TECHNISCHE UNIVERSITÄT MÜNCHEN

Fakultät für Chemie

Professur für computergestützte Biokatalyse

Function and Design of Protein Conformational Switches

Mona Baumgart, M.Sc.

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende: Prof. Dr. Corinna Hess

Prüfende der Dissertation:

1. Prof. Dr. Ville R. I. Kaila
2. Prof. Dr. Michael Groll

Die Dissertation wurde am 01.07.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 04.08.2021 angenommen.

Abstract

Proteins show a manifold of different structures and functions in nature. These can involve conformational changes in buried charged networks that regulate the biological activity. Charged networks play an important role in many intricate natural proteins, such as, complex I or heat-shock protein 90 (Hsp90), although the functional principles are not fully understood yet. In this thesis, we designed artificial *de novo* protein systems with buried charged elements embedded in their hydrophobic core, and we further analyzed the conformational switching in Hsp90 by engineering a charged linker region. The aim was to design model systems to study the effect of charged residues with a simple and easy to handle system to obtain a better insight into complex mechanisms. To this end, we developed artificial 4-α-helical bundles as starting framework and combined molecular dynamics (MD) simulations with experimental characterization. We stepwise introduced charged residues that gave molecular insights into the influence of ion-pairs and possibilities how a protein can stabilize them and stay folded. The work highlighted how proteins can overcome the desolvation effects of buried ion-pairs in natural and in our artificial protein models. The combination of these findings resulted in a hyperstable protein model system with a buried charged core. For the natural systems, we investigated the conformational changes in a charged linker region between N-terminal domain (NTD) and M-domain (M-D) of the molecular chaperone Hsp90. We designed mutants that lack the linker region and analyze them by small-angle X-ray scattering (SAXS) and computational simulations. The combined results suggest that the charged linker region is important for the conformation of Hsp90.

Our findings can help to understand the functional and conformational effects of ion-pairs in complex natural proteins. A better understanding of the protein function can have wide-ranging benefits, not only for design, but also in medicine, as perturbations in such charged elements are also involved in severe diseases like cancer or neurodegenerative disorders.

Zusammenfassung

In der Natur vorkommende Proteine besitzen eine Vielfalt an Strukturen und Funktionen. Diese können Veränderungen in der Konformation eingebetteter geladener Netzwerke, welche die biologische Aktivität regulieren, umfassen. Geladene Netzwerke spielen in vielen komplexen, natürlichen Proteinen, wie Komplex I oder Hitze-Schock Protein 90 (Hsp90), eine wichtige Rolle. Deren funktionelle Prinzipien sind dagegen noch nicht vollständig verstanden. In dieser Arbeit haben wir künstliche *de novo* Proteinsysteme mit eingeschlossenen, im hydrophoben Kern eingebetteten geladenen Elementen designt und zudem den Konformationswechsel in Hsp90 durch eine geladene Linkerregion analysiert.

Ziel war es, Modellsysteme zur Untersuchung des Effektes geladener Reste an einem einfachen und leicht handhabbaren System zu designen, um einen besseren Einblick in diese komplexen Mechanismen zu erhalten. Zu diesem Zweck haben wir als Ausgangsstruktur ein künstliches 4-α-helikales Bundle entwickelt und haben Molekulardynamik (MD) Simulationen mit experimenteller Charakterisierung kombiniert. Schrittweise wurden geladene Reste eingeführt, um einen molekularen Einblick in den Einfluss von Ionenpaaren und die Möglichkeiten der Proteine, diese zu stabilisieren und gefaltet zu bleiben, zu erhalten. Die Arbeit zeigt wie Proteine die Desolvatisierungseffekte eingeschlossener Ionenpaare in natürlichen und unseren künstlichen Proteinmodellen überwinden können. Die Kombination aus diesen Erkenntnissen führte zu einem hyperstabilen Proteinmodellsystem mit einem eingeschlossenen, geladenen Kern. Unter den natürlichen Systemen haben wir die konformationellen Veränderungen im geladenen Linker zwischen N-terminaler Domäne (NTD) und M-Domäne (M-D) im molekularen Chaperon Hsp90 untersucht. Wir haben Mutanten ohne diese Linkerregion designt und haben diese mithilfe von Kleinwinkel-Röntgenstreuung (SAXS) und Computersimulationen analysiert. Die kombinierten Ergebnisse deuten darauf hin, dass die geladene Linkerregion für die Konformation von Hsp90 wichtig ist.

Unsere Erkenntnisse können dazu beitragen die funktionalen und konformationellen Effekte von Ionenpaaren in komplexen natürlichen Proteinen zu verstehen. Ein besseres Verständnis der Proteinfunktion kann weitreichende Vorteile nicht nur für das Proteindesign, sondern auch für die Medizin haben, da Störungen in solchen geladenen Netzwerken auch an schweren Erkrankungen wie Krebs oder neurodegenerativen Beschwerden beteiligt sind.

Table of Contents

List of abbreviations

Å	Angstrom
ATP	Adenosine triphosphate
CD	Circular dichroism
CHARMM	Chemistry HARward Molecular Mechanics (force field and software)
CTD	C-terminal domain
CV	Column volume
Da	Dalton
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DSS	Trimethylsilylpropanesulfonic acid sodium salt
E. coli	*Escherichia coli*
GdnHCl	Guanidinium hydrochloride
hGH	Human growth hormone
Hsp90	Heat shock protein 90
HSQC	Heteronuclear single quantum coherence
HISQC	Heteronuclear in-phase single quantum coherence
IL	Interleukine
Ip	Ion-pair
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K	Kelvin
LB	Lysogeny Broth
M	mol l^{-1}
MD	Molecular dynamics
M-D	M-domain
MW	Molecular weight
MZH3	Metal, zinc, and heme binding maquette
NADH	Nicotinamide adenine dinucleotide
NAMD	Nanoscale molecular dynamics
NMR	Nuclear magnetic resonance

NTD	N-terminal domain
Ni-NTA	Nickel column
PBC	Periodic boundary conditions
PCR	Polymerase chain reaction
PDB	Protein data bank
PME	Particle-mesh Ewald
PS1	Porphyrin-binding sequence
PTM	Post-translational modification
Q	Quinone
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
Rop	Repressor of primer
ROS	Reactive oxygen species
SAXS	Small-angle X-ray scattering
SDS	sodium dodecyl sulfate
SEC	Size exclusion chromatography
sl	Short loops
SNase	Staphylococcal nuclease
SUMO	Small ubiquitin-like modifier
TEV	Tobacco etch virus
VdW	Van der Waals
VMD	Visual molecular dynamics (software)

1. Introduction

1.1 Proteins and their conformation

Proteins form the basis of life and are essential for all living cells. The human proteome is estimated to have more than 1 million proteins.[1] The primary structure of a protein is its amino acid sequence, which consists of the 20 natural amino acids. Protein function and their conformation are determined in addition to their amino acid composition, also by post-translational modifications (PTM) such as acetylation, and phosphorylation.[1] Protein function can also be determined by its dynamics. Many proteins are dynamical flexible and variably populated in different microstates.[2] The timescale for such events extends from nanoseconds to hours. Bond vibrations take place in the range of femtoseconds (10^{-15} s), while sidechain rotations and loop motions occur within the picoseconds (10^{-12} s), nanoseconds (10^{-9} s), and microseconds (10^{-6} s) timescales, whereas larger conformational changes occur within ten to hundreds of nanoseconds. The protein folding process can range from microseconds up to hours.[2] Protein conformational switches are rearrangements of the protein structure that allow structural response to a given biological signal.[3] Very often the change is induced by environmental stimuli like pH or temperature,[2] and the stimulus can lead to the activation of the protein and alteration of its biological function.[4] It is therefore not uncommon that multiple conformational changes contribute, *e.g.*, in many enzymatic processes. For example, the binding affinity of an enzyme substrate can be modulated by conformational changes or an optimized structure for a specific reaction step is obtained due to conformational changes.[5] The different conformational states are energetically described by the (free) energy landscape with different minima and corresponding barriers, that are populated by their respective Boltzmann factor, $e^{-E/RT}$.[6]

1.2 4-α-helical bundle frameworks and applications

Artificial helical bundle proteins, also called *Maquettes* that were introduced by Dutton and coworkers[7] and also DeGrado and coworkers[8,9], are often used as framework for different design approaches.[10] These *Maquette* proteins have served as platforms for different design applications due to their simpler structure as compared to natural systems. According to Richard Feynman's "What I cannot create, I do not understand"[11] every mechanism which can be designed is fully understood. Otherwise, it is not possible to properly design the system. *Maquette* proteins can help to understand natural complex processes in a minimalistic model system. For example, there are helical bundles with oxidoreductase activity[12], oxygen transport activity[13] or bundles that incorporate hemes, chlorines, and flavins have been designed.[7,14–18] As it has already been shown[7,10,12–18], artificial bundles can be applied to answer several biochemical questions. Another advantage is the robustness of *Maquette* proteins that can often be overexpressed in *E. coli*.[12] Some bundle proteins show high stability in a wide range of pH, which also make them redox-titration feasible.[16]

One of the first published bundle structures by Regan and DeGrado[8,9] served as a starting point for the design of many other 4-helix bundle proteins. They used motifs of naturally occurring 4-helix bundles like cytochrome *c'*, followed by the design of peptides that self-assemble into tetramers with a sequence consisting of glutamate, lysine and leucine residues. Afterwards, a loop (proline-arginine-arginine) was inserted to connect two peptides, which resulted in a dimerizing peptide. Glutamate and lysine residues were placed in with the aim to obtain a position pointing outwards to the aqueous solution, while leucines were placed towards the protein interior. The amino acids on the surface created favorable salt bridges. This first model showed a good expression level and a high stability against guanidine hydrochloride (GdnHCl) during circular dichroism (CD) spectroscopy measurements.[8] Shifman and coworkers[16] used the *Maquette* published by DeGrado *et al*. in 1989[9] and modified the system to study the coupling of proton exchange and heme oxidation/reduction. Due to the less complex model they were able to work on the relatively unknown mechanism in natural proteins.[16] Another application of *Maquettes* is to construct a flavo-heme protein.[18] The flavin is bound via cysteine and coordinated by histidines in the hydrophobic core of the protein. Sharp *et al*.[18] were able to investigate the light-inducible intramolecular electron transfer as it takes place in natural oxidoreductases with this model. Koder *et al*.[13] designed and engineered an O_2 transport protein based on an artificial 4-α-helical bundle. They used a basic-heme binding starting structure consisting only of leucine, glutamate and lysine residues and modified it step by step to an O_2 binding bundle. Lombardi *et al*.[17] built a simplified model for the active site of metalloproteins in 2019. They had a closer look on diiron and dimanganese proteins. All these examples show how rich the applications for α-helical bundles are and, that complex systems can be mimicked by simple artificial proteins to understand their function or in other words the complexity of natural systems is often not necessary for function.

1.3 Naturally occurring 4-α-helical bundles

The world of proteins is rich with regard to different folds and function. In this thesis, minimal *de novo* designed 4α-helical bundle proteins were studied, but there are also naturally occurring 4α-helical bundle proteins such as cytochrome b_{562}[19], repressor of primer (Rop)[20] and cytokines[21] that can serve as templates for designing artificial bundles. Some of these proteins consist of a single chain, whereas other are homo dimers. The natural 4-α-helical proteins are involved in many processes like signaling, electron transfer, or ligand binding.[19–25] In the following section a few natural bundle proteins are described.

1.3.1 Cytochrom b_{562}

Cytochrom b_{562} from *Escherichia coli* (*E. coli*) is a naturally occurring 4-α-helical bundle protein. The crystal structure of this cytochrome family shown in *Figure 1*, was solved in 1979 at a resolution of 2.5 Å.[19] The heme group in cytochrome b_{562} is non-covalently bound by a histidine and a methionine. The four helices are almost parallel and connected by loops. The monomeric protein is found in the periplasm of *E. coli*[22] and its 106 amino acids single chain has a molecular weight of 12 kDa.[22,23] Cytochrome b_{562} is often used to study the role of ligand binding in protein folding.[23,24] The electron transfer to the ferric center in the oxidized, unfolded state induces folding to the native folded state.[24] Members of the cytochrome family can serve as template for designing new *Maquettes* as they show naturally occurring bundle motifs.[8]

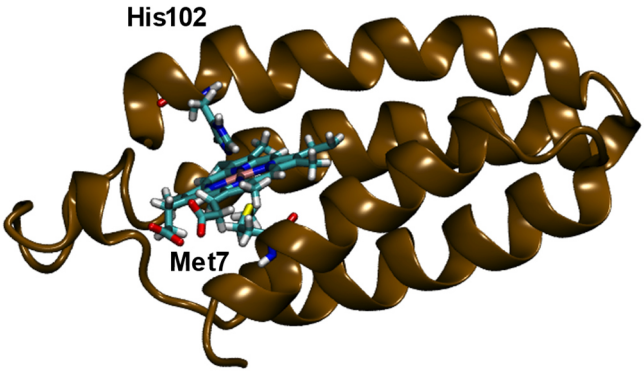

Figure 1: Crystal structure of cytochrome b_{562} from *E. coli* at a resolution of 2.5 Å.[19] The *b*-heme group is non-covalently bound to His102 and Met7. PDB ID: 1QPU.

1.3.2 Repressor of primer (Rop) protein dimer

The Rop protein is a naturally occurring dimeric form of an α-helical bundle that is involved in plasmid replication. The structure of the Rop protein from plasmid ColE1 was solved in 1987.[20] Two units create a homo dimer, which forms an antiparallel bundle. Helix 1 is 30 residues long and

connected by a loop to the second 24 amino acids long helix[20] (*Figure 2*). It shows a typical heptad repeat (**a**bc**d**efg)*n with apolar residues at positions **a** and **d** and a well packed hydrophobic core.[20,25] Steif *et al*.[25] showed by chemical unfolding experiments that Rop is highly stable with an unfolding at 5.3 M of GdnHCl.

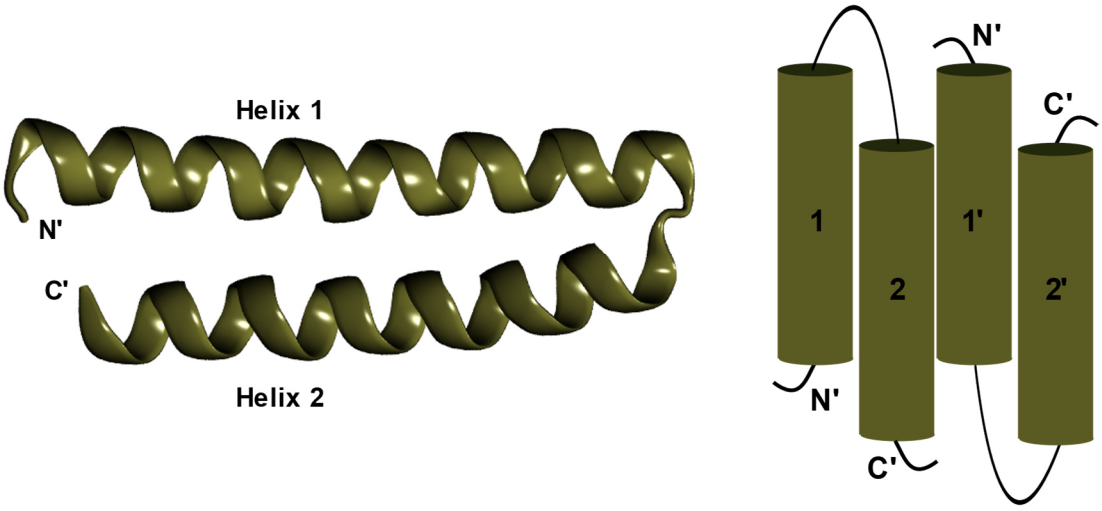

Figure 2: Structure of the Rop protein and topology of the 4-α-helical bundle. PDB ID: 1ROP.

The Rop protein has served as template in the *de novo* design of bundle proteins, which were called Rop-like proteins (RLP).[26] The influence of the heptad positions **a** and **d** on protein folding and stability was analyzed in previous work.[27] The residues at these positions were replaced by alanine and leucine residues to redesigned simpler hydrophobic core. The substitutions led to an increased melting temperature from 54°C to 91°C and a native-like protein is obtained.[27] The RPLs have also been subjected to optimization of positions **b, c, e,** and **g** in order to create a better stabilized bundle protein through additional interactions.[26] These residues shield the hydrophobic residues at positions **a** and **d** and expected to make hydrogen bonds that contribute to favorable electrostatic interactions at the same time. Amphiphilic residues like glutamate, lysine, arginine or tyrosine were placed at positions **e** and **g** to form hydrogen bonds or electrostatic interactions with the residues at position **b** and **c**.[26]

1.3.3 Cytokines

The cytokines superfamily comprises naturally occurring 4α-helical bundles that function, *e.g.,* as chemical messengers.[21] The cytokines are structurally divided into interleukine-1 (IL-1) like, IL-8 like, and growth hormone like bundle proteins. The bundle family is divided into long and short chain proteins that show an untypical *up-up-down-down* topology (*Figure 3*). Members of the long chain group typically have helices of 20 to 30 residues as compared to 10 to 20 residues for the short chain bundles.[28,29] The short chain bundles show a short two-stranded β-sheet. This antiparallel β-sheet is

involved in the core, whereas the long chain bundles show small α-helical parts in the connecting loop regions that do not, however contribute to the protein core.[28] In *Figure 3*, IL-4 is shown as an example for a short chain bundle and human growth hormone (hGH) for the long chain version. Members of the cytokine family can serve as template for designing new 4-α-helical bundles.

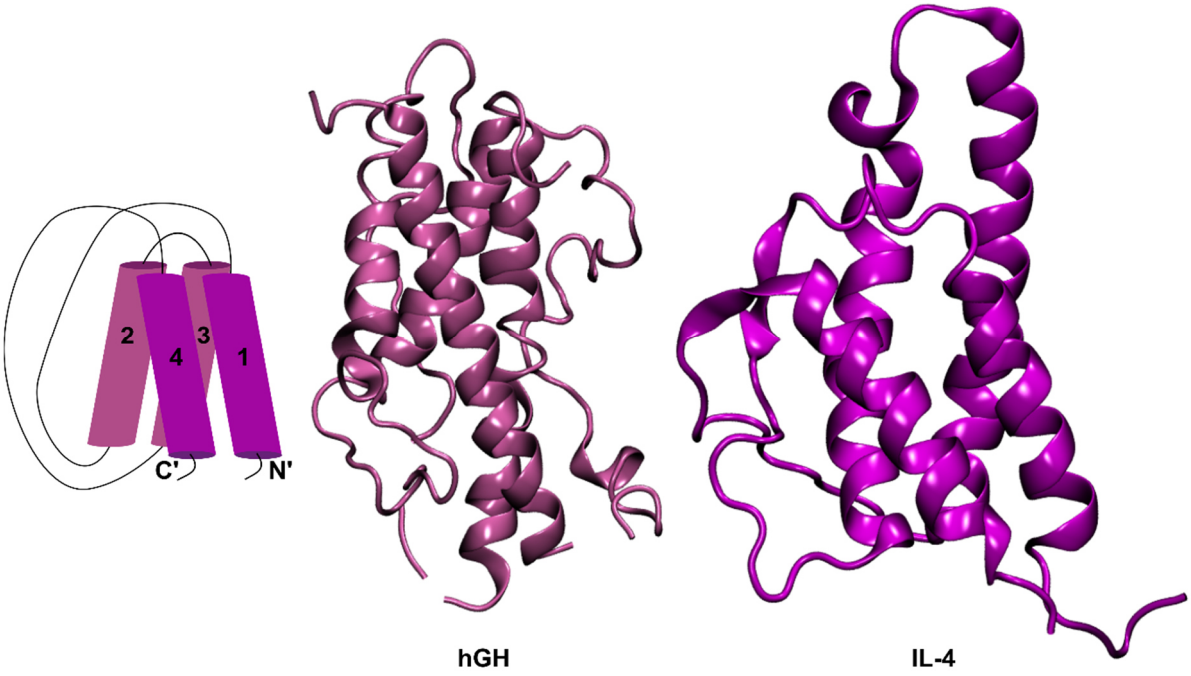

Figure 3: General topology of cytokines. hGH is an example for a long chain cytokine and IL-4 is a short chain one. PDB IDs: 1HGU (hGH), 1BBN (IL-4).

1.4 Long-range conformational switches and charged networks

The activity of some proteins, such as heat shock protein 90 (Hsp90) or respiratory complex I, is regulated by long-range conformational changes, up to 100 to 200 Å away from the active site. These processes can be triggered by oxido-reduction, protonation changes or ligand-binding, which induce the switching by conformational or electrostatic changes, or a combination thereof. Protein conformational switches control the function of molecular chaperones, enzymes, proteins involved in signaling or ion pumps.[30,31] A prominent example is the molecular chaperone Hsp90, which is shortly described in the section below. Another example for long-range controlled systems is complex I of the respiratory chain. Both proteins show a network of charged amino acids, which are involved in the long-range signal propagation. However, the *action-at-a-distance* mechanisms in both systems are poorly understood and the molecular origin of the long-range effects still needs to be investigated.

1.4.1 Heat shock protein (Hsp) 90

Hsp90 is one of the most important molecular chaperones in eukaryotic cells (*Figure 4*).[32] It is involved in the assembly and regulation of signaling pathways that help to activate other proteins (so-called *client proteins*) by combined enzymatic ATP hydrolysis activity and chaperone function.[30] A network of charged amino acids is crucial for both functions.[33] Hsp90 is responsible for the folding and activation of many different client proteins like kinases.[34] The homodimeric Hsp90 consists of an N-terminal domain (NTD), a middle domain (M-D), and a C-terminal domain (CTD). The NTD and M-D are connected via a charged linker region, which is possibly involved in ATP hydrolysis and could be part of the interaction site of client proteins or co-chaperones, but the function is not fully understood yet.[35] The CTD is essential for the dimerization, whereas the NTD is responsible for binding and hydrolysis of ATP.[30] A pair of helices at the C-terminal end of the CTD form a four-helical bundle in the dimerized form.[36] The chaperone function is driven by the ATPase activity, while the binding of ATP induces large conformational changes. Hsp90 changes its conformation from open to closed when it binds ATP, where the N-domains are dimerized.[37] ATP binds at the NTD and an arginine of the M-D (Arg380 in Hsp82, yeast homologue) also contributes to the binding site.[30] The closed state shows a different orientation of N-M-domain and a closed lid region. After all these conformational changes, a slow hydrolysis of ATP takes place.[34,38] In yeast Hsp90 a glutamate residue (Glu33) is important for hydrolysis.[39,40] Glu33 forms a dynamic ion-pair with Arg32 and the conformational change is therefore connected to an ion-pair interaction, with the ion-pair opening up in the presence of ATP and during hydrolysis.[33] One cycle is completed after ATP hydrolysis and Hsp90 changes again its conformation back to the open state.[37] Consequently, the ATP hydrolysis is connected to coupled large conformational changes, and an *action-at-a-distance* event is involved in the regulation of the ATPase activity. Serine and threonine residues 70 Å away from the active site become phosphorylated.[41] Also, several conserved ion-pairs (including Glu381-Arg380[33]) are located across the complete Hsp90 structure, that could stabilize the different conformational states or be involved in the dimerization. The connection between the ion-pairs and the ATPase activity is still not fully clear, but it is believed that the long-range coupling is induced by a combination of electrostatics and conformational changes.

Since Hsp90 acts in assembling, maturation and activation of many client proteins like kinases involved in important cellular pathways, a dysregulation of Hsp90 leads to diseases like cancer or neurological disorders.[42,43] Therefore, inhibitors of Hsp90 can act as cancer therapeutics and it is essential to understand the molecular principle behind the Hsp90 chaperone mechanism.[44]

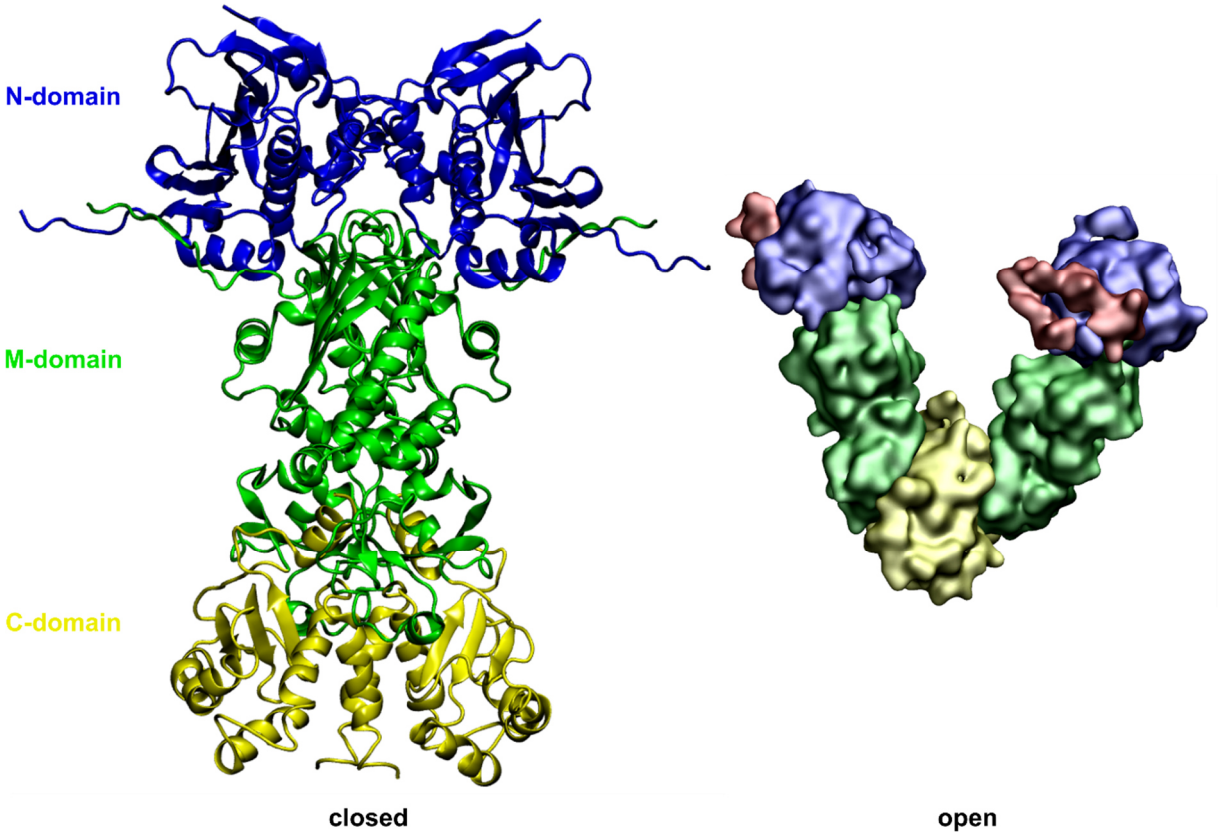

Figure 4: Crystal structure of Hsp90. The N-domain is labeled in blue, M-domain in green, and C-domain in yellow. The left part shows the Hsp90 dimer in the closed state. On the right, the Hsp90 dimer in the open conformation is shown in the same color code. The modeled charged linker region is shown in red. The V-shaped structure results from a coarse-grained simulation performed by Alexander Jussupow (Kaila Lab). PDB ID: 2CG9.

1.4.2 Respiratory Complex I

Complex I of the mitochondrial and bacterial respiratory chains is another example of long-range controlled proteins (*Figure 5*). Interestingly, long-range coupling effects consisting of combined electrostatics and conformational changes take place and control the proton-coupled electron transfer reactions in this membrane protein.[31] The mammalian complex I is one of the largest enzymes with a molecular weight of 1 MDa and up to 45 subunits.[45] The prokaryotic protein is simpler with a molecular mass of ~ 0.5 MDa. The 14 core subunits are highly conserved between the prokaryotic and eukaryotic enzymes. The protein can be divided into a hydrophilic and a membrane domain, where the electron transfer and the proton pumping take place.[46]

Complex I serves as the access point for electrons to the respiratory chain.[47] Reduced nicotinamide adenine dinucleotide (NADH) delivers electrons to complex I, which transfers them to quinone (Q). The released energy is used for pumping protons across the membrane.[48] Complex I is a unique enzyme with an impressive 300 Å *action-at-a-distance* between the electron entry point and the proton pumping events, which manage the enzyme. The mechanism is however not fully

understood.[31] The proton pumping takes place with the help of several ion-pairs distributed across the membrane domain. The conserved ion-pairs show different states, open or closed, during the signal propagation.[31]

Mutations in respiratory complex I are harmful for the whole organism and often causes diseases like neurodegenerative disorders.[49] Complex I is also a source of reactive oxygen species (ROS) in the organism. ROS can lead to DNA damage in the mitochondria and is involved in the aging process.[50,51] Therefore, in addition to providing fundamental functional insight, mechanistic studies can also be of biomedical importance.

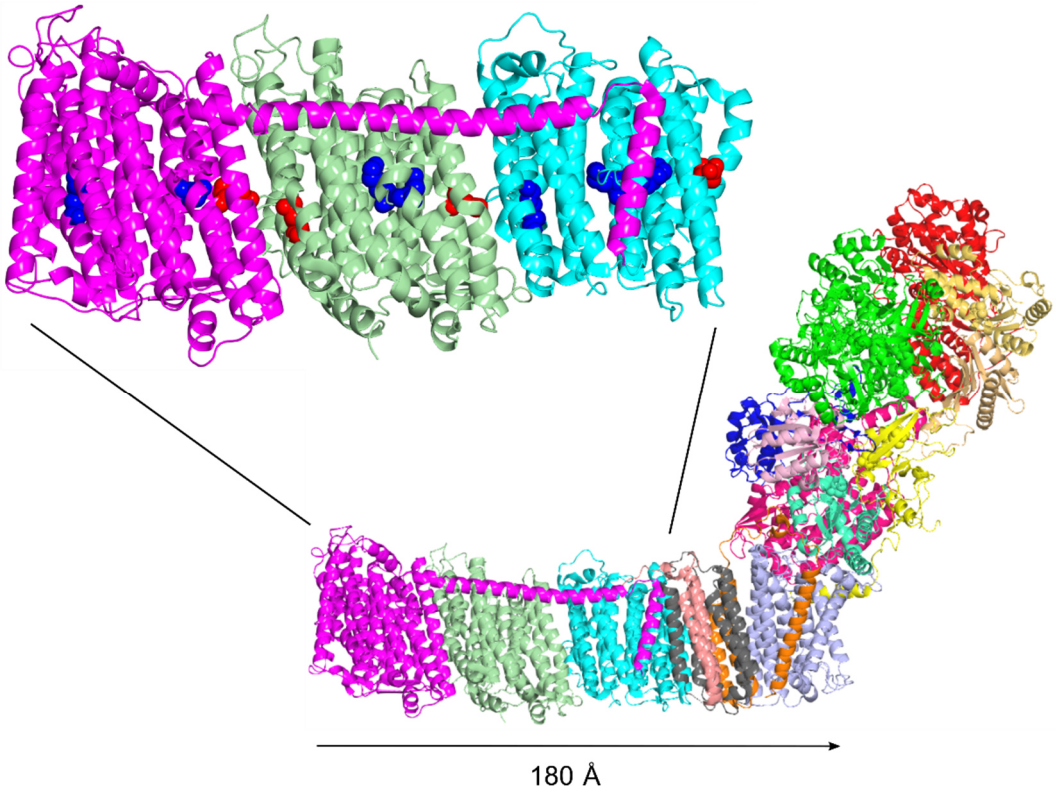

Figure 5: Crystal structure of respiratory **c**omplex I from *Thermus thermophilus* at a resolution of 3.3 Å. The subunits are shown in different colors. The membrane part of the protein is 180 Å long. Charged residues are marked in van der Waals representation in the enlarged membrane part. Selected buried lysines are highlighted in blue and glutamates in red in the inset. PDB ID: 4HEA.

1.5 Protein design

Protein design is an important field in science that aims to answer open questions on how proteins work and to elucidate the mechanisms behind their function. Protein design is also used to create new proteins or improve the properties of already existing ones. Engineering of natural proteins can be used to tune their function, *e.g.* by directed evolution.[52] However, some natural proteins are sensitive to sequence changes and the mutations lead to aggregation or unfolding.[53] *De novo* protein design is an approach where new proteins are created without sequence homology to natural ones. The design of a protein sequence which should have a distinct fold is also called *de novo* design.[53] There are two important requirements for successful design of a uniquely folded protein. One is that the folded state must be energetically favored as compared to the unfolded state, also known as *positive design*.[54,55] The other requirement is that of *negative design*[54,56], which means that there is an energy gap between the desired fold and any other possible fold.[57] It is important to not only stabilize the desired fold, but also to destabilize other possible folds. For example, hydrophobic residues are often buried in the protein interior in the desired fold.[57]

An important target of biochemistry is to develop an accurate method to design a given protein with distinct functions and fold.[53] However, it is challenging to reach this ultimate goal, as there are still many open questions regarding protein structure and function. Proteins show a high complexity[58] and can be affected by many factors in the surroundings, such as interactions other proteins, pH, or ions, that can influence both the structure and function of the target protein. It is often difficult to understand even simple processes within complex natural proteins. It would therefore be a major advantage to build simple protein model systems to study the function of the natural protein counterpart.[9] The protein folding problem *i.e.* how the primary and tertiary structure are connected, is also not understood. Protein design can help to better understand protein stability and folding, and it opens up possibilities for a broad variety of medical and industrial applications and help to understand complex biological systems.[59]

Special focus in the field has been on the design of 4-α-helical proteins because of their simpler structure and their experimental behavior. Different design strategies have been developed to create 4-α-helical proteins: DeGrado and co-workers followed a minimalistic approach to design the first *de novo* 4-α-helical proteins.[9] The α-helices in these constructs are stabilized by molecular interactions defined by distance (*Figure 6*). Some stabilizing amino acids, such as glutamate, alanines, and methionines, establish stabilizing short-range interactions.[9] Medium-range interactions are established via side-chain/side-chain and side-chain/backbone interactions, including hydrogen bonds, electrostatics, and π-stacking of aromates. The dipole moment of the α-helices can also be used to achieve stability.[9] Possible stabilizing electrostatic interactions located on the protein surface are shown in *Figure 6*. The most important long-range interactions for stabilization are hydrophobic interactions that often form the thermodynamic driving force for building secondary structure

elements.[60] Hydrophobic interactions occur between the positions **a** and **d** that often comprise leucine and alanine residues. Another important aspect is the packing of residues within the protein core. The arrangement of the helices in 4-α-helical proteins is controlled by a close side-chain packing.[61,62] For helix packing the "knobs into holes" principle is often used, which includes that the side-chain form knobs and the space between them is called holes. The knobs fit into the holes of a second helix.[63] Charged residues such as lysine or glutamate are often placed on the water accessible surface to stabilize the 4-α-helical protein by electrostatic interactions.[9] The first *de novo* designed protein utilized the principles discussed above, and formed a starting point for further design applications.[8]

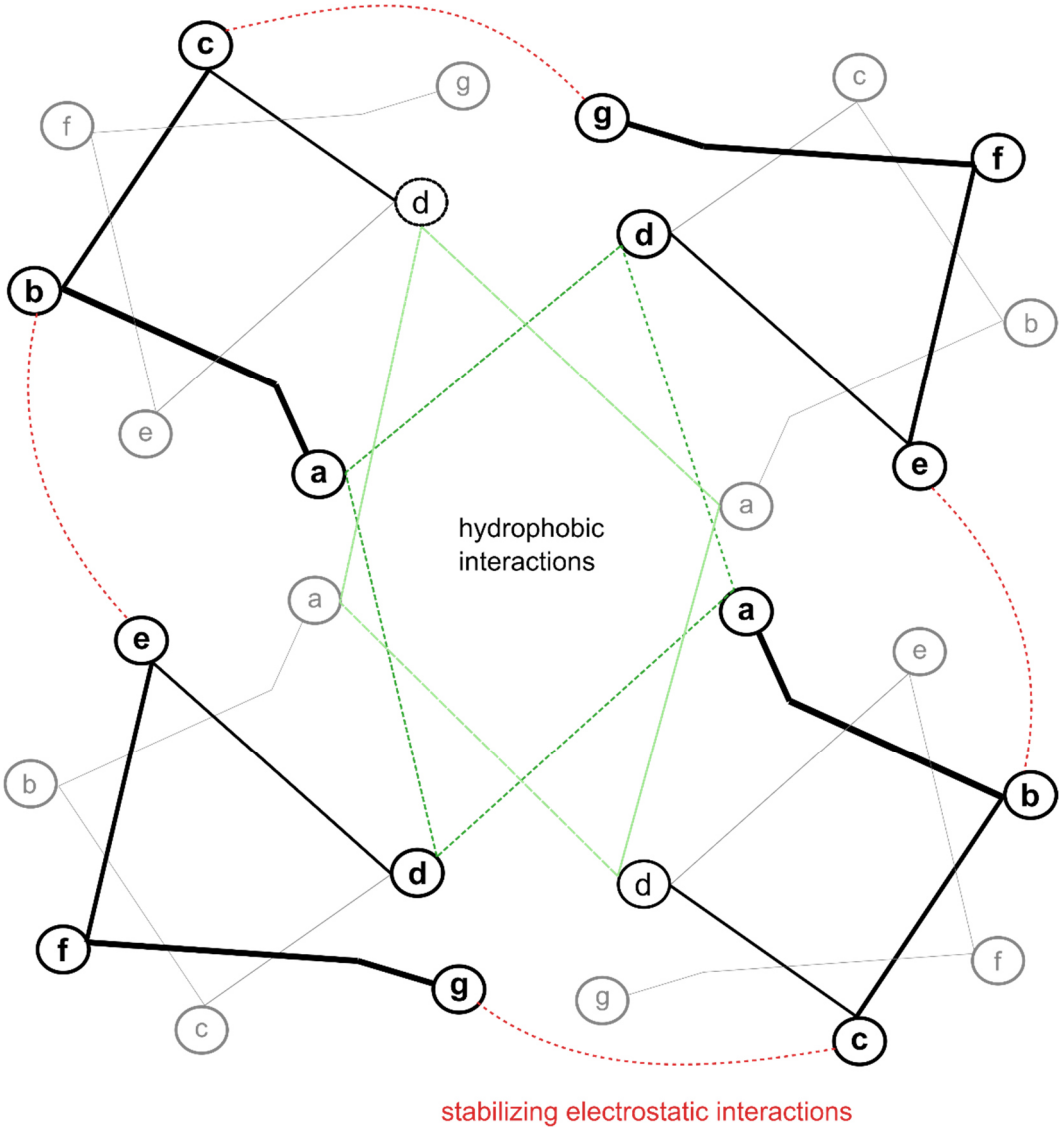

Figure 6: Stabilizing interactions between the helices in a 4-α-helical protein. One heptapeptide is shown for each helix. The interactions inside the hydrophobic core are shown in green and light green. The stabilizing electrostatic interactions on the protein surface are highlighted in red. Redrawn from [9,20].

Proteins can also be built *de novo* using hierarchic design[64] followed by computational design attempts. Starting point for the hierarchic design process is a minimal model, where more specific interactions such as a metal binding site are introduced.[64] Many research groups have focused on the development of computational design approaches already during the 1990s.[65] The algorithms often focus on the hydrophobic core[66], but there are also examples where the whole protein is taken into account.[67] One of the most prominent current algorithms is ROSETTA,[68,69] in which short fragments of known proteins are combined using a Monte-Carlo strategy to design new proteins with high accuracy.[70] Based on ongoing development, machine learning approaches are likely to become highly useful in protein design. For example, AlphaFold[71] or CASP[72] are used for protein structure predictions from the sequence.

1.6 Artificially designed 4-α-helical bundles

Three α-helical bundle topologies were used as starting point to investigate the function of switchable charged elements studied in this thesis. All three models show different advantages and disadvantages, but due to their properties, they were suitable for different experimental structure-solving techniques.

The following section focuses on the design topologies and the applications, for which they were originally designed. All three constructs were *de novo* designed without any sequence homology to naturally occurring bundle proteins.

1.6.1 4-α-helical bundle protein with oxidoreductase functions

One of the bundle topologies studied in this work, is a small, 17 kDa, artificial stable 4-helix-bundle from Dutton and coworkers[12] created for studies of oxidoreductase functions. The proteins were originally designed using a simple framework without functionality, and subsequently combined with functional design without copying natural enzymes. Only small changes are necessary to create different oxidoreductase functions like light-activated electron transfer or photochemical charge separation. Starting point for the design process was a homodimer consisting of two helix-loop-helix units. The units were connected via disulfide bridges between the loops. First, cysteine residues were removed from the loops to produce a single chain 4-α-helical bundle with a reduced subset of 12 amino acids.[12] The helices were connected by artificial loops, which are 9 amino acids long and consist of glycine and serine. After substitution of four histidines, the whole protein sequence of 132 amino acids is the following:

```
G    EIWKQFE DALQKFE EALNQFEDLKQL GGSGSGSGG
     EIWKQFE DALQKFE EALNQFEDLKQL GGSGSGSGG
     EIWKQFE DALQKFE EALNQFEDLKQL GGSGSGSGG
     EIWKQFE DALQKFE EALNQFEDLKQL
```

The loops are marked in grey. The designed protein consists of sequentially four identical helices. After investigating different loop lengths, loops of 9 amino acids resulted in a highly overexpressed monomeric 15 kDa protein. Therefore, these loops were kept for all constructs and different mutations were introduced, yielding in total 12 unique models.[12] *Figure 7* shows the assumed topology of the designed *Maquette*, but the three-dimensional structure remained experimentally unresolved. The protein interior is characterized by hydrophobic residues, while the surface contains more charged residues.

The melting temperature of the designed *apo* bundle was tuned by substitution of four histidine residues that were replaced by phenylalanines in the most stable model E[12], where the heme-ligation was eliminated. This so-called model E, shows the highest melting temperature with ~ 95°C. Another tested concept for raising the melting temperature was adding salt bridge favorable amino acids such as lysine or arginine on the protein surface, which leads to an increase of the melting temperature of almost 20°C. Most of the constructs show an additionally increased melting temperature after ligand binding.[12]

The conducted NMR studies of the *apo* models A and E show relatively undispersed spectra indicating a possible molten-globule like state. The addition of a heme group to model A improves the dispersion and indicates a more structured system, while it does not lead to a change in the spectra for model E. As shown by subsequent results, the designed 4-α-helical bundle protein provides a scaffold framework for multiple different applications.[12]

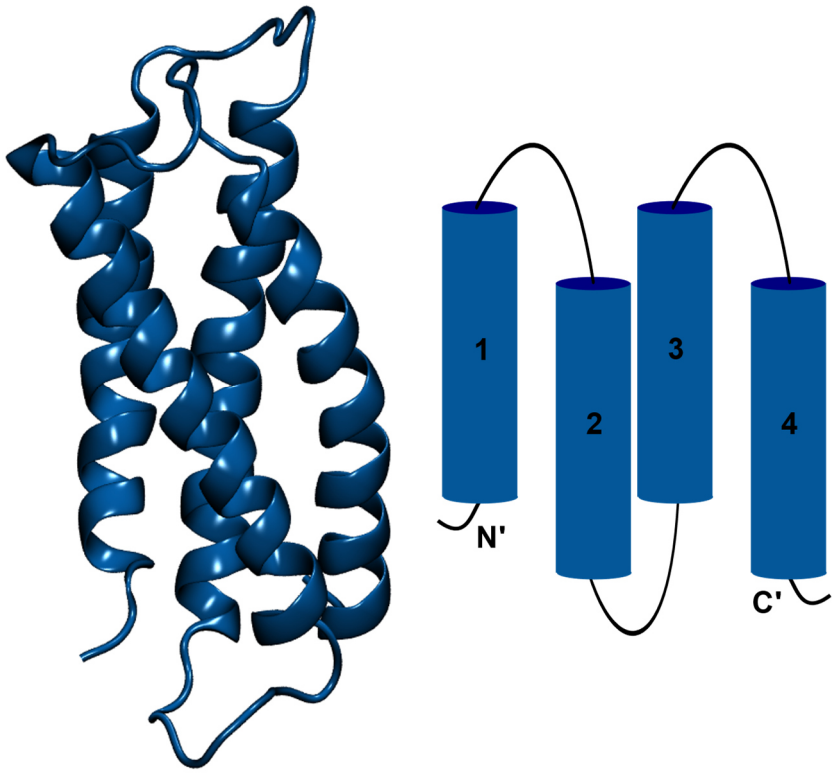

Figure 7: Computationally designed model of a 4-α-helical bundle protein and topology of the construct. The structure was modeled *in silico* in our work (see results) based on the model E sequence.[12]

1.6.2 Porphyrin-binding 4-α-helical bundle

Another artificial protein scaffold studied in this work is a hyperstable bundle from DeGrado and coworkers that has been originally designed for porphyrin-binding.[73] This *de novo* designed bundle called porphyrin-binding sequence 1 (PS1) is monomeric and binds a non-natural porphyrin (*Figure 8*).

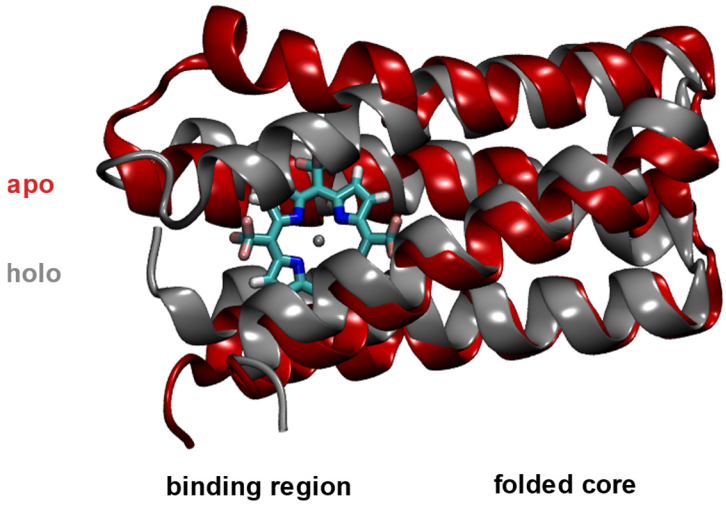

Figure 8: Structure of the porphyrin-binding 4-α-helical bundle. The three-dimensional structure was solved by NMR. The apo state is shown in red. The porphyrin bound structure in grey shows the holo state. The bundle can be divided into the porphyrin-binding region, which is more flexible and into the folded core. PDB IDs: 5TGW (apo), 5TGY (holo).

The porphyrin-bound protein-ligand complex is stable up to a temperature of 120°C.[73] Analysis of natural proteins has shown that side-chain packing up to 20 Å away from the binding site has an effect on the binding. Previous *de novo* designed cofactor-binding proteins like SCRPZ-2[74] or α2[75] were missing a well-packed apolar core. DeGrado and coworkers followed this concept for the first time in the *de novo* design process and tried to optimize also the *apo* protein and create an apolar, folded core. Naturally occurring ligand-binding bundle proteins like cytochrome b_{562} show a folded core combined with a binding region. The starting structure for PS1 was SCRPZ-2, also a *de novo* designed cofactor binding bundle protein. They first designed an antiparallel coiled coil followed by side-chain packing. *Figure 9* shows the antiparallel topology of this bundle.

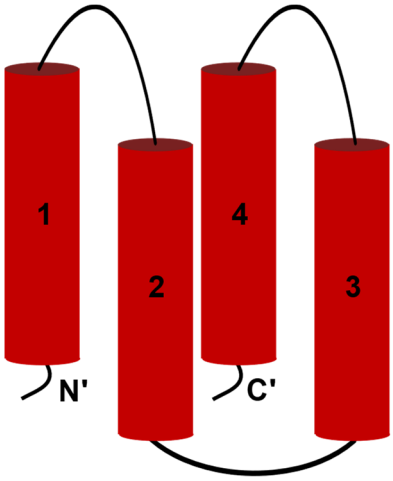

Figure 9: Topology of the porphyrin-binding 4-α-helical bundle PS1. The construct has an antiparallel topology.

The design process resulted in a ~ 13 kDa protein with the following 109 amino acids long sequence:

```
SEFEKLRQ TGDELVQ AFQRLREIFDK  GD
DDSLEQV  LEEIEEL IQKHRQLFDNR  QEAA
DTEAAKQ  GDQWVQL FQRFREAIDK   GD
KDSLEQL  LEELEQA LQKIRELAEKKN
```

After DeGrado *et al.*[73] successfully designed this porphyrin binding protein PS1, they were able to solve the *apo* (PDB ID: 5TGW) and *holo* (PDB ID: 5TGY) structure with NMR in very high agreement with the designed model. The binding region is more flexible in the *apo* state that leads to open and closed conformational states. The binding region is water accessible in both conformations, with an open conformation that is probably more favored for ligand binding due to the larger space. The folded core of the *apo* and *holo* structures do not show significant structural differences.[73]

1.6.3 Metal, zinc, and heme binding maquette MZH3

Another starting topology used in this thesis is a *Maquette* scaffold also developed on the Dutton group.[76] The designed protein is called MZH3, which stands for the three cofactors metal, zinc, and heme that was created to mimic the electron transfer process in photosystem II. The protein sequence was based on DF3, another metal-ion binding *Maquette* protein[77,78]:

```
GSPELRQEHQQLAQEFQQLLQEIQQLGRELLKGELQGIKQLREASEK  ARN
PEKKSVLQKILEDEEKHIELLETLQQTGQEAQQLLQELQQTGQELWQL GGSGG
PELRQKHQQLAQKIQQLLQKHQQLGAKILEDEEKHIELLETIL      GGSGG
DELRELLKGELQGIKQYRELQQLGQKAQQLVQKLQQTGQKLWQLG
```

Dutton and coworkers combined different observations in other designed *Maquettes* like Due Ferro proteins[79] or also in natural systems like cytochrome *b* to design MZH3.[76] The result of the design process was a 22.5 kDa 4-α-helical protein with an antiparallel topology (*Figure 10*).

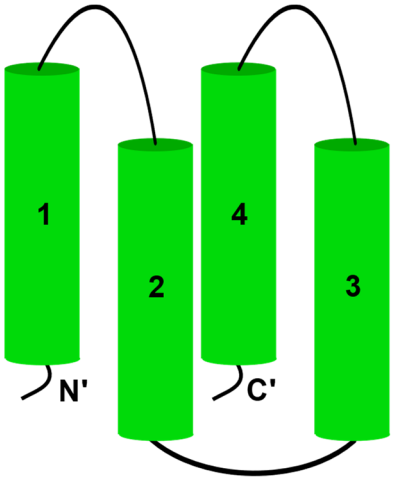

Figure 10: Topology of the bundle protein MZH3. The metal, zinc, and heme binding 4-α-helical bundle shows an antiparallel topology.

The helices of this *Maquette* are 45 residues longer as compared to the other two starting structures, and also compared to the size of natural 4-α-helical bundles proteins that normally show helices between 15 and 24 amino acids.[80] The MZH3 model was a crystallized to a high 1.45 Å resolution for the *holo* form of the protein, where heme and two metal ions are bound (*Figure 11*).[76]

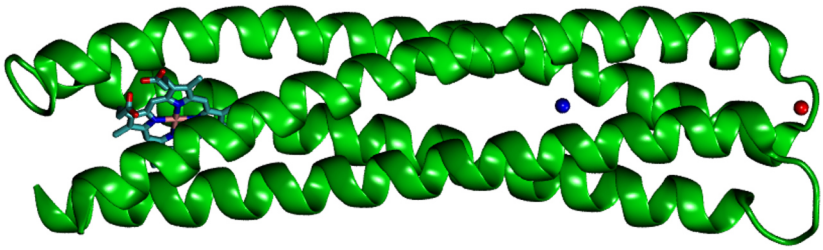

Figure 11: Crystal structure of MZH3 at 1.45 Å resolution. A zinc and a chloride ion are bound additionally to a heme. The heme is shown in licorice. The zinc ion is shown in blue and the chloride ion in red. PDB ID: 5VJT.

1.7 Electrostatics in natural proteins

The hydrophobic effect provides an important driving force for protein folding, but also electrostatics play an important role for the function of several proteins. Besides hydrophobic interactions and a close-packing electrostatics are also important for building the tertiary structure of a protein.[81,82] Fersht and co-workers[83,84] suggested that a partially buried ion-pair is formed in barnase very early during the folding process. Hydrogen bonds and salt bridges also called ion-pairs are common in proteins. The hydrogen-bond network contributes to achieving the correct conformation. Electrostatics due to hydrogen-bonding also play a role in enzyme catalysis.[81]

Ion-pairs are also involved in many biological processes. A salt bridge can be defined as an interaction between two oppositely charged residues in hydrogen-bonding vicinity. Salt bridges are formed between Asp or Glu and His, Arg, or Lys when the two charged side-chain groups are located within 4 Å of each other and when one carboxyl oxygen (Asp/Glu) and one nitrogen atom (Lys/Arg/His) of the side-chains are also within this region.[82,85] Ion-pairs can be involved in improving protein stability, signal transduction, photobiological processes, or ligand binding to list a few examples.[85,86] Salt bridges on the protein surface stabilize the protein and can lead to an increased thermal stability, but in a few cases, buried ion-pairs have been suggested to provide stabilizing contribution.[87] In general, however, ion-pairs are favorable for proteins because they help stabilizing the three-dimensional structure or help during the folding process.[85,88] Nevertheless, a single buried charged residue has a destabilizing effect unless the residue can interact with the surrounding protein residues, as also shown in this thesis work. An example for such an interaction is hydrogen bonding with other residues in the protein interior.[89] Buried charges increase the desolvation penalty, but due to the formation of salt bridges a partial compensation can take place. A second possibility is charge compensation by polar groups.[90] The desolvation energy of an ion is calculated with the Born equation,

$$E_{solv} = \frac{Z^2 e^2}{8\pi\varepsilon_0 a}\left(1 - \frac{1}{\varepsilon}\right)N_A$$

where Z is the charge, e is the elementary charge 1.6022×10^{-19} C, ε_0 is the vacuum permittivity, ε is the dielectric constant of the solvent (*e.g.* water or protein), a is the radius of the ion, and N_A is the Avogadro constant. The desolvation penalty is higher the more buried an ion-pair is. On the other hand, the salt bridge formation and the interactions between the charged residues and the rest of the protein is more favorable for buried ion-pairs and that partially to compensates for the desolvation penalty.

The electrostatic interaction energy between two charged particles can be calculated with Coulomb's law,

$$V_{ele} = \sum_{i,j\neq i} \frac{q_i q_j}{4\pi\varepsilon r_{ij}}$$

where q is the charge of particle i or j, ε is the dielectric constant of the solvent and r is the distance between the two charged particles.

The energetics of transferring an oppositely charged ion-pair from water inside a protein calculated with a simple model from Robinson *et al.*[91] is visualized in *Figure 12*. For the original discussion of these effects see Warshel *et al.*[92] and references therein. The black curve represents the free energy as a function of ε. The Born desolvation energy is shown in green and the Coulomb energy in red. The distance r_{ij} between the two charged residues is assumed to be 3.5 Å, while their radius is 1.75 Å. The Coulomb interaction between the positive and negative charges is stronger inside the protein than in water. The plot also shows that the transfer energy, resulting from the Born desolvation energy, of the charges into the protein interior (ε ~ 4) is less favorable as compared to solvated charged residues (ε = 80).

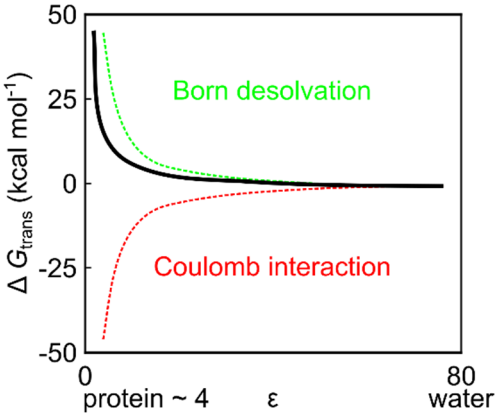

Figure 12: Schematic representation of energetic effects of transferring an oppositely charged ion-pair from water into the protein. The free energy of the transfer calculated as a function of ε is shown in black. The interaction between the positively and negatively charged residues is calculated with Coulomb's law (red). The desolvation energy is obtained with Born's formula (green). The effects are qualitatively estimated by using the model described by Robinson *et al.*[91]

Buried ion-pairs often have destabilizing effects on a protein that cannot fully compensate for the high energetic cost.[93] The position of the ion-pair within the protein can determine its stabilizing contribution on the overall protein structure. For example, a generally destabilizing ion-pair can lead to an increased specificity in functionality and helps to create a more unique protein. In this case, the protein has to pay a prize in stability for becoming more specific.[93]

In natural α-helices, most ion-pairs are often located the same helix at positions i and i +/- 4 or i and i +/- 3.[85] Negatively charged residues, such as Glu or Asp are favored in the vicinity of the N-terminus, while Lys, Arg, or His are better stabilized on the C-terminus. This way, the residues may contribute to helix-capping motifs. Ion-pairs have also been reported to contribute varying stabilizing effect on α-helices.[82]

1.8 Buried charged residues in the hydrophobic protein core

The folding of soluble proteins is directed by building a hydrophobic core and a polar surface, where the driving force originates from packing the hydrophobic core.[94] Packing of charged residues inside the hydrophobic core are generally energetically unfavorable. Nevertheless, many proteins comprise charged residues or even networks of charged residues that are important, *e.g.*, for enzymatic activity[30,95] or proton transport.[31,46] The energetic cost of bringing a charged residue from high dielectric environment (ε_{water} = 80) to a lower protein dielectric ($\varepsilon_{protein}$ = 4-10) can be significant and destabilizes the protein due to the (Born) desolvation effects (see above). The same principles also apply to an ion-pair, consisting of two charged residues, as the Coulomb interaction between the two opposite charges does not fully compensate for the penalty of desolvating the charged residues.[91,93]

Considering typical destabilizing effects of 3.5 kcal mol^{-1} in natural proteins[93], large proteins like respiratory Complex I, with hundreds of ion-pairs inside the protein would therefore be expected to unfold.[31,46] Interestingly, proteins have evolved different approaches to compensate for burying charged residues inside their hydrophobic parts.

Warshel studied protein electrostatics during the 1970s,[96,97] and identified key electrostatic principles discussed above. Generally, ion-pairs located on the polar protein surface are favored, and would become unstable if buried within the nonpolar protein interior. Nevertheless, some naturally occurring buried ion-pairs may require a polar surrounding,[86] and for example hydrogen-bonds and dipoles similarly as solvent molecules.[97] Key design principles to create stable ion-pairs are studied and described in the results section of this thesis work and publication Baumgart *et al.*[98].

In recent studies, where artificial ion-pairs were introduced into the hydrophobic core of the natural protein *Staphylococcal* nuclease (SNase).[91,99] The studied system had ion-pairs introduced in different positions, and destabilizing effects of up to 5 kcal mol^{-1} were found. Further studies suggested that the cost of bringing in two complementary charged residues together into the hydrophobic core is around 2 kcal mol^{-1} less than introducing only a single charge.[91] In a recent study[100], a glutamate-lysine ion-pair was introduced on a single helix in SNase at positions *i* and *i*+4, and the effect of the ion-pair reversal was also investigated. It was found that the protein undergoes conformational changes to compensate for the energetic penalty in introducing buried charged residues, and the results were further compared to charged residues inserted on an α-helix with the second charge placed on a loop region. It was suggested that the protein may to reorganize to provide a more polar environment for the buried ion-pair to allow for better solvation of the charged residues. The degree of reorganization may depend on the placement of the charged residues and varies from subtle side-chain rearrangements to partial backbone unfolding. The protein compensates the penalty by reorganization.[100]

2. Aims of the study

The main goal of the thesis is to design a simplified model system to study the mechanisms behind charged networks in complex proteins. Detailed insight from experiments is still missing due to the high complexity of, for example, respiratory complex I or Hsp90. Also, the yield and behavior of these protein machineries are often not ideal for detailed experimental characterizations, and the complex protein surroundings may not always be necessary to study the biological function. We developed here a combined approach by integrating computational work and biochemical and biophysical experiments with protein design. In another related project, the conformational changes enabled by a charged linker region in the natural Hsp90 chaperone were investigated.

To achieve these goals, three different α-helical *Maquette* frameworks were used as model protein scaffolds. These proteins provided experimentally well-behaved systems with high yields obtained during the expression and purification process. To systematically study the influence of charged residues built in the hydrophobic protein core, we developed mutation series started with single glutamate or lysine residue, and ending up with networks with up to three ion-pairs.

To develop a conformational switch, we built ion-pairs in a so-called *Maquette* 1 model, that was based on earlier designed scaffolds by Dutton and coworkers.[12] These models lack an experimentally resolved structure, which was attempted both computationally and experimentally. The *Maquette* 2 model was based on a protein scaffold designed by DeGrado and coworkers.[73] In *Maquette* 2, we probed the effect of introducing charged residues and different ion-pairs, in various micro-environments. These engineered protein models and their overall stability were analyzed by NMR and chemical unfolding studies. One of the designed models allowed us to resolve the three-dimensional structure of *Maquette* 2 with a buried ion-pair. Another aim of the thesis was to understand how buried ion-pairs are stabilized in the protein interior. To this end, we identified charge shielding motifs and tested how these help in stabilizing the buried ion-pair. For this purpose, we built a larger *Maquette* 3 bundle protein that we could structurally characterize using x-ray crystallography and molecular simulations.

In the last part of the thesis work, we examined the folding of the buried ion-pairs under *in vivo* conditions using force profile analysis and co-translational folding assay.[101,102] To this end, we created additional variations of the *Maquette* 2 models with buried ion-pairs.

3. Materials and methods

In this section, the employed materials and methods of the thesis are described. The first part of the chapter covers the experimental materials and procedures. The computational methods and models are explained in the second part.

3.1 Laboratory materials

3.1.1 Chemicals

Substance	**Manufacturer**
4-(2-hydroxyethyl)-1-piperazineethanesulfonic Acid (HEPES)	Carl Roth, Karlsruhe (Germany)
Acetic acid	Sigma-Aldrich, St. Louis (USA)
Acrylamide/bis solution, 29:1	Serva Electrophoresis GmbH, Heidelberg (Germany)
Ammonium-^{15}N chloride	Sigma-Aldrich, St. Louis (USA)
Ampicillin sodium salt	Carl Roth, Karlsruhe (Germany)
Biotin	Sigma-Aldrich, St. Louis (USA)
D-(+)-glucose	Sigma-Aldrich, St. Louis (USA)
D-glucose-$^{13}C_6$	Sigma-Aldrich, St. Louis (USA)
DNase I	Roche, Basel (Switzerland)
Disodium phosphate	Acros organics, New Jersey (USA)
Dithiothreitol (DTT)	AppliChem GmbH, Darmstadt (Germany)
Ethanol	Carl Roth, Karlsruhe (Germany)
Ethylenediaminetetraacetic acid (EDTA)	AppliChem GmbH, Darmstadt (Germany)
Guanidine hydrochloride (GdnHCl) ≥ 99.7%	Carl Roth, Karlsruhe (Germany)
High precision cell	Hellma Analytics, Müllheim (Germany)
Imidazole	Carl Roth, Karlsruhe (Germany)
Isopropyl-b-D-thiogalactopyranoside (IPTG)	Carl Roth, Karlsruhe (Germany)
Lysogeny broth (LB)	Fisher Scientific, Hampton (USA)
Magnesium chloride (MgCl2)	Sigma-Aldrich, St. Louis (USA)
Monosodium phosphate	Sigma-Aldrich, St. Louis (USA)
Nickel slurry	GE Healthcare, Chicago (USA)
N,N,N',N'-tetramethyl ethylenediamine (TEMED)	VWR Funding Inc, Westchester (USA)
Page Blue Protein Staining Solution	Thermo Fisher Scientific, Waltham (USA)

Phenylmethylsulfonylfluorid (PMSF)	Sigma-Aldrich, St. Louis (USA)
SOC medium	New England Biolabs, Ipswich (USA)
Sodium dodecyl sulfate (SDS)	Sigma-Aldrich, St. Louis (USA)
Sodium chloride (NaCl)	Merck, Darmstadt (Germany)
Sodium hydroxide (NaOH)	Sigma-Aldrich, St. Louis (USA)
Sodium monophosphate	Sigma-Aldrich, St. Louis (USA)
Sodium diphosphate	Sigma-Aldrich, St. Louis (USA)
Sodium sulfide	Sigma-Aldrich, St. Louis (USA)
Thiamine hydrochloride	Sigma-Aldrich, St. Louis (USA)
Tris(hydroxymethyl)-aminomethan (Tris)	SERVA Electrophoresis GmbH, Heidelberg (Germany)
Urea ≥ 99.5%	Carl Roth, Karlsruhe (Germany)

3.1.2 Equipment and consumables

Equipment	**Manufacturer**
Äkta pure	GE Healthcare, Chicago (USA)
Acculab-Atilon	Sartorius AG, Göttingen (Germany)
Amicon Ultra-15 Centrifuge Filter Ultracel-10K	Merck Millipore, Burlington (USA)
Benchtop pH Meter	Mettler-Toledo, Columbus (USA)
Big Squid	IKA, Staufen im Breisgau (Germany)
Centrifuge 5424, rotors FA-45-24-11, T29-8x50	Eppendorf AG, Hamburg (Germany)
Centrifuge 5810R, rotor S-4-104	Eppendorf AG, Hamburg (Germany)
CO8000 Cell Density Meter	WPA, Cambridge (UK)
Entris Scale	Sartorius AG, Göttingen (Germany)
Erlenmeyer flasks	Schott, Mainz (Germany)
Falcon™ Tubes 50ml, 15 ml	Corning, Corning (USA)
HiLoad™ 16/600m Superdex ™ 75 pg	GE Healthcare, Chicago (USA)
Mighty Small™ II Mini Vertical Electrophoresis Systems	Hoefer, Richmond (USA)
NanoPhotometer N60	Implen GmbH, Munich (Germany)
New Brunswick™ Innova 40	Eppendorf AG, Hamburg (Germany)
New Brunswick™ Innova 44	Eppendorf AG, Hamburg (Germany)
Pipettes 2.5 µl ,10 µl, 100 µl, 1ml	Eppendorf AG, Hamburg (Germany)

Pipette Tips	BRAND GmbH + CO KG, Wertheim (Germany)
PowerPac Basic	Bio-Rad Laboratories, Inc., Hercules (USA)
PTC-348WI Peltier Effect Temperature Control Device	Jasco Deutschland GmbH, Pfungstadt (Germany)
Roller 6 digital	IKA, Staufen im Breisgau (Germany)
Serological pipettes	Greiner Bio-One, Kremsmünster (Austria)
Shigemi tube	Wilmad, Vineland (USA)
Slide-A-Lyser Mini Dialysis Units 10,000 MWCO	Thermo Fisher Scientfic, Waltham (USA)
Sonifier 250 D	G. Heinemann, Schwäbisch Gmünd (Germany)
Sorvall Lynx 6000 Centrifuge, rotor F9-6x1000 LEX	Thermo Fisher Scientific, Waltham (USA)
Spectropolarimeter J-715	Jasco Deutschland GmbH, Pfungstadt (Germany)
Spectropolarimeter Power Supply PS-150J	Jasco Deutschland GmbH, Pfungstadt (Germany)
ThermoMixer C	Eppendorf AG, Hamburg (Germany)
Vacuum pump	Vacuubrand GmbH and Co, Olching (Germany)
Whatman™ Cellular Nitrate Membrane Filter (0:2 µm, diameter: 47mm)	GE Healthcare, Chicago (USA)
ZelluTrans	Carl Roth, Karlsruhe (Germany)

3.1.3 Primer list

Primer Name	**Sequence**	**Description**	**Manufacturer**
T7 promotor	TAATACGACTCACTATAGGG	Complementary to the T7 promotor sequence of pET21a (+); used for DNA sequencing	Sigma Aldrich (Steinheim, Germany)

T7 terminator	GCTAGTTATTGCTCAGCGG	Complementary to the T7 terminator sequence of pET21a (+); used for DNA sequencing	Sigma Aldrich (Steinheim, Germany)
K72F_1/2	TGATCAATGGGTGCAGCTGTTC CAACGTTTTCGTGAAGCGATC	Maquette 2 K72_E17 point mutant, (reverse primer exactly complementary)	Sigma Aldrich (Steinheim, Germany)
E17F_1/2	GAGCTGGTGCAGGCGTTC CAACGTCTGCGTGAGATCTTC	Maquette 2 K72_E17 point mutant, (reverse primer exactly complementary)	Sigma Aldrich (Steinheim, Germany)
210_274_1/2	CTACCCAATCCAATTAGTCGTC AAGCCTTTGTGGACTAGAAACC	Hsp90 full length/ NM-domain Δlinker, (reverse primer exactly complementary)	Sigma Aldrich (Steinheim, Germany)

3.1.4 Expression plasmids

Name of plasmid	**Description**	**Manufacturer**
Hsp82 full length	Expression vector, Spectinomycin^r	inhouse cloned
Hsp82$_{fl}$ Δlinker	Expression vector, Spectinomycin^r	inhouse cloned
NM-domain Hsp82	Expression vector, Can+Cam^r	inhouse cloned
NM-domain Hsp82 Δlinker	Expression vector, Can+Cam^r	inhouse cloned
pET21a+_target_gene	Expression vector, Amp^r	Genscript
pET21a+_Lys49	Expression vector, Amp^r	Genscript
pET21a+_Glu49	Expression vector, Amp^r	Genscript

pET21a+_bundle_1ionpair	Expression vector, Amp^r	Genscript
pET21a+_1ip_switched	Expression vector, Amp^r	Genscript
pET21a+_bundle_2ionpairs	Expression vector, Amp^r	Genscript
pET21a+_bundle_lysglu_3ionpairs	Expression vector, Amp^r	Genscript
pET21a+_bundle_glulys_3ionpairs	Expression vector, Amp^r	Genscript
pET21a+_bundle_3ip_allhelices	Expression vector, Amp^r	Genscript
pET21a+_bundle_3ip_allhelices_GLN	Expression vector, Amp^r	Genscript
pET21a+_bundle_shorter_loops	Expression vector, Amp^r	Genscript
pET21a+_sl_lys49	Expression vector, Amp^r	Genscript
pET21a+_sl_glu49	Expression vector, Amp^r	Genscript
pET21a+_SL_1ip_mutPhe	Expression vector, Amp^r	Genscript
pET21a+_SL_2ip_mutPhe	Expression vector, Amp^r	Genscript
pET21a+_SL_3ip_mutPhe	Expression vector, Amp^r	Genscript
pET21a+_SL_3ip_mutPhe_GLN	Expression vector, Amp^r	Genscript
pET21a+_degrado_WT	Expression vector, Amp^r	Genscript
pET21a+_ F72K	Expression vector, Amp^r	inhouse cloned
pET21a+_ F17E	Expression vector, Amp^r	inhouse cloned
pET21a+_ F72K_F17E	Expression vector, Amp^r	Genscript
pET21a+_ E17_K72_N69	Expression vector, Amp^r	Genscript
pET21a+_K50_E101	Expression vector, Amp^r	Genscript
pET21a+_ Dutton_long_2ip	Expression vector, Amp^r	Genscript

3.1.5 Enzymes

Enzyme	Description	Manufacturer
DpnI	Restriction enzyme, digests methylated DNA	Agilent Technologies, Santa Clara (USA)
Lysozyme	Used for cell lysis	Carl Roth, Karlsruhe (Germany)
TEV	Restriction enzyme for removing His-tag	In house produced
Quik Polymerase	Polymerase for QuikChange Mutagenesis	Agilent Technologies, Santa Clara (USA)

3.1.6 Protein standard

Name	Manufacturer
Serva Dual Color Protein Standard 3	SERVA Electrophoresis GmbH, Heidelberg (Germany)

3.1.7 Kits

Kit	Manufacturer
Wizard® Plus SV Minipreps DNA Purification System	Promega, Madison (USA)
QuikChange Lightening Mutagenesis	Agilent Technologies, Santa (USA)

3.1.8 Bacterial strains and growth media

Bacterial strains

Name	Manufacturer
E. coli BL21 DE3	Agilent Technologies, Santa Clara (USA)
E. coli XL 10 Gold	Agilent Technologies, Santa Clara (USA)
E. coli NEB5α	New England Biolabs, (USA)

Growth media

Lysogeny broth (LB) medium	Tryptone	1% (w/v)
	Yeast extract	0.5% (w/v)
	NaCl	0.5% (w/v)
SOC medium	Peptone	2% (w/v)
	Yeast extract	0.5% (w/v)
	Glucose	20 mM
	$MgSO_4$	10 mM
	NaCl	10 mM
	KCl	2.5 mM
M9 minimal medium	Na_2HPO_4	42 mM
	KH_2PO_4	22 mM
	NaCl	8.5 mM
	NH_4Cl	18 mM
	$MgSO_4$	2 mM

	$CaCl_2$	0.1 mM
	Glucose	2% (w/v)
	Biotin	1 mg
	Thiamine	1 mg
	Trace elements (1000x)	1 ml

3.1.9 Buffers

Hsp90 full length constructs

Components	**Description**
50 mM Tris, 300 mM NaCl, 5 mM Imidazole pH 8	Cell lysis
	Equilibration buffer
	Dialysis after TEV cleavage
+ 1000 mM NaCl	Washing buffer
+ 500 mM Imidazole	Elution buffer
50 mM Tris, 50 mM NaCl, 1 mM EDTA, 2 mM DTT pH 8	TEV cleavage
Buffer A: 20 mM Tris pH 8	Anion exchange chromatography
Buffer B: 20 mM Tris, 1000 mM NaCl pH 8	Anion exchange chromatography
40 mM HEPES, 150 mM KCL, 2 mM $MgCl_2$ pH 7.5	SEC
	SAXS

NM-domain Hsp90 constructs

Components	**Description**
50 mM Tris, 150 mM NaCl, 10 mM Imidazole pH 8	Cell lysis
	Equilibration buffer
	Dialysis after SUMO cleavage
	Reverse Ni-column
+ 1000 mM NaCl	Washing buffer
+ 500 mM Imidazole	Elution buffer
50 mM Tris, 150 mM NaCl, 2 mM DTT pH 8	SUMO cleavage
Buffer A: 20 mM Tris pH 8	Anion exchange chromatography
Buffer B: 20 mM Tris, 1000 mM NaCl pH 8	Anion exchange chromatography
40 mM HEPES, 150 mM KCL, 2 mM $MgCl_2$ pH 7.5	SEC
	SAXS

***Maquette* 1 and 3 constructs**

Components	Description
20 mM Tris, 100 mM NaCl pH 8	Cell lysis
	Ni-column
20 mM Tris, 50 mM NaCl, 1 mM EDTA pH 8	SEC
20 mM Tris, 50 mM NaCl, 1 mM EDTA, 2 mM DTT pH 8	TEV cleavage
20 mM Tris, 50 mM NaCl pH 8	Dialysis after TEV cleavage
	Reverse Ni-column
	CD measurements
20 mM HEPES, 300 mM NaCl, 0.5 mM EDTA pH 7	Thermal unfolding, CD measurements
20 mM NaPi, 50 mM NaCl, 0.5 mM EDTA pH 7	NMR measurements

***Maquette* 2 constructs**

50 mM NaPi, 150 mM NaCl pH 7	Cell lysis
	Ni-column
	SEC
50 mM Tris, 1 mM EDTA pH 8	Dialysis after SEC
50 mM Tris, 1 mM EDTA, 2 mM DTT pH 8	TEV cleavage
50 mM NaPi, 100 mM NaCl pH 7.5	Dialysis after TEV cleavage
	Reverse Ni-column
	CD measurements
	NMR measurements

3.1.10 Software and programs

Name	Manufacturer
Microsoft Office	Microsoft, Redmond (USA)
Multalign	http://multalin.toulouse.inra.fr/multalin (France)
Python	Python Software Foundation (Netherlands)
Protparam Expasy	https://web.expasy.org/protparam/
Reverse Complement	https://www.bioinformatics.org/sms/rev_comp.html
Serial Cloner	http://serialbasics.free.fr/Serial_Cloner.html
Sparky 3.135	[103]
TopSpin 3.5	Bruker, Massachusetts (USA)

3.2 Experimental methods

3.2.1 Transformation

E. coli cells were thawed on ice and 1 µl plasmid (~ 30 – 100 ng μl^{-1}) was added. The mixture was incubated for half an hour on ice. The incubation was followed by a 30 s long heat shock at 42°C. After heat shock, 300 µl SOC medium were added and the cells were incubated for 1 hour at 37°C and 150 rpm. Different types of *E. coli* cells depending on the application were used. BL21 DE3 were used for pre-cultures for expression while XL10 were used for plasmid amplifications. NEB5α were used for transformation of PCR products. In this case 3 µl of PCR reaction were added to the cells. After incubation, the cells were plated on agar plates containing the required antibiotic or transferred in LB medium containing 2% of glucose and the required antibiotic. The plates were incubated at 37°C and the pre-cultures in the shaker at 37°C and 150 rpm overnight.

3.2.2 Culturing of prokaryotic cells

LB medium containing the required antibiotic and 2% glucose was used for 50 ml pre-cultures and for 10 ml cultures for plasmid isolation. Ampicillin (100 µg ml^{-1}) was used for pET21a(+) constructs. The cells were cultured overnight at 37°C and 150 rpm.

3.2.3 Plasmid amplification

The transformation was performed as described under 3.2.1 using XL10 Gold cells. A 10 ml overnight culture was used for plasmid amplification containing 2% glucose and the required antibiotic. The plasmid isolation was conducted following the manual of the Wizard® Plus SV Minipreps DNA Purification System kit described in section 3.2.13.

3.2.4 Protein expression in BL21 (DE3) cells

LB or M9 minimal medium was used for expression depending on if a labeled protein expression was performed. Expression cultures contained the required antibiotic (100 µg ml^{-1}) and in case of labeled expression $^{13}C_6$ D-glucose and ^{15}N ammonium chloride. The cultures were inoculated 1:40 with the overnight pre-culture. When the cells reached an OD_{600} of 0.6-0.8 at 37°C, the protein over expression was induced by adding 0.5 mM IPTG. After 4 hours of expression at 37°C the cells were harvested by centrifugation at 7822 x g for 20 min. Afterwards, the pellet was washed with buffer (20 mM Tris, 50 mM NaCl, pH 8) at 3214 x g for 20 min. Cell pellets were stored at -80°C.

3.2.5 Cell lysis

First step in cell lysis was to slowly thaw the pellet on ice and resuspend it in lysis buffer. 1 mM PMSF and a spatula tip of lysozyme was added to the homogenic solution. The mixture was incubated for 30 min at + 4°C. Next step was sonication for 30 min. The used program was 1 s pulse with an amplitude of 30% and 2 s pause. After cell disruption, 5 mM $MgCl_2$ and 1 µl/10 ml lysis solution DNase I (stock 10 U μl^{-1}) were added and incubated again for 30 min at + 4°C. The cell solution was centrifuged for 20 min at 38,828 x g to pellet the cell fragments. The supernatant contained the soluble protein and was ready for next purification steps.

3.2.6 Ni-NTA and reverse Ni-NTA

A Ni-NTA affinity chromatography was performed during purification to separate the overexpressed protein from *E. coli* own proteins. After cell lysis, the supernatant was loaded onto the Ni-NTA. Due to the His-tag at N- or C-terminus of the protein, the protein stayed attached to the Ni slurry. Before loading the Ni-NTA an equilibration with buffer was performed. After loading, the column was incubated for 1 h at +4°C. A wash step with one column volume (CV = 20 ml) was performed with 10 mM imidazole to remove weakly bound proteins with natural His-tags before 400 mM imidazole were added to the buffer to elute the protein. The elution was again one CV.

After the TEV cleavage a so-called reverse Ni-column was performed to remove the cut His-tag and the TEV protease from the protein solution. In this case the protein was in the flow-through fraction.

The columns were regenerated after usage with a solution of 6 M GdnHCl and 0.2 mM acetic acid (one CV) and two CV water. They were stored in 20% ethanol at +4°C.

3.2.7 TEV digestion

A TEV digestion site was placed between the protein of interest and the His-tag to make the His-tag removable after protein purification. The TEV protease recognize ENLYFQ-G/S and cuts between QG or QS. The buffer during TEV cleavage need to contain 2 mM DTT and 1 mM EDTA. The digestion was performed at + 4°C overnight. Afterwards, EDTA and DTT were removed from the buffer before loading it onto the Ni-NTA to remove the His-tag and TEV protease. To this end, a dialysis in the target buffer was performed overnight.

3.2.8 SUMO cleavage

A SUMO cleavage site was placed between the NM-domain of Hsp90 and the His-tagged SUMO fusion protein. The protease recognizes the tertiary structure of SUMO and cleaves very specific. The required buffer for the optimal reaction was 50 mM Tris pH 8, 150 mM NaCl and 2 mM DTT. The digestion took place at +4°C overnight.

3.2.9 Size exclusion chromatography (SEC)

After running the Ni-NTA, a size exclusion chromatography on an ÄKTA pure system was performed. Due to the size and amount of the proteins a HiLoad 16/600 Superdex 75 pg column was used. Before the actual SEC run was started, the column was equilibrated using the buffer in which the protein should be eluted. The protein was automatically injected to the column from a 5 ml capillary loop. The flow rate was 1 ml min^{-1}. During elution the absorption was monitored at 280 nm and 2 ml fractions were collected. After elution, the column was washed with one CV water and stored in 20% ethanol.

The peaks were integrated to calculate the molecular weight (MW) of the proteins in Dalton (Da). Therefore, a calibration curve (*Figure 13*) was performed and the following formula for converting was obtained:

$$\log(MW) = \frac{K_{av} - 2.540905}{-0.50338} \qquad K_{av} = \frac{V_e - V_0}{V_c - V_0}$$

where V_c is the column volume (= 120 ml), V_e is the elution volume, and V_o is the dead volume of the column (= 44.3 ml).

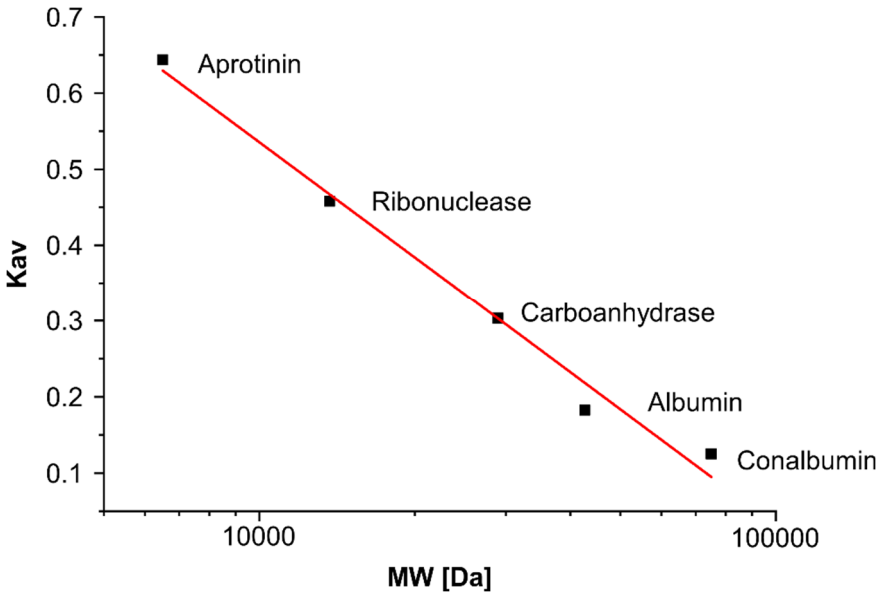

Figure 13: Calibration curve for S75p column. Proteins with different sizes were used for calibration.

3.2.10 Anion exchange chromatography

An anion exchange chromatography was necessary for the Hsp90 full length and the NM-domain of Hsp90. This step was performed before the final SEC run. To this end, the protein was dialyzed against 5 l of 20 mM Tris pH 8 to completely remove the salt to enable protein binding to the column. The column was equilibrated with 20 mM Tris pH 8 (buffer A) and the protein was loaded on the column with a superloop. After the protein bound to the column material, a washing step with buffer A was conducted. Afterwards, the salt concentration was slowly increased up to 1 mM NaCl by mixing buffer A with buffer B (20 mM Tris pH 8, 1 M NaCl).

3.2.11 Circular dichroism (CD) spectroscopy

The secondary structure of the different constructs was analyzed by circular dichroism (CD) spectroscopy. The measurements were performed on a Jasco J-715 system with a cooling-heating unit using a 1-mm-path quartz cuvette. The protein concentrations were between 5 and 10 µM. To analyze the secondary structure five spectra from 190 nm to 260 nm were measured at 20°C. Temperature sensitive unfolding was measured at 222 nm from 20 to 100°C to calculate the melting point.

The CD raw data in mdeg was converted according to following formula to mean residue molar ellipticity θ:

$$\theta = \frac{CD\ [mdeg]\ x\ 10^6}{1\ [mm]\ x\ c_{protein}\ [\mu M]\ x\ peptide\ bonds}$$

where CD is the CD raw signal, 1 [mm] is the path length of the cuvette, $c_{protein}$ is the protein concentration in µM, and peptide bonds is the number of peptide bonds.

3.2.12 Chemical unfolding

The thermodynamic stability was measured by following the chemical unfolding on a Jasco J-715 system with two denaturants: GdnHCl and urea. The measurements were performed at 20°C with denaturant concentrations up to 8 M using 1-mm-path quartz cuvette. The protein concentrations were between 5 and 10 µM. The unfolding was monitored by detection of the CD signal at 222 nm for 30 s. The raw data was converted according to 3.2.11 and plotted against denaturant concentration, which results in a sigmoid shaped unfolding curve. Each unfolding series was measured in triplicates to get statistics and the obtained error bars were added to the plots. The turning point (= unfolding concentration) was determined by fitting it to a sigmoid function. To obtain the unfolding energy, ΔG, ln(K) was calculated and plotted against the denaturant concentration. ΔG is defined as -RTln(K). The equilibrium constant K was defined as $\frac{[D]}{[N]}$ and was obtained from the unfolded protein fraction

$$fD = \frac{y - yN}{yD - yN}$$

where y is the CD raw signal and yN and yD corresponds to native and denatured protein state. ΔG in kcal mol^{-1} was determined by using a linear least-square function and extrapolation to 0 M denaturant.[104]

3.2.13 Site-directed mutagenesis

Point mutations were introduced by using the QuikChange Lightning Mutagenesis Kit. To this end, primers with a length of about 20 bases in front of and after the point mutation were designed and ordered from Sigma Aldrich. Up to 50 ng DNA plasmid were used per reaction. The total reaction volume was 25 µl consisting of 2.5 µl Quik Buffer, 0.75 µl Quik Solution, 0.65 µl primer forward and reverse, 0.5 µl dNTP mix, 0.5 µl Quik Polymerase and x µl plasmid. The missing volume was filled up with autoclaved water. The table below shows the PCR program which was used for mutagenesis.

Step	Temperature	Term	Cycle number
Initial denaturation	95°C	2 min	
Denaturation	95°C	30 s	
Annealing	67°C	30 s	25 cycles
Elongation	68°C	5 min	
Finale elongation	68°C	10 min	
Storing	10°C	∞	

After the PCR reaction was finished a transformation using NEB5α cells were performed according to section 3.2.1.

3.2.14 SDS-PAGE

For visualizing proteins and controlling purity during the purification process a denaturing SDS-gel was run. The gels consisted of a 6% stacking gel and a 15% separating gel. The samples were mixed with 2x Lämmli buffer. The running buffer was 1x SDS buffer and the Serva Dual Color Protein Standard 3 served as marker. 50 mA per gel were used while the running time was 30 min. After cooking the gel in water Page Blue Protein Staining Solution was added and let incubate for 30 min. The gels were destained in water.

3.2.15 Miniprep

The plasmid DNA isolation was performed accordingly to manufacturer's instructions. The first step included spinning down 5 to 10 ml of an overnight culture by centrifuging it at 20238 x g for 5 min at room temperature. The supernatant was removed and the pellet was resuspended with 250 µl of cell resuspension buffer. The solution was then transferred to a fresh Eppendorf tube and 250 µl of cell lysis solution were added. After mixing the solution, 10 µl of alkaline phosphatase were added and the suspension was mixed and incubated for 5 min at room temperature. To fell the cell fragments, 350 µl of neutralization solution were added and the mixture was centrifuged at 20238 x g for 10 min. The supernatant was transferred to the DNA binding column and after a 1 min centrifugation step, the flow through was discarded. The column was washed with 750 µl of washing solution containing ethanol. After 1 min of centrifugation at 20238 x g the wash step was repeated with 250 µl. Last step before elution was 2 min centrifugation at full speed to remove the rest of the washing solution and transferring the column to a fresh Eppendorf tube. The DNA was eluted with

60 µl of nuclease-free water by centrifuging it for 1 min at 20238 x g. Plasmid DNA was stored at 20°C.

3.2.16 Concentration measurements

A NanoPhotometer N60 was used for concentration measurements. 1 µl of protein or DNA solution was applied to Nanodrop. Water was used for blanking when DNA concentration at 260 nm was measured. Protein concentration was detected at 280 nm and the corresponding buffer served as blank. The concentrations in M were calculated with the corresponding extinction coefficient (ε) and accordingly in mg ml^{-1} with the molecular weight (MW) in Dalton (Da).

3.2.17 Sequencing of plasmid DNA

The sequencing of plasmid DNA was performed by Eurofins or Genewiz. Total sample volume was 10 µl per reaction for Genewiz and 17 µl for Eurofins. 2/2.5 µl sequencing primer were added to 50-100 ng $µl^{-1}$ DNA. For all pET21a (+) constructs which have a T7 promotor, the T7 promotor or terminator primer was used. The rest volume was filled up with nuclease-free water.

3.2.18 Gene synthesis and cloning

The starting constructs for all three *Maquette* series were ordered from Genscript from the amino acid sequence. A His-tag was added to N- or C-terminus of the different proteins with a TEV cleavage site between tag and protein to have to the possibility to remove the His-tag. After the protein sequence a stop codon was added. Genscript performed codon optimization for expression in *E. coli*. The synthesized gene was cloned into pET21a (+) via NdeI and HindIII or BamHI and HindIII. The constructs with BamHI and HindIII have an 11 amino acid long T7 tag attached to the N-terminus which is not removable. The plasmids arrived lyophilized and were solved in nuclease-free water. Plasmids were stored at -20°C.

3.2.19 Nuclear magnetic resonance (NMR) spectroscopy and structure calculation

The NMR measurements were performed on a Bruker Avance III spectrometers operating at a ^{1}H Larmor frequency of 600 MHz (14.09 T) and 950 MHz (22.31 T) using a CPTCI triple-resonance cryoprobe. The $^{1}H^{15}N$ HSQCs and 3D experiments for the assignment were performed at 25°C in a 5 mm tube. The following experiments were used for the assignment of backbone and side-chains: 3D HNCA, HN(CO)CA, HNCO, HN(CA)CO, HNCACB, CBCA(CO)NH, HBHA(CO)NH,

H(CCO)NH, CC(CO)NH, HC(C)H/(H)CCH TOCSY, $^{15}N/^{13}C^{ali}/^{13}C^{aro}$-edited $^{1}H,^{1}H$ NOESY (120 ms mixing time) as well as 2D (HB)CB(CGCD)HD. 2D $^{1}H,^{15}N$ HISQCs were measured at 2°C in a Shigemi tube to detect $^{15}NH_3$-groups of lysines. The protein concentration for *Maquette* 1 models was 400 µM and for *Maquette* 2 models 800 µM. The proteins were dissolved in the buffers described in section 3.1.9. 0.1% trimethylsilylpropanesulfonic acid sodium salt (DSS) was used as standard for referencing the spectra for structure solving. The raw data was processed with Bruker TopSpin Software. For the following analysis and the assignment Sparky[103] was used. CARA[105] was utilized for peak picking. For the structure calculation CYANA[106,107] was applied to obtain automated NOE assignments. These generated restraints were used to calculate a *de novo* structure of 1ip (Lys72-Glu17) *Maquette* 2. TALOS-N[108] was used to get additional distance and dihedral angle restraints from chemical shifts. For energy minimization of the bundle of 20 structures, OPALp and the AMBER force field were used.[109,110] Distance limits between Glu17-OE1 and Lys72-NZ were restrained to calculate the closed ion-pair structure. The distance limits between 2.8 Å and 3.0 Å were obtained based on MD simulations (see below).

3.2.20 X-ray crystallography

The sitting drop vapor diffusion method was used for crystallization of *Maquette* 3. The crystals were grown at 20°C. The protein concentration was 40 mg ml^{-1} in 20 mM Tris pH 8.0, 50 mM NaCl buffer. One drop consisted of a 1:1 mixture of protein solution and reservoir buffer which contained 150 mM HEPES buffer pH 7.5 and 1.6 M di-potassium hydrogen phosphate. The crystals were measured at the Paul Scherrer Institute, Switzerland (λ = 1.0 Å). The structure calculation and the refinement were conducted with REFMAC5[111] and COOT[112].

3.2.21 Force profile analysis (FPA)

To study the protein folding under *in* vivo conditions, we designed constructs for force profile analysis (FPA).[101,102] The general construct structure is visualized in *Figure 14*. A random coil linker of different length between 21 and 35 amino acids and the arrest peptide MsSup1 (HPPIRGSP[113]) were added to the C-terminus of the protein of interest. Each data point in the force profile is a unique genetic construct. The length *L* is defined as the total length in amino acids between the C-terminus of the model protein and the final residue of the arrest peptide sequence. After the arrest peptide a C-terminal tail derived from a soluble loop in the LepB protein (*E. coli*) consisting of 23 residues was added to allow the arrested and the full-length protein products to be separated on an SDS-PAGE gel.

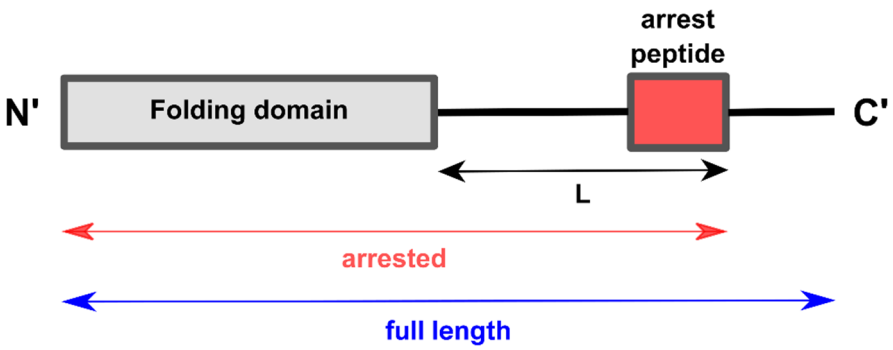

Figure 14: Schematic representation of the constructs for FPA.

The constructs are cloned into a pET21a+ vector (Genscript) and the T7 open reading frame is amplified by PCR. This PCR product, which provide the only template for transcription/translation, along with ^{35}S-methionie is added to the *in vitro* PURExpress™ (New England Biolabs) according to the manufacturer's instructions. While PURExpress contains several proteins, all newly synthesized protein is radiolabeled allowing sensitive and specific detection of only our model protein. For each construct two products were produced: a short "arrested" product that is produced when the force is acting on the nascent chain; and a longer "full-length" product that was produced when force is taking place. The latter effect can result from too short construct, where the domain cannot efficiently fold; or too long construct, for which the domain has already folded before the arrest peptide is translated, and does not result in a pulling force acting on the chain, resulting in the arrested product. The linker lengths where robust folding occurs, result in an increased full-length product. These two products can be resolved by SDS-PAGE and the amounts quantified to give a fraction full-length (f_{FL}), which is a proxy for the probability that protein folds during the translation process. A higher f_{FL} value indicates a higher probability of protein folding. With f_{FL} plotted against the linker length, L, allows to derive the point in translation with the highest probability of folding. SDS-PAGE gel bands were visualized with a phosphoimager (Fuji FLA9000) and the intensities were measured from the TIFF image using the ImageJ software. The intensities were quantified by fitting with Gaussian curves using the in-house developed EasyQuant (Rickard Hedman, Gunnar von Heijne, Stockholm University). The force profile f_{FL} was calculated as $\frac{I_{FL}}{I_{FL}+I_A}$, where I_{FL} is the intensity of the band on the gel of the full-length protein and I_A is the intensity of the band of the arrested protein.

3.3 Computational methods

3.3.1 Molecular mechanics and biomolecular force fields

In common biomolecular force fields used in molecular dynamics (MD) simulations, the energy function of the system is calculated as a sum of bonded and non-bonded interactions.

$$V = V_{bonded} + V_{non-bonded}$$

The bonded and non-bonded interactions are divided into different terms.

$$V = V_{bond} + V_{angle} + V_{dihedral} + V_{improp} + V_{VdW} + V_{elec}$$

Each potential is a sum over molecular interactions depending on the molecular composition and structure. k_x is the corresponding harmonic force constant, while b, θ, ϕ, and φ describe geometrical properties in the following equation given below.

The sum of bond stretching, angle bending and dihedral rotation energies represents the bonded term and describes the intramolecular interactions. Bond stretching and angle bending terms depend on the connection and the angle between the atoms. These potentials are modeled as harmonic in most force fields and oscillate around the equilibrium. The rotation around bonds in a molecule is described by dihedral potentials. In typical force fields the potential energy function is an expansion of periodic functions. Many force fields also consider out-of-plane motions known as improper torsions, where not all atoms are connected covalently.

$$V_{bond} = \sum_{bonds} k_b(b - b_0)^2 \qquad V_{angle} = \sum_{angles} k_\theta(\theta - \theta_0)^2$$

$$V_{dihedral} = \sum_{torsions} k_\phi[cos(n\phi + \delta) + 1)] \qquad V_{improper} = \sum_{impropers} k_\varphi(\varphi - \varphi_0)^2$$

The non-bonded interactions consist of van der Waals (VdW) and electrostatic interactions. The latter describe the intermolecular electrostatic contribution to the energy of the system, while the VdW term describes the non-bonded interactions without a permanent electric dipole moment. To describe the VdW interactions most of the force fields use the Lennard-Jones-potential.[114] The potential is divided into repulsive ($1/r^{12}$) and attractive ($1/r^6$) interactions, where σ is the interatomic distance where the potential is zero, and ε is the depth of the energy minimum. The VdW energy rapidly decays with distance, and therefore normally cut-off distances of around 10 Å can be used for reducing the computational cost. Interactions at larger distances are ignored for the potential energy calculations and a pair list for the calculations is generated and updated.[115]

$$V_{VdW} = \sum_{i,j\neq 1} \varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \qquad V_{elec} = \sum_{i,j\neq i} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}}$$

In contrast, interactions with a non-zero electric moment are represented by the electrostatics term. The electrostatic interaction between two atoms is described by Coulomb's law where a partial charge is assigned to each atom. The point charges are fixed in common force fields. In this case, the interaction energy decreases as $1/r$. There are two possibilities on how to computationally deal with such long-range contributions. First, one can choose a large cut-off distance, but the cost of required computational resources increases rapidly and becomes unfeasible for larger systems, and the neglect of long-range contribution leads to significant underestimation of the electrostatic effects. The second and commonly used way in infinite periodic systems is to use the Particle-mesh Ewald (PME) summation, where the scaling is $N\log N$ instead of N^2 and N is the total number of atoms. In the PME approach, the electrostatic potential is divided into short- and long-range interactions. Here, the energies are calculated in real (short-range) and reciprocal (long-range) space.[116]

Chemistry at HARvard using Molecular Mechanics (CHARMM) is one of the commonly used all-atom force field for biomolecules, where all atoms also hydrogens are incorporated. There are also other force fields where only polar hydrogens are included explicitly.[117] The CHARMM force field was developed in the early 1980s[118] and was stepwise improved during the following decades. The first version without explicit hydrogen atoms was followed by CHARMM19. In CHARMM19, hydrogen atoms bonded to nitrogen and oxygen atoms were explicitly incorporated. Further improvement of the force field parameters ended up in the CHARMM22 version with a protein parameter set of 55 defined atom types.[117] The CHARMM27 force field provides almost the same parameters for proteins, but the focus of iteration was laid on improving the parameters for nucleic acids.[119] Most of the simulations in this thesis were performed using the latest CHARMM force field version CHARMM36.[120] This version is refined for proteins regarding the backbone and the side-chain dihedral parameters.[121]

3.3.2 Molecular dynamics (MD)

During the last decades Molecular dynamics (MD) simulations have become a powerful method to understand biological processes on an atomic level. MD simulations provide molecular insight into phenomena that are difficult to study experimentally. Another advantage is the application for protein design and interpretation of experimental results. MD simulations can today be carried out by several commonly available implementations, *e.g.* CHARMM[122] or NAMD[123]. In this thesis work, NAMD was used for classical MD simulations of the designed proteins.

In MD simulations, Newton's second law of motion is integrated over all atoms.[122]

$$F_i = m_i a_i = -\nabla V_i = m_i \frac{d^2 r_i\,(t)}{dt^2}$$

where F is the force acting on the particle i, m is its mass, and a is the acceleration. V is the potential, in this case the force field which is acting on the particles. Newton's law connects the motion of a particle with the forces acting on them. The force field parameters relate to the acceleration, velocity and position of the atoms. The position of all particles at any time t' can be calculated, if the position at time t and the affecting forces are known. Newton's second law of motion is propagated using discrete timesteps, because the large number of atoms in an MD simulation would make the analytical calculation too complex. The forces and energies are only calculated at defined time points depending on the chosen size of the timestep. The equation of motion is expanded as a Taylor series which leads in combination with the *Verlet* algorithm to[124]

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + \delta t^2 a(t) + O(\delta t^4)$$

The coordinates in the next time step $t+\delta t$ are calculated using the actual position $2r(t)$, the position in the previous timestep $r(t-\delta t)$, and the force acting on the particle $\delta t^2 a(t) + O(\delta t^4)$.

The size of the time step for the simulation is important: The time step should not be too large as this may lead to a decreased stability of the system, but on the other hand, the reduced computational cost, could allow for longer the total simulation time should be increased. On the other hand, a short time step improves the accuracy, but also demands more computing time. It is important to find a compromise between the length of the time step and the stability of the system. The time step of choice should be around 10 times smaller than the fastest motion in the system, such as the X-H bond vibrations (10^{-14} s). Therefore, it is common to use a 1-2 fs timestep for biomolecules. If the investigation of hydrogen vibrations is not of interest, algorithms such as the SHAKE algorithm can be used, where the vibrations of hydrogen atoms are frozen.[125] This algorithm allows a 2 fs timestep without losing too much accuracy. MD simulations today are normally limited to µs time scales due to these small timesteps and large number of pair-wise interactions that must be evaluated during each timestep. Moreover, some larger systems that occur at larger time scales demand for different simulation methods, such as coarse-grained simulation methods.

Another often employed method to decrease the computing time is to use Langevin dynamics where stochastic dynamics are applied. A continuum model of the solvent, which interacts with the protein, is used. The temperature and the pressure are controlled by Langevin dynamics during the MD simulation. A friction coefficient γ and a random force with a mean of zero are added to the equation of motion.[126–128]

Another important aspect is the treatment of system boundaries. Normally, Periodic Boundary Conditions (PBC) are used in MD simulations to avoid surface artifacts. The simulation system is modeled in a way that the water box is a unit cell of an infinite lattice consisting of identical replicas like an ideal crystal.[129] With PBC a more realistic *in vivo* like environment is modeled due to using surrounding virtual systems. The PBC are used for estimating long-range electrostatic effects in the PME summation.

3.3.3 Classical molecular dynamics (MD) simulations of the designed bundle proteins

Classical MD simulations of the designed proteins were performed with NAMD using the CHARMM27 or CHARMM36 force field. The temperature was set to 310 K and a 1 or 2 fs integration timestep was used. Before the simulations were initiated the designed protein was solvated. Therefore, a water box consisting of TIP3P water molecules[130], was placed within 20 Å around the protein. The water box also contained neutralizing ions which corresponds to a strength of 0.15 M NaCl. The solvation and ionizing step were performed with a tcl-script using VMD. After the system was solvated the first step was a 0.1 ns minimization of the solvent with fixed protein. The relaxation of the solvent was followed by a protein relaxation (0.5 ns) where the Cα atoms of the protein were kept fixed and only the side-chains could move. After the minimization, the MD production simulations with freely moving protein were performed. The simulation lengths varied between 100 ns and 1 µs.

Construct	**System size in atoms**
Maquette1	
hydrophobic core	43,912
Lys49	43,919
Glu49	43,908
K49_E84	43,915
E49_K84	43,915
2ip	43,912
3ip	43,903
3ip_switched	43,903

3ip_all_helices	43,904
3ip_all_helices_E112prot	46,742
3ip_all_helices_Q112	46,743
sl_wt	43,645
sl_lys49	43,652
sl_glu49	43,641
sl_1ip	47,392
sl_2ip	47,401
sl_3ip	43,626
sl_3ip_E112prot	43,629
sl_3ip_E112	51,168
Maquette 2	
hydrophobic core	66,932
Lys72	65,039
Glu17	65,028
E17_K72	44,429
K50_E101	51,953
E17_K72_N69	48,059
Maquette 3	
2ip/4Q	289,717

3.3.4 PACMAN

The in-house PACMAN python code was used for stabilizing the introduced ion-pairs. To this end, random mutations were introduced followed by a short MD simulation step (10 ps) in explicit solvent. The mutations rated by a score considering the following terms: self-interaction between the ion-pair (weight = 0.48), ion-pair interaction with the surrounding protein (weight = 0.48), and all non-bonded interactions within the protein (weight = 0.02). Depending on the Monte Carlo-score, a mutation was accepted or rejected. The cycle was started again if a mutation was rejected. The models with an accepted mutation were minimized for 5000 steps followed by a 5 ps long restrained relaxation. The protein backbone was harmonically restrained with a force constant of 1 kcal mol^{-1} $Å^{-2}$. The CHARMM36 force field was applied to model the interactions.[120] The MD simulations were performed with NAMD using a 1 fs time step.[123]

3.3.5 Visualization and analysis of simulations

Visual molecular dynamics (VMD)[131] was used for visualization and analysis of MD simulations. The integrated root-mean-square deviation (RMSD) trajectory tool was used for analyzing the convergence of the simulations. Different representations of the models were created with VMD for showing simulation results and creating figures.

4. Results

This results chapter is divided in four sections, which are below. In the first part, the conformational influence of charged residues is analyzed in the natural protein Hsp90. In the second part, the artificial helical bundle protein systems are used to design and characterize the influence of buried charged residues. The design and characterization of *Maquette* 1, *Maquette* 2, and *Maquette* 3 models are described. Molecular dynamics simulations and experimental characterization are presented in a combined manner for each model.

4.1 Hsp90 conformational dynamic via the charged-linker region

Hsp90 undergoes large conformational changes during the chaperone cycle that are coupled to its ATPase activity, but the structural intermediates have not yet been fully characterized. Therefore, the conformational change related to the opening of the Hsp90 dimer that involves the charged linker region connecting the NTD and M-D is investigated here. It was detected from SAXS measurements that the experimentally observed atom-pair distances (D_{max}) of the Hsp90 dimer are not consistent with the value calculated from the crystal structure (PDB ID: 2CG9) (*Figure 15a*). For the evaluation of the crystal structure, the missing linker region in the crystal structure was modeled *in silico*, and a D_{max} of 120 Å and a radius of gyration (R_g) of 41 Å was detected.[30] The larger D_{max} of up to 250 Å and the R_g with > 60 Å measured in the SAXS experiments indicate that there exist several highly extended conformation states during the Hsp90 cycle. A possible idea was that these extended conformations result from the separation of the N- and M-domains which are connected by the charged linker. Coarse-grained simulations, performed by Alexander Jussupow (Kaila Lab), were used to investigate these conformational dynamics on a larger timescale. In these simulations initiated from the crystal structure of the closed conformation of the Hsp90 dimer, a separation of the two NTDs was observed after 10 µs that leads to a more extended, V-shaped conformation (*Figure 15b*). However, these findings do not explain the molecular dimensions observed in the SAXS experiments. Additionally, in the coarse-grained simulations, a separation of the N- and M-domain via the charged linker was observed. To test if this observation is the reason for the extended conformation detected in the SAXS experiments, we studied the influence of the charged linker region on the Hsp90 conformation, by engineering a monomeric Hsp90-NM-construct. The SAXS measurements were performed by Dr. Abraham Lopez (Sattler Lab).

To investigate the effect of the linker experimentally, the amino acids forming the linker were fully deleted by QuikChange mutagenesis in the NM-construct (ΔLinker), *i.e.* residues 211 to 273 were cut out. The construct was expressed in good yields and was purified as described in the method section. *Figure 16* shows an SDS gel during the purification process. The protein is detected at the expected size of ~ 54 kDa after cleavage of the fusion protein. *Figure 17* presents the experimental

results for the pair distance distribution (*P(r)*) and the radius of gyration (R_g). Indeed, the D_{max} is significantly smaller for the NM-domain without the charged linker (*Figure 17*, light orange) as compared to the experimental results of the wild type NM-construct (*Figure 17*, orange). During the coarse-grained simulations, the N- and M-domains separate via the linker region and both partially and fully open conformations are detected. Since these deletion studies were successfully applied to the monomer NM-domain, the idea of such extended conformations was also tested with full length (fl) Hsp90. The $Hsp90_{fl}$ mutant without the charged linker was successfully expressed and purified. *Figure 18* shows an SDS-PAGE gel after the final SEC. The protein runs at the correct size and shows a high purity. A similar effect as for the NM model is detected and $Hsp90_{fl}$ ΔLinker shows a reduced R_g (~ 60 Å) and D_{max} (~ 233 Å) in SAXS measurements (*Figure 19*).

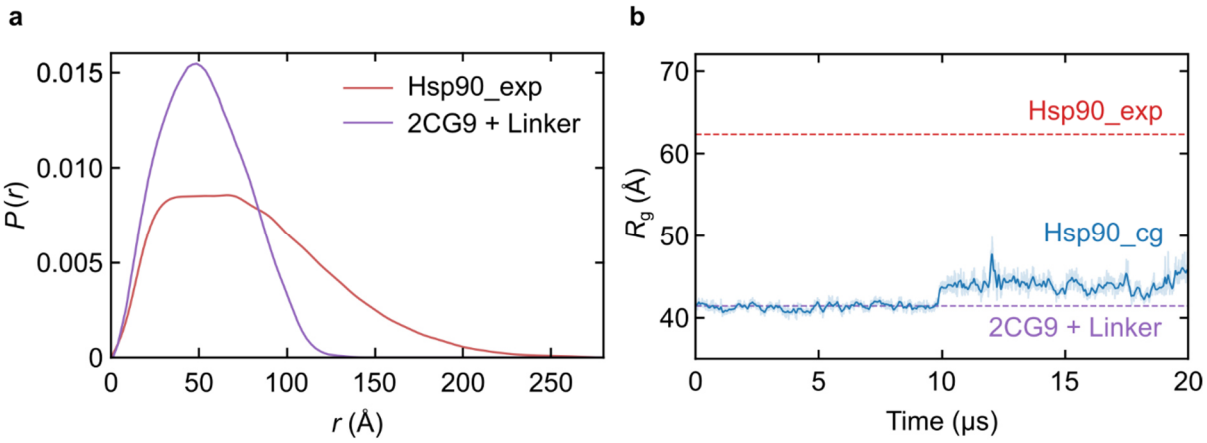

Figure 15: a, pair distance distribution (*P(r)*) of full length Hsp90 from SAXS experiments (red) and from the crystal structure with *in silico* modeled linker (purple); **b**, radius of gyration (R_g) of full length Hsp90 from experiments (red), coarse-grained simulations (blue), and crystal structure (purple).

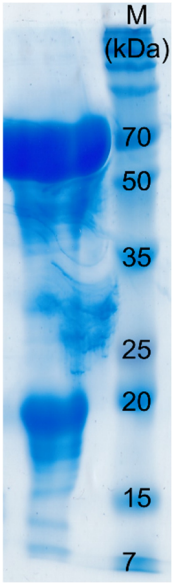

Figure 16: SDS gel of NM-domain without charged linker (ΔLinker) after expression at the beginning of the purification process. The strongest band around 70 kDa corresponds to the fusion protein.

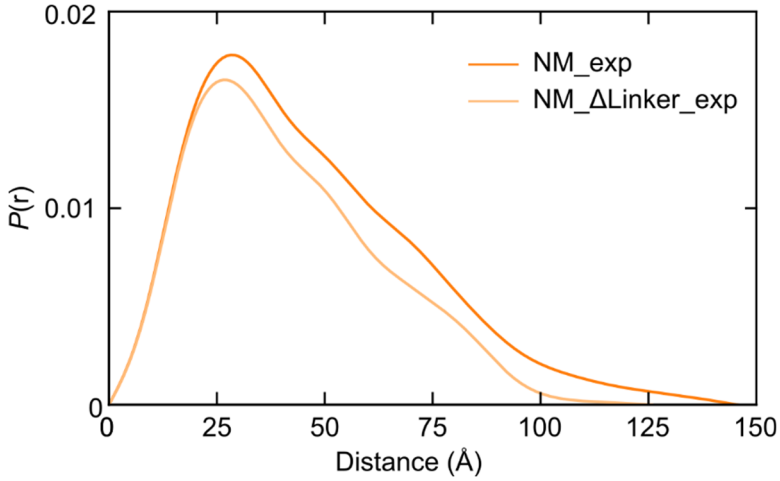

Figure 17: Pair distance distribution (*P(r)*) of NM-domain. The experimental result is shown in orange, while NM-domain without linker is light orange.

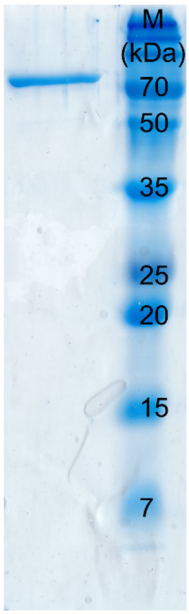

Figure 18: SDS gel of Hsp82 full length without linker (ΔLinker) used for the SAXS measurements at the end of the purification process.

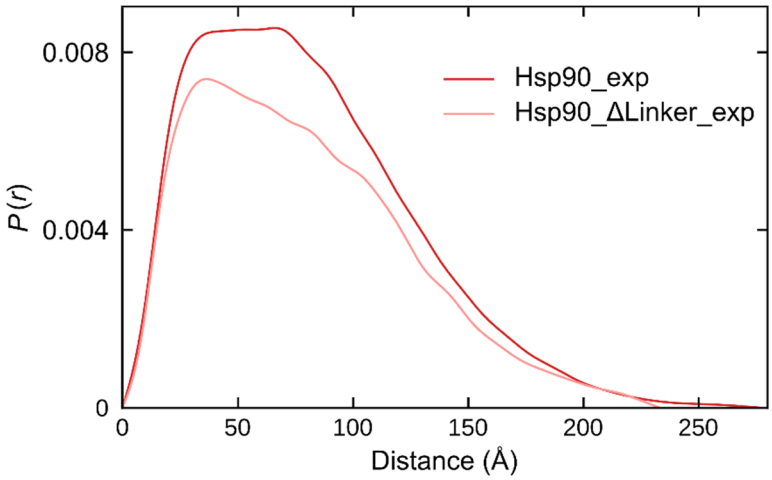

Figure 19: Experimental pair distance distribution (*P(r)*) of full length Hsp90 with (red) and without linker (coral).

4.2 *Maquette* 1

To construct an initial model for creating a protein conformational switch, we adapted the sequence of a *Maquette* scaffold, model E, as described by Farid *et al*.[12] The construct used here is named *Maquette* 1 in the following sections. The initial atomistic model was created *in silico* with first atomic coordinates modeled using CHARMM (with Dr. Ana Gamiz Hernandez, Kaila Lab).[122] To obtain the desired α-helical secondary structure restraints were placed on the residues 1-27 (helix 1), 36-62 (helix 2), 71-97 (helix 3), and 106-130 (helix 4) according to the PsiPred[132] secondary structure prediction. The packing of the helices had the goal to create a hydrophobic core and a hydrophilic surface according to known computational modelling principles.[133] After generating atomic coordinates, the structure was relaxed with constraints. The constraints were removed stepwise during the relaxation process. To maximize the non-polar interactions restraints on Phe15 and Phe21 were placed to form a π-stacking interaction. After this first relaxation process, the connecting loops were placed followed by a new restrained minimization step. After the relaxation process was finished the model was simulated in vacuum to check the stability. The following MD simulations in water-ion surroundings were performed using NAMD[123] and the CHARMM36[120] force field at 300 K using an 1fs timestep.

Mutations were introduced stepwise into *Maquette* 1. The mutant series is described in the following subsections. After 100 ns simulations for each construct, the RMSD, RMSF, water content and the ion-pair distance were analyzed. The constructs were expressed and purified experimentally, and their stability were characterized by CD spectroscopy. For the hydrophobic core constructs, single buried charges, and one ion-pair models chemical unfolding with GdnHCl were performed to calculate the protein stability.

4.2.1 Hydrophobic core

The first section introduces the hydrophobic core and starting structure for the first mutant series. *Figure 20* shows the results of 100 ns MD simulation of the *Maquette* 1 model. The RMSD value stabilizes after 10 ns of simulation until the end (*Figure 20a*) indicating that the system has converged. The RMSF (*Figure 20b*) represents the dynamic fluctuation of the residues, where one can see that the three loop regions (residues 28-36; residues 63-71; residues 98-106) and the N and C termini are more flexible as compared to the four α-helices and reach RMSF values around 5 Å as compared to around 1 Å for the helices. There is low increase in hydration during the MD simulation indicating a more or less water sealed hydrophobic protein core. *Figure 21* presents the first (grey) and the last simulation snapshot (blue) of the 100 ns simulation, where the helices do not show significant structural differences. Only the flexible loops undergo some conformational changes.

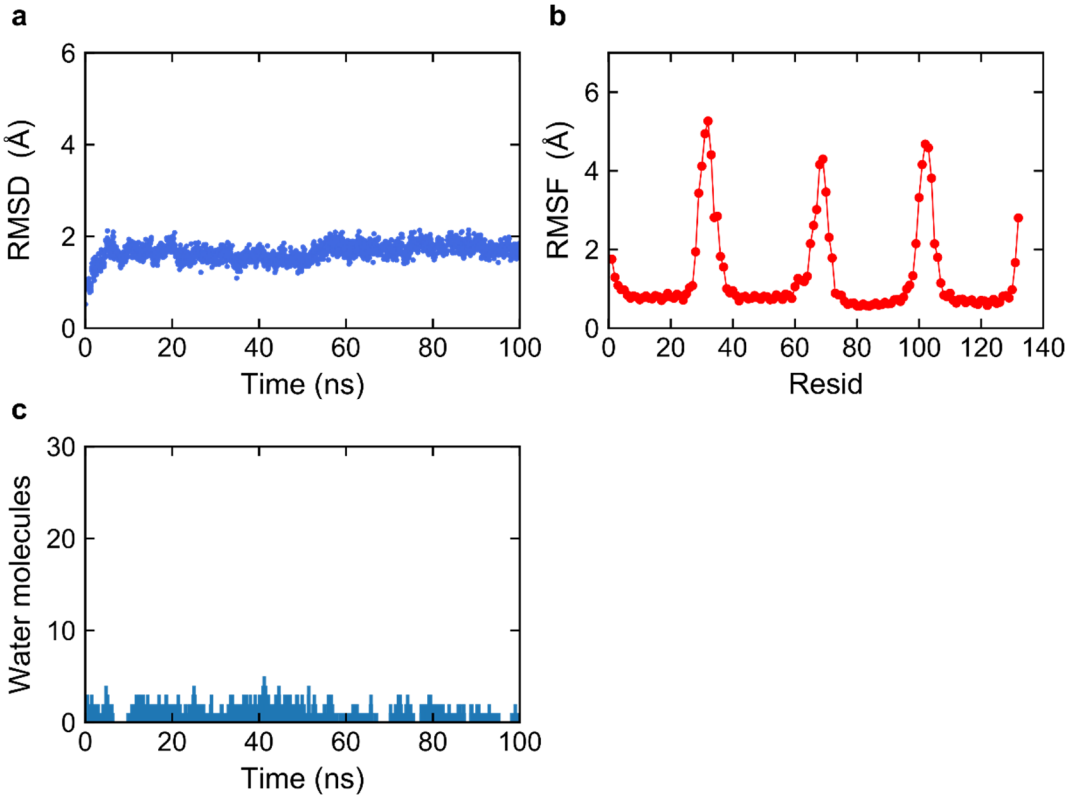

Figure 20: Analysis of 100 ns MD simulation of hydrophobic core *Maquette* 1. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content.

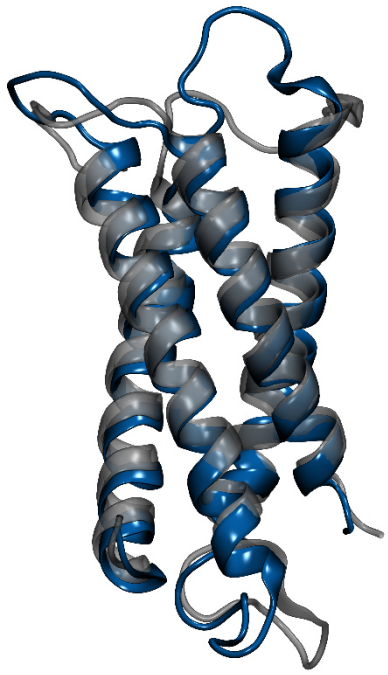

Figure 21: Snapshots from 100 ns MD simulation of hydrophobic core *Maquette* 1. The starting structure (grey) is aligned with the last frame (blue) of the simulation.

The following section focuses on the experimental data and *Figure 22a* shows an SDS-PAGE gel during the expression and purification process of the hydrophobic core protein. The marked band proves the successful expression of the designed protein. The protein is already very pure after the first affinity chromatography round. Afterwards, a size exclusion chromatography (SEC) step was performed showing that most of the protein is monomeric, but also a small dimer peak is observed. The protein elutes a bit too early and the calculated corresponding molecular weight (MW) is 22 kDa instead of the expected 17 kDa based on the protein sequence. This could be an effect of the highly negatively charged surface of the bundle. The monomer fractions are collected and pooled to conduct the CD spectroscopic analysis and the chemical unfolding experiments.

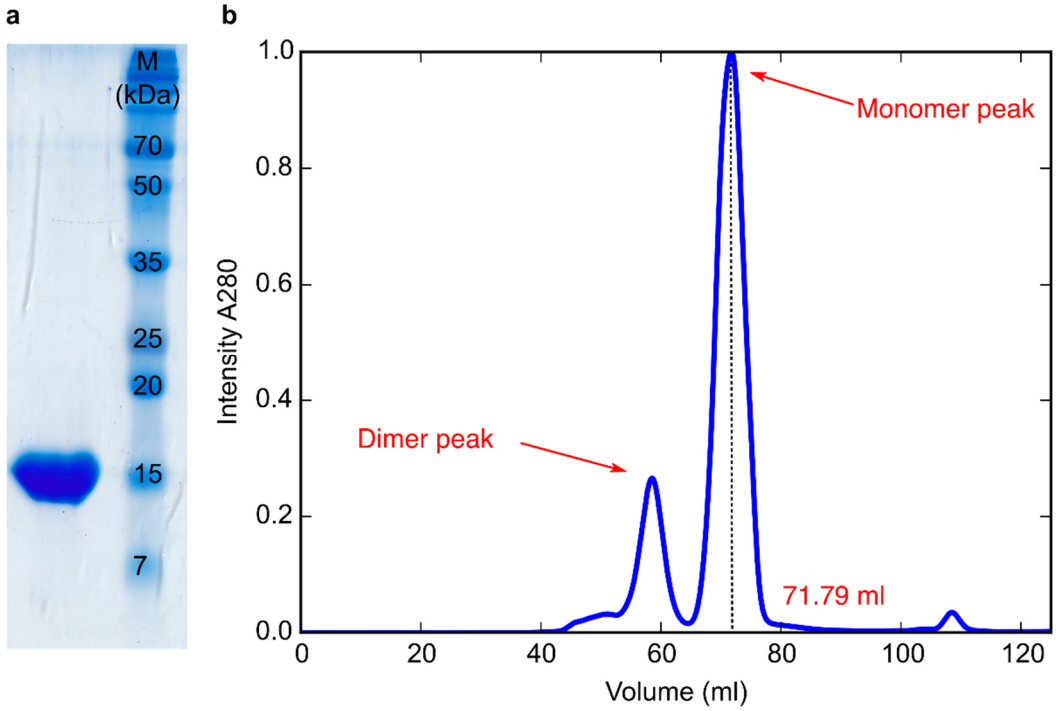

Figure 22: a, SDS gel of hydrophobic core *Maquette* 1; **b**, size exclusion chromatography (SEC) profile of hydrophobic core *Maquette* 1. The major peak corresponds to monomeric protein, but there is also a small dimer peak. The monomer has a retention volume of 71.79 ml, which corresponds to a molecular weight of 22 kDa.

We analyzed the secondary structure, thermal stability, and chemical unfolding using CD spectroscopy (*Figure 23*). The measured values and curves were processed and analyzed as described in section 3.2.11 and 3.2.12. The hydrophobic core construct has a well-defined α-helical secondary structure showing the typical minima at 208 and 222 nm when measuring the molar ellipticity between 190-260 nm (*Figure 23a*). The bundle is a very temperature robust protein since it does not show melting up to 100°C following the CD signal at 222 nm (*Figure 23b*). Also, the chemical unfolding experiment using GdnHCl indicates that the protein is highly stable and the unfolding is detected at 5.4 M GdnHCl with a calculated ΔG of 8.8 ± 0.7 kcal mol^{-1}.

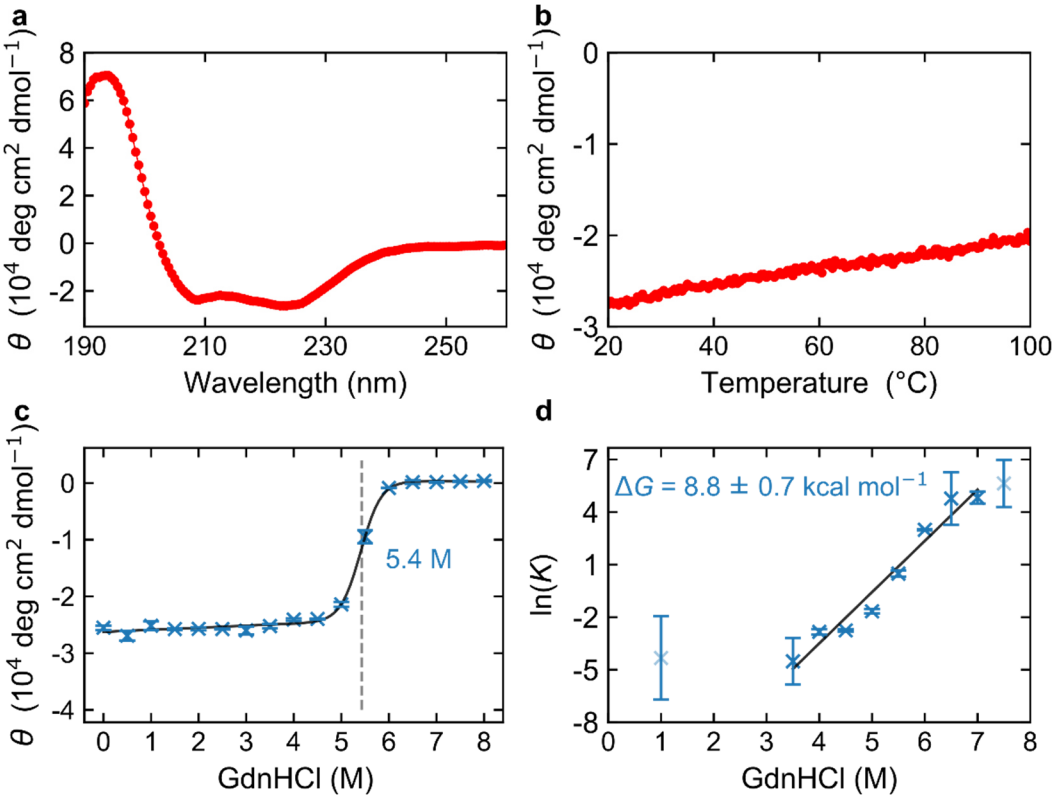

Figure 23: Experimental analysis of hydrophobic core *Maquette* 1. **a**, CD spectrum; **b**, melting curve; **c**, chemical unfolding with GdnHCl; **d**, ln*K* plotted against denaturant concentration.

To summarize the computational and experimental results, it is shown that the hydrophobic core *Maquette* 1 bundle is a very stable, robust α-helical protein, which overexpresses and behaves experimentally well.

4.2.2 Introduction of a single charged residue – glutamate at position 49

After the stability of the hydrophobic core was analyzed, a single charged residue was inserted into the protein core. For the first mutated construct a glutamate was placed at the position 49, which is located directly in the hydrophobic core. The phenylalanine at position 49 was mutated after the hydrophobic core model is converged during the MD simulation. *Figure 24* summarizes the results of a 100 ns long MD simulation. The RMSD is around 4 Å after 10 ns and slowly increases until it goes down a bit after 90 ns (*Figure 24a*). The protein structure does not fully converge during the MD, as can be observed from the small drift of the RMSD. The RMSF plot shows that, the loops and termini are again more flexible as compared to the α-helices (*Figure 24b*). The low content of buried water molecules implicates that the system remains almost water sealed (*Figure 24c*). Two snapshots from the simulation are shown in *Figure 25* where the first frame shows Glu49 pointing towards the hydrophobic core (grey) while helix 2 with Glu49 is slightly rotated in a way that the Glu49 side-chain moves towards the water during the simulation (blue). The other helices do not undergo significantly structural changes.

The results of the purification process described in the method section are not reported here in detail, but the protein showed a high purity on the SDS-PAGE gel and has a retention volume of 67.04 ml during SEC corresponding to a MW of 31.8 kDa. The MW is closer to the dimer (34 kDa) as to the monomer MW (17 kDa). The point mutation leads most likely to some structural changes which results in the dimerization of the protein. *Figure 26* presents the results of the experimental analysis which was performed with the dimeric protein. The secondary structure is characterized by the two typical α-helical minima in the CD spectrum (*Figure 26a*) and the melting curve indicates a thermal transition of the protein structure at around 67°C (*Figure 26b*). The chemical unfolding experiment leads to unfolding at 3.4 M GdnHCl and yields a ΔG of 3.2 kcal mol^{-1}.

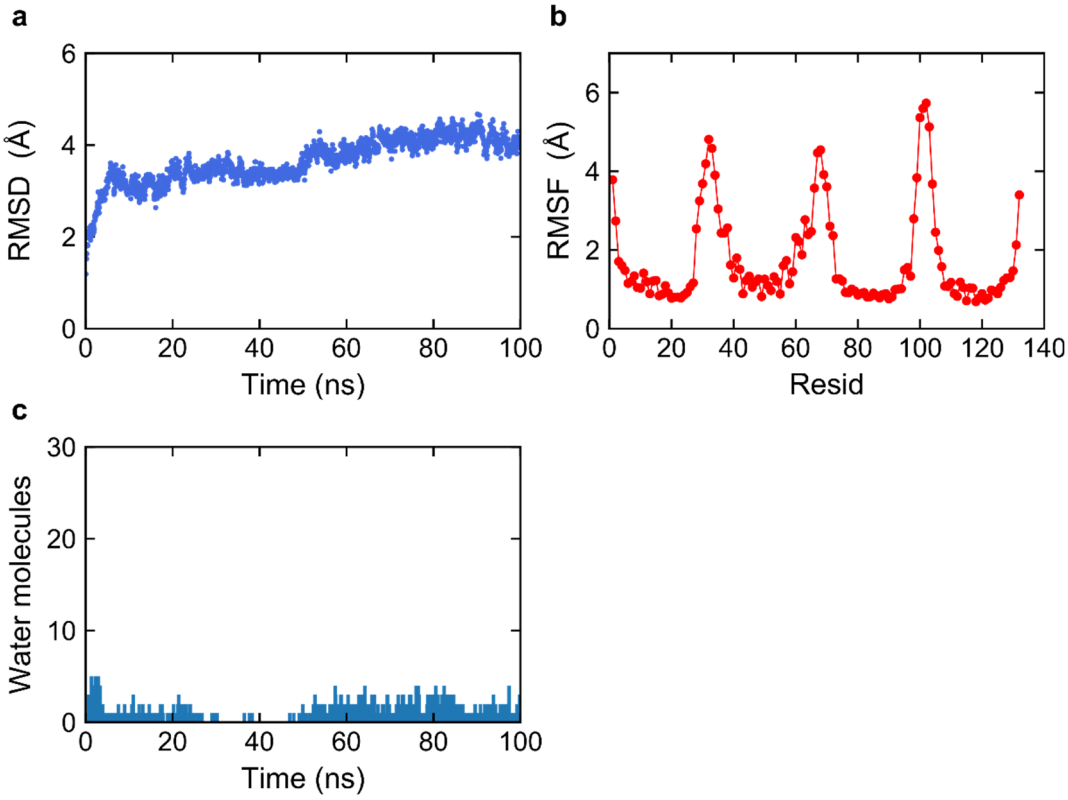

Figure 24: Analysis of 100 ns MD simulation of Glu49 *Maquette* 1. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content.

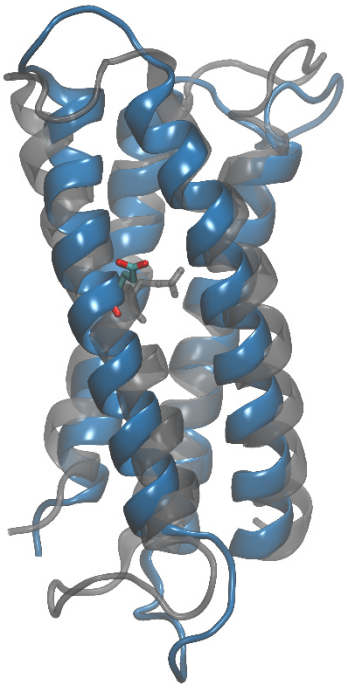

Figure 25: Snapshots from 100 ns MD simulation of Glu49 *Maquette*1. The starting structure (grey) is aligned with the last frame (blue) of the simulation. The Glu49 is shown in licorice.

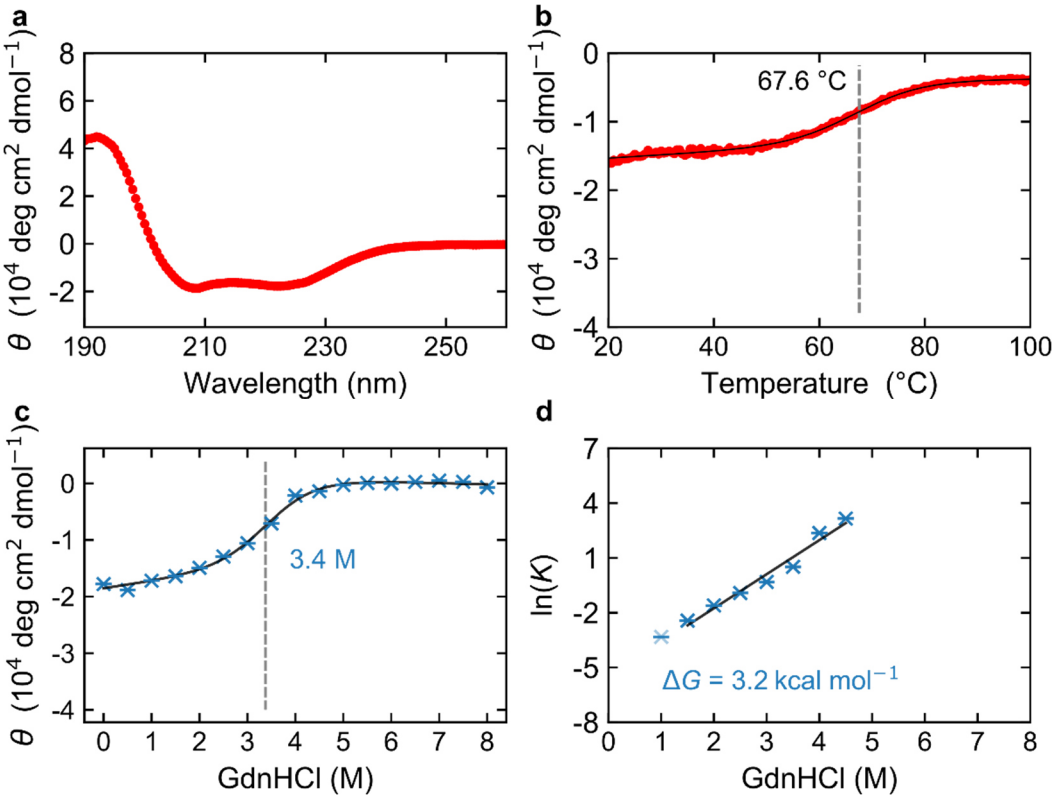

Figure 26: Experimental analysis of Glu49 *Maquette* 1. **a**, CD spectrum; **b**, melting curve; **c**, chemical unfolding with GdnHCl; **d**, ln*K* plotted against denaturant concentration.

Summarizing the result for the buried charged residue construct, we find in the simulation that the protein partially compensates for the penalty by partially moving Glu49 from the hydrophobic core towards the water-soluble part of the surface. This mutation and the subsequent structural changes also lead to a partial dimerization of the protein. Nevertheless, the protein still remains folded with a high stability.

4.2.3 Introduction of a single buried charged residue – lysine at position 49

For the *Maquette* 1 model, we also probed the effect of substituting phenylalanine at position 49 with a lysine to generate another construct with a single buried charged residue. The RMSD of this protein during 100 ns MD simulations fluctuates around 4 Å after the first 10 ns (*Figure 27a*). The RMSF plot indicates a higher flexibility in the loops as compared to the helices and the N-terminus of the protein shows the largest movement reaching 6 Å (*Figure 27b*). The *Maquette* 1/Lys49 mutant remains still water sealed and the hydrophobic core remains intact. *Figure 28* presents two snapshots from the simulation, with the starting structure colored in grey and the last frame in blue. Lys49 points into the hydrophobic core of the bundle in the starting structure, but during the 100 ns simulation, the side-chain of Lys49 moves partially towards the water due to a small helix rotation to compensate for the buried charge, whereas the other helices remain at similar positions.

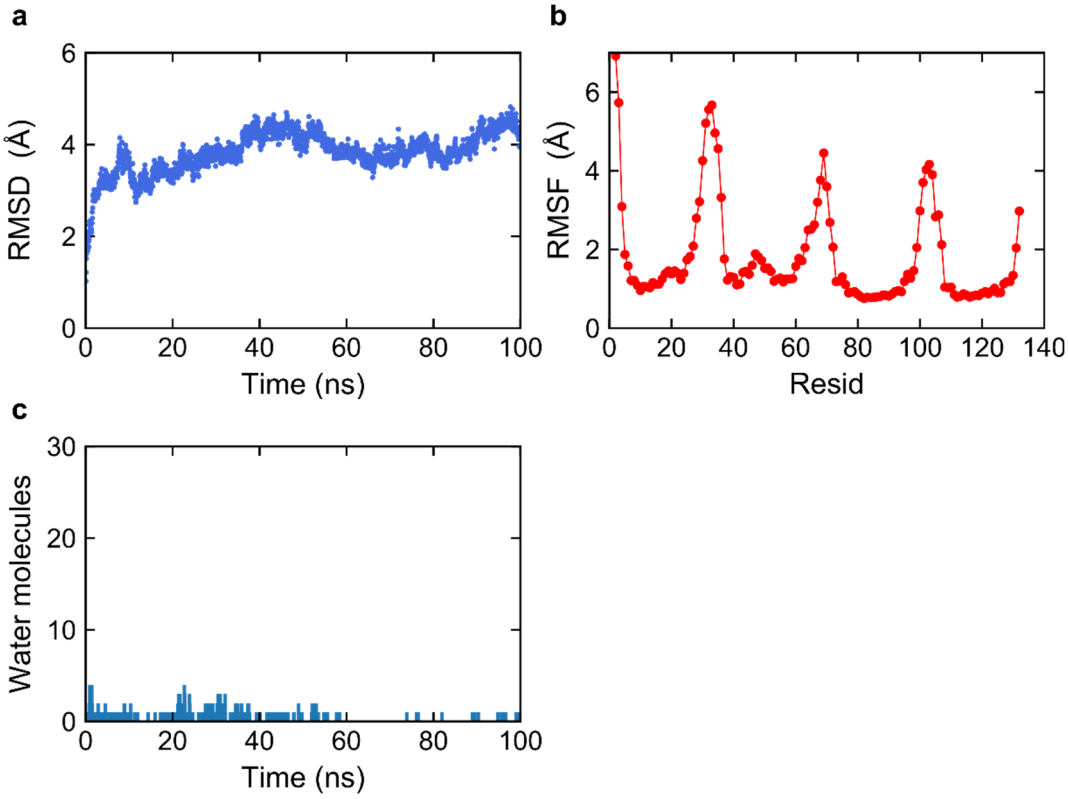

Figure 27: Analysis of 100 ns MD simulation of Lys49 *Maquette* 1. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content.

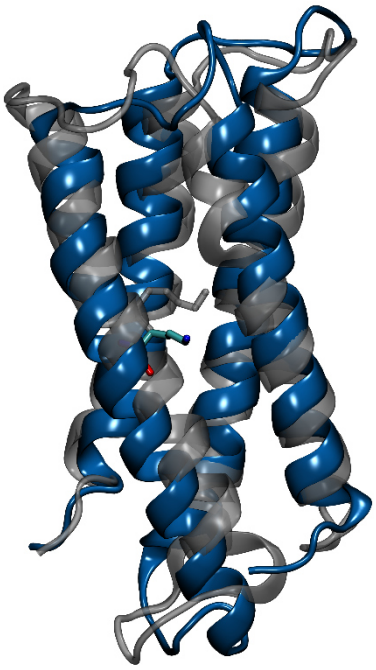

Figure 28: Snapshots from 100 ns MD simulation of Lys49 *Maquette* 1. The starting structure (grey) is aligned with the last frame (blue) of the simulation. Lys49 is shown in licorice.

The data of the purification process is not shown here, but the very pure protein has a retention volume of 72 ml during SEC which corresponds to a MW of 24.8 kDa. The calculated MW is between the expected weight of the monomeric (17 kDa) and dimeric (34 kDa) protein. *Figure 29* shows the experimental analysis where the CD spectrum proves a well-defined α-helical secondary structure (*Figure 29a*) and the melting curve presents a two-state unfolding with the transition at 72.4°C (*Figure 29b*). The chemical unfolding measurement results in a turning point at 3.2 M GdnHCl and a corresponding Δ*G* value of 3.4 kcal mol^{-1}.

The Lys49 *Maquette* 1 behaves very similar to the *Maquette* 1/Glu49 construct with no significant differences detected.

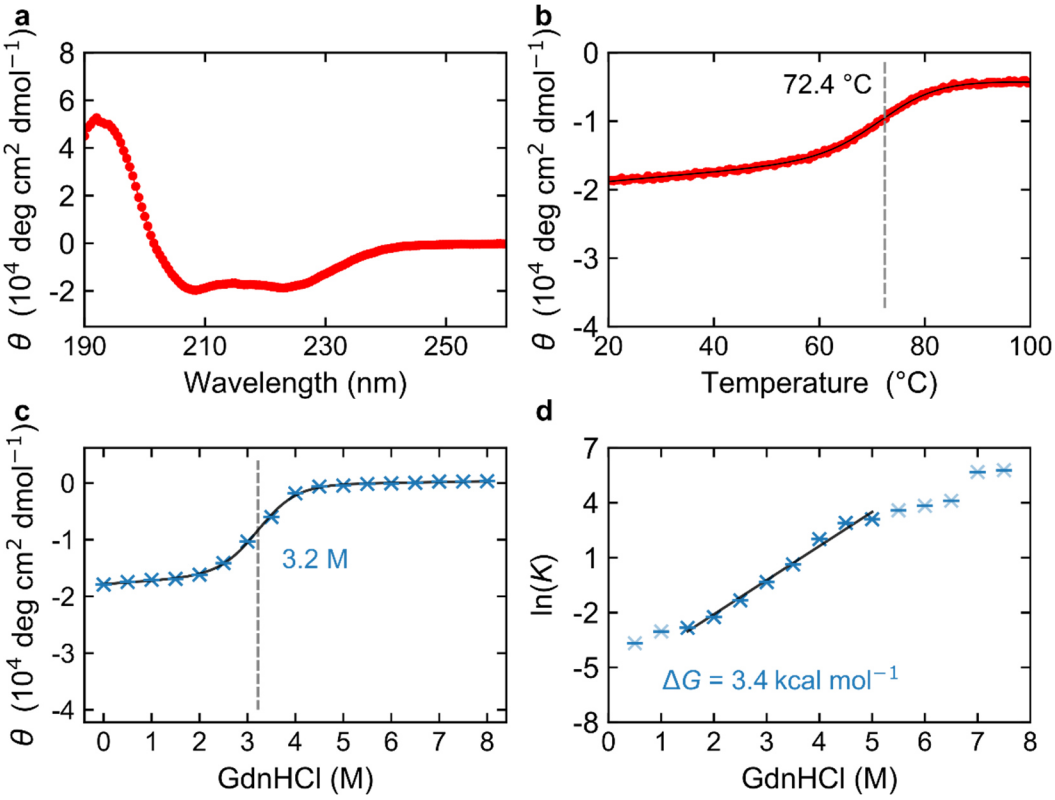

Figure 29: Experimental analysis of Lys49 *Maquette* 1. **a**, CD spectrum; **b**, melting curve; **c**, chemical unfolding with GdnHCl; **d**, ln*K* plotted against denaturant concentration.

4.2.4 Introduction of a buried ion-pair (Lys49-Glu84)

As a next step towards creating a protein conformational switch, a buried ion-pair was introduced into *Maquette* 1. The two phenylalanines at positions 49 and 84 were mutated to a lysine and a glutamate residue, respectively. The obtained ion-pair (1ip) construct was simulated for 100 ns that shows a stable RMSD around 2 Å after 10 ns (*Figure 30a*). The loops show slightly less movement as compared to the constructs with a single charged residue described above (*Figure 30b*), but as expected, the loops are dynamically more flexible than the helices. The protein interior has a maximum of 11 buried water molecules along the complete protein scaffold, and higher as compared to the previous described constructs with around 3-5 water molecules (*Figure 30c*). The CD-NZ distance between Lys49 and Glu84 indicates a fluctuating ion-pair (*Figure 30d*), but remains in a closed conformation during most of the simulation. *Figure 31* provides two snapshots from the simulation with the starting frame in grey and the ending frame in blue. The ion-pair is represented in licorice and is closed in both frames while the overall protein structure does not change significantly.

The protein shows a high purity and has a MW of 32.4 kDa determined by SEC. The MW indicates that the construct is between a monomer and a dimer, but closer to the dimer. The secondary structure is α-helical as shown by CD spectroscopy (*Figure 32a*) and the melting temperature is decreased to 61°C for (*Figure 32b*) from > 100°C for the hydrophobic core. Also, the turning point detected by chemical unfolding is reduced to 2.1 M GdnHCl with a Δ*G* of 2.7 kcal mol^{-1}.

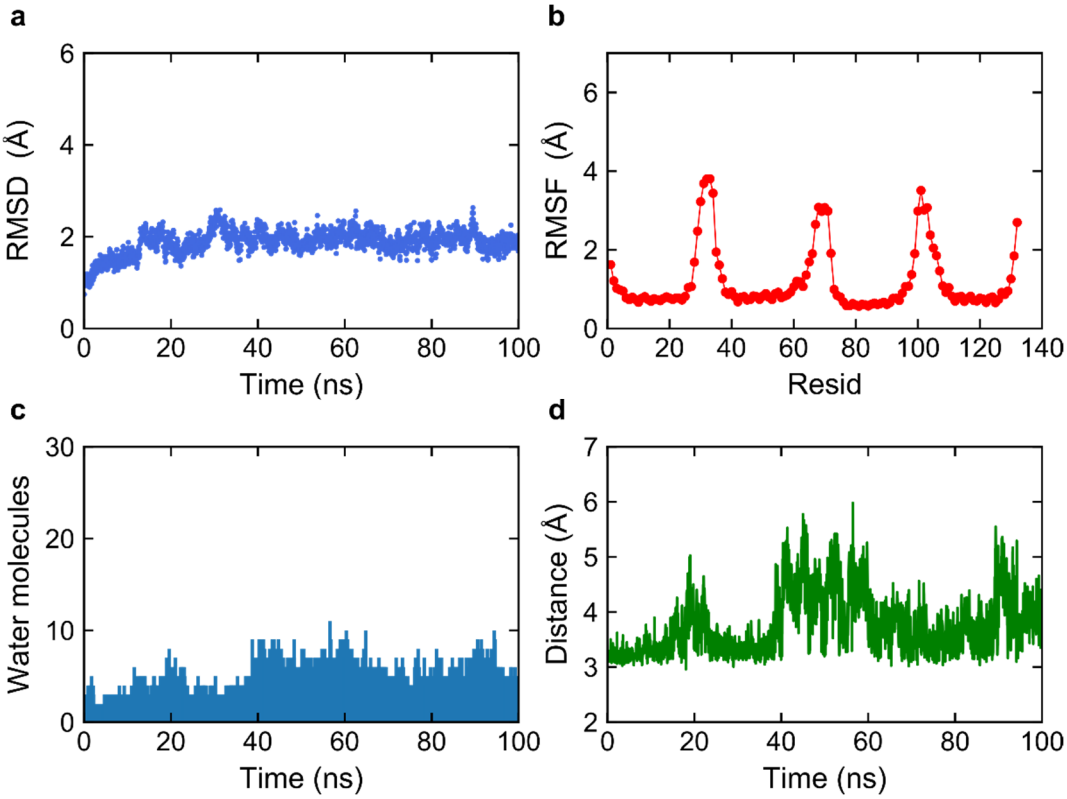

Figure 30: Analysis of 100 ns MD simulation of 1ip (Lys49-Glu84) *Maquette* 1. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content; **d**, distance profile between CD-NZ of Glu84 and Lys49.

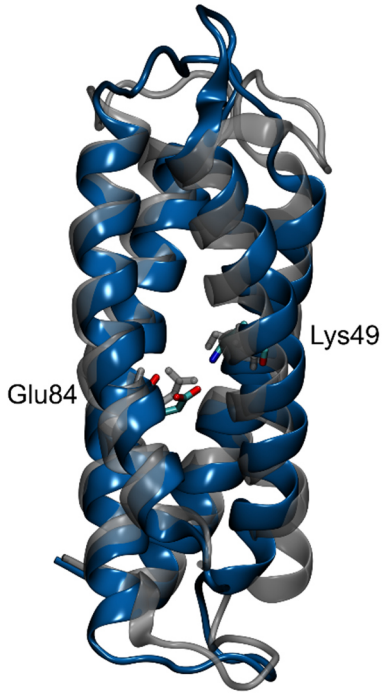

Figure 31: Snapshots from 100 ns MD simulation of 1ip *Maquette* 1. The starting structure (grey) is aligned with the last frame (blue) of the simulation. The Lys49 and Glu84 are shown in licorice.

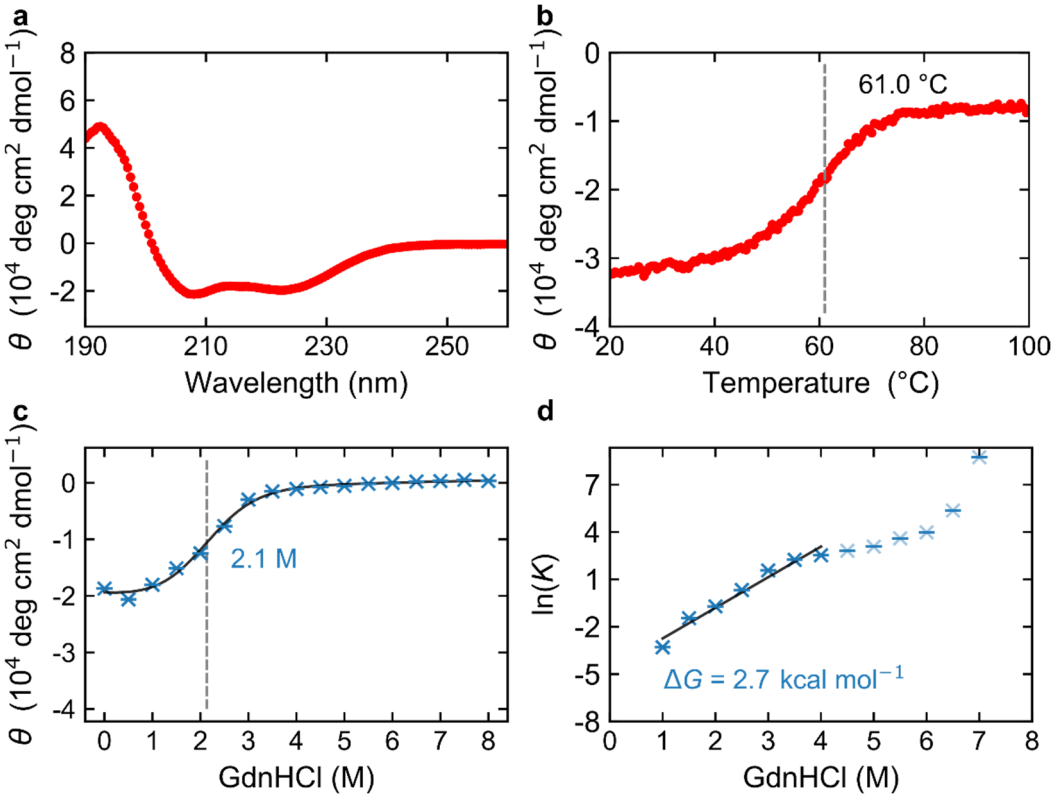

Figure 32: Experimental analysis of 1ip (Lys49-Glu84) *Maquette* 1. **a**, CD spectrum; **b**, melting curve; **c**, chemical unfolding with GdnHCl; **d**, ln*K* plotted against denaturant concentration.

The introduction of a buried ion-pair leads to a partial destabilization of the protein, which can be clearly seen from the experimental results, but is also indicated by the simulation results. The ion-pair pulls some water molecules inside the hydrophobic core to compensate the charged residues during the simulation. The hydrophobic core is not fully water sealed, which may affect the protein stability.

4.2.5 Introduction of another buried ion-pair (Glu49-Lys84)

We next investigated the effect of reverting the polarity of the buried ion-pair. This construct is important to investigate the influence of the polarity of the helices since each helix carries a dipole moment. The RMSD of a 100 ns MD simulation shows a slowly increasing value, which reaches a plateau during the last 30 ns (*Figure 33a*). The loops are more flexible as compared to the helices for all constructs described above (*Figure 33b*). The protein interior becomes partially solvated on the simulation timescale (*Figure 33c*). The plotted CD-NZ distance between Glu49 and Lys84 shows that the introduced ion-pair forms a stable salt bridge with a few transient dissociations over the course of a simulation (*Figure 33d*). Between 40 ns and 60 ns, we observe a few frames where the ion-pair opens and the distance between Glu49 and Lys84 reaches maximum values between 5 and 6 Å. The starting (grey) and ending (blue) frames in *Figure 34* indicate no significant structural changes during the MD simulation.

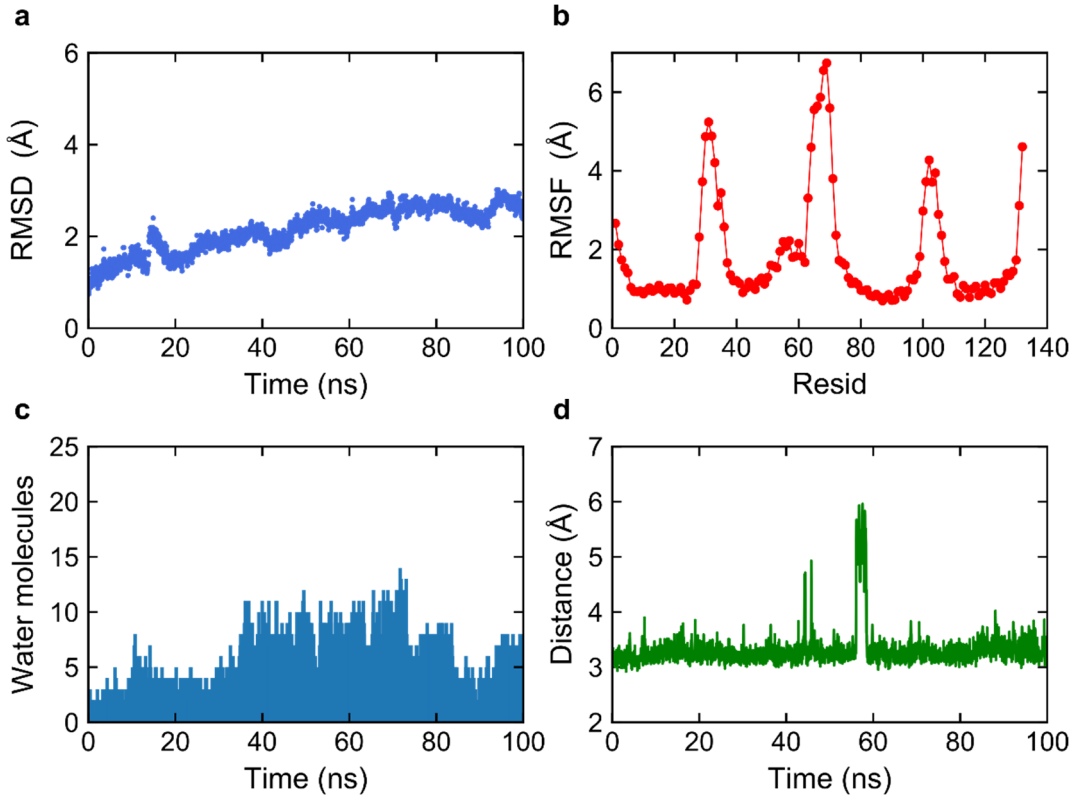

Figure 33: Analysis of 100 ns MD simulation of 1ip (Glu49-Lys84) *Maquette* 1. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content; **d**, distance profile between CD-NZ of Glu49 and Lys84.

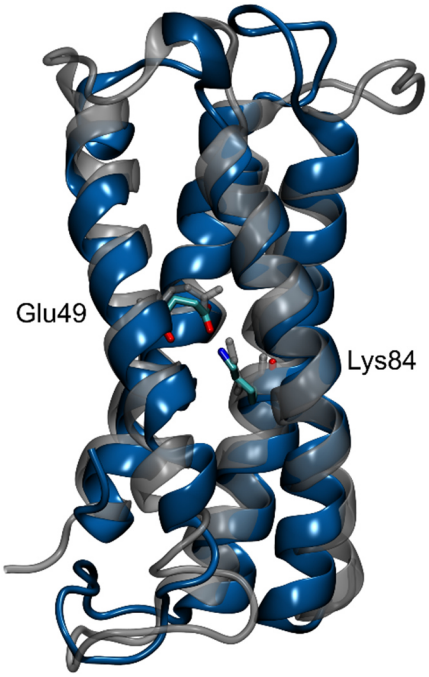

Figure 34: Snapshots from 100 ns MD simulation of 1ip (Glu49-Lys84) *Maquette* 1. The starting structure (grey) is aligned with the last frame (blue) of the simulation. Glu49 and Lys84 are shown in licorice.

Experimentally, the strong band around 15 kDa on the SDS-PAGE gel shows the successful expression and that a high purity was reached during the purification process (*Figure 35a*). Unfortunately, a monomer peak cannot be detected in SEC (*Figure 35b*). Only a dimer peak at a retention volume of 63.5 ml is visible, which corresponds to a MW of ~ 36 kDa. The following experiments were performed using the dimer construct.

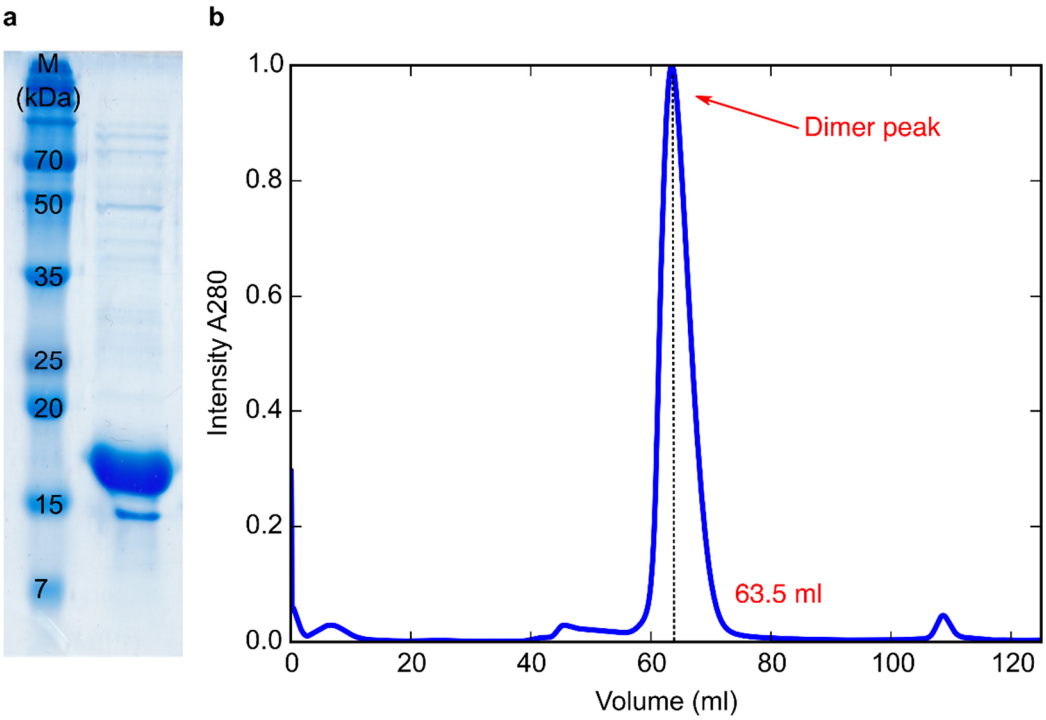

Figure 35: a, SDS gel of the 1ip (Glu49-Lys84) *Maquette* 1; **b**, **s**ize exclusion chromatography (SEC) profile of the 1ip (Glu49-Lys84) *Maquette* 1. The major peak corresponds to dimeric protein. The dimer has a retention volume of 63.5 ml, which corresponds to molecular weight of 36 kDa.

The CD spectroscopic measurements show an α-helical secondary structure (*Figure 36a*) with a melting temperature at 60.9°C (*Figure 36b*). The chemical unfolding with GdnHCl results in a turning point at 2.2 M (*Figure 36c*) and yields a ΔG of 2.9 kcal mol^{-1} (*Figure 36d*).

Summarizing the results for the Glu49-Lys84 *Maquette* 1 construct, suggest that the introduced ion-pair leads to solvation of the bundle interior and to a decreased stability. In comparison to the Lys49-Glu84 *Maquette* 1, the reversed ion-pair has a similar stability, suggesting that the dipole-charge interactions are similar in both constructs.

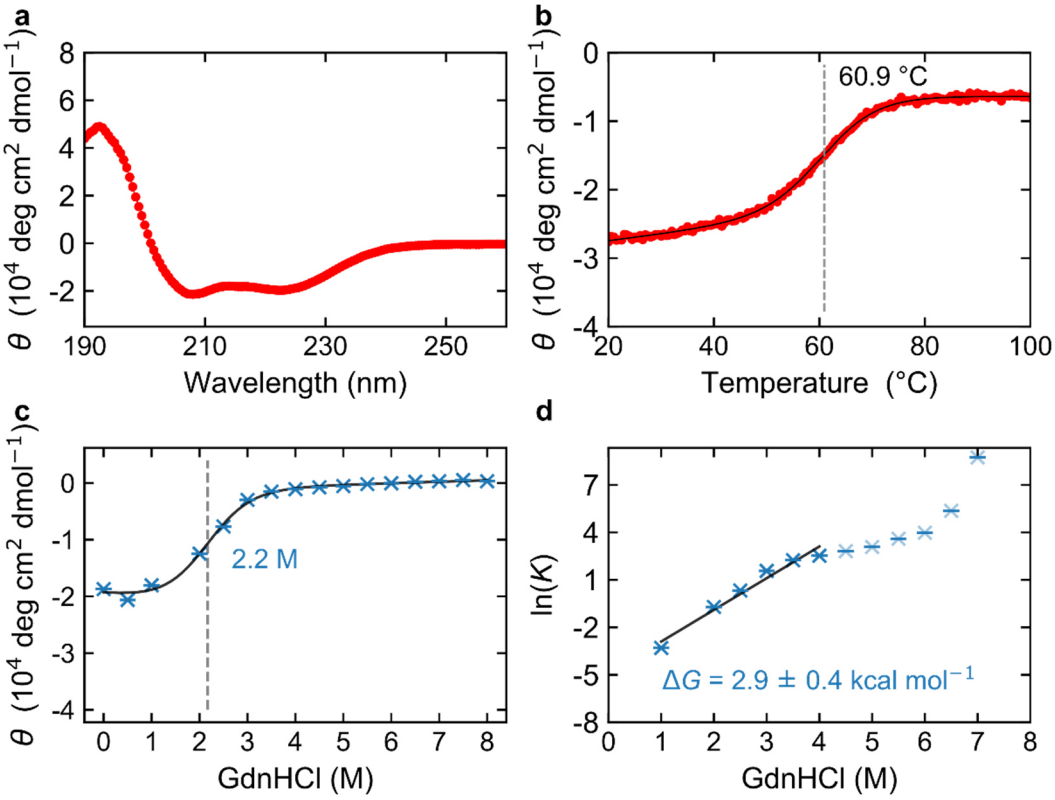

Figure 36: Experimental analysis of 1ip (Glu49-Lys84) *Maquette* 1. **a**, CD spectrum; **b**, melting curve; **c**, chemical unfolding with GdnHCl; **d**, ln*K* plotted against denaturant concentration.

4.2.6 Introduction of two buried ion-pairs (Lys42-Glu91/Lys49-Glu84)

To obtain the next construct of the series, a second ion-pair was introduced at positions 42 and 91. The two phenylalanines were mutated to Lys42 and Glu91 together with the Lys49-Glu84 ion-pair described in the last section. The RMSD for this double ion-pair model increases continuously from 1 Å to 3 Å during the MD simulation, which indicates that the structure has not fully converged during the simulation timescale (Figure *37a*). The RMSF shows a high flexibility of the termini and loops (Figure *37b*). Panel c presents the water content of the bundle interior, which is drastically higher as compared to the constructs with a single ion-pair with > 10 water molecules diffusing into the buried protein interior (Figure *37c*). The bundle scaffold thus becomes significantly more solvated by introduction of the second ion-pair. *Figure 37e* shows the CD-NZ distances of the double ion-pair model (2ip), where the blue trace corresponds to the first ion-pair Lys42-Glu91 indicating that the salt bridge breaks during the first nanoseconds. Glu91 makes a contact to the Lys49 of the second ion-pair (light blue), whereas Lys49 switches between Glu91 and Glu84 (orange), alternating salt bridges with both charges during the 100 ns MD simulation. We observe some conformational switching between these two ion-pairs, illustrated from two snapshots extracted from the simulation (*Figure 37d*). Helices 1 and 3 move away from their original positions and the bundle structure becomes less compact after 100 ns of simulation, which also explains the water molecules diffusing

into the interior of the bundle. The structure partially compensates for the large electrostatic penalty, introduced by the charged amino acids in the hydrophobic protein core. The SEC profile and the SDS gel are not shown here, but the construct elutes as oligomer during SEC with a MW of 59.8 kDa. The construct shows a high purity and the correct size on the denaturing SDS-PAGE gel. *Figure 37f* and *g* show the CD spectroscopic measurements, which were performed with the oligomer. The α-helical content of the secondary structure decreases as compared to the other *Maquette* 1 constructs, but it still forms the main secondary structural element (*Figure 37f*). The melting behavior is also perturbed, indicating that the protein becomes more temperature sensitive and melts at 53.2°C (*Figure 37g*).

As indicated by the combined computational and experimental results, the hydrophobic core may become disordered by the introduction of the four charged residues that may pull in water molecules, and result in the decreased stability.

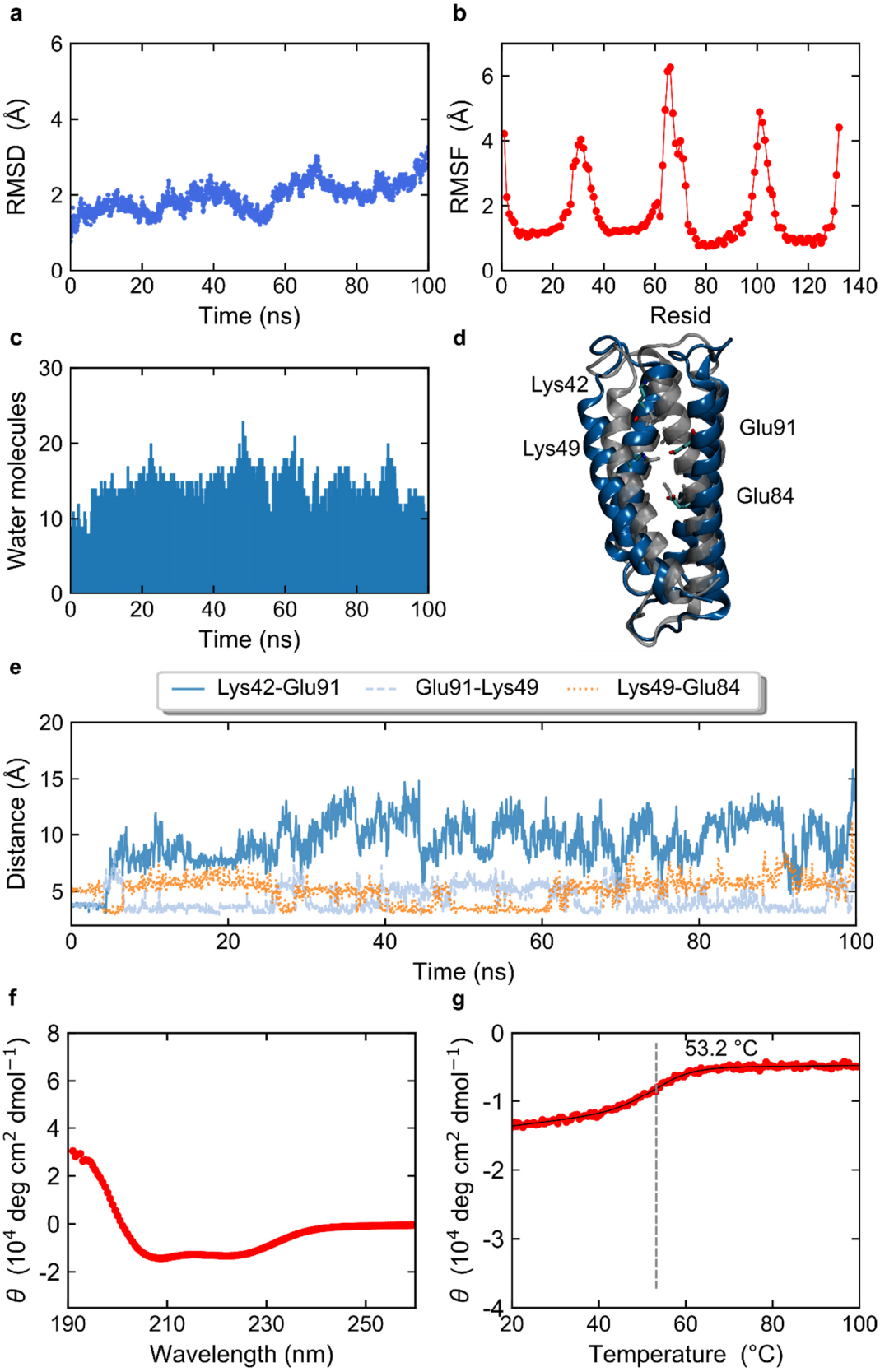

Figure 37: Analysis of 100 ns MD simulation and experimental analysis of 2ip (Lys42-Glu91/Lys49-Glu84) *Maquette* 1. **a**, RMSD plot; **b**, RMSF plot; **c**, Water molecule content; **d**, snapshots from 100 ns MD simulation. The starting structure (grey) is aligned with the last frame (blue) of the simulation. The charged residues are shown in licorice; **e**, Distance plot during 100 ns MD simulation. CD-NZ distance of Lys42-Glu91 (blue), CD-NZ distance of Glu91-Lys49 (light blue), and CD-NZ distance of Lys49-Glu84 (orange); **f**, CD spectrum; **g**, melting curve.

4.2.7 Three buried ion-pairs introduced on two helices (Lys42-Glu91/Lys49-Glu84/Lys56-Glu77)

Despite the challenges to introduce a second ion-pair, we also studied the influence of building a third ion-pair into the hydrophobic core of the bundle framework in order to create a whole charged network across the complete protein structure. The additional ion-pair was placed at positions 56 and 77 by substitution of two phenylalanines. The three ion-pairs (3ip) were placed diagonally on helices 2 and 3, but without using helices 1 and 4. As expected, the RMSD shown in *Figure 38a* indicates that the system does not converge: helices 1 and 3 become more flexible as compared to the other two helices, but still show less movement than the termini and loops in the RMSF plot (*Figure 38b*). The model becomes drastically solvated during the first 10 ns, with around 25 molecules solvating the protein interior (*Figure 38c*). *Figure 38e* shows the CD-NZ distances between the charged residues: after 40 ns, two of the introduced glutamates and lysines find their counter charges, and create stable salt bridges between Lys49-Glu84 (orange) and Lys56-Glu77 (green). The third ion-pair, Lys42-Glu91 (blue), remains closed during the first 15 ns, but opens up afterwards. Towards the end of the simulation, also this third ion-pair finds a closed conformation, but Glu91 does not interact with Lys49 during the simulation (light blue). The Lys56-Glu77 ion-pair opens up at the beginning of the simulation (green) and between 20-40 ns, Lys56 forms transient interactions with Glu84 (light orange). The starting (grey) and ending (blue) snapshots from the 100 ns MD simulation illustrate the afore mentioned findings (*Figure 38d*). Helix 1 moves slightly away from the other helices and the hydrophobic core becomes more accessible to water molecules. The second helix, which hosts three charged lysine residues, forms a kink in the middle of the helix. The three introduced ion-pairs push the helices apart and the bundle loses its tight packing, suggesting that the *Maquette* 1 scaffold is unable to stabilize three buried ion-pairs at least with the current placement. The purification process is not reported here, but the MW derived from SEC retention volume and the high purity was proved by SDS-PAGE. The experimental MW is 65 kDa, which indicates a high oligomeric state. The CD spectrum shows a clear loss of secondary structure as indicated by the less sharp and shallow minima (*Figure 38e*). The shape of the CD signal at 222 nm from 20°C to 100°C results in a temperature sensitivity with a melting point at 52.7°C (*Figure 38f*).

The combined computational and experimental data show that the three introduced ion-pairs on two helices induce solvation of the structure and a decreased stability, and a high oligomeric state. The three ion-pairs introduced inside the hydrophobic are not favorable and the hydrophobic core becomes disordered.

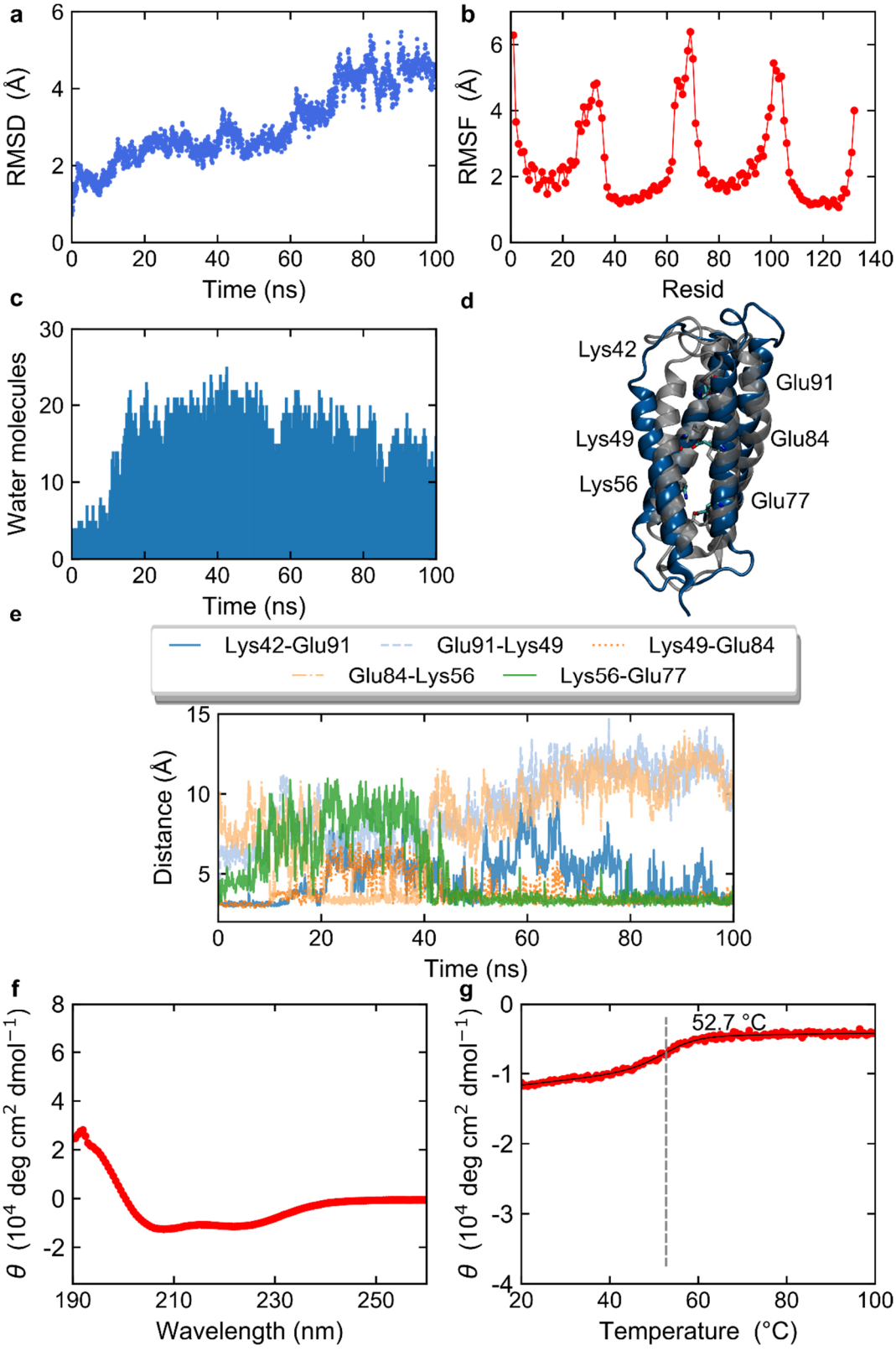

Figure 38: Analysis of 100 ns MD simulation and experimental analysis of 3ip (Lys42-Glu91/Lys49-Glu84/Lys56-Glu77) *Maquette* 1. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content; **d**, snapshots from 100 ns MD simulation. The starting structure (grey) is aligned with the last frame (blue) of the simulation. The charged residues are shown in licorice; **e**, distance plot during 100 ns MD simulation. CD-NZ distance of Lys42-Glu91 (blue), CD-NZ distance of Glu91-Lys49 (light blue), CD-NZ distance of Lys49-Glu84 (orange), CD-NZ distance of Glu84-Lys56 (light orange), and CD-NZ distance of Lys56-Glu77 (green); **f**, CD spectrum; **g**, melting curve.

4.2.8 Three buried ion-pairs introduced on two helices (Glu42-Lys91/Glu49-Lys84/Glu56-Lys77)

In order to improve the stability of the model with three buried ion-pairs, we inverted the position of the charged residues. The RMSD changes significantly during the first 40 ns, but thereafter remains more stable until the end of the simulation (*Figure 39a*). *Figure 39b* shows the RMSF, indicating flexible loops and termini, but more stable helices. However, despite the somewhat improved structural stability, this model also shows a high hydration of the protein (*Figure 39c*). Interestingly, as shown in *Figure 39e*, the first one is Lys91-Glu49 (in light blue) and Lys84-Glu56 (in light orange) remain tightly bound, whereas Lys49 forms a contact with Glu49. We also observe transient contacts between Lys84 with Glu56 that result in coupled conformational changes of Lys77 (green). The starting structure displayed in grey presents three introduced closed ion-pairs, while the snapshot after 100 ns shows two middle ion-pairs in the closed conformation (blue). Glu42 and Lys77 point towards the solvent, while helix 2 becomes partially unstructured near the second loop (*Figure 39d*). The results of the purification process are not shown here in detail, but this construct showed a high oligomer with a MW of 66.8 kDa determined by SEC. The CD spectroscopic analysis indicates loss of α-helicity and a melting point at 49.7°C compared to 52.7°C for the 3ip model described above (*Figure 39f, g*).

The combined computational and experimental results thus indicate that the protein loses stability by introduction of the highly charged residues inside the hydrophobic core. The results are very similar to the reversed 3ip *Maquette* 1 described in section 4.2.7.

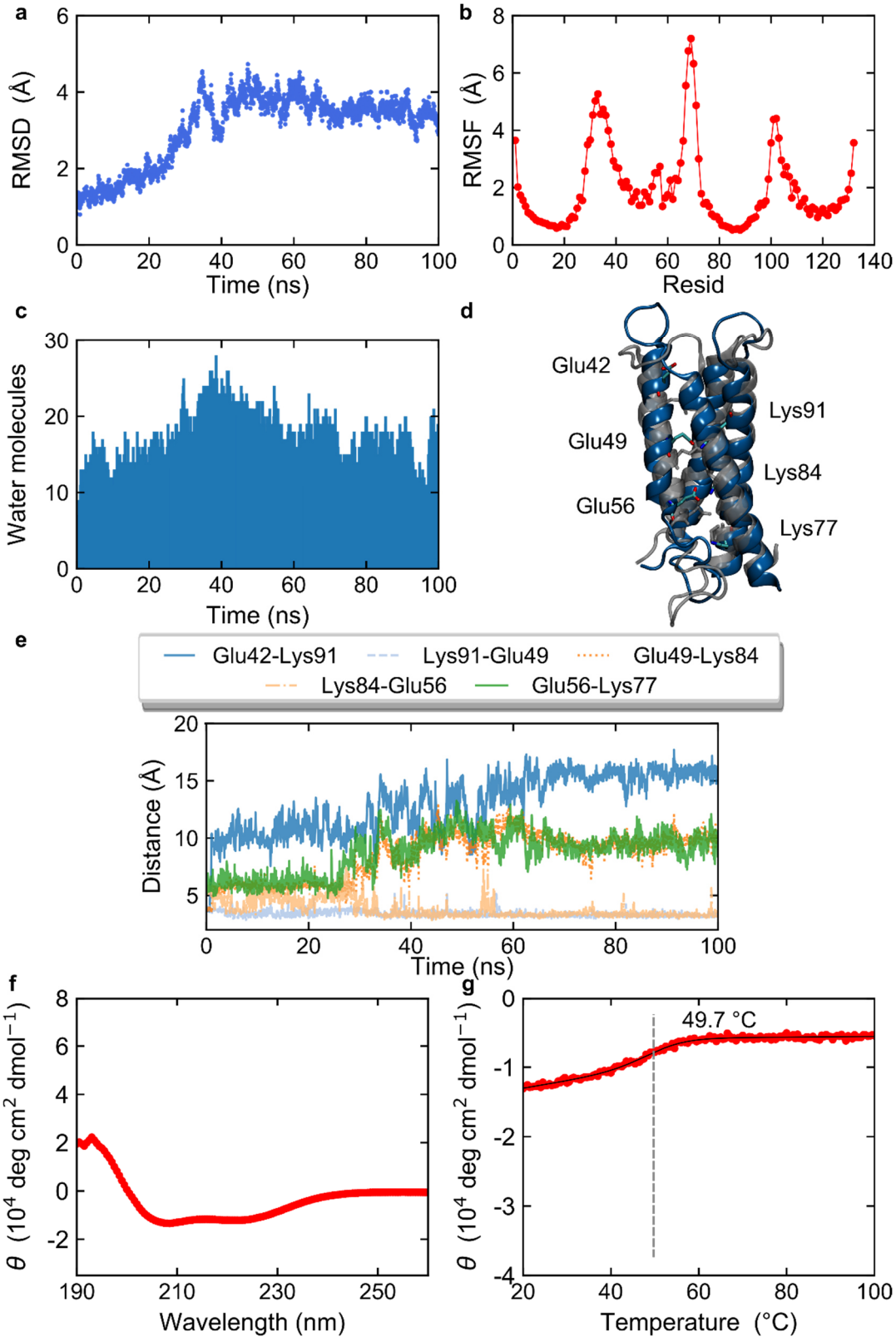

Figure 39: Analysis of 100 ns MD simulation and experimental analysis of 3ip (Glu42-Lys91/Glu49-Lys84/Glu56-Lys77) *Maquette* 1. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content; **d**, snapshots from 100 ns MD simulation. The starting structure (grey) is aligned with the last frame (blue) of the simulation. The charged residues are shown in licorice; **e**, distance plot during 100 ns MD simulation. CD-NZ distance of Glu42-Lys91 (blue), CD-NZ distance of Lys91-Glu49 (light blue), CD-NZ distance of Lys49-Glu84 (orange), CD-NZ distance of Glu84-Lys56 (light orange), and CD-NZ distance of Lys56-Glu77 (green); **f**, CD spectrum; **g**, melting curve.

4.2.9 Three buried ion-pairs introduced on four helices (Lys24-Glu112/Lys84-Glu49/Lys77-Glu56)

As an additional attempt to create a more stable bundle protein with three ion-pairs, we created charged residues on all four helices, instead of using only two helices. The middle and the third ion-pair were built on their previously described position, while the new first ion-pair was built at the top of the bundle by replacing phenylalanines at positions 24 and 112 by lysine and glutamate residues. The RMSD increases during the first 50 ns, and reaches values around 3 Å towards the end of the simulation (*Figure 40a*). The RMSF is similar to previous models (*Figure 40b*), and water penetration is slightly less compared to the previous model with the three introduced ion-pairs (*Figure 40c*). However, the bundle becomes extensively solvated during the simulation timescale. The initial ion-pairs between Glu112-Lys24 (blue), Glu49-Lys84 (orange), and Glu56-Lys77 (green) break (*Figure 40d*), and during first few nanoseconds Lys24 forms a stable salt bridge with Glu49, while Glu112 moves away (light blue). The same takes place at the other end of the bundle, where Lys77 flips down and Glu56 forms a fluctuating ion-pair with Lys84 (light orange). The opening of the Glu112-Lys24 ion-pair could result from the interaction of the lysine with the two bulky phenylalanines at positions 91 and 119. The structural changes indicate conformational changes in (grey) helices 1 and 4 that lead to a partial opening of the protein structure and diffusion of water molecules into the protein interior. At the end of the MD simulation, the two middle ion-pairs are closed and Lys77 and Glu112 point towards the bulk (*Figure 40d*). The calculated MW of this construct is 55.7 kDa from SEC, *i.e.*, approximately 10 kDa lower compared to the 3ip *Maquette* 1, and the protein is in a higher oligomeric form. The CD spectrum indicates loss of α-helical structure (*Figure 40e*) and the second minimum around 222 nm disappears almost completely. The first minimum at 208 nm is also less pronounced and is shifted towards the typical random coil minimum around 200 nm. The melting curve recorded at 222 nm indicates a high loss of secondary structure since no thermal transition is detectable anymore (*Figure 40f*). Thus, similar to the two other 3ip *Maquette* 1 models, the protein scaffold becomes unstructured by insertion of six buried charges.

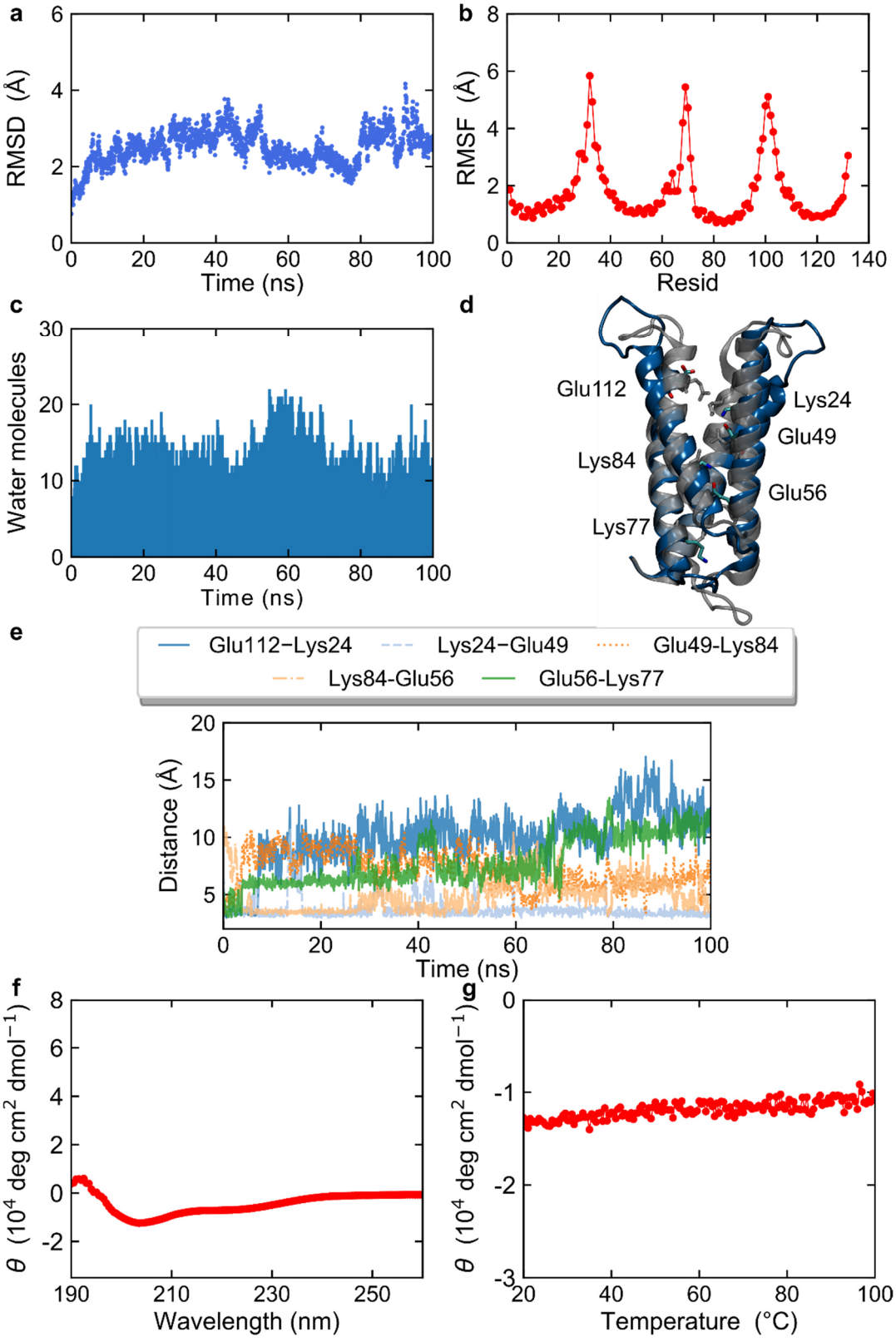

Figure 40: Analysis of 100 ns MD simulation and experimental analysis of 3ip (Lys24-Glu112/Lys84-Glu49/Lys77-Glu56) *Maquette* 1. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content; **d**, snapshots from 100 ns MD simulation. The starting structure (grey) is aligned with the last frame (blue) of the simulation. The charged residues are shown in licorice; **e**, distance plot during 100 ns MD simulation. CD-NZ distance of Glu112-Lys24 (blue), CD-NZ distance of Lys91-Glu49 (light blue), CD-NZ distance of Lys49-Glu84 (orange), CD-NZ distance of Glu84-Lys56 (light orange), and CD-NZ distance of Lys56-Glu77 (green); **f**, CD spectrum; **g**, melting curve.

4.2.10 Three ion-pairs introduced on four helices with protonated glutamate 112 (Lys24-Glu112prot/Lys84-Glu49/Lys77-Glu56)

Although a fully stable design with the three buried ion-pair residues could not be experimentally achieved with the current scaffold model, we wanted to nevertheless study the effect of conformational switching connecting the framework to an on-off signal. To this end, we protonated the first Glu112, and studied the effect of the conformational switching in the remaining ion-pair network. We expected that this protonation could trigger opening of the Glu112-Lys24 ion-pair, while Lys24 moves down to the next Glu49. Glu49 could then loose its contact with Lys84 and form interaction with Glu56 of the third ion-pair. This could trigger a conformational change in the terminal Lys77 that can flip "down", and trigger "signal". The model was simulated for 100 ns to test for switching function. *Figure 41a* shows that RMSD stabilizes to around 3.5 Å towards the end of the simulation, with the RMSF comparable to previous 3ip-models (*Figure 41b*) but interestingly, the bundle undergoes less interior solvations (< 20 water molecules) as compared to the deprotonated model with 22 water molecules (*Figure 41c*). The ion-pair dynamics indicates that the protonated Glu112 moves away from Lys24 and the first ion-pair dissociates during the first nanoseconds (blue). Lys24 indeed finds a new countercharge, creating an ion-pair with Glu49 (light blue). This ion-pair remains stable until it opens after 80 ns. However, Glu49 and Lys84 (orange) do not form contacts during the simulation, as Glu49 prefers a salt bridge with Lys24 and Lys84, in turn, forms a stable contact with Glu56 (in light orange). Lys77 flips out and points towards the bulk water (green). Based on the distance plot, protonation of the initial glutamate indeed leads to a switch function, where the terminal Lys77 undergoes a conformational change (*Figure 41e*). These findings are illustrated by structural snapshots from the simulation (*Figure 41d*). Although, helix 1 moves somewhat away from the other three helices, the model is in general more water sealed as compared to the other 3ip-constructs.

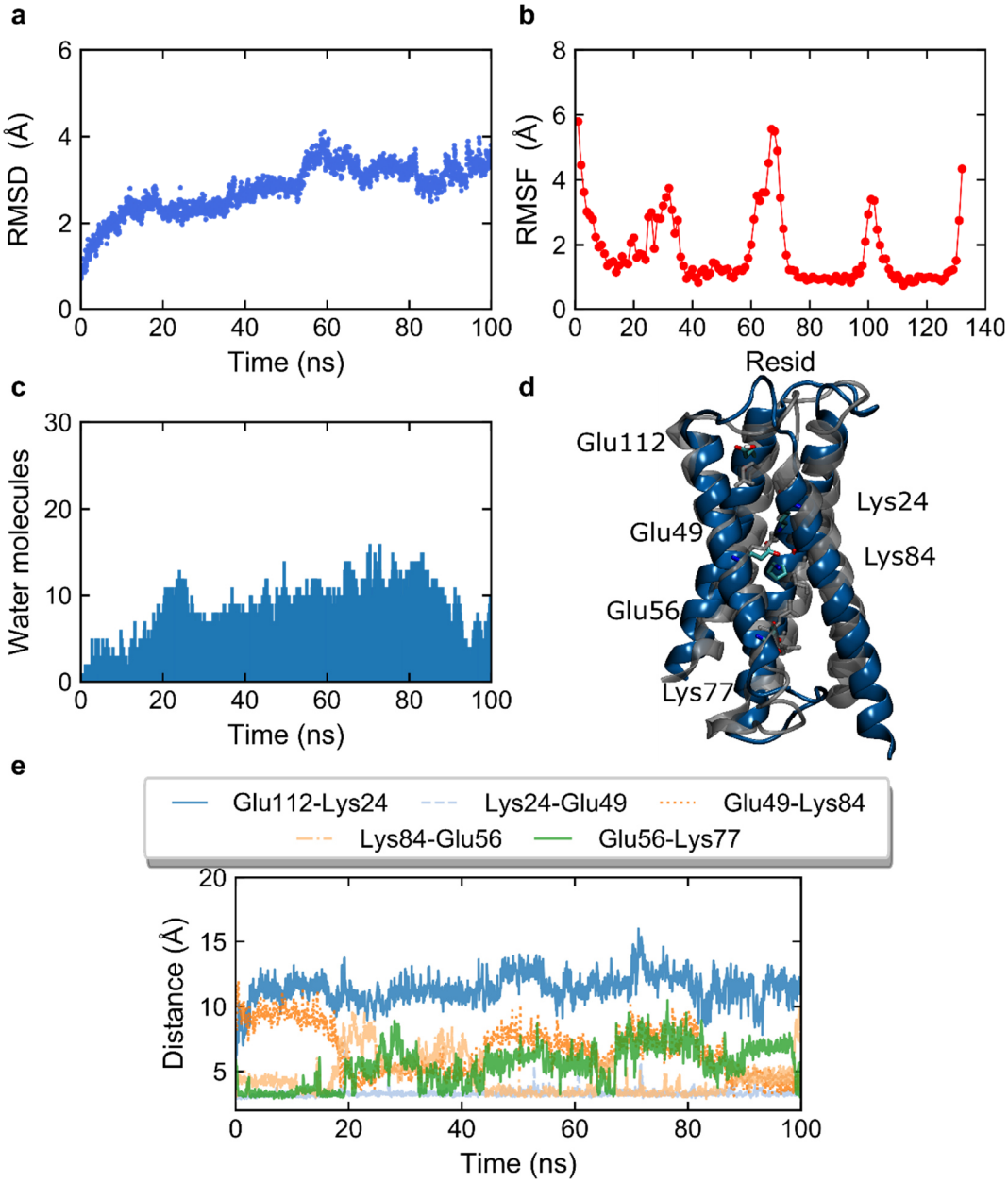

Figure 41: Analysis of 100 ns MD simulation and experimental analysis of 3ip (Lys24-Glu112prot/Lys84-Glu49/Lys77-Glu56) *Maquette* 1. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content; **d**, snapshots from 100 ns MD simulation. The starting structure (grey) is aligned with the last frame (blue) of the simulation. The charged residues are shown in licorice; **e**, distance plot during 100 ns MD simulation. CD-NZ distance of Glu112prot-Lys24 (blue), CD-NZ distance of Lys91-Glu49 (light blue), CD-NZ distance of Lys49-Glu84 (orange), CD-NZ distance of Glu84-Lys56 (light orange), and CD-NZ distance of Lys56-Glu77 (green).

4.2.11 Three buried ion-pairs introduced on four helices with glutamine 112 (Lys24-Gln112/Lys84-Glu49/Lys77-Glu56)

Since experimental validation is important for the protein design, and the site-specific protonation of Glu112 is not feasible to achieve experimentally, we built a new model based on the previous 3ip-*Maquette* 1, but with Glu112 replaced by a glutamine to mimic a protonated glutamic acid. The MD analysis indicates that the Gln-model becomes stable after 10 ns (*Figure 42a*), with the RMSF and water plots resembling the protonated Glu112 simulations described in section 4.2.10 (*Figure 42b, c*). The ion-pair dynamics closely resembles the simulations with protonated Glu112 (*Figure 42d*), with the exception that the Gln112 and Lys24 (blue) stays relatively close together until the end of the simulation. Two salt bridges form contacts at around 3.5 Å due to the flipped charged residues in the middle of the bundle (Lys24-Glu49 in light blue and Lys84-Glu56 in light orange), as also shown in *Figure 42e* . Helix 3 shows a kink in the final structure. This construct was also expressed and characterized in the laboratory, but the purification process is not reported. The protein elutes also as oligomer with a MW of ~ 56 kDa derived from SEC. *Figure 42f* shows the results of CD spectroscopical analysis performed with the oligomer, which look similar to the construct with Glu112. The α-helicity decreases (*Figure 42f*) and a thermal transition is not detectable anymore (*Figure 42e*). Despite the promising conformational switching observed computationally, the *Maquette* 1 scaffold seems to be unlikely to stabilize the high amount of buried charges for the experimental characterization.

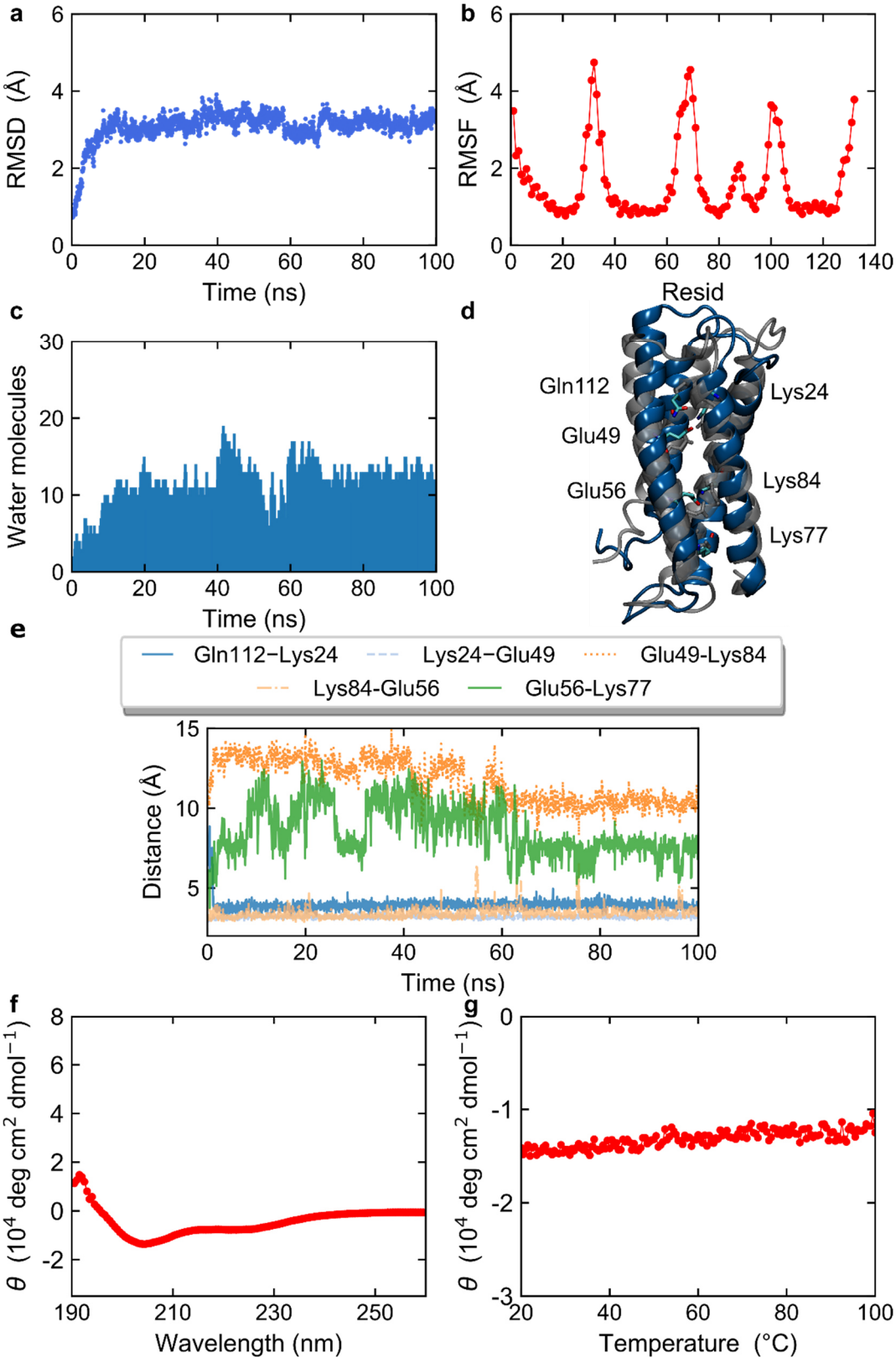

Figure 42: Analysis of 100 ns MD simulation and experimental analysis of 3ip (Lys24-Gln112/Lys84-Glu49/Lys77-Glu56) *Maquette* 1. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content; **d**, snapshots from 100 ns MD simulation. The starting structure (grey) is aligned with the last frame (blue) of the simulation. The charged residues are shown in licorice; **e**, Distance plot during 100 ns MD simulation. CD-NZ distance of Gln112-Lys24 (blue), CD-NZ distance of Lys24-Glu49 (light blue), CD-NZ distance of Lys84-Glu49 (orange), CD-NZ distance of Lys84-Glu56 (light orange), and CD-NZ distance of Lys77-Glu56 (green); **f**, CD spectrum; **g**, melting curve.

4.2.12 *Maquette* 1 models with short loops

To increase the overall stability of the *Maquette* 1 models, we next studied the effect of the long loop regions that were highly flexible in MD simulations, and possibly prevented, *e.g.*, crystallization despite several trials with many different buffer conditions (pH, salt concentration, etc.). To this end, we decided to follow the same approach as in the Hsp90 subproject, and cut the loop regions to obtain a more stable system. The loops comprise nine amino acids (GGSGGSGGS) in the *Maquette* 1 starting structure and the described mutant series. Based on MD simulation trials, we empirically found that by keeping only four amino acids (GGSG) in the loops created more stable models and a next generation of *Maquette* 1 models. The four residues formed the minimal loop lengths, as shorter loops resulted in dissociation of the helices during MD trial simulations. The first construct was based on the previous *Maquette* 1 hydrophobic core model but with the short loops. However, we also replaced Phe91 and Phe119 with alanines as these residues formed interactions with Lys25 and resulted in the dissociation of the Lys24-Glu112 salt bridge.

The RMSD obtained from a 100 ns MD shows stabilization toward the end of the simulation (*Figure 43a*). The RMSF (*Figure 43b*) confirms decreased flexibility of the loops, with a RMSF of around 3 Å as compared to approximately 5 Å for *Maquette* 1 with the long loops (*Figure 20b*). The new construct also remains more sealed from water molecules (*Figure 43c*). Two structural snapshots from the simulation are shown in *Figure 44*, indicating no significant structural changes.

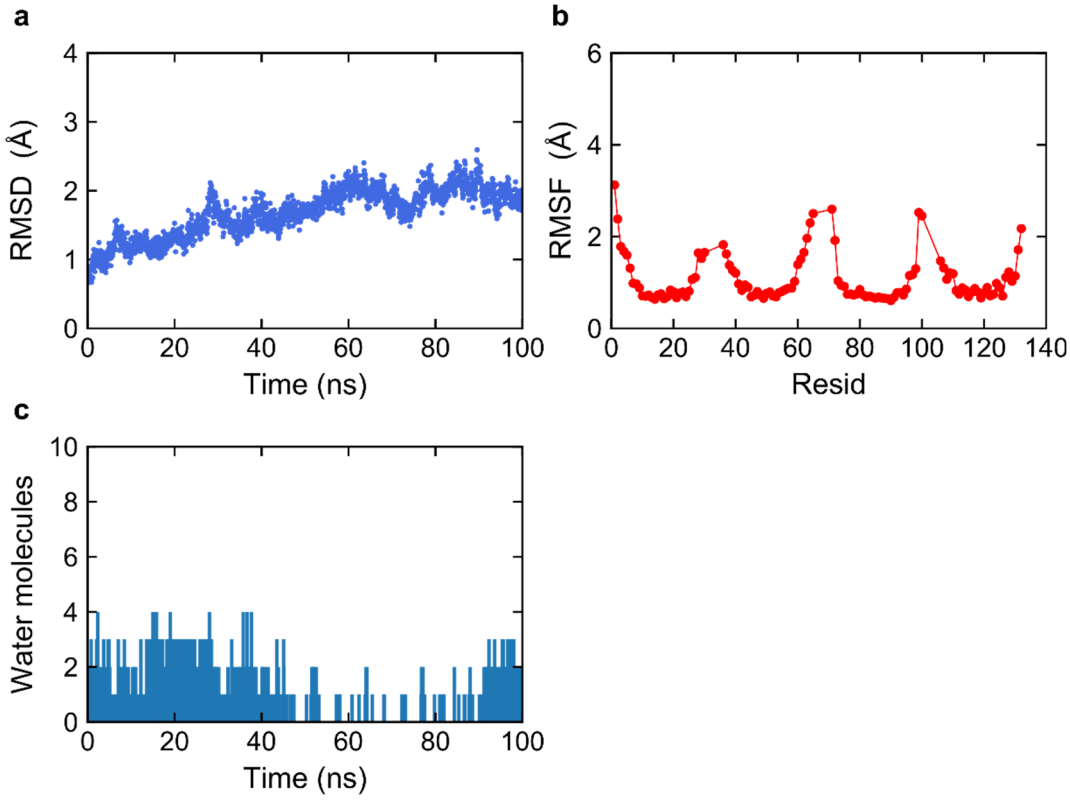

Figure 43: Analysis of 100 ns MD simulation of *Maquette* 1 with short loops. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content.

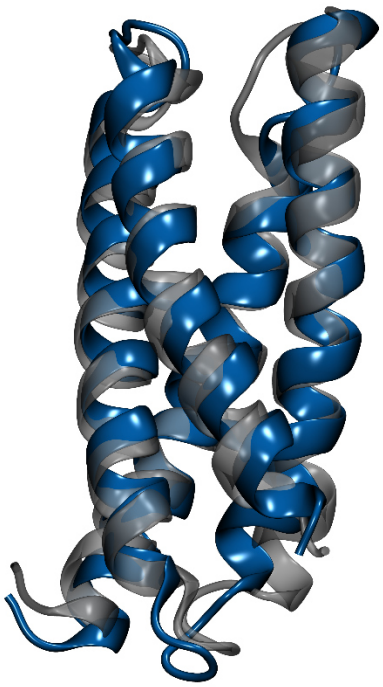

Figure 44: Snapshots from 100 ns MD simulation of hydrophobic core *Maquette* 1 with short loops. The starting structure (grey) is aligned with the last frame (blue) of the simulation.

The new protein model was expressed, purified, and characterized experimentally. The MW for the new construct calculated from the SEC profile is 22 kDa, which is slightly larger than the expected monomeric mass of 16 kDa. A high purity was detected on SDS-PAGE gel (data not shown) and the CD spectra indicate a well folded α-helical structure, which is temperature stable and shows no unfolding up to 100°C (*Figure 45a, b*).

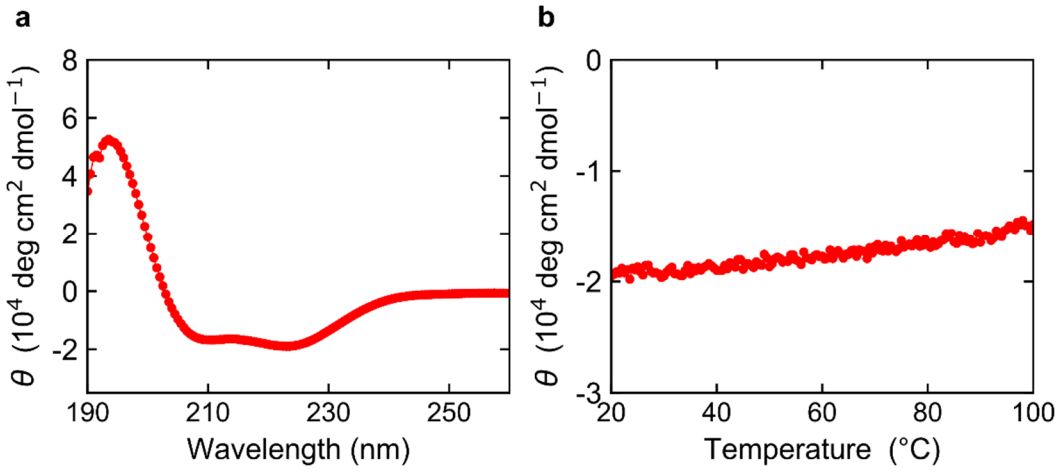

Figure 45: Experimental analysis of *Maquette* 1 with short loops. **a**, CD spectrum; **b**, melting curve.

We next studied the short loop constructs by systematically introducing a single buried Glu or Lys at position 49, a buried ion-pair at positions 49 and 84, a double ion-pair, and a triple ion-pair and its Gln112 mutation thereof. The models behave in general similar to the models with the long loops and therefore, detailed data from the simulation analysis is not shown here. Introduction of charged residues into the hydrophobic core leads to solvation during the MD simulation and to oligomerization and loss of the α-helical structure found by experiments. Snapshots from the 100 ns MD simulations and the results of the experimental characterization are summarized in *Figure 46* and *Figure 46* and only the major findings are shortly discussed as the only difference to the previously described series is the loop length. In both models with a single buried charge, the side-chain points towards the water at the end of the simulation (*Figure 46a, b*). The model with two ion-pairs already shows some kind of double ion-pair in the first frame (*Figure 46d*, grey). Another interesting aspect is that apparently helices 3 and 4 start to unfold while helix 1 reorganizes and the kink disappears. The second helix does not significantly change during the simulation and the bundle structure tries to compensate the buried ion-pairs starting to reorientate. In the three ion-pair model the ion-pairs partially switch and the terminal Lys77 flips out *(Figure 46e*, blue). The fourth helix where Glu112 is located has a less α-helical part which is flexible during the simulation. The protonation of Glu112 acts as starting signal for the mostly irreversible switch (*Figure 46f*) and the substitution by a glutamine leads to the identical switching (*Figure 46g*).

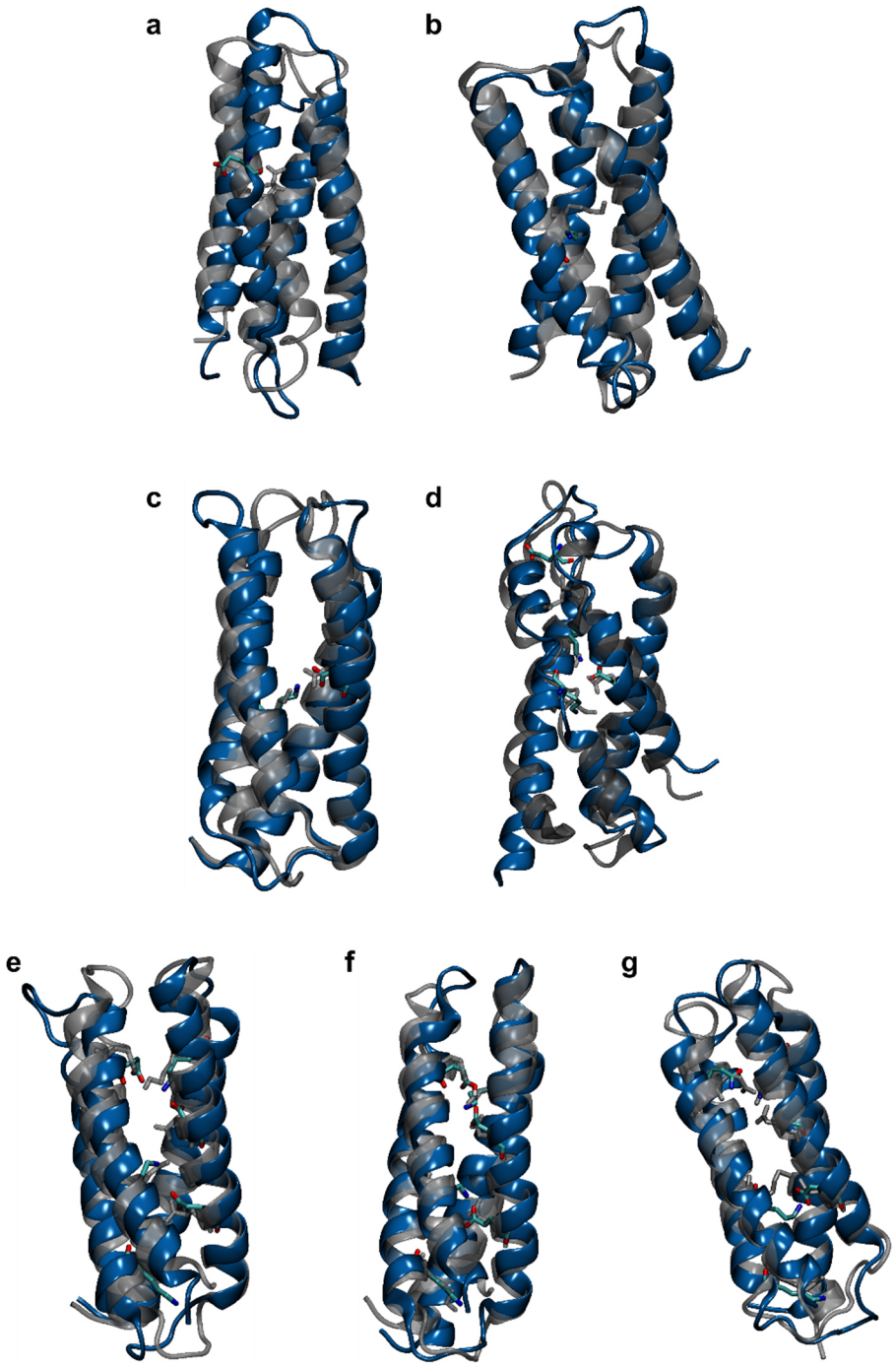

Figure 46: Snapshots from 100 ns MD simulation of models with buried, charged residues and short loops of *Maquette* 1. The starting structure (grey) is aligned with the last frame (blue) of the simulation. The charged residues are shown in licorice. **a**, Glu49; **b**, Lys49; **c**, 1ip (Glu49-Lys84); **d**, 2ip (Glu112-Lys24/Glu49-Lys84); **e**, 3ip (Glu112-Lys24/Glu49-Lys84/Glu56-Lys77); **f**, 3ip (Glu112prot-Lys24/Glu49-Lys84/Glu56-Lys77); **g**, 3ip (Gln112-Lys24/Glu49-Lys84/Glu56-Lys77).

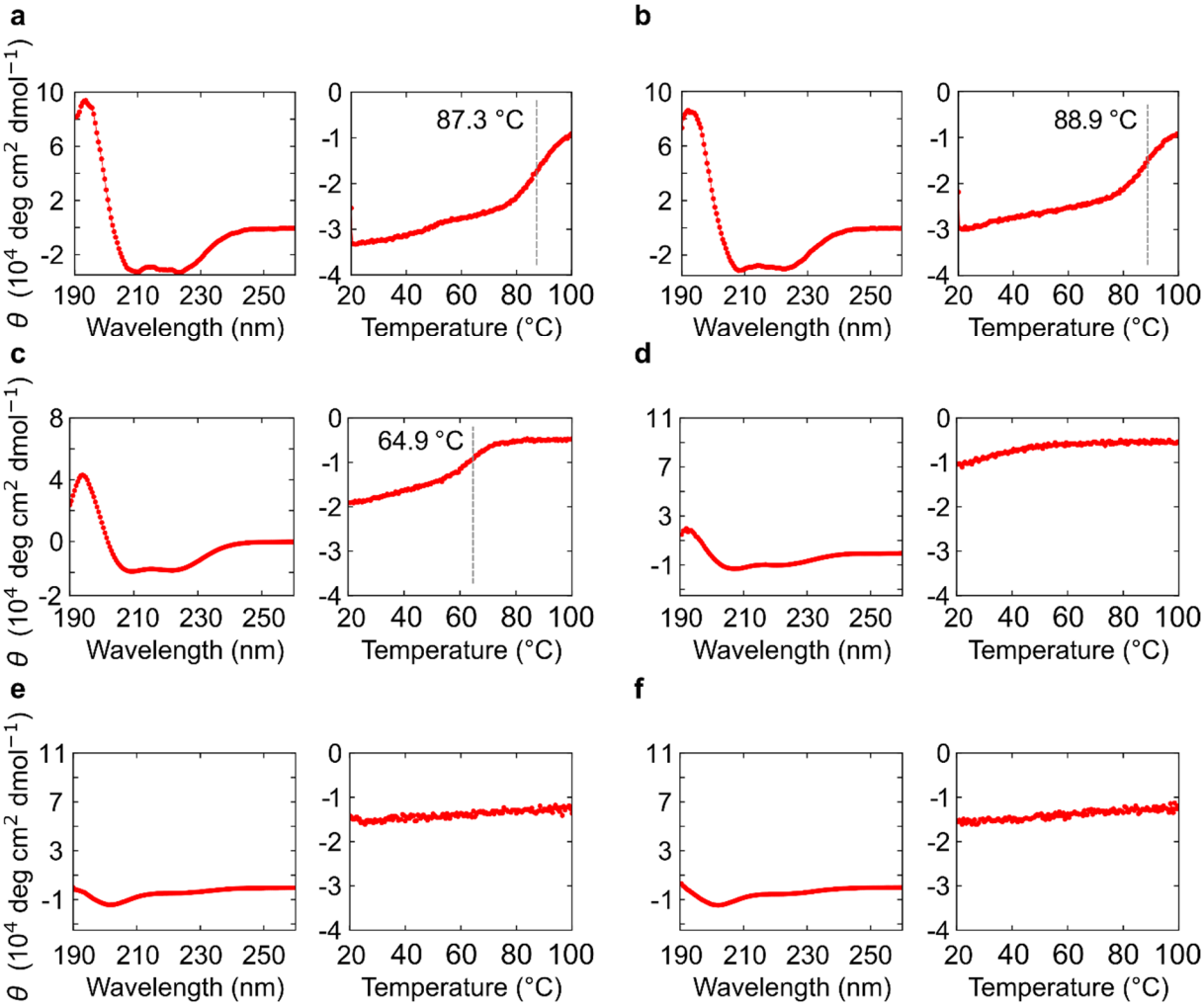

Figure 47: Experimental analysis of models with buried, charged residues and short loops of *Maquette* 1. CD spectrum (left) and melting curve (right) is shown. **a**, Glu49; **b**, Lys49; **c**, 1ip (Glu49-Lys84); **d**, 2ip (Glu112-Lys24/Glu49-Lys84); **e**, 3ip (Glu112-Lys24/Glu49-Lys84/Glu56-Lys77); **f**, 3ip (Gln112-Lys24/Glu49-Lys84/Glu56-Lys77).

The short loop models with buried, charged residues were experimentally analyzed. The purification process, that is not reported here, resulted in oligomeric proteins. The oligomers were further investigated with CD spectroscopy. The CD spectra and the melting curves for all models are reported in *Figure 47*. As already seen for the models with long loops described before, the introduction of charged residues into the hydrophobic core leads to decreasing melting points and further to the loss of α-helical structure in the triple ion-pair models.

4.2.13 Analysis of switching in short loops 3ip models

To study the switching in the 3ip (Glu112-Lys24/Glu49-Lys84/Glu56-Lys77) *Maquette* 1 with short loops in greater detail, 1000 ns MD simulations of the non-protonated and the protonated variant were performed. *Figure 48* shows the CD-NZ distance between Glu56 and Lys77 of the third ion-pair. Here the switch is considered to be "on" when Glu112 is modeled in the protonated state (*Figure 48*, light blue). The terminal ion-pair indeed remains open during the whole 1000 ns trajectory. However, when the switch is "off", here achieved by modelling Glu112 in the deprotonated state, the Glu56-Lys77 remains in a closed conformation during almost entire 1000 ns MD simulation (*Figure 48*, blue). Even when the ion-pair is open, the distance between CD and NZ is smaller than in the protonated model indicates a clear difference in the sampled distance distribution, suggesting the construct has an effective high-fidelity switching function in the computational models.

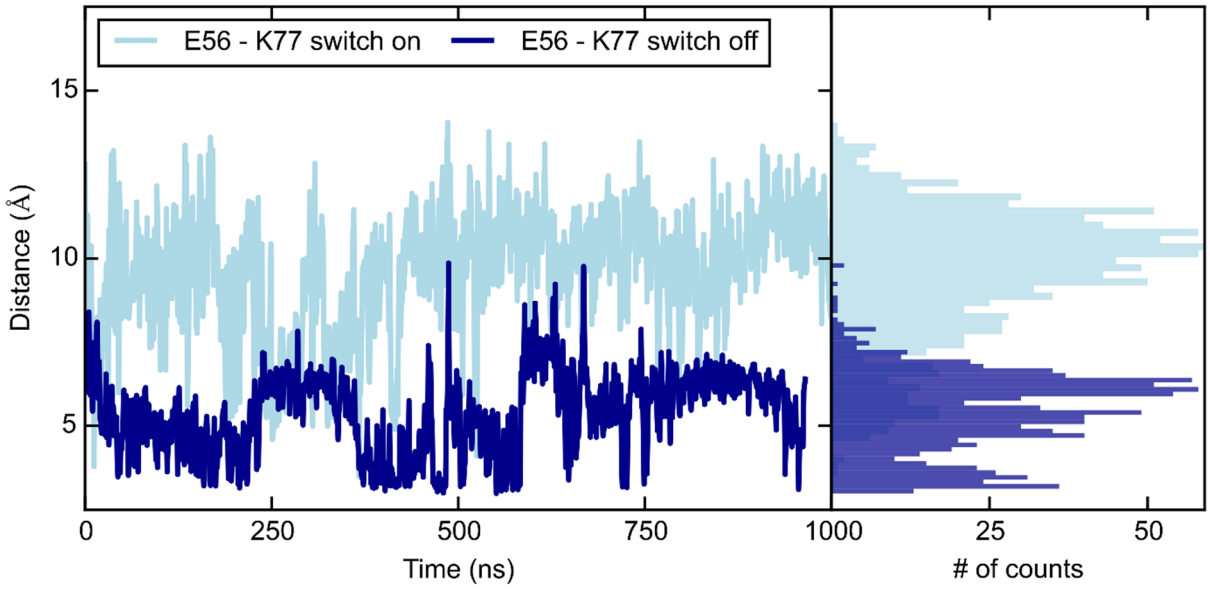

Figure 48: Distance plot of the protonated and the not-protonated 3ip (Gln112-Lys24/Glu49-Lys84/Glu56-Lys77) *Maquette* 1 with short loops. The left part shows the distance between the CD-NZ of Glu56 and Lys77 during almost 1000 ns MD simulation. The switch is on in the protonated construct (light blue), while the switch is off in the unprotonated model (blue). The left part shows the populations, where the ion-pair is closed or open. The protonated state is again in light blue and the unprotonated one in blue.

4.2.14 NMR studies on *Maquette* 1 series

To obtain structural insight into *Maquette* 1, we attempted to crystalize the hydrophobic core construct and the mutant series. However, possibly due to the very flexible loops, the attempts were not successful. We also studied the system using NMR, but the high homogeneity of the bundle did not allow us to assign the spectra. As shown in *Figure 49a*, the quality of HSQC spectrum of the hydrophobic core with long loops is poor. The signals show a low dispersion and are difficult to

identify since they overlap. In the next step, the mutants 1ip, 2ip, and 3ip *Maquette* 1 with long loops were measured. *Figure 49b* shows the HSQCs of 1ip (red) and 3ip (on all three helices, grey) compared to the hydrophobic core (blue). The quality and dispersion increase a bit with higher number of introduced charged amino acids. The chemical surrounding becomes more differentiated due to the mutations and the 3ip mutant shows more separated peaks, but still the dispersion is not ideal. The HSQCs of the hydrophobic core with long (blue) and short loops (turquoise) are compared in panel c. The shorter loops lead to a few additional peaks in the spectrum and slightly improves the spectral quality. The ion-pair series with the short loops, shown in *Figure 49d* indicate that the introduction of the ion-pairs leads to an improved spectrum. Interestingly, for these models, the 1ip shows more additional signals and a higher dispersion as compared to 3ip. It seems that the 3ip mutant partially loses its fold, as indicated by its less dispersed spectrum. In spite of that, the peaks of the 3ip mutant are better separated as in the hydrophobic core due to the changed chemical surrounding. Panel e shows the comparison of 1ip with long (red) and short loops (orange), where, an improved spectrum is observed with the shorter loops.

The NMR studies thus suggest that cutting the loops and the introduction of the mutations lead to a drastically improved quality and dispersion of the spectra due to a change in the chemical environment. Nevertheless, it was not possible to solve the structure of these constructs, because of the high homogeneity of the helices. Therefore, another bundle framework, where the starting structure was already solved by NMR is used for further investigations.

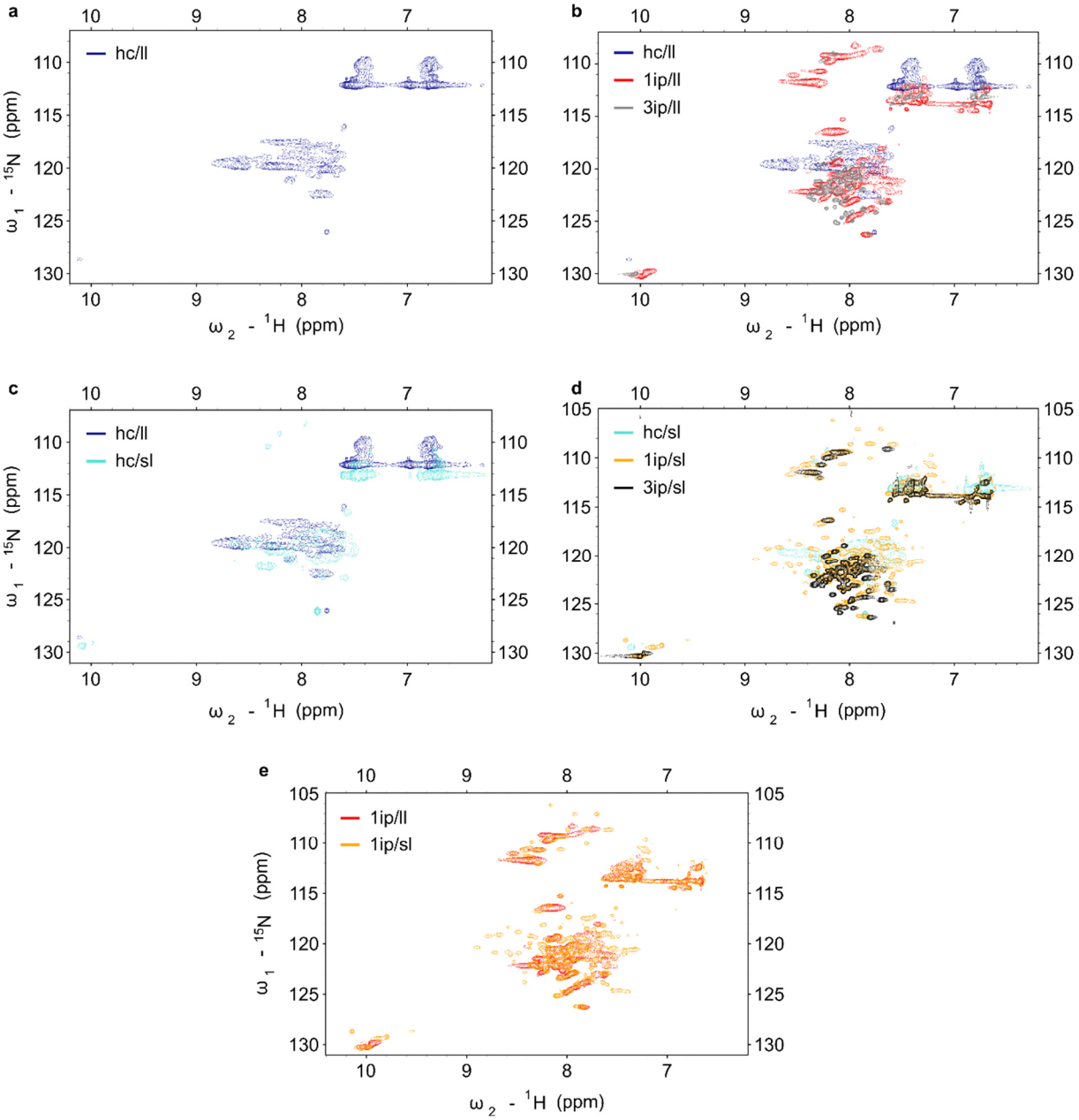

Figure 49: HSQCs of the bundle series *Maquette* 1 with long (ll) and short loops (sl). **a**, HSQC of hydrophobic core with long loops; **b**, HSQCs of hydrophobic core (blue), 1ip (red), and 3ip with long loops (grey); **c**, HSQCs of hydrophobic core with long (blue) and short loops (turquoise); **d**, HSQCs of hydrophobic core (turquoise), 1ip (orange), and 3ip (black) with short loops; **e**, HSQCs of 1ip with long (red) and short (orange) loops.

4.3 *Maquette* 2

The second starting framework that we used to study the influence of introduced charged residues on the structure and stability of a protein comes from a published NMR structure by DeGrado's group.[73] In the following sections these constructs are named *Maquette* 2. The focus for this series is on the analysis of the stability and conformational changes induced by introducing an ion-pair in either the hydrophilic or the hydrophobic part of the protein. This model system is more suitable to study structural changes that result from the introduced ion-pair as a high-quality NMR structure is already known so changes observed in the spectra can be assigned to residues in *Maquette* 2. Here the mutations of interest were also introduced stepwise, but in this case the series was limited to a single ion-pair. All constructs were simulated for 200 ns and the RMSD, RMSF, water content, and ion-pair distance were analyzed. The secondary structure and stability of the model proteins were experimentally investigated by CD spectroscopy following thermal melting and chemical unfolding, while the structure of the *Maquette* 2 model was solved using NMR.

4.3.1 Hydrophobic core

200 ns MD simulations were carried out based on an NMR structure model (PDB ID: 5TGW) from Polizzi *et al*.[73] by embedding the protein in a waterbox with ions around the protein. *Figure* 50 shows the results of the simulation. During the 200 ns MD simulations, the RMSD value stabilizes after a few nanoseconds and the system remains dynamically stable (*Figure 50a*). The RMSF plot indicates higher flexibility in the C-terminus as compared to the rest of the protein, but in general, the value is relatively low and without significant deviations from the rest of the protein (*Figure 50b*). The loop regions are flexible, but less as compared to *Maquette* 1 (*Figure 20b*). The protein interior has a low hydration level, which does not increase during the simulation (*Figure 50c*).

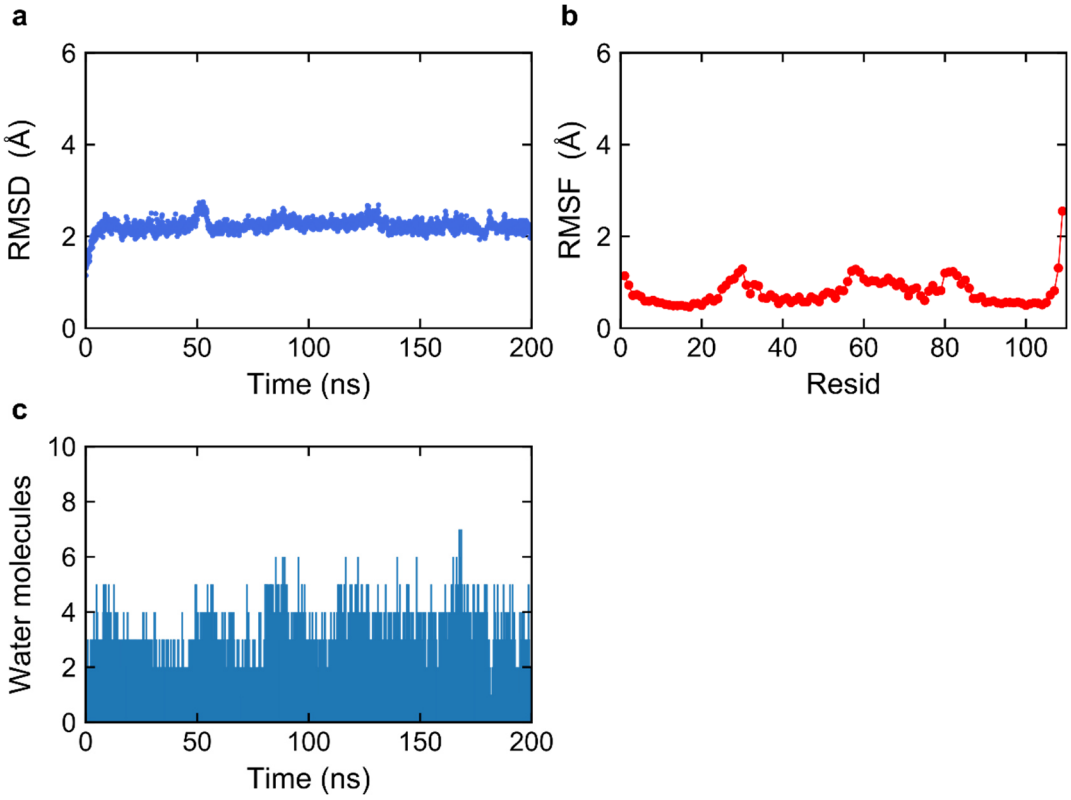

Figure 50: Analysis of 200 ns MD simulation of hydrophobic core *Maquette* 2. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content.

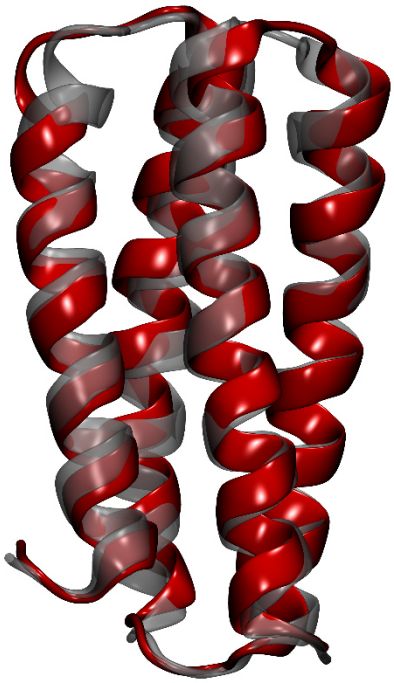

Figure 51*:* Snapshots from 200 ns MD simulation of hydrophobic core *Maquette* 2. The starting structure (grey) is aligned with the last frame (red) of the simulation.

Two snapshots from the 200 ns MD simulation are presented in *Figure 51* with the starting structure in grey and the last frame in red, indicating the absence of large-scale structural differences. In the next step, the *Maquette* 2 models were characterized experimentally. The expression and purification were conducted as described in the method section. The purity of the protein was tested by SDS-PAGE. *Figure 52*a shows a protein band of the expected size for a monomer and proves a high purity of the sample. A SEC performed during the purification process, indicated a monomeric protein with a retention volume of 74.02 ml corresponding to a MW of 18 kDa (*Figure 52b*). While this is larger than the expected based on the MW of ~ 13 kDa, Polizzi *et al*. also noted that the protein elutes a bit early due to the highly negatively charged protein surface.[73] We next analyzed the stability of the monomeric protein using CD spectroscopy.

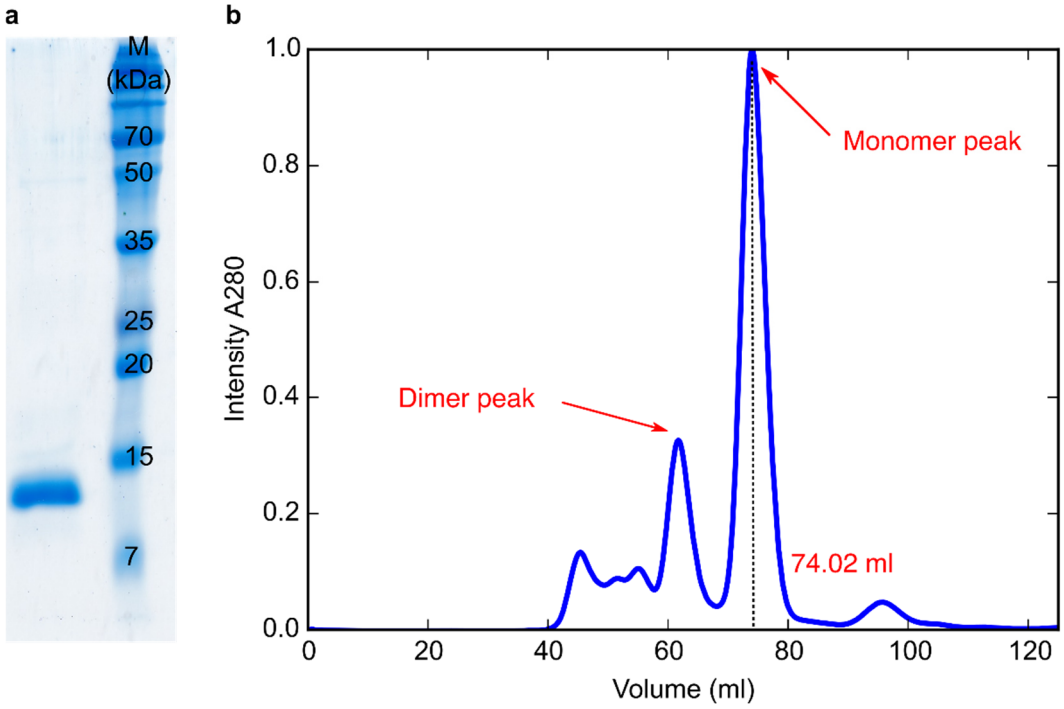

Figure 52: a, SDS-PAGE of hydrophobic core *Maquette* 2; **b**, size exclusion chromatography (SEC) profile of hydrophobic core *Maquette* 2. The major peak corresponds to monomeric protein, but there is also a small dimer peak. The monomer has a retention volume of 74.02 ml which corresponds to a molecular weight of 18 kDa.

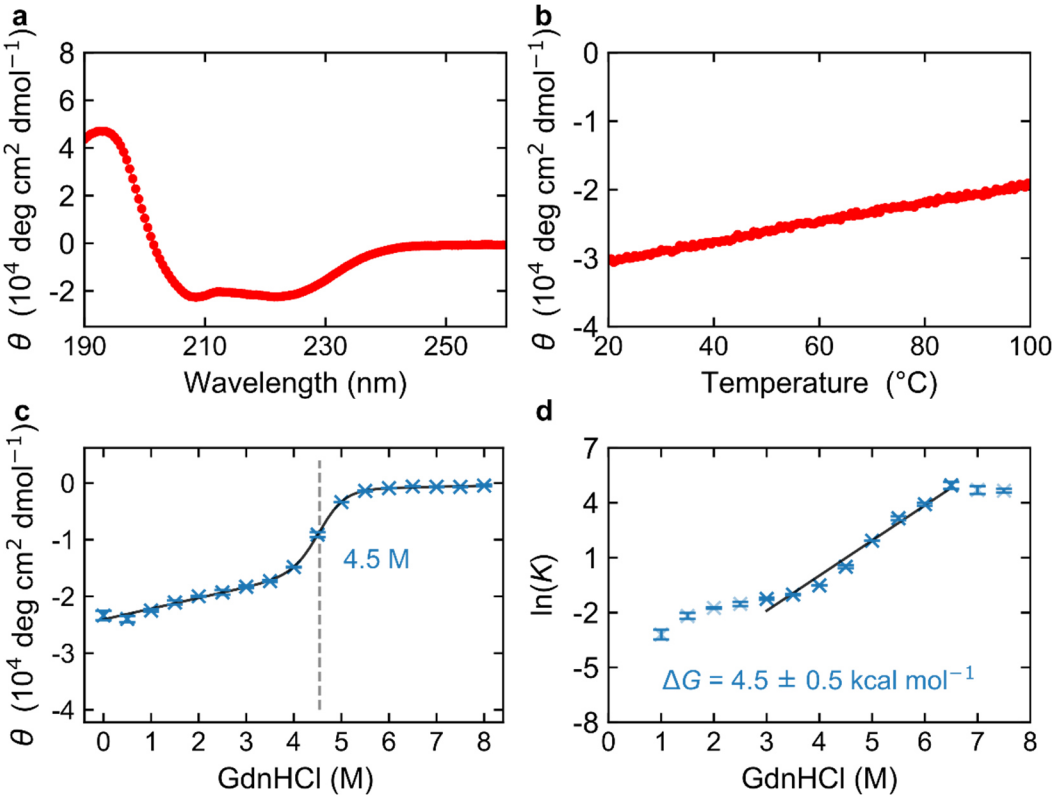

Figure 53: Experimental analysis of hydrophobic core *Maquette* 2. **a**, CD spectrum; **b**, melting curve; **c**, chemical unfolding with GdnHCl; **d**, ln*K* plotted against denaturant concentration.

The results of the CD spectroscopic measurements are presented in *Figure 53*. Panel a shows the molar ellipticity measurements between 190 – 260 nm that indicate a well-folded α-helical secondary structure. Melting analysis performed at the 222 nm spectrum minimum was unable to detect a transition to an unfolded state due to the protein's high thermal stability (*Figure 53b*). However, chemical unfolding experiments with GdnHCl were able to observe an unfolding transition at 4.5 M GdnHCl (*Figure 53c*), yields a Δ*G* of 4.5 ± 0.5 kcal mol^{-1} (*Figure 53d*).

4.3.2 Introduction of a single buried charged residue – glutamate at position 17

After the stability of the hydrophobic core model was determined, we next inserted an ion-pair by first creating a single buried charge in the *Maquette* 2 by substituting Phe17 with a glutamate. In our simulations, the mutation is introduced based on the converged structure of the hydrophobic core model, followed by further 200 ns MD simulation of the F17E model (*Figure 54*). Comparing the results of this construct to those of the hydrophobic core (*Figure 50*), the RMSD remains stable at around 2.5 Å after the first 10 ns (*Figure 54a*) and the RMSF reaches similar values as in the hydrophobic core model (*Figure 54b*). The water molecule content also does not change during the simulation, and remains moderate as in the hydrophobic core model (*Figure 54c*). The starting and end snapshots from the simulation are presented in *Figure 55*, and indicate that the side-chain of

Glu17 moves towards the solvent during the simulation (grey to red). Structural changes of the α-helices were not detected.

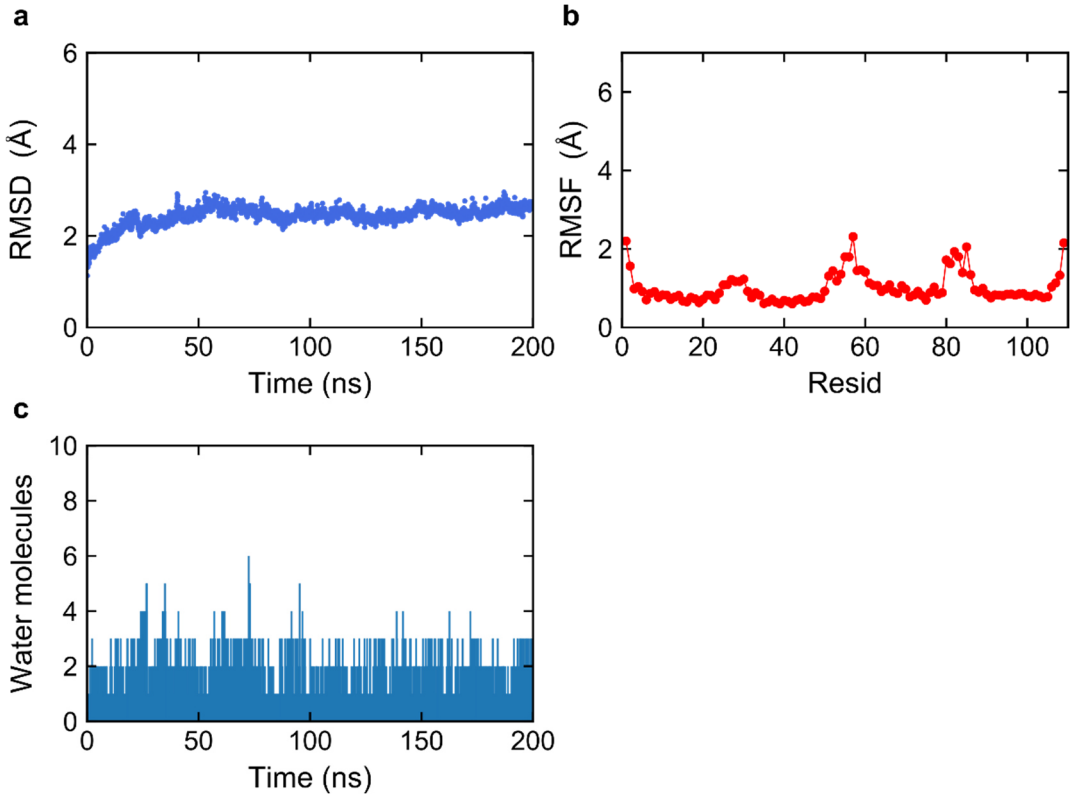

Figure 54: Analysis of 200 ns MD simulation of Glu17 *Maquette* 2. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content.

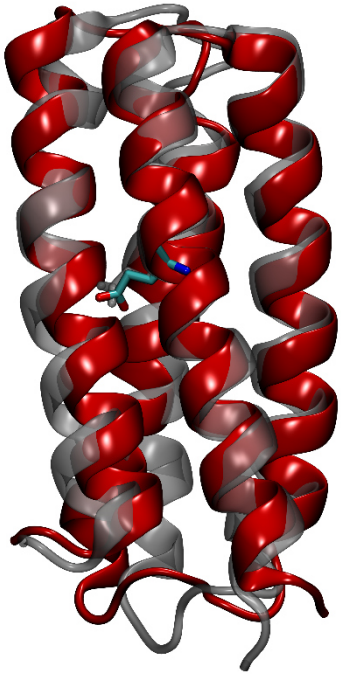

Figure 55: Snapshots from 200 ns MD simulation of Glu17 *Maquette* 2. The starting structure (grey) is aligned with the last frame (red) of the simulation. Glu17 is shown in licorice.

The results of the expression and the subsequent purification process of the half ion-pair were similar to the hydrophobic core model and therefore not reported here; the protein showed high purity with an apparent MW of 19 kDa. CD spectroscopy indicates a typical α-helical spectrum (*Figure 56a*), and thermal melting experiments indicated very high thermal stability (*Figure 56b*), although the precise melting point could not be accurately determined due to the lack of a clear sigmoidal inflection point, which is hinted at around 80°C and estimated here as the melting point (*Figure 56b*). The protein stability was also studied by chemical unfolding with GdnHCl resulting in an observed unfolding concentration of 3.1 M GdnHCl and a Δ*G* of 3.9 kcal mol^{-1} (*Figure 56c, d*). Therefore, the charged residue Glu17 destabilizes the *Maquette* 2 model by 0.6 kcal mol^{-1}.

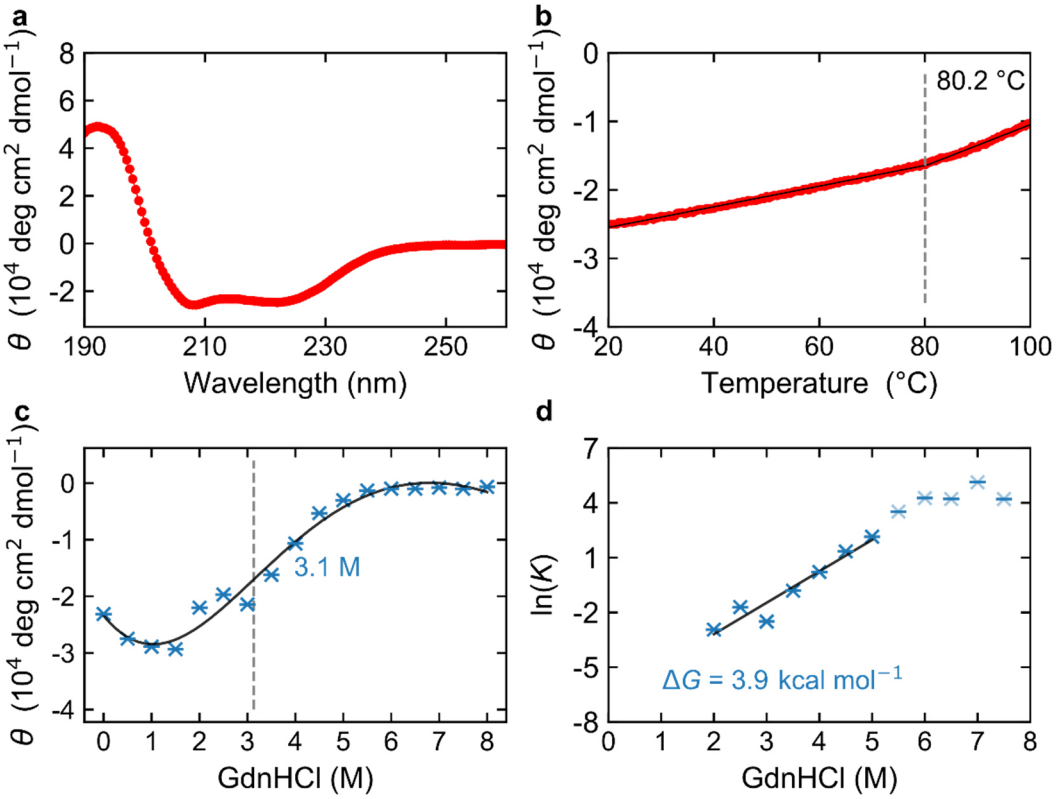

Figure 56: Experimental analysis of Glu17 *Maquette* 2. **a**, CD spectrum; **b**, melting curve; **c**, chemical unfolding with GdnHCl; **d**, ln*K* plotted against denaturant concentration.

4.3.3 Introduction of a single buried charged residue – lysine at position 72

Next, the other half of the ion-pair was introduced into the *Maquette* 2 model, by substituting Phe72 with a lysine residue. As for the Glu17 model, the Lys72 is introduced into the hydrophobic core *in silico*, based on the converged simulation model, followed by 200 MD simulation (*Figure 57*). Both the RMSD and RMSF are similar to the two other *Maquette* 2 models, although the loop connecting helices 2 and 3 appears to be slightly more flexible than the other loops (*Figure 57a,b*). The water penetration is low during the simulation, which indicates the hydrophobic core remains intact and

structured (*Figure 57*c). *Figure 58* presents the first and the last frames from the 200 ns simulation (grey and red, respectively). Lys72, points towards the bulk water during the entire simulation and no other structural changes are detectable.

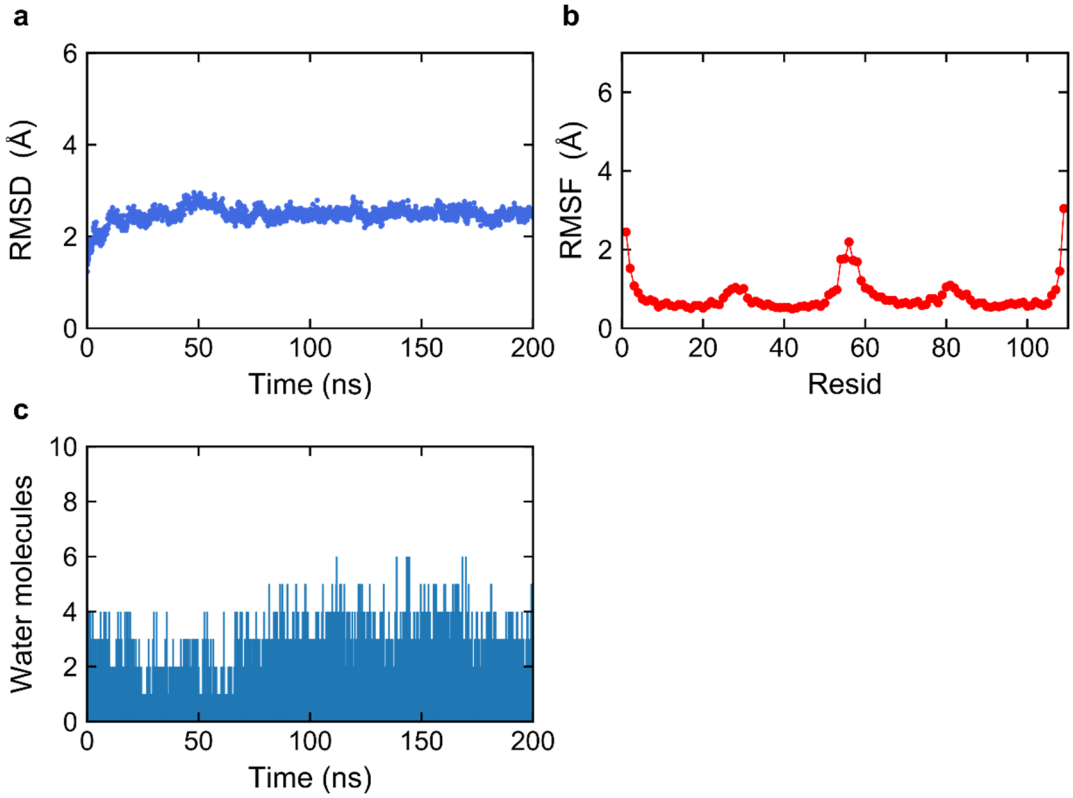

Figure 57: Analysis of *2*00 ns MD simulation of Lys72 *Maquette* 2. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content.

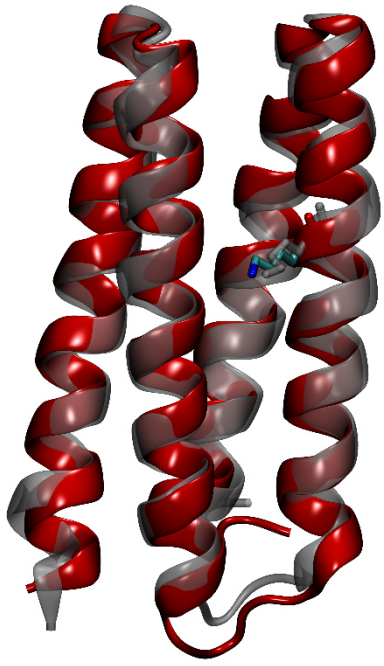

Figure 58: Snapshots from 200 ns MD simulation of Lys72 *Maquette* 2. The starting structure (grey) is aligned with the last frame (red) of the simulation. Lys72 is shown in licorice.

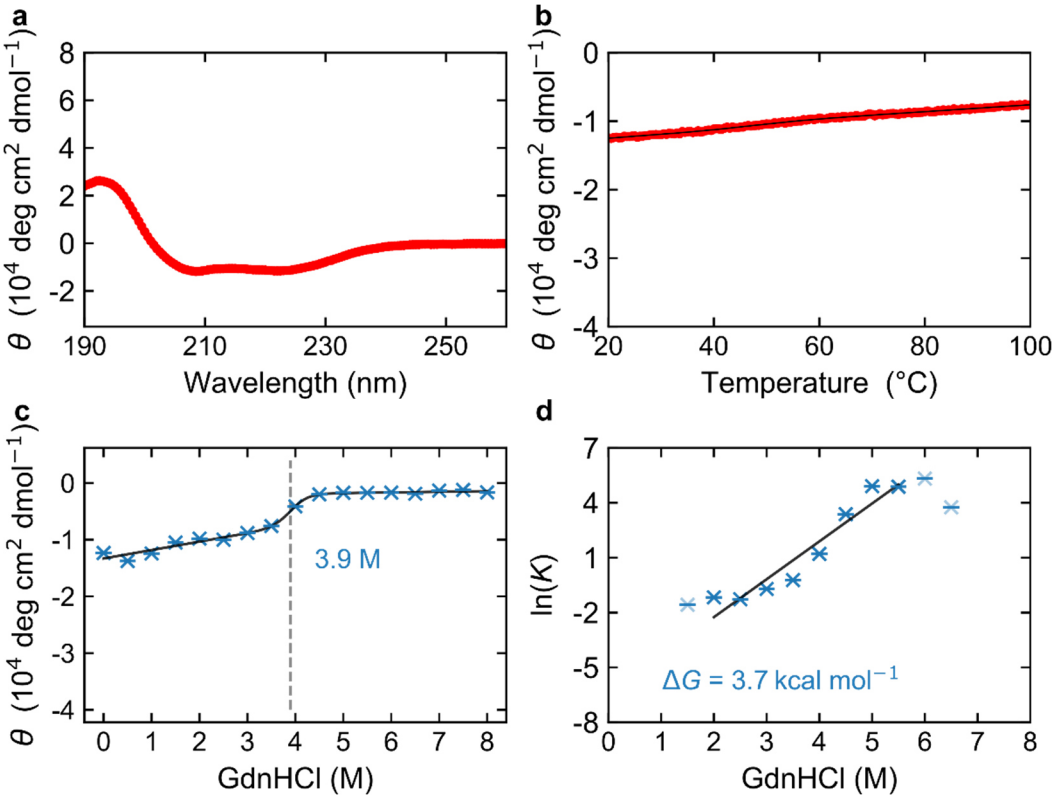

Figure 59: Experimental analysis of Lys72 *Maquette* 2. **a**, CD spectrum; **b**, melting curve; **c**, chemical unfolding with GdnHCl; **d**, ln*K* plotted against denaturant concentration.

Again, the results of the expression and purification process were similar to the constructs described above and are not presented here, but the protein was very pure and had a MW of 19 kDa as determined by SEC. CD analysis of the Lys72 model yielded less deep spectral minima, but still indicate an α-helical secondary structure (*Figure 59a*). The determination of the thermal stability was not possible for this model because while, the CD signal at 222 nm did decrease with increasing temperature, a two-state transition was not observed (*Figure 59b*). The chemical unfolding experiments result in an unfolding concentration of 3.9 M GdnHCl and a Δ*G* of 3.7 kcal mol^{-1} (*Figure 59c, d*) making the model 0.8 kcal mol^{-1} less stable than the hydrophobic core. Another important finding is, that the minima in the CD spectrum are less deep at the same sample concentration as for the hydrophobic core and Glu17.

4.3.4 Introduction of a buried ion-pair in the hydrophobic core (Glu17-Lys72)

Two different 1ip *Maquette* 2 models were investigated. The first one, described in this section, carries a Glu17-Lys72 ion-pair located in the hydrophobic part of the bundle. 200 ns of MD simulation were carried out as described for the single substitution constructs above (*Figure 61*).

The RMSD value reaches slightly higher values around 3 Å compared to the models discussed above, which show RMSDs around 2.5 Å,but this stays relatively stable until end of the simulation (*Figure 60a*). The RMSF plot indicates that the termini are the most flexible parts of the protein followed by the region around residue 60, which corresponds to loop 2 (*Figure 60b*). Although the solvation of the protein is higher compared to the previous *Maquette* 2 models (*Figure 60c*), the protein stays intact and the hydrophobic core remains more or less water-sealed. In panel d the distance between the δ – carbon and the ζ- nitrogen of Glu17 and Lys72 is plotted showing that the introduced ion-pair is closed during the whole simulation (*Figure 60d*). Unfortunately, in a subsequent 200 ns MD simulation the salt bridge does not form at all (*Figure 60e*). A possible explanation for this behavior would be that it is a dynamic process and that both a closed and open population exist.

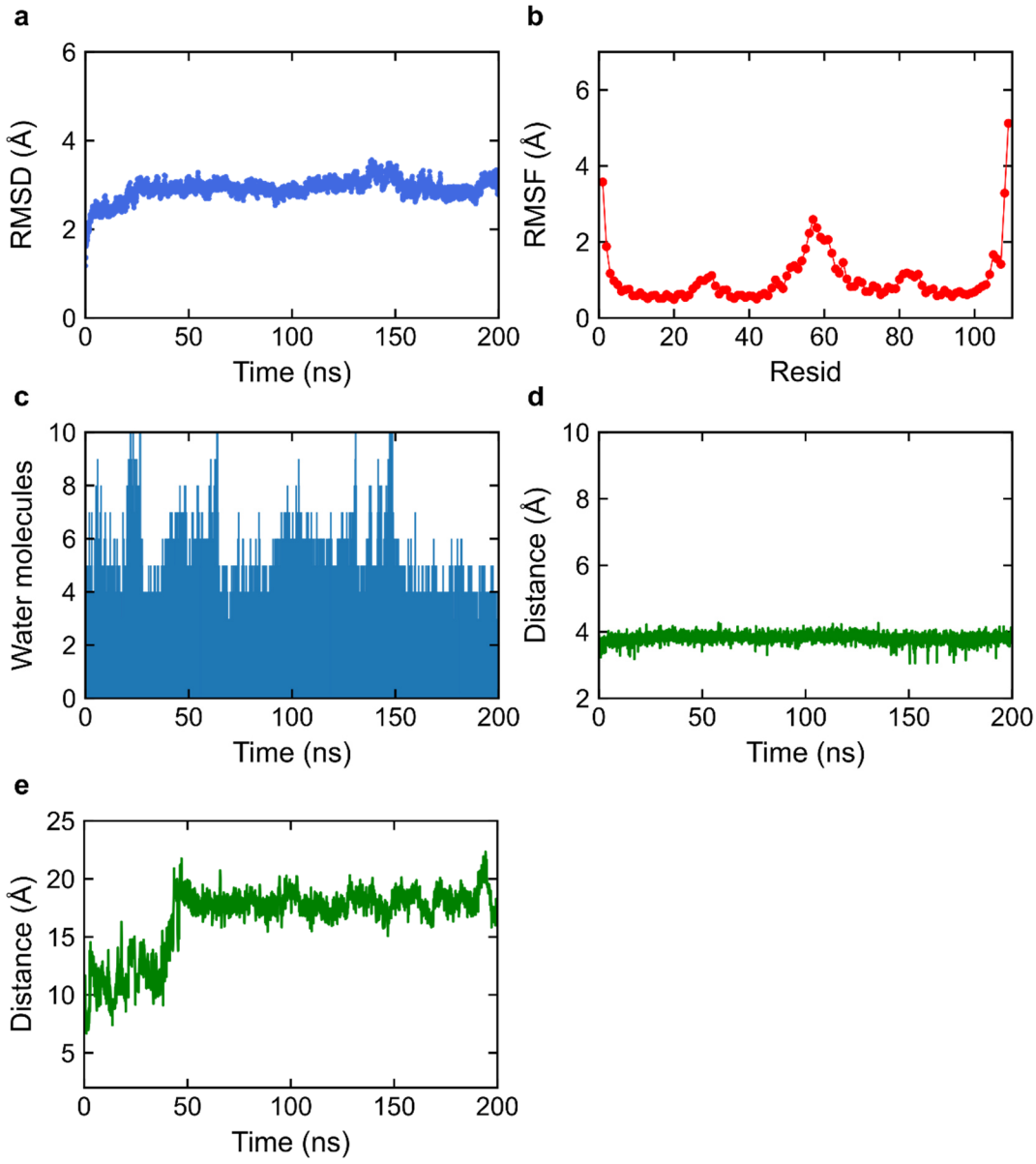

Figure 60: Analysis of *2*00 ns MD simulation of 1ip (Glu17-Lys72) *Maquette* 2. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content; **d,** distance profile between CD-NZ of Glu17 and Lys72; **e**, distance profile between CD-NZ of Glu17 and Lys72 of a second simulation run.

Two structural snapshots from the first 200 ns MD simulation are depicted in *Figure 61*. The closed ion-pair is visualized in licorice in both frames. The most interesting finding from the comparison of the starting structure (grey) and the last frame (red) is, that helix 3 which hosts Lys72, becomes kinked during the simulation while the other helices only shift slightly.

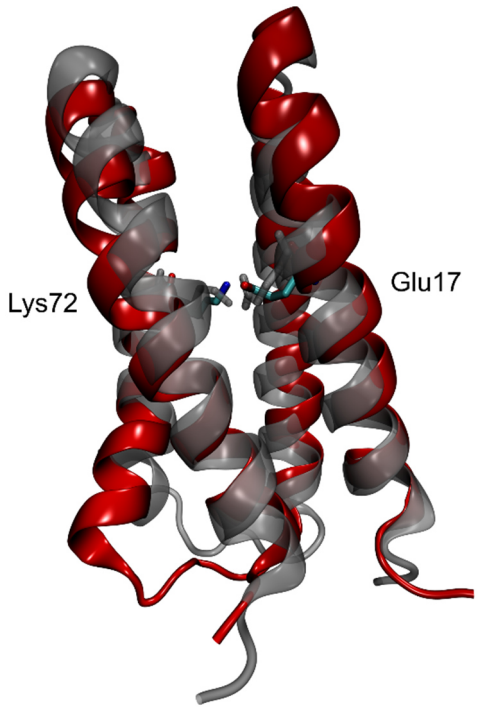

Figure 61: Snapshots from 200 ns MD simulation of 1ip (Glu17-Lys72) *Maquette* 2. The starting structure (grey) is aligned with the last frame (red) of the simulation. Charged residues are shown in licorice.

The 1ip *Maquette* 2 was also analyzed experimentally. *Figure 62a* demonstrates the high protein purity and proves the expected size on the SDS gel. The MW is determined by a SEC during the purification process (*Figure 62b*) where the main peak corresponds to a MW of 19 kDa and indicates monomeric protein. The secondary structure analyzed by CD spectroscopy shows a well-folded α-helical structure (*Figure 63a*) that is very temperature stable and only starts to melt above 90°C, but the transition point is not detectable by 100°C (*Figure 63b*). Therefore, to determine the unfolding energy, a GdnHCl series was measured at 222 nm (*Figure 63c*) and a ΔG of 2.6 kcal mol^{-1} and an unfolding concentration of 3.2 M GdnHCl were obtained (*Figure 63d*).

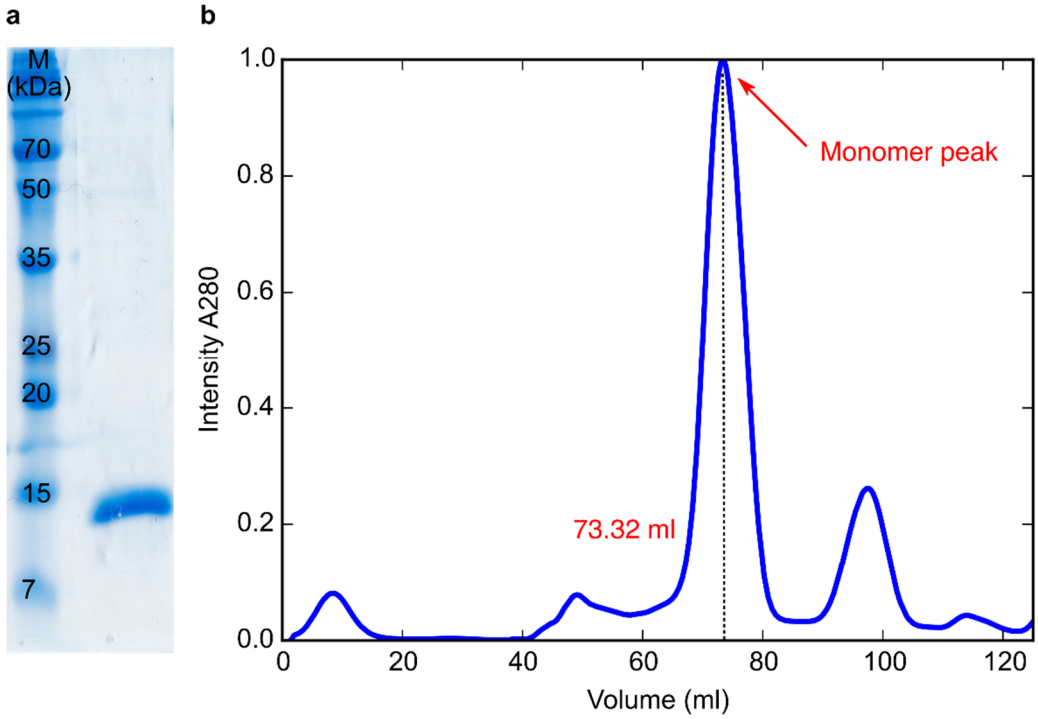

Figure 62: a, SDS-PAGE of 1ip (Glu17-Lys72) *Maquette* 2 during purification process; **b**, size exclusion chromatography (SEC) profile of 1ip (Glu17-Lys72) *Maquette* 2. The major peak corresponds to monomeric protein and has a retention volume of 73.32 ml which corresponds to a molecular weight of 19 kDa.

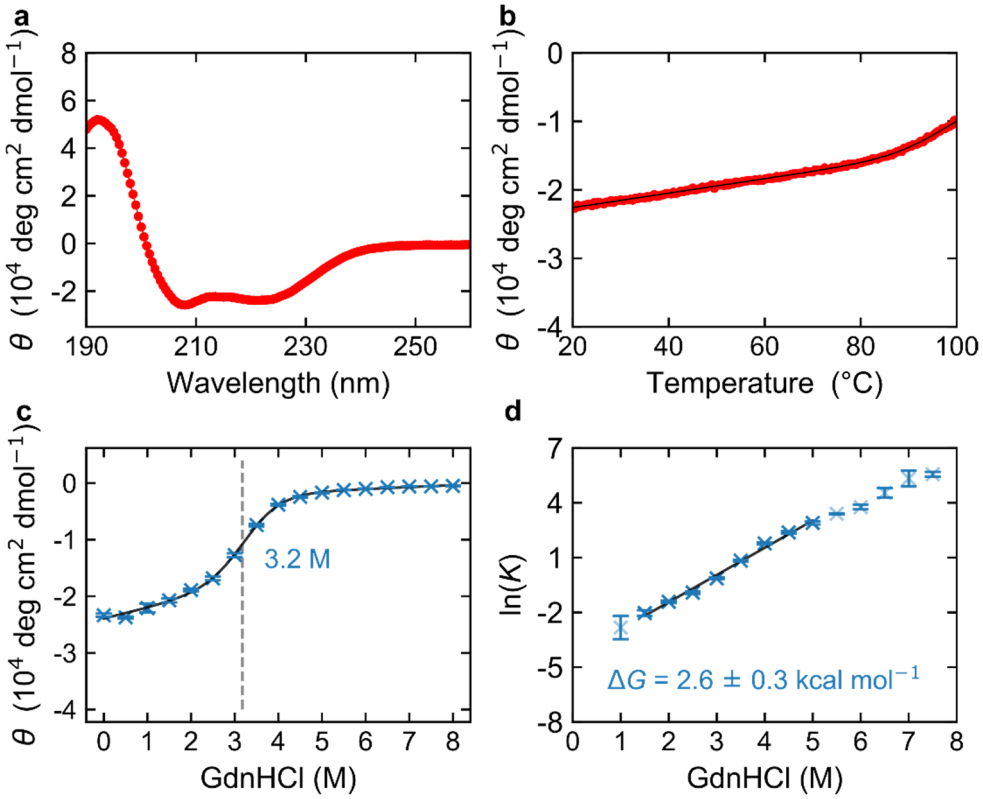

Figure 63: Experimental analysis of 1ip (Glu17-Lys72) *Maquette* 2. **a**, CD spectrum; **b**, melting curve; **c**, chemical unfolding with GdnHCl; **d**, ln*K* plotted against denaturant concentration.

These findings indicate a stable protein which is 1.9 kcal mol^{-1} less stable compared to the hydrophobic core. The introduced ion-pair is detectable in the simulation but seems to fluctuate. In order to investigate the protein's reaction to introducing two charged residues in the hydrophobic core and to study how exactly the ion-pair is buried in reality the protein was prepared for analysis by NMR (see section 4.3.6).

4.3.5 Introduction of an ion-pair in the hydrophilic part (Lys50-Glu101)

In order to contrast the effect of introducing an ion-pair in the hydrophilic part and not in the hydrophobic core, a second 1ip model was created in the water accessible part of the bundle where the porphyrin binds in the original model. Again, a 200 ns long MD simulation was performed after introducing the charged pair into the converged hydrophobic core model (*Figure 64*). Unlike with the other constructs, the RMSD value increases from 2 Å to 4 Å over the simulation timescale and the system does not converge and is still in the adaption process. The RMSF plot does not show significant differences compared to the previously described *Maquette* 2 models with the termini and the second loop being the most dynamic parts of the protein. The solvation of the bundle reaches a maximum of 9 water molecules (*Figure 64c*). Interestingly, the number of water molecules decreases significantly during the last 50 ns coinciding with when the introduced ion-pair Lys50-Glu101 opens having been primarily closed during the first 150 ns (*Figure 64d*). The beginning and concluding structural snapshots from the simulation are presented in *Figure 65* in grey and red. The change from the closed to the open state of the ion-pair is visible with both Lys50 and Glu101 pointing towards the water in the final (red) structure. The movement of the side-chains to the surroundings to compensate the charges results in slight rearrangements of all helices.

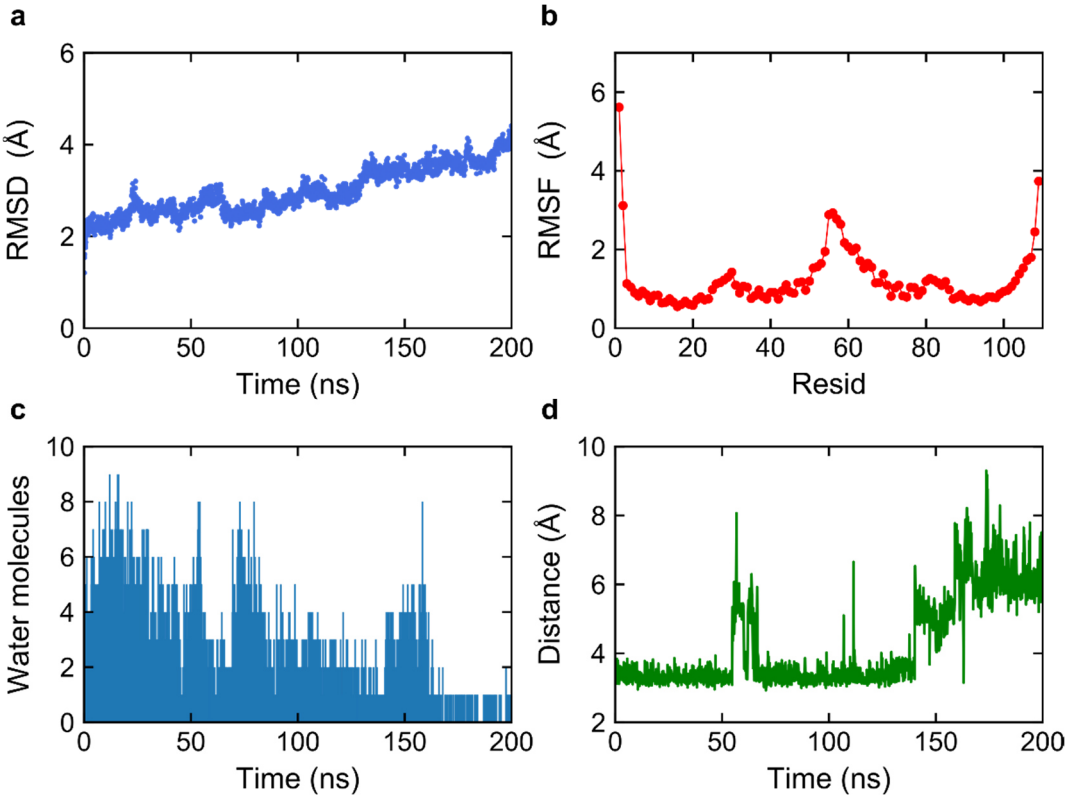

Figure 64: Analysis of 200 ns MD simulation of 1ip (Lys50-Glu101) *Maquette* 2. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content; **d**, distance profile between CD-NZ of Lys50 and Glu101.

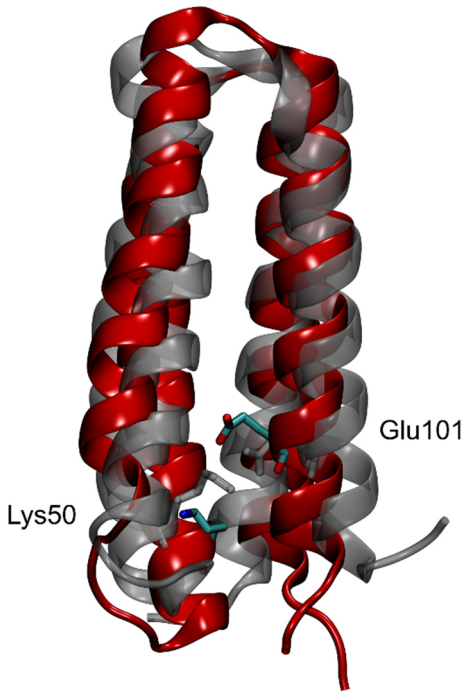

Figure 65: Snapshots from 200 ns MD simulation of 1ip (Lys50-Glu101) *Maquette* 2. The starting structure (grey) is aligned with the last frame (red) of the simulation. Charged residues are shown in licorice.

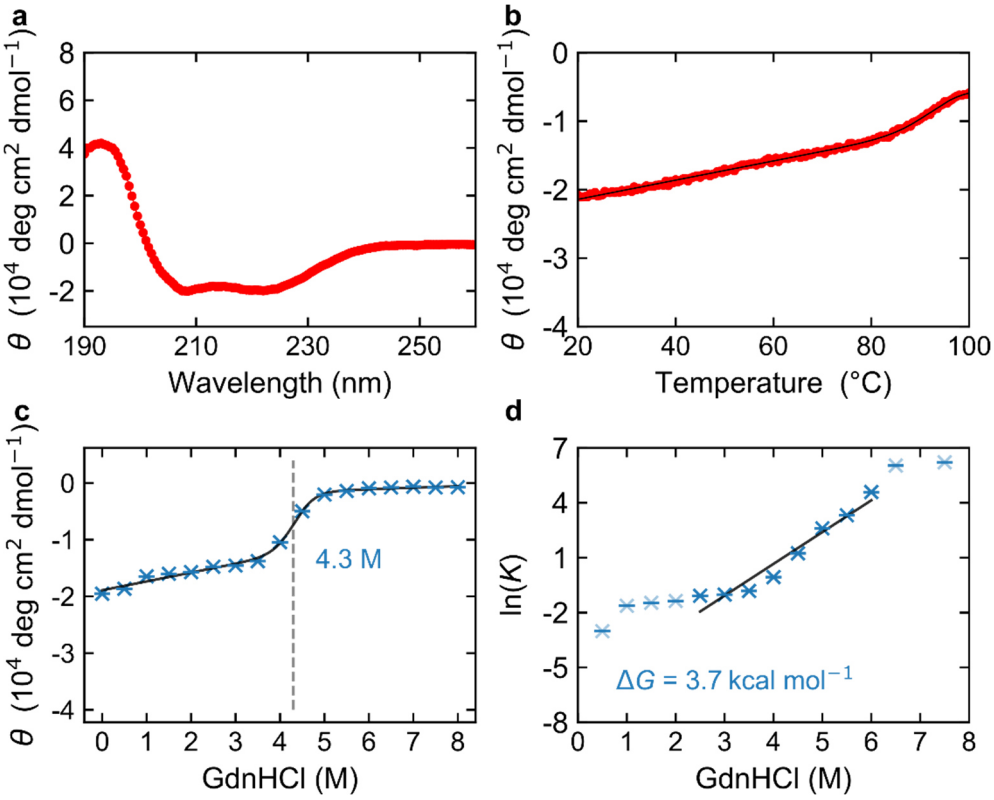

Figure 66: Experimental analysis of 1ip (Lys50-Glu101) *Maquette* 2. **a**, CD spectrum; **b**, melting curve; **c**, chemical unfolding with GdnHCl; **d**, ln*K* plotted against denaturant concentration.

The 1ip (Lys50-Glu101) *Maquette* 2 was also analyzed experimentally. Similar results for the expression and purification were obtained and so are not reported here. The high purity and the expected size of 20 kDa were shown by both SDS-PAGE and SEC. Secondary structure analysis and protein stability determined by CD spectroscopy show a typical α-helical spectrum with distinct minima at 208 and 222 nm and high thermal stability with unfolding only beginning over 90°C (*Figure 66a, b*). The chemical unfolding analysis leads to an unfolding transition at a concentration of 4.3 M GdnHCl, which is similar to the hydrophobic core (*Figure 66c*). The unfolding energy, Δ*G*, was calculated to be 3.7 kcal mol^{-1}, which resembles the values of the half ion-pair constructs (3.7 kcal mol^{-1} and 3.9 kcal mol^{-1}) (*Figure 66d*).

Taken all together, introducing an ion-pair (Lys50-Glu101) into the hydrophilic part of *Maquette* 2 resulted in a very stable protein, as indicated by MD simulation and experiments. This model is 1.1 kcal mol^{-1} more stable compared to the ion-pair in the hydrophobic core (Glu17-Lys72) of *Maquette* 2, but nonetheless still 0.8 kcal mol^{-1} less than the hydrophobic core *Maquette* 2 indicating that introducing an ion-pair still comes at an energetic cost.

4.3.6 NMR studies on *Maquette* 2

In order to better understand the structure and dynamics of the models, different isotopically-labeled NMR samples of the hydrophobic core and 1ip (Glu17-Lys72) *Maquette* 2 were prepared with a protein concentration of 800 µM. The NMR measurements were conducted by Dr. Sam Asami (Sattler Lab). First, $^{1}H^{15}N$-HSQC spectra of both models were recorded using a 600 MHz Bruker Avance x at 25°C. Both proteins show very well dispersed HSQC spectra, which indicate that they are folded (*Figure 67*). The measured HSQC of the hydrophobic core is only shifted globally due to a different spectrometer compared to the published apo PS1 HSQC (*Figure 67*, black).[73] *Figure 67* also presents the HSQC of 1ip (Glu17-Lys72) *Maquette* 2 with 98 peaks (red). The spectra are superimposed to highlight the peak shifts that result from the two introduced charged residues. Additionally, some peaks completely disappear and some new peaks appear. The mutated positions are labeled in both spectra. Mutation of the Phe17 and Phe72 to Glu17 and Lys72 shifts the corresponding peaks by ~ 0.2 ppm (^{1}H). The replacement of two aromatic residues inside the hydrophobic protein core leads to significant shifts of other resonances as well, which can be seen by the difference between the spectra (*Figure 67*). The chemical shifts in the proton dimension (dH) between hydrophobic core and 1ip (Glu17-Lys72) *Maquette* 2 for all residues that could be assigned in both spectra are visualized in *Figure 68* with the two mutation sites marked in red. Interestingly, these two residues do not show the greatest shift. Residue 73 (Gln) shows the highest shift due to the replacement of the aromatic neighbor by a charged lysine.

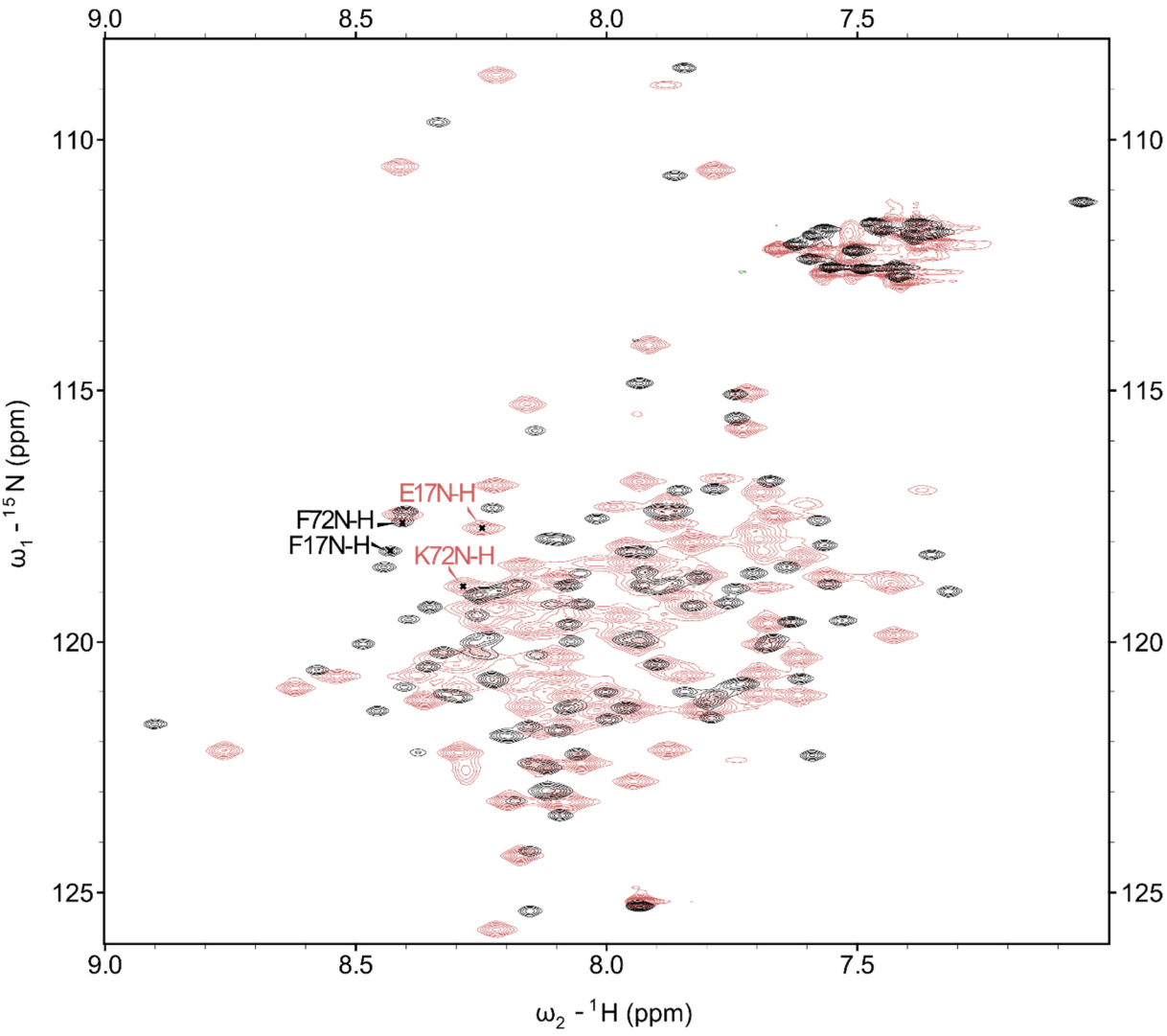

Figure 67: HSQCs of hydrophobic core and 1ip (Glu17-Lys72) *Maquette* 2. The hydrophobic core spectrum is shown in black and the 1ip in red. The mutated positions F72K and F17E are labeled.

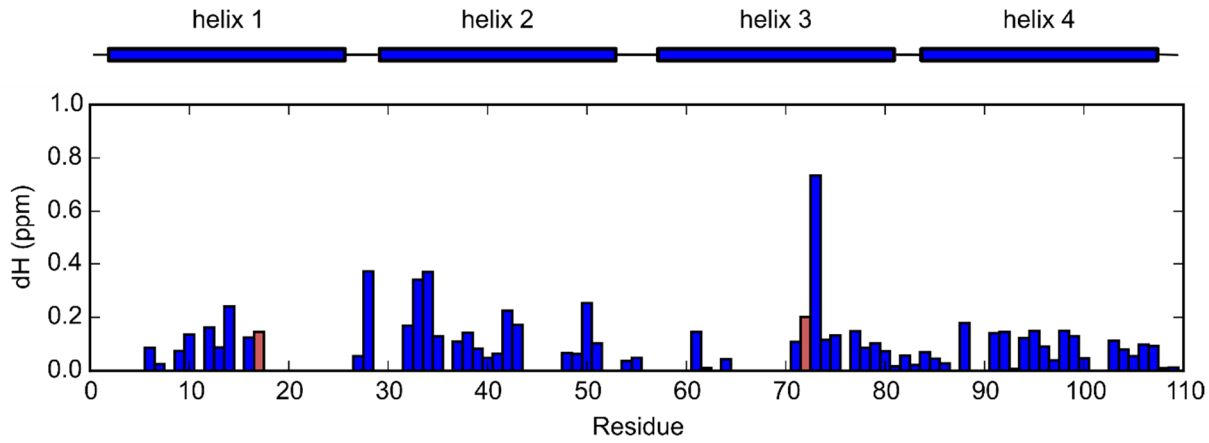

Figure 68: Chemical shift perturbation plot (CSP). The secondary structure elements are shown at the top. Existing dH in ppm is plotted for each residue, which is only possible for assigned residues. The two mutated positions 17 and 72 are marked in red.

4.3.7 Structure determination of 1ip (Glu17-Lys72) *Maquette* 2

One of the main reasons for selecting *Maquette* 2 as the basis for this work is that it has an experimentally solved structure.[73] Therefore, in order to answer the main question of this thesis, the structure of the 1ip (Glu17-Lys72) model need to be solved. It is necessary to know the structure to study and understand the consequences of introducing charged residues into the protein's hydrophobic core. Therefore, a series of 3D NMR spectra was recorded with a $^{15}N^{13}C$ labeled sample. The experiments were again measured at a 600 MHz Bruker Avance x at 25°C. The sequence specific backbone assignment was conducted with HNCA, HNCACO, HNCACB, CBCA(CO)NH, HBHA(CO)NH, H(CCO)NH, and CC(CO)NH using Sparky.[103] Additionally, HC(C)H/(H)CCH TOCSYs, $^{15}N/^{13}C_{ali}/^{13}C_{aro}$-NOESYs, and a 2D (HB)CB(CGCD)HD were measured and automatically assigned to obtain as much as possible structure information and side-chain assignments. *Figure 69* presents the backbone assignment of 1ip (Glu17-Lys72) *Maquette* 2. 78 peaks out of 98 are successfully assigned. The structure calculation itself was performed together with Dr. Sam Asami (Sattler lab) using CYANA.[106] There were two different ways of calculating the structure. One approach was to calculate a *de novo* structure while the second approach was to use restraints obtained from the MD simulations to create a closed ion-pair in the NMR structure. Both results are shown in *Figure 70*. For getting the closed structure the distance between Glu17-OE1 and Lys72-NZ was restrained to approximately 3 Å. *Figure 70* presents the whole *de novo* calculated NMR structure bundle with 20 different structures. The model with the smallest CD-NZ distance between Glu17 and Lys72 is marked in red and the mutated residues are highlighted in licorice. The other 19 models are shown in light grey. The backbone RMSD of the obtained structures is ~ 0.75 Å. The 20 different models are very similar and create a very compact, converged bundle of structures. The two charged residues try to reach the surrounding solvent and the side-chains point towards the water. The most dynamic parts with the highest variance are the loop connecting helices 2 and 3 as well as the termini. The positions of the individual helices are almost identical. The same effects can be seen in *Figure 70* for the closed ion-pair NMR structures which were obtained by putting a distance restraint on the ion-pair. Again, the frame with the closest ion-pair distance is marked in red, while the other 19 structures are colored in light grey. The mutated residues are visualized in licorice and the distance between them fluctuates around 3 Å in all 20 models. The backbone RMSD of this structure bundle is also around 0.75 Å. An interesting finding in the solved structure is the π-cation interaction between Trp68 and Lys72 (*Figure 71*). Trp68 is ~ 3 Å away from Lys72 and stabilizes the lysine in the closed conformation.

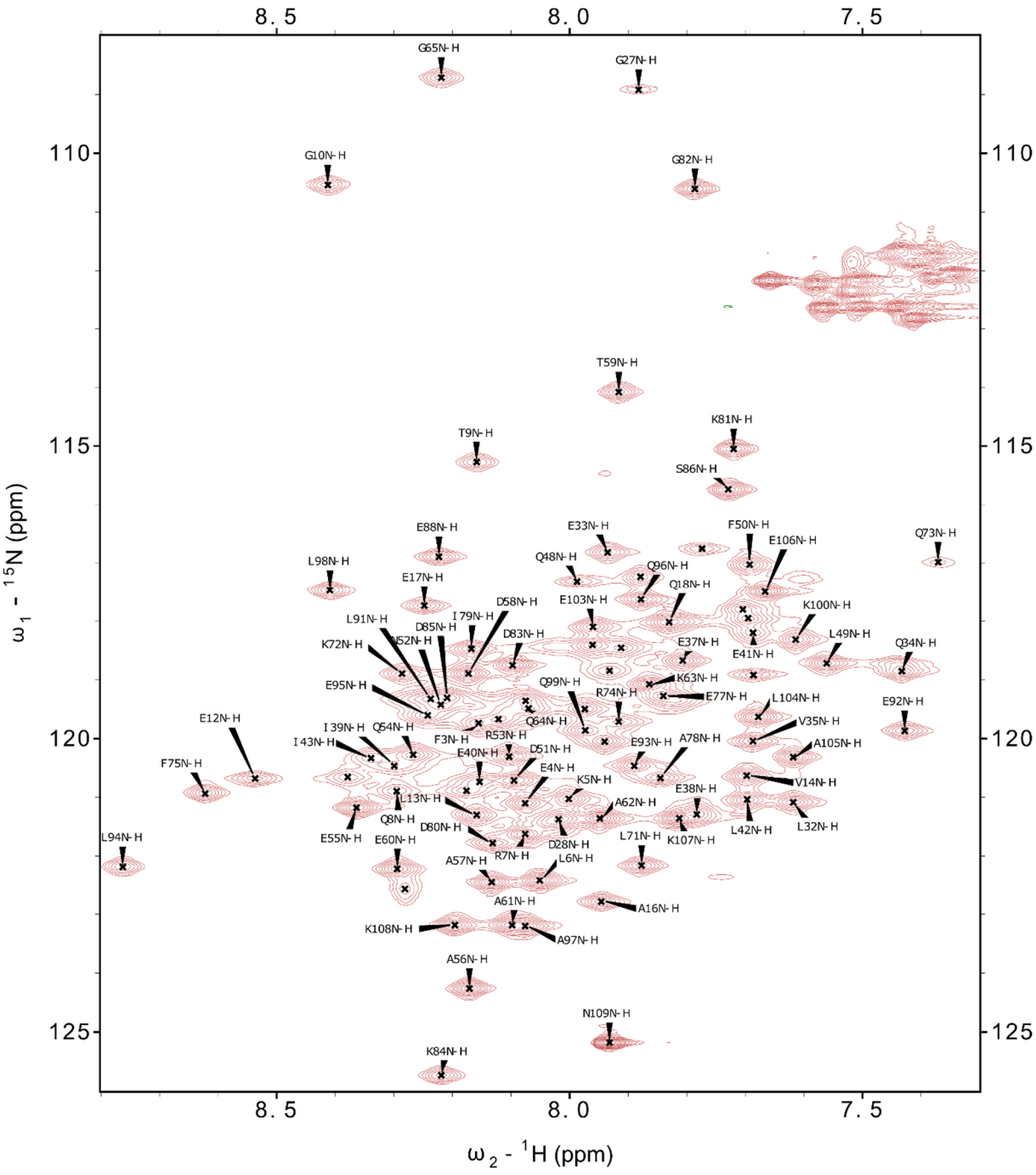

Figure 69: Assigned HSQC of 1ip (Glu17-Lys72) *Maquette* 2.

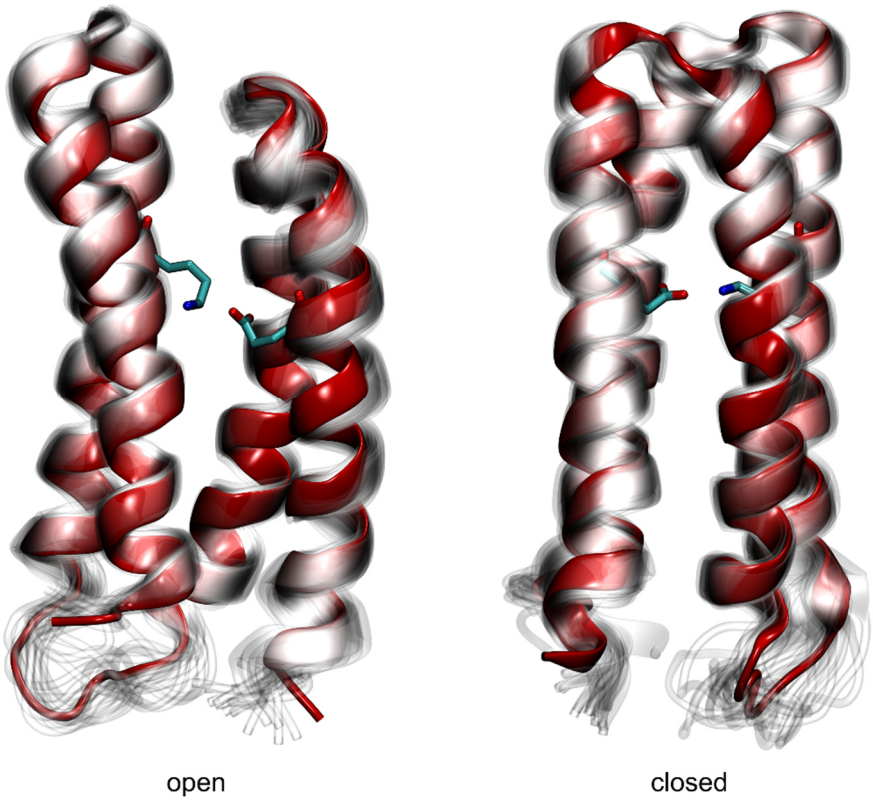

Figure 70: NMR structure bundles for the *de novo* open structure and the restraint closed structure. Each bundle consists of 20 single structures. The model with the lowest CD-NZ distance between Glu17 and Lys72 is highlighted in red and the mutated residues in licorice. All other 19 models are shown in light grey.

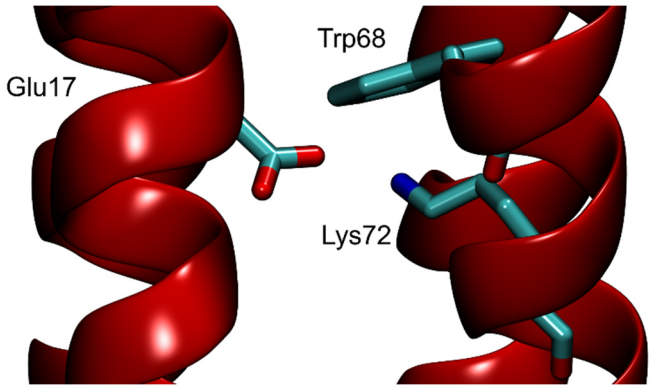

Figure 71: Zoom into the closed NMR structure. The interacting residues are shown in licorice. Trp68 is ~ 3 Å away from Lys72.

The calculated NMR structures are also compared to the MD models. *Figure 72* shows the open/closed MD model (grey) and the open *de novo* NMR model (PDB ID: 6Z35; red). The helices are located in slightly different positions especially helices 1 and 3. It seems that a small inwards rotation of helices 1 and 3 would be enough to close the ion-pair (*Figure 72b*). Also, a very interesting finding is that the unusual kink in helix 3 in the closed MD simulation is not present in the NMR model and therefore seems to be an artefact. Comparing the NMR structure with the closed MD model, helix 1 is also rotated slightly while helix 3 shows less rotation.

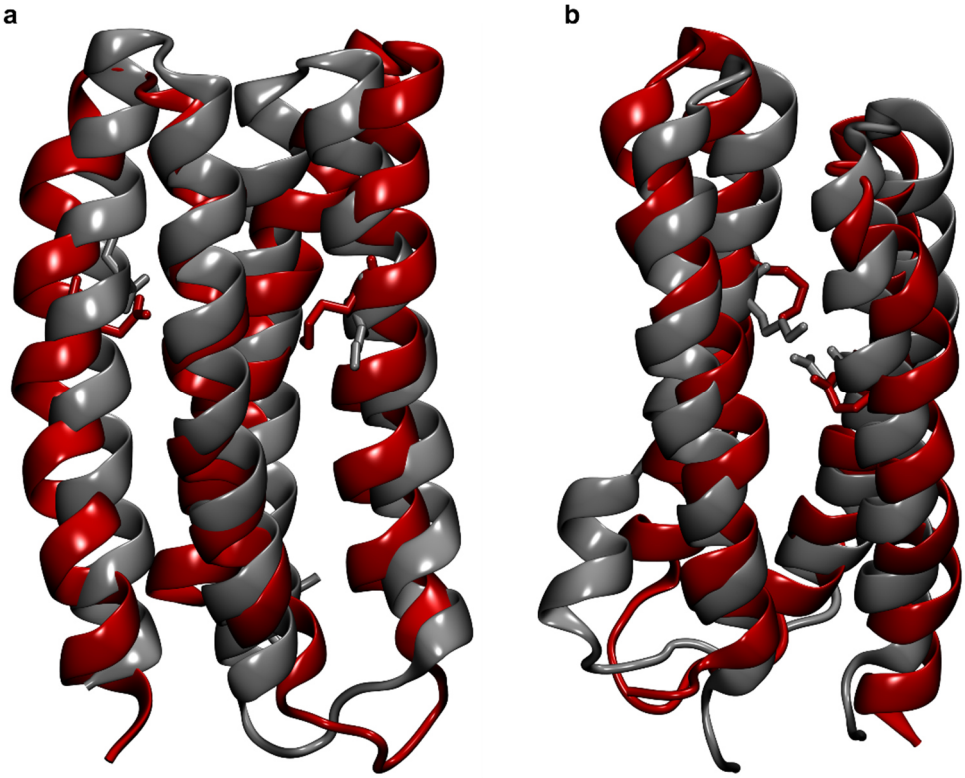

Figure 72: *De novo* open NMR structure compared to two MD simulation models. The NMR structure is always shown in red and the MD model in grey. **a**, NMR structure compared to open ion-pair MD model; **b**, NMR structure compared to closed ion-pair MD model. PDB ID: 6Z35.

These structural findings along with the MD simulation results indicate that the closed ion-pair conformation exists, but is less populated than the open state. The system shows strong dynamics and it seems that the surrounding of the ion-pair is not ideal for stabilizing it. The introduced ion-pair leads to a structural rearrangement. The possible reorganization to close the ion-pair in the NMR structure is obtained by adding a restraint on the CD-NZ distance of Glu17 and Lys72 during the structural calculation. The restraint resulted from analysis of the MD simulations. The structure (red) is also compared to the closed MD model (grey) (*Figure 73*). The helices are located in similar positions, but helix 3 shows an improved α-helical secondary structure in the NMR model. The kink of the MD structure disappeared as mentioned before. It was not possible to solve an NMR structure with the closed ion-pair conformation. However, the MD simulations suggest that there is a population where Glu17 and Lys72 are in contact and therefore, another method to experimentally detect the salt bridge is used.

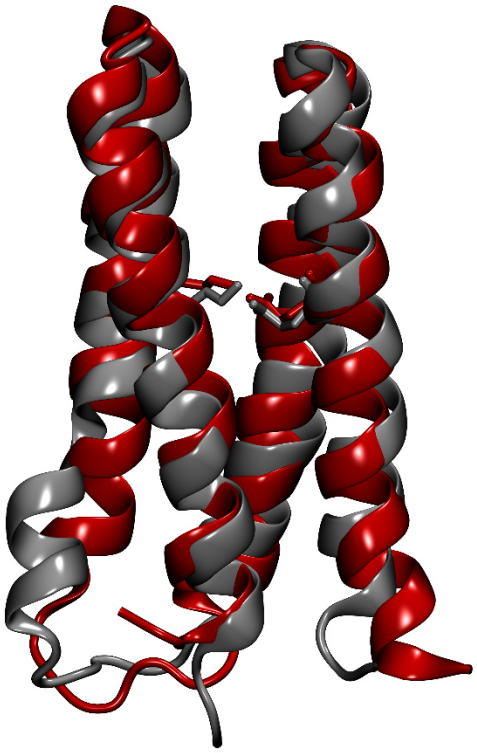

Figure 73: Closed NMR structure compared to MD simulation model. The NMR model is shown in red and the MD model in grey. The charged residues are marked in licorice.

4.3.8 HISQC spectrum for salt bridge detection

The *de novo* structure calculations failed to obtain a closed ion-pair conformation. To check the possibility of an existing closed ion-pair Glu17-Lys72, another detection method called heteronuclear in-phase single quantum coherence spectroscopy (HISQC)[134] is applied. This detects the NH_3-groups of lysines, which is normally very challenging due to the fast hydrogen exchange with water. The exchange is slowed down depending on the chemical surrounding of the lysine for example through electrostatic interaction in a salt bridge.[134] As the hydrophobic core also shows an NH_3 group peak at ~ 32 (^{15}N) ppm a series of experiments were performed to prove the existence of the introduced salt bridge (*Figure 74*, black). First, the 1ip (Glu17-Lys72) model was analyzed and compared to the hydrophobic core. Here, a second peak around 33 (^{15}N) ppm appears, which could be the introduced salt bridge. To confirm this the single buried charge models Lys72 (light orange) and Glu17 (orange) were also measured. The obtained HISQC spectra are overlaid in *Figure 74*. The additional peak is only present in the spectrum of the 1ip model. All spectra show the same main peak, which corresponds to a possible surface ion-pair. Interestingly, the additional peak is less intense compared to the second peak, which means that the introduced peak is not present in the whole population and there is in fact a small population with a closed Glu17-Lys72 salt bridge. This finding is consistent with the obtained NMR structures and the MD simulation results.

The most promising candidate for the surface ion-pair is Lys100-Glu103, which is present in MD simulations of hydrophobic core and 1ip. Both NMR structures were simulated for 200 ns and described in the next section.

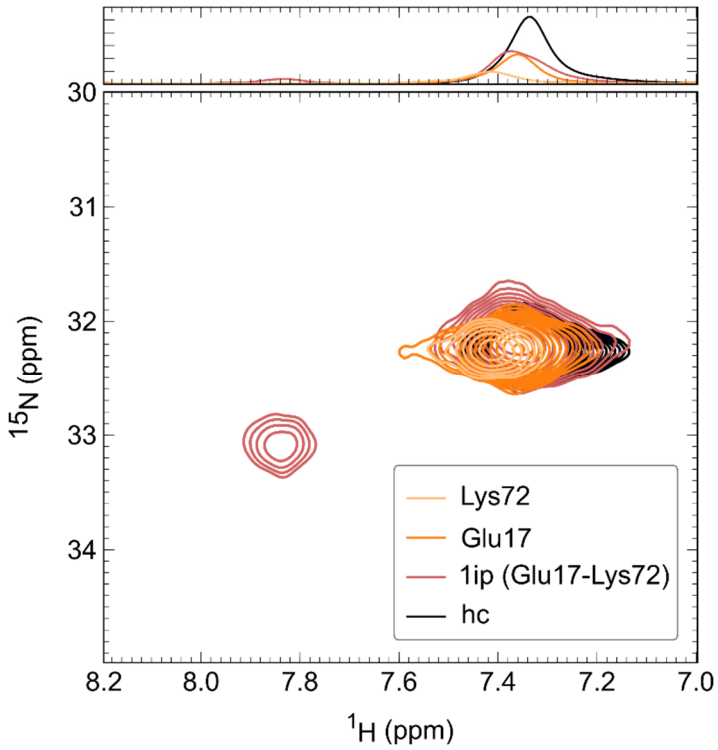

Figure 74: HISQC spectra of hydrophobic core, Glu17, Lys72, and 1ip (Glu17-Lys72) *Maquette* 2 for salt bridge detection. Lys72 is shown in light orange, Glu17 in orange, 1ip (Glu17-Lys72) in red, and hydrophobic core in black.

4.3.9 MD simulations of NMR structures

MD simulations were also performed for the calculated NMR structures. Model 17 was selected for the closed structure and model 14 for the open one. The structures were solvated and ions were added to the system. After relaxing the water molecules and the protein, a 200 ns production run was performed.

Figure 75 presents the analysis of the simulation of the open *de novo* structure. The system converges and the RMSD is stable around 3 Å after a few nanoseconds (*Figure 75a*). The RMSF plot indicates the three connecting loops and the termini to be more dynamic than the helices (*Figure 75b*). The water content indicates that the structure stays relatively water sealed and on average only 3 to 4 water molecules were detected during the 200 ns (*Figure 75c*). Therefore, the stability of the hydrophobic core seems to be maintained. The CD-NZ distance between the introduced Glu17 and Lys72 indicates that the ion-pair does not close with the distance stabilizing around 18 Å. The charged side-chains remain pointing towards the solvent instead of traversing the hydrophobic core

to form a salt bridge. The protein compensates for the charges by reorganizing the side-chains. These findings are also visualized with two structural snapshots presented in *Figure 76* where the first (grey) and the last (red) frames of the simulation do not show significant differences. The system stays stable and holds its conformation during the simulation.

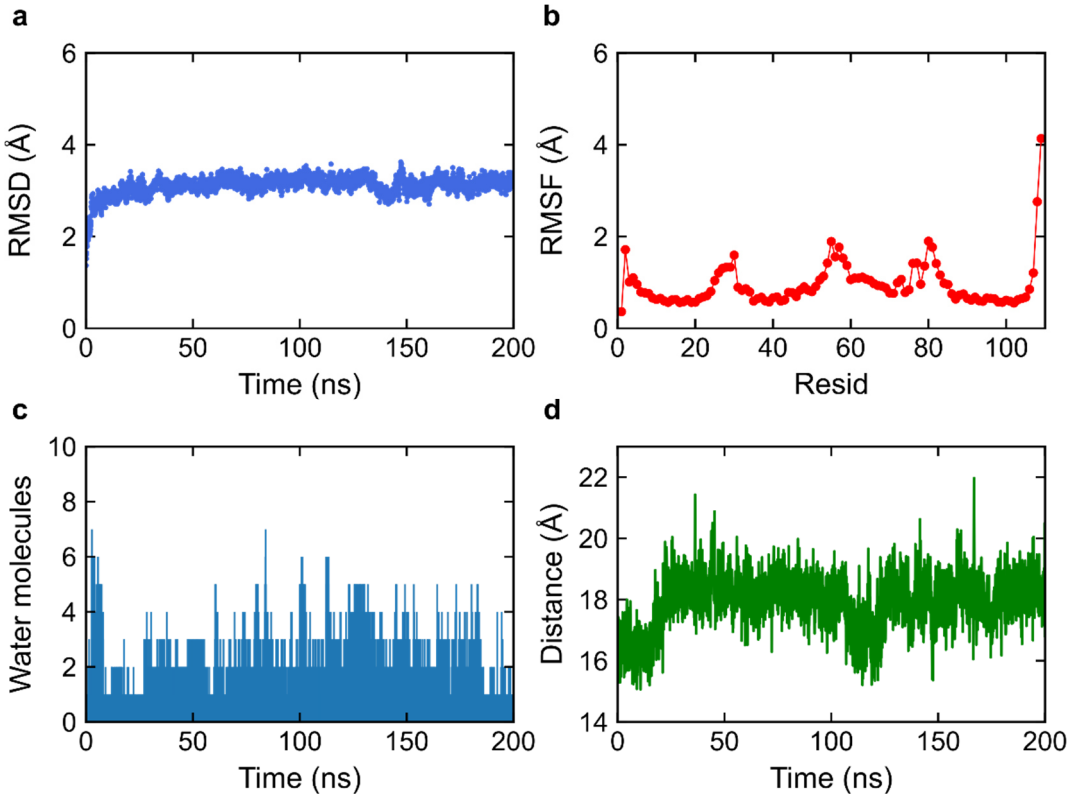

Figure 75: Analysis of *2*00 ns MD simulation of the *de novo* 1ip (Glu17-Lys72) *Maquette* 2 NMR structure. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content; **d**, distance profile between CD-NZ of Glu17 and Lys72.

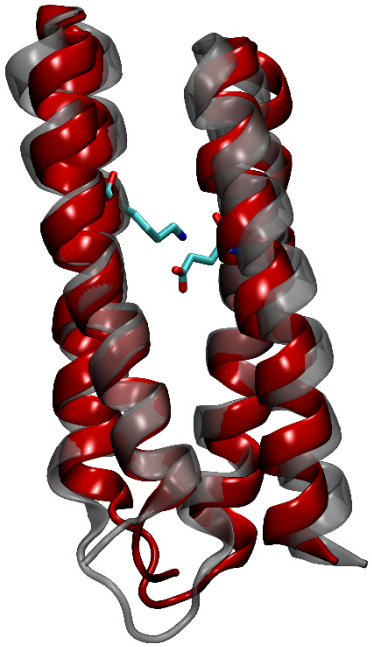

Figure 76: Two structural snapshots during MD simulation of the *de novo* NMR structure. The starting structure is shown in grey and the last frame in red. The charged residues are marked in licorice.

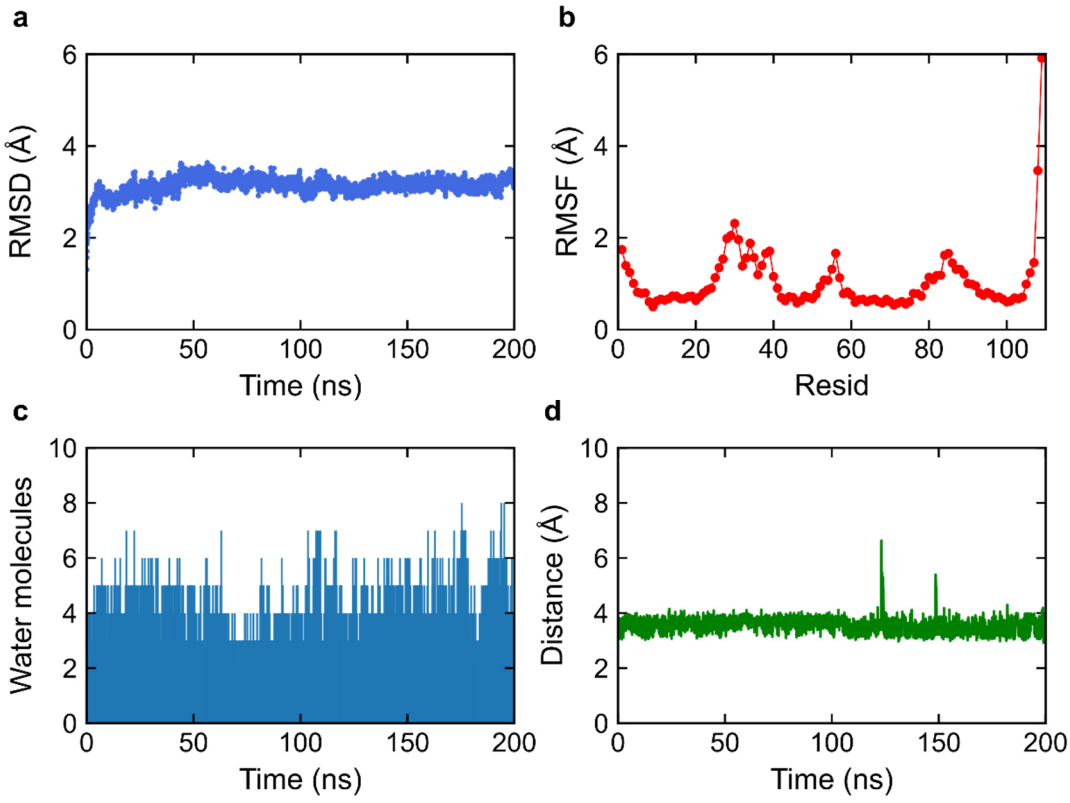

Figure 77: Analysis of *2*00 ns MD simulation of the restrained 1ip (Glu17-Lys72) *Maquette* 2 NMR structure. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content; **d**, distance profile between CD-NZ of Glu17 and Lys72.

Figure 77 shows the simulation results of the restrained closed NMR structure. The RMSD and RMSF values behave almost identically to the *de novo* structure (*Figure 77a, b*), but the internal solvation is different in this system. The total number of water molecules is similar, the two charged residues attract a few water molecules into the protein interior (*Figure 77c*). The distance between the ion-pair as shown in panel d indicates that Glu17 and Lys72 stay in close proximity to each other at 3-4 Å, which is a quite ideal distance for a salt bridge. The restrained NMR structure behaves similar to the simulated MD model where the ion-pair is closed. These findings are also highlighted in *Figure 78* where the first (grey) and the last (red) frame of the simulation are almost identical. Therefore, the system is stable during the MD simulation and the ion-pair is closed.

Both calculated NMR structures reflect the results of different MD simulations of the *in silico* 1ip (Glu17-Lys72) *Maquette* 2 model.

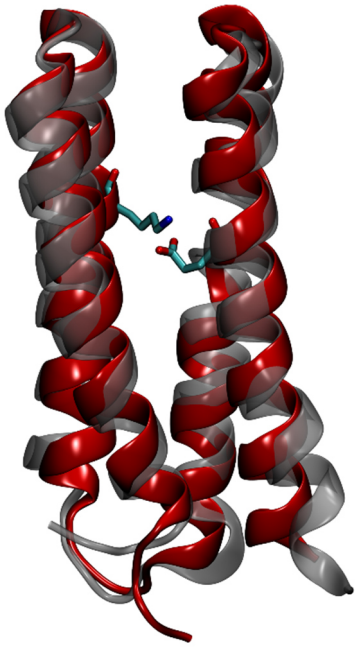

Figure 78: Two structural snapshots during MD simulation of the restrained NMR structure. The starting structure is shown in grey and the last frame in red. The charged residues are marked in licorice.

4.3.10 Computational design with PACMAN

Our combined results strongly suggest that the closed ion-pair model forms a stable structure, but the closed form of the salt bridge may exist in small proportion that it is dynamically flexible. As the next goal, we aimed to optimize the surrounding of the Glu17-Lys72 in order to stabilize it and increase the population of the closed state. A computational design approach called PACMAN, which is described in section 3.3.4 was performed by Max Mühlbauer (Kaila Lab). Briefly, the design was directed by MD simulations, random *in silico* mutations were inserted and a Metropolis Monte Carlo (MC)-criterion was applied to improve the ion-pair interaction and to optimize the protein structure. The PACMAN simulations resulted in structures where 80% of the introduced amino acids are polar or charged, in addition to charged amino acids (Gln/Asp, Tyr, Ser/Thr). These amphiphilic residues were placed next to the ion-pair, which resulted in its stabilization (*Figure 79*). To verify these results, the environment of ion-pairs in natural membrane proteins was also analyzed (Michael Röpke, Kaila Lab), resembling the predicted substitutions. Of particular interest is the finding that Trp residues are often located in vicinity of lysine residues where they can form stabilizing π-cation interactions, resembling the interaction between Trp68 and Lys72 in the closed ion-pair conformation of the *Maquette* 2 model. The suggestion that an amphiphilic residue next to an ion-pair can have a stabilizing effect is analyzed in the following section.

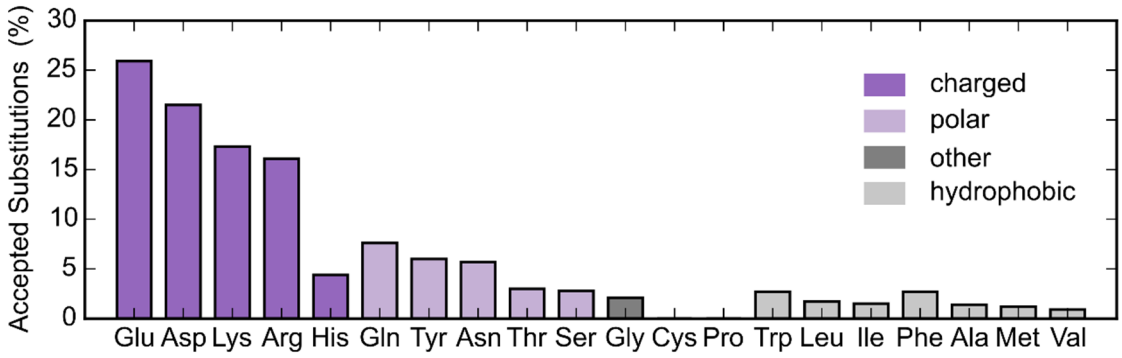

Figure 79: Distribution of accepted substitutions during PACMAN simulations.

4.3.11 Charge shielding motif to stabilize the ion-pair

The next step was to test, if one can stabilize the ion-pair by introducing a charge shielding motif as suggested by the PACMAN simulations and the statistical analysis of natural proteins. To this end, an amphiphilic asparagine was placed at position 69, followed by 200 ns MD simulation (*Figure 80*).

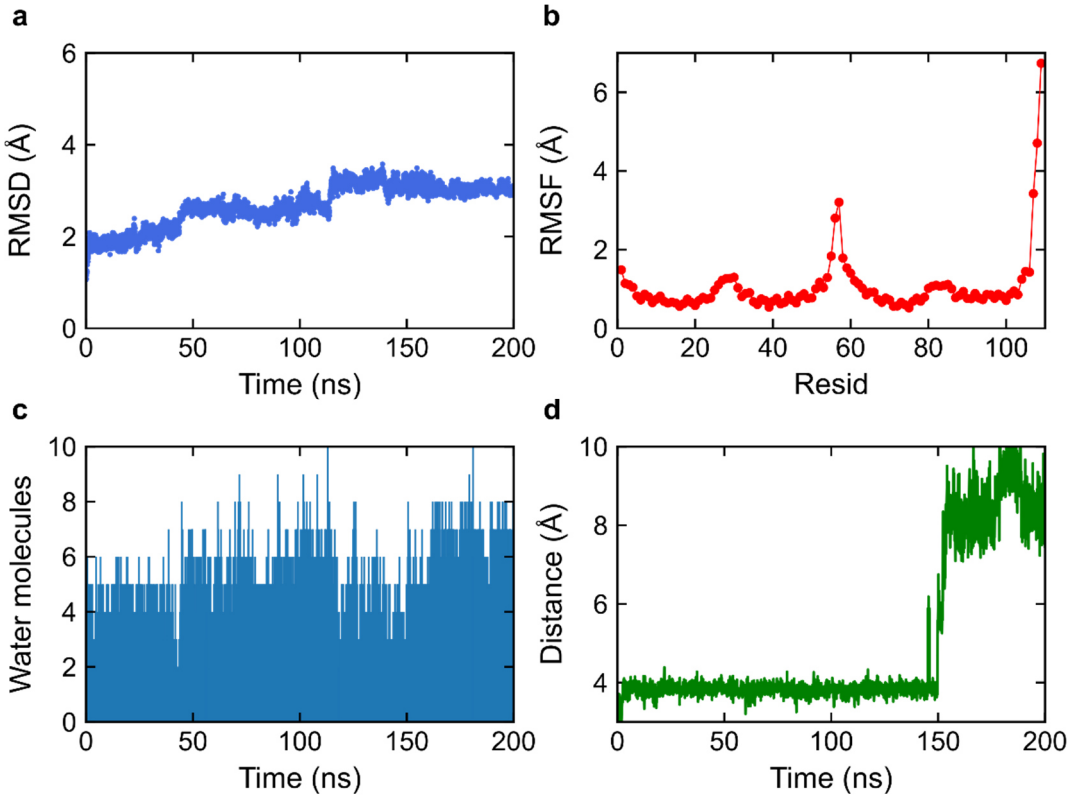

Figure 80: Analysis of *2*00 ns MD simulation of the 1ip (Glu17-Lys72)-N69 *Maquette* 2 NMR structure. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content; **d**, distance profile between CD-NZ of Glu17 and Lys72.

The RMSD is shown in panel a. During the first 100 ns the value continuously increases until it stabilizes during the last 50 ns (*Figure 80a*). The C-terminus forms the most flexible part of the protein with an RMSF of around 6 Å (*Figure 80b*). The solvation of the system resembles the closed ion-pair NMR structure and shows an average water content of around five water molecules. The water content does not significantly change during the simulation (*Figure 80c*). The CD-NZ distance between Glu17 and Lys72 is plotted in panel d. The ion-pair is stable during the first 150 ns, followed by opening of the ion-pair. This result fits our previous findings of the dynamically flexible ion-pair with multiple conformations.

The experimental analysis of the stabilized ion-pair model was done in the same way as presented above and expression and purification of the 1ip (Glu17-Lys72)-N69 *Maquette* 2 results again in high sample purity assayed by SDS-PAGE and an apparent MW via SEC of ~ 26 kDa, which is slightly higher than the 1ip *Maquette* 2. *Figure 81* shows the CD spectrum and results from chemical unfolding experiments. This construct exhibits a typical α-helical CD spectrum (*Figure 81a*) and is

highly temperature stable, with a melting transition that initiates at around 90°C, but does not reach the thermal transition by 100°C (*Figure 81b*). The chemical unfolding with GdnHCl resulted in a well-shaped sigmoidal curve with an inflection point at 3.0 M of GdnHCl and a corresponding unfolding energy of 2.9 ± 0.2 kcal mol^{-1}, which is slightly higher than 1ip (Glu17-Lys72) *Maquette* 2. Following this successful preliminary characterization, an ^{15}N labeled NMR sample is produced to measure an HISQC, which is shown in *Figure 82*.

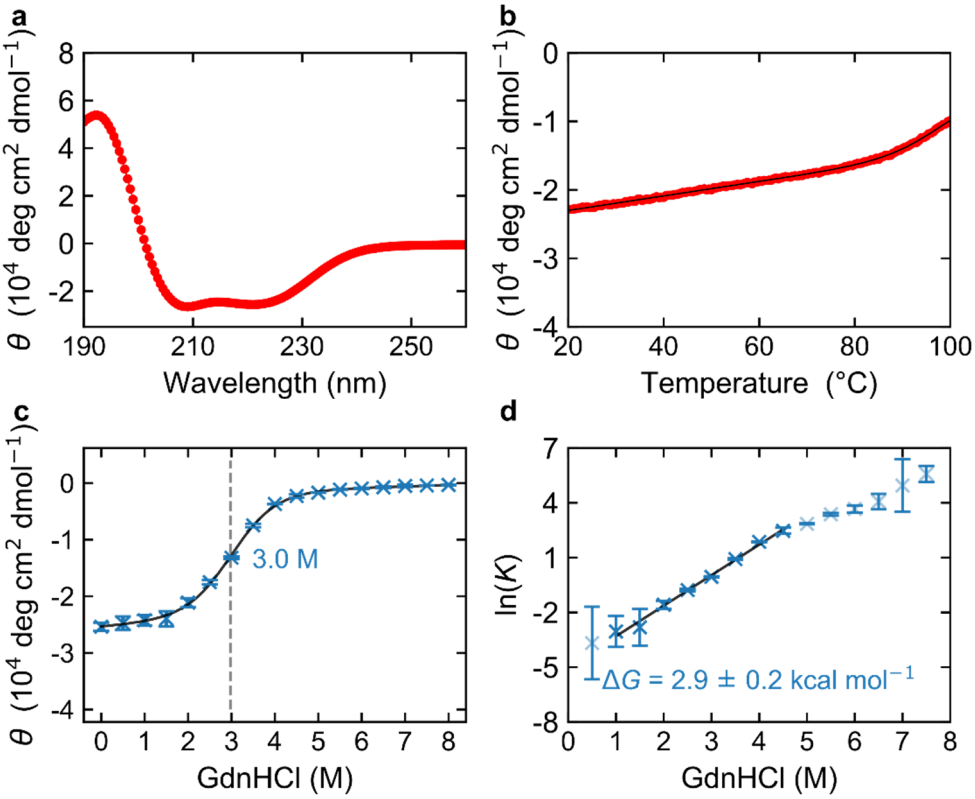

Figure 81: Experimental analysis of 1ip (Glu17-Lys72)-N69 *Maquette* 2. **a**, CD spectrum; **b**, melting curve; **c**, chemical unfolding with GdnHCl; **d**, ln*K* plotted against denaturant concentration.

Comparison of the spectra of 1ip (Glu17-Lys72) (red) and 1ip-N69 (grey) suggest that the ion-pair peak of interest is significantly more intense in the latter model. Due to the substitution of a valine the peak is slightly shifted, but still corresponds to the same lysine. Taken together, these findings suggest the ion-pair is indeed stabilized by the asparagine at position 69.

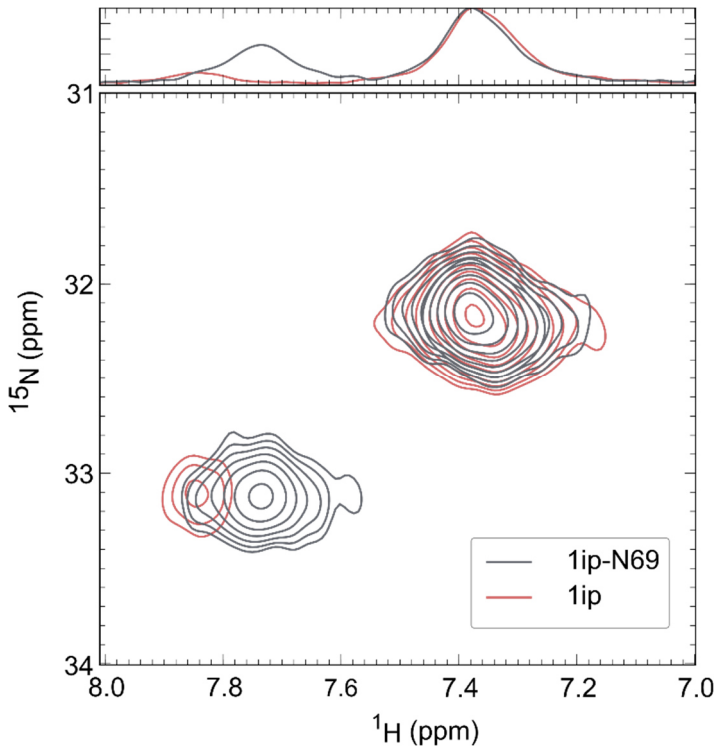

Figure 82: HISQC spectra of 1ip (Glu17-Lys72) and 1ip (Glu17-Lys72)-N69 *Maquette* 2. 1ip is shown in red and 1ip-N69 in grey.

4.3.12 Destabilization series of *Maquette* 2

The *Maquette* 2 models show a high thermal stability and a sigmoidal melting curve could not always be obtained between 20°C and 100°C. The protein starts to melt above 90°C but does not reach the fully unfolded state until 100°C. Therefore, the idea was to decrease the melting point of the protein in order to obtain the complete transition between the native and unfolded states in the range between 20 and 100°C, by replacing leucines and isoleucines at the top of the bundle with alanines (*Figure 83a*). While alanine is smaller, it is also a non-polar amino acid weakening the packing and therefore the overall stability of the protein by creating small "holes", without disrupting the overall protein fold (*Figure 83b*). The alanine substitutions were made stepwise beginning with Leu32, then subsequently adding a second and third alanine at positions 79 and 87. The destabilization series was introduced into the hydrophobic core, 1ip (Glu17-Lys72), and 1ip-N69 *Maquette* 2 models and the protein stability was analyzed by thermal and chemical unfolding experiments (*Figure 84*). The effect was most apparent for the hydrophobic core series (panel a), which clearly showed a decrease in the stability with increasing alanine substitutions, in the order of 0.5 – 0.7 kcal mol^{-1} for each additional alanine substitution. For the 1ip (Glu17-Lys72) model, the results are less clear: the first alanine substitution leads to a small decrease of 0.2 kcal mol^{-1} in the protein stability, but no measurable difference could be detected when adding the second alanine. When the third alanine substitution is introduced, another significant decrease of 0.9 kcal mol^{-1} takes place.

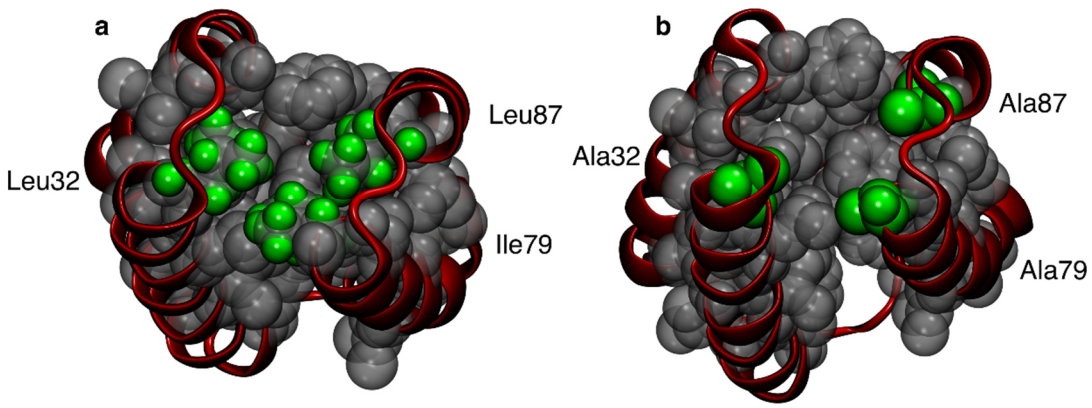

Figure 83: Hydrophobic packing of 1ip (Glu17-Lys72) *Maquette* 2. The hydrophobic packing is visualized in grey and Van der Waals representation. The substitution positions are highlighted in green, while the protein is shown in new cartoon representation and red. **a**, original 1ip model with Leu and Ile at positions 32, 79 and 87; **b**, 1ip model with Ala substitutions at positions 32, 79 and 87.

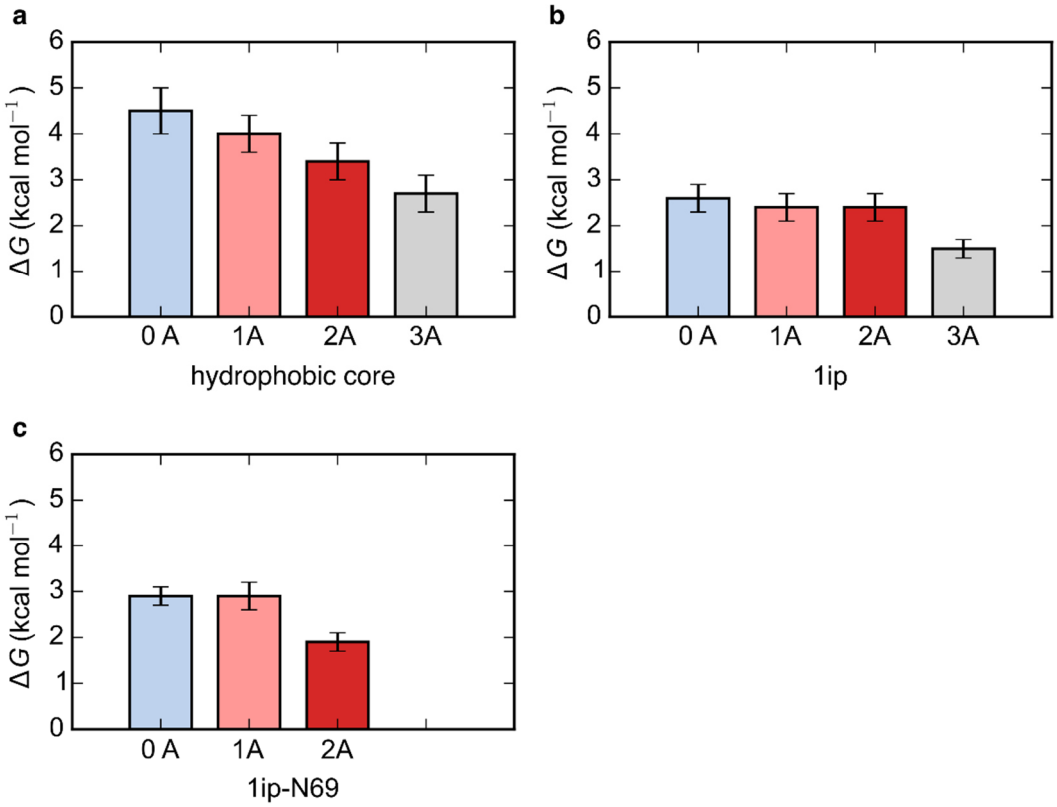

Figure 84: Comparison of the protein stability of hydrophobic core, 1ip (Glu17-Lys72), and 1ip-N69 *Maquette* 2 from without alanine (blue) substitutions up to three alanines (grey) resulting from chemical unfolding experiments with GdnHCl. One and two alanine substitutions are shown in coral and red, respectively. **a**, Δ*G* values for hydrophobic core *Maquette* 2; **b**, Δ*G* values for 1ip (Glu17-Lys72) *Maquette* 2; **c,** Δ*G* values for 1ip-N69 *Maquette* 2.

For the 1ip model with the charge shielding residue Asn69, a very similar picture is obtained, except that no stability difference is observed with one alanine substitution and a larger decrease of 1.0 kcal mol^{-1} is observed after the introduction of the second alanine substitution. Unfortunately, it was not possible to obtain reasonable ΔG for the three alanines model. Overall, our rational destabilization approach was successfully introduced. The hydrophobic core series was the most amenable to this approach.

As a next step, the influence of the substitutions on the thermal stability was analyzed. Melting curves for the hydrophobic core, 1ip (Glu17-Lys72), and 1ip-N69 from 20 to 100°C were measured by CD at 222 nm (*Figure 85*). As described above, due to the high overall thermal stability, ideal sigmoidal curves with clear transition points could not be accurately assigned. Nevertheless, a general trend towards reduced stability with increasing alanine content is visible. In all three series the 0A model was the most stable showing no sign of melting by 100°C. With each additional alanine substitution the melting curves become more sigmoidal. For the hydrophobic core, 1A, and 2A constructs the shape of the melting curves suggest that the transition takes place slightly above 100°C, above which data collection was not possible. In the case of 3A, the transition is shifted into the detectable temperature range. These findings are consistent with the trend of the chemical unfolding experiments. In the case of 1ip and 1ip-N69, the transition point is already shifted to the detectable range in the 2A construct.

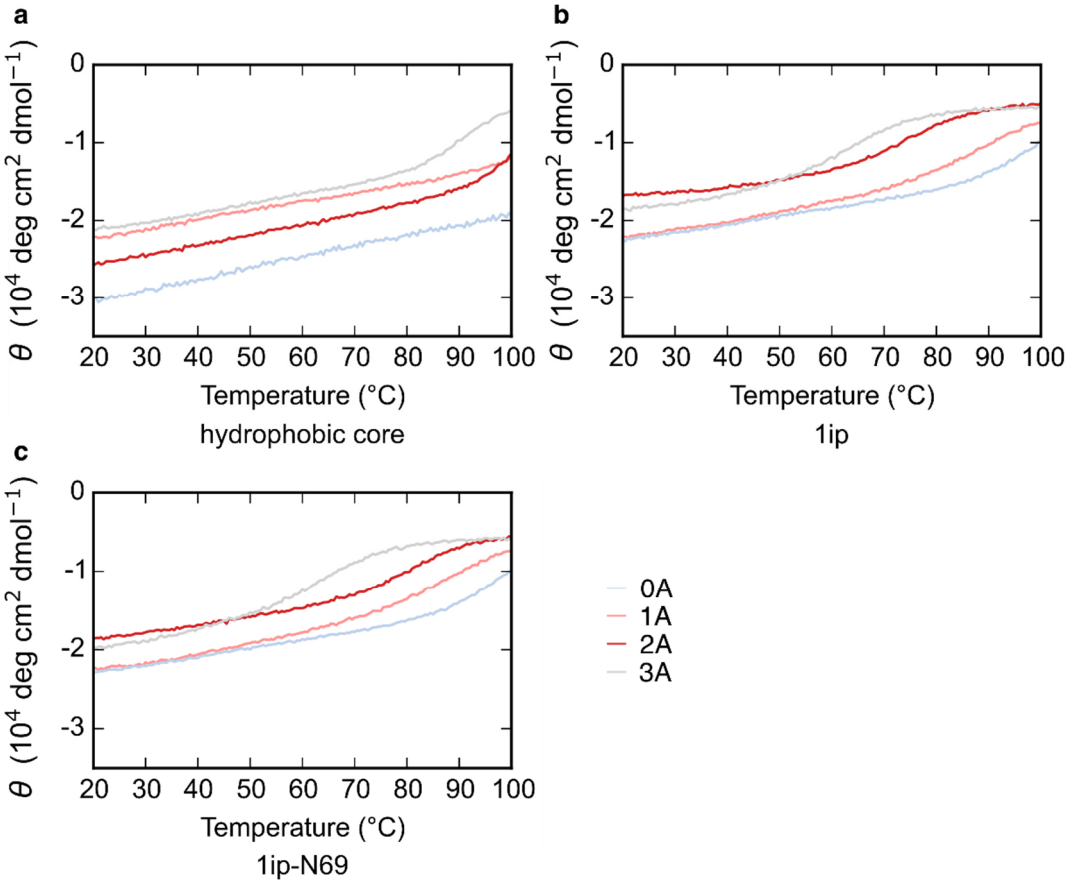

Figure 85: Comparison of the thermal stability of hydrophobic core, 1ip (Glu17-Lys72), and 1ip-N69 *Maquette* 2 from without alanine (blue) substitutions up to three alanine substitutions (grey). One and two alanines are shown in coral and red. **a**, melting curves for hydrophobic core *Maquette* 2; **b**, melting curves for 1ip (Glu17-Lys72) *Maquette* 2; **c**, melting curves for 1ip-N69 *Maquette* 2.

The models of the alanine series were also simulated for 100 ns and the water content around the ion-pair and the substituted positions was analyzed (*Figure 86*). In the case of the hydrophobic core, the water content does not clearly correlate with the amount of alanine mutations (*Figure 86a*). All four models show a similar internal solvation, which indicates a stable, water sealed hydrophobic core. The 1ip model with three alanines show the highest water content in the 1ip series (*Figure 86b*), but in general, the number of water molecules is only slightly higher as compared to the hydrophobic core models. The 1ip-N69 series does not either show a trend in solvation with the number of alanine substitutions (*Figure 86c*).

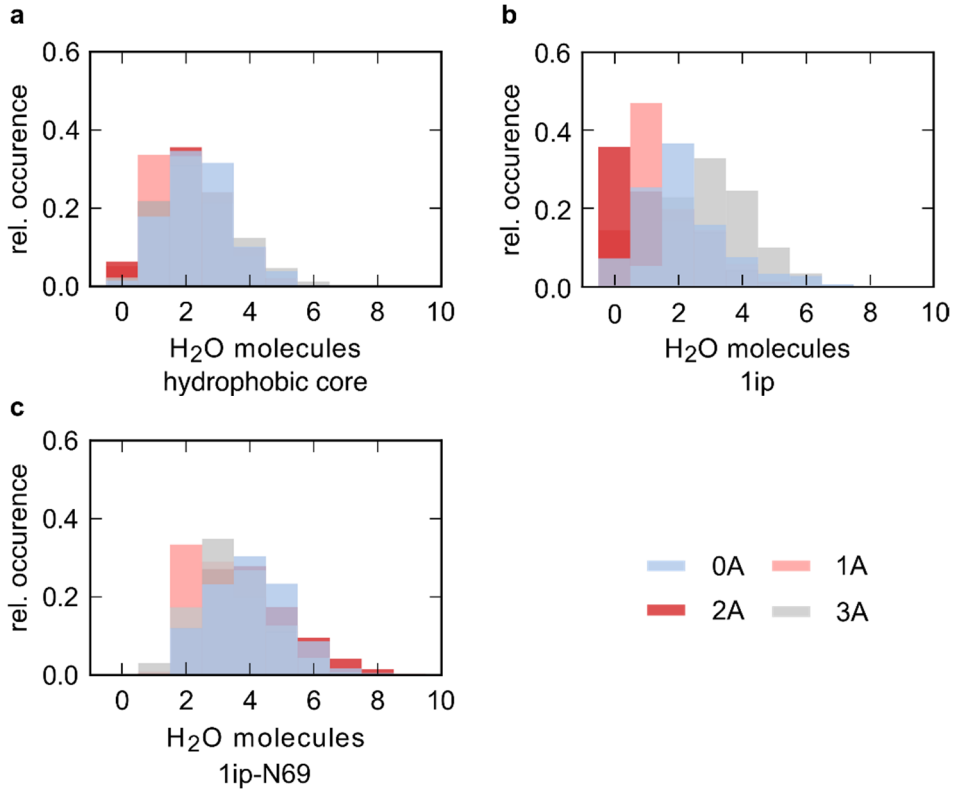

Figure 86: Water content of the hydrophobic core during 100 ns MD simulation. For each construct series zero (blue), one (coral), two (red), and three (grey) alanines were substituted and simulated. **a**, hydrophobic core; **b**, 1ip (Glu17-Lys72); **c**, 1ip-N69.

4.3.13 Force profile analysis (FPA)

Force profile analysis (FPA) provides a powerful tool to study co-translational events, *i.e.* folding in the ribosomal exit tunnel. Previous analysis has examined the stability of various protein models by unfolding them. With FPA, we determined if the artificial protein *Maquettes* fold differently and reflect the introduced elements. FPA was performed by Dr. Grant Kemp (Kaila Lab). The experiments were conducted using the PURExpress™ *in vitro* translation system. The translated proteins are labeled with ^{35}S-methionine for ease of detection and analysis by the SDS-PAGE.[101] The FPA technique involves engineering a force-sensitive translational arrest peptide at different linker lengths to the protein of interest.[102] The concept behind these measurements is that the translation stops when the last codon of the arrest peptide is reached. During the expression of the fusion protein, translation stops when the last codon of the arrest peptide is reached due to unfavorable conformation of the arrest peptide at the peptidyl transferase center. However, this arrest can be overcome if a pulling force acts on the stuck nascent protein. One potential pulling force is the tension introduced in the nascent protein chain by co-translationally folding inside or very close to the ribosomal exit tunnel.[101] By systemically varying the distance between the *Maquette* 2 model and the arrest peptide a picture of how much force is acting on the nascent chain at any point during translation can be

obtained – a so-called force profile. It has been noted previously that for a two-state folding domain, the stability can be correlated with the characteristics of the primary peak in the force profile.[135]

In order to determine whether FPA could detect differences in the stability of the various *Maquette* 2 models have been studied, measurements were performed for the hydrophobic core, 1ip (Glu17-Lys72), and the stabilized 1ip-N69 construct. Experimentation with different arrest peptides of increasing strength was necessary, which indicates that the bundles are very strong folders. The length of the linker (including the arrest peptide sequence) fused to the model was varied from 21 up to 35 amino acids (*L*) and the amount of force (via fraction full-length (f_{FL})) for each *L* was plotted to give the force profiles for all three models (*Figure 87*). f_{FL} is defined as $\frac{I_{FL}}{I_{FL}+I_A}$, where *I* is the intensity of the radiolabeled band on the SDS-PAGE gel, *FL* corresponds to the full length protein, and *A* to the arrested protein. The hydrophobic core construct shows a distinct peak at a linker length of 27 amino acids and the f_{FL} is 0.5 (*Figure 87*, blue). In comparison the force profile of the 1ip construct shows a peak shifted towards higher lengths (*Figure 87*, red). The full-length protein peak occurs at 29 amino acids and the f_{FL} is reduced to around 0.35 corresponding to a lower protein stability. Regarding the 1ip-N69 model, the peak also occurs at a length of 29 amino acids, but the intensity is higher than for 1ip around 0.43 placing it in between the hc and the 1ip profile (*Figure 87*, grey). This illustrates how Asn69 is able to increase the folding stability of the 1ip model, though not the same level as a hydrophobic core, which is consistent with the unfolding results.

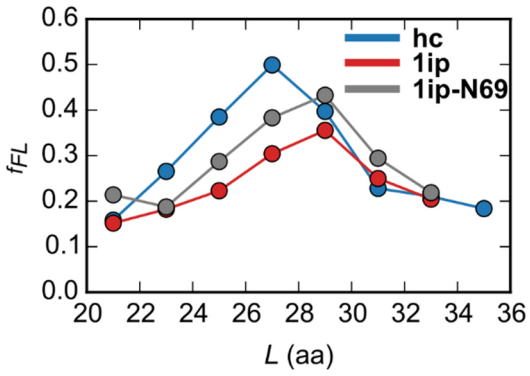

Figure 87: Force profiles for hydrophobic core, 1ip, and 1ip-N69 *Maquette* 2 models. The fraction of full-length protein (f_{FL}) is plotted against the length of the linker in amino acids (*L*). The profile of the hydrophobic core construct is shown in blue, 1ip in red, and 1ip-N69 in grey.

FPA analysis was also performed for the destabilizing alanine series for the hc, 1ip, and 1ip-N69 models (*Figure 88)*. In general, the observed peak intensity correlates with the number of introduced alanine residues, with increasing alanine substitutions corresponding to decreasing f_{FL} values. *Figure 88a* presents the results for the hc *Maquette* 2 series. The peak height is reduced to around 0.35 f_{FL} due to one and two introduced alanine residues, but the peaks are still located at the same length and lies under the hc peak. The 3A construct shows the highest reduction in peak height at 0.29. For the 1ip 1A and 2A profiles the folding peaks shift of the peaks to longer linker lengths. The peak height

of 1A and 2A is very similar to the 1ip profile with 0.36 and 0.3. The 3A model essentially no longer exhibits a peak (*Figure 88b*). For the 1ip-N69 series (*Figure 88c*), the introduction of 1A does not lead to a significant difference in the peak height. This time even the peak position is still the same. The 2A model has a shifted peak towards higher linker length with 31 amino acids and again for the 3A construct there is no significant peak detectable. These results are also consistent with the observations of bundle unfolding.

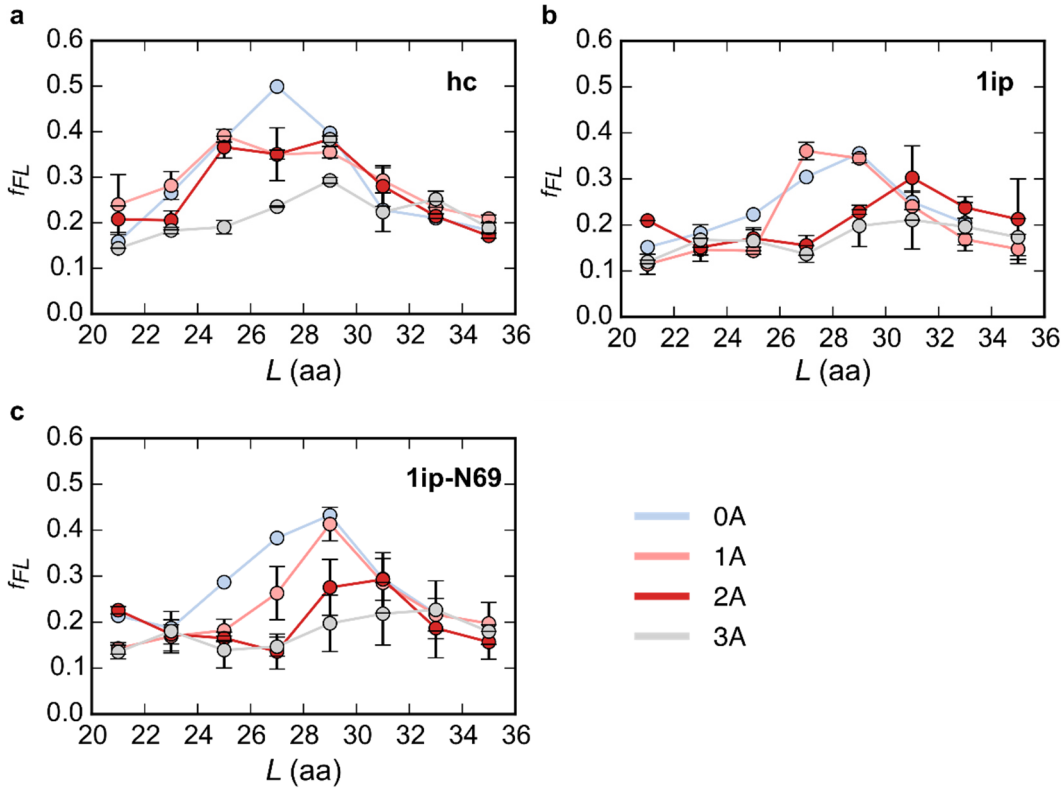

Figure 88: Force profiles of the alanine series for hc, 1ip, and 1ip-N69 *Maquette* 2. For all models with zero up to three alanine substitutions the force profiles are plotted. **a**, hc *Maquette* 2; **b**, 1ip *Maquette* 2; **c**, 1ip-N69 *Maquette 2*.

4.4 *Maquette* 3

As a third model framework an almost 70 Å long bundle template from Dutton's group was used as a starting point for studying buried ion-pairs.[76] The original structure was solved experimentally using X-ray crystallography, yet bound to some cofactors. In the following, several mutations were introduced to seal the hydrophobic core. After obtaining a water sealed hydrophobic core, a double ion-pair was placed and four additional glutamine residues surrounding the ion-pairs. The glutamines should serve as charge shielding motif. The glutamates are located at positions 65 and 136 and the lysines at 35 and 162, while the surrounding glutamines were placed at positions 61, 68, 132, and 166. The resulting construct is called 2ip *Maquette* 3 in the following sections. As for the previously discussed constructs, it was also simulated for 200 ns and the standard analysis was performed afterwards. The protein was also investigated experimentally including analysis of the stability. To solve the structure of *Maquette* 3, crystallography was applied.

Figure 89 presents the results of the MD simulation. The RMSD is highly stable around 2 Å after a few nanoseconds and the system seems to be stable and converged (*Figure 89a*). Panel b shows the RMSF value where the N-terminus is the most flexible part. Interestingly, the dynamics of the loops slightly increases from loop one to three.

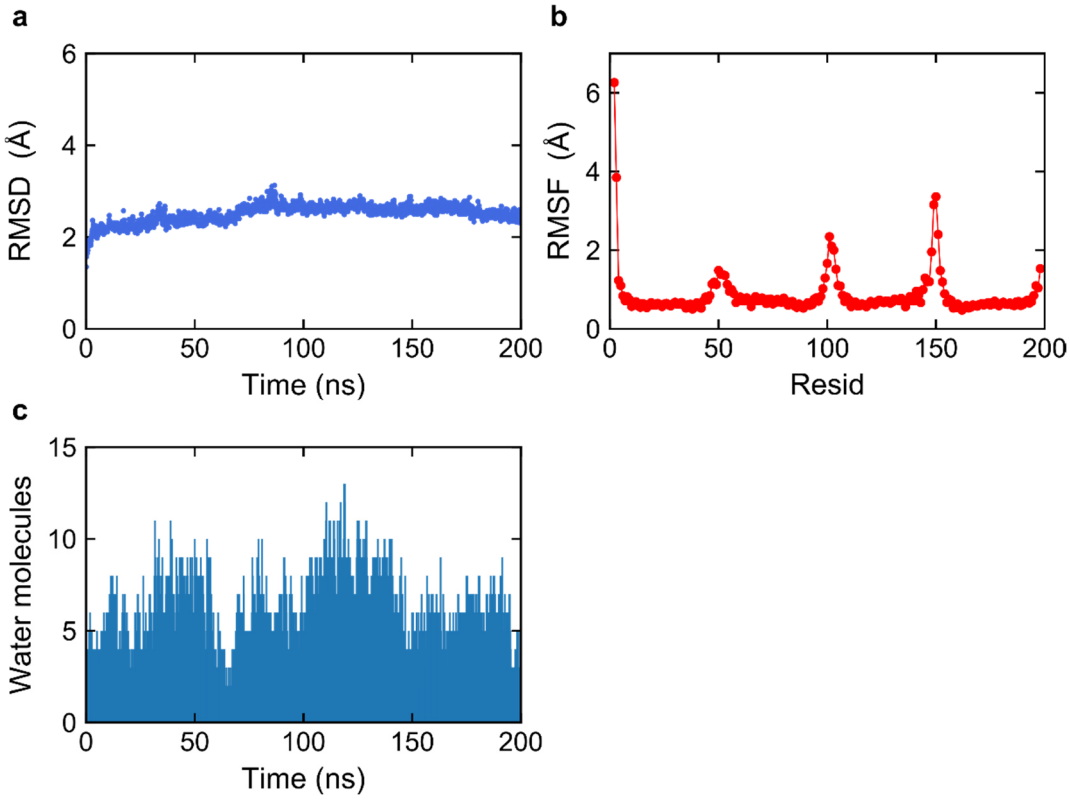

Figure 89: Analysis of 200 ns MD simulation of 2ip *Maquette* 3. **a**, RMSD plot; **b**, RMSF plot; **c**, water molecule content.

The solvation of the bundle is relatively low and the hydrophobic core stays water sealed. The maximum number of water molecules inside the helices is 10, but most of them are located at the top and the bottom of the bundle (*Figure 89c*).

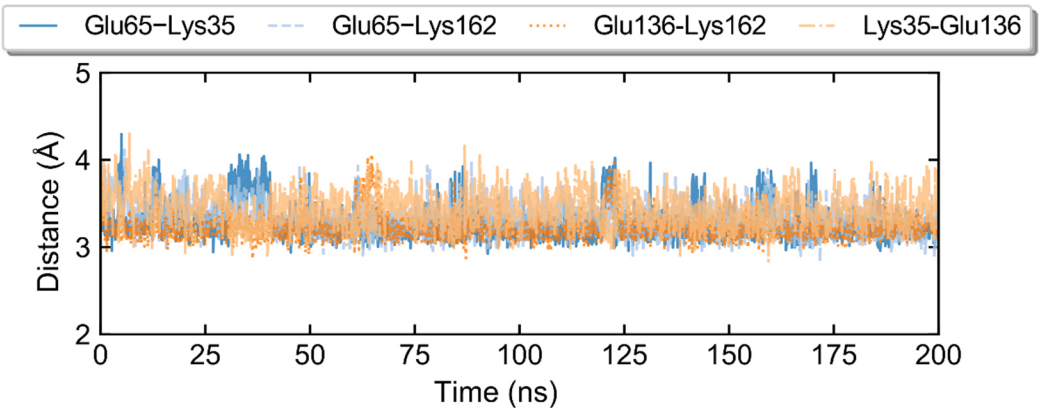

Figure 90: Distance plot of 2ip *Maquette* 3 during 200 ns MD simulation. CD-NZ distance of Glu65-Lys35 (blue), CD-NZ distance of Glu65-Lys162 (light blue), CD-NZ distance of Glu136-Lys162 (orange), and CD-NZ distance of Lys35-Glu136 (light orange).

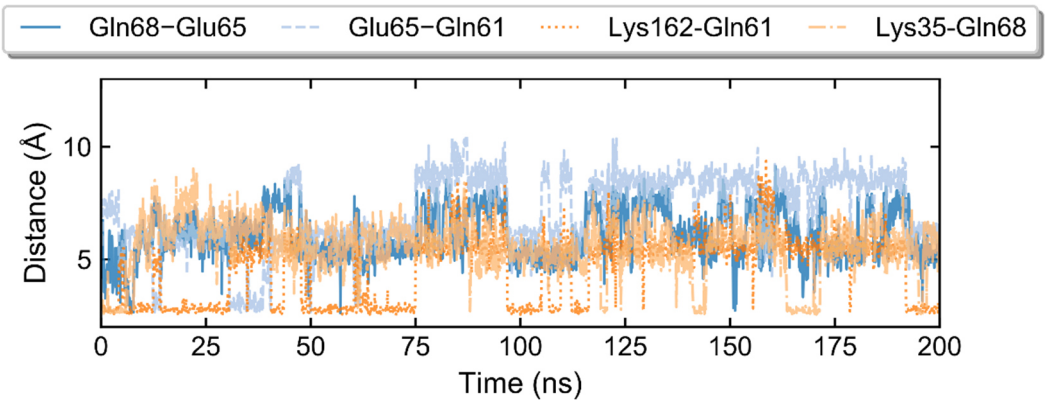

Figure 91: Distance plot of 2ip *Maquette* 3 during 200 ns MD simulation. CD-OE distance of Gln68-Glu65 (blue), OE-CD distance of Glu65-Gln61 (light blue), NZ-CD distance of Lys162-Gln61 (orange), and NZ-CD distance of Lys35-Gln68 (light orange).

Figure 90 presents the CD-NZ distances between the ion-pair residues. The four charged residues form a highly stable double ion-pair, where each glutamate interacts with each lysine and vice versa. The distance stays between 3 and 4 Å, which is an ideal distance for a salt bridge. The charge shielding glutamines are analyzed in *Figure 91*. Only Gln61 and Gln68 form direct hydrogen bonds with the double ion-pair and helps to stabilize the charge. The other two glutamines (Gln132 and Gln169) point towards the water during the simulation, while Gln61 and Gln68 stabilizes Lys35 and Lys162 in its position.

This construct was also experimentally analyzed. Therefore, the protein was overexpressed in *E. coli* and purified according to the described protocol in the method section. The gel in *Figure 92a* proves the high purity and the right size of the protein. A SEC was performed during the purification process to determine the oligomerization state and calculate the MW (*Figure 92b*).

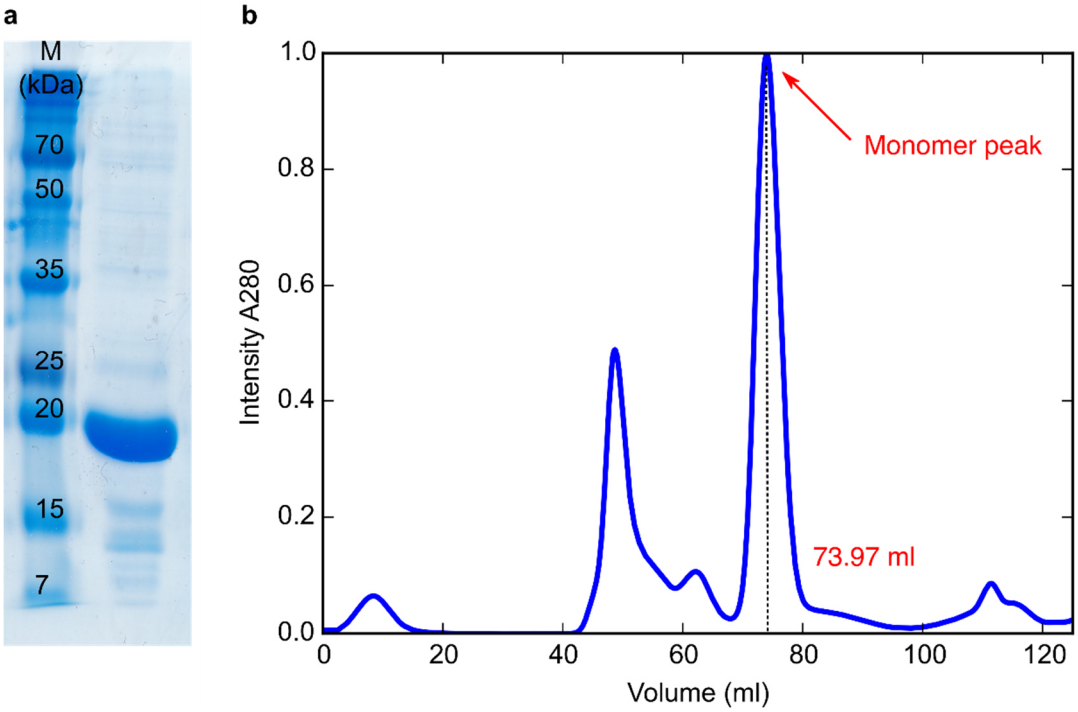

Figure 92: a, SDS-PAGE gel of 2ip *Maquette* 3; **b**, size exclusion chromatography (SEC) profile of 2ip *Maquette* 3. The major peak corresponds to monomeric protein and has a retention volume of 73.97 ml which corresponds to a molecular weight of 19 kDa.

The obtained molecular weight is somewhat too low with 19 kDa instead of the expected 23 kDa, but the main peak corresponds to the monomeric protein.

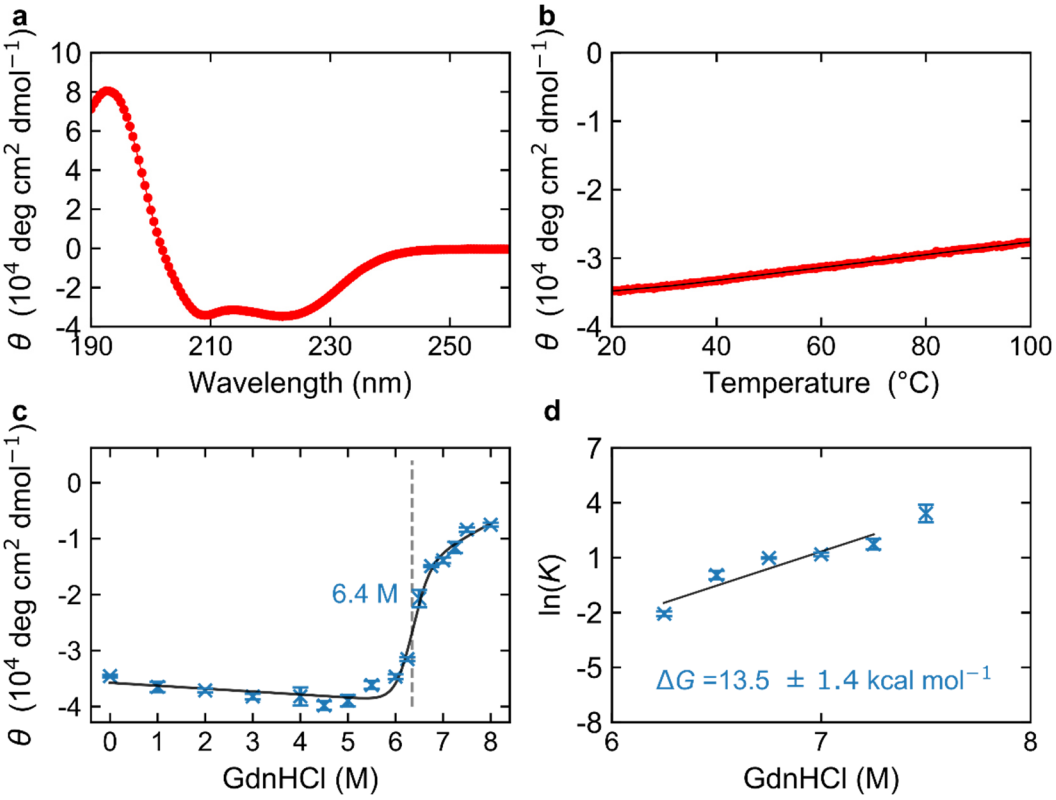

Figure 93: Experimental analysis of 2ip *Maquette* 3. **a**, CD spectrum; **b**, melting curve; **c**, chemical unfolding with GdnHCl; **d**, ln*K* plotted against denaturant concentration.

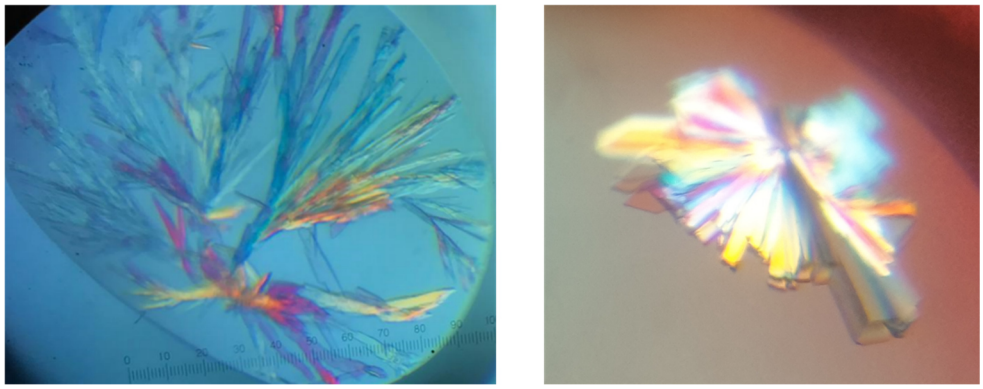

Figure 94: Crystals of 2ip *Maquette* 3.

The secondary structure was analyzed by CD spectroscopy (*Figure 93a*) and it shows a typical spectrum for an α-helical structure. The thermal stability was investigated by heating up the protein sample to 100°C, but there is no sign of losing its structure due to higher temperature (*Figure 93b*). The unfolding energy was determined by chemical unfolding with GdnHCl in triplicates. The *Maquette* 3 model stays stable until 6 M of GdnHCl and has its transition point at 6.4 M (*Figure 93c*). The corresponding Δ*G* is 13.5 ± 1.4 kcal mol^{-1} (*Figure 93d*).

The structure calculation was performed by Prof. Ville Kaila and Prof. Michael Groll. A highly concentrated protein sample (40 mg ml^{-1}) was prepared for the crystallization. The protein

crystallized extremely well (*Figure 94*) and several crystals were measured to obtain a crystal structure with high resolution. The best structure was solved at 1.85 Å resolution (PDB ID: 6Z0C) where it is possible to detect water molecules. *Figure 95* presents the solved crystal structure of 2ip *Maquette* 3 (green) compared to the *in silico* MD model (grey). The experimental model closely resembles the computational model. Inside the hydrophobic core, tight ion-pairs are formed as predicted by the MD simulations. Also, the stabilizing motif is detected in contact with the ion-pairs and the ion-pairs are tightly sealed from water molecules in the crystal structure. The hydrophobic core stays intact and the amphiphilic residues shield the charged ion-pairs from the hydrophobic surrounding.

The design of the 2ip *Maquette* 3 successfully results in an exceptionally stable protein with a double ion-pair. The protein shows a high stability despite the charged residues inside the hydrophobic core. It is resistant towards unfolding until 100°C and up to 6 M GdnHCl. Further, the predicted *in silico* model is almost identical to the solved crystal structure.

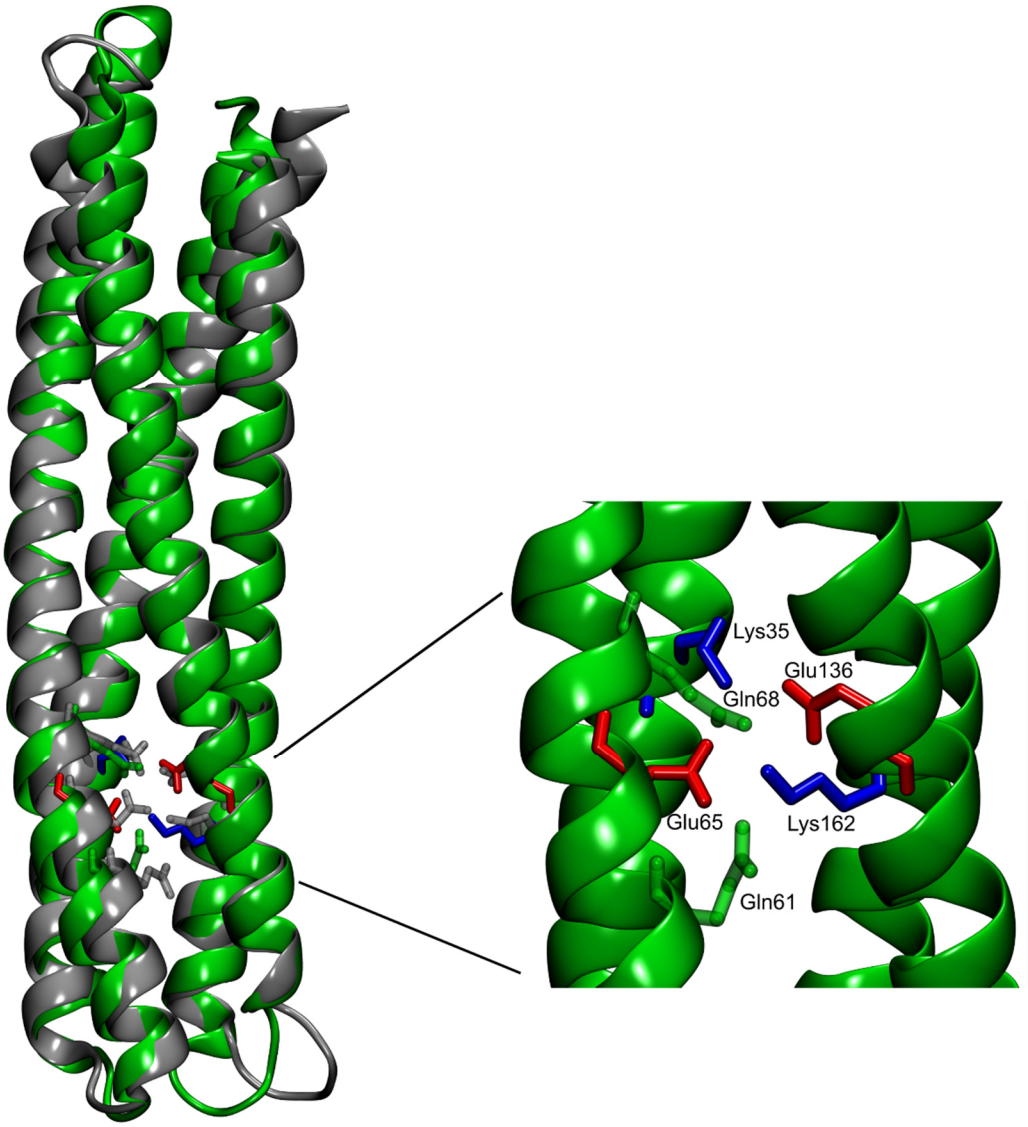

Figure 95: Resolved crystal structure (green) and structural snapshot from the *in silico* model of the 2ip *Maquette* 3 protein (grey). The ion-pairs are shown in licorice. Glu65 and Glu136 are highlighted in red and Lys35 and Lys162 in blue. The charge shielding Gln68 and Gln61 are green. A zoom into the resolved crystal structure is presented on the right side. PDB ID: 6Z0C.

To quantify the stabilizing effect of the charge shielding glutamines surrounding the two ion-pairs, Gln61 and Gln68 were mutated to alanine residues. Gln61 and Gln68 were chosen for the substitution because both show the closest contact and most interaction of the four introduced glutamines with the ion-pairs during the MD simulations and in the crystal structure. The construct was expressed and purified in the same way as the 2ip *Maquette* 3. The data of the purification process was similar to the 2ip *Maquette* 3 and therefore is not reported here. At the end a very pure monomeric protein with a MW of 19 kDa according to the SEC run is obtained and the MW is in agreement with the 2ip *Maquette* 3. The protein concentration for the following CD spectroscopic analysis was 5 µM. First, a CD spectrum was measured and the typical curve for an α-helical secondary structure was obtained (*Figure 96a*) which is very similar to the 2ip *Maquette* 3 spectrum. The melting curve monitored at

222 nm does not show a transition and therefore, the protein is stable up to 100°C (*Figure 96b*). For the quantification of the stabilizing effect of the glutamines, chemical unfolding with up to 8 M GdnHCl was measured (*Figure 96c*). The unfolding series shows a transition point at 7.3 M GdnHCl and the corresponding Δ*G* is 8.3 kcal mol^{-1}. Consequently, the stabilizing effect of Q61 and Q68 together is around 5.2 kcal mol^{-1}.

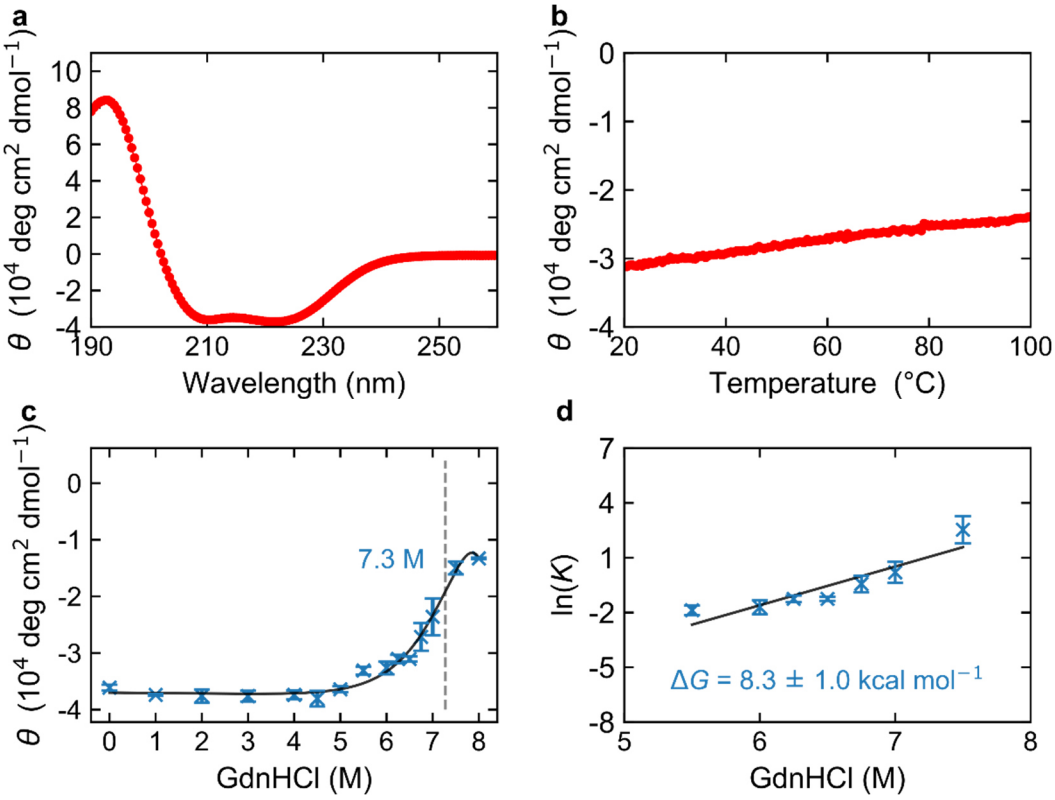

Figure 96: Experimental analysis of 2ip-Q61A-Q68A *Maquette* 3. **a**, CD spectrum; **b**, melting curve; **c**, chemical unfolding with GdnHCl; **d**, ln*K* plotted against denaturant concentration.

5. Discussion

5.1 Conformational studies on Hsp90

The molecular chaperone Hsp90 shows rich conformational dynamics between open and closed states during the catalytic cycle, but the underlying conformational changes are not fully understood. Hsp90 recognize misfolded client proteins in its extended *apo* state.[136] ATP-binding induces a large conformational change that finally leads to the closed dimerized state.[137] The influence and motion of a charged linker region between NTD and M-D was studied here by using a combined computational and experimental approach. For the molecular simulation studies, the charged linker was modeled *in silico* because the linker region is missing in the crystal structure.[138] This further implies that the linker is dynamically flexible and that it can contribute to the chaperone dynamics.[35,139] The sequence of the linker region consists of a repetitive motif of aspartate, glutamate, and lysine residues,[140] but the function of the linker region is not clear. Our findings from the combined simulations and SAXS experiments suggested that highly extended states of Hsp90 are involved in the chaperone cycle and could be responsible for the interaction with client proteins and in the recognition of misfolded proteins. These findings are consistent with previous results indicating that the linker deletion mutants affect cell viability and are important for the proper function.[139] Based on *in vitro* studies, it was found that the ATPase activity is lowered by linker deletions (residues 211-273), while the binding affinity for ATP is similar to that of the wild type Hsp90 with an intact linker.[139] It is also assumed that cutting the linker region reduces the conformational flexibility that can affect the closure of the NTDs and therefore decrease the ATPase activity.[139] The specific linker sequence is important, as an artificially designed linker does not result in full functionality of the chaperone.[138] The charged linker makes contact to the NTD, leading to both docked and undocked states. In the docked state, the linker region can form secondary structure elements, while in the undocked conformation, the NTD rearranges.[35] The linker is important for the dynamics between NTD and M-D and also affects the NTD dimerization.[35] In addition to our findings that the charged linker is involved in the conformational cycle of Hsp90, previous studies show that the linker region can also serve as rheostat that control the chaperone activity,[138] enable conformational switching by increasing the flexibility of the system, as well as modulate the chaperone activity.[35,138,139]

5.2 *Maquette* 1

The results for the *Maquette* 1 models with long and short loops are presented in section 4.2 of this thesis. The reported findings for the hydrophobic core starting structure (hc) are in good agreement with previous results of similar protein model.[12] The *Maquette* 1 has a high stability and shows a high expression yield with a non-functionalized interior and a highly negative charged surface. Our NMR spectrum indicates a possible molten-globule like ensemble and a relatively undispersed HSQC spectrum.[12] The *Maquette* 1 were used to design a series of mutants which were simulated and experimentally characterized. Generally, both the computational and experimental data indicate that the protein stability decreases with increasing charged residues placed within the hydrophobic core. For these constructs, we initially placed a single charged residue in the protein interior, and built models with a chain of up to three ion-pairs, based on *in silico* design. Although the overall simulation timescale is too short to observe complete unfolding of the protein structure, the unfavorable effects can also be clearly observed in the MD simulations. The CD measurements further confirm the loss of the α-helical structure and decrease in the melting temperature for the unstable protein models. For the constructs with a single buried charge, the computational and the experimental results are in good agreement and show well-folded proteins. The six charged residues in the 3ip model experimentally destabilize the bundle structure and lead to oligomerization, but the simulation timescale is most likely not long enough to observe complete unfolding or rearrangement of the protein structure. Another interesting finding is that the 1ip model also oligomerizes, suggesting that some structural changes may take place in the protein due to the introduced charged residues. It seems likely that the bundle protein compensates the penalty of introducing charged residues inside the hydrophobic core and changes its conformation, favoring the oligomerization. The goal of the conformational change is to lower the solvation penalty of introducing buried charged residues.[100] The surroundings of the *Maquette* 1 models are most likely not ideal for hosting charged residues. A network of hydrogen bonding partners could help stabilize the ion-pair within the protein, but it is possible that further modifications in the protein interior are needed to stabilize the buried charged residues in the hydrophobic core.

Interestingly, the HSQC spectrum of the 3ip construct seems to be of slightly better quality and the peaks are more dispersed than for the hydrophobic core model. Nevertheless, the signal dispersion is too small and we cannot make conclusions about the conformational state of the protein. The introduced mutations change the chemical surroundings and introduce more diversity in the otherwise highly symmetric and repetitive bundle sequence. Even introduction of a single ion-pair leads to improved signal dispersion in the HSQC, but the sequence is still too homogeneous and results in overlapping signals. Therefore, NMR is not the experimental method of choice to analyze buried ion-pairs for this system.

In a second step, the nine amino acids long loops were reduced to four amino acids. The long loops are highly flexible and presumably enable expansion of the otherwise compact hydrophobic core upon loop elongation. *Figure 97* shows a comparison of the thermal stability of the hydrophobic core, with a single introduced charged residue, and one ion-pair constructs with long and short loops. In general, the melting temperature behaves as expected and decreases with more introduced charged residues in both series. Interestingly, the melting temperature for the short loop series are higher than for the long loops, suggesting that the constructs with shorter loops are more stable. The 2ip and 3ip models with short loops, however, also have lower secondary structure content and show a high oligomerization state. Nevertheless, we successfully reached the goal of a more stable *Maquette* 1 by shortening the loops in case of the hydrophobic core, a single buried charge, and one ion-pair constructs regarding the thermal stability. The reversal of the ion-pair leads only to a small shift in the melting temperature of 0.1°C, suggesting that helix dipole-sidechain charge interaction do not significantly change between the two models.[141–143]

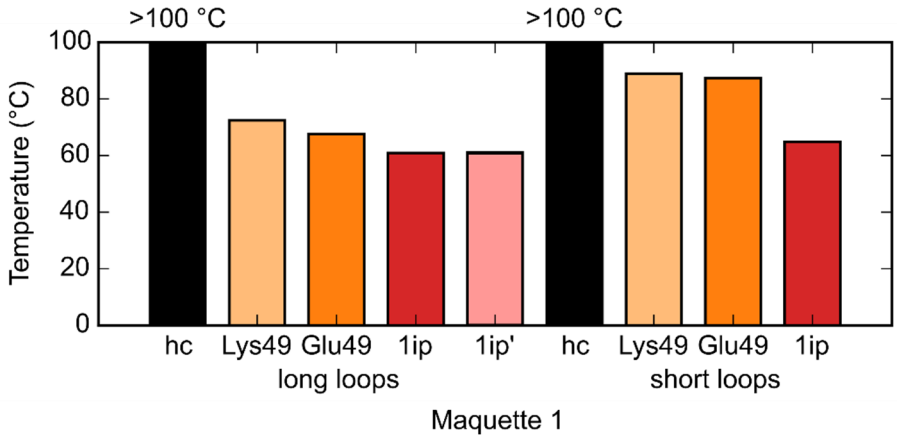

Figure 97: Melting temperatures of hydrophobic core, Lys49, Glu49, and 1ip models with long and short loops.

In computational studies of the 3ip model with short loops, we observed conformational switching of the ion-pairs during the 1 µs long simulation. The protonation of Glu112 serves as "input signal" for the ion-pair flipping, while without protonation of this glutamate, the ion-pairs stay in their initial conformation (section 4.2.20, *Figure 48*). Experimentally, we mimicked the protonation of Glu112 by a glutamine mutation at position 112, which resulted in a similar behavior in MD simulations. Unfortunately, we could not to detect the conformational switch experimentally due to rather weak protein stability and/or oligomerized state (see above).

Accurate design of *Maquette* 1 models with buried charged residues was hampered by the lack of an experimentally solved structure. Instead, we modeled the structure *in silico* by assuming a possible fold, but were not able to solve an NMR or X-ray structure during the project, and is not an ideal for designing a protein conformational switch. The *in silico* model is stable and the MD simulations show a well-packed protein, but uncertain whether the protein sequence adopts the modeled fold.

The chemical unfolding experiments show a two-state unfolding for the hydrophobic core and the 1ip constructs, whereas the models with a higher number of charged residues showed loss of α-helical structure, as described above. During the simulations, the water content of the bundle increases with more charged residues inside the hydrophobic core. The charged residues attract water molecules to buried protein interior, which may explain the loss of stability, and lead to possible unfold on a longer simulation timescale.

The *Maquette* 1 bundle protein might be too small to accommodate up to six charged residues inside the hydrophobic core. Normally, proteins have a polar surface to make them soluble and a hydrophobic closed-packed interior that provides thermodynamic driving force during the folding process.[94] The protein has to compensate for the rather high desolvation penalty when charged residues are placed in the hydrophobic core.[91,99] It is possible that this penalty is too high, which leads to drastically decreased protein stability with inserted ion-pairs. Moreover, the interaction between the charged residues and the interaction with possible other groups inside the protein might not compensate the desolvation energy of the charged residues. A larger protein system could help to overcome this high desolvation penalty, or by stabilizing the charges, *e.g.* by introducing polar amino acids which could interact with the charges. The introduced glutamate and lysine residues presumably lead to a hole in the hydrophobic packing that may lead to water entry and a less stable protein.

We learned from the *Maquette* 1 project that the system size and the surroundings of the ion-pairs are important. The work also suggested that structure-guided design would help to build more stable protein models. Therefore, to improve the design approach, we next employed the *Maquette* 2, with a resolved NMR structure of the protein scaffold.

5.3 *Maquette* 2

The next *Maquette* scaffold used for designing buried ion-pairs was based on a design template and NMR structure obtained from the DeGrado group.[73] First, a model with a hydrophobic core was expressed and analyzed experimentally. The protein behaves well and as expected. The apo form of the protein did not show a melting transition up to 100°C and our HSQC spectrum reproduces previous NMR spectra.[73] Therefore, we used this system as a starting point to design buried ion-pairs. Two different 1ip constructs were designed: Lys50-Glu101 was placed in the hydrophilic part of the protein and Glu17-Lys72 in the hydrophobic core. The ion-pair in the water accessible part did not affect the overall protein stability, as analyzed by thermal stability and chemical unfolding experiments, as water molecules surround and stabilize the introduced charged residues and the desolvation penalty is minimal as the charges are in the solvent-accessible area of the protein. The *Maquette* 2 model with introduced Glu17-Lys72 behaves very differently, since the buried residues are within the hydrophobic core of the protein. The protein is ~ 2 kcal mol^{-1} ($\Delta\Delta G$) less stable as compared to the hydrophobic core construct, but the overall protein structure remains stable.

In the next step, we solved the structure of the 1ip (Glu17-Lys72) model to study the structural influence of the introduced ion-pair. In our NMR structure, we could not experimentally refine the structure of the closed ion-pair state (but see below) that seems to be dynamically flexible with a small population in closed form under the studied conditions. This small population was detected by HISQC experiments and also the MD simulations suggest that there are different populations with open and closed states, suggesting that the computational and the experimental results are consistent. The detected dynamic flexibility is very important for the functionality of ion-pairs in natural proteins such as in respiratory complex I or Hsp90.[31,33] In our case, the introduced ion-pair leads to a subtle rotation of helices 1 and 3 that may partially compensate for the thermodynamic penalty introduced from the buried ion-pair. The closed ion-pair conformation in the NMR structure can be achieved by a small distance restraint between Glu17 and Lys72, which leads to a < 1 Å inwards motion of the reorganized helices. Similar findings are detected in a natural protein such as SNAse with modeled ion-pairs at different positions in the protein, whereas also this protein reorganizes to compensate the buried charged residues.[100]

The suggestions of the PACMAN design algorithm and the analysis of natural proteins were tested with the buried 1ip (Glu17-Lys72) model. We successfully showed that Asn69 slightly improves the protein stability and increases the population of the closed ion-pair, which was detected by HISQC experiments. The amphiphilic residue form hydrogen bonds with the charged residues and at the same time, establish non-polar contacts with the hydrophobic protein core. Our findings provide important clues in understanding how proteins may stabilize buried ion-pairs. It also advances methods of protein design and establish an approach to study complex biological questions with

simple model systems. Our findings are consistent with results of buried ion-pairs in 2600 different protein structures, suggesting that buried polar residues form hydrogen bonds and salt bridges with other polar or ionizable groups.[144] Our electrostatic calculations further suggest that the buried polar residues are usually located in an electrostatically optimized environment by counter ions, hydrogen bonds, or favorable backbone dipole interactions.[145]

The stability of the *Maquette* 2 model was studied by introducing destabilizing residues. To this end, we mutated a series of leucine and isoleucine residues into alanine residues. These substitutions lead to a weaker hydrophobic packing and allow the entry of water molecules, as indicated by a decrease in the melting temperature. Similar effects have also been reported for other 5-helical bundles, where substitution of a leucine with an alanine in the hydrophobic core of the α-helical HemK NTD leads to a 20°C shift of the thermal transition point, from 50°C for the wild type to 30°C in the mutated protein.[146] For the mutant series studied in this thesis, the effect is less pronounced as compared to HemK NTD, but also detectable. A possible explanation for the smaller effect could be that the bundle used for the destabilizing series is an artificial protein. The *Maquette* 2 construct was designed as a well-packed protein and is much more compact than many natural proteins as a result of iterative rounds of design to make it hyperstable. The hydrophobic core constructs show such a tight packing and the introduced alanine residues therefore lead to only a small destabilizing effect. Natural proteins are often only marginally stabile, as often only their specific function or specificity has been optimized during evolution.

In addition to the *in vitro* studies of the protein stability using chemical unfolding experiments, we performed a force profile analysis for the hc, 1ip (Glu17-Lys72), and 1ip-N69 models and also for the associated alanine mutation series to obtain insights into the folding process under native conditions in the ribosome exit tunnel. It seems, that the bundle protein is a strong folder and therefore a strong arrest peptide with a second proline was used. The modification improves the arrest probability due to the ribosomal ability for translating prolines only very slowly. The bundle protein exerts a high folding force on the nascent chain, therefore stronger arrest peptides were used to improve the resolution of the force profiles and to study differences in the folding pathways of the three models. Interestingly, the 1ip and 1ip-N69 profiles show a shifted peak towards longer linker. It looks like that these two proteins fold later during the translational process, *i.e.*, towards the outside of the ribosomal exit tunnel. A possible explanation for this effect can be that the hydrophobic protein core needs to be partially folded in order that the ion-pair can be formed. The free energy from packing the hydrophobic core may help to compensate the energetic penalty of introducing charged residues inside the hydrophobic core.

Another interesting aspect is that we were able to show the stabilizing effect of the V69N mutation also with this method. The fraction of the full-length protein is significantly increased and the protein seems to be more stable than the 1ip model alone, but does not fully reach the level of the hc

construct. The 1ip-N69 protein folds later than the hc but prior to 1ip protein models suggesting that the introduced charge stabilizing residue (V69N) may improve the protein stability. The alanine substitutions perturb the non-polar interactions within the hydrophobic core as can also be observed from the force profiles. The peaks are decreased and shifted towards longer linkers with increasing number of alanines, supporting the notion that packing of the hydrophobic core is used to overcome the energetic cost of a buried ion-pair.

5.4 *Maquette* 3

Based on our combined findings for the *Maquette* 1 and 2 models, we designed the *Maquette* 3 models with a double ion-pair surrounded by four glutamine residues. The principle of a larger system size is here combined with stabilization of ion-pairs by charged residues and additional charge shielding motifs. Originally, the system was designed to bind cofactors, but the binding sites were successfully remodeled to create a densely packed hydrophobic core.[76] *Maquette* 3 is almost 70 Å long, and as discussed in the results section, it can compensate all four introduced charged residues. We were also able to show that the stabilization of the ion-pairs by other polar residues and the surrounding glutamines help to stabilize the hydrophobic core as no water molecules between the charged residues are observed in the crystal structure or MD simulations. The glutamines shield the charged residues from the non-polar surrounding and the protein stability is extremely high with an intact water sealed hydrophobic core, as indicated by both MD simulations and experiments. We successfully designed a hyperstable bundle with stable ion-pairs and showed that our proposed modifications increase the thermodynamic stability.

Another important aspect is the quantification of the stabilizing effect of the introduced glutamines. Therefore, a construct with Q61A and Q68A mutations was expressed and analyzed. The chemical unfolding measurements led to a transition point at 7.3 M GdnHCl, which is slightly higher as compared to 6.4 M for the 2ip *Maquette 3* construct with all four glutamines. The corresponding ΔG is with 8.3 kcal mol^{-1} smaller and indicates a 5.2 kcal mol^{-1} less stable protein. We were able to quantify the charge shielding effect of Q61 and Q68. Interestingly, the transition point of 2ip-Q61A-Q68A is higher compared to 2ip *Maquette* 3, but the protein stability is lower. Based on chemical unfolding experiments, the protein unfolds slower and the unfolding transition takes place at higher GdnHCl concentrations. The sigmoidal curve is broadened. In an ideal case, the signal of the fully folded protein stays constant without showing a slope in the beginning of the plot until the unfolding takes place. For the *Maquette* 3 model, the unfolding curve shows a slowly increasing slope before the transition point and the slope towards the transition point is also smaller, resulting in a lower ln$K$ slope, and a smaller ΔG value.

5.5 Stability comparison of *Maquette* 1, 2, and 3

Figure 98 shows the stability of all three tested *Maquette* models for the hydrophobic core, models with a single introduced charged residue, one ion-pair, and for the *Maquette* 3 model, also the double ion-pair. In general, the hydrophobic core of *Maquette* 1 is more stable than the *Maquette* 2 one. It seems that the hydrophobic core of *Maquette* 1 is more tightly packed in general, which is consistent with the observation that *Maquette* 2 has a water accessible dynamic part and a hydrophobic, packed core region. The flexible region was originally designed to bind a cofactor,[73] but used here to study the effect of introducing charged residues into solvent accessible parts of a protein. A possibly more stable variant of *Maquette* 2 could be designed by expanding the hydrophobic packed core region of the model. Comparing the models with only a single introduced charge (*Figure 98*, orange and light orange), suggest that *Maquette* 2 constructs are slightly more stable as compared to the 1ip models. The cost of introducing a buried ion-pair into the hydrophobic protein core of *Maquette* 1 is 5.9 kcal mol^{-1} which is consistent with the in literature reported 5 kcal mol^{-1} for natural proteins.[91] The solvation penalty of introducing buried ion-pair in *Maquette* 2 is $\Delta\Delta G \sim 1.9$ kcal mol^{-1}, but the hydrophobic core structure is less stable by 5.3 kcal mol^{-1} as compared to *Maquette* 1. Interestingly, the stabilizing effect of the charge shielding Asn69 is small but visible. The *Maquette* 3 2ip model is the most stable protein in this thesis, with a ΔG of 13.5 kcal mol^{-1}, and provides a basis for the successful design of a stable buried charged network with two ion-pairs.

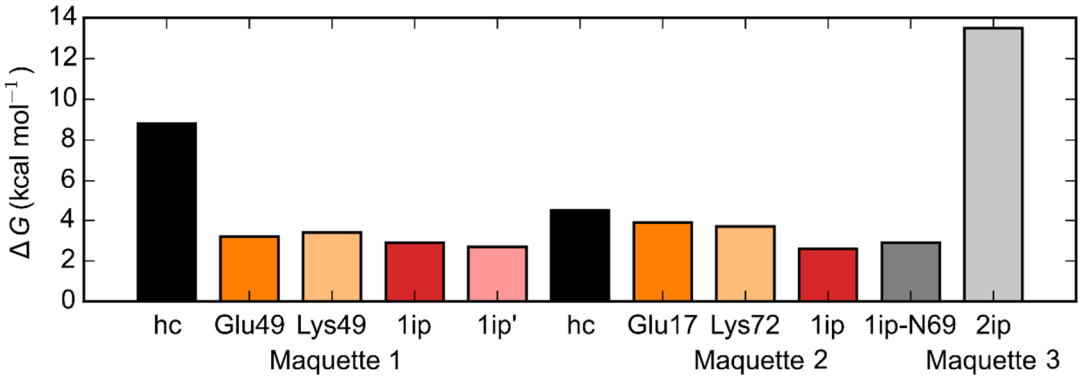

Figure 98: Protein stability of different *Maquette* 1, 2, and 3 models. The hydrophobic cores are colored in black, single buried charges in orange and light orange, 1ip in red, 1ip' in coral, 1ip-N69 in grey, and 2ip *Maquette* 3 in light grey.

The *Maquette* 1 and 2 models with a packed hydrophobic core are all stable as indicated by the thermal and chemical unfolding with GdnHCl, and consistent with the hyperstable nature of artificially designed α-helical bundle proteins. Already one of the first designed proteins by DeGrado's group, called α2, showed a high resistance against GdnHCl and unfoldings at 4.5 M,[8] which could be improved for α4 model that denatures at 6-7 M GdnHCl.[9] Also other artificially designed bundle proteins have high transition points with 5.5 M GdnHCl and a corresponding free energy of unfolding of 18.1 ± 0.8 kcal mol^{-1}.[18] In contrast, small natural proteins normally unfold

between 1 M and 3 M GdnHCl.[9] The designed bundle protein α4 has several stabilizing elements unified in a single protein. For example, multiple electrostatic interactions were introduced to increase the protein stability in combination with a tightly packed hydrophobic core that increase the stability over natural proteins.[9]

To summarize our findings, we successfully characterized buried ion-pairs using three different artificial model systems *Maquettes* 1-3 by combining computational, biophysical, and structural experiments. Furthermore, we discovered by analysis of natural membrane proteins and independently by developing a protein design model (PACMAN algorithm) that shielding amphiphilic residues contribute to compensate the energetic cost of a buried ion-pair. Our designed models may contribute to a better understanding of charged networks in natural proteins, how these proteins remain folded,[30,95,147] and how disease-causing mutations may influence the protein structure.[148–150]

6. Bibliography

1. Duan, G. & Walther, D. The Roles of Post-translational Modifications in the Context of Protein Interaction Networks. *PLoS Comput. Biol.* **11**, (2015) DOI: 10.1371/journal.pcbi.1004049.

2. Skjærven, L., Reuter, N. & Martinez, A. Dynamics, flexibility and ligand-induced conformational changes in biological macromolecules: A computational approach. *Future Medicinal Chemistry* **3**, 2079–2100 (2011) DOI: 10.4155/fmc.11.159.

3. Ha, J. H. & Loh, S. N. Protein conformational switches: From nature to design. *Chemistry - A European Journal* **18**, 7984–7999 (2012) DOI: 10.1002/chem.201200348.

4. Grant, B. J., Gorfe, A. A. & McCammon, J. A. Large conformational changes in proteins: Signaling and other functions. *Current Opinion in Structural Biology* **20**, 142–147 (2010) DOI: 10.1016/j.sbi.2009.12.004.

5. Hammes, G. G. Multiple conformational changes in enzyme catalysis. *Biochemistry* **41**, 8221–8228 (2002) DOI: 10.1021/bi0260839.

6. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007) DOI: 10.1038/nature06522.

7. Robertson, D. E. *et al.* Design and synthesis of multi-haem proteins. *Nature* **368**, 425–432 (1994) DOI: 10.1038/368425a0.

8. Regan, L. & Degrado, W. F. Characterization of a helical protein designed from first principles. *Science* **241**, 976–978 (1988) DOI: 10.1126/science.3043666.

9. DeGrado, W. F., Wasserman, Z. R. & Lear, J. D. Protein design, a minimalist approach. *Science* **243**, 622–628 (1989) DOI: 10.1126/science.2464850.

10. Anderson, J. L. L. R., Koder, R. L., Moser, C. C. & Dutton, P. L. Controlling complexity and water penetration in functional de novo protein design. *Biochem. Soc. Trans.* **36**, 1106–1111 (2008) DOI: 10.1042/BST0361106.

11. Feynman's Office: The Last Blackboards. *Phys. Today* **42**, 88 (1989) DOI: 10.1063/1.2810904.

12. Farid, T. A. *et al.* Elementary tetrahelical protein design for diverse oxidoreductase functions. *Nat. Chem. Biol.* **9**, 826–833 (2013) DOI: 10.1038/nchembio.1362.

13. Koder, R. L. *et al.* Design and engineering of an O2 transport protein. *Nature* **458**, 305–309 (2009) DOI: 10.1038/nature07841.

14. Kodali, G. *et al.* Design and engineering of water-soluble light-harvesting protein maquettes. *Chem. Sci.* **8**, 316–324 (2016) DOI: 10.1039/c6sc02417c.

15. DeGrado, W. F., Summa, C. M., Pavone, V., Nastri, F. & Lombardi, A. De novo design and structural characterization of proteins and metalloproteins. *Annual Review of Biochemistry* **68**, 779–819 (1999) DOI: 10.1146/annurev.biochem.68.1.779.

16. Shifman, J. M., Moser, C. C., Kalsbeck, W. A., Bocian, D. F. & Dutton, P. L. Functionalized de Novo designed proteins: Mechanism of proton coupling to oxidation/reduction in heme protein maquettes. *Biochemistry* **37**, 16815–16827 (1998) DOI: 10.1021/bi9816857.

17. Lombardi, A., Pirro, F., Maglio, O., Chino, M. & DeGrado, W. F. De novo design of four-helix bundle metalloproteins: One scaffold, diverse reactivities. *Acc. Chem. Res.* **52**, 1148–1159 (2019) DOI: 10.1021/acs.accounts.8b00674.

18. Sharp, R. E., Moser, C. C., Rabanal, F. & Dutton, P. L. Design, synthesis, and characterization of a photoactivatable flavocytochrome molecular maquette. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 10465–10470 (1998) DOI: 10.1073/pnas.95.18.10465.

19. Mathews, F. S., Bethge, P. H. & Czerwinski, E. W. The structure of cytochrome b562 from Escherichia coli at 2.5 A resolution. *J. Biol. Chem.* **254**, 1699–1706 (1979) DOI: 10.1016/s0021-9258(17)37829-8.

20. Banner, D. W., Kokkinidis, M. & Tsernoglou, D. Structure of the ColE1 Rop protein at 1.7 Å resolution. *J. Mol. Biol.* **196**, 657–675 (1987) DOI: 10.1016/0022-2836(87)90039-8.

21. Nicola, N. A. & Hilton, D. J. General classes and functions of four-helix bundle cytokines. *Advances in Protein Chemistry* **52**, 1–65 (1998) DOI: 10.1016/s0065-3233(08)60432-5.

22. Itagaki, E. & Hager, L. P. Studies on cytochrome b-562 of Escherichia coli. I. Purification and crystallization of cytochrome b-562. *J. Biol. Chem.* **241**, 3687–3695 (1966) DOI: 10.1016/S0021-9258(18)99827-3.

23. Arnesano, F. *et al.* The solution structure of oxidized Escherichia coli cytochrome b562. *Biochemistry* **38**, 8657–8670 (1999) DOI: 10.1021/bi982785f.

24. Wittung-Stafshede, P., Gray, H. B. & Winkler, J. R. Rapid formation of a four-helix bundle. Cytochrome b562 folding triggered by electron transfer. *J. Am. Chem. Soc.* **119**, 9562–9563 (1997) DOI: 10.1021/ja971855n.

25. Steif, C. *et al.* Subunit Interactions Provide a Significant Contribution to the Stability of the Dimeric Four-α-Helical-Bundle Protein ROP. *Biochemistry* **32**, 3867–3876 (1993) DOI: 10.1021/bi00066a005.

26. Betz, S. F., Liebman, P. A. & DeGrado, W. F. De novo design of native proteins: Characterization of proteins intended to fold into antiparallel, rop-like, four-helix bundles. *Biochemistry* **36**, 2450–2458 (1997) DOI: 10.1021/bi961704h.

27. Munson, M., Regan, L., O'Brien, R. & Sturtevant, J. M. Redesigning the hydrophobic core of a four-helix-bundle protein. *Protein Sci.* **3**, 2015–2022 (1994) DOI: 10.1002/pro.5560031114.

28. Rozwarski, D. A. *et al.* Structural comparisons among the short-chain helical cytokines. *Structure* **2**, 159–173 (1994) DOI: 10.1016/S0969-2126(00)00018-6.

29. Hill, E. E., Morea, V. & Chothia, C. Sequence conservation in families whose members have little or no sequence similarity: The four-helical cytokines and cytochromes. *J. Mol. Biol.* **322**, 205–233 (2002) DOI: 10.1016/S0022-2836(02)00653-8.

30. Ali, M. M. U. *et al.* Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* **440**, 1013–1017 (2006) DOI: 10.1038/nature04716.

31. Kaila, V. R. I. Long-range proton-coupled electron transfer in biological energy conversion: Towards mechanistic understanding of respiratory complex i. *Journal of the Royal Society Interface* **15**, (2018) DOI: 10.1098/rsif.2017.0916.

32. Schopf, F. H., Biebl, M. M. & Buchner, J. The HSP90 chaperone machinery. *Nature Reviews Molecular Cell Biology* **18**, 345–360 (2017) DOI: 10.1038/nrm.2017.20.

33. Mader SL, Lopez A, Lawatschek J, Luo Q, Rutz DA, Gamiz-Hernandez AP, Sattler M, Buchner J, K. V. Conformational dynamics modulate the catalytic of the molecular chaperone Hsp90. *Nat. Commun.* **11**, 1410 (2020) DOI:10.1038/s41467-020-15050-0.

34. Li, J. & Buchner, J. Structure, function and regulation of the Hsp90 machinery. *Biomedical Journal* **36**, 106–117 (2013) DOI: 10.4103/2319-4170.113230.

35. Jahn, M. *et al.* The charged linker of the molecular chaperone Hsp90 modulates domain contacts and biological function. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 17881–17886 (2014) DOI: 10.1073/pnas.1414073111.

36. Pearl, L. H. & Prodromou, C. Structure and Mechanism of the Hsp90 Molecular Chaperone Machinery. *Annu. Rev. Biochem.* **75**, (2006). DOI:10.1146/annurev.biochem.75.103004.142738.

37. Hessling, M., Richter, K. & Buchner, J. Dissection of the ATP-induced conformational cycle of the molecular chaperone Hsp90. *Nat. Struct. Mol. Biol.* **16**, 287–293 (2009) DOI: 10.1038/nsmb.1565.

38. Shiau, A. K., Harris, S. F., Southworth, D. R. & Agard, D. A. Structural Analysis of E. coli hsp90 Reveals Dramatic Nucleotide-Dependent Conformational Rearrangements. *Cell* **127**, 329–340 (2006) DOI: 10.1016/j.cell.2006.09.027.

39. Panaretou, B. *et al.* ATP binding and hydrolysis are essential to the function of the Hsp90 molecular chaperone in vivo. *EMBO J.* **17**, 4829–4836 (1998).

40. Zierer, B. K. *et al.* Importance of cycle timing for the function of the molecular chaperone Hsp90. *Nat. Struct. Mol. Biol.* **23**, 1020–1028 (2016) DOI: 10.1038/nsmb.3305.

41. Soroka, J. *et al.* Conformational Switching of the Molecular Chaperone Hsp90 via Regulated Phosphorylation. *Mol. Cell* **45**, 517–528 (2012) DOI: 10.1016/j.molcel.2011.12.031.

42. Morra, G., Verkhivker, G. & Colombo, G. Modeling signal propagation mechanisms and ligand-based conformational dynamics of the Hsp90 molecular chaperone full-length dimer. *PLoS Comput. Biol.* **5**, (2009) DOI: 10.1371/journal.pcbi.1000323.

43. Isaacs, J. S., Xu, W. & Neckers, L. Heat shock protein 90 as a molecular target for cancer therapeutics. *Cancer Cell* vol. 3 213–217 (2003) DOI: 10.1016/S1535-6108(03)00029-1.

44. Whitesell, L. & Lindquist, S. L. HSP90 and the chaperoning of cancer. *Nature Reviews Cancer* **5**, 761-772 (2005) DOI: 10.1038/nrc1716.

45. Vinothkumar, K. R., Zhu, J. & Hirst, J. Architecture of mammalian respiratory complex I. *Nature* **515**, 80–84 (2014) DOI: 10.1038/nature13686.

46. Baradaran, R., Berrisford, J. M., Minhas, G. S. & Sazanov, L. A. Crystal structure of the entire respiratory complex i. *Nature* **494**, 443–448 (2013) DOI: 10.1038/nature11871.

47. Hirst, J. Mitochondrial complex i. *Annual Review of Biochemistry* **82**, 551–575 (2013) DOI: 10.1146/annurev-biochem-070511-103700.

48. Sazanov, L. A. A giant molecular proton pump: Structure and mechanism of respiratory complex I. *Nature Reviews Molecular Cell Biology* **16**, 375–388 (2015) DOI: 10.1038/nrm3997.

49. Schapira, A. H. V. Human complex I defects in neurodegenerative diseases. *Biochim. Biophys. Acta - Bioenerg.* **1364**, 261–270 (1998) DOI: 10.1016/S0005-2728(98)00032-2.

50. Murphy, M. P. How mitochondria produce reactive oxygen species. *Biochemical Journal* **417**, 1–13 (2009) DOI: 10.1042/BJ20081386.

51. Balaban, R. S., Nemoto, S. & Finkel, T. Mitochondria, oxidants, and aging. *Cell* **120**, 483–495 (2005) DOI: 10.1016/j.cell.2005.02.001.

52. Dougherty, M. J. & Arnold, F. H. Directed evolution: new parts and optimized function. *Current Opinion in Biotechnology* **20**, 486–491 (2009) DOI:10.1016/j.copbio.2009.08.005.

53. Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016) DOI: 10.1038/nature19946.

54. Koga, N. *et al.* Principles for designing ideal protein structures. *Nature* **491**, 222–227 (2012) DOI: 10.1038/nature11600.

55. Kuhlman, B. *et al.* Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **302**, 1364–1368 (2003) DOI: 10.1126/science.1089427.

56. Hecht, M. H., Richardson, J. S., Richardson, D. C. & Ogden, R. C. De novo design, expression, and characterization of Felix: A four-helix bundle protein of native-like sequence. *Science* **249**, 884–891 (1990) DOI: 10.1126/science.2392678.

57. Degrado, W. F., Summa, C. M., Pavone, V., Nastri, F. & Lombardi, A. De Novo Design and Structural Characterization of Proteins and Metalloproteins. *Annu. Rev. Biochem.* **68**, 779–819 (1999).

58. Frauenfelder, H. Proteins: Paradigms of complexity. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 2479–2480 (2002) DOI: 10.1073/pnas.012579999.

59. Kuhlman, B. & Bradley, P. Advances in protein structure prediction and design. *Nature Reviews Molecular Cell Biology* **20**, 681–697 (2019) DOI: 10.1038/s41580-019-0163-x.

60. DeGrado, W. F. & Lear, J. D. Induction of Peptide Conformation at Apolar/Water Interfaces. 1. A Study with Model Peptides of Defined Hydrophobic Periodicity. *J. Am. Chem. Soc.* **107**, 7684–7689 (1985) DOI: 10.1021/ja00311a076.

61. Richards, F. M. Areas, volumes, packing and protein structure. *Annual review of biophysics and bioengineering* **6**, 151–176 (1977) DOI: 10.1146/annurev.bb.06.060177.001055.

62. Richmond, T. J. & Richards, F. M. Packing of α-helices: Geometrical constraints and contact areas. *J. Mol. Biol.* **119**, 537–555 (1978) DOI: 10.1016/0022-2836(78)90201-2.

63. Crick, F. H. C. The packing of α-helices: simple coiled-coils. *Acta Crystallogr.* **6**, 689–697 (1953) DOI: 10.1107/s0365110x53001964.

64. Bryson, J. W. *et al.* Protein design: A hierarchic approach. *Science* **270**, 935–941 (1995) DOI: 10.1126/science.270.5238.935.

65. Korendovych, I. V. & DeGrado, W. F. De novo protein design, a retrospective. *Quarterly Reviews of Biophysics* **53**, (2020) DOI: 10.1017/S0033583519000131.

66. Desjarlais, J. R. & Handel, T. M. De novo design of the hydrophobic cores of proteins. *Protein Sci.* **4**, 2006–2018 (1995) DOI: 10.1002/pro.5560041006.

67. Dahiyat, B. I. & Mayo, S. L. De novo protein design: Fully automated sequence selection. *Science* **278**, 82–87 (1997) DOI: 10.1126/science.278.5335.82.

68. Das, R. & Baker, D. Macromolecular modeling with Rosetta. *Annual Review of Biochemistry* **77**, 363–382 (2008) DOI: 10.1146/annurev.biochem.77.062906.171838.

69. Alford, R. F. *et al.* The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017) DOI: 10.1021/acs.jctc.7b00125.

70. Rohl, C. A., Strauss, C. E. M., Misura, K. M. S. & Baker, D. Protein Structure Prediction Using Rosetta. *Methods Enzymol.* **383**, 66–93 (2004) DOI: 10.1016/S0076-6879(04)83004-0.

71. Senior, A. W. *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* **577**, (2020) DOI: 10.1038/s41586-019-1923-7.

72. Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K. & Moult, J. Critical assessment of methods of protein structure prediction (CASP)—Round XIII. *Proteins: Structure, Function and Bioinformatics* **87**, (2019) DOI: 10.1002/prot.25823.

73. Polizzi, N. F. *et al.* De novo design of a hyperstable non-natural protein-ligand complex with sub-Å accuracy. *Nat. Chem.* **9**, 1157–1164 (2017) DOI: 10.1038/nchem.2846.

74. Fry, H. C. *et al.* Computational de novo design and characterization of a protein that selectively binds a highly hyperpolarizable abiological chromophore. *J. Am. Chem. Soc.* **135**, 13914–13926 (2013) DOI: 10.1021/ja4067404.

75. Choma, C. T. *et al.* Design of a heme-binding four-helix bundle. *J. Am. Chem. Soc.* **116**, 856–865 (1994) DOI: 10.1021/ja00082a005.

76. Ennist, N. M. Design, structure, and action of an artificial photosynthetic reaction center. (2016).

77. Faiella, M. *et al.* An artificial di-iron oxo-protein with phenol oxidase activity. *Nat. Chem. Biol.* **5**, 882–884 (2009) DOI: 10.1038/nchembio.257.

78. Torres Martin De Rosales, R. *et al.* Spectroscopic and metal-binding properties of DF3: An artificial protein able to accommodate different metal ions. *J. Biol. Inorg. Chem.* **15**, 717-728 (2010) DOI: 10.1007/s00775-010-0639-9.

79. Calhoun, J. R. *et al.* Artificial diiron proteins: From structure to function. *Biopolymers - Peptide Science Section* **80**, 264–278 (2005) DOI: 10.1002/bip.20230.

80. Cohen, C. & Parry, D. A. D. α-Helical coiled coils and bundles: How to design an α-helical protein. *Proteins: Structure, Function, and Bioinformatics* **7**, 1–15 (1990) DOI: 10.1002/prot.340070102.

81. Perutz, M. F. Electrostatic effects in proteins. *Science (80-.).* **201**, 1187–1191 (1978) DOI: 10.1126/science.694508.

82. Kumar, S. & Nussinov, R. Close-range electrostatic interactions in proteins. *ChemBioChem* **3**, 604–617 (2002) DOI: 10.1002/1439-7633(20020703)3:7<604::AID-CBIC604>3.0.CO;2-X.

83. Oliveberg, M. & Fersht, A. R. Formation of electrostatic interactions on the protein-folding pathway. *Biochemistry* **35**, 2726–2737 (1996) DOI: 10.1021/bi9509661.

84. Oliveberg, M. & Fersht, A. R. A new approach to the study of transient protein conformations: The formation of a semiburied salt link in the folding pathway of barnase. *Biochemistry* **35**, 6795–6805 (1996) DOI: 10.1021/bi9529317.

85. Donald, J. E., Kulp, D. W. & DeGrado, W. F. Salt bridges: Geometrically specific, designable interactions. *Proteins Struct. Funct. Bioinforma.* **79**, 898–915 (2011) DOI: 10.1002/prot.22927.

86. Braun-Sand, S. & Warshel, A. Electrostatics of Proteins: Principles, Models and Applications. in *Protein Folding Handbook* **1**, 163–200 (2008). DOI: 10.1002/9783527619498.ch7.

87. Perutz, M. F. & Raidt, H. Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. *Nature* **255**, 256–259 (1975) DOI: 10.1038/255256a0.

88. Berezovsky, I. N., Zeldovich, K. B. & Shakhnovich, E. I. Positive and negative design in stability and thermal adaptation of natural proteins. *PLoS Comput. Biol.* **3**, 0498–0507 (2007) DOI: 10.1371/journal.pcbi.0030052.

89. Loladze, V. V., Ermolenko, D. N. & Makhatadze, G. I. Thermodynamic consequences of burial of polar and non-polar amino acid residues in the protein interior. *J. Mol. Biol.* **320**, 343–357 (2002) DOI: 10.1016/S0022-2836(02)00465-5.

90. Sindelar, C. V., Hendsch, Z. S. & Tidor, B. Effects of salt bridges on protein structure and design. *Protein Sci.* **7**, 1898–1914 (1998) DOI: 10.1002/pro.5560070906.

91. Robinson, A. C., Castañeda, C. A., Schlessman, J. L. & Bertrand García-Moreno, E. Structural and thermodynamic consequences of burial of an artificial ion pair in the hydrophobic interior of a protein. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 11685–11690 (2014) DOI: 10.1073/pnas.1402900111.

92. Warshel, A. What about protein polarity? *Nature* **330**, 15–16 (1987) DOI: 10.1038/330015a0.

93. Hendsch, Z. S. & Tidor, B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**, 211–226 (1994) DOI: 10.1002/pro.5560030206.

94. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973) DOI: 10.1126/science.181.4096.223.

95. Mader, S. L. *et al.* Conformational dynamics modulate the catalytic activity of the molecular chaperone Hsp90. *Nat. Commun.* **11**, (2020) DOI: 10.1038/s41467-020-15050-0.

96. Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **103**, 227–249 (1976) DOI: 10.1016/0022-2836(76)90311-9.

97. Warshel, A. Energetics of enzyme catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 5250–5254 (1978) DOI: 10.1073/pnas.75.11.5250.

98. Baumgart M., Röpke M., Mühlbauer M. E., Asami S., Mader S.L., Fredriksson K., Groll M., Gamiz-Hernandez A. P., Kaila V. R. I. Design of Buried Charged Networks in Artificial Proteins. *Nat. Commun.* **12**, 1895 (2021).

99. Isom, D. G., Castañeda, C. A., Cannon, B. R., Velu, P. D. & García-Moreno E, B. Charges in the hydrophobic interior of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16096–16100 (2010) DOI: 10.1073/pnas.1004213107.

100. Christos M. Kougentakis, Lauren Skerritt, Ananya Majumdar, Jamie L. Schlessman, B. G.-M. E. The properties of buried ion pairs are governed by the propensity of proteins to reorganize. (2020) DOI: https://doi.org/10.1101/2020.02.03.932012.

101. Notari, L., Martínez-Carranza, M., Farías-Rico, J. A., Stenmark, P. & von Heijne, G. Cotranslational Folding of a Pentarepeat β-Helix Protein. *J. Mol. Biol.* **430**, 5196–5206 (2018) DOI: 10.1016/j.jmb.2018.10.016.

102. Marino, J., Von Heijne, G. & Beckmann, R. Small protein domains fold inside the ribosome exit tunnel. *FEBS Lett.* **590**, 655–660 (2016) DOI: 10.1002/1873-3468.12098.

103. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015) DOI: 10.1093/bioinformatics/btu830.

104. Fersht, A. Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding; series in structural biology **9**. *World Scientific* (2017).

105. Keller, R. The computer aided resonance assignment tutorial. Goldau, Switzerland: *Cantina Verlag* (2004).

106. Güntert, P. & Buchner, L. Combined automated NOE assignment and structure calculation with CYANA. *J. Biomol. NMR* **62**, 453–471 (2015) DOI: 10.1007/s10858-015-9924-9.

107. Güntert, P., Mumenthaler, C. & Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298 (1997) DOI: 10.1006/jmbi.1997.1284.

108. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241 (2013) DOI: 10.1007/s10858-013-9741-y.

109. Ponder, J. W. & Case, D. A. Force fields for protein simulations. *Adv. Protein Chem.* **66**, 27–85 (2003) DOI: 10.1016/S0065-3233(03)66002-X.

110. Koradi, R., Billeter, M. & Güntert, P. Point-centered domain decomposition for parallel molecular dynamics simulation. *Comput. Phys. Commun.* **124**, 139–147 (2000) DOI: 10.1016/S0010-4655(99)00436-1.

111. Vagin, A. A. *et al.* REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. Sect. D Biol. Crystallogr.* D**60**, 2184-2195 (2004) DOI: 10.1107/S0907444904023510.

112. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* D**66**, 486–501 (2010) DOI: 10.1107/S0907444910007493.

113. Yap, M. N. & Bernstein, H. D. The Plasticity of a Translation Arrest Motif Yields Insights into Nascent Polypeptide Recognition inside the Ribosome Tunnel. *Mol. Cell* **34**, (2009) DOI: 10.1016/j.molcel.2009.04.002.

114. Jones, J. E. On the determination of molecular fields.—I. From the variation of the viscosity of a gas with temperature. *Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character* **106**, 441–462 (1924) DOI: 10.1098/rspa.1924.0081.

115. Petrella, R. J., Andricioaei, I., Brooks, B. R. & Karplus, M. An improved method for nonbonded list generation: Rapid determination of near-neighbor pairs. *J. Comput. Chem.* **24**, 222–231 (2003) DOI: 10.1002/jcc.10123.

116. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993) DOI: 10.1063/1.464397.

117. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998) DOI: 10.1021/jp973084f.

118. Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983) DOI: 10.1002/jcc.540040211.

119. Foloppe, N. & MacKerell, A. D. All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J. Comput. Chem.* **21**, 86–104 (2000) DOI: 10.1002/(SICI)1096-987X(20000130)21:2<86::AID-JCC2>3.0.CO;2-G.

120. Huang, J. & Mackerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013) DOI: 10.1002/jcc.23354.

121. Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 Dihedral Angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012) DOI: 10.1021/ct300400x.

122. Brooks, B. R. *et al.* CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009) DOI: 10.1002/jcc.21287.

123. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781–1802 (2005) DOI: 10.1002/jcc.20289.

124. Verlet, L. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **159**, 98 (1967) DOI: 10.1103/PhysRev.159.98.

125. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977) DOI: 10.1016/0021-9991(77)90098-5.

126. Pastor, R. W., Brooks, B. R. & Szabo, A. An analysis of the accuracy of langevin and molecular dynamics algorithms. *Mol. Phys.* **65**, 1409–1419 (1988) DOI: 10.1080/00268978800101881.

127. Lzaguirre, J. A., Catarello, D. P., Wozniak, J. M. & Skeel, R. D. Langevin stabilization of molecular dynamics. *J. Chem. Phys.* **114**, 2090–2098 (2001) DOI: 10.1063/1.1332996.

128. Barth, E. & Schlick, T. Overcoming stability limitations in biomolecular dynamics. I. Combining force splitting via extrapolation with Langevin dynamics in LN. *J. Chem. Phys.* **109**, 1617–1632 (1998) DOI: 10.1063/1.476736.

129. Alder, B. J. & Wainwright, T. E. Studies in molecular dynamics. I. General method. *J. Chem. Phys.* **31**, 459–466 (1959) DOI: 10.1063/1.1730376.

130. 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.* **79**, 926–935 (1983) DOI: 10.1063/1.445869.

131. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996) DOI: 10.1016/0263-7855(96)00018-5.

132. Buchan, D. W. A. & Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019) DOI: 10.1093/nar/gkz297.

133. Gámiz-Hernández, A. P., Kieseritzky, G., Galstyan, A. S., Demir-Kavuk, O. & Knapp, E. W. Understanding properties of cofactors in proteins: Redox potentials of synthetic cytochromes b. *ChemPhysChem* **11**, 1196–1206 (2010) DOI: 10.1002/cphc.200900889.

134. Iwahara, J., Jung, Y. S. & Clore, G. M. Heteronuclear NMR spectroscopy for lysine NH3 groups in proteins: Unique effect of water exchange on 15N transverse relaxation. *J. Am. Chem. Soc.* **129**, 2971–2980 (2007) DOI: 10.1021/ja0683436.

135. Farías-Rico, J. A., Selin, F. R., Myronidi, I., Frühauf, M. & Von Heijne, G. Effects of protein size, thermodynamic stability, and net charge on cotranslational folding on the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E9280–E9287 (2018) DOI: 10.1073/pnas.1812756115.

136. Oroz, J., Kim, J. H., Chang, B. J. & Zweckstetter, M. Mechanistic basis for the recognition of a misfolded protein by the molecular chaperone Hsp90. *Nat. Struct. Mol. Biol.* **24**, 407–413 (2017) DOI: 10.1038/nsmb.3380.

137. Wegele, H., Muschler, P., Bunck, M., Reinstein, J. & Buchner, J. Dissection of the contribution of individual domains to the ATPase mechanism of Hsp90. *J. Biol. Chem.* **278**, (2003) DOI: 10.1074/jbc.M305751200.

138. Tsutsumi, S. *et al.* Charged linker sequence modulates eukaryotic heat shock protein 90 (Hsp90) chaperone activity. *Proc. Natl. Acad. Sci. U. S. A.* **109**, (2012) DOI: 10.1073/pnas.1114414109.

139. Hainzl, O., Lapina, M. C., Buchner, J. & Richter, K. The charged linker region is an important regulator of Hsp90 function. *J. Biol. Chem.* **284**, 22559–22567 (2009) DOI: 10.1074/jbc.M109.031658.

140. Farrelly, F. W. & Finkelstein, D. B. Complete sequence of the heat shock-inducible HSP90 gene of Saccharomyces cerevisiae. *J. Biol. Chem.* **259**, 5745–5751 (1984) DOI: 10.1016/s0021-9258(18)91077-x.

141. Hwang, J. K. & Warshel, A. Why ion pair reversal by protein engineering is unlikely to succeed. *Nature* **334**, 270–272 (1988) DOI: 10.1038/334270a0.

142. Šali, D., Bycroft, M. & Fersht, A. R. Stabilization of protein structure by interaction of α-helix dipole with a charged side chain. *Nature* **335**, 740–743 (1988) DOI: 10.1038/335740a0.

143. Robinson, A. C., Schlessman, J. L. & García-Moreno, B. E. Dielectric Properties of a Protein Probed by Reversal of a Buried Ion Pair. *J. Phys. Chem. B* **122**, 2516–2524 (2018) DOI: 10.1021/acs.jpcb.7b12121.

144. Bush, J. & Makhatadze, G. I. Statistical analysis of protein structures suggests that buried ionizable residues in proteins are hydrogen bonded or form salt bridges. *Proteins Struct. Funct. Bioinforma.* **79**, 2027–2032 (2011) DOI: 10.1002/prot.23067.

145. Rashin, A. A. & Honig, B. On the environment of ionizable groups in globular proteins. *J. Mol. Biol.* **173**, 515–521 (1984) DOI: 10.1016/0022-2836(84)90394-2.

146. Holtkamp, W. *et al.* Cotranslational protein folding on the ribosome monitored in real time. *Science* **350**, 1104–1107 (2015) DOI: 10.1126/science.aad0344.

147. Di Luca, A., Gamiz-Hernandez, A. P. & Kaila, V. R. I. Symmetry-related proton transfer pathways in respiratory complex I. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E6314–E6321 (2017) DOI: 10.1073/pnas.1706278114.

148. Boczek, E. E. *et al.* Conformational processing of oncogenic v-Src kinase by the molecular chaperone Hsp90. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3189–E3198 (2015) DOI: 10.1073/pnas.1424342112.

149. Fiedorczuk, K. *et al.* Atomic structure of the entire mammalian mitochondrial complex i. *Nature* **538**, 406–410 (2016) DOI: 10.1038/nature19794.

150. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003) DOI: 10.1038/nature02261.

7. Appendix

Amino acid sequence of *de novo* proteins

Construct	Sequence
***Maquette* 1** **Hydrophobic core**	G EIWKQFE DALQKFE EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQKFE EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQKFE EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQKFE EALNQFEDLKQL
***Maquette* 1** **Lys49**	G EIWKQFE DALQKFE EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQK**K**E EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQKFE EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQKFE EALNQFEDLKQL
***Maquette* 1** **Glu49**	G EIWKQFE DALQKFE EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQK**E**E EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQKFE EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQKFE EALNQFEDLKQL
***Maquette* 1** **1ip (Lys49-Glu84)**	G EIWKQFE DALQKFE EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQK**E**E EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQK**K**E EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQKFE EALNQFEDLKQL
***Maquette* 1** **1ip' (Glu49-Lys84)**	G EIWKQFE DALQKFE EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQK**K**E EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQK**E**E EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQKFE EALNQFEDLKQL
***Maquette* 1** **2ip (Lys42-Glu81/** **Lys49-Glu84)**	G EIWKQFE DALQKFE EALNQFEDLKQL GGSGSGSGG EIWKQ**K**E DALQK**K**E EALNQFEDLKQL GGSGSGSGG EIWKQFE DA**E**QK**E**E EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQKFE EALNQFEDLKQL
***Maquette* 1** **3ip (Lys42-Glu91/** **Lys49-Glu84/Lys56-Glu77)**	G EIWKQFE DALQKFE EALNQFEDLKQL GGSGSGSGG EIWKQ**K**E DALQK**K**E EALNQ**K**EDLKQL GGSGSGSGG EIWKQ**E**E DA**E**QK**E**E EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQKFE EALNQFEDLKQL
***Maquette* 1** **3ip (Glu42-Lys91/** **Glu49-Lys84/Glu56-Lys77)**	G EIWKQFE DALQKFE EALNQFEDLKQL GGSGSGSGG EIWKQ**E**E DALQK**E**E EALNQ**E**EDLKQL GGSGSGSGG EIWKQ**K**E DA**K**QK**K**E EALNQFEDLKQL GGSGSGSGG EIWKQFE DALQKFE EALNQFEDLKQL
***Maquette* 1** **3ip (Glu112-Lys24/** **Glu49-Lys84/Glu56-Lys77)**	G EIWKQFE DALQKFE EALNQFED**K**KQL GGSGSGSGG EIWKQFE DALQK**E**E EALNQ**E**EDLKQL GGSGSGSGG EIWKQ**K**E DA**K**QK**K**E EALNQFEDLKQL GGSGSGSGG EIWKQ**E**E DALQKFE EALNQFEDLKQL
***Maquette* 1** **3ip (Gln112-Lys24/** **Glu49-Lys84/Glu56-Lys77)**	G EIWKQFE DALQKFE EALNQFED**K**KQL GGSGSGSGG EIWKQFE DALQK**E**E EALNQ**E**EDLKQL GGSGSGSGG EIWKQ**K**E DA**K**QK**K**E EALNQFEDLKQL GGSGSGSGG EIWKQ**Q**E DALQKFE EALNQFEDLKQL

***Maquette* 1** **Hydrophobic core**	G	EIWKQFE DALQKFE EALNQFEDLKQL GGSG EIWKQFE DALQKFE EALNQFEDLKQL GGSG EIWKQFE DALQKFE EALNQFEDLKQL GGSG EIWKQFE DALQKFE EALNQFEDLKQL
***Maquette* 1** **Lys49**	G	EIWKQFE DALQKFE EALNQFEDLKQL GGSG EIWKQFE DALQK**K**E EALNQFEDLKQL GGSG EIWKQFE DALQKFE EALNQFEDLKQL GGSG EIWKQFE DALQKFE EALNQFEDLKQL
***Maquette* 1** **Glu49**	G	EIWKQFE DALQKFE EALNQFEDLKQL GGSG EIWKQFE DALQK**E**E EALNQFEDLKQL GGSG EIWKQFE DALQKFE EALNQFEDLKQL GGSG EIWKQFE DALQKFE EALNQFEDLKQL
***Maquette* 1** **1ip (Glu49-Lys84)**	G	EIWKQFE DALQKFE EALNQFEDLKQL GGSG EIWKQFE DALQK**K**E EALNQFEDLKQL GGSG EIWKQFE DALQK**E**E EALNQFEDLKQL GGSG EIWKQFE DALQKFE EALNQFEDLKQL
***Maquette* 1** **2ip (Glu42-Lys81/** **Glu49-Lys84)**	G	EIWKQFE DALQKFE EALNQFEDLKQL GGSG EIWKQ**E**E DALQK**E**E EALNQFEDLKQL GGSG EIWKQFE DA**K**QK**K**E EALNQFEDLKQL GGSG EIWKQFE DALQKFE EALNQFEDLKQL
***Maquette* 1** **3ip (Glu112-Lys24/** **Glu49-Lys84/Glu56-Lys77)**	G	EIWKQFE DALQKFE EALNQFED**K**KQL GGSG EIWKQFE DALQK**E**E EALNQ**E**EDLKQL GGSG EIWKQ**K**E DA**K**QK**K**E EALNQFEDLKQL GGSG EIWKQ**E**E DALQKFE EALNQFEDLKQL
***Maquette* 1** **3ip (Gln112-Lys24/** **Glu49-Lys84/Glu56-Lys77)**	G	EIWKQFE DALQKFE EALNQFED**K**KQL GGSG EIWKQFE DALQK**E**E EALNQ**E**EDLKQL GGSG EIWKQ**K**E DA**K**QK**K**E EALNQFEDLKQL GGSG EIWKQ**Q**E DALQKFE EALNQFEDLKQL
***Maquette* 2** **Hydrophobic core**	S	EFEKLRQ TGDELVQ AFQRLREIFDK GD DDSLEQV LEEIEEL IQKHRQLFDNR QEAA DTEAAKQ GDQWVQL FQRFREAIDK GD KDSLEQL LEELEQA LQKIRELAEKKN
***Maquette* 2** **1ip (Glu17-Lys72)**	S	EFEKLRQ TGDELVQ A**E**QRLREIFDK GD DDSLEQV LEEIEEL IQKHRQLFDNR QEAA DTEAAKQ GDQWVQL **K**QRFREAIDK GD KDSLEQL LEELEQA LQKIRELAEKKN
***Maquette* 2** **1ip-N69**	S	EFEKLRQ TGDELVQ A**E**QRLREIFDK GD DDSLEQV LEEIEEL IQKHRQLFDNR QEAA DTEAAKQ GDQW**N**QL **K**QRFREAIDK GD KDSLEQL LEELEQA LQKIRELAEKKN
***Maquette* 2** **1ip (Glu101-Lys50)**	S	EFEKLRQ TGDELVQ AFQRLREIFDK GD DDSLEQV LEEIEEL IQKHRQL**K**DNR QEAA DTEAAKQ GDQWVQL FQRFREAIDK GD KDSLEQL LEELEQA LQK**E**RELAEKKN

Maquette 3 **2ip**	M	ASPELRQ EFQQLIQ EFQQLLQEIQQL IRELLKI **K**LQIIKQ LREASEK ARN PEKKSVL QK**Q**LEL**E** EK**Q**IELLETLQQ TAQEAQQ LLQELQQ TGQELWQL GGSGG PELRQKF QQLAQKI QQLLQKFQQLVA K**Q**LED**E**E KFIELLE TIL GGSGG DELRELL KG**K**LQVI KQ**Q**RELLQLVQK AQQLVQK LQQTGQK LW

NMR and refinement statistics of *Maquette* 2

	***Maquette* 2 1ip open**		***Maquette* 2 1ip closed**	
	CYANA Result	**Energy Minimized[b]**	**CYANA Result**	**Energy Minimized[b]**
Automated NOE assignment[a]				
^{15}N-resolved NOESY cross-peaks	1758		1758	
^{13}C-resolved aliphatic NOESY cross-peaks	1938		1938	
^{13}C-resolved aromatic NOESY cross-peaks	237		237	
Total no. of NOESY cross-peaks	3933 (100%)		3933 (100%)	
Assigned cross-peaks	2861 (72.7%)		2854 (72.6%)	
Unassigned cross-peaks	1072 (27.3%)		1079 (27.4%)	
Structural restraints				
Assigned NOE distance restraints	1918 (100%)		1923 (100%)	
Short range ($\|i-j\| \leq 1$)	1109 (57.8%)		1119 (58.2%)	
Medium-range ($1 < \|i-j\| < 5$)	533 (27.8%)		512 (26.6%)	
Long-range ($\|i-j\| \geq 5$)	276 (14.4%)		292 (15.2%)	
Dihedral angle restraints TALOS-N (ϕ/ψ)	190		190	
Structure statistics				
Average CYANA target function value (Å^2)	2.25 ± 0.41	1.80 ± 0.27	9.48 ± 16.27	2.23 ± 0.45
Average AMBER energy (kcal/mol)	-3829.55 ± 81.08	-4847.10 ± 132.53	-3863.46 ± 159.03	-4850.34 ± 148.53
Restraint violations				
Max. distance restraint violation (Å)	0.95	0.12	2.76	0.13
No. of violated distance restraints >0.2 Å	13	0	10	1
Max. dihedral angle restraint violation (°)	6.67	3.59	5.68	3.41
No. of violated dihedral angle restraints >5°	4	1	4	1
Ramachandran plot				
Residues in most favored regions	92.1%	93.5%	92.6%	93.4%
Residues in additionally allowed regions	7.8%	6.2%	7.3%	6.3%
Residues in generously allowed regions	0.1%	0.3%	0.1%	0.3%
Residues in disallowed regions	0%	0%	0%	0%
RMSD (residues 1-109)**				
Average backbone RMSD to mean (Å)	0.75 ± 0.16	0.76 ± 0.15	0.75 ± 0.19	0.77 ± 0.18
Average heavy atom RMSD to mean (Å)	1.27 ± 0.14	1.29 ± 0.13	1.23 ± 0.18	1.27 ± 0.17

[a]Using structure calculation functionalities of CYANA.

[b]After restrained energy minimization with OPALp.

**Pairwise r.m.s. deviation was calculated among 20 structures.

Crystallographic data collection and refinement statistics of *Maquette* 3

	***Maquette* 3**
Data collection	
Space group	P1
Cell dimensions	
a, *b*, *c* (Å)	62.7, 65.9, 68.3
α, β, γ (°)	89.9, 90.3, 117.5
Resolution (Å)	30-1.85 (1.85-1.75)
R_{merge}	0.051 (0.545)
$I / \sigma I$	11.3 (2.1)
Completeness (%)	97.6 (98.5)
Redundancy	3.4 (3.6)
Refinement	
Resolution (Å)	30-1.85
No. reflections	76730
R_{work} / R_{free}	0.189 / 0.223
No. atoms	
Protein	6488
Ligand/ion	2
Water	123
B-factors	
Protein	45
Ligand/ion	41
Water	46
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	1.1

*Data have been collected from two crystals, values in parentheses are for highest-resolution shell.

List of publications

Baumgart M., Röpke M., Mühlbauer M. E., Asami S., Mader S. L., Fredriksson K., Groll M., Gamiz-Hernandez A. P., Kaila V. R. I. **Design of Buried Charged Networks in Artificial Proteins**. *Nat Commun*. **12**, 1895 (2021).

Jussupow A.[+], Lopez A.[+], Baumgart M.[+], Mader S. L., Sattler M., Kaila V. R. I. **Extendend conformational dynamics is crucial for the function of Hsp90**. Submitted at *Angew. Chem.*

[+] Authors contributed equally.

Acknowledgements

I would like to thank my supervisor Prof. Ville Kaila for the opportunity to be the first experimentalist in the group and for giving me insights into the computational world. Further I want to thank you for your support, helpful discussions and the trust in my work. I'm deeply grateful for the opportunity to learn so much and to try many different things during my PhD. It was a great time!

Thanks to Kai for being my mentor and the support during my PhD project. You always had an open ear for me even after you left TUM. You have always been the one to motivate me again even if the experiments took some time longer.

Many thanks to the whole Kaila group, located in Munich and Sweden for their help during my project. Anyone of you contributed somehow to the bundle project! It was a great adventure with anyone of you!

Thank you very much to all my experimentalist lab colleagues for the helpful discussions and the great time which we spent together. Especially to Andrea (now Dr. Andrea Steiner), you went along with me the last 10 years from our very first lecture until today. Also, thanks to Prof. Franz Hagn who gave me the opportunity to be part of his lab.

I would like to thank Prof. Michael Groll, Astrid König and Dr. Sam Asami for the successful collaborations and all their help with my project. I was very impressed by your patience and the amount of time that you committed to our project.

Last but not least I also want to express my thankfulness towards my family and friends for their infinite love, support and strong believe in me and my abilities. Without you I would surely not be in the place that I am today. A very special gratitude belongs to Basti who has gone through all the ups and downs at my side, also in situations were things did not go as I hoped for.

Zu guter Letzt möchte ich meiner Familie und Freunden für die unendliche Liebe, die Unterstützung und den unerschütterlichen Glauben an mich und meine Fähigkeiten danken. Ohne euch würde ich heute nicht hier stehen. Ein besonderer Dank geht hier an Basti, der all meine Höhen und Tiefen, wenn es mal nicht so lief wie ich mir das gewünscht habe, miterlebt hat und immer an meiner Seite steht.

Anhang I

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die bei der promotionsführenden Einrichtung
Technische Universität München – Fakultät Chemie

der TUM zur Promotionsprüfung vorgelegte Arbeit mit dem Titel:
Function und Design of Protein Conformational Switches

in Fakultät Chemie, Lehrstuhl für computergestützte Biokatalyse
Fakultät, Institut, Lehrstuhl, Klinik, Krankenhaus, Abteilung

unter der Anleitung und Betreuung durch: Professor V.R.I. Kaila ohne sonstige Hilfe erstellt und bei der Abfassung nur die gemäß § 6 Ab. 6 und 7 Satz 2 angebotenen Hilfsmittel benutzt habe.

☒ Ich habe keine Organisation eingeschaltet, die gegen Entgelt Betreuerinnen und Betreuer für die Anfertigung von Dissertationen sucht, oder die mir obliegenden Pflichten hinsichtlich der Prüfungsleistungen für mich ganz oder teilweise erledigt.

☒ Ich habe die Dissertation in dieser oder ähnlicher Form in keinem anderen Prüfungsverfahren als Prüfungsleistung vorgelegt.

☒ Die vollständige Dissertation wurde in englischer Sprache veröffentlicht. Die promotionsführende Einrichtung
Technische Universität München – Fakultät Chemie
hat der Veröffentlichung zugestimmt.

☒ Ich habe den angestrebten Doktorgrad noch nicht erworben und bin nicht in einem früheren Promotionsverfahren für den angestrebten Doktorgrad endgültig gescheitert.

☐ Ich habe bereits am _____ bei der Fakultät für _____ der _____ Hochschule _____ unter Vorlage einer Dissertation mit dem Thema _____ die Zulassung zur Promotion beantragt mit dem Ergebnis: _____

Die öffentlich zugängliche Promotionsordnung der TUM ist mir bekannt, insbesondere habe ich die Bedeutung von § 28 (Nichtigkeit der Promotion) und § 29 (Entzug des Doktorgrades) zur Kenntnis genommen. Ich bin mir der Konsequenzen einer falschen Eidesstattlichen Erklärung bewusst.

Mit der Aufnahme meiner personenbezogenen Daten in die Alumni-Datei bei der TUM bin ich

☒ einverstanden, ☐ nicht einverstanden.

Ort, Datum, Unterschrift