two functionally deviating segments: the flexible C-terminal fragment ($R_{38}-E_{66}$) was shown to be responsible for the inhibitory activity, whereas the disulfide-linked N-terminal segment (K_1-C_{37}) confers conformational stability and contributes to the binding to FXIIIa.[9,10] As a result of an oxidative self-folding approach, only three out of 15 possible disulfide-bridged isomers were identified and their disulfide connectivities were confirmed.[10] Accordingly, tridegin neither folds into one major isomer like BPTI (bovine pancreatic trypsin inhibitor),[11,12] or the conotoxins μ-SIIIA[13] and κ-RIIIK,[3] nor tends to form several of the 15 possible isomers, such as hirudin[11] and the conotoxins μ-PIIIA[4,14] and μ-SmIIIA.[3] All of the three identified tridegin isomers, however, share the disulfide bond $C_{19}-C_{25}$, suggesting its conservation as well as its functional importance for the peptide. During synthesis, it was also observed that partially oxidized tridegin species formed a

Received: December 18, 2018
Published: March 9, 2019

DOI: 10.1021/acs.jmedchem.8b01982
J. Med. Chem. 2019, 62, 3513−3523

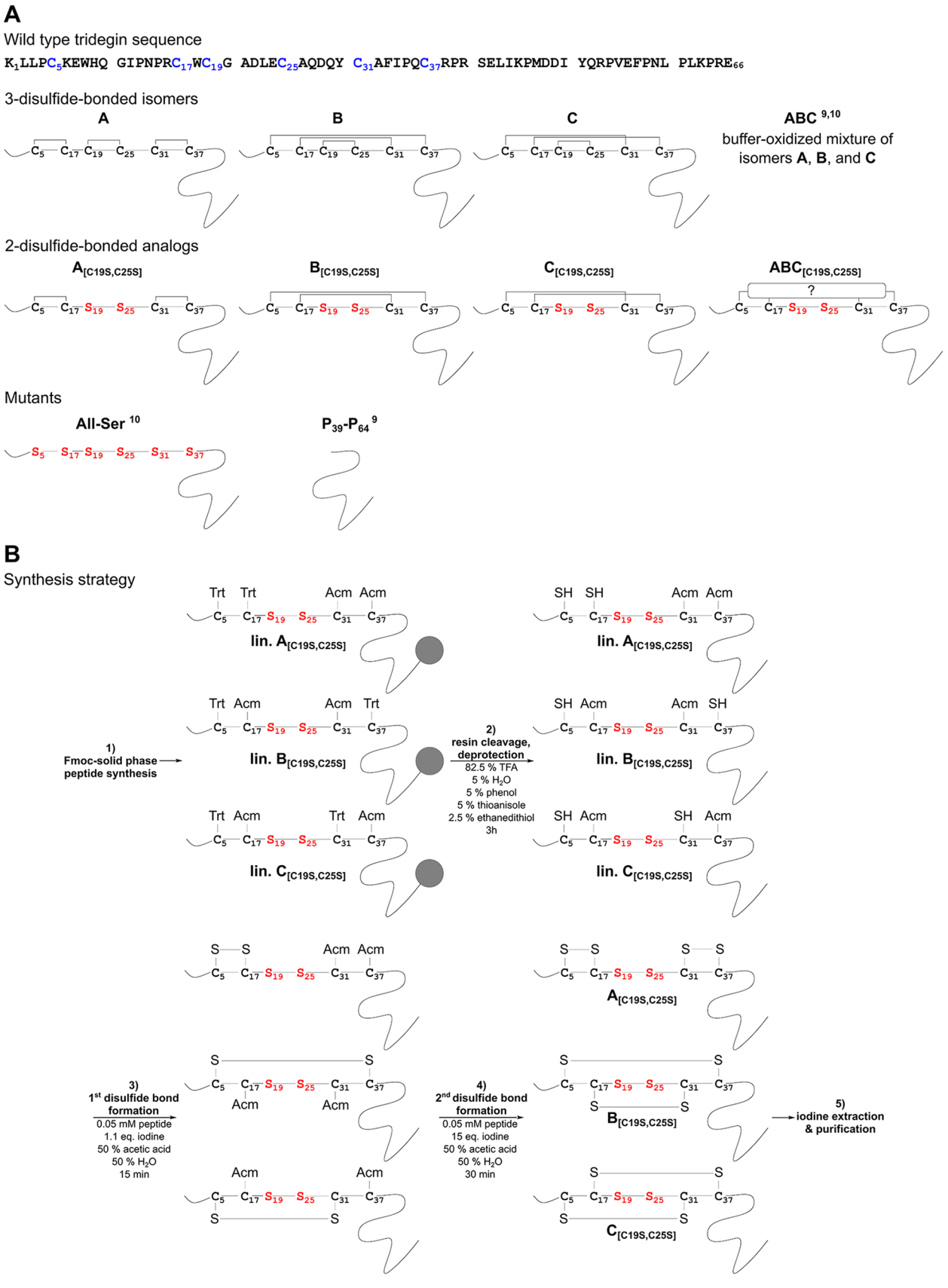


Figure 1. (A) Tridegin peptide sequence (cysteines are marked in blue) and disulfide connectivities of analyzed disulfide isomers and analogues. Cys−Ser exchanges are marked in red. Peptides **ABC**,[9,10] **All-Ser**,[10] and $\mathbf{P_{39}}$−$\mathbf{P_{64}}$[9] were described in previous studies. All peptides were produced as C-terminal amides. (B) Synthesis strategy of sequential SPPS and oxidative folding in solution for the tridegin mutants $\mathbf{A_{[C19S,C25S]}}$−$\mathbf{C_{[C19S,C25S]}}$. lin., linear precursor; TFA, trifluoroacetic acid; and eq., equivalents.

DOI: 10.1021/acs.jmedchem.8b01982
J. Med. Chem. 2019, 62, 3513−3523

bridge involving C_{25} and C_{37}, which seemed to trap the peptide in an irreversibly misfolded state.[10] One could therefore suppose that the disulfide bond C_{19}−C_{25} is important for correct oxidative folding in buffer. Structure elucidation by multidimensional NMR spectroscopy or X-ray crystallography was not successful so far due to the occurrence of aggregates and different isomers in the mixture.[10] Consequently, the three isomer structures were modeled and subjected to docking studies with FXIIIa,[15] revealing different docking poses for the three isomers.

Targeted synthesis of cysteine-rich peptides with three disulfide bonds is a challenging process.[14] Hence, with respect to a possible future medical application, a simplification of cysteine-rich peptides is highly desirable.[16−19] Because both truncation of tridegin and removal of all disulfide bonds resulted in a significant loss of activity,[9,10] our present approach was to reduce the complexity of the peptide structure by reducing the number of disulfide bonds. Several studies have elucidated the role of individual disulfide bonds in different cysteine-rich peptides, with different outcomes regarding the respective biological activity of the targeted peptide. For example, in the scorpion toxin leiurotoxin I,[16] the μ-conotoxin KIIIA,[17] the leech carboxypeptidase inhibitor,[18] and the α-conotoxin Vc1.1,[19] one disulfide bond was removed without a significant loss of inhibitory potential. In contrast, Han et al. showed that in the case of μ-GIIIA, all three disulfide bridges were essential for bioactivity.[20]

The aim of the present study was thus to explore the influence of the common disulfide bridge C_{19}−C_{25} on the structure and function of FXIIIa inhibitor tridegin. In a combined approach of experimental and computational studies, the three 2-disulfide-bonded tridegin isomers (bead, ribbon, and globular fold)[21] were synthesized and evaluated for their FXIIIa inhibitory activity and their structural stability based on earlier established models of the 3-disulfide-bonded tridegin.[10] Molecular dynamics (MD) simulations as well as docking studies with FXIIIa were used additionally to substantiate and also correlate with the in vitro data. Our study defines the role of the C_{19}−C_{25} disulfide bond for the structure−activity relationships of tridegin, the results of which may also be applicable to other disulfide-rich peptides and proteins.

RESULTS AND DISCUSSION

Design and Synthesis of Peptides. According to the 3-disulfide-bonded tridegin isomers (**A**−**C**) identified in our previous report, three novel disulfide-deficient analogues $\mathbf{A}_{[C19S,C25S]}$, $\mathbf{B}_{[C19S,C25S]}$, and $\mathbf{C}_{[C19S,C25S]}$, respectively, were synthesized by a protecting group strategy (Figure 1, Table S1).[10] Herein, the disulfide bond between C_{19}−C_{25} was omitted by replacing cysteines with serines. In addition, another C_{19}−C_{25}-deficient analogue of tridegin (peptide $\mathbf{ABC}_{[C19S,C25S]}$), that is, an unknown mixture of 2-disulfide-bonded isomers, was synthesized in an oxidative self-folding approach performed in buffer as described earlier for wild-type tridegin.[9,10] Furthermore, the 3-disulfide-bonded tridegin (**ABC**)[10] and a linear tridegin analogue (**All-Ser**)[10] with serine replacements for all six cysteines, which have been described in earlier studies,[10] were used as controls in the fluorogenic enzyme activity assay. Apart from the **All-Ser** mutant, other peptides still showed acceptable, yet reduced, inhibitory activity in our former study, for example, four of the peptides derived from the C-terminal segment such as $\mathbf{P}_{39}$−$\mathbf{P}_{64}$.[9] These peptides can be synthesized with a higher yield in comparison to the **All-Ser** mutant because of their shorter length. As the whole-blood assays required considerably more substance, herein $\mathbf{P}_{39}$−$\mathbf{P}_{64}$ was used as a control instead of the **All-Ser** mutant and the parent peptide **ABC**. The sequences and the respective disulfide bond patterns of the peptides are summarized in Figure 1A.

Synthesis of disulfide-containing peptides can be achieved by a variety of approaches. Depending on the peptide sequence and the number of cysteines, diverse methods, such as buffer oxidation or regioselective synthesis by different protecting-group strategies, can be applied.[22] The targeted synthesis of the different 2-disulfide-bonded analogues was performed using a stepwise protecting group strategy with acetamidomethyl (Acm)- and trityl (Trt)-protecting groups (Figure 1B, Table S1). Partially Cys-protected linear precursors (Figure S1, **1**) were synthesized and subjected to stepwise oxidation in the following order according to their side chain protection: (1) oxidation of formerly Trt-protected Cys-pair, (2) deprotection and oxidation of the Acm-protected Cys-pair, and (3) RP HPLC purification. In the process, increasing amounts of iodine were used for selective formation of the first and second disulfide bond (Figure 1B).[23] The oxidation reaction was stopped by the addition of an excess of ascorbic acid[14] or by iodine extraction and monitored via HPLC and MS. The reaction generating the first disulfide bond was carried out more efficiently when iodine extraction was applied (Figure S1). Termination of the reaction by adding ascorbic acid led to byproducts (**4**) besides the desired peptides (**3**) (Figure S1). The occurrence of byproducts could, however, be diminished upon extraction of iodine from the aqueous reaction mixture in an organic solvent instead of adding ascorbic acid. The observed byproducts were supposedly a result of side reactions, such as iodination or oxidation of sensitive residues, which are often associated with iodine-catalyzed oxidation.[24−26] Particularly during the lyophilization step after oxidation, a concentration of residual iodine can occur, which may lead to the formation of undesired byproducts.[24] The efficient removal of excessive iodine by extraction was therefore crucial for the outcome of the synthesis. Thus, an increased yield of the extracted variants was achieved after purification. Analytical HPLC and MS data of the synthesized peptides are shown in Figure S2. The oxidation yield of each peptide ranged from 33% for $\mathbf{B}_{[C19S,C25S]}$, which was the most efficient synthesis, to 14% for $\mathbf{A}_{[C19S,C25S]}$ and 19% for $\mathbf{C}_{[C19S,C25S]}$ (Table S2). These results are comparable to what has been described earlier for other peptides[24,25] yet are remarkably higher compared to the synthesis of the 3-disulfide-bonded tridegin (<10%).

Analysis of Disulfide Connectivities of Tridegin Isomers. In an earlier study, we analyzed the elution behavior of different disulfide isomers of μ-conotoxin PIIIA[14] and realized that most of the isomers elute with the same retention time independent of the chromatography setup used. In order to find out about the elution profile of mixtures of the tridegin isomers of the present study, equimolar amounts of analogues $\mathbf{A}_{[C19S,C25S]}$−$\mathbf{C}_{[C19S,C25S]}$ in water were mixed and injected on a C18 column for RP HPLC analysis. The coelution profile (Figure S3) resulted in a single peak at a retention time of 19.3 min. Accordingly, the isomers in the mixture possess similar physicochemical properties, rendering it difficult to separate them by conventional RP HPLC methods. This confirms the hypothesis from our earlier study that these disulfide isomers cannot be separated by HPLC,[10] and other methods (e.g. MS/

DOI: 10.1021/acs.jmedchem.8b01982
J. Med. Chem. 2019, 62, 3513−3523

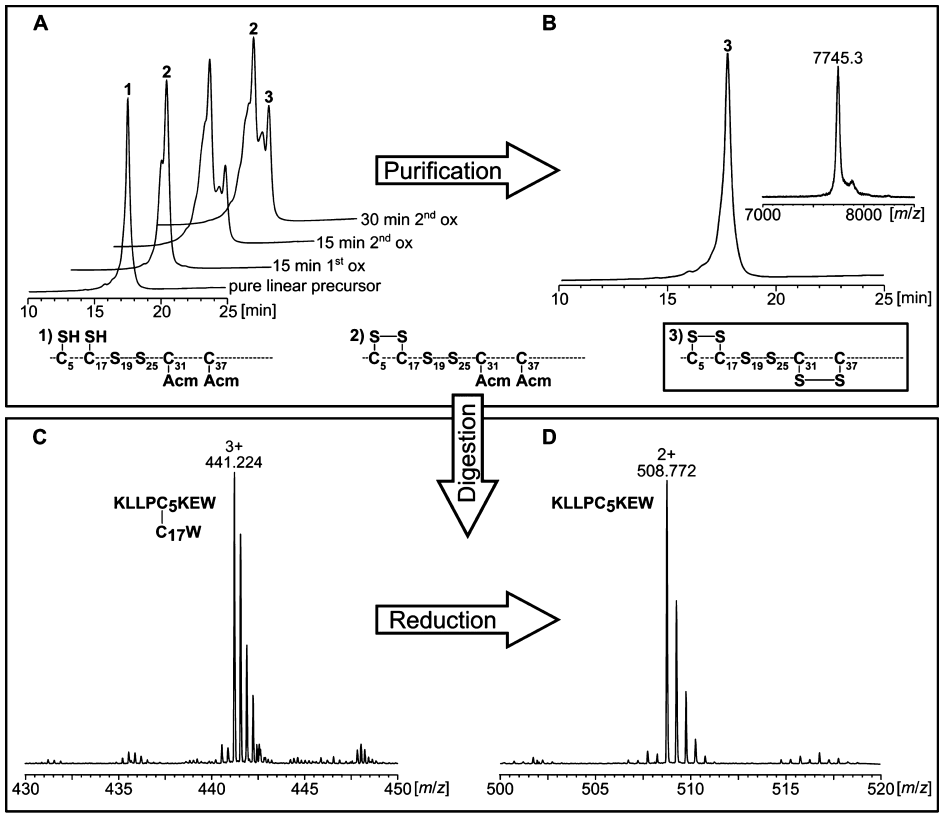


Figure 2. Analysis of $\mathbf{A}_{[C19S,C25S]}$ during synthesis (A,B) and chymotryptic digest for mass spectrometry analysis (C,D). (A) Reaction control during first and second oxidation analyzed by HPLC. (B) MS and HPLC analysis of purified $\mathbf{A}_{[C19S,C25S]}$. (C) The chymotryptic digest of $\mathbf{A}_{[C19S,C25S]}$ was analyzed by HPLC, and the collected fractions were subjected to MS analysis. Peptide fragments containing disulfide bonds were identified as exemplified here for the C_5–C_{17} bond (C,D). (D) The fractions containing disulfide bonds were reduced with DTT, and the identity of the reduced fragments was confirmed by subsequent tandem mass spectrometry (example: reduced fragment containing C_5).

MS, see below) have to be employed in order to confirm the identity of the peptides.

The verification of the correct disulfide-bridge formation of the three different tridegin analogues ($\mathbf{A}_{[C19S,C25S]}$–$\mathbf{C}_{[C19S,C25S]}$) was performed by tandem mass spectrometry analysis of a chymotryptic digest according to an earlier reported protocol.[10] Chymotryptic proteolysis of the linear precursors as well as the fully oxidized analogues and separation of the digested fragments via HPLC were carried out. The fragments were analyzed by LC–ESI-MS and –MS/MS before and after reduction by DTT (Figure 2). In case of the linear peptides, the correct positions of the acetamidomethyl-protecting groups were confirmed by MS/MS analysis (Table S3, Figures S4–S6). The evaluation of disulfide-linked fragments confirmed the disulfide connectivity of the different oxidized analogues as well. This is exemplarily shown for analogue $\mathbf{A}_{[C19S,C25S]}$ (Figure 2). Several linked fragments belonging to the disulfide bonds C_5–C_{17} and C_{31}–C_{37} were identified in the chymotryptic digest, confirming the disulfide connectivity (Table S4).

The oxidative self-folding approach peptide $\mathbf{ABC}_{[C19S,C25S]}$ was also subjected to chymotryptic digest and subsequent MS and MS/MS analysis. All three possible isomers ($\mathbf{A}_{[C19S,C25S]}$–$\mathbf{C}_{[C19S,C25S]}$) were identified by specific linked fragments as depicted in Table S5.

Fluorogenic Enzyme Activity Assay. In former studies using a chromogenic and a fluorogenic assay, we analyzed the role of the disulfide-stabilized N-terminal and the flexible C-terminal part as well as the general impact of disulfide bonds on the inhibitory potential of tridegin.[9,10] A comparison of fully oxidized tridegin (mixture of isomers **A**–**C**, **ABC**) with the reduced variant as well as with the Cys-mutated version **All-Ser** revealed a 3- to 4-fold decrease of inhibitory activity, which was explained by the structural changes induced by oxidative folding of the peptide.[9,10] Here, we investigated the special effect of the disulfide bond C_{19}–C_{25} on the inhibitory potential of tridegin by testing the three disulfide-deficient isomers $\mathbf{A}_{[C19S,C25S]}$, $\mathbf{B}_{[C19S,C25S]}$, and $\mathbf{C}_{[C19S,C25S]}$ (Figure 1). The earlier reported 3-disulfide-bonded wild-type **ABC** as well as the **All-Ser** mutant were used as controls.[10] All tested inhibitors showed linear progress curves pointing to stable inhibition of FXIIIa (Figure S7). The tridegin analogues $\mathbf{A}_{[C19S,C25S]}$, $\mathbf{B}_{[C19S,C25S]}$, $\mathbf{C}_{[C19S,C25S]}$, and $\mathbf{ABC}_{[C19S,C25S]}$ displayed IC_{50} values of 0.55 ± 0.05 μM (n = 4), 0.50 ± 0.05 μM (n = 3), 0.48 ± 0.06 μM (n = 4), and 0.51 ± 0.02 μM (n = 2), respectively (Figure 3A–E). The control peptides **ABC** and **All-Ser** inhibited FXIIIa with IC_{50} values of 0.45 ± 0.03 μM (n = 3) and 0.72 ± 0.05 μM (n = 4) (Figure 3E). The difference between the control peptide **ABC** and the tridegin analogue $\mathbf{A}_{[C19S,C25S]}$ was minor, whereas $\mathbf{B}_{[C19S,C25S]}$, $\mathbf{C}_{[C19S,C25S]}$, and $\mathbf{ABC}_{[C19S,C25S]}$ did not differ significantly from **ABC** (Figure 3A–E, Tables S6 and S7). Our experimental data thus suggest almost full retention of the inhibitory activity upon removal of the disulfide bond C_{19}–C_{25}.

Whole Blood Clotting Assay. We also tested the inhibitory effect of $\mathbf{A}_{[C19S,C25S]}$, $\mathbf{B}_{[C19S,C25S]}$, $\mathbf{C}_{[C19S,C25S]}$, and the previously described variant $\mathbf{P}_{39}$–$\mathbf{P}_{64}$[9] on the FXIIIa activity in a whole blood milieu. This assay is based on the

DOI: 10.1021/acs.jmedchem.8b01982
J. Med. Chem. 2019, 62, 3513–3523

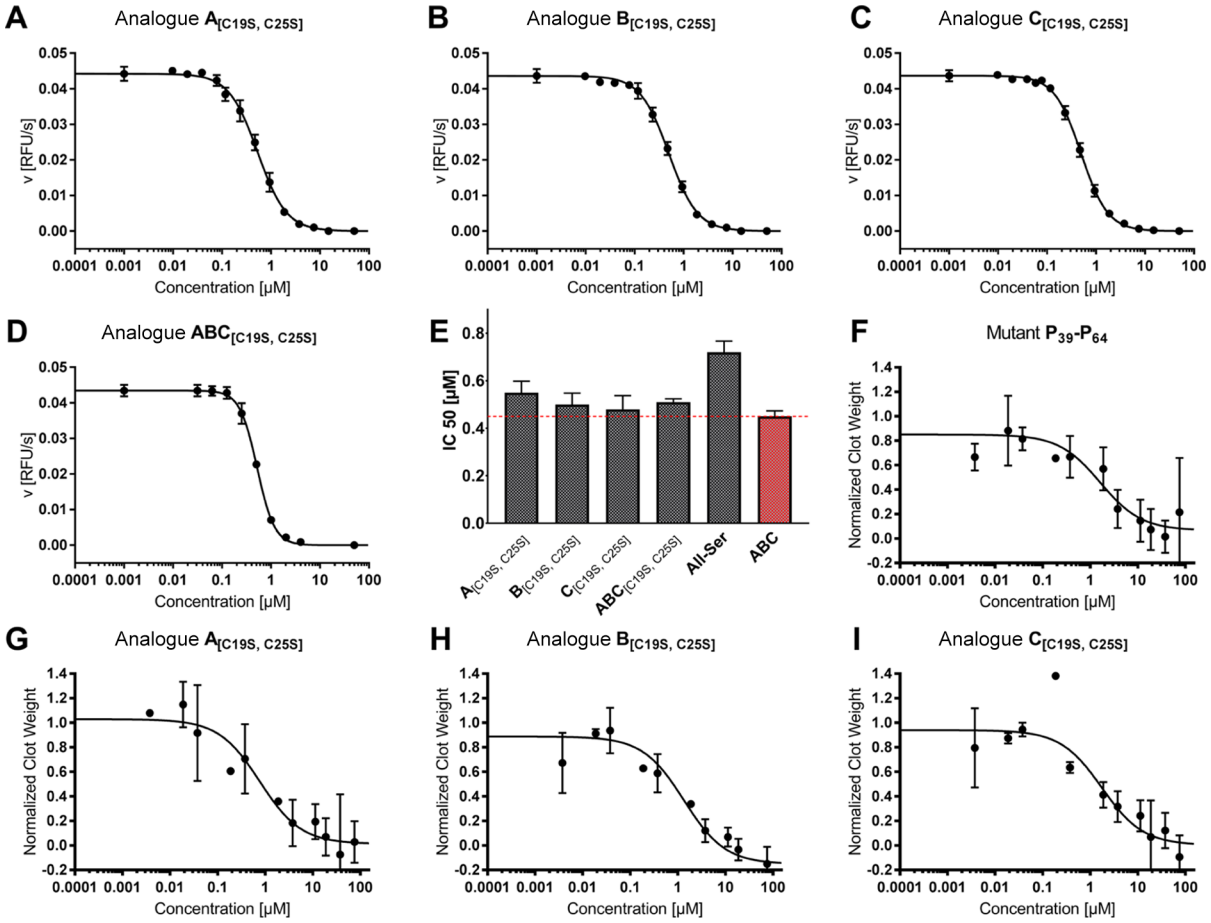


Figure 3. Activity assays of different tridegin variants. (A–E) In vitro FXIII assay in a purified system: exemplary IC_{50} plots of tridegin analogues $\mathbf{A}_{[C19S,C25S]}$, $\mathbf{B}_{[C19S,C25S]}$, $\mathbf{C}_{[C19S,C25S]}$, $\mathbf{ABC}_{[C19S,C25S]}$, and corresponding bar chart of averaged IC_{50} values in comparison to controls **All-Ser** and **ABC**. The dashed red line marks the IC_{50} value of wild-type **ABC**. (F–I) Whole blood clot contraction assay of $\mathbf{A}_{[C19S,C25S]}$, $\mathbf{B}_{[C19S,C25S]}$, and $\mathbf{C}_{[C19S,C25S]}$ in comparison to control $\mathbf{P}_{39}$–$\mathbf{P}_{64}$.

discovery that FXIIIa promotes red blood cell retention in contracted whole blood clots[27,28] and therefore is a major determinant of whole blood clot size. $\mathbf{A}_{[C19S,C25S]}$, $\mathbf{B}_{[C19S,C25S]}$, $\mathbf{C}_{[C19S,C25S]}$, and $\mathbf{P}_{39}$–$\mathbf{P}_{64}$ each reduced red blood cell retention in clots and clot weight with IC_{50} values of 0.7 ± 0.4, 1.1 ± 0.1, 2.3 ± 2.2, and 2.2 ± 2.0 μM, respectively (n = 3, Figure 3F–I, Table S7). Consistent with prior findings,[8] the tridegin-derived peptides did not alter the kinetics of fibrin formation or final clot turbidity in plasma (Figure S8). Effects were similar to that seen with the FXIIIa active site covalent inhibitor T101 (IC_{50} 0.5 ± 0.1 μM, data not shown), a small-molecule imidazolium derivative,[28] and within 1.5–4.5-fold of values obtained in the enzyme kinetic studies with FXIIIa. Given the lack of efficient and specific compounds for inhibiting FXIIIa, the novel peptides $\mathbf{A}_{[C19S,C25S]}$, $\mathbf{B}_{[C19S,C25S]}$, and $\mathbf{C}_{[C19S,C25S]}$ represent promising leads for this function.

Conformational Analysis Using MD Simulations. Interestingly, $\mathbf{A}_{[C19S,C25S]}$, $\mathbf{B}_{[C19S,C25S]}$, and $\mathbf{C}_{[C19S,C25S]}$ did not show significantly different inhibitory activities toward FXIIIa in both assays (Figure 3, Tables S6, S7). This is surprising at first glance, due to the theoretical impact of disulfide bonds on peptide conformation and thus structure–function properties. In particular, tridegin variant $\mathbf{A}_{[C19S,C25S]}$ harbors the shortest distances between linked cysteines and therefore was expected to represent a considerably more flexible structure than variants $\mathbf{B}_{[C19S,C25S]}$ and $\mathbf{C}_{[C19S,C25S]}$ and thus to differ in its inhibitory potential. The contribution of the different disulfide bridges on the peptide's inhibitory activity, however, seems to be of a more complex nature, which was analyzed by MD and docking studies.

MD-based conformational analysis was conducted on previously reported threaded models for tridegin isomers **A**–**C**[10] owing to the non-availability of NMR or X-ray crystallographic structures for these peptides. Two independent rounds of MD simulations were conducted for each isomer and its variants. In order to improve the level of equilibration in simulation, the final frame from the first 100 ns simulation was used as the starting structure for a second simulation for 300 ns. All results are discussed from this longer 300 ns simulation. The structural variabilities of the simulated isomers, variants with in silico removal of the C_{19}–C_{25} bond, and their corresponding serine-mutated analogues were compared (Figure 4). Moreover, root-mean square deviation (RMSD) as well as per-residue root mean square fluctuation (RMSF) profiles during the initial 100 ns (Figure S9) and an extended 300 ns MD simulations were calculated (Figures S10–S14).

Isomer **A** showed a moderate level of structural deviation from its starting conformation with a time-averaged RMSD value of 3.33 Å computed on the backbone of the whole molecule (Figures 4B, S10, Table S8). The replacement of cysteines C_{19} and C_{25} by serines brought this deviation down

DOI: 10.1021/acs.jmedchem.8b01982
J. Med. Chem. 2019, 62, 3513–3523

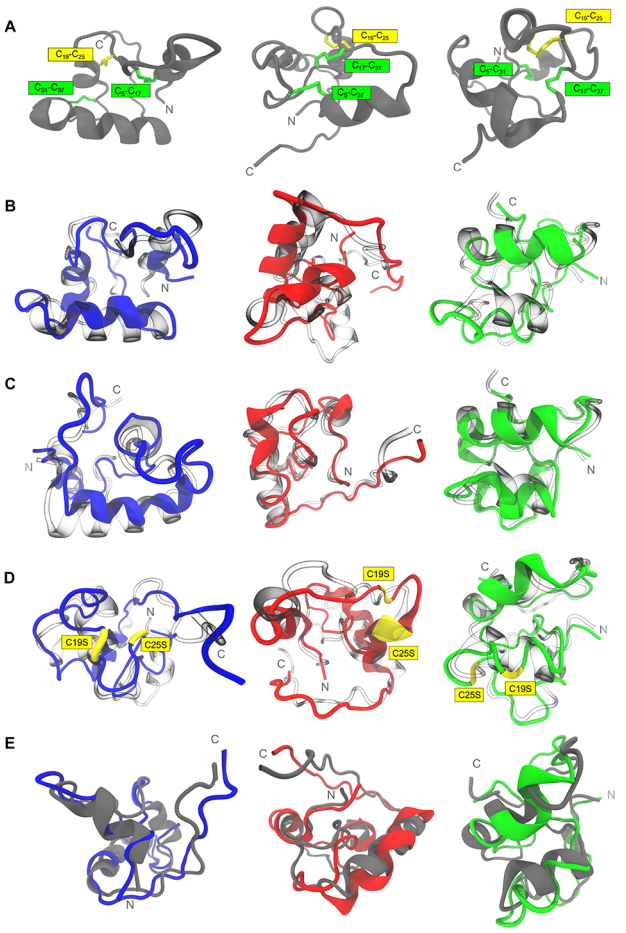


Figure 4. Comparison of structural variability of tridegin isomers in simulation. (A) Structures of the three isomers **A**, **B**, and **C** shown in a (grey) cartoon representation. The disulfide bonds are marked and highlighted. (B) Simulated structures of **A** (blue), **B** (red), and **C** (green) superimposed on their respective starting structures (transparent/blown glass representation). (C) Simulated structures of **A** (blue), **B** (red), and **C** (green) with in silico removal of the C_{19}−C_{25} disulfide bond superimposed on their respective starting structures (transparent/blown glass representation). (D) Simulated structures of $\mathbf{A}_{[C19S,C25S]}$ (blue), $\mathbf{B}_{[C19S,C25S]}$ (red), and $\mathbf{C}_{[C19S,C25S]}$ (green) superimposed on their respective starting structures (transparent/blown glass representation). Serine mutations of cysteines (C19S and C25S) are depicted in yellow. (E) Simulated structures of $\mathbf{A}_{[C19S,C25S]}$ (blue), $\mathbf{B}_{[C19S,C25S]}$ (red), and $\mathbf{C}_{[C19S,C25S]}$ (green) superimposed on their respective parent 3-disulfide bonded isomers (grey cartoons from row (A)). All simulated structures presented here refer to the energy-minimized structures of the final snapshot in the 300 ns MD simulation.

to 2.59 Å as observed for the novel analogue $\mathbf{A}_{[C19S,C25S]}$ (Figures 4D, S10, Table S8). The average structural deviation between isomer **A** and its novel analogue $\mathbf{A}_{[C19S,C25S]}$ differed by 0.74 Å global RMSD only.

The structure of isomer **B** displayed minor deviation from its initial conformation in simulation, that is, average backbone RMSD of 2.70 Å (Figures 4B, S10, Table S8). The highest structural deviation between all three isomers was for the serine-mutated novel analogue $\mathbf{B}_{[C19S,C25S]}$ (Figures 4D, S10, Table S8). The overall backbone RMSD was 10.74 Å of which 8.25 Å was attributed to deviations arising from the unstructured C-terminal region (residues 38−66) (Table S8). The disulfide-linked part of the peptide (residues 1−37) on the other hand, remained comparatively stable with a meagre 1.54 Å backbone RMSD (Table S8) in comparison to 1.77 Å in the parent isomer **B**. Hence, in the case of the analogue $\mathbf{B}_{[C19S,C25S]}$, though the overall molecule deviates significantly from its 3-disulfide-bonded version, the region constrained by the two disulfide bonds remained almost unaffected, that is, only 0.23 Å difference (Table S8), on the constrained segment of residues 1−37.

The isomer **C** displayed the greatest amount of structural rigidity in simulation (Figure 4B). With a backbone RMSD of 1.71 Å, the entire molecule, including the C-terminal region, remained structurally rigid (Table S8, Figure S11G). This could be attributed to the constrained conformation rendered by the cross-linked disulfide bonds C_5−C_{37} and C_{17}−C_{31} (Figures 1 and 4A). Similar to the behavior observed in isomer **B**, the isomer $\mathbf{C}_{[C19S,C25S]}$ deviated significantly in backbone RMSD, that is, 7.03 Å compared to its 3-disulfide-bonded parent isomer (Table S8). Here again 6.52 Å of the overall structural deviation arose from the conformations sampled by the unstructured C-terminal residues (38−66) (Table S8). The overall structural change between the isomer **C** and its serine-mutated analogue $\mathbf{C}_{[C19S,C25S]}$ within the disulfide-stabilized N-terminal region (residues 1−37) was still as low as 1.09 Å.

To gain deeper insights into the regions responsible for the conformational flexibility in isomers **A**−**C**, RMSF was related to conformation snapshots throughout the simulation (Figures S12 and S14). From the RMSD-based conformational analysis of the three parent isomers and their disulfide-deficient analogues from the present study, two clear inferences can be made. First, the removal of the C_{19}−C_{25} disulfide bonds in all three isomers does not significantly impact the structural integrity of the region bracketed by the two remaining disulfide bonds. Second, at the same time, a marked impact on the structural conformation of the C-terminal region between residues 38−66 is observed. The distances sampled in simulation between the Sγ atoms of the unbound cysteine residues by in silico disulfide bond removal did not increase significantly (Figure S13). This further confirms that these partially unfolded states do not experience further structural disruption or unfolding. Videos of these simulations as well as additional discussions of the results of the MD studies are available in the Supporting Information.

Overall, it can be stated that neither the in silico opening of the disulfide bond C_{19}−C_{25} nor the serine mutations of corresponding cysteine residues seem to promote unfolding or structural disruption of the peptides, which goes in line with the experimental data generated.

Docking of Disulfide-Deficient Isomers to FXIIIa. Blind docking simulations were conducted wherein the resultant structures from the MD simulations of the fully bridged isomers of wild-type tridegin (**A**, **B**, and **C**),[10] and the corresponding 2-disulfide-bonded isomers were docked to FXIIIa (PDB 4KTY).[15] Blind docking protocols are usually used as a means to scan for possible binding sites and binding modes of peptide ligands over the surface of target proteins.[29] Following an "ensemble docking" protocol (see Experimental Section), the best docked poses of tridegin−FXIIIa complexes based on the predicted binding energies were chosen and analyzed. It is known from previous studies that the active site of FXIIIa is defined by the catalytic triad C_{314}, H_{373}, and D_{396}.[15] The most interesting observation from the docked conformations of the different analogues with in silico removal of

DOI: 10.1021/acs.jmedchem.8b01982
J. Med. Chem. 2019, 62, 3513−3523

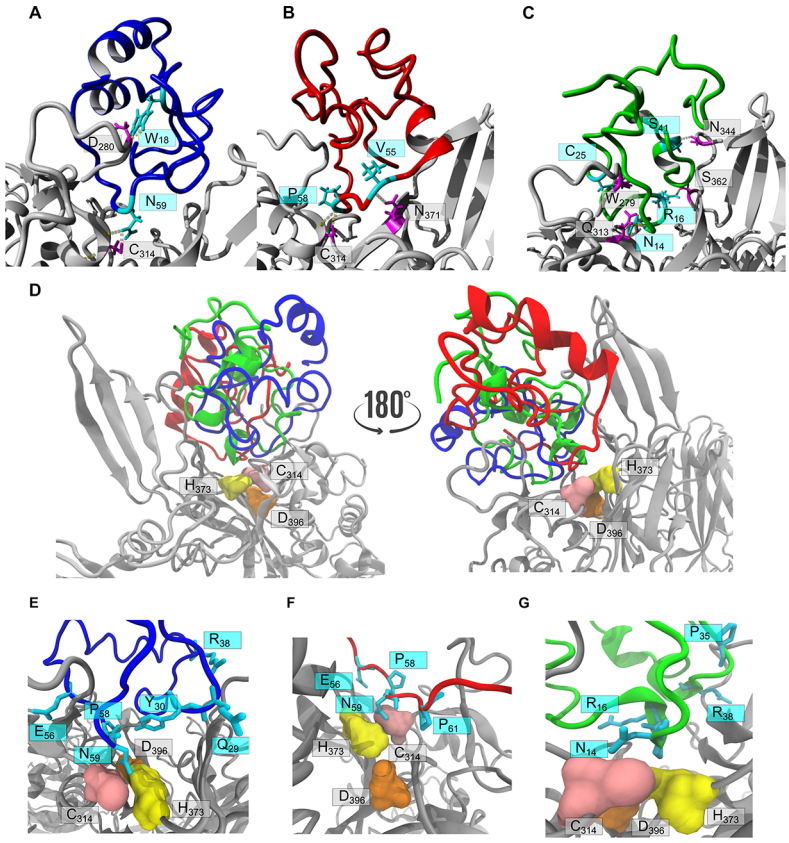


Figure 5. Docking of isomers on blood coagulation factor XIIIa. FXIIIa (PDB 4KTY) is represented as a grey ribbon structure in all panels. Hydrogen bonds are depicted as dashed lines. Docking of the three tridegin isomers with the in silico opened C_{19}–C_{25} bond on FXIIIa: (A) isomer **A** (blue), (B) isomer **B** (red), and (C) isomer **C** (green). Interacting residues of the isomers (cyan) and FXIIIa (magenta) are labeled. (D) Two 180° rotated orientations showing $\mathbf{A}_{[C19S,C25S]}$ (blue ribbon), $\mathbf{B}_{[C19S,C25S]}$ (red ribbon), and $\mathbf{C}_{[C19S,C25S]}$ (green ribbon) docked in close proximity to active site C_{314} (pink surface), H_{373} (yellow surface), and D_{396} (orange surface). Zoom-in on analogue $\mathbf{A}_{[C19S,C25S]}$ (E, blue), $\mathbf{B}_{[C19S,C25S]}$ (F, red), and $\mathbf{C}_{[C19S,C25S]}$ (G, green) docked on FXIIIa with the interacting residues labeled as in (D).

the C_{19}–C_{25} bond and their corresponding serine-mutated analogues ($\mathbf{A}_{[C19S,C25S]}$, $\mathbf{B}_{[C19S,C25S]}$, and $\mathbf{C}_{[C19S,C25S]}$) is that the best docked conformation among all docked variants bound to approximately the same region on FXIIIa as was observed earlier for wild-type isomers **A** and **B**.[10] For example, in all of the docked complexes, the C_{314} residue of FXIIIa was found in close proximities to the docked tridegin isomer. This result is in agreement with previous studies using the starting structures of the simulations with respect to isomers **A** and **B**.[10] Regarding isomer **C**, however, the present study reveals that, with the C_{19}–C_{25} disulfide bond opened, conformations adopted by the isomer were able to bind close to the active site of FXIIIa. This docked conformation of isomer **C** deviates from the previously described docking studies wherein the isomer **C** bound far away from the active site of FXIIIa.[10]

The docking simulations of the tridegin analogues $\mathbf{A}_{[C19S,C25S]}$, $\mathbf{B}_{[C19S,C25S]}$, and $\mathbf{C}_{[C19S,C25S]}$ showed that a vast majority of the conformations sampled by these analogues bound at close proximity to the active site (Figure 5). This key finding supports the experimental data suggesting comparable inhibitory activities of all three novel tridegin analogues ($\mathbf{A}_{[C19S,C25S]}$, $\mathbf{B}_{[C19S,C25S]}$, and $\mathbf{C}_{[C19S,C25S]}$).

Based on the docking experiments and the experimental data, one can hypothesize that the disulfide-deficient tridegin analogues inhibit FXIIIa by blocking the active site, rather than by direct interactions with individual active site residues. The fact that multiple conformations of the three tridegin analogues in this study dock at very similar poses, that is, close to the catalytic triad in a manner that blocks the active site, may serve as a strong indicator to this suggested mechanism of inhibition. The presented docking simulations provide visual evidence to the experimentally determined consensus that all three isomers synthesized and tested in this study lacking the C_{19}–C_{25} disulfide bond possess comparable inhibitory activities.

CONCLUSIONS

In the present study, we were able to elucidate the role of the disulfide bridge C_{19}–C_{25} of the previously identified 3-disulfide-bonded tridegin isomers.[10] Targeted synthesis of the three different 2-disulfide-bonded tridegin analogues (peptides $\mathbf{A}_{[C19S,C25S]}$, $\mathbf{B}_{[C19S,C25S]}$, $\mathbf{C}_{[C19S,C25S]}$) was accomplished via a protecting group strategy. Moreover, a mixture of these isomers ($\mathbf{ABC}_{[C19S,C25S]}$) was produced by an oxidative self-folding approach. The fluorogenic activity assay revealed a similar inhibitory potential of all analogues $\mathbf{A}_{[C19S,C25S]}$, $\mathbf{B}_{[C19S,C25S]}$, and $\mathbf{C}_{[C19S,C25S]}$ toward FXIIIa and a comparable activity in relation to the formerly described fully oxidized 3-disulfide-bonded isomer mixture (**ABC**). This points to a minor influence of the central disulfide bond on the overall inhibitory activity. Furthermore, the disulfide-deficient analogues were tested in a whole-blood milieu and significantly reduced red blood cell retention in clots and clot weight. These

DOI: 10.1021/acs.jmedchem.8b01982
J. Med. Chem. 2019, 62, 3513–3523

findings hint at the peptides' possible potential in anticoagulant therapy. Surprisingly, in both assay systems, no significant differences in inhibitory potential were observed among the three different 2-disulfide-bonded variants $\mathbf{A}_{[C19S,C25S]}$, $\mathbf{B}_{[C19S,C25S]}$, and $\mathbf{C}_{[C19S,C25S]}$. This is in line with previous findings[10] accrediting that the N-terminal disulfide-bridged part of tridegin plays a role in stabilization and binding, whereas the flexible C-terminal part is mostly responsible for the inhibitory action.

MD studies on models of the variants in comparison with unmodified tridegin support the experimental data, suggesting only minor impairment of the N-terminal fragment overall structure. The simulations conducted with the in silico opened $C_{19}-C_{25}$ disulfide bond enabled the examination of the peptides' folding behavior while the corresponding simulations with mutated cysteine residues C_{19} and C_{25} aided in illustrating the structural and dynamic properties of the peptides synthesized in this study. The disulfide connectivity of the disulfide-deficient tridegin seems to be of little importance regarding its globular structure and hence its activity. Molecular docking of selected MD simulation snapshots suggests only slightly different binding modes of the different disulfide-deficient variants toward FXIIIa, which matches the experimental data.

In summary, we thoroughly characterized three different 2-disulfide-bonded tridegin analogues without significant loss of inhibitory potential while facilitating large-scale synthesis. The presented tridegin variants may serve as valuable tools for future research on FXIIIa and as lead structures for the inhibitor development. Moreover, the reduction of the number of disulfide bonds might also be applicable to other cysteine-rich molecules, rendering them more suitable for drug development.

EXPERIMENTAL SECTION

Materials. Fmoc-amino acids, HBTU, and resins for solid-phase peptide synthesis (SPPS) were purchased from Orpegen Peptide Chemicals (Heidelberg, Germany) and IRIS Biotech (Marktredwitz, Germany), respectively. Other chemicals used for peptide synthesis including reagent-grade *N*-methylmorpholine, piperidine, trifluoroacetic acid (TFA), and *N*,*N*-dimethylformamide (DMF) were obtained from Sigma-Aldrich Chemie GmbH (Munich, Germany), Alfa Aesar (Karlsruhe, Germany), abcr GmbH (Karlsruhe, Germany), and VWR International GmbH (Darmstadt, Germany). Solvents (analytical grade: acetonitrile, water, and methanol and reagent grade: ethyl acetate and diethyl ether) and chemicals (iodine and acetic acid) used for purification and oxidation were purchased from VWR International GmbH and Fisher Scientific GmbH (Schwerte, Germany). Fibrogrammin P was kindly supplied by CSL Behring GmbH (Marburg, Germany). Lipidated tissue factor (Innovin) was from Siemens (Newark, DE, USA). Phospholipids were provided by Synapse Research Institute (Maastricht, Netherlands).

Peptide Synthesis and Purification. All peptides were synthesized according to a standard Fmoc solid-phase peptide synthesis protocol by automated peptide synthesis using a ResPep SL peptide synthesizer from Intavis Bioanalytical Instruments GmbH (Cologne, Germany) as described previously.[9] The linear peptides were assembled using Rink-Amide MBHA resin with a loading capacity of 0.53 mmol/g. The following side-chain-protected amino acids were employed: Arg(Pbf), Asn(Trt), Asp(O*t*Bu), Cys(Acm), Cys(Trt), Gln(Trt), Glu(O*t*Bu), His(Trt), Lys(Boc), Ser(*t*Bu), Trp(Boc), and Tyr(*t*Bu). Peptide cleavage from the resin and side-chain deprotection was accomplished as described earlier.[9] The purification of the crude peptides was done by semipreparative RP-HPLC using a JASCO PV-987 instrument (Gross-Umstadt, Germany) equipped with a Knauer Eurospher column (C18, 250 × 32 mm, 5 μm particle size, 100 Å pore size) and using a gradient elution system employing 0.1% TFA in water (eluent A) and 0.1% TFA in acetonitrile/water (90:10) (eluent B). The peptides were eluted with an increase of eluent B from 10 to 60% over 120 min at a flow rate of 10 mL/min. The detection was at λ = 220 nm. The collected fractions were combined, freeze-dried, and stored at −20 °C. The purity of the linear precursors was >95% as assessed by analytical RP-HPLC on a Shimadzu LC-20AD system equipped with a Vydac 218TP column (C18, 250 × 4.6 mm, 5 μm particle size, 300 Å pore size). Gradient elution was carried out with a gradient of 20−50% eluent B in 30 min with eluent A: 0.1% TFA in water and eluent B: 0.1% TFA in acetonitrile. The flow rate was 1 mL/min. The detection of the peptides was at λ = 220 nm. The yield of linear peptides after purification was in the range of 5−10%.

Oxidation of Linear Precursor Peptides. The selective oxidation of the 2-disulfide-bonded isomers proceeded in a one-pot reaction. The linear precursors (1 equiv) were dissolved in 50% acetic acid/water to yield a final concentration of 0.05 mM. 1.1 equiv iodine (0.1 M in methanol) was added to the peptide solution, and the reaction was stirred at room temperature and under argon atmosphere (first oxidation). After 15 min, further 13.9 equiv iodine (0.1 M in methanol) was added and the reaction was stirred for an additional 30 min (second oxidation). The oxidation reaction was stopped by addition of an excess of ascorbic acid (1 M in water) or by removal of iodine by extraction. During the extraction procedure, the aqueous reaction mixture was extracted 3 times with the same volume of ethyl acetate. The peptide-containing aqueous solutions were combined, freeze-dried, and stored at −20 °C.

Oxidation of peptide $\mathbf{ABC}_{[C19S,C25S]}$ was accomplished by self-folding of the linear precursor in a buffer solution as described earlier.[9,10] The product was freeze-dried and stored at −20 °C.

The peptide powder was dissolved in water and purified by semipreparative RP-HPLC using a Shimadzu LC-8A system (Duisburg, Germany) equipped with a Vydac 218TP1022 column (C18, 250 × 22 mm, 5 μm particle size, 100 Å pore size) using gradient elution employing 0.1% TFA in water as eluent A and 0.1% TFA in acetonitrile/water (90:10) as eluent B. The peptides were eluted with a gradient of 10−60% of eluent B over 120 min at a flow rate of 10 mL/min. The detection was at λ = 220 nm. The collected fractions were combined, freeze-dried, and stored at −20 °C. The purity of the peptides was >95% as assessed by analytical RP-HPLC on a Shimadzu LC-20AD system equipped with a Vydac 218TP column (C18, 250 × 4.6 mm, 5 μm particle size, 300 Å pore size). The elution was performed with a gradient of 20−50% eluent B in 30 min. The flow rate was 1 mL/min with eluent A: 0.1% TFA in water and eluent B: 0.1% TFA in acetonitrile. The detection of the peptides was at λ = 220 nm.

Peptide Analytical Characterization. Peptides were characterized by analytical HPLC (see above), mass spectrometry, amino acid analysis, and thin layer chromatography (TLC). The amino acid composition of the peptides was verified by amino acid analysis using an LC 3000 system from Eppendorf-Biotronik (Hamburg, Germany). Full peptide hydrolysis was carried out in 6 N HCl at 110 °C in sealed tubes for 24 h. Afterwards, the samples were dried in a vacuum concentrator and redissolved. The concentrations were determined by comparison with an amino acid standard. The results were in the expected range for peptide contents between 50 and 80%.

For characterization of peptide molar masses, matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) mass spectrometry were used. MALDI mass spectra were measured on an ultrafleXtreme, an autoflex III smartbeam, and an autoflex II instrument (Bruker Daltonics, Bremen, Germany); ESI mass spectra were produced on a micrOTOF-Q III device (Bruker Daltonics, Bremen, Germany) as earlier described.[14]

Determination of Disulfide Connectivity. Disulfide connectivity of the fully oxidized peptides was determined by MS and MS/MS analysis of a chymotryptic digest as earlier described.[10]

Enzyme Activity Assay. Enzyme activity was measured using a fluorogenic enzyme activity assay.[30] The measurements were carried out with the substrate H−Tyr(3-NO_2)−Glu(NH−$(CH_2)_4$−NH−

DOI: 10.1021/acs.jmedchem.8b01982
J. Med. Chem. 2019, 62, 3513−3523

Abz)−Val−Lys−Val−Ile−NH_2 as described previously.[10] IC_{50} plots were fitted in Origin Pro 8G (Origin Lab Corporation, Northampton, MA, USA) by nonlinear regression using the equation $y = v_{max}/(1 + (x/x_0)^s)$ with y = response, v_{max} = maximum velocity, x = inhibitor concentration, x_0 = relative IC_{50}, and s = slope factor. Results were subsequently illustrated in GraphPad v 7.04 (Synergy Software, Reading, PA, USA). IC_{50} values are shown as the means ± standard deviation. Statistical analysis was performed using an unpaired t-test (Holm−Sidak method) by means of GraphPad v 7.04 (Synergy Software, Reading, PA, USA). P values less than 0.01 were considered statistically significant.

Whole Blood and Plasma Preparation. Phlebotomy was performed on consenting healthy donors in accordance with the Declaration of Helsinki and the University of North Carolina Institutional Review Board. Blood was collected by venipuncture using a 21G butterfly needle (Becton, Dickenson and Company, Franklin Lakes, NJ) into 0.105 M sodium citrate, pH 5.5 (10% v/v, final concentration). Normal pooled plasma (NPP) was prepared from four healthy donors. For each donor, the first 5 mL were discarded. Platelet-poor plasma was prepared by sequential centrifugation (150g for 25 min, then 20 000g for 20 min), pooled, aliquoted, flash frozen with liquid nitrogen, and stored at −80 °C.

Whole Blood Clot Contraction Assay. Clotting was triggered in recalcified (10 mM, final concentration) whole blood via the addition of a tissue factor (Innovin diluted 1:12 000, 1 pM, final concentration). Final reaction volumes were 200 μL (85% whole blood; 10% tridegin variant, T101, or HBS; 2.5% Innovin; and 2.5% calcium chloride). Clot contraction proceeded at 37 °C for 120 min in siliconized multiwell plates. Contracted clots were removed and weighed. Analysis of clot weight was performed in GraphPad v 7.02 (Synergy Software, Reading, PA, USA) using the equation y = bottom + (top − bottom)/(1 + (x/IC_{50})) with y = response and x = inhibitor concentration. IC_{50} values are shown as the means ± standard deviation. Statistical analysis was performed using an unpaired t-test (Holm−Sidak method) by means of GraphPad v 7.04 (Synergy Software, Reading, PA, USA). P values less than 0.01 were considered statistically significant.

MD Simulations. Because no resolved structures were present for any of the three tridegin isomers, threaded optimized models from an earlier work[10] of the same served as starting structures for the simulations. GROMACS 5.1.4[31−34] was used for all the MD simulations in this study. An individual peptide was placed in the center of a cubic box, ensuring a separation of at least 2 nm between the peptide and the edges of the box. The TIP3P[35] water model was used as the solvent to fill the box. No ions were added to balance the charges because all three isomers had a net neutral charge. Simulations were run using the AMBER99SB-ILDN[36] force field. Structures with the opened disulfide bonds were created via the method described in earlier work.[12] 5000 steps of steepest descents were carried out as the energy minimization protocol. A thermal equilibration at 300 K using a velocity rescaling Berendsen thermostat[37] and a constant pressure equilibration using the Parrinello−Rahman barostatat[38,39] at 1 atm were carried out for 10 ns each, prior to the production MD. During both the constant temperature (NVT) and constant pressure (NPT) ensemble simulations, position restraints on all bonds were applied using the LINCS[40] algorithm. Two sets of MD simulations were conducted for each variant of the three tridegin isomers. From the first set of simulations conducted for 100 ns, it was decided that the peptides required further unrestrained equilibration, based on the large RMSD deviations observed. Because the observations from these first 100 ns do not directly contribute to the discussions of results, they have been presented in the Supporting Information (Figure S9). The final frame from the 100 ns simulation was independently simulated for an additional 300 ns for each isomer with all of the pre-processing steps such as energy minimization as well as temperature and pressure equilibration applied again. Thus, a total of 3.6 μs of MD simulations were conducted during the course of this study. The results are discussed from the observations in the final 300 ns simulations. The production MD for all simulations in this study was done with a 2 fs time. Periodic boundary conditions were applied to the system. Long-range electrostatics were accounted by the particle mesh Ewald method.[41,42] 10 000 frames were written to the trajectory simulation per 100 ns of simulation time. The effect of periodic boundary conditions was adjusted by suppression of the center of mass movement from the trajectory prior to analysis. Visualizations of conformations for the analysis were performed using Visual Molecular Dynamics (VMD).[43]

Molecular Docking Studies. Molecular docking studies of the tridegin analogues were conducted with the dock_runensemble.mcr macro embedded in YASARA[44] suite (version 18.4.2). This macro utilizes the Vina[45] algorithm to dock a ligand to a receptor ensemble with flexible side-chains. In this ensemble docking approach,[46] the tridegin analogue was docked 400 times on a receptor ensemble created by 20 high scoring side chain orientations of FXIIIa (PDB 4KTY). The procedure was adopted from a previous work[47] that employed the same protocol. The activated FXIII structure was optimized first by removing the small molecule peptide-like inhibitor (ID: PRD_001125) and filling missing loops and residues on the FREAD loop modeling server (http://opig.stats.ox.ac.uk/webapps/fread/php/).[48] The structure was energy minimized in the YASARA suite pre-docking using a combination of steepest descent and simulated annealing minimization runs.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jmedchem.8b01982.

Details of peptide synthesis and analytics; details of activity assays, MD simulation data, additional analysis of MD, and experimental details turbidity assay (PDF)

Dynamics of tridegin isomer A with in silico removal of disulfide bond C_{19}−C_{25} (AVI)

Dynamics of tridegin isomer B with in silico removal of disulfide bond C_{19}−C_{25} (AVI)

Dynamics of tridegin isomer C with in silico removal of disulfide bond C_{19}−C_{25}(AVI)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: dimhof@uni-bonn.de. Phone: +49-228-735254. Fax: +49-228-736829.

ORCID

Torsten Steinmetzer: 0000-0001-6523-4754

Diana Imhof: 0000-0003-4163-7334

Author Contributions

C.A.B. and Thomas Schmitz contributed equally. The study has been designed by A.B. and D.I. Thomas Schmitz and C.A.B. performed synthesis and analysis of the peptides. Thomas Schmitz, C.A.B., K.H., and Torsten Steinmetzer performed and analyzed the enzymatic activity assay. L.A.H. and A.S.W. designed, conducted, and analyzed the turbidity and whole blood contraction assays. M.S., A.A.P.G., and A.B. carried out the computational studies. Torsten Steinmetzer, A.S.W., A.B. and D.I. analyzed and interpreted all data. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

The Bruker micrOTOF-Q instrument (to D.I.) was funded by the University of Bonn, the Ministry of Innovation, Science and Research of North-Rhine Westphalia and the DFG. This study was financially supported by the German Foundation of

DOI: 10.1021/acs.jmedchem.8b01982
J. Med. Chem. 2019, 62, 3513−3523

Heart Research and by the University of Bonn (to D.I.) as well as the National Institutes of Health (R01HL126974, to A.S.W.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Authors are grateful to Dr. M. Sylvester (Core Facility Mass Spectrometry, Institute of Biochemistry and Molecular Biology, University of Bonn) and Dr. M. Engeser (Department of Mass Spectrometry, Institute of Chemistry, University of Bonn) for access to the MALDI instruments. We thank Dr. P. Heimer for his support during peptide synthesis and analysis as well as Dr. M. M. Aleman for analyzing T101 in the whole blood clot clotting assay.

■ ABBREVIATIONS

DMF, dimethylformamide; ESI, electrospray ionization; FXIII, blood coagulation factor XIII; HPLC, high-performance liquid chromatography; MALDI, matrix-assisted laser desorption/ionization; MD, molecular dynamics; NPP, normal pooled plasma; RMSD, root-mean square deviation; RMSF, root-mean square fluctuation; SPPS, solid-phase peptide synthesis; TLC, thin-layer chromatography

■ REFERENCES

(1) Correnti, C. E.; Gewe, M. M.; Mehlin, C.; Bandaranayake, A. D.; Johnsen, W. A.; Rupert, P. B.; Brusniak, M.-Y.; Clarke, M.; Burke, S. E.; De Van Der Schueren, W.; Pilat, K.; Turnbaugh, S. M.; May, D.; Watson, A.; Chan, M. K.; Bahl, C. D.; Olson, J. M.; Strong, R. K. Screening, large-scale production and structure-based classification of cystine-dense peptides. *Nat. Struct. Mol. Biol.* **2018**, *25*, 270−278.

(2) Grütter, M. G.; Priestle, J. P.; Rahuel, J.; Grossenbacher, H.; Bode, W.; Hofsteenge, J.; Stone, S. R. Crystal Structure of the Thrombin-Hirudin Complex: A Novel Mode of Serine Protease Inhibition. *EMBO J.* **1990**, *9*, 2361−2365.

(3) Fuller, E.; Green, B. R.; Catlin, P.; Buczek, O.; Nielsen, J. S.; Olivera, B. M.; Bulaj, G. Oxidative Folding of Conotoxins Sharing an Identical Disulfide Bridging Framework. *FEBS J.* **2005**, *272*, 1727−1738.

(4) Tietze, A. A.; Tietze, D.; Ohlenschläger, O.; Leipold, E.; Ullrich, F.; Kühl, T.; Mischo, A.; Buntkowsky, G.; Görlach, M.; Heinemann, S. H.; Imhof, D. Structurally Diverse μ-Conotoxin PIIIA Isomers Block Sodium Channel NaV1.4. *Angew. Chem., Int. Ed.* **2012**, *51*, 4058−4061.

(5) Almeida, J. R.; Mendes, B.; Lancellotti, M.; Marangoni, S.; Vale, N.; Passos, Ó.; Ramos, M. J.; Fernandes, P. A.; Gomes, P.; Da Silva, S. L. A Novel Synthetic Peptide Inspired on Lys49 Phospholipase A2 from Crotalus Oreganus Abyssus Snake Venom Active against Multidrug-Resistant Clinical Isolates. *Eur. J. Med. Chem.* **2018**, *149*, 248−256.

(6) Moga, M.; Dimienescu, O. G.; Arvătescu, C. A.; Ifteni, P.; Pleş, L. Anticancer Activity of Toxins from Bee and Snake Venom − An Overview on Ovarian Cancer. *Molecules* **2018**, *23*, 692.

(7) Sarfo-Poku, C.; Eshun, O.; Lee, K. H. Medical Application of Scorpion Venom to Breast Cancer: A Mini-Review. *Toxicon* **2016**, *122*, 109−112.

(8) Finney, S.; Seale, L.; Sawyer, R. T.; Wallis, R. B. Tridegin, a New Peptidic Inhibitor of Factor XIIIa, from the Blood-Sucking Leech Haementeria Ghilianii. *Biochem. J.* **1997**, *324*, 797−805.

(9) Böhm, M.; Kühl, T.; Hardes, K.; Coch, R.; Arkona, C.; Schlott, B.; Steinmetzer, T.; Imhof, D. Synthesis and Functional Characterization of Tridegin and Its Analogues: Inhibitors and Substrates of Factor XIIIa. *ChemMedChem* **2012**, *7*, 326−333.

(10) Böhm, M.; Bäuml, C. A.; Hardes, K.; Steinmetzer, T.; Roeser, D.; Schaub, Y.; Than, M. E.; Biswas, A.; Imhof, D. Novel Insights into Structure and Function of Factor XIIIa-Inhibitor Tridegin. *J. Med. Chem.* **2014**, *57*, 10355−10365.

(11) Chang, J.-Y. Diverse Pathways of Oxidative Folding of Disulfide Proteins: Underlying Causes and Folding Models. *Biochemistry* **2011**, *50*, 3414−3431.

(12) Paul George, A. A.; Heimer, P.; Maaß, A.; Hamaekers, J.; Hofmann-Apitius, M.; Biswas, A.; Imhof, D. Insights into the Folding of Disulfide-Rich μ-Conotoxins. *ACS Omega* **2018**, *3*, 12330−12340.

(13) Steiner, A. M.; Bulaj, G. Optimization of Oxidative Folding Methods for Cysteine-Rich Peptides: A Study of Conotoxins Containing Three Disulfide Bridges. *J. Pept. Sci.* **2011**, *17*, 1−7.

(14) Heimer, P.; Tietze, A. A.; Bäuml, C. A.; Resemann, A.; Mayer, F. J.; Suckau, D.; Ohlenschläger, O.; Tietze, D.; Imhof, D. Conformational μ-Conotoxin PIIIA Isomers Revisited: Impact of Cysteine Pairing on Disulfide-Bond Assignment and Structure Elucidation. *Anal. Chem.* **2018**, *90*, 3321−3327.

(15) Stieler, M.; Weber, J.; Hils, M.; Kolb, P.; Heine, A.; Büchold, C.; Pasternack, R.; Klebe, G. Structure of Active Coagulation Factor XIII Triggered by Calcium Binding: Basis for the Design of Next-Generation Anticoagulants. *Angew. Chem., Int. Ed.* **2013**, *52*, 11930−11934.

(16) Zhu, Q.; Liang, S.; Martin, L.; Gasparini, S.; Ménez, A.; Vita, C. Role of Disulfide Bonds in Folding and Activity of Leiurotoxin I: Just Two Disulfides Suffice. *Biochemistry* **2002**, *41*, 11488−11494.

(17) Han, T. S.; Zhang, M.-M.; Walewska, A.; Gruszczynski, P.; Robertson, C. R.; Cheatham, T. E.; Yoshikami, D.; Olivera, B. M.; Bulaj, G. Structurally Minimized μ-Conotoxin Analogues as Sodium Channel Blockers: Implications for Designing Conopeptide-Based Therapeutics. *ChemMedChem* **2009**, *4*, 406−414.

(18) Arolas, J. L.; Castillo, V.; Bronsoms, S.; Aviles, F. X.; Ventura, S. Designing Out Disulfide Bonds of Leech Carboxypeptidase Inhibitor: Implications for Its Folding, Stability and Function. *J. Mol. Biol.* **2009**, *392*, 529−546.

(19) Yu, R.; Seymour, V. A. L.; Berecki, G.; Jia, X.; Akcan, M.; Adams, D. J.; Kaas, Q.; Craik, D. J. Less Is More: Design of a Highly Stable Disulfide-Deleted Mutant of Analgesic Cyclic α-Conotoxin Vc1.1. *Sci. Rep.* **2015**, *5*, 13264.

(20) Han, P.; Wang, K.; Dai, X.; Cao, Y.; Liu, S.; Jiang, H.; Fan, C.; Wu, W.; Chen, J. The Role of Individual Disulfide Bonds of μ-Conotoxin GIIIA in the Inhibition of Nav1.4. *Mar. Drugs* **2016**, *14*, 213.

(21) Wu, X.; Wu, Y.; Zhu, F.; Yang, Q.; Wu, Q.; Zhangsun, D.; Luo, S. Optimal Cleavage and Oxidative Folding of α-Conotoxin Txib as a Therapeutic Candidate Peptide. *Mar. Drugs* **2013**, *11*, 3537−3553.

(22) Akaji, K.; Kiso, Y. Synthesis of Cystine Peptides. *Synthesis of Peptides and Peptidomimetics*; Goodman, M., Ed.; Georg Thieme Verlag: Stuttgart, Germany, 2004; 101−177.

(23) Mochizuki, M.; Tsuda, S.; Tanimura, K.; Nishiuchi, Y. Regioselective Formation of Multiple Disulfide Bonds with the Aid of Postsynthetic S-Tritylation. *Org. Lett.* **2015**, *17*, 2202−2205.

(24) Andreu, D.; Albericio, F.; Solé, N. A.; Munson, M. C.; Ferrer, M.; Barany, G. Formation of Disulfide Bonds in Synthetic Peptides and Proteins. *Peptide Synthesis Protocols*; Humana Press: New York City, USA, 1995; 91−169.

(25) Zhang, S.; Lin, F.; Hossain, M. A.; Shabanpoor, F.; Tregear, G. W.; Wade, J. D. Simultaneous Post-Cysteine(S-Acm) Group Removal Quenching of Iodine and Isolation of Peptide by One Step Ether Precipitation. *Int. J. Pept. Res. Ther.* **2008**, *14*, 301−305.

(26) Chikwana, E.; Davis, B.; Morakinyo, M. K.; Simoyi, R. H. Oxyhalogen-Sulfur Chemistry − Kinetics and Mechanism of Oxidation of Methionine by Aqueous Iodine and Acidified Iodate. *Can. J. Chem.* **2009**, *87*, 689−697.

(27) Aleman, M. M.; Byrnes, J. R.; Wang, J.-G.; Tran, R.; Lam, W. A.; Di Paola, J.; Mackman, N.; Degen, J. L.; Flick, M. J.; Wolberg, A. S. Factor XIII Activity Mediates Red Blood Cell Retention in Venous Thrombi. *J. Clin. Invest.* **2014**, *124*, 3590−3600.

(28) Byrnes, J. R.; Duval, C.; Wang, Y.; Hansen, C. E.; Ahn, B.; Mooberry, M. J.; Clark, M. A.; Johnsen, J. M.; Lord, S. T.; Lam, W. A.; Meijers, J. C. M.; Ni, H.; Ariëns, R. A. S.; Wolberg, A. S. Factor XIIIa-

DOI: 10.1021/acs.jmedchem.8b01982
J. Med. Chem. 2019, 62, 3513−3523

Dependent Retention of Red Blood Cells in Clots Is Mediated by Fibrin α-Chain Crosslinking. *Blood* **2015**, *126*, 1940−1948.
(29) Hetényi, C.; Van Der Spoel, D. Blind Docking of Drug-Sized Compounds to Proteins with up to a Thousand Residues. *FEBS Lett.* **2006**, *580*, 1447−1450.
(30) Hardes, K.; Hammamy, M. Z.; Steinmetzer, T. Synthesis and Characterization of Novel Fluorogenic Substrates of Coagulation Factor XIII-A. *Anal. Biochem.* **2013**, *442*, 223−230.
(31) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.
(32) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(33) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(34) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(35) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926.
(36) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber Ff99SB protein force field. *Proteins* **2010**, *78*, 1950−1958.
(37) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity-Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(38) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(39) Nosé, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **1983**, *50*, 1055−1076.
(40) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(41) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(42) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(43) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(44) Krieger, E.; Vriend, G. YASARA View − Molecular Graphics for All Devices − from Smartphones to Workstations. *Bioinformatics* **2014**, *30*, 2981−2982.
(45) Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 445−461.
(46) Novoa, E. M.; Ribas de Pouplana, L.; Barril, X.; Orozco, M. Ensemble Docking from Homology Models. *J. Chem. Theory Comput.* **2010**, *6*, 2547−2557.
(47) Peherstorfer, S.; Brewitz, H. H.; Paul George, A. A.; Wißbrock, A.; Adam, J. M.; Schmitt, L.; Imhof, D. Insights into Mechanism and Functional Consequences of Heme Binding to Hemolysin-Activating Lysine Acyltransferase HlyC from Escherichia Coli. *Biochim. Biophys. Acta, Gen. Subj.* **2018**, *1862*, 1964−1972.
(48) Choi, Y.; Deane, C. M. FREAD revisited: Accurate loop structure prediction using a database search algorithm. *Proteins* **2010**, *78*, 1431−1440.

DOI: 10.1021/acs.jmedchem.8b01982
J. Med. Chem. 2019, 62, 3513−3523

4.3.2 Summary

In the current study three 2-disulfide-bonded variants based on their corresponding 3-disulfide-bonded parent isomers of the anticoagulant peptide tridegin were analyzed by a combination of experimental and computational means. In the 2-disulfide-bonded variants, the cysteines forming the C19-C25 disulfide bond were mutated to serines. Structure models of the same (variants) were also created based on the 3-disulfide-bonded structures available from earlier studies.[203] Experimental results describing the inhibitory activity of the 2-disulfide-bonded tridegin variants on FXIIIa indicated that they were as active as their 3-disulfide bonded parental forms. It was therefore clear that, the removal of the C19-C25 disulfide bond in the three-disulfide bonded isomers had little impact on the structure and consequently the inhibitory activity. A complete explanation of the behavior observed in the experimental assays was provided by the MD simulation and docking results. It was obvious upon comparing the MD trajectories of the three-disulfide-bonded and two-disulfide-bonded tridegin variants explored virtually the same conformations in simulation and importantly, the removal of the C19-C25 disulfide bond did not disrupt the core structure of the peptide unlike what was observed with the previous studies using μ-conotoxins.[129,294] The docking simulations revealed that, all of the three 2-disulfide-bonded variants of tridegin used in this study bound directly over the active site of FXIIIa defined by the catalytic triad C314, H373 and D396 thereby inhibiting the activity of FXIIIa in similar fashion.

Part 2

4.4 Chapter IV

Insights into mechanism and functional consequences of heme binding to hemolysin-activating lysine acyltransferase HlyC from *Escherichia coli*

Original research article

Authors*

Sandra Peherstorfer, Hans Henning Brewitz, Ajay Abisheck Paul George, Amelie Wißbrock, Jana Maria Adam, Lutz Schmitt, Diana Imhof

This peer-reviewed research article was published in *Biochimica et Biophysica Acta - General Subjects*.

Citation

Biochim. Biophys. Acta - Gen. Subj., 1862, 1964–1972

DOI: 10.1016/j.bbagen.2018.06.012

4.4.1 Introduction

Many pathogenic bacteria that infect mammalian hosts use heme as an iron source, mostly from hemoglobin. This includes uropathogenic *Escherichia coli* (UPECs) which cause urinary tract infections (UTIs) and related conditions such as cystitis and pyelonephritis, thereby posing a severe threat to human health. UPEC strains such as UTI89 produce factors like exotoxin hemolysin A (HlyA) that lyse erythrocytes by pore formation in the cell membrane. HlyA requires activation from its inactive precursor form, which is mediated by the protein hemolysin activating acyltransferase (HlyC). It is proposed that heme transiently binds to the residue H151 on the proposed heme-binding motif (HBM) on HlyC that has a regulatory effect resulting in the inhibition of the enzymatic activity of the protein. An array of experimental (UV-vis spectroscopy, fluorescence quenching assay, surface plasmon resonance (SPR) and *in vitro* hemolysis assay) and computational (homology modeling, molecular docking, MD simulations) investigations are carried out to determine the mechanism and functional consequences of the proposed HlyC.

***Own contribution**

The computational study including creation of the homology models, molecular docking, and MD simulations and analysis of the results was performed by myself. The design and the content of the manuscript has been compiled by Diana Imhof and Lutz Schmitt. The layout and preparation of the figures related to the computational study has been planned and carried out by me in consultation and agreement with Diana Imhof. All authors contributed to writing of the manuscript.

BBA - General Subjects 1862 (2018) 1964–1972

Contents lists available at ScienceDirect

BBA - General Subjects

journal homepage: www.elsevier.com/locate/bbagen

Insights into mechanism and functional consequences of heme binding to hemolysin-activating lysine acyltransferase HlyC from *Escherichia coli*

Sandra Peherstorfer[a,1], Hans Henning Brewitz[b,1], Ajay Abisheck Paul George[b], Amelie Wißbrock[b], Jana Maria Adam[b], Lutz Schmitt[a,*], Diana Imhof[b,*]

[a] Institute of Biochemistry, University of Düsseldorf, 40255 Düsseldorf, Germany
[b] Pharmaceutical Biochemistry and Bioanalytics, Pharmaceutical Institute, University of Bonn, 53121 Bonn, Germany

ARTICLE INFO

Keywords:
Hemolysin A
HlyC
Heme binding
Heme regulation
Heme-regulatory motif

ABSTRACT

Background: Tight regulation of heme homeostasis is a critical mechanism in pathogenic bacteria since heme functions as iron source and prosthetic group, but is also toxic at elevated concentrations. Hemolysin-activating lysine-acyltransferase (HlyC) from *Escherichia coli* is crucial for maturation of hemolysin A, which lyses several mammalian cells including erythrocytes liberating large amounts of heme for bacterial uptake. A possible impact and functional consequences of the released heme on events employing bacterial HlyC have remained unexplored.

Methods: Heme binding to HlyC was investigated using UV/vis and SPR spectroscopy. Functional impact of heme association was examined using an in vitro hemolysis assay. The interaction was further studied by homology modeling, molecular docking and dynamics simulations.

Results: We identified HlyC as potential heme-binding protein possessing heme-regulatory motifs. Using wild-type protein and a double alanine mutant we demonstrated that heme binds to HlyC via histidine 151 (H151). We could show further that heme inhibits the enzymatic activity of wild-type HlyC. Computational studies illustrated potential interaction sites in addition to H151 confirming the results from spectroscopy indicating more than one heme-binding site.

Conclusions: Taken together, our results reveal novel insights into heme-protein interactions and regulation of a component of the heme uptake system in one of the major causative agents of urinary tract infections in humans.

General significance: This study points to a possible novel mechanism of regulation as present in many uropathogenic *E. coli* strains at an early stage of heme iron acquisition from erythrocytes for subsequent internalization by the bacterial heme-uptake machinery.

1. Introduction

Heme is an important iron source for many pathogenic bacteria [1–3]. In mammalian host organisms the vast majority of heme molecules are found in hemoglobin in erythrocytes [4]. The level of free hemoglobin in blood is efficiently reduced to a minimum by the scavenger protein haptoglobin that forms stable hemoglobin-haptoglobin complexes [5]. Various bacteria have developed the ability to use hemoglobin and hemoglobin-haptoglobin complexes as a potential heme and iron source [6, 7]. To capture hemoglobin that is sheltered in erythrocytes some bacteria produce hemolysins that are capable of erythrocyte lysis resulting in the liberation of hemoglobin [8, 9].

Uropathogenic *Escherichia coli* (UPECs) are a serious threat to human health as they cause a range of urinary tract infections (UTIs) including cystitis and pyelonephritis [10]. Some UPEC strains, e.g. UTI89, produce virulence factors like the exotoxin hemolysin A (HlyA) that is capable of inducing pore formation in the membrane of mammalian cells such as erythrocytes resulting in cell lysis [11, 12]. Secretion of HlyA is accomplished by a type I secretion system (T1SS) composed of the ABC transporter hemolysin B (HlyB), the membrane fusion protein HlyD, both located in the inner membrane, and the outer membrane factor TolC (Fig. 1) [13–15]. HlyA is expressed as an

Abbreviations: HlyC, hemolysin-activating lysine-acyltransferase; UPEC, uropathogenic *Escherichia coli*; UTIs, urinary tract infections; HlyA, hemolysin A; Hly B, hemolysin B; T1SS, type I secretion system; HRM, heme-regulatory motif; MD, molecular dynamics

* Corresponding author: L. Schmitt, Institute of Biochemistry, University of Düsseldorf, Universitätsstrasse 1, 40225 Düsseldorf, Germany; D. Imhof, Pharmaceutical Biochemistry and Bioanalytics, Institute of Pharmacy, University of Bonn, An der Immenburg 4, 53121 Bonn, Germany.
E-mail addresses: Lutz.Schmitt@hhu.de (L. Schmitt), dimhof@uni-bonn.de (D. Imhof).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.bbagen.2018.06.012
Received 28 February 2018; Received in revised form 12 June 2018; Accepted 13 June 2018
Available online 15 June 2018

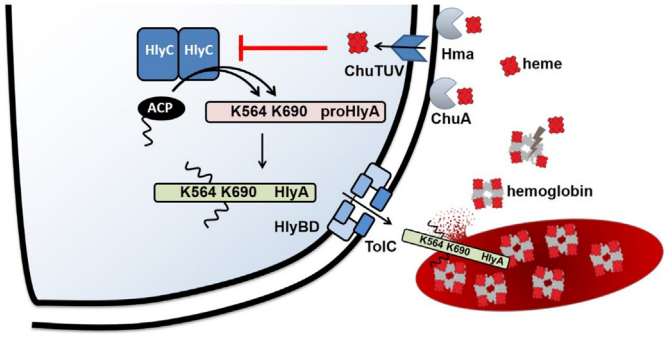


Fig. 1. Activation of proHlyA by HlyC, HlyA-mediated erythrocyte lysis and heme uptake by *E. coli*. ACP, acyl carrier protein; TolC, outer membrane protein; HlyB, hemolysin A translocation ATP-binding protein; HlyD, hemolysin-secretion membrane fusion protein D; Hma, haem-acquisition protein; ChuA, outer membrane heme/hemoglobin receptor. ChuTUV, ABC transporter system.

unfolded and inactive precursor protein, proHlyA [16]. Activation is mediated by the intracellular hemolysin-activating lysine-acyltransferase (HlyC) that catalyzes the transfer of two acyl groups from acyl carrier protein (ACP) to two lysine residues in proHlyA, K564 and K690, respectively [17]. Acylation of proHlyA is a prerequisite for the hemolytic activity on mammalian cells. Upon secretion, HlyA is folded into its mature form in the presence of extracellular Ca^{2+} ions [18]. The mature exotoxin causes pore formation in erythrocytes resulting in hemoglobin liberation [15, 19].

This hemoglobin is either utilized directly (as intact protein) or after proteolytic degradation, and heme is taken up by the bacterial high-affinity TonB-dependent heme receptors ChuA and Hma that facilitate transport across the outer cell membrane [20, 21]. Internalization from the periplasm to the cytoplasm is completed by the ABC-transport system ChuTUV (Fig. 1) and has been investigated in enterohemorrhagic *E. coli* [21].

Although heme is required for bacterial growth and survival, elevated levels of heme cause toxicity due to catalysis of Fenton-type reactions or peroxidase-like activity [22]. Thus, unbridled lysis of erythrocytes and subsequent liberation of hemoglobin and heme might have adverse effects on bacterial growth. Existence of a regulatory mechanism that adjusts heme release from mammalian cells to an appropriate amount is therefore conceivable.

Transient heme interactions based on so-called heme-regulatory motifs, i.e. short sequence stretches on a proteins surface possessing a heme coordination site, have been described in detail earlier [23–26]. In a previous study we suggested HlyC from *E. coli* to potentially bind heme via a histidine-based heme-binding (HBM) or heme-regulatory motif (HRM) (Fig. 2A) located in the C-terminal α-helix of HlyC [27]. Fe(III)-protoporphyrin (heme) binding to a 9mer peptide containing this suggested motif revealed strong heme-binding affinity (K_D ~0.24 μM) and a Soret-band shift to 415 nm [27]. Herein we demonstrate heme binding to wild-type HlyC (HlyC-wt) protein and confirm contribution of the suggested HRM using a double mutant (HlyC-mut), where both histidine residues (H151 and H152) in the HRM were replaced by alanines. We further show the inhibitory effect of heme on the enzymatic activity of HlyC and involvement of the suggested HRM as probed in an in vitro hemolysis assay. Molecular modeling and docking studies revealed further insight into the heme-binding mode, facilitating an improved conceptual understanding of the underlying biomolecular processes.

2. Materials and methods

2.1. Materials

Hemin and Co(III)PPIX were purchased from Sigma-Aldrich (Steinheim, Germany). Ga(III)PPIX, Zn(II)PPIX and PPIX were purchased from Frontier Scientific (Logan, UT, USA).

2.2. Methods

2.2.1. Protein expression and purification

Expression, purification and unfolding of proHlyA was carried out as described earlier [28]. The plasmid pET28b-hlyC-His based on the sequence of chromosomally encoded HlyC from UPEC strain UTI89 (Uniprot entry Q1R2T4) was produced according to an existing protocol [29]. The plasmid pET28b-hlyC-mut-His was revealed by PCR-amplification from pET28b-hlyC-His using the primer pair 5′-GCGAATAAAATTTTTAAACAATATGCCGCCGAGTTAATAACTGAAG-3′ and 5′-CTTCAGTTATTAACTCGGCGGCATATTGTTTAAAAATTTTATTCGC-3′ (MWG-Biotech). Plasmid isolation was performed using NucleoSpin Plasmid Miniprep Kit (Macherey-Nagel). Sequence identities were confirmed using DNA sequencing (GATC-Biotech). For protein expression, an overnight culture of *E. coli* BL21(DE3) (Novagen, Germany) with plasmid pET28b-hlyC-His or pET28b-hlyC-mut-His, respectively, was grown on LB medium substituted with kanamycin (final concentration in all cases 30 μg/ml) at 37 °C and 180 rpm and used to inoculate six 2 l Erlenmeyer flasks containing 600 ml of LB medium complemented with 1% w/v glucose and kanamycin. Cells were grown at 37 °C and 180 rpm to an OD600 of 0.4 and expression was subsequently induced using final concentrations of 1 mM IPTG (HlyC-wt) or 0.1 mM IPTG (HlyC-mut). Expression was performed at 18 °C and 180 rpm for 18 to 20 h. Cells were harvested by centrifugation at 4000 *g* for 15 min at 4 °C. Cell pellets were dissolved in 50 ml resuspension buffer 20 mM Hepes pH 8.0, 150 mM KCl, 20% (v/v) glycerin and one tablet of protease inhibitor mixture (cOmpleteTM) on ice. Cell disruption was performed using a cell disrupter (Microfluidics System M-110P) at 1.5 kbar for three times on ice. The mixture was centrifuged at 20000 *g* for 30 min at 4 °C and the supernatant (~300 ml) was immediately loaded on a 5 ml HighTrap HP (GE Healthcare) affinity column prepared with 100 mM $NiCl_2$ and equilibrated with resuspension buffer. For HlyC-wt, the column was washed with 100 ml resuspension buffer and 200 ml IMAC buffer A (resuspension buffer containing 10 mM imidazole) and eluted using 100 ml IMAC buffer B (resuspension buffer containing 500 mM imidazole) as a gradient elution (to 100% IMAC buffer B). For HlyC-mut, the column was washed with 100 ml resuspension buffer, 200 ml IMAC buffer A, 60 ml of 15% IMAC buffer B and eluted using 20 ml IMAC buffer B as a gradient elution (15% to 60% IMAC buffer B). Fractions were collected at a volume of 2 ml and analyzed using SDS-PAGE. Pure fractions were combined and immediately loaded on a HighLoadTMSuperdexTM75 10/60 SEC column equilibrated with SEC buffer A (20 mM Hepes pH 8.0, 150 mM KCl, 20% (v/v) glycerin). Separation was carried out at a flow rate of 0.7 ml/min, 2 ml fractions were collected and analyzed using SDS-PAGE. Pure fractions were combined and His-Tag removal was performed by adding TEV protease (molar ratio 1:25) and incubation at 4 °C overnight. The mixture was subjected to a second IMAC purification step to remove undigested HlyC-His protein and therefore loaded on 5 ml HighTrap HP (GE Healthcare) affinity column equilibrated with SEC buffer A. The column was washed with 100 ml SEC buffer A at a flow rate of 1.5 ml/min. Elution was performed using step elution with each 20 ml of 10%, 20%, 30%, 40%, 50% and 100% IMAC buffer B and fractions were collected at a volume of 1 ml. Pure fractions were combined and buffer exchange was carried out using a second SEC with SEC buffer B (20 mM Hepes pH 8.0, 150 mM KCl). For HlyC-wt, SEC was carried out using a Superdex 75 16/60 column at a flow rate of 0.7 ml/min, while for HlyC-mut, a Superdex 75 10/300 column at a flow rate of 0.35 ml/min was facilitated. Fractions were collected at a volume of 2 ml. Pure fractions were combined and concentrated to a concentration of 0.2 mg/ml using 15 ml Amicon (MWCO 10 kDa, Merck Millipore) centrifugation tubes. Protein aliquots were prepared and stored at −80 °C.

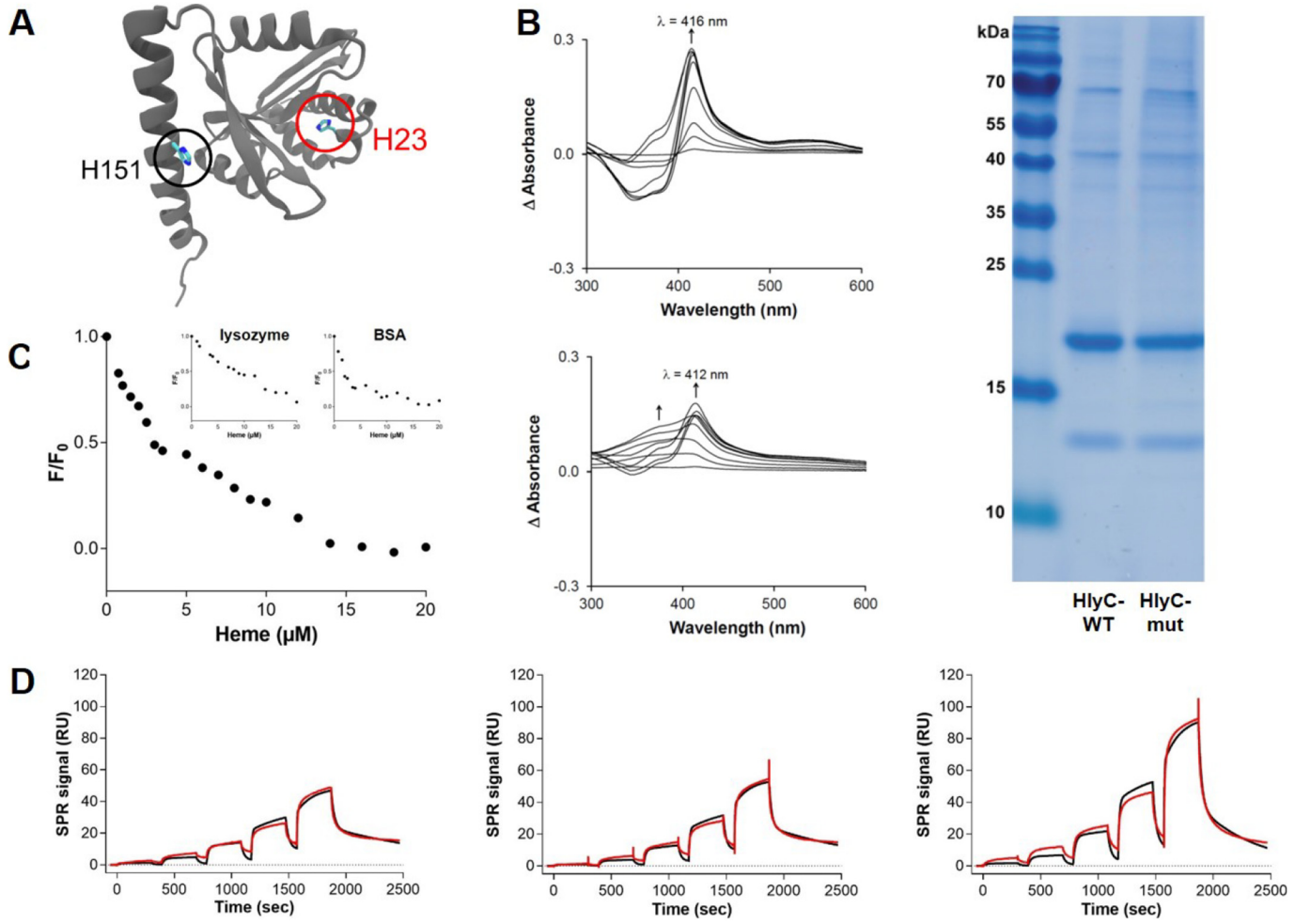


Fig. 2. Heme binding to HlyC as shown by UV–vis, fluorescence and SPR spectroscopy. (A) Chain A of the homodimeric homology model of HlyC is depicted with the proposed heme-coordinating histidine (H151) highlighted and labeled. (B) Left: UV–vis studies of HlyC-wt (upper spectrum) and HlyC-mut (lower spectrum) (each 2.5 μM) incubated with heme (0–40 μM), shown are the difference spectra. The K_d-value for heme binding was determined to be 0.51 ± 0.10 μM for HlyC-wt protein. Fitting the data for HlyC-mut was not possible and therefore no K_d value could be determined. Right: SDS-PAGE of HlyC-wt and HlyC-mut after expression and purification. The band at ~ 20 kDa corresponds to HlyC-wt and HlyC-mut, respectively. The band at ~13 kDa corresponds to ACP (C) Fluorescence quenching of lysozyme, HlyC-wt and BSA (all 0.8 μM) by heme (0–20 μM). (D) SPR analysis of heme interaction with HlyC-wt (left, K_{d1} ~2.9 ± 0.1 μM, K_{d2} ~2.3 ± 0.6 μM), HlyC-mut protein (center, K_{d1} ~3.3 ± 0.1 μM, K_{d2} ~4.2 ± 1.2 μM) and BSA (right, K_{d1} ~8.0 ± 0.1 μM, K_{d2} ~2.8 ± 0.3 μM). K_d-values were obtained assuming a complex binding model using a kinetic global fit (heterogenous ligand binding, red curve).

2.2.2. Preparation of heme solutions

Heme and other PPIX derivatives were dissolved in 30 mM NaOH to a concentration of 1 mM and incubated in the dark on ice for 30 min. The heme concentration of the solution was determined using the molar extinction coefficient of heme (ε385 = 58.440 M − 1 cm − 1) [30, 31]. The stock solutions were diluted in Hepes buffer (100 mM, pH 7.0) to the desired concentration directly before use.

2.2.3. UV–vis measurements

All measurements were performed in Hepes buffer (100 mM, pH 7.0) with a Thermo Fisher Scientific (Vanta, Finland) Multiskan Go spectrophotometer using the experimental set up established earlier [32]. Heme (final concentration of 0–40 μM) and HlyC proteins (final concentration of 2.5 μM) were mixed and incubated in the dark at room temperature for 60 min to ensure complete complex formation. Difference spectra were performed by subtracting pure heme and pure protein from heme-protein complex spectra for each heme concentration.

2.2.4. Fluorescence quenching assay

The intrinsic fluorescence of HlyC-wt, BSA and lysozyme was used to investigate dynamic and static fluorescence quenching by heme as a consequence of complex formation with the fluorescence quencher heme. All proteins were used in a final concentration of 0.8 μM and mixed with 0–20 μM heme. Incubation time was 60 min. Fluorescence intensity was measured at an excitation wavelength of 306 nm and an emission wavelength of 352 nm using a Jasco FP-8300 (Tokyo, Japan) spectrofluorometer. All values (F) were normalized to the fluorescence intensity of the protein without heme (F0).

2.2.5. SPR binding studies

SPR measurements were performed as in running buffer (10 mM Hepes (pH 7.4), 150 mM NaCl, 0.05% Tween 20) using a Biacore T200 instrument (GE Healthcare, 79111 Freiburg, Germany) at 25 °C based on Karnaukhova et al. [33]. The proteins HlyC-wt, HlyC-mut and BSA were covalently immobilized by amine coupling on a CM5 sensor chip (GE Healthcare). BSA was diluted to 2.5 μg/ml in acetate buffer (pH 5.0). HlyC-wt and HlyC-mut were diluted to 10 μg/ml in acetate buffer (pH 5.0). The proteins were severally injected at 10 μl/min on a EDC/NHS activated flow cell until immobilization levels of 810 RU (HlyC-wt), 750 RU (HlyC-mut) and 2217 RU (BSA) were achieved. For reference subtraction an activated/deactivated flow cell was utilized. The heme concentration was determined spectrophotometrically as

mentioned above. K_D value and kinetic parameter determinations were achieved by using titration series of five consecutive injections with varying heme concentrations (0.08 μM, 0.31 μM, 1.25 μM, 5 μM, 20 μM, samples were diluted in running buffer). The measurements were performed at a flow rate of 30 μl/min using the implemented (Biacore T200 Control Software (GE Healthcare)) standard single-cycle kinetics method. The surface was subsequently regenerated by two injections of NaOH (25 mM and 500 mM). An injection series of running buffer was subtracted from each curve for double referencing. All curves were globally fitted using the heterogeneous ligand analysis model.

2.2.6. In vitro hemolysis assay

The in vitro hemolysis assay was performed as reported earlier [29] with minor changes which are described in the following. Incubation of proHlyA and HlyC was performed for 5 min at 23 °C. Washing of defibrillated sheep blood cells (Oxid, United Kingdom) was performed as described earlier [29]. Reaction mixture was mixed with washed sheep blood cells and incubated at 37 °C for 30 min. Cellular particles and intact cells were removed by centrifugation at 14.000 rpm for one minute. The supernatant was transferred to a 96 well plate and diluted with water (1,32). The content of hemoglobin was determined at a wavelength of 544 nm using spectrophotometry.

For K_M-value determination HlyC-wt (constant concentration of 0.23 μM) was incubated with concentrations of 0 to 2 μM of proHlyA. Evaluation and data analysis was performed using GraphPad Prism 5 (GraphPad software, Inc.) software and eq. $Y = Vmax * X/(Km + X)$.

To investigate the effect of different concentrations of heme (0–100 μM) on HlyC-wt or HlyC-mut, respectively, the proteins were mixed with heme and incubated at 23 °C for 30 min prior to addition of proHlyA. Subsequent steps were carried out as mentioned above and described earlier [29]. IC_{50}-vaules were determined using QTiPlot (version 0.9.7.8) software and non-linear regression with equation $v = ((A - B) / (1 + ([I] / IC_{50})s)) + B$ (A, maximal enzymatic velocity, B, minimal enzymatic velocity, [I], concentration of the inhibitor, s, Hill slope). All IC_{50}-values are stated as MEAN ± SE.

Inhibition of HlyC-wt by heme (25 μM), PPIX (25 μM), Ga(III)PPIX (25 μM), Co(III)PPIX (25 μM) and Zn(II)PPIX (25 μM) was tested by mixing HlyC-wt with the respective PPIX derivative for 30 min at 23 °C prior to addition of proHlyA. Subsequent steps were carried out as mentioned above and described earlier [29].

2.2.7. Homology modeling of HlyC-wt and HlyC-mut proteins

Barring the change of residues 151 and 152 in the input sequences from histidine in HlyC-wt to alanine in HlyC-mut, the homology modeling experiment followed identical settings and steps as follows. First, 3 iterations of PSI-BLAST were run against UniRef90 [34] to build a position-specific scoring matrix (PSSM) from related sequences. This profile was then used to search the PDB. The method tries to avoid false positives by excluding common protein purification tags. An E-value threshold of 0.5 was set on PSI-BLAST to limit hits too remote to be considered as templates. For both HlyC-wt and HlyC-mut, this resulted in 1 distinct template hit which was an X-ray diffraction structure of a bacterial toxin-activating acyltransferase from pathogenic *Actinobacillus pleuropneumoniae* [35] (PDB-ID: 4WHN). Template ranking determined by the product of the BLAST alignment score, the WHAT_CHECK [36] structural quality score in the PDBFinder2 database and the target coverage, showed optimal values of 400.65 for HlyC-wt and 394.29 for HlyC-mut. The alignment target coverage for both HlyC-wt and HlyC-mut was 99%. Secondary structure predictions were obtained by running PSI-BLAST to create a target sequence profile and feeding it to the PSI-Pred secondary structure prediction server. With the default oligomerization state set at 4, the program appropriately built homodimeric models (Fig. 3). 50 conformations were used to model the loop sequence MN between the individual subunits of the homodimeric model. At this stage, overall z-score obtained as a sum of the z-scores of dihedrals, 1D packing and 3D packing was found to be 0.242 and 0.226 for the models of HlyC-wt and HlyC-mut respectively. Both these initial models were subjected to a full unrestrained simulated annealing minimization which produced the final models with improved z-scores of 0.257 for HlyC-wt and 0.305 for HlyC-mut.

2.2.8. Molecular docking of heme to HlyC-wt and HlyC-mut

Molecular docking studies [37] were performed on both HlyC-wt and HlyC-mut to predict, observe and elucidate the mechanism of heme binding to these proteins, especially at the experimentally determined HRMs. Since, the HlyC-wt and HlyC-mut homology models were homodimers, only one subunit from the dimeric model was used as receptor in docking studies (Fig. 5). The ligand heme was downloaded from the ChemSpider [38] database (ChemSpider ID 16739951). Preparation of the docking setup involved first running separate energy minimizations of the receptor subunits of both the HlyC-wt and HlyC-mut homology models followed by a separate energy minimization of the heme ligand. For the first phase of the docking experiments a 20 Å × 20 Å × 20 Å cubic simulation cell was built around the nine amino acid long HRM (residues 147 to 155). This was done to narrow down the search space for docking around the experimentally determined sites. The AutoDock Vina [39] algorithm was used to run docking simulations via YASARA. Docking was conducted using the "ensemble docking" approach, a method that could improve the validity of docking results obtained from homology models [40]. Using this approach, receptor flexibility was induced by creating a receptor ensemble of 20 high-scoring solutions of the side-chain rotamer network at a set temperature of 298 K. To each member of this ensemble, the ligand was docked 400 times resulting in a total of 8000 docking runs per experiment. The second phase of docking simulations was done to check for other potential binding sites in the receptor than what has been experimentally described (H151). In this case, the only change to the experimental setup was an increase in size of the simulation cell to 50 Å × 50 Å × 50 Å, which had the entire A subunit of HlyC-wt (and HlyC-mut) inside the search space. For each run, docking results were clustered to produce a ranked list of receptor-ligand docked complexes, based on predicted binding energies. An r.m.s.d value of 5 Å of the ligand was used as threshold for membership within a particular cluster.

2.2.9. Molecular dynamics simulations of heme-bound complexes of Hlyc-wt and Hlyc-mut

Although results from docking studies produce geometrically optimized, energy minimized conformations of receptor-ligand complexes, one cannot comment on the stability of the bound conformations in solution or in a dynamic environment with certainty. This often results in the validity of docking studies being viewed with skepticism [41]. To address this concern, molecular dynamics (MD) simulations of selected docked complexes were run again via YASARA [42]. Since the docking experiments were conducted in YASARA, it was an easy choice to conduct the MD simulations using the same program, especially considering the ease of parameterizing the docked complex with the force field. The prime intention here, was to check if the ligand was stable in its predicted binding site in a dynamic environment. Selected docked complexes of HlyC-wt-heme and Hlyc-mut-heme were loaded into YASARA as starting structures for the MD simulations. A set of preliminary operations such as cleaning of the structure, preparing the topology, optimizing of the hydrogen bond network, creating the (85 Å × 85 Å × 85 Å) cubic simulation cell, assigning force field parameters, filling the cell with water, placing counter ions, predicting pKa values and assigning protonation states, running an energy minimization and setting initial atom velocities according to a Boltzmann distribution were done automatically by YASARA. A significant part of these preliminary operations includes the parameterization of the simulated system with the force field. This was done by the AutoSMILES method (http://www.yasara.org/autosmiles.htm) in YASARA that implements automatic force field parameterization for almost any system including those containing metal ions. For the current system, the

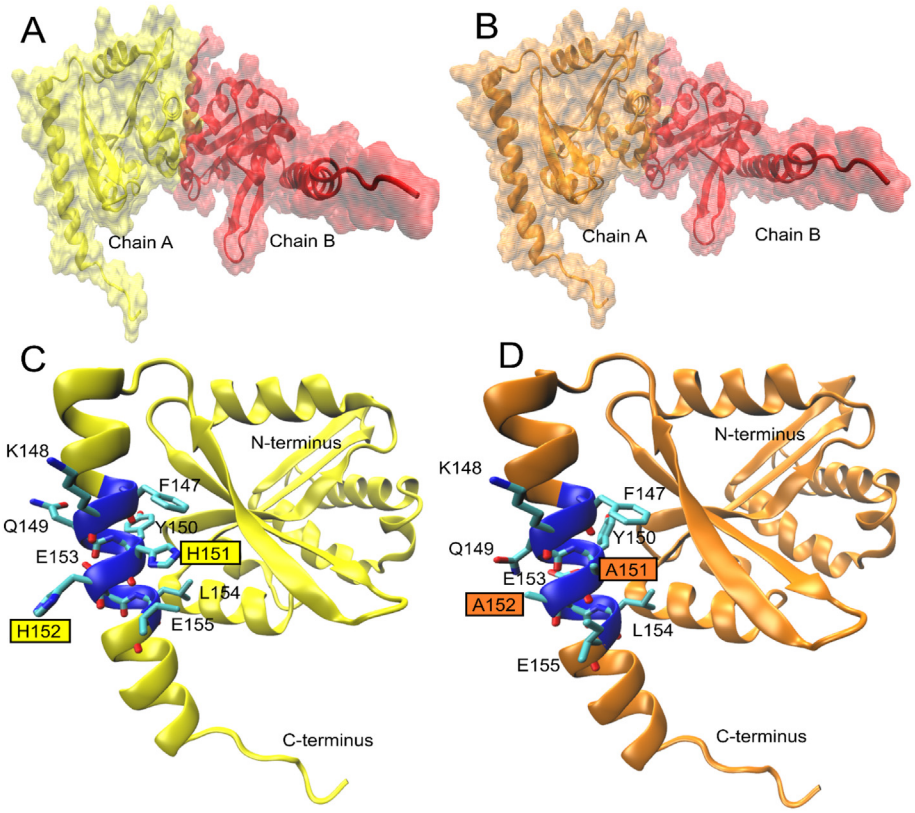


Fig. 3. Homology models of the HlyC-wt and HlyC-mut structures. (A) Surface representation of the homology model of homodimeric HlyC-wt. Chain A (yellow) was used as the biological subunit for docking and molecular dynamics studies. (B) Surface representation of the homology model of homodimeric HlyC-mut. Chain A (orange) was used as the biological subunit for docking and molecular dynamics studies. (C) Ribbon representation (yellow) of the chain A subunit of HlyC-wt with the proposed HRM shown in blue and the residues shown in sticks. The labels of H151 and H152 are highlighted since they were replaced by alanines for creating HlyC-mut. (D) Ribbon representation (orange) of the chain A subunit of HlyC-mut with the proposed HRM shown in blue and the residues shown in sticks. The mutated residue labels A151 and A152 are highlighted.

AutoSMILES force field parameters for the protoporphyrin ring were assigned using GAFF [43] and AM1-BCC [44] with Fe3+ VdW parameters from Li et al. [45]. Since the simple electrostatic interactions in empirical force fields do not usually reproduce dative bonds to metal ions, explicit bonds were (automatically) added between the iron and the coordinating nitrogens. With the knowledge that AutoSMILES uses GAFF (General AMBER Force Field) to derive force field parameters, a compatible all-atom force field namely AMBER 11 (aka AMBER99SB-ILDN) [46] was used to conduct the MD simulations. Simulations were conducted for 50 ns at 298 K under the constant pressure NPT ensemble with the pressure maintained at 1 bar. Periodic boundary conditions were used, long range forces were computed with a cutoff of 8 Å and the particle-mesh Ewald method [47] was used for long range coulomb interactions. Simulations were run with a 2 fs time step, pair lists updated every 10 fs and snapshots saved every 100 fs thus accounting to 500 snapshots saved per simulation.

3. Results

3.1. Protein expression

The expression of HlyC-wt and HlyC-mut in *E. coli* yielded pellets of 4 to 5 g wet cells per liter of cell culture after expression for 18 to 20 h at 18 °C. Protein purification was performed from the soluble fraction after cell disruption via a two-step purification protocol. The C-terminal His_6-tag allowed the purification of the protein via an immobilized metal ion affinity chromatography (IMAC), followed by size exclusion chromatography (SEC) to remove imidazole from the buffer. HlyC containing fractions with purity larger than 90% as judged by SDS-PAGE, were combined and the His_6-tag was cleaved off by TEV protease. After TEV protease cleavage, the products were separated from the non-cleaved proteins via a reverse IMAC and a buffer exchange was performed using SEC. Protein identity and cleavage of the His_6-tag were verified by Western blotting using a penta-His antibody as described previously [28]. The purified proteins were finally analyzed by SDS-PAGE (Fig. 2B). Two major bands were detected on the gel, corresponding to the size of HlyC-wt/HlyC-mut monomer (between the 15 and 25 kDa marker bands) and a second band of unknown protein identity between the 10 and 15 kDa marker bands. A subsequent matrix-assisted laser desorption ionization - time of flight (MALDI-TOF) mass spectrometric (MS) analysis of the unknown protein species identified several peptides of the acyl-carrier protein (ACP) from *E. coli* strain K12 (UniProtKB entry P0A6A8) among peptides from HlyC (*E. coli* strain UTI89/UPEC) and a few peptides from other *E. coli* K12 proteins (data not shown). The co-purification of ACP with HlyC provided sufficient amounts of acyl-ACP for the in vitro acylation of proHlyA and allowed to perform the previously described in vitro HlyA acylation assay without the addition of crude *E. coli* extract as external source for fatty acids carried out by ACP [28]. It also supports the view that the C-terminal helix harboring the two mutated histidine residues is not involved in ACP binding. The yield of purified HlyC-wt and HlyC-mut protein without His_6-tag out of 1 l cell culture were approximately 0.1 mg and 0.06 mg, respectively. Both proteins could be concentrated up to ~ 0.35 mg/ml and remained active after storage at −80 °C for several weeks in 50 mM Hepes, pH 8.0, 150 mM KCl.

3.2. Heme binding to bacterial protein HlyC

Heme (Fe(III)-PPIX) binding to a peptide representing the predicted HRM (FKQYH151H152ELI) of HlyC has been investigated earlier [35] and revealed a K_D value of 0.24 μM and a Soret-band shift to 415 nm [27]. This sequence stretch, which is located in a long α-helix at the C-terminus of HlyC, is surface-exposed and the suggested axial heme ligand, i.e. histidine 151 (H151), is easily accessible as can also be seen from the homology model of HlyC (Fig. 2A) created based on a previous study [35] using the crystal structure of the toxin-activating acyltransferase from *Actinobacillus pleuro-pneumoniae*.

The heme-binding ability of HlyC-wt protein was first investigated using UV–vis spectroscopy. Titrations of HlyC-wt (2.5 μM) with heme (0–40 μM) revealed a K_D value of 0.51 ± 0.10 μM for heme and a

Soret-band shift to 416 nm (Fig. 2B), which is typical for histidine and a water molecule as axial heme ligands [27] at the central iron(III) ion within the protoporphyrin ring system. It is rather unlikely that due to the experimental setup and the fact that heme binding occurs at the protein surface, and not within a heme-binding pocket, the oxidation state of iron(III) in heme is changed to iron(II).

BSA binds heme and can be used as positive control in heme-binding studies, while lysozyme is frequently used as a negative control in the same context [30]. Heme binding to HlyC-wt was also confirmed using a fluorescence quenching assay. Therefore, HlyC-wt (0.8 μM), bovine serum albumin (BSA, 0.8 μM) and lysozyme (0.8 μM) were titrated with different concentrations of heme (0–20 μM).

HlyC-wt and BSA showed non-linear fluorescence quenching curves caused by heme-protein complex formation (static and dynamic quenching), while lysozyme exhibited no heme binding as indicated by a linear decrease of fluorescence quenching (dynamic quenching) corresponding to earlier studies [29].

To investigate the importance of the suggested HRM for heme binding, the double mutant H151A, H152A (HlyC-mut) was prepared and analyzed under the same conditions as the wild-type protein. Preparation of this double mutant was carried out to avoid a potential ligand shift of the heme molecule from histidine 151 to histidine 152 upon mutation of H151 by alanine. Additionally, no expression was detected for one of the single H-A mutants (H152A). Thus, we did not investigate the single mutants. Histidines were first considered as potential interaction partners rather than the adjacent tyrosine residue, since it was found earlier that tyrosine in Tyr-based HRMs represents a weaker heme-coordination partner than histidine in His-based ones [27]. This was also supported by molecular docking studies performed herein (see below, Fig. 5).

The differential spectrum of the HlyC-mut was found to differ significantly in absorbance intensity and binding mode from the spectrum of HlyC-wt protein (Fig. 2B). Fitting the obtained data based on the equation established by Bogdan et al. [48] to determine a K_D value was not possible. The spectra were less intense and showed a broad shoulder at ~370 nm for heme concentrations up to 15 μM heme, while a band shift to 412 nm appeared for concentrations exceeding 20 μM. With respect to the earlier reported heme-binding modes [27, 49, 50], the former indicates a pentacoordinated state involving a tyrosine, whereas the latter refers to histidine as an axial heme ligand. Nevertheless, the differences between wild-type and mutant protein indicate the important role of the suggested HRM for heme binding to HlyC.

In order to get deeper insight into the binding characteristics of heme association with HlyC, surface plasmon resonance (SPR) measurements of both complexes (Fig. 2D), i.e. heme-HlyC-wt and heme-HlyC-mut, were performed inspired by the study of Karnaukhova et al. [33], in which BSA was again used as positive control. Heme binding was observed for both, HlyC-wt and HlyC-mut and the heme-protein interaction occurred in a biphasic mode according to the observed binding curves. Fitting the data with a 1:1 kinetic model was not possible. Instead the data hinted at a more complex binding mode. Therefore a global heterogeneous ligand model that assumes two independent binding events was chosen as it offered the best fit. Binding affinity was denoted as K_{D1} and K_{D2}. K_D values for both HlyC proteins were in a similar range (HlyC-wt: K_{D1} ~2.9 ± 0.1 μM, K_{D2} ~2.3 ± 0.6 μM vs. HlyC-mut: K_{D1} ~3.3 ± 0.1 μM, K_{D2} ~4.2 ± 1.2 μM). BSA exhibited slightly higher K_D values (K_{D1} ~8.0 ± 1.2 ± 0.3 μM, K_{D2} ~2.8 μM). This confirmed heme binding to HlyC-wt but at the same time revealed that HlyC-mut interacts with heme indicating the presence of further potential interaction site(s) (Fig. 2D). Comparing the K_D value of HlyC-wt obtained from UV/Vis spectroscopy (0.51 ± 0.10 μM) with the K_Ds determined by SPR analysis (K_{D1} 2.9 μM K_{D2} 2.3 μM) revealed slighty deviating binding affinities. This can be explained by the differing experimental set up. While both interactions partners are in solution during UV/Vis spectroscopy, the protein is immobilized on a chip when SPR analysis is performed. This may spatially block distinct heme binding epitopes of the proteins and therefore interfere with the potential heme-protein interaction.

3.3. Impact of heme on the biological activity of HlyC-wt and HlyC-mut

To examine whether heme binding to HlyC has an impact on its biological activity, an in vitro hemolysis assay was performed as described earlier [28]. Hereby, the activity of HlyC is detected indirectly by the amount of hemoglobin that is liberated from sheep blood erythrocytes as a consequence of proHlyA activation through HlyC-mediated acylation. Quantification of hemoglobin is performed by measuring the absorbance increase at 544 nm. The enzymatic activation of proHlyA is stopped by the addition of washed sheep blood cells containing 20 μM $CaCl_2$ to the reaction mixture as proHlyA folds in the presence of Ca^{2+} ions and is no longer acylated in this form [28].

Suitability of the assay performed was tested in several preliminary experiments. Fe^{2+} ions did not inhibit this reaction as tested at concentrations of 33 μM, 66 μM and 100 μM (data not shown). Heme did not interfere with HlyA, since pre-activated HlyA was fully active in the presence or absence of 33 μM heme (data not shown).

Heme alone also had no lytic effect on sheep blood cells. The absorbance of added heme can be neglected as its absorbance at 544 nm is, due to several dilution steps, below the detection limit of the spectrometer. However, it was included into the background correction of the absorbance values. HlyC-wt and HlyC-mut were tested and exhibited similar enzymatic activities. The K_M-value of HlyC-wt for proHlyA was determined to be ~0.2 μM (Fig. 4A), which is in the order of earlier reported values [51]. Increasing heme concentrations (0–100 μM) showed a strong inhibitory effect on the enzymatic reaction. Already a concentration of 12.5 μM abolished the liberation of hemoglobin into the supernatant. The IC_{50}-value of HlyC-wt for heme was determined as 5.6 ± 0.4 μM (Fig. 4B). An intracellular heme concentration of about 5–10 μM is reasonable since several bacterial heme exporters are only active in the presence of ~10 μM heme [52]. This further supports the hypothesis of a heme-mediated HlyC regulation. Although elevated heme concentrations showed an inhibitory effect on HlyC-mut, the effect was weaker compared to the wildtype protein (IC_{50}-value 8.3 ± 0.2 μM). These results thus revealed a contribution of the suggested HRM to the heme-dependent inhibition of HlyC (Fig. 4B), however, also implied the contribution of additional motif(s) besides the H151/H152 sequence. This was therefore examined in more detail in the molecular docking studies described below.

Apart from heme other PPIX derivatives, e.g. Co(III)PPIX and Ga(III)PPIX, exhibited a comparable inhibitory effect on HlyC-wt at a concentration of 25.0 μM. For PPIX and Zn(II)PPIX the inhibition was significantly weaker compared to heme (Fe(III)PPIX) at the same concentration (Fig. 4c). This demonstrates the importance of the central metal atom in PPIX. The comparatively weaker interaction of PPIX and Zn(II)PPIX might be driven by the porphyrin-protein interactions caused by e.g. π-π stacking and hydrophobic interactions. PPIX derivatives with the same charge compared to Fe(III)PPIX, i.e. Co(III)PPIX and Ga(III)PPIX exhibit similar inhibitory effects likely due to similar coordination between the PPIX-metal atom and the protein. This seems to be especially the case for Ga(III)PPIX where even the atom radius is similar to iron(III). These findings suggest that the inhibitory effect is not unique to heme but also to suitable PPIX derivatives (Co(III)PPIX and Ga(III)PPIX). These compounds might lead to the same structural change and eventually result in a similar decrease of enzymatic activity. Ga(III)PPIX was successfully used in our earlier studies as surrogate for Fe(III)-heme in order to enable solution NMR structure analysis by preventing the paramagnetic effect of iron(III) species on nuclear magnetic relaxation of protons [27, 49, 50].

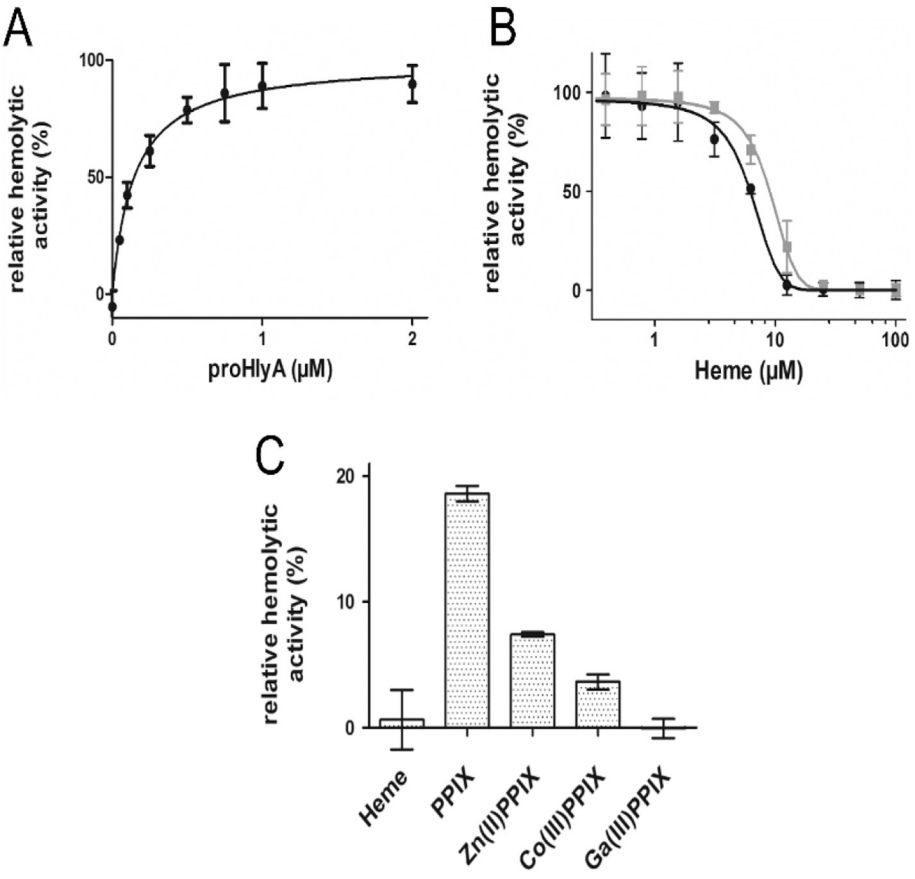


Fig. 4. Impact of heme on the enzymatic activity of HlyC-wt and HlyC-mut. (A) Michaelis-Menten kinetics of HlyC-wt. K_M of HlyC for proHlyA (0–2 μM) was determined to be 0.15 ± 0.02 μM. (B) Inhibition of HlyC-wt (black circles) and HlyC-mut (gray squares) by heme (0–100 μM). IC_{50} values were determined as 5.6 ± 0.4 μM and 8.3 ± 0.2 μM, respectively. Values are normalized to the hemolytic effect of 0.21% SDS. (C) Inhibition of HlyC-wt by heme, PPIX and Zn(II)-, Co(III)- and Ga(III)-PPIX derivatives (each 25.0 μM). Values are normalized to the hemolytic effect of 0.21% SDS.

3.4. Molecular docking and molecular dynamics simulation studies of HlyC-wt- and HlyC-mut-heme complexes

The unavailability of an experimentally derived structure for HlyC led us to employ homology modeling within this study. A homology model of a bacterial toxin-activating acyltransferase was earlier introduced by Green et al. [35]. In contrast to the aforementioned study, our work required the modeling of a double mutant of the HlyC structure (HlyC-mut). In order to keep the computational setup uniform for both the wild type (HlyC-wt) and mutant (HlyC-mut), new homology models were built. The 3D structures of both proteins were produced using the homology modeling module of YASARA [53] (version 17.4.17) based on the X-ray structure of a toxin-activating acyltransferase from pathogenic *Actinobacillus pleuro-pneumoniae* (PDB-ID: 4WHN). Homodimeric models with alignment coverage of 99% were built for both, HlyC-wt and HlyC-mut (Fig. 3). It can be seen that the envisaged heme-binding residues H151 and H152 are located in the C-terminal α-helix of the protein (Fig. 3), thus enabling easy access of the heme ligand. In the following section, results observed from two separate rounds (see methods) of molecular docking simulations performed using AutoDock Vina [39] are described. From the molecular docking runs focused on the C-terminal α-helix (containing the proposed H151) of HlyC-wt, 70 distinct HlyC-wt-heme complex conformations were formed. The membership of a certain bound conformation within a cluster had a threshold of 5 Å heavy ligand atom root mean square deviation (r.m.s.d). Despite being automatically sorted by predicted binding energies, each cluster was manually inspected to select only those complexes that had coordination interactions between HlyC-wt and a surface-bound heme molecule that facilitates transient heme binding. Interestingly, out of the 70 complexes investigated, only one docked complex showed a coordinative interaction between H151 and the heme iron ion (Fig. 5). The docked pose was formed by a combination of heme coordination via H151 along with a hydrogen bond between one oxygen atom of the protoporphyrin ring and H152. The bindings additionally stabilized by hydrophobic interactions between heme and the surrounding residues N144, K145, F147, K148, I155 and π-stacking induced by F147 (Fig. 5A, Fig. 5B). A 50 ns MD simulation of this complex conformation indicated that this binding mode of heme to HlyC-wt was extremely stable. Snapshots taken at different intervals of the simulation showed that the coordination between H151 and heme along with the hydrogen bond between heme and H152 persisted throughout the course of simulation time, adding evidence to the strength of this binding event (Fig. 5C, Fig. 5D). In addition, HlyC-mut docking with heme revealed 63 complex clusters, however, none of these showed coordination interactions. This observation added further evidence to experimentally found heme coordination via H151/H152 of HlyC.

In the second round of docking simulations, when the search space for the docking runs was increased to cover the entire molecule (see Methods), 75 complexes were formed for both HlyC-wt and HlyC-mut. Here, three other potential binding sites, namely Y44, Y52, and Y104 emerged in both proteins exhibiting identical behavior observed from the MD simulations. The key difference between the interactions of heme with these sites in comparison to heme interaction with H151 is that, in a dynamic environment, the coordinative interaction with heme was not permanent and was lost in a few nanoseconds of MD simulation time. This leads us to welcome the idea that there exist further sites for heme to interact but at a weaker magnitude. A detailed description of how these additional sites behave is available in the Supplementary Material.

4. Discussion

The present study demonstrates for the first time that HlyC is a potential heme-binding protein and that heme is capable of inhibiting its enzymatic activity as observed in an in vitro hemolysis assay. The K_D

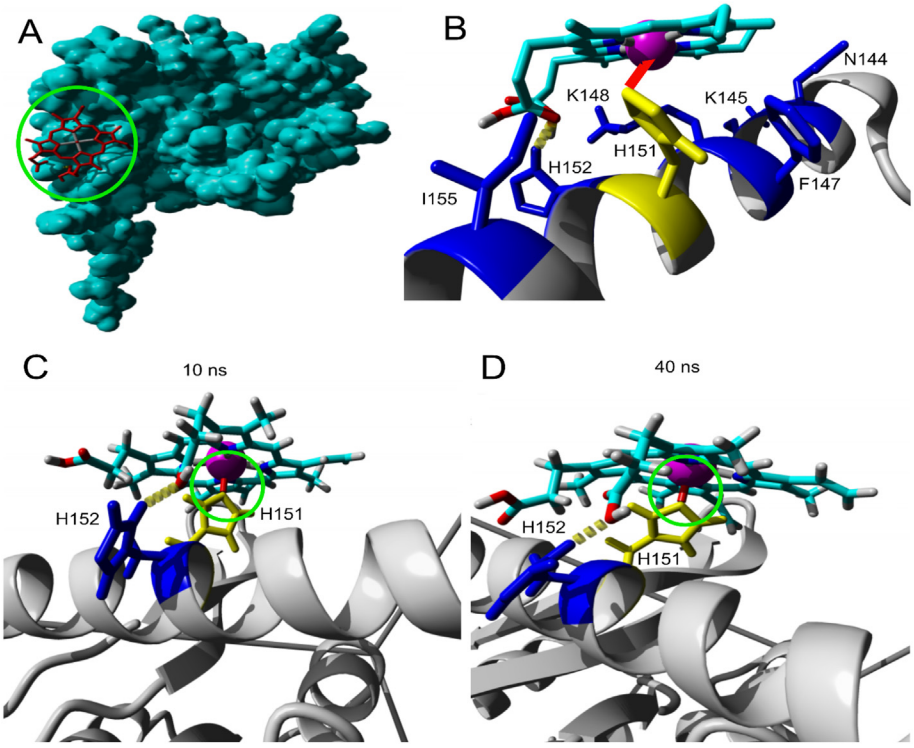


Fig. 5. Heme coordination by HlyC-wt via H151. (A) Docked pose of heme to HlyC-wt. The HlyC-wt is represented as van der Waals surface (cyan) and the bound heme as sticks (red). The area of heme binding on the surface of HlyC-wt is highlighted by a green circle. (B) Close-up of the circled region shown in A. Heme is bound to the C-terminal α-helix (gray ribbon) of HlyC-wt. The H151 residue (yellow stick) is involved in coordination (red solid line) of the heme iron(III) ion (pink sphere). The H152 residue (blue stick) forms H-bonds (yellow dotted line) with the oxygen atom of the protoporphyrin ring. Residues N144, K145, F147, K148 and I155 (blue) are additional contributors to the binding by their hydrophobic and π-stacking (F147) interactions with heme. (C) and (D) MD simulation snapshots of heme bound to HlyC-wt at 10 and 40 ns, respectively. The coordination interaction between H151 (yellow stick) and the iron(III) ion is shown by a solid red line highlighted by a green circle. H-bond formation between H152 (blue stick) and an oxygen atom of the protoporphyrin ring is shown as a yellow dotted line.

value of HlyC-wt for heme was determined to be in the lower micromolar range. Such a value is typical for heme-regulated proteins as was demonstrated e.g. for heme-regulated transporter regulator (HrtR) from *Lactococcus lactis* [54]. HrtR, that acts as a transcriptional repressor of a heme efflux pump, is binding heme via two histidine motifs with a K_D value of ~0.4 μM. Heme-regulated proteins are interacting with heme in a transient fashion and are usually characterized by comparatively weak heme-binding affinities with K_D values up to the lower micromolar range [20, 55, 56]. Absorbance spectra of HlyC-wt incubated with heme showed a Soret-band shift which is typical for histidine residues as axial heme-iron ligand [27].

In contrast, HlyC-mut showed a broad absorbance ranging from ~370 nm to 412 nm for heme concentrations up to 15 μM. This finding indicates that loss of the suggested motif in HlyC-wt results in a different heme-binding behavior. The broad absorbance range might indicate multiple, rather non-specific heme interactions with the HlyC-mut protein. Heme concentrations of 20 μM to 40 μM result in formation of a spectrum with a sharper band at 412 nm. The Soret-band shift could be indicative for a second heme-binding motif. Based on the docking and MD results we suspect that this could be one of the three tyrosine-based motifs that were detected (see Supplementary Material). This second site possesses a lower heme-binding affinity according to UV measurements. The significance of the heme binding is reflected by the 1.5-fold lower IC_{50} value of HlyC-wt. The higher IC_{50} value for the mutant indicates a role of the suggested HRM for heme interaction. However, the data also indicates the presence of an additional interaction site.

The findings of this study verify our earlier work on heme-binding peptides that eventually led to the prediction of heme-binding motifs in proteins. We were able to show that our approach can successfully be applied and resulted in the identification of another potential heme-regulated enzyme. A regulatory mechanism on the protein level of HlyC is imaginable and would suggest a heme-mediated regulation at the earliest stage of heme acquisition, i.e. liberation of heme from erythrocytes for subsequent internalization by the bacterial heme-uptake machinery. One can speculate that the posttranslational regulation of hemolysin A activation in dependence of the intracellular heme concentration allows fast adaptation of hemolytically active *E. coli* to the availability of heme. This might be important to avoid heme toxicity as a result of uncontrolled heme liberation because internalization of extracellular heme occurs highly efficient.

Obviously, erythrocytes are not the main target cells of HlyA. However, uropathogenic *E. coli* strains pass the blood-urine barrier in the late stages of urinary tract infection which ultimately results in sepsis. For example, Smith et al. [57] demonstrated that two toxins of UPEC strains, HlyA and CNF-1, are major determinants of cystitis. This suggests that HlyA gets also in its natural habitat into contact with erythrocytes. More importantly, however, Martin et al. recently demonstrated a clear relation between the presence of HlyA and increased heme plasma concentrations in a mouse model [58]. This study and our results on the inhibition of heme on the enzymatic function of HlyC in activating the pore-forming toxin HlyA revelead a new level of regulation and adaption of UPEC strains on their environment. Moreover, antibacterial strategies based on transient heme-protein interactions in bacteria should be taken into consideration exploiting knowledge on the molecular basis of heme-affected processes in pathogenic bacteria [3].

Acknowledgements

The authors like to thank Kornelia Hampel and Bastian Zimmermann (Biaffin GmbH Kassel) for Biacore measurements. Financial support by the Deutsche Forschungsgemeinschaft within SFB813 (to D.I.) and the Jürgen Manchot Graduate School ‘Molecules of Infections’ (to L.S.) is gratefully acknowledged.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft within SFB813 (to D.I.) and the Jürgen Manchot Graduate School ‘Molecules of Infections’ (to L.S.).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagen.2018.06.012.

References

[1] J.R. Sheldon, D.E. Heinrichs, Recent developments in understanding the iron acquisition strategies of gram positive pathogens, FEMS Microbiol. Rev. 39 (2015) 592–630.
[2] J.E. Choby, E.P. Skaar, Heme synthesis and acquisition in bacterial pathogens, J. Mol. Biol. 428 (2016) 3408–3428.
[3] H.H. Brewitz, G. Hagelueken, D. Imhof, Structural and functional diversity of transient heme binding to bacterial proteins, Biochim. Biophys. Acta, Gen. Subj. 1861 (2016) 683–697.
[4] F. Richter, B.H. Meurers, C. Zhu, V.P. Medvedeva, M.F. Chesselet, Neurons express hemoglobin α- and β-chains in rat and human brains, J. Comp. Neurol. 515 (2009) 538–547.
[5] T.D. Hull, A. Agarwal, J.F. George, The mononuclear phagocyte system in homeostasis and disease: a role for heme oxygenase-1, Antioxid. Redox Signal. 20 (2014) 1770–1788.
[6] C.E. Allen, M.P. Schmitt, Utilization of host iron sources by *Corynebacterium diphtheriae*: multiple hemoglobin-binding proteins are essential for the use of iron from the hemoglobin-haptoglobin complex, J. Bacteriol. 197 (2015) 553–562.
[7] C.S. Bates, G.E. Montañez, C.R. Woods, R.M. Vincent, Z. Eichenbaum, Identification and characterization of a *Streptococcus pyogenes* operon involved in binding of hemoproteins and acquisition of iron, Infect. Immun. 71 (2003) 1042–1055.
[8] F. Vandenesch, G. Lina, T. Henry, *Staphylococcus aureus* hemolysins, bi-component leukocidins, and cytolytic peptides: a redundant arsenal of membrane-damaging virulence factors? Front. Cell. Infect. Microbiol. 2 (2012) 12.
[9] T.J. Wiles, M.A. Mulvey, The RTX pore-forming toxin α-hemolysin of uropathogenic *Escherichia coli*: progress and perspectives, Future Microbiol 8 (2013) 73–84.
[10] J. Bien, O. Sokolova, P. Bozko, Role of uropathogenic escherichia coli virulence factors in development of urinary tract infection and kidney damage, Int. J. Nephrol. 2012 (2012).
[11] I.S.S.S.J. Cavalieri, Cytotoxic activity of partially purified *Escherichia coli* alpha haemolysin, J. Med. Microbiol. 15 (1982) 11–21.
[12] I. Gentschev, G. Dietrich, W. Goebel, The *E. coli* alpha-hemolysin secretion system and its use in vaccine development, Trends Microbiol. 10 (2002) 39–45.
[13] S. Létoffé, P. Delepelaire, C. Wandersman, Protein secretion in gram-negative bacteria: assembly of the three components of ABC protein-mediated exporters is ordered and promoted by substrate binding, EMBO J. 15 (1996) 5804–5811.
[14] T. Thanabalu, E. Koronakis, C. Hughes, V. Koronakis, Substrate-induced assembly of a contiguous channel for protein export from *E. coli*: reversible bridging of an inner-membrane translocase to an outer membrane exit pore, EMBO J. 17 (1998) 6487–6496.
[15] S. Thomas, I.B. Holland, L. Schmitt, The type 1 secretion pathway - the hemolysin system and beyond, Biochim. Biophys. Acta, Mol. Cell Res. 1843 (2014) 1629–1641.
[16] P.J. Bakkes, S. Jenewein, S.H.J. Smits, I.B. Holland, L. Schmitt, The rate of folding dictates substrate secretion by the *Escherichia coli* hemolysin type 1 secretion system, J. Biol. Chem. 285 (2010) 40573–40580.
[17] P. Stanley, L.C. Packman, V. Koronakis, C. Hughes, Fatty acylation of two internal lysine residues required for the toxic activity of *Escherichia coli* hemolysin, Science 80 (266) (1994) 1992–1996.
[18] P. Stanley, V. Koronakis, C. Hughes, Acylation of *Escherichia coli* hemolysin: a unique protein lipidation mechanism underlying toxin function, Microbiol. Mol. Biol. Rev. 62 (1998) 309–333.
[19] G. Menestrina, C. Moser, S. Pellet, R. Welch, Pore-formation by *Escherichia coli* hemolysin (HlyA) and other members of the RTX toxins family, Toxicology 87 (1994) 249–267.
[20] E.C. Hagan, H.L.T. Mobley, Haem acquisition is facilitated by a novel receptor Hma and required by uropathogenic *Escherichia coli* for kidney infection, Mol. Microbiol. 71 (2009) 79–91.
[21] A.G. Torres, S.M. Payne, Haem iron-transport system in enterohaemorrhagic *Escherichia coli* O157:H7, Mol. Microbiol. 23 (1997) 825–833.
[22] H. Lin, J. Everse, The cytotoxic activity of hematoheme: evidence for two different mechanisms, Anal. Biochem. 161 (1987) 323–331.
[23] L. Zhang, L. Guarente, Heme binds to a short sequence that serves a regulatory function in diverse proteins, EMBO J. 14 (1995) 313–320.
[24] J.T. Lathrop, M.P. Timko, Regulation by heme of mitochondrial protein transport through a conserved amino acid motif, Science 259 (1993) 522–525.
[25] X.D. Tang, R. Xu, M.F. Reynolds, M.L. Garcia, S.H. Heinemann, T. Hoshi, Haem can bind to and inhibit mammalian calcium-dependent Slo1 BK channels, Nature 425 (2003) 531–535.
[26] L. Zhang, Heme Biology: The Secret Life of Heme in Regulating Diverse Biological Processes, (2011), p. 228.
[27] H.H. Brewitz, N. Goradia, E. Schubert, K. Galler, T. Kühl, B. Syllwasschy, J. Popp, U. Neugebauer, G. Hagelueken, O. Schiemann, O. Ohlenschläger, D. Imhof, Heme interacts with histidine- and tyrosine-based protein motifs and inhibits enzymatic activity of chloramphenicol acetyltransferase from *Escherichia coli*, Biochim. Biophys. Acta, Gen. Subj. 1860 (2016) 1343–1353.
[28] S. Thomas, S.H.J. Smits, L. Schmitt, A simple in vitro acylation assay based on optimized HlyA and HlyC purification, Anal. Biochem. 464 (2014) 17–23.
[29] J. Shen, X. Sheng, Z.N. Chang, Q. Wu, S. Wang, Z. Xuan, D. Li, Y. Wu, Y. Shang, X. Kong, L. Yu, L. Li, K. Ruan, H. Hu, Y. Huang, L. Hui, D. Xie, F. Wang, R. Hu, Iron metabolism regulates p53 signaling through direct Heme-p53 interaction and modulation of p53 localization, stability, and function, Cell Rep. 7 (2014) 180–193.
[30] I. Kirschner-Zilber, E. Rabizadeh, N. Shaklai, The interaction of hemin and bilirubin with the human red cell membrane, BBA - Biomembr. 690 (1982) 20–30.
[31] K.M. Dawson, M.C. Rex, Daphne C. Elliott, William H. Elliott, Jones, Data for Biochemical Research, 2nd ed., Oxford University Press, 1969.
[32] T. Kühl, N. Sahoo, M. Nikolajski, B. Schlott, S.H. Heinemann, D. Imhof, Determination of hemin-binding characteristics of proteins by a combinatorial peptide library approach, Chembiochem 12 (2011) 2846–2855.
[33] E. Karnaukhova, S. Rutardottir, M. Rajabi, L.W. Rosenlöf, A.I. Alayash, B. Åkerström, Characterization of heme binding to recombinant α1-microglobulin, Front. Physiol. 5 (2014).
[34] B.E. Suzek, H. Huang, P. McGarvey, R. Mazumder, C.H. Wu, UniRef: comprehensive and non-redundant UniProt reference clusters, Bioinformatics 23 (2007) 1282–1288.
[35] N.P. Greene, A. Crow, C. Hughes, V. Koronakis, Structure of a bacterial toxin-activating acyltransferase, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) E3058–E3066.
[36] R.W.W. Hooft, G. Vriend, C. Sander, E.E. Abola, Errors in protein structures, Nature 381 (1996) 272.
[37] X.-Y. Meng, H.-X. Zhang, M. Mezei, M. Cui, Molecular docking: a powerful approach for structure-based drug discovery, Curr. Comput. Aided. Drug Des. 7 (2011) 146–157.
[38] Royal Society of Chemistry, ChemSpider Search and Share Chemistry, R. Soc. Chem., 2015.
[39] O. Trott, A.J. Olson, AutoDock Vina, J. Comput. Chem. 31 (2010) 445–461.
[40] E.M. Novoa, L.R. De Pouplana, X. Barril, M. Orozco, Ensemble docking from homology models, J. Chem. Theory Comput. 6 (2010) 2547–2557.
[41] Y.C. Chen, Beware of docking!, Trends Pharmacol. Sci. 36 (2015) 78–95.
[42] E. Krieger, G. Vriend, New ways to boost molecular dynamics simulations, J. Comput. Chem. 36 (2015) 996–1007.
[43] J.M. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem. 25 (2004) 1157–1174.
[44] A. Jakalian, D.B. Jack, C.I. Bayly, Fast, efficient generation of high-quality atomic charges AM1-BCC model: II parameterization and validation, J. Comput. Chem. 23 (2002) 1623–1641.
[45] P. Li, L.F. Song, K.M. Merz, Parameterization of highly charged metal ions using the 12-6-4 LJ-type nonbonded model in explicit water, J. Phys. Chem. B 119 (2015) 883–895.
[46] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, D.E. Shaw, Improved side-chain torsion potentials for the Amber ff99SB protein force field, Proteins Struct. Funct. Bioinforma. 78 (2010) 1950–1958.
[47] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577–8593.
[48] A. Pîrnău, M. Bogdan, Investigation of the interaction between naproxen and human serum albumin, Rom. J. Biophys. 18 (2008) 49–55.
[49] T. Kühl, A. Wißbrock, N. Goradia, N. Sahoo, K. Galler, U. Neugebauer, J. Popp, S.H. Heinemann, O. Ohlenschläger, D. Imhof, Analysis of Fe(III) heme binding to cysteine-containing heme-regulatory motifs in proteins, ACS Chem. Biol. 8 (2013) 1785–1793.
[50] H.H. Brewitz, T. Kühl, N. Goradia, K. Galler, J. Popp, U. Neugebauer, O. Ohlenschläger, D. Imhof, Role of the chemical environment beyond the coordination site: structural insight into FeIII protoporphyrin binding to cysteine-based heme-regulatory protein motifs, Chembiochem 16 (2015) 2216–2224.
[51] M.S. Trent, L.M.S. Worsham, M. Lou Ernst-Fonberg, The biochemistry of hemolysin toxin activation: characterization of HlyC, an internal protein acyltransferase, Biochemistry 37 (1998) 4644–4652.
[52] L.L. Anzaldi, E.P. Skaar, Overcoming the heme paradox: heme toxicity and tolerance in bacterial pathogens, Infect. Immun. 78 (2010) 4977–4989.
[53] E. Krieger, G. Vriend, YASARA view - molecular graphics for all devices - from smartphones to workstations, Bioinformatics 30 (2014) 2981–2982.
[54] P. Stanley, V. Koronakis, K. Hardie, C. Hughes, Independent interaction of the acyltransferase HlyC with two maturation domains of the *Escherichia coli* toxin HlyA, Mol. Microbiol. 20 (1996) 813–822.
[55] D. Lechardeur, B. Cesselin, U. Liebl, M.H. Vos, A. Fernandez, C. Brun, A. Gruss, P. Gaudu, Discovery of intracellular heme-binding protein HrtR, which controls heme efflux by the conserved HrtB-HrtA transporter in *Lactococcus lactis*, J. Biol. Chem. 287 (2012) 4752–4758.
[56] V.A. Villareal, R.M. Pilpa, S.A. Robson, E.A. Fadeev, R.T. Clubb, The IsdC protein from *Staphylococcus aureus* uses a flexible binding pocket to capture heme, J. Biol. Chem. 283 (2008) 31591–31600.
[57] Y.C. Smith, S.B. Rasmussen, K.K. Grande, R.M. Conran, A.D. O'Brien, Hemolysin of uropathogenic *Escherichia coli* evokes extensive shedding of the uroepithelium and hemorrhage in bladder tissue within the first 24 hours after intraurethral inoculation of mice, Infect. Immun. 76 (2008) 2978–2990.
[58] R. Martins, J. Maier, A.D. Gorki, K.V.M. Huber, O. Sharif, P. Starkl, S. Saluzzo, F. Quattrone, R. Gawish, K. Lakovits, M.C. Aichinger, B. Radic-Sarikas, C.H. Lardeau, A. Hladik, A. Korosec, M. Brown, K. Vaahtomeri, M. Duggan, D. Kerjaschki, H. Esterbauer, et al., Heme drives hemolysis-induced susceptibility to infection via disruption of phagocyte functions, Nat. Immunol. 17 (2016) 1361–1372.

4.4.2 Summary

Transient heme binding to the surface of the protein HlyC via H151 on the proposed HBM was investigated by both experimental and computational means. To be certain that the binding indeed took place via the proposed HBM, a mutant version of HlyC (HlyC-mut) wherein the alanines replaced the histidine residues in the motif (H151A and H152A), was also expressed along with the wild type protein (HlyC-wt). Spectroscopic measurements established clear differences between the heme-binding behavior of HlyC-wt and HlyC-wt suggesting that heme binding to the HBM via H151. The biological activity tested between HlyC-wt and HlyC-mut using an *in vitro* hemolysis assay revealed that at elevated concentrations, heme induced a larger inhibitory effect on the enzymatic activity of HlyC-wt over HlyC-mut. The observations and inferences from the experimental results were enhanced by the computational studies performed. High quality homology models of HlyC-wt and HlyC-mut were first built on which molecular docking simulations were conducted. Molecular docking experiments focused on the proposed nonapeptide HBM revealed that heme bound to the H151 residue in the motif. The binding was described by the heme molecule aligned over the motif with the Fe^{3+} ion and the nitrogen atom of H151 found in a distance that favors coordination (under 3Å). The binding was supported by hydrogen bond interactions between heme and the neighboring H152 residue and further hydrophobic and π-π stacking interactions from other residues that make up the motif. No heme binding could be observed when heme was docked to the same region on HlyC-wt proving the importance of the H151 residue for heme binding. The docking results were further validated by performing MD simulations of the Hlyc-wt-heme complexes and the heme binding was found to be strong and consistent in the dynamic, solvent-surrounded environment that the MD simulation induces. The docking simulations considering the entire HlyC protein subunit (for both HlyC-wt and HlyC-mut) found that other residues namely Y44, Y52 and Y104 could possibly coordinate heme. However, MD simulations of these docked complexes revealed that the heme-protein interaction did not persist for beyond few nanoseconds suggesting that these residues might contribute to some weak binding that might be seen in both HlyC-wt and HlyC-mut. Thus, heme's role as a regulatory molecule was demonstrated by studying its binding to HlyC using a combination of experimental and computational approaches with the computational studies adding further clarity to the experimental observations.

4.5 Chapter V

The molecular basis of transient heme-protein interactions: analysis, concept and implementation

Review article

Authors*

Amelie Wißbrock, Ajay Abisheck Paul George, Hans Henning Brewitz, Toni Kühl, Diana Imhof

This peer-reviewed research article was published in *Bioscience Reports*.

Citation

Biosci. Rep. (2019) 39(1), BSR20181940

DOI: 10.1042/BSR20181940

4.5.1 Introduction

Heme is usually considered to be an intrinsic component of many proteins belonging to the families of globins and cytochromes. In these proteins, heme is present as a prosthetic group that is permanently bound into a deep hydrophobic groove or pocket in the protein and enable the proteins perform a wide variety of functions such as oxygen transport (in hemoglobin, myoglobin) and electron transfer (in cytochrome b, cytochrome c) to name a few. However, in the early 1990s the ability of heme to perform alternative functions as effector and signaling molecule were identified. It was realized that, to perform these functions, heme bound (transiently as opposed to permanently) to the surface of proteins via heme-binding or heme-regulatory (HBMs/HRMs) motifs, which were essentially short stretches of amino acids containing a heme coordinating amino acid (C, H or Y). This review is an account of current knowledge on how to analyze sequence and structural recognition requirements for transient heme binding to proteins. The work discusses the use of a combinatorial peptide library screening method that enabled the establishment of sequence criteria for transient heme binding. A classification of heme-binding motifs based on the specific axial ligand is presented along with an in-depth analysis of heme-binding motifs using various spectroscopic methods is shown. Finally, building up to the fact that, there exists several yet to be discovered heme-regulated proteins, the work introduces a novel computational solution in the form of a program called *SeqD-HBM* (sequence based detection of transient heme-binding motifs), to enable rapid and efficient detection of HBMs from protein sequence datasets in order to explore how heme's regulatory impact affects the protein's function.

***Own contribution**

I performed all the computational investigation including the algorithmic design, Python implementation and testing of the program *SeqD-HBM*. The manuscript was written through the contribution of all authors involved.

Bioscience Reports (2019) **39** BSR20181940
https://doi.org/10.1042/BSR20181940

PORTLAND PRESS

Review Article

The molecular basis of transient heme-protein interactions: analysis, concept and implementation

Amelie Wißbrock, Ajay Abisheck Paul George, Hans Henning Brewitz, Toni Kühl and Diana Imhof

Pharmaceutical Biochemistry and Bioanalytics, Pharmaceutical Institute, University of Bonn, An der Immenburg 4, Bonn, Germany

Correspondence: Diana Imhof (dimhof@uni-bonn.de)

Deviant levels of available heme and related molecules can result from pathological situations such as impaired heme biosynthesis or increased hemolysis as a consequence of vascular trauma or bacterial infections. Heme-related biological processes are affected by these situations, and it is essential to fully understand the underlying mechanisms. While heme has long been known as an important prosthetic group of various proteins, its function as a regulatory and signaling molecule is poorly understood. Diseases such as porphyria are caused by impaired heme metabolism, and heme itself might be used as a drug in order to downregulate its own biosynthesis. In addition, heme-driven side effects and symptoms emerging from heme-related pathological conditions are not fully comprehended and thus impede adequate medical treatment. Several heme-regulated proteins have been identified in the past decades, however, the molecular basis of transient heme-protein interactions remains to be explored. Herein, we summarize the results of an in-depth analysis of heme binding to proteins, which revealed specific binding modes and affinities depending on the amino acid sequence. Evaluating the binding behavior of a plethora of heme-peptide complexes resulted in the implementation of a prediction tool (SeqD-HBM) for heme-binding motifs, which eventually led and will perspectively lead to the identification and verification of so far unknown heme-regulated proteins. This systematic approach resulted in a broader picture of the alternative functions of heme as a regulator of proteins. However, knowledge on heme regulation of proteins is still a bottomless barrel that leaves much scope for future research and development.

Introduction

Heme is a valued, versatile, and vital molecule [1–3]. As a prosthetic group of hemoglobin, heme was initially described by Fritz Ludwig Hünefeld in the 1840s [4]. Two Nobel Prizes awarded to Hans Fischer for the synthesis of hemin in 1930, and to Max Perutz and John Kendrew in 1962, who explored the structure of hemoglobin and myoglobin, honored the eminent role of the molecule which was already recognized in the early 19th century (*cf.* www.Nobelprize.org). However, even though the nature of heme and related physiological processes had been investigated for many decades, another substantial function was only identified in the 1990s: Heme may act as an effector and signaling molecule [5]. Part of this function includes transient heme-protein interactions as found for the human Aminolevulinic acid synthase (ALAS) by Lathrop and Timko in 1993 [6]. Lathrop and Timko are the eponyms of the nowadays broadly used term 'heme-regulatory motif' (HRM) that originally described a distinct conserved motif involved in heme-mediated regulation of ALAS [6]. Over the years, the term HRM was refined and meanwhile describes a short amino acid sequence that includes a heme-coordination site and is located on the protein surface [7]. Heme binding to

Received: 17 October 2018
Revised: 18 December 2018
Accepted: 02 January 2019

Accepted Manuscript Online: 08 January 2019
Version of Record published: 30 January 2019

Bioscience Reports (2019) **39** BSR20181940
https://doi.org/10.1042/BSR20181940

such motifs may alter protein stability and/or function or it can result in the formation of a catalytically active heme-peptide/protein complex [5,8]. Motifs including a cysteine-proline (CP) dipeptide are specified by the term 'CP motif' [7]. The latter one is the most prominent representative amongst HRMs [7,9–14]. Today, after more than two decades of research, CP motifs are still the best explored HRMs, nevertheless, there is no doubt about HRMs occurring in much more versatile ways considering also other coordinating amino acids such as histidine- and tyrosine-based motifs. If no functional impact occurs upon heme binding to a protein, the term 'heme-binding motif' (HBM) is used to describe a protein sequence stretch that interacts with heme. An intriguing question that arises at this point is, what are the specific requirements for transient heme-protein interactions?

Given the chemical nature of heme, one can easily imagine that a heme-protein interaction occurs at various molecular levels. A transient interaction requires fast heme association and disassociation in order to allow for a situation-dependent response. For regulatory heme, a coordinative bond of the central iron ion to a heteroatom-containing amino acid side chain is observed in the first place. The most prominent heme-coordinating residues are cysteine, histidine, and tyrosine, while methionine and lysine are less frequently found [10]. In addition, hydrophobic interactions and π-π stacking conveyed by the porphyrin ring system as well as electrostatic interactions and hydrogen bonding via the propionate side chains contribute to heme binding and influence binding mode and affinity [10,15]. Therefore, not only the coordinative residue but also the surrounding amino acids are responsible for the decision if and how heme interacts with a specific protein [10,15]. Following the aforementioned initial studies, several heme-regulated proteins have been identified in the last 25 years [5,8]. These heme-regulated proteins take part in diverse biological processes including transcription and translation (e.g. DGCR8 [16], Rev-Erbβ [17]), ion channel modulation (e.g. hSlo1 [18]), circadian rhythm (Per2 [19]), and cell-cycle regulation (p53 [20]) [5,8]. Moreover, the function of various extracellular proteins such as complement factors (e.g. C1q [21], C3 [22]), and coagulation factors (e.g. FVIII [23]) is altered by transient heme binding. Heme-binding to eminent medical targets as, for instance, the cystathionine-β-synthase [24] or the amyloid β (Aβ) peptide [25], known for its crucial role in Alzheimer's disease, promoted further interest in potential (patho-)physiological implications of transient heme-protein interplays. Heme interaction with proteins may be of particular interest in the case of amplified hemolysis, for example, as a consequence of vascular injury or action of bacterial hemolysins, resulting in ominously augmented concentrations of biologically available heme. Extending the observations of a heme-mediated change of a protein's activity and/or stability, a peroxidase-like activity [25] was shown for the Aβ-heme complex raising the question of heme-mediated physiological responses depending on the cellular milieu in specific events, e.g. oxidative stress. On the other hand, heme-mediated protein regulation can be used for therapy as in the case of ALAS mentioned above [26]. In an acute attack, porphyria patients may receive heme preparations that are thought to downregulate ALAS activity in the liver and thereby decrease toxic heme intermediates [26]. Due to the fundamental role of heme, as an oxygen-binding molecule in e.g. hemoglobin and its omnipresence in the blood, it is inevitably necessary to understand basic heme-associated processes as well as heme-mediated protein regulation in order to take appropriate measures as required in the case of risk-bearing pathophysiological conditions.

Even though the awareness of alternative roles of heme increased over the years, a systematic approach to explore the molecular basis of transient heme-protein interactions was missing until 2010, when we started the search for specific interaction patterns, binding motifs, and structural insights concerning the transient binding of heme to peptides and proteins.

Sequence criteria for heme binding identified by a combinatorial peptide library screening

Since short protein-derived sequences (∼9 amino acids) were shown to be suitable to study heme-binding behavior [9,12,27,28], our initial studies included the construction of a combinatorial nonapeptide library based on histidine, tyrosine, and cysteine as heme axial ligand (at position P^0) as these are the most striking heme-coordinating residues (Figure 1A) [29]. Besides the coordinating amino acid all positions were randomized, and the library was constructed as $X_4(C/H/Y)^0X_4$ (X: all amino acids except Cys and Met, but including Nle). While methionine was required for technical reasons (special peptide elimination procedure) and thus could not be used at the randomized positions [30], additional cysteines were excluded to avoid intramolecular disulfide formation. After library synthesis, the peptide-bound resin beads were incubated with varying concentrations of heme (0.01 to 100 nM) [29]. Upon incubation a yellow-green color occurred in the case of heme binding and allowed to manually pick the respective beads using a stereomicroscope (Figure 1A).

Sequence elucidation by PED-MALDI-TOF mass spectrometry [30] as well as on-bead automated Edman degradation revealed distinct sequence characteristics [29]. Evaluation of the obtained peptides revealed a predominance

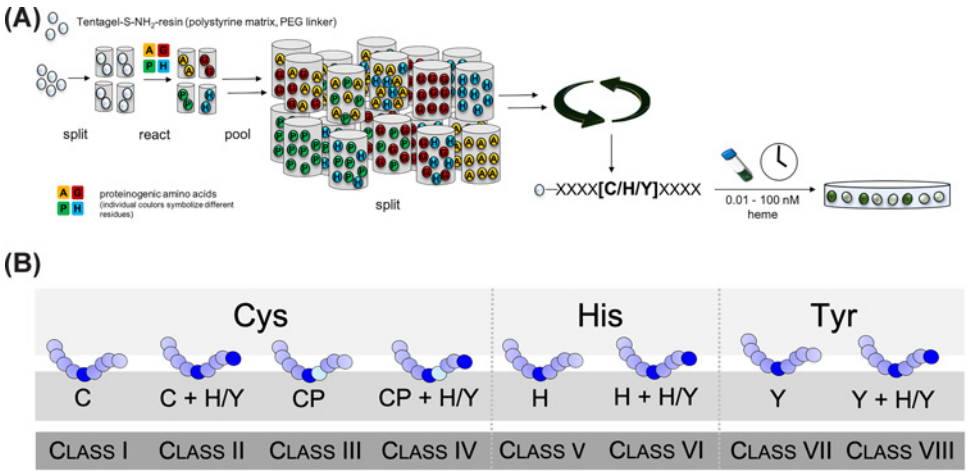


Figure 1. Investigation and classification of heme-binding sequences by means of a combinatorial peptide library

(**A**) A combinatorial peptide library $X_4(C/H/Y)^0X_4$ was constructed in order to investigate sequence specificities of heme-binding peptides. As a result of the screening, a classification system for HRMs/HBMs (**B**) was compiled for cysteine-, histidine-, and tyrosine-based motifs including extra classes for CP motifs [13,29,33,34].

of histidine and tyrosine residues (∼40% each) as heme axial ligands over cysteines (∼20%). In addition, corresponding sequence specificities at the termini were identified for all peptides: primarily polar residues as e.g. E, D, Q, N, R, K, H, Y emerged and also, to a lesser extent, hydrophobic amino acids like L, V, F, and Y. Such residues facilitate the interaction with the functional groups of the porphyrin ring. Interestingly, the appearance of additional coordination sites (His/Tyr) was observed in more than 50% of the peptide hits.

The identified hit sequences were analyzed to derive consensus sequences for the different classes of HRMs. To identify potential heme-regulated proteins, a database screening with the consensus sequences was subsequently performed by means of the ScanProsite tool [31] (ExPASy Proteomics server). Several search runs revealed potential heme-binding bacterial as well as human proteins suggesting that the underlying molecular concept is evolutionarily conserved [5,8,32]. Moreover, already published HRMs were evaluated for similarities and sequence characteristics, too. Based on these findings, further fine-tuned consensus sequences were derived and screened as described before. Among the proposed heme-binding sequences, the human dipeptidyl-peptidase 8 (DPP8) was identified as an interesting target protein that was later shown to be regulated by heme [13,29].

Classifying heme-binding motifs

The findings of the peptide library screening inspired us to develop a classification system for HBMs according to the axial ligand (H, Y, C) (Figure 1B) [33,34]. The observation that more than 50% of the library-derived heme-binding peptides exhibited ancillary coordination sites in close proximity [29], justified a further division of the three classes considering the presence or absence of additional potential coordination sites beyond position P^0. Thus, a total number of eight classes of HBMs was established (Figure 1B). To allow for profound analysis, suitable peptide representatives of each class as well as proteins carrying the corresponding motif were selected to study their heme-binding behavior with the help of sophisticated spectroscopic methods [13,29,33,34]. The subsequent investigations were based on peptide sequences as model compounds bearing in mind that the gained knowledge was to be transferred to the protein level in following studies.

In-depth analysis of heme-binding motifs by different spectroscopic methods

The spectroscopic methods applied within our studies complemented each other and allowed to draw a comprehensive picture of the heme-peptide interaction taking place (Figure 2). The fact that relatively large substance quantities are required is common to all of these methods, restricting the practical applicability for molecules with limited availability (e.g. proteins). In the case of nonapeptides, the amount available is usually not the limiting factor, but physicochemical properties affecting solubility, for example, may prevent the implementation of the individual method. Hereafter, a short, simplified insight into the individual spectroscopic methods is given (Figure 2).

Bioscience Reports (2019) **39** BSR20181940
https://doi.org/10.1042/BSR20181940

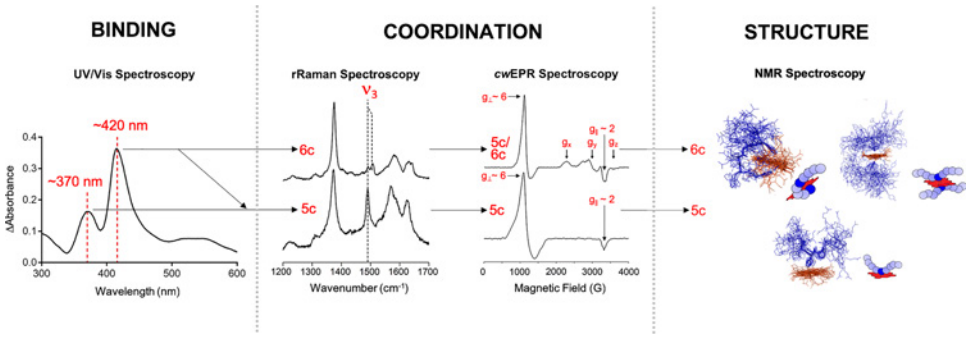


Figure 2. Investigating heme binding to peptides or proteins by various spectroscopic methods

Complex formation can be detected by UV/Vis spectroscopy, in particular by a shift of the heme-characteristic Soret-band. Upon binding the heme-iron coordination state, i.e. the occurrence of a penta- and/or hexa-coordinated complex, can be investigated by rRaman and *cw*EPR spectroscopy, focusing amongst others on the ν_3 band (rRaman) and the signals around $g \sim 6$ and $g \sim 2$ (*cw*EPR). The topology and structure of the formed complexes can be clarified by applying 2D- or 3D-NMR spectroscopy [32,33].

Due to the number of conjugated double bonds found in the porphyrin scaffold, heme shows a characteristic absorbance spectrum [35,36]. Especially worth mentioning is the B-band, called Soret-band after its discoverer Jacques-Louis Soret [37]. This band is visible at ∼400 nm. Besides the Soret-band, there are less pronounced Q-bands in the range of ∼450-700 nm [38]. The exact position of the bands depends on the oxidation and spin state of the heme-iron ion as well as the immediate surrounding of the heme molecule [39,40]. As described in the next section, a shift of the Soret-band to ∼370 nm seems to correlate primarily with a penta-coordinated complex, while a shift to ∼420 nm is found for penta- and hexa-coordinated complexes [33,34]. Therefore, the formation of heme-peptide/protein complexes is detectable in the UV/Vis spectrum, in particular with regard to a shift of the Soret-band as is commonly observed for heme binding to amino acid sequences. Moreover, titrations with different concentrations of either interaction partner enable the determination of binding constants such as the dissociation constant K_D.

Once a heme-peptide/protein interaction is confirmed by UV/Vis spectroscopy, the heme-iron coordination state (complex geometry), i.e. a penta- or hexa-coordinated iron ion, is of great interest. The iron ion is bound to four nitrogen atoms of the planar porphyrin ring system while there are two open positions remaining that allow for one or two additional coordinative bonds. These positions can be occupied by ligands possessing sulfur, oxygen or nitrogen atoms such as trifunctional amino acids. Coordination to these ligands will result in the formation of a penta- or hexa-coordinated heme-complex. It is worth mentioning that in hemoproteins the sixth ligand can be a solvent or gas molecule such as water, oxygen, carbon monoxide or nitrogen monoxide, which is usually classified as a penta-coordination with respect to the protein ligand [41]. Methods that allow to determine the coordination state of the heme iron are *resonance* Raman (rRaman) spectroscopy and *continuous wave* electron spin resonance (*cw*EPR) spectroscopy (Figure 2) [42,43]. rRaman spectroscopy is based on laser-induced vibrations (370-430 nm) characteristic for the heme molecule [42,44]. The so-called ν_3 band is of special interest regarding the iron coordination state, since it shifts according to the coordination occurring during complex formation [44]. The ν_3 band emerges around ∼1491 cm^{-1} in the case of hemin only (chloride acts as ligand) and a penta-coordinated heme-peptide/protein complex, while in case of a hexa-coordinated complex, the band appears around 1505 cm^{-1} (Figure 2) [44]. Moreover, mixtures of penta- and hexa-coordinated complexes can be detected as a double band [42]. Additional bands such as the ν_7 band can give further insight into the complex geometry present [45].

*cw*EPR spectroscopy is based on the spin state of the iron ion [43]. Simplified, free hemin as well as a penta-coordinated heme complex exhibit a high-spin state (S = 5/2), while a hexa-coordinated complex results in a low-spin state (S = 6/2) [43]. The respective signals appear at $g \sim 6$ and $g_{||} \sim 2$ in the case of penta-coordination, whereas a hexa-coordination leads to the occurrence of three signals with g-values (g_x, g_y, g_z) in the range of $g \sim 1.5$ to $g \sim 3$ (Figure 2) [34,43]. In-depth information on the structure and topology of heme-peptide/protein complexes can be obtained by applying 2D/3D-NMR spectroscopy (Figure 2) [13]. Changes upon heme complexation are identified by comparing the complex structure to the unbound peptide/protein structure, i.e. evaluation of the chemical shifts of residues before and after heme incubation is necessary [13,46]. NMR spectroscopy of longer sequences is extremely time consuming and challenging primarily due to high structural flexibility [47,48]. Depending on the sequence composition (^{1}H, ^{13}C) HSQC (heteronuclear single quantum coherence) spectra that are based on the natural

abundance of ^{13}C are used among other experiments. NMR structural analysis of large peptides and proteins is also possible, yet recombinant expression of ^{13}C and/or ^{15}N labeled molecules is required, which is usually achieved by adding e.g. $^{15}NH_4Cl$ and $^{13}C_6$-glucose to the respective growth media. The fact that the paramagnetic Fe^{3+} interferes with the surrounding amino acids leads to opposing effects on the intensities and to a broadening of the resonances. Therefore, in many studies other metal porphyrins, e.g. Ga^{III}-PPIX, were used instead, which are supposed to interact with peptides/proteins in a similar manner as heme [49–51].

Furthermore, it is possible to use the knowledge obtained regarding sequence requirements for transient heme-peptide/protein interactions to identify HBMs within known heme-binding/heme-regulated proteins. Therefore, we developed a computational tool that allows for the evaluation of potential motifs upon input of the protein sequence.

Sequence-dependent characteristics of heme-binding motifs

Since the most prominent motifs so far are the aforementioned CP-motifs, our initial studies focused on cysteine-based sequences representing HRM-classes I-IV (Figure 1) [13,33]. Li *et al.* showed that the sole occurrence of a heme-coordination site as well as the presence of a CP-motif does not necessarily result in an interaction with heme [10]. This observation was confirmed within all of our studies, since nonapeptides consisting of a CP-dipeptide surrounded by four respectively three alanine residues did not interact with heme [13,29,33,34]. These findings support the original idea that specific sequence features are required for heme binding. For the cysteine-based motifs (class I-IV) high to moderate binding affinities with K_D values ranging from 0.40 ± 0.19 μM to 6.36 ± 2.61 μM were determined by UV/Vis spectroscopy [13,29,33]. The cysteine-based sequences with additional coordinating residues (tyrosine, histidine) generally appeared to lead to higher binding affinities [13,33]. The affinities determined seem plausible because a transient interaction requires a fast and uncomplicated association and dissociation of the respective molecules and, in addition, heme is a rather small interaction partner [25,47,52]. The K_D values obtained are also supported by examples of regulatory heme binding described by other groups [52,53].

Furthermore, evaluation of the observed UV spectra revealed the occurrence of four different kinds of spectra which we categorized as UV-groups I-IV [54]. While a shift of the Soret-band to ∼370 nm primarily represented a penta-coordinated complex as was predominantly found for CP-peptides, a shift to ∼420-430 nm cannot be uniquely assigned to a distinct coordination state although spectra with both maxima frequently exhibited a mixture of penta- and hexa-coordinated complexes as revealed by rRaman and *cw*EPR spectroscopy (Figure 2) [13,33]. In contrast to the CP-peptides, for cysteine-based motifs (without proline) all forms of coordination states were observed [33].

Structural investigation of selected representatives of classes I-IV by NMR spectroscopy gave insight into the particular role of the proline residue within the CP-motif [13,33]. On the peptide level there was a clear difference between C- and CP-based peptides [33]. The proline residue seemed to reinforce a more defined backbone structure of the free peptide compared to the rather flexible structure of the cysteine-based peptide [33]. Application of heme did not lead to an increase of rigidity in the case of the cysteine-based motifs, however, CP-based peptides showed increased backbone rigidity upon heme binding, in particular in close proximity of the CP-motif. It was found that the proline residue confers a distinct conformation to the subsequent backbone, which - as a consequence - is directed away from the porphyrin ring [33]. NMR spectroscopy also revealed penta-coordinated heme complexes for CP-motifs (class III and IV) independently of the presence or absence of an additional possible heme coordination site [33]. Analysis of the cysteine-based motifs without a proline displayed hexa-coordination for both classes (I and II) (Figures 1 and 3). On the one hand, the cysteine-based peptide with no additional coordination site revealed binding of two peptide molecules to one heme molecule in a 'sandwich-like' structure (class I) (Figure 3). On the other hand, for class II (e.g. HXXXC) it was shown that a spacer length of three amino acids is required to obtain a 'loop-like' (respectively 'clamp-like') hexa-coordinated complex with one coordinating residue being the central cysteine and the other one being a distal histidine residue [33] (Figure 3).

To complete the picture of heme-binding sequences, histidine- and tyrosine-based motifs representing classes V-VIII (Figure 1) were investigated in the same manner as described above [34]. In contrast to the earlier findings for cysteine-based motifs, several sequences did not interact with heme or did not show saturation upon increasing heme concentrations, thereby hampering the determination of K_D values [34]. The K_D values determined ranged from 0.24 ± 0.17 μM to 6.25 ± 1.44 μM [34] again revealing high to moderate binding affinities as expected. In-depth analysis of selected sequences revealed that histidine-based motifs usually formed mixed or hexa-coordinated complexes with e.g. an HXXXH-motif exhibiting a loop-like structure, while tyrosine-based motifs predominantly occurred in a penta-coordinated fashion [34]. In the case of an additional coordination site in tyrosine-based peptides no loop

Bioscience Reports (2019) **39** BSR20181940
https://doi.org/10.1042/BSR20181940

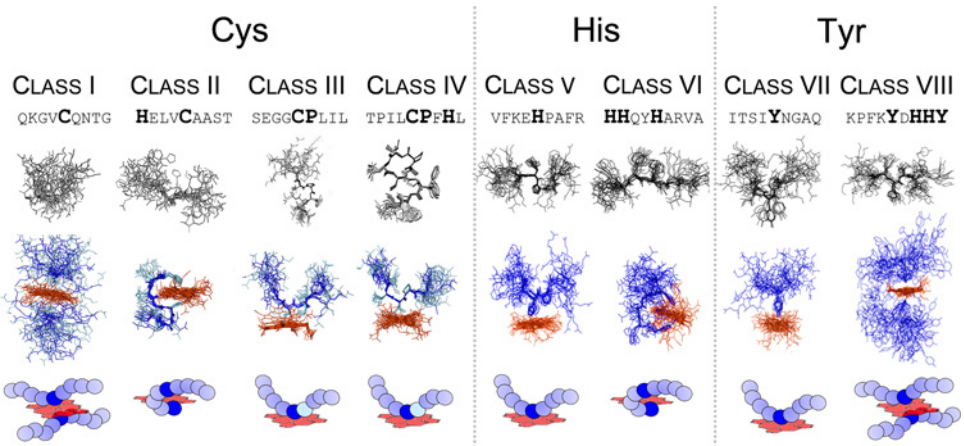


Figure 3. Structural elucidation of heme-binding peptides (classes I–VIII) using 2D-NMR spectroscopy

Different binding modes occurred depending on the peptide sequence composition and the formation of penta-coordinated (III, IV, V, and VII) and hexa-coordinated complexes (I, II, VI, and VIII). The latter ones emerged in different forms: a sandwich-like complex including two peptides interacting with one heme molecule (I and VIII) and a loop/clamp-like complex for peptides that exhibit additional coordination sites (II and VI) [32,33].

formation was found [34]. In general, no increased backbone rigidity as found for the CP-motifs was observed within these studies [34].

Comparing the amino acid composition of the heme-binding sequences revealed a crucial role of the net charge of the nonapeptides. While a negative net charge appeared to inhibit heme interaction, a positive net charge was usually accompanied by a comparably high binding affinity [34]. All the information gained from UV/Vis, rRaman, *cw*EPR, and 2D-NMR spectroscopy revealed insight into the specific characteristics of heme binding to peptides/proteins on the level of primary sequences and secondary structures. In the presented study, peptides served as model system to examine a broad range of primary sequence motifs of heme-binding peptides and proteins. It is worth noting that various other studies have applied peptide-based approaches to investigate heme binding and hemoproteins. These studies addressed functional, structural, stability, and specificity issues of the respective heme complexes [28]. The sequences and secondary structural elements of the examined peptides vary broadly. Whereas some studies use protein-derived sequences, others utilize specifically designed peptides which exhibit desired functional and structural properties. In contrast to the study using nonapeptides, specially designed peptides are often characterized by secondary structures that facilitate distinct functions, e.g. intended heme binding [28,55–59]. Among these are heme-binding multi-stranded β-sheet peptides [55], β-hairpin conformation [56], heme-Cage β-Sheet miniproteins [57], as well as helical sequence stretches [59]. Besides basic research these heme-binding peptides and miniproteins are intended for industrial and biomedical applications [58].

Subsequently the established consensus sequences mentioned above were screened against the protein data base ScanProsite tool [31] (ExPASy Proteomics server) aiming to identify so far unknown heme-regulated proteins. The hits obtained were further assessed taking into account the accessibility of the suggested HBM/HRM for heme binding, the protein structure if available, and the possibility to experimentally test the impact of heme on the protein activity. Several potential heme-regulated proteins were identified using this approach. Heme binding to these proteins was verified by spectroscopic methods and the functional impact of heme was shown *in vitro* [13,34,54]. This approach was successful for bacterial proteins such as FeoB [54], chloramphenicol-acetyltransferase (Cat) [32], and hemolysin C [60] as well as human proteins as the aforementioned dipeptidylpeptidase 8 [13]. To summarize the knowledge obtained regarding distinct sequence features of HBMs/HRMs, a procedure that enables evaluation of a protein sequence to comprise HBMs in a stepwise manner was generated (see below). A first assessment of the sequence can be drawn on the basis of the primary sequence. Additional experiments such as UV/Vis spectroscopy will then facilitate pre-evaluation of the heme-binding mode based on general structural features of the complex formed (pre-selection). In order to facilitate the HBM/HRM evaluation process for other users, we recently developed an algorithm termed *SeqD-HBM*. The basics of the *SeqD-HBM* are explained below.

Evaluating heme-binding capacity of protein sequences using *SeqD-HBM*

A handful of computational tools are available to predict ligand binding to proteins. Some of these have a broader scope of being able to predict the binding of multiple ligands to proteins e.g. *TargetS* [61], whereas other tools are specific for the prediction of heme binding (e.g. *HemeBIND+* [62] and *HemeBIND* [63]). While these tools use complex algorithms and base their predictions on extracting several novel features (e.g. Depth index, Protrusion index, Surface complementarity etc.), the nature of binding usually described is the strong irreversible binding of heme (mostly within a binding pocket) rather than a transient interaction. Moreover, most of these tools rely heavily on the availability of structural information as a basis for their predictions [63]. With this in mind, we introduce a novel tool named "*SeqD-HBM"* for sequence-based identification of HBM using the protein sequence as its primary input (Supplementary Material). The program processes the input sequence through a systematic stepwise validation extracting at each step relevant features from the sequence to produce a tabulated list of the possible heme-coordination sites available for the given sequence. Besides reporting the potential heme-coordination sites (Cys, His, Tyr), the program outputs the 9mer motifs associated with the coordination site and the net charge. Finally, in a column named "comment" useful hints regarding the predicted 9mer motif, such as identification of a CP motif, are provided to the user. This comprehensive output further guides the user to make informed judgments on the nature of heme interaction with the protein of interest.

SeqD-HBM in its current stand-alone form has two distinct modes of operation. The *default mode* assumes that the user has no information on the structure related to the input sequence. This consequently means that there is no possibility to determine if a predicted coordination site is "surface-exposed" or "buried". Prediction of a buried residue as a potential coordination site would be a false positive, defeating the purpose of this tool. To overcome this roadblock, we pass the input sequence through a sequence-based solvent accessibility meta-predictor namely *WESA* (Weighted Ensemble Solvent Accessibility) [64,65]. *WESA* determines the solvent accessibility of each residue in a sequence using an ensemble of five methods: Bayesian statistics, multiple linear regression, decision tree, neural network, and support vector machine. A weighted sum of individual predictions determines the final prediction. This deems *WESA* to be a robust and reliable tool to distinguish between the buried and exposed states of residues and has a published accuracy of 80%. *WESA* is invoked from within *SeqD-HBM* and only those coordination sites that are predicted to be "exposed" are considered for the final tabulated output of the *SeqD-HBM* program.

The second mode of operation called the *structure mode* assumes that the user is aware of the structure or is in possession of the structure data of a sequence that is passed as input to *SeqD-HBM*. In this case, *SeqD-HBM* does not invoke WESA and the HBM validation checks are done on every possible coordination site available in the sequence. The user is expected to use the known structural information to manually filter out false positives (i.e. ignore a coordination site prediction if it is known from the structure that the site is a buried residue) from the *SeqD-HBM* prediction. The operation of the application is presented as a flowchart in Figure 4. The implementation and testing details of *SeqD-HBM* are discussed in the Supplementary Material.

Conclusion

To gain a deeper insight into the molecular basis of transient heme binding, nonapeptides were used as models in order to allow for a global investigation of heme-binding characteristics. More than 200 heme-peptide complexes based on cysteine, histidine, or tyrosine as heme axial ligand have been examined so far using UV/Vis spectroscopy and, in part, methods such as rRaman, *cw*EPR, and 2D-NMR spectroscopy. Evaluation of the data obtained revealed specific sequence features such as a positive net charge of the heme-binding sequence or the existence of hydrophobic residues which have a positive effect on the heme-peptide/protein interplay. Depending on the sequence composition, different binding modes have been observed, e.g. penta- vs. hexa-coordination or mixtures thereof. The knowledge derived from the detailed analysis may first be used to predict heme binding to proteins based on consensus sequences and respective data base searches. Prediction and verification of unknown potentially heme-binding proteins based on such a consensus sequences search has been successful in several cases, i.e. bacterial FeoB, and HlyC. Second, it may be utilized to assess the heme-binding capacity of proteins which were shown to bind heme and to identify HBMs in such sequences. To make the evaluation of heme-binding sequences available to the public, our knowledge was incorporated into the program '*SeqD-HBM*' for the determination of HBMs in proteins. The software evaluates the motifs contained in a protein on the basis of the primary sequence and, if possible, takes structural features into account. We expect that such a tool will be useful to decipher molecular details on heme-binding/regulated proteins and in this way support basic research concerning the previously mentioned heme-related pathological scenarios.

Bioscience Reports (2019) **39** BSR20181940
https://doi.org/10.1042/BSR20181940

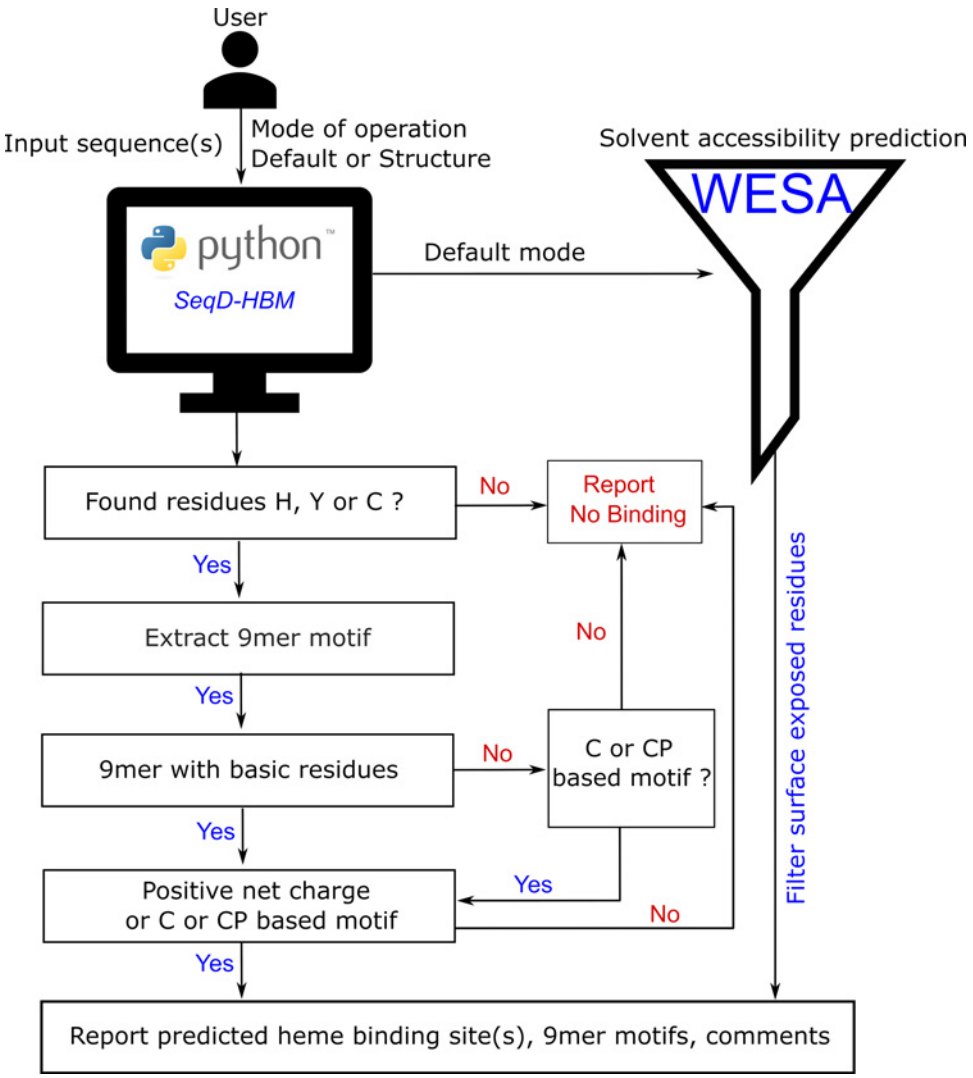


Figure 4. Computational prediction of heme-binding protein sequences

Computational evaluation of the heme-binding potential of various peptide/protein sequences based on the knowledge gained from in-depth spectroscopic studies on heme-peptide complexes.

Acknowledgements

We acknowledge Oliver Ohlenschläger (FLI Jena) for the performance of NMR structure analysis and his useful suggestions and scientific discussions. The authors also thank Jürgen Popp (IPHT Jena) and Ute Neugebauer (University Hospital Jena) for support with rRaman spectroscopy and Olaf Schiemann and Gregor Hagelüken (University of Bonn) for support with *cw*EPR spectroscopy.

Funding

This work was financially supported by the Deutsche Forschungsgemeinschaft (DFG) within [FOR1738 and SFB 813 (to D.I.)].

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Author Contribution

D.I. designed and directed the project. A.A.P.G. developed the algorithm of SeqD-HBM. All authors contributed to the manuscript.

Abbreviations

ALAS, aminolevulinic acid synthase; Aβ, amyloid β; *cw*EPR, continuous-wave electron spin resonance spectroscopy; CP, cysteine-proline; HBM, heme-binding motif; HBP, heme-binding protein; HRM, heme-regulatory motif; MALDI-TOF, matrix-assisted laser desorption/ionization time of flight; PED, partial Edman degradation; rRaman spectroscopy, resonance Raman spectroscopy; SeqD-HBM, sequence-based detection of heme binding motifs; WESA, weighted ensemble solvent accessibility.

References

1 Mense, S.M. and Zhang, L. (2006) Heme: a versatile signaling molecule controlling the activities of diverse regulators ranging from transcription factors to MAP kinases. *Cell Res.* **16**, 681–692, https://doi.org/10.1038/sj.cr.7310086
2 Severance, S. and Hamza, I. (2009) Trafficking of heme and porphyrins in metazoa. *Chem. Rev.* **109**, 4596–4616, https://doi.org/10.1021/cr9001116
3 Poulos, T.L. (2014) Heme enzyme structure and function. *Chem. Rev.* **114**, 3919–3962, https://doi.org/10.1021/cr400415k
4 Hünefeld, F.L. (1840) *Der Chemismus in der thierischen Organization*, F. A. Brockhaus, Leipzig
5 Zhang, L. (2011) *Heme Biology: The Secret Life of Heme in Regulating Diverse Biological Processes*, World Sci., Singapore; Hackensack, NJ
6 Lathrop, J.T. and Timko, M.P. (1993) Regulation by heme of mitochondrial protein transport through a conserved amino acid motif. *Science* **259**, 522–525, https://doi.org/10.1126/science.8424176
7 Shimizu, T. (2012) Binding of cysteine thiolate to the Fe(III) heme complex is critical for the function of heme sensor proteins. *J. Inorg. Biochem.* **108**, 171–177, https://doi.org/10.1016/j.jinorgbio.2011.08.018
8 Kühl, T. and Imhof, D. (2014) Regulatory FeII/III heme: the reconstruction of a molecule's biography. *Chem. Bio. Chem.* **15**, 2024–2035, https://doi.org/10.1002/cbic.201402218
9 Zhang, L. and Guarente, L. (1995) Heme binds to a short sequence that serves a regulatory function in diverse proteins. *EMBO J.* **14**, 313–320, https://doi.org/10.1002/j.1460-2075.1995.tb07005.x
10 Li, T., Bonkovsky, H.L. and Guo, J. (2011) Structural analysis of heme proteins: implications for design and prediction. *BMC Struct. Biol.* **11**, 13, https://doi.org/10.1186/1472-6807-11-13
11 Hou, S., Reynolds, M.F., Horrigan, F.T., Heinemann, S.H. and Hoshi, T. (2006) Reversible binding of heme to proteins in cellular signal transduction. *Acc. Chem. Res.*, https://doi.org/10.1021/ar040020w
12 Igarashi, J., Murase, M., Iizuka, A., Pichierri, F., Martinkova, M. and Shimizu, T. (2008) Elucidation of the heme binding site of heme-regulated eukaryotic initiation factor 2α kinase and the role of the regulatory motif in heme sensing by spectroscopic and catalytic studies of mutant proteins. *J. Biol. Chem.* **283**, 18782–18791, https://doi.org/10.1074/jbc.M801400200
13 Kühl, T., Wißbrock, A., Goradia, N., Sahoo, N., Galler, K., Neugebauer, U. et al. (2013) Analysis of Fe(III) heme binding to cysteine-containing heme-regulatory motifs in proteins. *ACS Chem. Biol.* **8**, 1785–1793, https://doi.org/10.1021/cb400317x
14 Westberg, J.A., Jiang, J. and Andersson, L.C. (2011) Stanniocalcin 1 binds hemin through a partially conserved heme regulatory motif. *Biochem. Biophys. Res. Commun.* **409**, 266–269, https://doi.org/10.1016/j.bbrc.2011.05.002
15 Schneider, S., Marles-Wright, J., Sharp, K.H. and Paoli, M. (2007) Diversity and conservation of interactions for binding heme in b-type heme proteins. *Nat. Prod. Rep.* **24**, 621–630, https://doi.org/10.1039/b604186h
16 Faller, M., Matsunaga, M., Yin, S., Loo, J.A. and Guo, F. (2007) Heme is involved in microRNA processing. *Nat. Struct. Mol. Biol.* **14**, 23–29, https://doi.org/10.1038/nsmb1182
17 Raghuram, S., Stayrook, K.R., Huang, P., Rogers, P.M., Nosie, A.K., McClure, D.B. et al. (2007) Identification of heme as the ligand for the orphan nuclear receptors REV-ERBα and REV-ERBβ. *Nat. Struct. Mol. Biol.* **14**, 1207–1213, https://doi.org/10.1038/nsmb1344
18 Tang, X.D., Xu, R., Reynolds, M.F., Garcia, M.L., Heinemann, S.H. and Hoshi, T. (2003) Haem can bind to and inhibit mammalian calcium-dependent Slo1 BK channels. *Nature* **425**, 531–535, https://doi.org/10.1038/nature02003
19 Yang, J., Kim, K.D., Lucas, A., Drahos, K.E., Santos, C.S., Mury, S.P. et al. (2008) A novel heme-regulatory motif mediates heme-dependent degradation of the circadian factor period 2. *Mol. Cell. Biol.* **28**, 4697–4711, https://doi.org/10.1128/MCB.00236-08
20 Shen, J., Sheng, X., Chang, Z., Wu, Q., Wang, S., Xuan, Z. et al. (2014) Iron metabolism regulates p53 signaling through direct Heme-p53 interaction and modulation of p53 localization, stability, and function. *Cell Rep.* **7**, 180–193, https://doi.org/10.1016/j.celrep.2014.02.042
21 Roumenina, L.T., Radanova, M., Atanasov, B.P., Popov, K.T., Kaveri, S.V., Lacroix-Desmazes, S. et al. (2011) Heme interacts with C1q and inhibits the classical complement pathway. *J. Biol. Chem.* **286**, 16459–16469, https://doi.org/10.1074/jbc.M110.206136
22 Frimat, M., Tabarin, F., Dimitrov, J.D., Poitou, C., Halbwachs-Mecarelli, L., Fremeaux-Bacchi, V. et al. (2013) Complement activation by heme as a secondary hit for atypical hemolytic uremic syndrome. *Blood* **122**, 282–292, https://doi.org/10.1182/blood-2013-03-489245
23 Repessé, Y., Dimitrov, J.D., Peyron, I., Moshai, E.F., Kiger, L., Dasgupta, S. et al. (2012) Heme binds to factor VIII and inhibits its interaction with activated factor IX. *J. Thromb. Haemost.* **10**, 1062–1071, https://doi.org/10.1111/j.1538-7836.2012.04724.x
24 Kumar, A., Wißbrock, A., Goradia, N., Bellstedt, P., Ramachandran, R., Imhof, D. et al. (2018) Heme interaction of the intrinsically disordered N-terminal peptide segment of human cystathionine-β-synthase. *Sci. Rep.* **8**, 2474
25 Atamna, H. and Boyle, K. (2006) Amyloid-beta peptide binds with heme to form a peroxidase: relationship to the cytopathologies of Alzheimer's disease. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3381–3386, https://doi.org/10.1073/pnas.0600134103
26 Balwani, M. and Desnick, R.J. (2012) The porphyrias: advances in diagnosis and treatment. *Blood.* **120** ((23)), 4496–4504, https://doi.org/10.1182/blood-2012-05-423186
27 Goodfellow, B.J., Dias, J.S., Ferreira, G.C., Henklein, P., Wray, V. and Macedo, A.L. (2001) The solution structure and heme binding of the presequence of murine 5-aminolevulinate synthase. *FEBS Lett.* **505**, 325–331, https://doi.org/10.1016/S0014-5793(01)02818-6

28 Lombardi, A., Nastri, F. and Pavone, V. (2001) Peptide-based heme–protein models. *Chem. Rev., Am Chem. Soc.* **101**, 3165–3190, https://doi.org/10.1021/cr000055j
29 Kühl, T., Sahoo, N., Nikolajski, M., Schlott, B., Heinemann, S.H. and Imhof, D. (2011) Determination of hemin-binding characteristics of proteins by a combinatorial peptide library approach. *Chem. Bio. Chem.* **12**, 2846–2855, https://doi.org/10.1002/cbic.201100556
30 Sweeney, M.C., Wavreille, A.S., Park, J., Butchar, J.P., Tridandapani, S. and Pei, D. (2005) Decoding protein-protein interactions through combinatorial chemistry: sequence specificity of SHP-1, SHP-2, and SHIP SH2 domains. *Biochemistry* **44**, 14932–14947, https://doi.org/10.1021/bi051408h
31 de Castro, E., Sigrist, C. J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E. et al. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **34**, W362–W365, https://doi.org/10.1093/nar/gkl124
32 Brewitz, H.H., Hagelueken, G. and Imhof, D. (2016) Structural and functional diversity of transient heme binding to bacterial proteins. *Biochim. Biophys. Acta - Gen. Subj.* **1861**, 683–697, https://doi.org/10.1016/j.bbagen.2016.12.021
33 Brewitz, H.H., Kühl, T., Goradia, N., Galler, K., Popp, J., Neugebauer, U. et al. (2015) Role of the chemical environment beyond the coordination site: structural insight into Fe(III) protoporphyrin binding to cysteine-based heme-regulatory protein motifs. *Chem. Bio. Chem.* **16**, 2216–2224, https://doi.org/10.1002/cbic.201500331
34 Brewitz, H.H., Goradia, N., Schubert, E., Galler, K., Kühl, T., Syllwasschya, B. et al. (2016) Heme interacts with histidine- and tyrosine-based protein motifs and inhibits enzymatic activity of chloramphenicol acetyltransferase from E. coli. *Biochim. Biophys. Acta - Gen. Subj.* **1860**, 1343–1353, https://doi.org/10.1016/j.bbagen.2016.03.027
35 Marcelli, A., Jelovica Badovinac, I., Orlic, N., Salvi, P.R. and Gellini, C. (2013) Excited-state absorption and ultrafast relaxation dynamics of protoporphyrin IX and hemin. *Photochem. Photobiol. Sci.* **12**, 348–355, https://doi.org/10.1039/C2PP25247C
36 Nienhaus, K. and Nienhaus, G.U. (2005) Probing heme protein-ligand interactions by UV/visible absorption spectroscopy. *Methods Mol. Biol.* **305**, 215–242
37 Soret, J.-L. (1883) Analyse spectrale: Sur le spectre d'absorption du sang dans la partie violette et ultra-violette. *Comptes rendus l'Académie des Sci* **97**, 1269–1270
38 Uttamlal, M. and Sheila Holmes-Smith, A. (2008) The excitation wavelength dependent fluorescence of porphyrins. *Chem. Phys. Lett.* **454**, 223–228, https://doi.org/10.1016/j.cplett.2008.02.012
39 Papadopoulos, P.G., Walter, S.A., Li, J. and Baker, G.M. (1991) Proton interactions in the resting form of cytochrome oxidase. *Biochemistry* **30**, 840–850, https://doi.org/10.1021/bi00217a038
40 Luthra, A., Denisov, I.G. and Sligar, S.G. (2011) Spectroscopic features of cytochrome P450 reaction intermediates. *Arch. Biochem. Biophys.* **507** ((1)), 26–35, https://doi.org/10.1016/j.abb.2010.12.008
41 Sono, M., Roach, M.P., Coulter, E.D. and Dawson, J.H. (1996) Heme-containing oxygenases. *Chem. Rev.* **96**, 2841–2888, https://doi.org/10.1021/cr9500500
42 Spiro, T.G. (1985) Resonance Raman spectroscopy as a probe of heme protein structure and dynamics. *Adv. Protein Chem.* **37**, 111–159, https://doi.org/10.1016/S0065-3233(08)60064-9
43 Nakamura, M., Ikeue, T., Ohgo, Y., Takahashi, M. and Takeda, M.T. (2002) Highly saddle shaped (porphyrinato)iron(III) iodide with a pure intermediate spin state. *Chem. Commun.* 1198–1199, https://doi.org/10.1039/b202768b
44 Spiro, T.G. and Burke, J.M. (1976) Protein control of porphyrin conformation. comparison of resonance raman spectra of heme proteins with mesoporphyrin IX analogs. *J. Am. Chem. Soc.* **98**, 5482–5489, https://doi.org/10.1021/ja00434a013
45 Kitagawa, T., Abe, M., Kyogoku, Y., Ogoshi, H., Watanabe, E. and Yoshida, Z. (1976) Resonance Raman spectra of metallooctaethylporphyrins. Low frequency vibrations of porphyrin and iron-axial ligand stretching modes. *J. Phys. Chem.* **80**, 1181–1186, https://doi.org/10.1021/j100552a012
46 Bagai, I., Sarangi, R., Fleischhacker, A.S., Sharma, A., Hoffman, B.M., Zuiderweg, E.R.P. et al. (2015) Spectroscopic studies reveal that the heme regulatory motifs of heme oxygenase-2 are dynamically disordered and exhibit redox-dependent interaction with heme. *Biochemistry* **54**, 2693–2708, https://doi.org/10.1021/bi501489r
47 Yin, L., Dragnea, V., Feldman, G., Hammad, L.A., Karty, J.A., Dann, C.E. et al. (2013) Redox and light control the heme-sensing activity of AppA. *MBio* **4**, e00563–13, https://doi.org/10.1128/mBio.00563-13
48 Ishikawa, H., Nakagaki, M., Bamba, A., Uchida, T., Hori, H., O'Brian, M.R. et al. (2011) Unusual heme binding in the bacterial iron response regulator protein: spectral characterization of heme binding to the heme regulatory motif. *Biochemistry* **50**, 1016–1022, https://doi.org/10.1021/bi101895r
49 Caillet-Saguy, C., Piccioli, M., Turano, P., Lukat-Rodgers, G., Wolff, N., Rodgers, K.R. et al. (2012) Role of the iron axial ligands of heme carrier HasA in heme uptake and release. *J. Biol. Chem.* **287**, 26932–26943, https://doi.org/10.1074/jbc.M112.366385
50 Caillet-Saguy, C., Piccioli, M., Turano, P., Izadi-Piuneyre, N., Delepierre, M., Bertini, I. et al. (2009) Mapping the interaction between the hemophore HasA and its outer membrane receptor HasR using CRINEPT-TROSY NMR spectroscopy. *J. Am. Chem. Soc.* **131**, 1736–1744, https://doi.org/10.1021/ja804783x
51 Moriwaki, Y., Caaveiro, J.M.M., Tanaka, Y., Tsutsumi, H., Hamachi, I. and Tsumoto, K. (2011) Molecular basis of recognition of antibacterial porphyrins by heme-transporter IsdH-NEAT3 of Staphylococcus aureus. *Biochemistry* **50**, 7311–7320, https://doi.org/10.1021/bi200493h
52 Hu, R.-G., Wang, H., Xia, Z. and Varshavsky, A. (2008) The N-end rule pathway is a sensor of heme. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 76–81, https://doi.org/10.1073/pnas.0710568105
53 Lechardeur, D., Cesselin, B., Liebl, U., Vos, M.H., Fernandez, A., Brun, C. et al. (2012) Discovery of intracellular heme-binding protein HrtR, which controls heme efflux by the conserved HrtB-HrtA transporter in Lactococcus lactis. *J. Biol. Chem.* **287**, 4752–4758, https://doi.org/10.1074/jbc.M111.297531
54 Schubert, E., Florin, N., Duthie, F., Brewitz, H.H., Kühl, T., Imhof, D. et al. (2015) Spectroscopic studies on peptides and proteins with cysteine-containing heme regulatory motifs (HRM). *J. Inorg. Biochem.* **148**, 49–56, https://doi.org/10.1016/j.jinorgbio.2015.05.008

55 D'Souza, A., Mahajan, M. and Bhattacharjya, S. (2016) Designed multi-stranded heme binding β-sheet peptides in membrane. *Chem. Sci. R. Soc. Chem.* **7**, 2563–2571
56 Nagarajan, D., Sukumaran, S., Deka, G., Krishnamurthy, K., Atreya, H.S. and Chandra, N. (2018) Design of a heme-binding peptide motif adopting a β-hairpin conformation. *J. Biol. Chem.* **293**, 9412–9422, https://doi.org/10.1074/jbc.RA118.001768
57 D'Souza, A., Wu, X., Yeow, E.K.L. and Bhattacharjya, S. (2017) Designed heme-cage β-sheet miniproteins. *Angew. Chem.* **129**, 5998–6002
58 Rai, J. (2017) Mini Heme-Proteins: designability of structure and diversity of functions. *Curr. Protein Pept. Sci.* **18**, 1132–1140, https://doi.org/10.2174/1389203718666170515144037
59 Shifman, J.M., Gibney, B.R., Sharp, R.E. and Dutton, P.L. (2000) Heme redox potential control in de novo designed four-α-helix bundle proteins. *Biochem. Am. Chem. Soc.* **39**, 14813–14821
60 Peherstorfer, S., Brewitz, H.H.B., Paul George, A.A., Wißbrock, A., Adam, J.M., Schmitt, L. et al. (2018) Insights into mechanism and functional consequences of heme binding to hemolysin-activating lysine acyltransferase HlyC from Escherichia coli. *Biochim. Biophys. Acta* **1862**, 1964–1972, https://doi.org/10.1016/j.bbagen.2018.06.012
61 Yu, D.J., Hu, J., Yang, J., Shen, H.B., Tang, J. and Yang, J.Y. (2013) Designing template-free predictor for targeting protein-ligand binding sites with classifier ensemble and spatial clustering. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **10**, 994–1008, https://doi.org/10.1109/TCBB.2013.104
62 Liu, R. and Hu, J. (2011) Computational prediction of heme-binding residues by exploiting residue interaction network. *PLoS ONE* **6**, e25560, https://doi.org/10.1371/journal.pone.0025560
63 Liu, R. and Hu, J. (2011) HemeBIND: a novel method for heme binding residue prediction by combining structural and sequence information. *BMC Bioinformatics* **12**, 207–219, https://doi.org/10.1186/1471-2105-12-207
64 Chen, H. and Zhou, H.X. (2005) Prediction of solvent accessibility and sites of deleterious mutations from protein sequence. *Nucleic Acids Res.* **33**, 3193–3199, https://doi.org/10.1093/nar/gki633
65 Shan, Y., Wang, G. and Zhou, H.X. (2001) Fold recognition and accurate query-template alignment by a combination of PSI-BLAST and threading. *Proteins Struct. Funct. Genet.* **42**, 23–37, https://doi.org/10.1002/1097-0134(20010101)42:1%3c23::AID-PROT40%3e3.0.CO;2-K

4.5.2 Summary

In the previous chapter a molecular docking and MD simulation derived approach was used to detect and prove the binding of heme to a predicted HBM on the protein HlyC.[234] In this chapter, the fact that several such heme-regulated proteins are yet to be discovered and the need for the establishment of molecular basis to determine transient heme-protein interaction is emphasized. Based on a consensus from several existing studies, a set of generic rules are established that sequences must satisfy in ordered to be considered as a potential HBM. These include, the requirement of a nonapeptide motif with a central C, Y or H residue, containing basic residues in its neighborhood resulting in a motif with an overall net positive charge. C- and CP-based motifs are assigned special privileges, again based on consensus derived from existing studies and results. Till now, the heme-binding potential of any given protein could only be analyzed by experimental methods on a case-by-case basis, (i.e., individual proteins had to be exclusively subjected to experimental studies) and the current study produced a computational solution that changes this situation. The program *SeqD-HBM* (sequence based detection of transient heme-binding motifs) written in Python takes in large datasets of sequences in the FASTA[300] format for the rapid and efficient identification of heme binding motifs. The program despite using only sequence data, tries to discriminate 'solvent exposed' sites from 'buried' ones by automatically passing them to an external, machine learning based solvent accessibility predictor WESA (weighted ensemble solvent accessibility).[301] The final output provided to the user indicates the potential heme binding residues (if present), the associated nonapeptide sequence, the information on the net charge of the motif and additional information such as the possible existence of a disulfide bond in the motif. *SeqD-HBM* is touted to be used in the prediction of several novel heme-regulated proteins in the future.

5 Conclusions

This section of conclusions attempts to concisely summarize how this thesis titled "*In silico* facets of biochemical research: accounts from protein folding and protein ligand interaction studies" has served as yet another demonstration of the unquestionable significance and the synergistic usage of computational methods along with experimental practice in biochemical research. The computational component employed here is an amalgam of methods drawn from the overlapping fields of molecular modeling, computational (bio)chemistry and bioinformatics. In particular, MD simulations and molecular docking are extensively used in a variety of forms to study the folding of multiple-disulfide-bonded peptides and the mechanism and mode of interaction of these peptides with their target receptor proteins. In multiple instances, computationally generated 3D structure models are used due to the lack of experimentally resolved structures. This is on the one hand is another feather in the hat for the field of computational structure prediction but at the same time points at the blaring reality that despite the Protein Data Bank (PDB) database[302] containing currently over 157000 experimentally resolved structures (as of October 2019), in several instances structures need to be computationally predicted by either homology modeling[19] or fold recognition[303] techniques. In this final section of this thesis, a retrospective evaluation of the five individual chapters presented in the preceding sections is done with a goal of understanding the general biochemical significance of the study, the computational approach employed, the results that were produced from the computational component of the study and how this in turn added to existing knowledge in the field by either aiding in a clearer interpretation of experimental results or by producing novel results that were inaccessible to experimental investigation. Wherever possible, an outlook for possible future work is suggested.

In chapter I, μ-conotoxins namely, μ-GIIIA, μ-KIIIA, μ-PIIIA, μ-SIIIA and μ-SmIIIA are studied for their oxidative folding behavior.[304] The availability of NMR structural ensembles serves as an advantage to start with. The computational investigation primarily using MD simulations is done with a few distinct goals with large biochemical significance. The simplest question answered here is that of the minimum number of disulfide bonds required by a three-disulfide-bonded conopeptide to retain native-like structures since it is possible that native-like structures might retain to some extent the functionality of the native protein. A possibility to reduce the number of disulfide bonds in the synthesized peptide is always a welcome move from a peptide chemist's point of view. From the get go, by observing the structural stability of the 3-disulfide-bonded native folds of the five different conopeptides, it becomes evident that the underlying mechanism might be different. The RMSD profiles of μ-GIIIA, μ-KIIIA indicate that while these peptides were relatively stable in their folded states, those of μ-SIIIA and μ-SmIIIA indicate some inherent instability in their structures which seems rather uncharacteristic for a small peptide compactly folded and constrained by three disulfide bonds.[47] The case of μ-PIIIA especially falls in clear line of agreement that alternative-disulfide-bonded folds are possible for this peptide as the stability of the so-called native state seems poor. Interestingly there is experimental backing to this proposal as an earlier study identified that multiple isomers accumulate upon oxidative self-folding of this peptide.[192] This is a non-trivial inference since, it clearly points to the differences in the underlying folding mechanism of five peptides of relatively the same sequence length, overall net charge and identical disulfide connectivity. Without having to experimentally synthesize 2-disulfide-bonded versions of these peptides, by carefully comparing the RMSD and RMSF profiles between the 3- and 2-disulfide-bonded states, a favorability profile could be generated. Independent verifications of some of the results of the study by other

published experimental work add to the validity of the method, results and interpretation of the results of this work. In detail, the fact that µ-KIIIA could be relatively stable and functional with just two of its disulfide bonds and the that µ-PIIIA would need all of its disulfide bonds for stability and function were proved by independent experimental studies by other researchers.[193,195] The interpretation and assignment of the five peptides relative to existing opposite and extreme folding models of BPTI and hirudin,[54] is the first attempt by a purely computational approach doing so, and this effort has been duly recognized as an impactful addition to the field.[177] As with any MD-based studies, future work in this regard could always use longer simulations (than the 300 ns used in this study) while one could also experiment other orders of disulfide bond opening than the C2-C5, C2-C5/C3-C6 strategy used in this study.

Chapter II made use of a remarkable effort of Heimer *et al.* to synthesize and structurally resolve 12 of the 15 theoretically possible isomers of the conotoxin µ-PIIIA using NMR spectroscopy and 3 of isomers derived by computer generated means.[152] In the current study, all 15 isomers and 3 additional 2-disulfide-bonded mutants based on the native isomeric fold of µ-PIIIA (C1-C4/C2-C5/C3-C6) were subject to electrophysiological testing for their blocking activity against the VGSC Nav1.4. The computational investigation using MD simulations was once again employed to relate MD-derived measures of structural stability to explain the activity profiles of the individual isomers as observed from the electrophysiological experiments. A clear connection between structural stability induced by folded configurations, resulting from disulfide bonds crosslinking distant ends of the sequence and good blocking activity at Nav1.4 was established. For exceptional cases (µ-PIIIA 10) wherein, reasonable structural stability was observed from MD simulations but the isomer was a poor blocker of the channel, could be explained clearly by closer examination of the side-chain fluctuations of its functionally significant basic residues,[305] that were found to form intramolecular hydrogen bonds for large parts of the simulation hence taking them out of play from interacting with the channel. Besides the usual MD-derived measures of structural stability such as RMSD, RMSF and Rg, this study determined an estimate of the solvation free energy derived from the solvent accessible surface areas computed from the simulation.[306,307] The estimation of the solvation free energy provided an explanation as to why not all (or even many) of the fifteen possible isomers are formed in oxidative self-folding experiments, since they have much higher solvation free energies in comparison to for example the native isomeric fold of µ-PIIIA (C1-C4/C2-C5/C3-C6) which is formed as the major isomer in experimental investigation. The docking studies explaining the interaction of each of the isomers to the channel Nav1.4 again served as a valuable tool in understanding the mechanism and mode of binding, providing a link to the experimentally obtained bioactivity results. In essence, the isomers with good bioactivity measured in experiment were found to block the channel pore more efficiently while the ones that were regarded as inactive were observed to leave large portions of the channel pore exposed. An outlook for a possible extension of this study is to test some of the isomers (e.g., µ-PIIIA 8) that were found to have stable conformations but poor blocking activity on Nav1.4 on other voltage gated sodium or potassium channels for their blocking activity.

Chapter III is somewhat of a combination of both chapters I and II only that here, we introduce a different animal-derived, disulfide-rich bioactive peptide, namely, tridegin.[53] Tridegin is by far the only known naturally-derived peptide to be able to efficiently bind to and inhibit the human coagulation factor FXIIIa.[202] Here again, earlier studies established that, in an oxidative folding experiment, three 3-disulfide-bonded isomers of tridegin are formed in buffer solution.[203] The absence of an experimentally resolved structure in this case meant that a computational approach had to be undertaken and accordingly, structure models of the three isomers were produced as a part of earlier work using fold recognition and

threading techniques.[203] In the current study,[308] experimental results proved that all three disulfide isomers of tridegin which interestingly contain a common disulfide linkage between the cysteines 19 and 25 in the sequence, retained bioactivity comparable to their 3-disulfide-bonded parental counterparts even when this particular disulfide bond was omitted. The cysteines 19 and 25 were replaced by serine residues in this case. The key role of the computational component of this study was to establish a molecular and structural basis of how and why the removal of the C19-C25 disulfide bond had little negative effect on the bioactivity of the peptides as opposed to previous instances with conotoxins that displayed loss of bioactivity upon disulfide bond number reduction.[294] MD simulation based conformational analysis established that despite the removal of the C19-C25 disulfide bond, the structure of the mutated isomers did not experience significant distortion in solution as observed from the simulations. All three structures maintained comparable conformations to their three-disulfide bonded counterparts providing a clear structural basis to the observed bioactivity. Additionally, in this study, multiple conformations of the isomers from the MD simulations were used as ligands in the molecular docking experiments all of which bound directly at over the active site of their receptor FXIIIa. Hereby, yet again computational methods in the form of molecular simulations augmented experimental findings by providing both quantitative and qualitative explanations to the experimentally determined findings.

The final two chapters comprising part 2 of the thesis, move on from using animal-derived peptides as models to investigating the role of heme as a regulatory molecule. In chapter IV,[234] homology models of the structures of wild type and mutant versions of the protein hemolysin-activating lysine acyltransferase (HlyC) are used to study the significance of the proposed nonapeptide heme-binding motif (HBM) for its role in heme binding and its functional consequences. Once again, molecular docking combined with MD simulations clearly distinguish the impact of the mutation of the heme-coordinating histidine residue (H151) to alanine. Additional binding sites found on the surface of HlyC were also evaluated and based on MD simulations of the heme-protein complexes obtained, an explanation on the qualitative nature of the binding and could be established. In the final chapter V,[209] which is indeed a review of the often underrepresented study of transient heme interactions to proteins, the computational contribution to the study takes a different turn. In this case, the potential of experimental data and observations forming a strong conceptual basis, that can be converted into usable software is demonstrated. Based on the accumulated knowledge on the sequence requirements for transient heme binding to protein surfaces, a computational solution *SeqD-HBM* acronymized for the Sequence-based Detection of transient Heme-Binding Motifs is developed. The program implemented in Python, takes a sequence or multiple sequences from files as input, performs a set of validations and lets the user know of the different transient heme binding residues (if present), the associated nonapeptide HBM and additional useful information such as the possibility of the occurrence of a disulfide bond in the motif. This in turn lets the user make informed decisions on the proteins they wish to analyze for its heme binding potential.

Though only the final chapter provides a standalone computational solution as an outcome, there was not a single instance in the study where existing software could be used in an “out-of-the-box” fashion. A solid understanding of molecular mechanics, molecular modeling and drug design and what simulations can and cannot do is required to produce meaningful results. Most importantly, at all points of time, the ‘big picture’ of the biochemical process had to be borne in mind to avoid falling into the trap of trusting false positive results which at times can be most of the results generated. In all of the docking simulations conducted, either manual examination or custom scripts had to be developed to analyze the huge set of ligand-protein complexes that the programs usually generate on each run. For example, while studying the binding of heme to the surface of HlyC,[234] docking simulations produced

up to 75 different heme-protein complexes as ‘good hits’ but careful manual examination produced only one good hit. In studies that involved conotoxins, MD force field parameters for the non-standard amino acid pyroglutamic acid had to be derived separately. In the case of heme binding studies, care had to be taken to have the non-bonded interactions in heme molecules parameterized correctly since existing molecular mechanics force fields are not built to model the coordination interaction between the Fe^{3+}/Fe^{2+} ion of the heme and the protein.

In conclusion, this thesis work serves as a testament to the impact and significance that computational methods hold in modern biochemical research. Computational methods are an indispensable component of the modern drug discovery process.[309] As one observes the obvious trend of exponential improvements in hardware and software capabilities, especially with the ongoing advancements in AI and machine learning[73,74] in biochemical research, it is very reasonable to conclude that the future of the field is bright. This is just not speculation but is proven by the fact that the Nobel prize in 2013 was shared by three scientists Martin Karplus, Michael Levitt and Arieh Warshel “for the development of multiscale models for complex chemical systems” (https://www.nobelprize.org/prizes/chemistry/2013/press-release/). It is therefore only fitting to end this thesis with a snippet from of the press release of the Nobel committee that honored these three great scientists. A snippet that should encourage every computational chemist, molecular modeler or bioinformatician to broaden their shoulder and move ahead with confidence.

“Chemists used to create models of molecules using plastic balls and sticks. Today, the modelling is carried out in computers. Computer models mirroring real life have become crucial for most advances made in chemistry today. Today the computer is just as important a tool for chemists as the test tube. Simulations are so realistic that they predict the outcome of traditional experiments.”

-Official press release from the Nobel committee, 2013.

Abbreviations

AI	Artificial intelligence
ACI	Ascaris carboxypeptidase inhibitor
ADC	Antibody drug conjugate
AIDS	Acquired immunodeficiency syndrome
ALAS	Aminolevulinic acid synthase
BPTI	Bovine pancreatic trypsin inhibitor
CQA	Critical quality attribute
EFG	Epidermal growth factor
FDA	Food and Drug Administration
FXIIIA	Activated factor XIII
GdmCl	Guanidinium chloride
HBM	Heme binding motif
HIV	Human immunodeficiency virus
HlyA	Exotoxin hemolysin A
HlyC	Hemoylsin-activating lysine acyltransferase
HPLC	High performance liquid chromatography
HRM	Heme regulatory motif
IGF-I	Insulin-like growth factor-I
IFN-α	Interferon-α
LCI	Leech carboxypeptidase inhibitor
LDTI	Leech derived tryptase inhibitor
MD	Molecular dynamics
PCI	Potato carboxypeptidase inhibitor
PDB	Protein data bank
NGS	Next-generation sequencing
NP	Non-deterministic polynomial
Rg	Radius of gyration
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RNase A	Bovine pancreatic ribonuclease A

SBBD	Structure-based drug discovery
SWOT	Strength, Weakness, Opportunity and Threat
SLPI	Secretory leucocyte protease inhibitor
TAP	Tick anticoagulant peptide
TCI	Tick carboxypeptidase inhibitor
UPEC	Uropathogenic *Escherichia coli*
UTI	Urinary tract infection
VGSC	Voltage-gated sodium channel
VMD	Visual molecular dynamics

List of Figures

Bibliography

1. Hagen, J. B. The origins of bioinformatics. *Nat. Rev. Genet.* **1**, 231–236 (2000).
2. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).
3. Levy, S. E. & Myers, R. M. Advancements in Next-Generation Sequencing. *Annu. Rev. Genomics Hum. Genet.* **17**, 95–115 (2016).
4. Mardis, E. R. Next-Generation Sequencing Platforms. *Annu. Rev. Anal. Chem.* **6**, 287–303 (2013).
5. Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). *http://www.genome.gov/sequencingcosts.* (2016). Available at: www.genome.gov/sequencingcostsdata/. (Accessed: 3rd August 2019)
6. Howard, K. The bioinformatics gold rush. *Sci. Am.* **283**, 58–63 (2000).
7. Abdurakhmonov, I. Y. Bioinformatics: Basics, Development, and Future. in *Bioinformatics - Updated Features and Applications* (InTech, 2016). doi:10.5772/63817
8. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science (80-.).* **181**, 223–230 (1973).
9. Dill, K. A. & MacCallum, J. L. The protein-folding problem, 50 years on. *Science (80-.).* **338**, 1042–1046 (2012).
10. Bai, X. chen, McMullan, G. & Scheres, S. H. W. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **40**, 49–57 (2015).
11. Nogales, E. The development of cryo-EM into a mainstream structural biology technique. *Nat. Methods* **13**, 24–27 (2015).
12. Karaca, E. & Bonvin, A. M. J. J. Advances in integrative modeling of biomolecular complexes. *Methods* **59**, 372–381 (2013).
13. Stone, J. E. *et al.* Accelerating molecular modeling applications with graphics processors. *J. Comput. Chem.* **28**, 2618–2640 (2007).
14. Stone, J. E., Hardy, D. J., Ufimtsev, I. S. & Schulten, K. GPU-accelerated molecular modeling coming of age. *J. Mol. Graph. Model.* **29**, 116–125 (2010).
15. Al-Lazikani, B., Hill, E. E. & Morea, V. Protein structure prediction. *Methods Mol. Biol.* **453**, 33–85 (2008).
16. Karplus, M. & Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. USA.* **102**, 6679–6685 (2005).
17. Hillisch, A., Pineda, L. F. & Hilgenfeld, R. Utility of homology models in the drug discovery process. *Drug Discov. Today* **9**, 659–669 (2004).
18. Young, A. R. J., Narita, M. & Narita, M. Homology Modeling. *Methods Mol. Biol.* **857**, 1–13 (2012).
19. Cavasotto, C. N. & Phatak, S. S. Homology modeling in drug discovery: current trends and applications. *Drug Discov. Today* **14**, 676–683 (2009).

20. Peng, J. & Xu, J. Low-homology protein threading. *Bioinformatics* **26**, (2010).

21. Peng, J. & Xu, J. A multiple-template approach to protein threading. *Proteins Struct. Funct. Bioinforma.* **79**, 1930–1939 (2011).

22. Wu, S. & Zhang, Y. LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Res.* **35**, 3375–3382 (2007).

23. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).

24. Morris, G. M. & Lim-Wilby, M. Molecular docking. *Methods Mol. Biol.* **443**, 365–382 (2008).

25. Meng, X.-Y., Zhang, H.-X., Mezei, M. & Cui, M. Molecular docking: a powerful approach for structure-based drug discovery. *Curr. Comput. Aided. Drug Des.* **7**, 146–57 (2011).

26. Shoichet, B. K., McGovern, S. L., Wei, B. & Irwin, J. J. Lead discovery using molecular docking. *Curr. Opin. Chem. Biol.* **6**, 439–446 (2002).

27. Lengauer, T. & Rarey, M. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.* **6**, 402–406 (1996).

28. Ganesan, A., Coote, M. L. & Barakat, K. Molecular dynamics-driven drug discovery: leaping forward with confidence. *Drug Discov. Today* **22**, 249–269 (2017).

29. Gelpi, J., Hospital, A., Goñi, R. & Orozco, M. Molecular dynamics simulations: advances and applications. *Adv. Appl. Bioinforma. Chem.* 37 (2015). doi:10.2147/aabc.s70333

30. Onuchic, J. N. & Wolynes, P. G. Theory of protein folding. *Current Opinion in Structural Biology* **14**, 70–75 (2004).

31. Baker, D. A surprising simplicity to protein folding. *Nature* **405**, 39–42 (2000).

32. Fraenkel, A. S. Complexity of protein folding. *Bull. Math. Biol.* **55**, 1199–1210 (1993).

33. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).

34. Thomas, P. J., Qu, B. H. & Pedersen, P. L. Defective protein folding as a basis of human disease. *Trends Biochem. Sci.* **20**, 456–459 (1995).

35. Valastyan, J. S. & Lindquist, S. Mechanisms of protein-folding diseases at a glance. *Dis. Model. Mech.* **7**, 9–14 (2014).

36. Dobson, C. M. Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24**, 329–332 (1999).

37. Rath, A., Tulumello, D. V. & Deber, C. M. Peptide models of membrane protein folding. *Biochemistry* **48**, 3036–3045 (2009).

38. Baum, J. & Brodsky, B. Folding of peptide models of collagen and misfolding in disease. *Curr. Opin. Struct. Biol.* **9**, 122–128 (1999).

39. Korhonen, H. & Pihlanto, A. Bioactive peptides: Production and functionality. *Int. Dairy J.* **16**, 945–960 (2006).

40. Park, A. J., Okhovat, J. P. & Kim, J. Antimicrobial peptides. in *Clinical and Basic Immunodermatology: Second Edition* (2017). doi:10.1007/978-3-319-29785-9_6

41. Gilad, Y., Firer, M. & Gellerman, G. Recent Innovations in Peptide Based Targeted

Drug Delivery to Cancer Cells. *Biomedicines* **4**, 11 (2016).

42. Lien, S. & Lowman, H. B. Therapeutic peptides. *Trends Biotechnol.* **21**, 556–562 (2003).

43. Goodwin, D., Simerska, P. & Toth, I. Peptides As Therapeutics with Enhanced Bioactivity. *Curr. Med. Chem.* **19**, 4451–4461 (2012).

44. Leader, B., Baca, Q. J. & Golan, D. E. Protein therapeutics: A summary and pharmacological classification. *Nat. Rev. Drug Discov.* **7**, 21–39 (2008).

45. Staats, P. S. *et al.* Intrathecal Ziconotide in the Treatment of Refractory Pain in Patients with Cancer or AIDS: A Randomized Controlled Trial. *J. Am. Med. Assoc.* **291**, 63–70 (2004).

46. Escoubas, P. & King, G. F. Venomics as a drug discovery platform. *Expert Rev. Proteomics* **6**, 221–224 (2009).

47. Góngora-Benítez, M., Tulla-Puche, J. & Albericio, F. Multifaceted roles of disulfide bonds. peptides as therapeutics. *Chem. Rev.* **114**, 901–926 (2014).

48. Northfield, S. E. *et al.* Disulfide-rich macrocyclic peptides as templates in drug design. *European Journal of Medicinal Chemistry* **77**, 248–257 (2014).

49. Arolas, J. L., Aviles, F. X., Chang, J. Y. & Ventura, S. Folding of small disulfide-rich proteins: clarifying the puzzle. *Trends Biochem. Sci.* **31**, 292–301 (2006).

50. Akondi, K. B. *et al.* Discovery, Synthesis, and Structure–Activity Relationships of Conotoxins. *Chem. Rev.* **114**, 5815–5847 (2014).

51. Prashanth, J. R. *et al.* Cone snail venomics: From novel biology to novel therapeutics. *Future Med. Chem.* **6**, 1659–1675 (2014).

52. Olivera, B. M. *et al.* Peptide neurotoxins from fish-hunting cone snails. *Science (80-.).* **230**, 1338–1343 (1985).

53. Finney, S., Seale, L., Sawyer, R. T. & Wallis, R. B. Tridegin, a new peptidic inhibitor of factor XIIIa, from the blood-sucking leech Haementeria ghilianii. *Biochem. J.* **324 (Pt 3**, 797–805 (1997).

54. Chang, J. Y. Diverse pathways of oxidative folding of disulfide proteins: Underlying causes and folding models. *Biochemistry* **50**, 3414–3431 (2011).

55. Chang, J.-Y., Lu, B.-Y. & Lai, P.-H. Oxidative folding of hirudin in human serum. *Biochem. J.* **394**, 249–257 (2006).

56. Fuller, E. *et al.* Oxidative folding of conotoxins sharing an identical disulfide bridging framework. *FEBS J.* **272**, 1727–1738 (2005).

57. Betz, S. F. Disulfide bonds and the stability of globular proteins. *Protein Sci.* **2**, 1551–1558 (1993).

58. Hogg, P. J. Disulfide bonds as switches for protein function. *Trends Biochem. Sci.* **28**, 210–214 (2003).

59. Bulaj, G. Formation of disulfide bonds in proteins and peptides. *Biotechnol. Adv.* **23**, 87–92 (2005).

60. Wedemeyer, W. J., Welker, E., Narayan, M. & Scheraga, H. A. Disulfide bonds and protein folding. *Biochemistry* **39**, 4207–4216 (2000).

61. Kühl, T. & Imhof, D. Regulatory FeII/III Heme: The Reconstruction of a Molecule's

Biography. *ChemBioChem* **15**, 2024–2035 (2014).

62. Wißbrock, A., Paul George, A. A., Brewitz, H. H., Kühl, T. & Imhof, D. The molecular basis of transient heme-protein interactions: analysis, concept and implementation. *Biosci. Rep.* **39**, BSR20181940 (2019).

63. Wishart, D. S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668–D672 (2006).

64. Wishart, D. S. *et al.* DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).

65. Wishart, D. S. *et al.* DrugBank: A knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **36**, D901–D906 (2008).

66. Van Arnum, P. Small molecules still leading in new drug approvals. *DCAT Value Chain Insights* (2019). Available at: https://bit.ly/2tRzZIp. (Accessed: 15th August 2019)

67. Schütte, J. Success of Biologics. *Biametrics* (2017). Available at: http://biametrics.com/success-of-biologics/. (Accessed: 20th September 2019)

68. Andrews, L., Ralston, S., Blomme, E. & Barnhart, K. A snapshot of biologic drug development: Challenges and opportunities. *Hum. Exp. Toxicol.* **34**, 1279–1285 (2015).

69. Buvailo, A. Will Biologics Surpass Small Molecules In The Pharma Race? *BiopharmaTrend* (2018). Available at: https://www.biopharmatrend.com/post/67-will-small-molecules-sustain-pharmaceutical-race-with-biologics/. (Accessed: 20th September 2019)

70. Pelaia, G., Vatrella, A. & Maselli, R. The potential of biologics for the treatment of asthma. *Nat. Rev. Drug Discov.* **11**, 958–972 (2012).

71. Smalley, E. AI-powered drug discovery captures pharma interest. *Nat. Biotechnol.* **35**, 604–605 (2017).

72. Zhang, L., Tan, J., Han, D. & Zhu, H. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discovery Today* **22**, 1680–1685 (2017).

73. Smith, J. S., Roitberg, A. E. & Isayev, O. Transforming Computational Drug Discovery with Machine Learning and AI. *ACS Med. Chem. Lett.* **9**, 1065–1069 (2018).

74. Yang, X., Wang, Y., Byrne, R., Schneider, G. & Yang, S. Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. *Chem. Rev.* (2019). doi:10.1021/acs.chemrev.8b00728

75. Fosgerau, K. & Hoffmann, T. Peptide therapeutics: Current status and future directions. *Drug Discov. Today* **20**, 122–128 (2015).

76. Bray, B. L. Large-scale manufacture of peptide therapeutics by chemical synthesis. *Nat. Rev. Drug Discov.* **2**, 587–593 (2003).

77. Harrington, D. A. *et al.* Branched peptide-amphiphiles as self-assembling coatings for tissue engineering scaffolds. *J. Biomed. Mater. Res. - Part A* **78**, 157–167 (2006).

78. Sreejalekshmi, K. G. & Nair, P. D. Biomimeticity in tissue engineering scaffolds through synthetic peptide modifications-Altering chemistry for enhanced biological response. *J. Biomed. Mater. Res. - Part A* **96 A**, 477–491 (2011).

79. Holmes, T. C. Novel peptide-based biomaterial scaffolds for tissue engineering. *Trends Biotechnol.* **20**, 16–21 (2002).

80. Zhang, S., Marini, D. M., Hwang, W. & Santoso, S. Design of nanostructured biological materials through self-assembly of peptides and proteins. *Curr. Opin. Chem. Biol.* **6**, 865–871 (2002).

81. Moyle, P. M. & Toth, I. Modern Subunit Vaccines: Development, Components, and Research Opportunities. *ChemMedChem* **8**, 360–376 (2013).

82. Moyle, P. & Toth, I. Self-Adjuvanting Lipopeptide Vaccines. *Curr. Med. Chem.* **15**, 506–516 (2008).

83. Croft, N. P. & Purcell, A. W. Peptidomimetics: Modifying peptides in the pursuit of better vaccines. *Expert Rev. Vaccines* **10**, 211–226 (2011).

84. Zeng, W. *et al.* A totally synthetic lipopeptide-based self-adjuvanting vaccine induces neutralizing antibodies against heat-stable enterotoxin from enterotoxigenic Escherichia coli. *Vaccine* **30**, 4800–4806 (2012).

85. Xin, H. *et al.* Self-adjuvanting glycopeptide conjugate vaccine against disseminated candidiasis. *PLoS One* **7**, (2012).

86. Zaman, M. *et al.* Structure-activity relationship for the development of a self-adjuvanting mucosally active lipopeptide vaccine against Streptococcus pyogenes. *J. Med. Chem.* **55**, 8515–8523 (2012).

87. Stewart, K. M., Horton, K. L. & Kelley, S. O. Cell-penetrating peptides as delivery vehicles for biology and medicine. *Org. Biomol. Chem.* **6**, 2242–2255 (2008).

88. Vivès, E., Schmidt, J. & Pèlegrin, A. Cell-penetrating and cell-targeting peptides in drug delivery. *Biochim. Biophys. Acta - Rev. Cancer* **1786**, 126–138 (2008).

89. Munyendo, W. L. L., Lv, H., Benza-Ingoula, H., Baraza, L. D. & Zhou, J. Cell penetrating peptides in the delivery of biopharmaceuticals. *Biomolecules* **2**, 187–202 (2012).

90. Fonseca, S. B., Pereira, M. P. & Kelley, S. O. Recent advances in the use of cell-penetrating peptides for medical and biological applications. *Adv. Drug Deliv. Rev.* **61**, 953–964 (2009).

91. Kersemans, V. & Cornelissen, B. Targeting the tumour: Cell penetrating peptides for molecular imaging and radiotherapy. *Pharmaceuticals* **3**, 600–620 (2010).

92. Liu, D. *et al.* In vivo imaging of human colorectal cancer using radiolabeled analogs of the uroguanylin peptide hormone. *Anticancer Res.* **29**, 3777–3783 (2009).

93. Kolmar, H. Engineered cystine-knot miniproteins for diagnostic applications. *Expert Rev. Mol. Diagn.* **10**, 361–368 (2010).

94. Hruby, V. J. & Cai, M. Design of Peptide and Peptidomimetic Ligands with Novel Pharmacological Activity Profiles. *Annu. Rev. Pharmacol. Toxicol.* **53**, 557–580 (2013).

95. Gracia, S. R., Gaus, K. & Sewald, N. Synthesis of chemically modified bioactive peptides: Recent advances, challenges and developments for medicinal chemistry. *Future Med. Chem.* **1**, 1289–1310 (2009).

96. Liskamp, R. M. J., Rijkers, D. T. S., Kruijtzer, J. A. W. & Kemmink, J. Peptides and Proteins as a Continuing Exciting Source of Inspiration for Peptidomimetics. *ChemBioChem* **12**, 1626–1653 (2011).

97. Zhang, M. Q. & Wilkinson, B. Drug discovery beyond the 'rule-of-five'. *Curr. Opin. Biotechnol.* **18**, 478–488 (2007).

98. Leeson, P. D. & Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov.* **6**, 881–890 (2007).

99. Lipinski, C. A. Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discov. Today Technol.* **1**, 337–341 (2004).

100. Mullard, A. Re-assessing the rule of 5, two decades on. *Nat. Rev. Drug Discov.* **17**, 777 (2018).

101. Walters, W. P. Going further than Lipinski's rule in drug design. *Expert Opin. Drug Discov.* **7**, 99–107 (2012).

102. Degoey, D. A., Chen, H. J., Cox, P. B. & Wendt, M. D. Beyond the Rule of 5: Lessons Learned from AbbVie's Drugs and Compound Collection. *J. Med. Chem.* **61**, 2636–2651 (2018).

103. Shultz, M. D. Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs. *J. Med. Chem.* **62**, 1701–1714 (2019).

104. Lee, A. C. L., Harris, J. L., Khanna, K. K. & Hong, J. H. A comprehensive review on current advances in peptide drug development and design. *Int. J. Mol. Sci.* **20**, (2019).

105. Henninot, A., Collins, J. C. & Nuss, J. M. The Current State of Peptide Drug Discovery: Back to the Future? *J. Med. Chem.* **61**, 1382–1414 (2018).

106. Rafferty, J. *et al.* Peptide Therapeutics and the Pharmaceutical Industry: Barriers Encountered Translating from the Laboratory to Patients. *Curr. Med. Chem.* **23**, 4231–4259 (2016).

107. Adessi, C. & Soto, C. Converting a Peptide into a Drug: Strategies to Improve Stability and Bioavailability. *Curr. Med. Chem.* **9**, 963–978 (2005).

108. Gentilucci, L., De Marco, R. & Cerisoli, L. Chemical Modifications Designed to Improve Peptide Stability: Incorporation of Non-Natural Amino Acids, Pseudo-Peptide Bonds, and Cyclization. *Curr. Pharm. Des.* **16**, 3185–3203 (2010).

109. Grauer, A. & König, B. Peptidomimetics - A versatile route to biologically active compounds. *European J. Org. Chem.* 5099–5111 (2009). doi:10.1002/ejoc.200900599

110. Foldvari, M. *et al.* Dermal and transdermal delivery of protein pharmaceuticals: Lipid-based delivery systems for interferon α. *Biotechnol. Appl. Biochem.* **30**, 129–137 (1999).

111. Wollmer, P., Pieber, T. R., Gall, M. A. & Brunton, S. Delivering needle-free insulin using AERx® iDMS (Insulin Diabetes Management System) technology. *Diabetes Technol. Ther.* **9**, S57-64 (2007).

112. Jesus Perez de Vega, M., Martin-Martinez, M. & Gonzalez-Muniz, R. Modulation of Protein-Protein Interactions by Stabilizing/Mimicking Protein Secondary Structure Elements. *Curr. Top. Med. Chem.* **7**, 33–62 (2006).

113. White, C. J. & Yudin, A. K. Contemporary strategies for peptide macrocyclization. *Nat. Chem.* **3**, 509–524 (2011).

114. Hruby, V. J. Designing peptide receptor agonists and antagonists. *Nat. Rev. Drug Discov.* **1**, 847–858 (2002).

115. Ahrens, V. M., Bellmann-Sickert, K. & Beck-Sickinger, A. G. Peptides and peptide conjugates: Therapeutics on the upward path. *Future Med. Chem.* **4**, 1567–1586 (2012).

116. Dombkowski, A. A., Sultana, K. Z. & Craig, D. B. Protein disulfide engineering. *FEBS Letters* **588**, 206–212 (2014).

117. Feyertag, F. & Alvarez-Ponce, D. Disulfide bonds enable accelerated protein evolution. *Mol. Biol. Evol.* **34**, 1833–1837 (2017).

118. Chang, R. J. Y. & Ventura, S. Folding of Disulfide Proteins. *Protein Rev.* **14**, 282 (2010).

119. Gidalevitz, T., Stevens, F. & Argon, Y. Orchestration of secretory protein folding by ER chaperones. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833**, 2410–2424 (2013).

120. Oka, O. B. V & Bulleid, N. J. Forming disulfides in the endoplasmic reticulum. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833**, 2425–2429 (2013).

121. Thornton, J. M. Disulphide bridges in globular proteins. *J. Mol. Biol.* **151**, 261–287 (1981).

122. Petersen, M. T. N., Jonson, P. H. & Petersen, S. B. Amino acid neighbours and detailed conformational analysis of cysteines in proteins. *Protein Eng.* **12**, 535–548 (1999).

123. Craig, D. B. & Dombkowski, A. A. Disulfide by Design 2.0: A web-based tool for disulfide engineering in proteins. *BMC Bioinformatics* **14**, (2013).

124. Qian, W. & Krimm, S. Energetics of the disulfide bridge: An ab initio study. *Biopolymers* **33**, 1591–1603 (1993).

125. Freedman, R. B. Proteins: Structures and molecular properties. *Trends Biochem. Sci.* **10**, 82 (1985).

126. Narayan, M., Welker, E., Wedemeyer, W. J. & Scheraga, H. A. Oxidative folding of proteins. *Acc. Chem. Res.* **33**, 805–812 (2000).

127. Hidaka, Y. Overview of the regulation of disulfide bond formation in peptide and protein folding. *Curr. Protoc. Protein Sci.* 1–6 (2014). doi:10.1002/0471140864.ps2806s76

128. Zavodszky, M. *et al.* Disulfide bond effects on protein stability: Designed variants of Cucurbita maxima trypsin inhibitor-V . *Protein Sci.* **10**, 149–160 (2001).

129. Paul George, A. A. *et al.* Insights into the Folding of Disulfide-Rich μ-Conotoxins. *ACS Omega* **3**, 12330–12340 (2018).

130. Pace, C. N., Grimsley, G. R., Thomson, J. A. & Barnett, B. J. Conformational stability and activity of ribonuclease T1 with zero, one, and two intact disulfide bonds. *J. Biol. Chem.* **263**, 11820–11825 (1988).

131. Tidor, B. & Karplus, M. The contribution of cross-links to protein stability: A normal mode analysis of the configurational entropy of the native state. *Proteins Struct. Funct. Bioinforma.* **15**, 71–79 (1993).

132. Czapinska, H., Otlewski, J., Krzywda, S., Sheldrick, G. M. & Jaskólski, M. High-resolution structure of bovine pancreatic trypsin inhibitor with altered binding loop sequence. *J. Mol. Biol.* **295**, 1237–1249 (2000).

133. Creighton, T. E. Intermediates in the refolding of reduced pancreatic trypsin inhibitor.

J. Mol. Biol. **87**, (1974).

134. Zhang, T., Bertelsen, E. & Alber, T. Entropic effects of disulphide bonds on protein stability. *Nat. Struct. Biol.* **1**, 434–438 (1994).

135. Pecher, P. & Arnold, U. The effect of additional disulfide bonds on the stability and folding of ribonuclease A. *Biophys. Chem.* **141**, 21–28 (2009).

136. Sauer, R. T. *et al.* An Engineered Intersubunit Disulfide Enhances the Stability and DNA Binding of the N-Terminal Domain of λ Repressor. *Biochemistry* **25**, 5992–5998 (1986).

137. Wetzel, R., Perry, L. J., Baase, W. A. & Becktel, W. J. Disulfide bonds and thermal stability in T4 lysozyme. *Proc. Natl. Acad. Sci. USA.* **85**, 401–405 (1988).

138. Wells, J. A. & Powers, D. B. In vivo formation and stability of engineered disulfide bonds in subtilisin. *J. Biol. Chem.* **261**, 6564–6570 (1986).

139. Matsumura, M., Becktel, W. J., Levitt, M. & Matthews, B. W. Stabilization of phage T4 lysozyme by engineered disulfide bonds. *Proc. Natl. Acad. Sci. USA.* **86**, 6562–6566 (1989).

140. Jacobson, R. H., Matsumura, M., Faber, H. R. & Matthews, B. W. Structure of a stabilizing disulfide bridge mutant that closes the active-site cleft of T4 lysozyme. *Protein Sci.* **1**, 46–57 (1992).

141. Wouters, M., George, R. & Haworth, N. 'Forbidden' Disulfides: Their Role as Redox Switches. *Curr. Protein Pept. Sci.* **8**, 484–495 (2007).

142. Biswas, S., Chida, A. S. & Rahman, I. Redox modifications of protein-thiols: Emerging roles in cell signaling. *Biochem. Pharmacol.* **71**, 551–564 (2006).

143. Wouters, M. A., Fan, S. W. & Haworth, N. L. Disulfides as redox switches: From molecular mechanisms to functional significance. *Antioxidants Redox Signal.* **12**, 53–91 (2010).

144. Richardson, J. S. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **34**, 167–339 (1981).

145. Lack, J. *et al.* Solution structure of the third TB domain from LTBP1 provides insight into assembly of the large latent complex that sequesters latent TGF-β. *J. Mol. Biol.* **334**, 281–291 (2003).

146. Siddiqui, K. S. *et al.* Role of disulfide bridges in the activity and stability of a cold-active α-amylase. *J. Bacteriol.* **187**, 6206–6212 (2005).

147. Gilbert, H. F. Molecular and cellular aspects of thiol-disulfide exchange. *Adv. Enzymol. Relat. Areas Mol. Biol.* **63**, 69–172 (1990).

148. Huber-Wunderlich, M. & Glockshuber, R. A single dipeptide sequence modulates the redox properties of a whole enzyme family. *Fold. Des.* **3**, 161–171 (1998).

149. Krause, G. & Holmgren, A. Substitution of the conserved tryptophan 31 in Escherichia coli thioredoxin by site-directed mutagenesis and structure-function analysis. *J. Biol. Chem.* **266**, 4056–4066 (1991).

150. Lin, T. Y. & Kim, P. S. Urea Dependence of Thiol-Disulfide Equilibria in Thioredoxin: Confirmation of the Linkage Relationship and a Sensitive Assay for Structure. *Biochemistry* **28**, 5282–5287 (1989).

151. Wunderlich, M. & Glockshuber, R. Redox properties of protein disulfide isomerase

(dsba) from escherichia coli. *Protein Sci.* **2**, 717–726 (1993).

152. Heimer, P. *et al.* Conformational μ-Conotoxin PIIIA Isomers Revisited: Impact of Cysteine Pairing on Disulfide-Bond Assignment and Structure Elucidation. *Anal. Chem.* **90**, 3321–3327 (2018).

153. Creighton, T. E. Disulfide Bonds as Probes of Protein Folding Pathways. *Methods Enzymol.* **131**, 83–106 (1986).

154. Dutton, J. L. *et al.* A New Level of Conotoxin Diversity, a Non-native Disulfide Bond Connectivity in α-Conotoxin AuIB Reduces Structural Definition but Increases Biological Activity. *J. Biol. Chem.* **277**, 48849–48857 (2002).

155. Tietze, A. A. *et al.* Structurally diverse μ-conotoxin PIIIA isomers block sodium channel Na V1.4. *Angew. Chemie - Int. Ed.* **51**, 4058–4061 (2012).

156. Olivera, B. M., Rivier, J., Scott, J. K., Hillyard, D. R. & Cruz, L. J. Conotoxins. *J. Biol. Chem.* **266**, 22067–22070 (1991).

157. Carstens, B. B. *et al.* Structure-Activity Studies of Cysteine-Rich α-Conotoxins that Inhibit High-Voltage-Activated Calcium Channels via GABAB Receptor Activation Reveal a Minimal Functional Motif. *Angew. Chemie - Int. Ed.* **55**, 4692–4696 (2016).

158. Liu, H. & May, K. Disulfide bond structures of IgG molecules: Structural variations, chemical modifications and possible impacts to stability and biological function. *MAbs* **4**, 17–23 (2012).

159. Dillon, T. M. *et al.* Structural and functional characterization of disulfide isoforms of the human IgG2 subclass. *J. Biol. Chem.* **283**, 16206–16215 (2008).

160. Moritz, B. & Stracke, J. O. Assessment of disulfide and hinge modifications in monoclonal antibodies. *Electrophoresis* **38**, 769–785 (2017).

161. Wiesner, J., Resemann, A., Evans, C., Suckau, D. & Jabs, W. Advanced mass spectrometry workflows for analyzing disulfide bonds in biologics. *Expert Rev. Proteomics* **12**, 115–123 (2015).

162. Neira, J. L. & Rico, M. Folding studies on ribonuclease A, a model protein. *Fold. Des.* **2**, (1997).

163. Raines, R. T. Ribonuclease A. *Chem. Rev.* **98**, 1045–1065 (1998).

164. Levinthal, C. Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44–45 (1968).

165. Pantoja-Uceda, D. *et al.* Deciphering the structural basis that guides the oxidative folding of leech-derived tryptase inhibitor. *J. Biol. Chem.* **284**, 35612–35620 (2009).

166. Yang, Y., Wu, J. & Watson, J. T. Probing the folding pathways of long R3 insulin-like growth factor-I (LR3IGF-I) and IGF-I via capture and identification of disulfide intermediates by cyanylation methodology and mass spectrometry. *J. Biol. Chem.* **274**, 37598–37604 (1999).

167. Lin, C. C. J. & Chang, J. Y. Pathway of oxidative folding of bovine α-interferon: Predominance of native disulfide-bonded folding intermediates. *Biochemistry* **46**, 3925–3932 (2007).

168. Chang, J. Y., Canals, F., Schindler, P., Querol, E. & Avilés, F. X. The disulfide folding pathway of potato carboxypeptidase inhibitor. *J. Biol. Chem.* **269**, 22087–22094 (1994).

169. Salamanca, S., Li, L., Vendrell, J., Aviles, F. X. & Chang, J. Y. Major kinetic traps for

the oxidative folding of leech carboxypeptidase inhibitor. *Biochemistry* **42**, 6754–6761 (2003).

170. Arolas, J. L., Sanglas, L., Lorenzo, J., Bronsoms, S. & Aviles, F. X. Insights into the two-domain architecture of the metallocarboxypeptidase inhibitor from the Ascaris parasite inferred from the mechanism of its oxidative folding. *Biochemistry* **48**, 8225–8232 (2009).

171. Arolas, J. L., Bronsoms, S., Ventura, S., Aviles, F. X. & Calvete, J. J. Characterizing the tick carboxypeptidase inhibitor: Molecular basis for its two-domain nature. *J. Biol. Chem.* **281**, 22906–22916 (2006).

172. Chang, J. Y. The disulfide folding pathway of tick anticoagulant peptide (TAP), a Kunitz-type inhibitor structurally homologous to BPTI. *Biochemistry* **35**, 11702–11709 (1996).

173. Chang, J. Y., Li, L. & Lai, P. H. A Major Kinetic Trap for the Oxidative Folding of Human Epidermal Growth Factor. *J. Biol. Chem.* **276**, 4845–4852 (2001).

174. Lin, C. C. J. & Chang, J. Y. Pathway of oxidative folding of secretory leucocyte protease inhibitor: An 8-disulfide protein exhibits a unique mechanism of folding. *Biochemistry* **45**, 6231–6240 (2006).

175. Price-Carter, M., Gray, W. R. & Goldenberg, D. P. Folding of ω-conotoxins. 1. Efficient disulfide-coupled folding of mature sequences in vitro. *Biochemistry* **35**, 15537–15546 (1996).

176. King, G. F. Venoms as a platform for human drugs: Translating toxins into therapeutics. *Expert Opin. Biol. Ther.* **11**, 1469–1484 (2011).

177. Mansbach, R. A. *et al.* Snails In Silico: A Review of Computational Studies on the Conopeptides. *Mar. Drugs 2019, Vol. 17, Page 145* **17**, 145 (2019).

178. Fuller, E. *et al.* Oxidative folding of conotoxins sharing an identical disulfide bridging framework. *FEBS J.* **272**, 1727–1738 (2005).

179. Gao, B. *et al.* Cone snails: A big store of conotoxins for novel drug discovery. *Toxins (Basel).* **9**, (2017).

180. Olivera, B. M. Conus venom peptides: Reflections from the biology of clades and species. *Annu. Rev. Ecol. Syst.* **33**, 25–47 (2002).

181. Hille, B. Ion Channel Excitable Membranes. *Book* 1–37 (2001). doi:10.1007/3-540-29623-9_5640

182. Prashanth, J. R. *et al.* The role of defensive ecological interactions in the evolution of conotoxins. *Mol. Ecol.* **25**, 598–615 (2016).

183. Rigo, F. K. *et al.* Effect of ω-conotoxin MVIIA and Phα1β on paclitaxel-induced acute and chronic pain. *Pharmacol. Biochem. Behav.* **114**–**115**, 16–22 (2013).

184. Eisapoor, S. S., Jamili, S., Shahbazzadeh, D., Ghavam Mostafavi, P. & Pooshang Bagheri, K. A New, High Yield, Rapid, and Cost-Effective Protocol to Deprotection of Cysteine-Rich Conopeptide, Omega-Conotoxin MVIIA. *Chem. Biol. Drug Des.* **87**, 687–693 (2016).

185. Lewis, R. J. Conotoxin: The venom that's good for you! *2SER* (2016). Available at: https://2ser.com/conotoxin-venom-thats-good/. (Accessed: 28th September 2019)

186. Kaas, Q., Westermann, J. C., Halai, R., Wang, C. K. L. & Craik, D. J. ConoServer, a database for conopeptide sequences and structures. *Bioinformatics* **24**, 445–446

(2008).

187. Kaas, Q., Yu, R., Jin, A. H., Dutertre, S. & Craik, D. J. ConoServer: Updated content, knowledge, and discovery tools in the conopeptide database. *Nucleic Acids Res.* **40**, 325–330 (2012).

188. Robinson, S. D. & Norton, R. S. Conotoxin gene superfamilies. *Mar. Drugs* **12**, 6058–6101 (2014).

189. Kaas, Q., Westermann, J. C. & Craik, D. J. Conopeptide characterization and classifications: An analysis using ConoServer. *Toxicon* **55**, 1491–1509 (2010).

190. Craik, D. J. & Adams, D. J. Chemical modification of conotoxins to improve stability and activity. *ACS Chem. Biol.* **2**, 457–468 (2007).

191. Slapcinsky, J. Cone shells. *Florida Museum of Natural History* (2017). Available at: https://www.floridamuseum.ufl.edu/100years/cone-shells/. (Accessed: 22nd September 2019)

192. Tietze, A. A. *et al.* Structurally diverse μ-conotoxin PIIIA isomers block sodium channel Na V1.4. *Angew. Chemie - Int. Ed.* **51**, 4058–4061 (2012).

193. Khoo, K. K. *et al.* Structure of the analgesic μ-conotoxin KIIIA and effects on the structure and function of disulfide deletion. *Biochemistry* **48**, 1210–1219 (2009).

194. Fan, C. *et al.* The Role of Individual Disulfide Bonds of μ-Conotoxin GIIIA in the Inhibition of NaV1.4. *Mar. Drugs* **14**, 213 (2016).

195. Xu, X. *et al.* Role of the disulfide bond on the structure and activity of μ-conotoxin PIIIA in the inhibition of Na V 1.4. *RSC Adv.* **9**, 668–674 (2019).

196. Bäuml, C. A. & Imhof, D. Tridegin as FXIIIa inhibitor. *Zedira* (2019). Available at: https://zedira.com/Blog/Tridegin-as-FXIIIa-inhibitor_118?oswsid=f2bef85f6cc533cf4182bdd03540d1ef. (Accessed: 26th October 2019)

197. Hyson, J. M. Leech therapy: a history. *J. Hist. Dent.* (2005).

198. Sig, A. K., Guney, M., Uskudar Guclu, A. & Ozmen, E. Medicinal leech therapy—an overall perspective. *Integr. Med. Res.* **6**, 337–343 (2017).

199. Sawyer, R. T. Thrombolytics and anticoagulants from leeches. *Bio/Technology* **9**, 513–518 (1991).

200. Arolas, J. L., Castillo, V., Bronsoms, S., Aviles, F. X. & Ventura, S. Designing Out Disulfide Bonds of Leech Carboxypeptidase Inhibitor: Implications for Its Folding, Stability and Function. *J. Mol. Biol.* **392**, 529–546 (2009).

201. Finney, S., Seale, L., Sawyer, R. T. & Wallis, R. B. Tridegin, a new peptidic inhibitor of factor XIIIa, from the blood-sucking leech Haementeria ghilianii. *Biochem. J.* **324**, 797–805 (1997).

202. Seale, L., Finney, S., Sawyer, R. T. & Wallis, R. B. Tridegin, a novel peptidic inhibitor of factor XIIIa from the leech, Haementeria ghilianii, enhances fibrinolysis in vitro. *Thromb. Haemost.* **77**, 959–963 (1997).

203. Böhm, M. *et al.* Novel insights into structure and function of factor XIIIa-inhibitor tridegin. *J. Med. Chem.* **57**, 10355–10365 (2014).

204. Marx, U. C., Daly, N. L. & Craik, D. J. NMR of conotoxins: Structural features and an analysis of chemical shifts of post-translationally modified amino acids. *Magn. Reson.*

Chem. **44**, (2006).

205. Roy, A., Xu, D., Poisson, J. & Zhang, Y. A Protocol for Computer-Based Protein Structure and Function Prediction. *J. Vis. Exp.* 1–10 (2011). doi:10.3791/3259

206. Donegan, R. K., Moore, C. M., Hanna, D. A. & Reddi, A. R. Handling heme: The mechanisms underlying the movement of heme within and between cells. *Free Radic. Biol. Med.* **133**, 88–100 (2019).

207. D'Alessandro, A. & Zolla, L. Proteomic analysis of red blood cells and the potential for the clinic: what have we learned so far? *Expert Rev. Proteomics* **14**, 243–252 (2017).

208. Hünefeld, F. L. Die Chemismus in der thierischen Organisation. *J.prakt.Chem.* **16**, 160–161 (1839).

209. Wißbrock, A., George, A. A. P., Brewitz, H. H., Kühl, T. & Imhof, D. The molecular basis of transient heme-protein interactions: Analysis, concept and implementation. *Biosci. Rep.* **39**, BSR20181940 (2019).

210. Zhang, L. *Heme biology: The secret life of heme in regulating diverse biological processes*. *Heme Biology: The Secret Life of Heme in Regulating Diverse Biological Processes* (World scientific, 2011). doi:10.1142/7484

211. Lathrop, J. T. & Timko, M. P. Regulation by heme of mitochondrial protein transport through a conserved amino acid motif. *Science (80-.).* **259**, 522–525 (1993).

212. Shimizu, T. Binding of cysteine thiolate to the Fe(III) heme complex is critical for the function of heme sensor proteins. *J. Inorg. Biochem.* **108**, 171–177 (2012).

213. Kühl, T. & Imhof, D. Regulatory FeII/III heme: The reconstruction of a molecule's biography. *ChemBioChem* **16**, 2024–2035 (2014).

214. Zhang, L. & Guarente, L. Heme binds to a short sequence that serves a regulatory function in diverse proteins. *EMBO J.* **14**, 313–320 (1995).

215. Li, T., Bonkovsky, H. L. & Guo, J. Structural analysis of heme proteins: implications for design and prediction. *BMC Struct. Biol.* **11**, 13 (2011).

216. Hou, S., Reynolds, M. F., Horrigan, F. T., Heinemann, S. H. & Hoshi, T. Reversible binding of heme to proteins in cellular signal transduction. *Accounts of Chemical Research* **39**, 918–924 (2006).

217. Igarashi, J. *et al.* Elucidation of the heme binding site of heme-regulated eukaryotic initiation factor 2α kinase and the role of the regulatory motif in heme sensing by spectroscopic and catalytic studies of mutant proteins. *J. Biol. Chem.* **283**, 18782–18791 (2008).

218. Kühl, T. *et al.* Analysis of Fe(III) heme binding to cysteine-containing heme-regulatory motifs in proteins. *ACS Chem. Biol.* **8**, 1785–1793 (2013).

219. Westberg, J. A., Jiang, J. & Andersson, L. C. Stanniocalcin 1 binds hemin through a partially conserved heme regulatory motif. *Biochem. Biophys. Res. Commun.* **409**, 266–269 (2011).

220. Schneider, S., Marles-Wright, J., Sharp, K. H. & Paoli, M. Diversity and conservation of interactions for binding heme in b-type heme proteins. *Nat Prod Rep* **24**, 621–630 (2007).

221. Faller, M., Matsunaga, M., Yin, S., Loo, J. A. & Guo, F. Heme is involved in microRNA processing. *Nat. Struct. Mol. Biol.* **14**, 23–29 (2007).

222. Raghuram, S. *et al.* Identification of heme as the ligand for the orphan nuclear receptors REV-ERBα and REV-ERBβ. *Nat. Struct. Mol. Biol.* **14**, 1207–1213 (2007).

223. Tang, X. D. *et al.* Haem can bind to and inhibit mammalian calcium-dependent Slo1 BK channels. *Nature* **425**, 531–535 (2003).

224. Repessé, Y. *et al.* Heme binds to factor VIII and inhibits its interaction with activated factor IX. *J. Thromb. Haemost.* **10**, 1062–1071 (2012).

225. Atamna, H. & Boyle, K. Amyloid-beta peptide binds with heme to form a peroxidase: relationship to the cytopathologies of Alzheimer's disease. *Proc. Natl. Acad. Sci. USA.* **103**, 3381–3386 (2006).

226. Goodfellow, B. J. *et al.* The solution structure and heme binding of the presequence of murine 5-aminolevulinate synthase. *FEBS Lett.* **505**, 325–331 (2001).

227. Lombardi, A., Nastri, F. & Pavone, V. Peptide-based heme-protein models. *Chem. Rev.* **101**, 3165–3189 (2001).

228. Kühl, T. *et al.* Determination of hemin-binding characteristics of proteins by a combinatorial peptide library approach. *ChemBioChem* **12**, 2846–2855 (2011).

229. Tame, J. R. H. & Vallone, B. The structures of deoxy human haemoglobin and the mutant Hb Tyrα42His at 120 K. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56**, 805–811 (2000).

230. Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins Struct. Funct. Bioinforma.* **23**, 566–579 (1995).

231. Varshney, A., Brooks, F. P. & Wright, W. V. Computing Smooth Molecular Surfaces. *IEEE Comput. Graph. Appl.* **14**, 19–25 (1994).

232. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

233. Brewitz, H. H., Hagelueken, G. & Imhof, D. Structural and functional diversity of transient heme binding to bacterial proteins. *Biochim. Biophys. Acta - Gen. Subj.* **1861**, 683–697 (2017).

234. Peherstorfer, S. *et al.* Insights into mechanism and functional consequences of heme binding to hemolysin-activating lysine acyltransferase HlyC from Escherichia coli. *Biochim. Biophys. Acta - Gen. Subj.* **1862**, 1964–1972 (2018).

235. Yu, D. J. *et al.* Designing template-free predictor for targeting protein-ligand binding sites with classifier ensemble and spatial clustering. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **10**, 994–1008 (2013).

236. Liu, R. & Hu, J. Computational prediction of heme-binding residues by exploiting residue interaction network. *PLoS One* **6**, (2011).

237. Ferreira, L. G., Dos Santos, R. N., Oliva, G. & Andricopulo, A. D. Molecular docking and structure-based drug design strategies. *Molecules* **20**, 13384–13421 (2015).

238. McConkey, B. J., Sobolev, V. & Edelman, M. The performance of current methods in ligand-protein docking. *Curr. Sci.* **83**, 845–856 (2002).

239. Taylor, R. D., Jewsbury, P. J. & Essex, J. W. A review of protein-small molecule docking methods. *J. Comput. Aided. Mol. Des.* **16**, 151–166 (2002).

240. Smith, G. R. & Sternberg, M. J. E. Prediction of protein-protein interactions by docking methods. *Curr. Opin. Struct. Biol.* **12**, 28–35 (2002).

241. Fischer, E. Einfluss der configuration auf die wirkung der enzyme [The influence of configuration on the effect of enzymes]. *Ber. Dtsch. Chem. Ges.* **27**, 2985–2993 (1894).

242. Koshland, D. E. Correlation of structure and function in enzyme action. *Science (80-.).* **142**, 1533–1541 (1963).

243. Hammes, G. G. Multiple conformational changes in enzyme catalysis. *Biochemistry* **41**, 8221–8228 (2002).

244. Moitessier, N., Englebienne, P., Lee, D., Lawandi, J. & Corbeil, C. R. Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go. *Br. J. Pharmacol.* **153**, (2008).

245. Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **261**, 470–489 (1996).

246. Perola, E., Walters, W. P. & Charifson, P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins Struct. Funct. Genet.* **56**, 235–249 (2004).

247. Meng, X.-Y., Zhang, H.-X., Mezei, M. & Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput. Aided-Drug Des.* **7**, 146–157 (2012).

248. Hetényi, C. & Van Der Spoel, D. Blind docking of drug-sized compounds to proteins with up to a thousand residues. *FEBS Lett.* **580**, 1447–1450 (2006).

249. Brint, A. T. & Willett, P. Algorithms for the Identification of Three-Dimensional Maximal Common Substructures. *J. Chem. Inf. Comput. Sci.* **27**, 152–158 (1987).

250. Norel, R., Fischer, D., Wolfson, H. J. & Nussinov, R. Molecular surface recognition by a computer vision-based technique. *Protein Eng. Des. Sel.* **7**, 39–46 (1994).

251. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **161**, 269–288 (1982).

252. Miller, M. D., Kearsley, S. K., Underwood, D. J. & Sheridan, R. P. FLOG: A system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. *J. Comput. Aided. Mol. Des.* **8**, 153–174 (1994).

253. Diller, D. J. & Merz, K. M. High throughput docking for library design and library prioritization. *Proteins Struct. Funct. Genet.* **43**, 113–124 (2001).

254. Burkhard, P., Taylor, P. & Walkinshaw, M. D. An example of a protein ligand found by database mining: Description of the docking method and its verification by a 2.3 Å X-ray structure of a thrombin-ligand complex. *J. Mol. Biol.* **277**, 449–466 (1998).

255. Ewing, T. J. A., Makino, S., Skillman, A. G. & Kuntz, I. D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided. Mol. Des.* **15**, 411–428 (2001).

256. Welch, W., Ruppert, J. & Jain, A. N. Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites. *Chem. Biol.* **3**, 449–462 (1996).

257. Schnecke, V. & Kuhn, L. A. Virtual screening with solvation and ligand-induced complementarity. *Perspect. Drug Discov. Des.* **20**, 171–190 (2000).

258. Bentham Science Publisher, B. S. P. eHiTS: An Innovative Approach to the Docking and Scoring Function Problems. *Curr. Protein Pept. Sci.* **7**, 421–435 (2006).

259. Goodsell, D. S., Lauble, H., Stout, C. D. & Olson, A. J. Automated docking in crystallography: Analysis of the substrates of aconitase. *Proteins Struct. Funct. Bioinforma.* **17**, 1–10 (1993).

260. Hart, T. N. & Read, R. J. A multiple-start Monte Carlo docking method. *Proteins Struct. Funct. Bioinforma.* **13**, 206–222 (1992).

261. Abagyan, R., Totrov, M. & Kuznetsov, D. ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **15**, 488–506 (1994).

262. McMartin, C. & Bohacek, R. S. QXP: Powerful, rapid computer algorithms for structure-based drug design. *J. Comput. Aided. Mol. Des.* **11**, 333–344 (1997).

263. Morris, G. M. *et al.* Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**, 1639–1662 (1998).

264. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727–748 (1997).

265. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins Struct. Funct. Genet.* **52**, 609–623 (2003).

266. Clark, K. P. & Ajay. Flexible ligand docking without parameter adjustment across four ligand–receptor complexes. *J. Comput. Chem.* **16**, 1210–1226 (1995).

267. Taylor, J. S. & Burnett, R. M. DARWIN: A program for docking flexible molecules. *Proteins Struct. Funct. Genet.* **41**, 173–191 (2000).

268. Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **93**, 2395–2417 (1993).

269. Carlson, H. A. & Jorgensen, W. L. An extended linear response method for determining free energies of hydration. *J. Phys. Chem.* **99**, 10667–10673 (1995).

270. Böhm, H. J. Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput. Aided. Mol. Des.* **12**, 309–323 (1998).

271. Head, R. D. *et al.* VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands. *J. Am. Chem. Soc.* **118**, 3959–3969 (1996).

272. Muegge, I. & Martin, Y. C. A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *J. Med. Chem.* **42**, 791–804 (1999).

273. Wallqvist, A., Jernigan, R. L. & Covell, D. G. A preference-based free-energy parameterization of enzyme-inhibitor binding. Applications to HIV-1-protease inhibitor design. *Protein Sci.* **4**, 1881–1903 (1995).

274. Chen, Y. C. Beware of docking! *Trends Pharmacol. Sci.* **36**, 78–95 (2015).

275. Novoa, E. M., De Pouplana, L. R., Barril, X. & Orozco, M. Ensemble docking from homology models. *J. Chem. Theory Comput.* **6**, 2547–2557 (2010).

276. Liu, X. *et al.* Molecular dynamics simulations and novel drug discovery. *Expert Opin. Drug Discov.* **13**, 23–37 (2018).

277. Fischer, M., Coleman, R. G., Fraser, J. S. & Shoichet, B. K. Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand

discovery. *Nat. Chem.* **6**, 575–583 (2014).

278. Ponder, J. W. & Case, D. A. Force fields for protein simulations. in *Advances in Protein Chemistry* **66**, 27–85 (2003).

279. Lopes, P. E. M., Guvench, O. & Mackerell, A. D. Current status of protein force fields for molecular dynamics simulations. *Methods Mol. Biol.* **1215**, 47–71 (2015).

280. Krieger, E., Darden, T., Nabuurs, S. B., Finkelstein, A. & Vriend, G. Making optimal use of empirical energy functions: Force-field parameterization in crystal space. *Proteins Struct. Funct. Genet.* **57**, 678–683 (2004).

281. Serafeim, A. P., Salamanos, G., Patapati, K. K. & Glykos, N. M. Sensitivity of Folding Molecular Dynamics Simulations to even Minor Force Field Changes. *J. Chem. Inf. Model.* **56**, 2035–2041 (2016).

282. Piana, S., Lindorff-Larsen, K. & Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* **100**, L47–L49 (2011).

283. Hornak, V. *et al.* Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Genet.* **65**, 712–725 (2006).

284. Freddolino, P. L., Park, S., Roux, B. & Schulten, K. Force field bias in protein folding simulations. *Biophys. J.* **96**, 3772–3780 (2009).

285. Mackay, D. H. J., Cross, A. J. & Hagler, A. T. The Role of Energy Minimization in Simulation Strategies of Biomolecular Systems. in *Prediction of Protein Structure and the Principles of Protein Conformation* 317–358 (1989). doi:10.1007/978-1-4613-1571-1_7

286. Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).

287. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).

288. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, (2007).

289. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).

290. Tribello, G. A. & Gasparotto, P. Using dimensionality reduction to analyze protein trajectories. *Front. Mol. Biosci.* **6**, (2019).

291. Brown, D. K. *et al.* MD-TASK: A software suite for analyzing molecular dynamics trajectories. *Bioinformatics* **33**, 2768–2771 (2017).

292. McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **109**, 1528–1532 (2015).

293. Lindahl, E. Molecular dynamics simulations. *Methods Mol. Biol.* **1215**, 3–26 (2015).

294. Paul George, A. A. *et al.* Effect of conformational diversity on the bioactivity of μ-conotoxin PIIIA disulfide isomers. *Mar. Drugs* **17**, 390 (2019).

295. Braun, E. *et al.* Best Practices for Foundations in Molecular Simulations [Article v1.0].

Living J. Comput. Mol. Sci. **1**, (2019).

296. Lemkul, J. From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]. *Living J. Comput. Mol. Sci.* **1**, (2019).

297. Grossfield, A. *et al.* Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations [Article v1.0]. *Living J. Comput. Mol. Sci.* **1**, 1–24 (2019).

298. Chang, J. Y., Li, L. & Bulychev, A. The underlying mechanism for the diversity of disulfide folding pathways. *J. Biol. Chem.* **275**, 8287–8289 (2000).

299. Böhm, M. *et al.* Synthesis and Functional Characterization of Tridegin and Its Analogues: Inhibitors and Substrates of Factor XIIIa. *ChemMedChem* **7**, 326–333 (2012).

300. Pearson, W. R. & Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA.* **85**, 2444–2448 (1988).

301. Chen, H. & Zhou, H. X. Prediction of solvent accessibility and sites of deleterious mutations from protein sequence. *Nucleic Acids Res.* **33**, 3193–3199 (2005).

302. Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

303. Shan, Y., Wang, G. & Zhou, H. X. Fold recognition and accurate query-template alignment by a combination of PSI-BLAST and threading. *Proteins Struct. Funct. Genet.* **42**, 23–37 (2001).

304. Paul George, A. A. *et al.* Insights into the Folding of Disulfide-Rich μ-Conotoxins. *ACS Omega* **3**, 12330–12340 (2018).

305. Akondi, K. B. *et al.* Discovery , Synthesis , and Structure − Activity Relationships of Conotoxins. (2013).

306. Eisenhaber, F., Lijnzaad, P., Argos, P., Sander, C. & Scharf, M. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.* **16**, 273–284 (1995).

307. Eisenberg, D. & Mclachlan, A. D. Solvation energy in protein folding and binding. *Nature* **319**, 199–203 (1986).

308. Bäuml, C. A. *et al.* Coagulation Factor XIIIa Inhibitor Tridegin: On the Role of Disulfide Bonds for Folding, Stability, and Function. *J. Med. Chem.* **62**, 3513–3523 (2019).

309. Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science (80-.).* **303**, 1813–1818 (2004).

Acknowledgement

Herein, I would like to express my earnest gratitude and appreciation to the numerous individuals without whose support, this PhD thesis would not have been possible.

First and foremost, I thank Prof. Dr. Diana Imhof who has been my supervisor, scientific mentor and guide from the time I started as a master student and through the years of my PhD studies. Her supervision of my work with unparalleled scientific rigor, while at the same time giving me the creative freedom to explore new ideas and approaches is something that I will be eternally thankful for.

I thank PD Dr. Arijit Biswas for agreeing to be the second supervisor of my thesis. I am thankful for the guidance and mentorship that Dr. Biswas has offered me over the years, through numerous fruitful scientific discussions, ideas and inputs for biomolecular simulation approaches, and for pointing me towards opportunities where my work could be presented.

I thank Prof. Dr. Martin Hofmann-Apitius for accepting my request to be on my PhD examination committee. I thank him for enabling and facilitating all of the collaborative work done with his team at Fraunhofer SCAI.

I am thankful to Nils Lichtenberg from the University of Koblenz, Dr. Astrid Maaß from Fraunhofer SCAI, and Dr. Daniel Tietze and Desirée Kaufmann from the Technical University of Darmstadt for the healthy scientific exchange and collaboration established between their organizations and mine over the course of my PhD studies.

I am thankful to the Deutsche Forschungsgemeinschaft (DFG) and the Deutscher Akademischer Austauschdienst (DAAD) for the financial support offered to conduct my research as well as to participate in conferences.

I am grateful and deeply indebted to my colleagues, both past and present members of the research group, Toni, Pascal, Amelie, Henning, Justin, Charlotte, Ben, Britta, Thomas, Marie, Max, Milena, Mauricio, Ria and Sven as well as all the master students and interns who have worked with me. Thank you all very much for the great times, from intense scientific discussions to unforgettable memories of team lunches, dinners, barbecues, Christmas parties, football games, retreats and team outings. I am especially grateful to Prof. Dr. Diana Imhof and the entire team for the incomparable support extended to me in the months following July 2017, while I faced a dire health predicament. My special thanks go out to Amelie and Pascal in this regard.

I thank my friends Alex, Tatjana, Tony, Prashant, Denny, Srirag and Sharan who have been major sources of inspiration, encouragement and support.

I thank my parents and my brother, who despite being thousands of miles away, always made me feel motivated, encouraged and cared for.

Lastly and most importantly, I thank Swathi, my wonderful wife for being my biggest source of strength, courage, motivation and love, at every step of the way.

Publications

Research articles

1. Hopp, M.; Alhanafi, N.; Paul George, A.A.; Hamedani, N.S.; Oldenburg, J.; Biswas, A.; Imhof, D.; Pötzsch, B. Molecular insights and functional consequences of the interaction of heme with activated protein C. *Antioxid. Redox Signal* ***2020****, doi: http://doi.org/10.1089/ars.2019.7992*

2. Bäuml, C.A.; Paul George, A.A.; Schmitz, T.; Sommerfeld, P.; Pietsch, M.; Podsiadlowski, L.; Steinmetzer, T.; Biswas, A.; Imhof, D. The Impact of Disulfide Pairing in Coagulation Factor XIIIa Inhibitor Tridegin on Its Structure and Function. *Eur. J. Med. Chem.* ***2020****, 112474*

3. Detzel, M.S.; Syllwasschy, B.F,; Steinbock, F.; Ramoji, A.; Hopp, M-T.; Paul George, A.A.; Neugebauer, U.; Imhof, D. Revisiting the interaction of heme with hemopexin: Recommendations for the responsible use of an emerging drug. *(Preprint) bioRxiv* ***2020****, doi: 2020.04.16.044321*

4. Paul George, A.A.; Lacerda, M.; Syllwasschy, B.F.; Hopp, M.-T.; Wißbrock, A.; Imhof, D. HeMoQuest: A webserver for qualitative prediction of transient heme binding to protein motifs. *BMC Bioinformatics,* ***2020****, 21, 124*

5. Humayun, F.; Domingo-Fernandez, D.; Paul George, A.A.; Hopp, M.-T.; Syllwasschy, B.F.; Detzel, M.S.; Hoyt, T.C.; Hofmann-Apitius, M.; Imhof, D. A computational approach for mapping heme biology in the context of hemolytic disorders. *Front. Bioeng. Biotechnol.,* ***2020****, 8, 74*

6. Wißbrock, A.; Goradia, N.B.; Kumar, A.; Paul George, A.A.; Kühl, T.; Bellstedt, P., Ramachandran, R.; Hoffmann, P.; Galler, K.; Popp, J.; Neugebauer, U.; Hampel, K.; Zimmermann, B.; Adam, S.; Wiendl, M.; Krönke, G.; Hamza, I.; Heinemann, S.H.; Frey, S.; Hueber, A.J.; Ohlenschläger, O.; Imhof, D. Structural insights into heme binding to IL-36α proinflammatory cytokine. *Sci. Rep.,* ***2019,*** *9, 16893*

7. Paul George, A.A.; Heimer, P.; Leipold, E.; Schmitz, T.; Kaufmann, D.; Tietze, D.; Heinemann, S.H.; Imhof, D. Effect of Conformational Diversity on the Bioactivity of μ-Conotoxin PIIIA Disulfide Isomers. *Mar. Drugs.* **2019***, 17(7), 390*

8. Bäuml, C.A.; Schmitz, T.; Paul George, A.A.; Sudarsanam, M.; Hardes, K.; Steinmetzer, T.; Holle, L.A.; Wolberg, A.S.; Pötzsch, B.; Oldenburg, J.; Biswas, A.; Imhof, D. Coagulation Factor XIIIa Inhibitor Tridegin: On the Role of Disulfide Bonds for Folding, Stability, and Function. *J. Med. Chem.* ***2019****, 62, 3513-3523*

9. Wißbrock, A.; Paul George, A.A.; Brewitz, H.H.; Kühl, T.; Imhof, D. The molecular basis of transient heme-protein interactions: analysis, concept and implementation. *Biosci. Rep.* ***2019****, 39, BSR20181940*

10. Paul George, A.A.; Heimer, P.; Maaß, A.; Hamaekers, J.; Hofmann-Apitius, M.; Biswas, A.; Imhof, D. Insights into the Folding of Disulfide-Rich μ-Conotoxins. *ACS Omega* ***2018****, 3, 12330-12340*

11. Peherstorfer, S.; Brewitz, H.H.; Paul George, A.A.; Wißbrock, A.; Adam, J.M.; Schmitt, L.; Imhof, D. Insights into mechanism and functional consequences of heme binding to hemolysin-activating lysine acyltransferase HlyC from Escherichia coli. *Biochim. Biophys. Acta - Gen. Subj.* ***2018****, 1862, 1964-1972*

12. Lichtenberg, N.; Menges, R.; Ageev, V.; Paul George, A.A.; Heimer, P.; Imhof, D.; Lawonn, K. Analyzing residue surface proximity to interpret molecular dynamics. *Comput. Graph. Forum.* ***2018****, 37(3), 379-390*

Oral presentations

1. Paul George, A.A.; Imhof, D. Computationally approaching biochemical research - From protein folding to ligand-protein interactions, (2019) International Interdisciplinary Virtual Conference on “Use of Artificial Intelligence and Machine Learning for Bio Engineering Applications”, VIT Vellore, India.

2. Paul George, A.A.; Maas, A., Imhof, D.; Lichtenberg, L.; Menges, R.; Lawonn, K. Computational study on peptide conformations of µ-conotoxins using molecular dynamics and computer visualization methods. (2018) 6th International work conference on Bioinformatics and Biomedical engineering (IWBBIO 2018), Granada, Spain.

3. Paul George, A.A.; Biswas, A.; Sudarsanam, M.; Imhof, D. Putative oxidative folding pathways of the disulfide rich inhibitory peptide Tridegin. (2018) 6th International work conference on Bioinformatics and Biomedical engineering (IWBBIO 2018), Granada, Spain.

Poster presentations

1. Bäuml, C.A.; Schmitz, T.; Paul George, A.A.; Sudarsanam, M.; Hardes, K.; Steinmetzer, T.; Holle, L.A.; Wolberg, A.S.; Pötzsch, B.; Oldenburg, J.; Biswas, A.; Imhof, D. "Impact of disulfide linkage alterations on folding, structure, and activity of FXIIIa inhibitor Tridegin". 27th International Society on Thrombosis and Haemostasis (ISTH) Congress, Melbourne, Australia, July 6-10, 2019.

2. Schmitz, T.; Bäuml, C.A.; Paul George, A.A.; Sudarsanam, M.; Hardes, K.; Steinmetzer, T.; Holle, L.A.; Wolberg, A.S.; Pötzsch, B.; Oldenburg, J.; Biswas, A.; Imhof, D. "Impact of modified disulfide connectivity on structure and function of analogs of the FXIIIa inhibitor tridegin". 14th German Peptide Symposium (GPS), Cologne, Germany, March 18-21, 2019.

3. Nubbemeyer, B.; Kühl, T.; Reher, R.; Paul George, A.A.; Mahgoub, M.; Annala, S.; Benkel, T.; Kaufmann, D.; Kehraus, S.; Crüsemann, M.; Kostenis, E; Tietze, D.; König, G.M.; Imhof, D. “Exploring the specific interaction of FR900359 with the Gαq subfamily”. 14th German Peptide Symposium (GPS), Cologne, Germany, March 18-21, 2019.

4. Paul George, A.A.; Heimer, P.; Bäuml, C.A.; Schmitz, T.; Imhof, D. "The folding of disulfide isomers of multiple disulfide-bonded conopeptides". 8th Peptide Engineering Meeting (PEM8), Berlin, Germany, November 8-10, 2018.

5. Schmitz, T.; Bäuml, C.A.; Paul George, A.A.; Sudarsanam, M.; Hardes, K.; Steinmetzer, T.; Holle, L.A.; Wolberg, A.S.; Pötzsch, B.; Oldenburg, J.; Biswas, A.; Imhof, D. “Factor XIIIA inhibitor tridegin: On the role of disulfide bonds for folding, stability and function”. 35th European Petide Symopsium (EPS), Dublin, Ireland, August 26-31, 2018.

6. Syllwasschy, B.F.; Hopp, M.-T.; Paul George, A.A.; Wißbrock, A.; Imhof, D. “Heme binding beyond Cys-Pro-motifs? – A systematic Analysis of His/Tyr-Based Motifs in Heme Binding peptides and proteins”. 35th European Petide Symopsium (EPS), Dublin, Ireland, August 26-31, 2018.