Conformation Prediction of Synthetic Ion Channels

From structure prediction to structure classification

By

BRONCIO AGUILAR SANJUAN

School of Mathematics
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Sciences.

8TH OF JULY 2019

Abstract

The range of conformational states that determine the function of a protein is often unknown. This is the case for most proteins. In particular, for membrane proteins, this is more challenging, as their structures are often difficult to resolve. At present, the computational prediction of membrane protein structures has become a powerful means to overcome this limitation. However, determination of the meaningful conformations that we can relate to observable function is unclear overall. This research aims to provide a modelling framework to determine those conformational states that are relevant to ion conduction for peptide-assembled ion channels. We focus on the particular case of *cWza Cysteine-mutant* ion channels. These are *synthetic* ion channels made from the parallel symmetric assembly of eight α-helical peptides with identical amino-acid sequences. Their peptide sequences (cWza Cysteine-mutant) are redesigned versions of the outer-membrane sequence of the *Wza* complex; the polysaccharide exporter in *E. coli*. Experiments measuring their conductive activity showed that these channels can either show a voltage-independent single-conductance or a voltage-dependent dual-conductance depending on their peptide sequence. A hypothesis is that a transition between two conformation states takes place for those channels with dual-conductance. While for those channels showing a single conductance, a single conformational state should only exist. However, three-dimensional crystal structures of these channels are unavailable.

Based on a combination of techniques, we employed *symmetric peptide-peptide docking*, for prediction of atomic three-dimensional structures of all cWza Cysteine-mutant ion channels; a proposed method for *conformational classification* of structures, via a description of their most probable inner Van der Waals pore dimensions; and *molecular dynamics* simulations, to test the robustness of conformational classifications. Overall, we found that two groups of conformations were found for those modelled channels showing dual-conductance in experiments, while for the single-conductance channels, a single conformational group was found. Thus, our results show that observed changes in conductance of synthetic ion channels can indeed be explained by changes in conformation. While when a single conductance is observed, a single conformation exists. We also found that inter-chain atomic interactions such as Hydrogen-bonds and Knobs-Into-Holes played a role in the specificity of conformations and the evolution of their simulated dynamics. Further research needs to verify the validity of our procedure to other synthetic peptide-assembled ion channels that could be a target for *de novo* design or redesign.

Additionally, we developed a molecular dynamics procedure to study with atomic-resolution the ion permeation process across peptide-assembled ion channels; embedded in a lipid-bilayer model and under applied voltage conditions. Details of the ion permeation mechanisms for individual channels are often difficult to resolve via experiments and computation. Application of our procedure to three different types of peptide-assembled ion channels revealed details of their very different ion permeation mechanisms, previously unknown. More time-intense simulations are needed to test the persistence of the identified permeation mechanisms.

Dedication and acknowledgements

It would not have been possible to write this doctoral thesis without the help and support of many persons. I would like to use this opportunity to thank them all. First and foremost, I would like to express my most sincere gratitude and appreciation to my supervisor, Prof. Tanniemola B. Liverpool, for admitting me to undertake the research work leading to this dissertation. It is through his guidance, advice and stimulating encouragements that this work could be accomplished.

I am also deeply indebted to my co-advisors, Prof. Dek Woolfson for their continuing help and support during my PhD. We have had many fruitful discussions about glasses and I learned a lot from them. Their constructive ideas and suggestions have certainly contributed a lot to the work.

I am thankful to Dr Sarah Harris from the University of Leeds, for accepting to be the examiner of my PhD thesis. I am also thankful to Dr John Russo of the University of Bristol, for accepting to be my co-examiner.

I am grateful to all my colleagues and members of the Fluids and Materials group of the School of Mathematics and also to the Woolfson group of the School of Chemistry for their support to my work and countless meeting from which I got invaluable advice. In particular, I would like to thank Dr Majid Mosayebi (School of Mathematics), Dr Fabio Parmeggiani (School of Biochemistry), and Dr Richard Sessions (School of Biochemistry).

I am also infinitely grateful to the National Council for Science and Technology (CONACyT) of Mexico, for funding the full extent of my research programme.

Finally, I want to dedicate this work to my many friends and family who kept me company, helped me and taught me so much throughout these years. In particular, to my father Agripino and my brother Miguel, who believed in my effort and celebrated my achievements. Also, to the two women who I have loved the most, my mother Herlinda, who always gave me her unconditional love, and to the love of my life Blanca, who although far from me, it has not stopped her from giving happiness to my days, making me a better person, and giving me her sincere love and company.

Finalmente, quiero dedicar este trabajo a mis muchos amigos y familia que me acompañaron, ayudaron, y eseñaron tanto durante estos años. Especialmente a mi padre Agripino y mi hermano Miguel, quienes han confiado en mi esfuerzo y celebrado mis logros. También, a las dos mujeres que mas he amado, mi madre Herlida, quien siempre me dio su amor incondicional, y al amor de mi vida Blanca, que aunque lejos de mi, no ha impedido que haya dado alegria a mis dias, me haya hecho una mejor persona y me haya dado su amor sincero y compañia.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE: .....

TABLE OF CONTENTS

LIST OF TABLES

TABLE **Page**

List of Figures

Figure **Page**

CHAPTER 1

WHAT IS IT ABOUT?

Alternate conformations are a key structural feature for the function of all proteins as a result of their flexibility. In transmembrane ion channels, conformational changes modulate their transport properties affecting the passage and/or blockage of solutes across the cell membrane. cWza Cys*-mutant ion channels are a family of synthetic α-helical peptide barrels proven to assemble and span lipid bilayer. These display electric activity with transitions between low and high conductance, as a result of suspected conformational changes. But, when Tyrosines (*Tyr*) in all peptide are swapped with Cysteines (*Cys*), a single stable structure seems to take place, as a single conductance is observed. Prediction of conformations of cWza* Cys*-mutant ion channels before peptide assembly is challenging, given the absence of structural information; often difficult to resolve for transmembrane proteins. Next, will introduce further the context of this problem and also how modelling, which is the main essence of our work, can provide an alternative to deal with the lack of structural information for prediction of conformations.*

1.1 Introduction

1.1.1 Transmembrane Ion Channels And Multiple Conductance

Proteins are organic molecules made of the assembly of *single chains* (units) of variable length made from the combination of 20 different types of *essential amino acids*. These specific amino-acids sequences will determine whether a protein can assemble either in an aqueous environment or a hydrophobic one; like the cell membrane. As a result, proteins will perform various tasks in living organisms, such as providing flexibility to tissues, catalysing chemical reactions, and

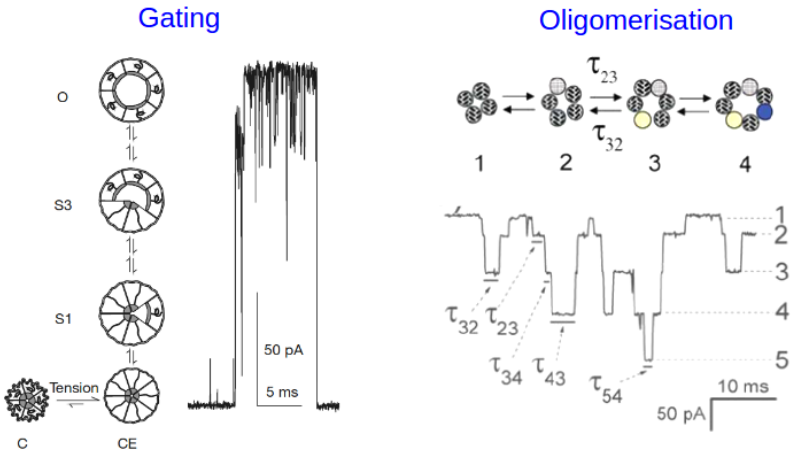


Figure 1.1: Recorded electric activity from electrophysiology experiments for *mechanosensitive channel* (Msc); *left*, and pore-forming activity of *alamethicin* peptides; *right*. These channels are representative examples that illustrate *gating* and *oligomerisation* as conductance changing mechanisms in transmembrane ion channels; for the first, a channel opens progressively in response to applied membrane tension, whereas for the second, peptides associate/dissociate dynamically to form channel pores of variable dimensions.

as transporters of substances for the self-maintenance of the cell. In particular, *transmembrane proteins* embedded in the cell membrane are responsible for the permeation of ions and charged biomolecules via diverse transport mechanisms [14].

Ion channels are a special class of transmembrane proteins that form single open pores spanning the membrane with dimensions ranging from 0.1 to 10 *nm* of diameter. Transport across ion channels generally occurs at no extra energy input as they form single pores in the membrane (passive transport), and thus their conductance will be a direct consequence of the pore dimensions and geometry, limiting the internal volume available for permeation. However, as revealed by *electrophysiology* measurements, for some ion channels their electric current activity may show a step-like dynamic variation - as shown in Fig., 1.1, characterised by sudden, random switching events between fixed, *multiple conductance* values, with durations ranging from seconds to minutes [41, 64].

To explain the appearance of multiple conductances for ion channels, *gating* and *oligomerisation* are commonly invoked mechanisms consistent with the premise that the conductance of a channel is the direct consequence of its structure. Thus, *gating* explains changes in conductance by changes in the *conformation* or geometry of the channel triggered by some external stimulus, such as pressure or electric potential; examples of this are the mechanosensitive channel (Msc) - responsible for pressure sensing in *E. coli* bacteria [80]; Fig., 1.1, and the potassium channel - responsible for transmitting electric impulses across neurons [127]. By contrast, *oligomerisation* explains conductance changes on the basis of changes in the number of chains or units that assemble forming a channel, leading to an expansion or reduction of its internal diameter; a

well-known example is the class of alamethicin channels - antibiotics produced by *Trichoderma viride* fungus [40]; Fig., 1.1.

1.1.2 Redesigning Transmembrane Proteins: The cWza Cys-Mutant Ion Channels

Recently, *Synthetic Biology* has become a growing branch of Biology with the general aim of improving or innovating functionalities of natural biological systems. Not surprisingly, ion channels have become a target for design and/or redesign for novel applications such as biosensing, filtering, and other potential ones. Some progress has been made in this direction inspired by mutation experiments of ion channels, where one or various individual amino acids in the protein original sequences are replaced with others to induce modifications to their original transport functionality, whilst preserving their pore formation ability in a membrane environment[53, 64, 72].

The cWza *Cys*-mutant sequences (Figure 1.2), is a set of engineered sequences obtained from punctual Cysteine-mutations of a base sequence called cWza; derived from mutation of the sequence of the outer-membrane domain of Wza (PDB *id*: 2J58), the natural sugar export machinery in *E. coli* [30]. When synthesised, these sequences lead to short α-helical *peptides* able to assemble in parallel to form ion channels in a lipid-bilayer, under the presence of an applied voltage difference ($\Delta V = +100mV$) between opposite sides of the bilayer. After being assembled, these channels are capable of transporting ions across the bilayer under different applied voltage conditions, as indicated by recordings of their conductive activity [64].

Recorded electric currents reveals that different peptide sequences will form pores showing three different types of conductive behaviour, hence, *three sequence types* can be categorically distinguished; type 0, 1, and 2. The *zeroth type* will correspond to those sequences for which one single negatively charged amino-acid in the cWza-base sequence (either Glutamate (E) or Aspartate (D)) has been mutated with a Cysteine (C), such as cWza-E369C and cWza-D366C; whose corresponding assembled pores will only display a *noisy electric activity* ($\Delta V = +100mV$). Next, the *first type* will correspond to the only sequence for which Tyrosine (Y, residue 373) has been mutated with a Cysteine (C), called cWza-Y373C, whose synthesised peptides will form pores showing a single steady conductance value regardless of the applied voltage ($|\Delta V| \leq 200mV$); i.e., *ohmic conductance*. Finally, by contrast, the *second type* will include sequences where Tyrosine has not been mutated while preserving negatively charged amino-acids, that is cWza, cWza-K375C, and cWza-S355C; whose corresponding pores will consistently display *dual conductance* at high applied voltage ($\Delta V = +200mV$), with the lifetime of each conductive state varying from seconds to minutes; which is consistent with a multiple conductive behaviour [64]. See summary in Fig., 1.2.

Ecoli-wzaD4	Ac-	APLVRWNRVISQLVPTISGVHDMTETVRYIKRWPN	-NH2
cWza (base sequence)	Ac-	APLVRWNRVISQLVPTITGVHDLTETVRYIKTWPN	-NH2
cWza-K375C	Ac-	APLVRWNRVISQLVPTITGVHDLTETVRYICTWPN	-NH2
cWza-S355C	Ac-	APLVRWNRVICQLVPTITGVHDLTETVRYIKTWPN	-NH2
cWza-Y373C	Ac-	APLVRWNRVISQLVPTITGVHDLTETVRCIKTWPN	-NH2

	Peptide used for pore formation	Low conductance,L (nS) Mean (± sd)	High conductance, H (nS). Mean (± sd)
Type 2	cWza	0.38 ± 0.020	0.97 ± 0.06
	cWza-K375C	0.39 ± 0.015	0.98 ± 0.06
	cWza-S355C	0.37 ± 0.020	0.97 ± 0.07
Type 0	cWza-E369C	Noisy channel	Not observed
	cWza-D366C	Noisy channel	Not observed
Type 1	cWza-Y373C	Not observed	0.75 ± 0.04

Figure 1.2: *Top*: Set of engineered cWza-based sequences [64]. cWza serves as a base sequence to carry out mutations of specific amino-acids with Cysteines (C, blue). cWza is obtained from mutations (red) of the natural peptide sequence of E-coli Wza-D4, whose peptides can naturally assemble in a parallel, regular arrangement with 8 units (*octamer*). *Bottom*: Classification of electric conductance activity for ion channels with cWza-based sequences.

1.1.3 Explaining the Conductance of cWza *Cys*-Mutant Ion Channels via Computation

As discussed above, it is clear that the location of single *Cysteine* mutations along the cWza-base sequence will have diverse effects on the electric activity of assembled channels; either displaying a *noisy conductive activity* (sequence type 0), *ohmic conductance* (sequence type 1) or a voltage-dependent *dual conductance* (sequence type 2). Thus, under the assumption that conductance will be the direct consequence of a channel's structure, we can infer that for sequence types 1 and 2, respectively, different conductive activities will be a consequence of having either a *single steady conformation* or *two alternate steady conformations*. As for sequence-type-3 channels, we do not have a clear interpretation; noisy electric activity has been suggested previously as indicative of an unstably assembled channel though [64]. Thus, we will focus on sequence-type-1 and -2 channels only and we will aim to explain the conductive behaviour their corresponding ion channels.

In particular, the appearance of dual conductance suggests that the type of pore assembly process and its dynamics may be driven either by *conformational changes*, consistent with a gating conductive mechanism. Experiments suggest that the number of assembling peptides is eight, for all cWza *Cys*-mutant ion channels; see blocking experimental assays in Reference [64]. Consequently, to change their conductance channels either may *expand*/*contract* in response

to the applied voltage. However, the formation process of these assemblies and their *atomic structures* cannot be revealed by the current experimental data from electrophysiology alone. Also, due to the need of a suitable hydrophobic environment for these peptides to assemble - as often expected for transmembrane proteins, it seems challenging to determine their precise atomic structure or their dynamic behaviour via standard experimental techniques such as X-ray crystallography or NMR (Nuclear Magnetic Resonance), respectively; which often require stringent experimental conditions too [68].

As an alternative to experimental means, modelling techniques such as *docking* and *molecular dynamics* aim to overcome their limitations in a complementary fashion [63]. *Docking* is recurrently used as an alternative to experiments like X-ray crystallography to model protein atomic structures with variable strength of association (binding affinity) between their constitutive units (amino-acid chains)[29, 124]. Although one of the main caveats of docking is its inability to study protein dynamics, *Molecular dynamics* (MD) is recurrently used to simulate the structure evolution of proteins via solving Newton's equations of motion, provided an initial protein structure, for time-scales not easily achievable in experiments like NMR [2, 100].

The main aim of our research is to reveal via computation whether alternate conformations are the mechanism explaining observed changes in conductance for some of the cWza Cys*-mutant channels. While for cWza-Y373C channels, whether a single conformation can only exist*. Thus, before introducing our main findings presented in this work, I will first introduce an overview of the general methods and concepts (Chapter 2) that make the essence of *Docking*, for structure prediction of membrane proteins; and *Molecular Dynamics*, to perform dynamic simulations with atomic detail. Next, in Chapter 3, I will present the results from the application of docking for modelling of candidate three-dimensional structures for all cWza *Cys*-mutant channels (Figure 1.2) and their classification according to their internal pore dimensions, from which groups of structures with similar conformations are identified from docking, and after Molecular Dynamics simulations of these predicted structures. In Chapter 4, we will introduce a Molecular Dynamics procedure that can allow us to study ion permeation in atomic-resolution, which aims to complement the study of peptide-assembled ion channels by providing insight into their ion permeation mechanisms. The main Conclusions and Outlook of our work constitute part of Chapter 5. A series of Appendices providing in-depth detail from various results of our work is featured at the end.

CHAPTER 2

THEORY AND METHODS

For conformation prediction, we first require a mean for modelling of plausible structures of the cWza Cys*-mutant ion channels per peptide sequence. Membrane-Protein Symmetric Docking will be the core technique allowing us to identify preferential cyclic peptide arrangements according to their estimated assembly energies; used later on as candidate structures for conformation identification analysis (*Chapter 3*). Here, we will briefly describe the general concepts and methods used for the implementation of docking. Additionally, we will give an overview of the concepts and methods employed for the setup of Atomistic Molecular Dynamics simulations; that we will employ later to test resulting docked models in terms of their stability under equilibrium conditions and for channel conductance calculation assays under the influence of an applied electric field; as a potential way to study conductance across docked models in a dynamic scenario (*Chapter 4*).*

2.1 Symmetric Docking of Membrane Proteins

2.1.1 Structure Prediction of Peptide-Assembles via Docking

As discussed in Subsection 1.1, trans-membrane proteins are crucial actors in a wide range of biological processes. However, understanding of their functioning is limited due to the lack of detailed structural information. Thus, computational techniques have become a powerful approach to *structure prediction*, which deals with the task of modelling the three-dimensional structures of a protein provided its amino acid sequence (a.k.a, *primary structure*).

At present, a variety of techniques and tools have become available to deal with the intricate task of predicting the structure of proteins according to their different levels of organisation,

ranging from the structure of local *folds* or segments (*secondary structure*) and their combination (*tertiary structure*) in a single protein chain to the resulting *complexes* or three-dimensional arrangements of several assembling units (*quaternary structure*) like *peptide* and/or *polypeptide* chains [14].

Consequently, different techniques for structure prediction have been developed and that are recurrently employed in combination for the resolution of transmembrane protein structures, which include (1) *ab initio* or *de novo* modelling, via physical principles for prediction of new folds from energy global optimisation of a suitable energy function in combination with Monte Carlo methods [74–78, 118]; *homology* or *template-based* modelling, where single or fold combinations are built according to sequence similarity with respect to segments from another structure, which has been experimentally determined [5, 20, 23, 50, 57, 70, 89, 103, 104]; and *docking* modelling, used for prediction of protein complexes via conformational search and energy scoring for identification of preferential protein-protein associations with units assumed either rigid or flexible [3, 5, 8, 22, 29, 35, 37, 69, 124]. However, application of these may be limited due to structure availability - usually the case for many membrane proteins (*homology*), high computational demand due to conformational sampling, and the unsuitability of scoring functions (*ab initio* and *docking*).

Ion channels often assemble forming symmetric complexes. Known structures include potassium channels (*KcsA* and *Kv*), mechanosensitive channels (*MscL* and *MscS*), α-hemolysing, outer-membrane porins (*OmpF* and *OmpC*), amongst others; see Figure 4 in Reference [63]. Particularly, the *Wza* complex also exhibits a cyclic symmetry for the outer-membrane D4 domain in its crystal structure [30, 64, 72]. Thus, we will focus on the modeling of complexes with cWza *Cys*-mutant sequences (Subsection 1.1), under the assumption of cyclic symmetry.

2.1.2 Protein Complexes and Symmetry

Symmetry is a recurrent pattern of assembly in both soluble and membrane proteins in nature, with complexes made of repeated and regularly arranged single units [34]. It has been argued that symmetry plays a key role in promoting stability and conformational changes, often associated with function [29, 92, 123].

Overall, relevant types of symmetry for a wide range of protein structures and crystallography include *i)* point symmetry, *ii)* helical symmetry, and *iii)* crystal symmetry. In particular, *point symmetries* encompass five essential sub-types: *i)* cyclic (C); e.g., transmembrane pores and chambers; *ii)* dihedral (D); e.g., clathrin cage (PDB *id*: 1xi4); and high-order symmetries *iii)* tetrahedral (T), octahedral (O), and icosahedral (I); e.g., soluble nano-cages like ferretin (PDB *id*: 1lb3) and viruses (PDB *id*: 1stm). Figure 2.1 illustrates these symmetries.

A more thorough discussion on the definitions and natural observations of the above symmetry types and sub-types can be found in References [8, 29, 34]. However, here we focus on the use of cyclic symmetry applied to docking. As this will be the only symmetry type relevant to the

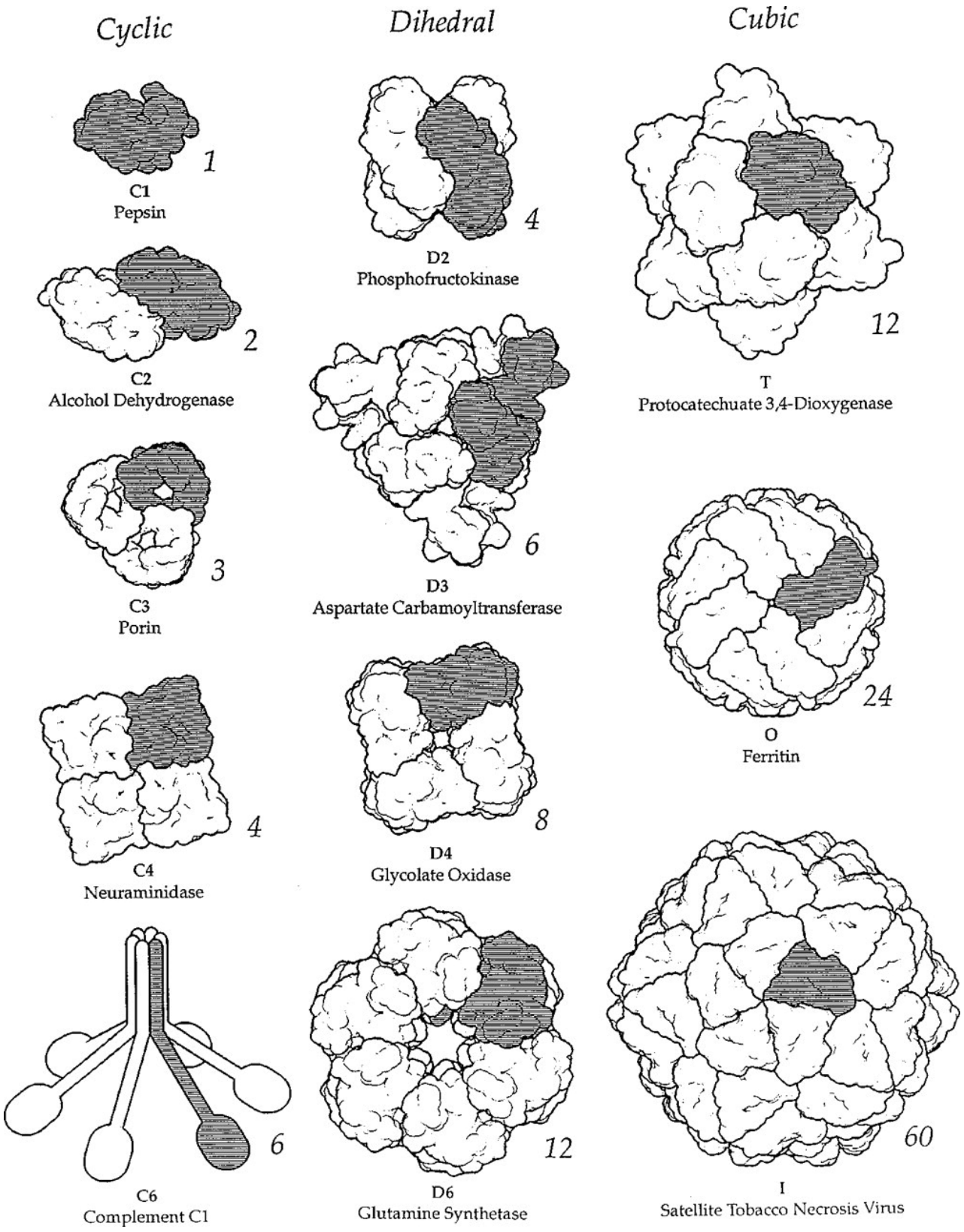


Figure 2.1: Examples of natural protein complexes illustrating the different types of *point symmetries*. Figures are taken from Reference [34].

peptide-assembled channels studied here; the cWza *Cys*-mutant ion channels.

Symmetry, in general, is an important and useful constraint for structure prediction. In particular, peptide-peptide docking takes advantage of this. Symmetry allows simplifying energy estimation (*scoring*) of a structure since interactions between assembling units will be redundant [29]. Also, the time-intensive task of conformational search, for backbone units and side-chain conformations, gets reduced, as atomic-coordinates of all assembling units can be obtained under symmetry operations applied to a single unit [8]. Consequently, symmetry enhances the efficiency of side-chain optimisation and energy minimisation tasks that lead to output structures via docking [29].

Next, we will provide further details on *conformation sampling* and *energy scoring*, which make up the two main components of docking, as implemented in *Rosetta*. Rosetta is a widely used software suite that features algorithms for computational modelling and analysis of protein structures [3].

2.1.3 Conformational Sampling for Symmetric Membrane Protein Complexes

For a protein complex, the degrees of freedom (*dof*) of its conformation comprehend: *i)* the set of *torsion angles* of backbone and side-chain atoms, and *ii)* the *rigid-body transformations* (translation and rotations) between assembling units. Sampling the range of values for all of these *dof*, for all assembling units, is challenging. However, symmetry simplifies this task.

According to symmetry, the *torsion angles* of all assembling units should be identical. Then, sampling gets reduced to only describing the *internal dof* of a single unit; in Rosetta terminology, the *master unit*. As a result, torsion angles for all other units are set according to this single unit. But, the identity of the master unit requires careful selection, as this will impact energy calculations [29]. Thus, for a peptide-assembled ion channel, a unique peptide will play the role of the master unit.

Similarly, symmetry reduces the sampling of *rigid-body dof*. Positions/orientations of all units will be set from a single unit via translations/rotations consistent with the imposed symmetry. Definition of an *intrinsic* reference frame for the protein complex is required though [35, 59]. For instance, for an ion channel made from the cyclic arrangement of n peptides (C_n-symmetry), coordinates of all units around a common axis, will be obtained via up to $n-1$ consecutive $\frac{2\pi}{n}$ rotations of the coordinates of a single unit, i.e., the *master unit*. Thus, owing to symmetry again, sampling of the *dof* of a single unit is only required.

2.1.4 Energy Scoring of Symmetric Membrane Protein Complexes

Energy evaluation in full-atomic resolution is the most time-consuming task in docking [8, 29]. Functions for evaluation of a protein structure usually feature one- to two-body distant-dependent/independent energy terms and whole-structure energy terms [59]. For membrane

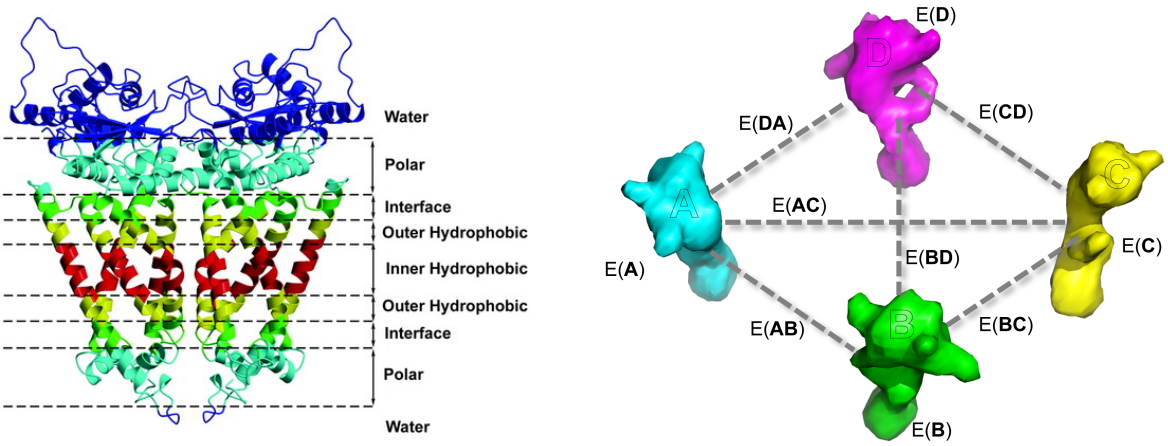


Figure 2.2: *Left*: Illustration of the layered-membrane model used for low-resolution energy evaluation during docking; Figure is taken from [126]. *Right* Schematic representation of the energy evaluation process between units for a protein complex made of four units assembled with C_4-symmetry; Figure taken from [29]

protein complexes, energy terms are featured according to the representation of the membrane environment [3]. Thus, typically low and high-resolution functions are employed. At low-resolution [126], the membrane is represented by nine discrete layers describing its hydrophobic interior/exterior regions, its polar regions, the interface between these, and the solvent regions on both sides of the membrane; see Figure 2.2 (*Right*). At high-resolution [58], the membrane is represented as a continuous dielectric medium, from the hydrophobic interior to the solvent exterior, according to an implicit membrane model (IMM1). A summary of the energy terms for low-/high-resolution score functions employed by Rosetta can be found in Reference [3] and its supplement.

Symmetry can considerably increase the efficiency of the *total energy* evaluation of protein complexes, as both *internal* (one unit alone) and *interface* (between units) energy terms become duplicated. Consider the example of a four-unit complex, with its units called, are A, B, C, and D, which assemble according to a C_4-symmetry. As illustrated in Figure 2.2 (*Right*), under symmetry, internal interactions for A are identical to those of the remaining three. Similarly, interactions between A and B are identical to B-C, C-D, and D-A, and AC interactions appear two times. Then, if we ignore whole-structure energies, the total energy E of the complex will be given by

$$\begin{aligned} E &= E_{internal} + E_{interface} \\ &= [E(A)+E(B)+E(C)+E(D)] + [E(AB)+E(BC)+E(CD)+E(DA)+E(AC)+E(BD)] \\ &= 4E(A) + 4E(AB) + 2E(AC) \end{aligned}$$

where, $E(X)$ refers to the internal energy of unit X, while $E(XY)$ will refer to the interaction energy between X and Y units. Thus, only three terms out of ten are required for energy

evaluation.

2.1.5 Sidechain Optimisation and Energy Minimisation

Sidechain optimisation and Energy minimisation are two crucial tasks involved in determining the sidechain and backbone atomic-coordinates of predicted structures via docking. These take advantage of symmetry to enhance the performance of their computation. The *sidechain optimisation* or *sidechain packing* refers to the problem of assigning the individual conformations or *rotameric states* [84] of all sidechains present in a protein, such that their contribution to the total energy of the protein is minimised. For an asymmetric protein complex, this problem is known to be NP-Complete [81]; there is not any deterministic algorithm that can solve the problem efficiently. So, a stochastic approach is typically employed [54]. However, as similarly described above (Figure 2.2, *Right*), symmetry simplifies the computation of internal and interacting sidechain energies. For a more detailed description featuring the explicit energy terms, see Reference [29].

Energy minimisation will deal with the task of minimising the contribution to the total energy due to rigid-body degrees of freedom (*dof*), discussed above. Thus, derivatives of the high-resolution score function are computed for the corresponding degrees of freedom. However, this process will become simplified for a symmetric complex. As derivatives will be done only for the degrees of freedom of the chosen *master unit*. For instance, for a complex made of n units with cyclic C_n-symmetry, the partial derivative for a degree of freedom, say x_i, will be written in terms of the master unit's copy of that dof, say x_i^c, then we have that

$$\frac{\partial E}{\partial x_i} = \sum_{k=1}^{n} \left(\mathbf{R}_k \frac{\partial E}{\partial x_i^k} \right) = n \frac{\partial E}{\partial x_i^c}$$

where x_i^k is the $k-th$-unit's symmetric copy of x_i, which is related to the master unit's dof, by the transformation $x_i^c = \mathbf{R}^k x_i^k$

2.1.6 Procedure for Symmetric Docking Simulation

So far, the core elements that comprehend the essence of docking have been introduced. Besides, we have seen how the imposition of symmetry, in particular, for the case of cyclic symmetry, impacts the formulation of energy functions and conformational search. And consequently, the optimisation tasks of sidechain packing and rigid-body energy minimisation become simplified. Next, we will provide an overview of the assembly protocol followed by docking simulations to predict symmetric protein assemblies out of using a single unit; the master unit.

In the initial phase of the complex assembly, a random configuration of the rigid-body degrees of freedom is taken as the starting point but respecting the imposed symmetry. Units are placed apart so that atomic contacts are avoided between them. Rotational degrees of freedom are taken randomly too. Next, the assembling units are translated along their allowed degrees of

Figure 2.3: Schematic representation of how docking explores the energy landscape during the various phases of the peptide-peptide assembly process. The figure is taken form [35]

freedom to establish contacts between the units. For instance, for a cyclic assembly, units are only allowed to move radially in the membrane plane around a common axis perpendicular to the membrane. Once contacts have been established, the energy of the assembly is minimised via a rigid-body Monte Carlo search, but under a low-resolution representation. This assumes the layered-membrane model described above, while the protein is represented by its heaviest backbone atoms (N, C_α, C, O) and its centre of geometry (*centroid*); taken as a "*pseudo-atom*". Once finished, this process is followed by a high-resolution search phase, assuming the high-resolution score function and an all-atom representation of the protein complex. Thus, sidechains are added and the energy of the structure is optimised using a Monte Carlo minimisation procedure. This consists of several rounds of rigid-body moves (*perturbations*) followed by symmetric sidechain optimisation [54]. All moves are accepted/rejected according to a standard Metropolis criterion [4]. Thus, the energy is minimised via a gradient-based algorithm [8, 29, 35]. Figure 2.3 illustrates how the energy of the docked complex is minimised throughout all these phases. The final energy-minimised structure of this process will be the output of docking.

2.2 Molecular Dynamics

In general, *molecular dynamics* is a recurrent simulation technique that by solving Newton's equations of motion aims to resolve the temporal evolution of a system of many particles; as it is the case for many biomolecular systems, such as lipid membranes [31, 83, 90, 128], enzymes [1, 60, 71, 113, 116], and protein assemblies [63]. It is worth noticing that *simulations* can only show how a system behaves according to approximations to the laws governing the actual system, but cannot be solely used instead of actual experiments in the real world. Thus, experiments

combined with computer modelling and theory can ideally help us to better understand the system in question.

In practice, it is usually distinguished between two forms of MD, all-atom (or atomistic) and coarse-grained. The former requires an atomistic approximation of particles and interactions, whereas the latter adopts effective particle representations and potentials. As a result, computational time and cost limit capabilities to simulate particular length and time scales; with atomistic MD restricted to lengths starting from 10^{-10}m (or Å) and times up to $10^{-}6$m (or μs), and coarse-grained MD length and times scales, both up to$10^{-}6$m (or μs) and $10^{-}3$m (or ms), respectively . Atomistic MD nonetheless is a recurrent approach to first study key mechanisms concerning ion conduction across protein pores that require atomic precision, such as conformational changes of the protein, like in the potassium channel, and protonation-deprotonation of channel amino-acids and polarisability of lipid membrane molecules [63].

2.2.1 Dynamics of biomolecules

Although bio-molecules are composed of atoms, which at the same time have nuclei and electrons, almost all of the mass of an atom is concentrated in its nucleus, that in consequence determines the position of the whole atom. Thus, as justified by the *Born-Oppenheimer approximation* [18], electronic degrees of freedom can be absorbed in an effective potential so that the dynamics of atoms can be reduced to a classical one, i.e. via Newton's equations of motion

$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i \tag{2.1}$$

where $\mathbf{F}_i$ denotes the *net force* acting upon the *i-th* particle with mass m_i, causing an acceleration $\ddot{\mathbf{r}}_i$ of this, at position $\mathbf{r}_i$, within a collection of N particles.

When the force $\mathbf{F}_i$ is considered to be conservative, this can be derived from a potential, U,

$$\mathbf{F}_i = -\nabla_{\mathbf{r_i}} U(\mathbf{r}_1, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_N) \tag{2.2}$$

which is the case for common inter-atomic interactions in bio-molecules such as, bonded, non-bonded, and long-range forces; thoroughly described in Subsection 2.2.2.

Thus, from the above, it is clear that in order to integrate Eqs., 2.1, and hence to solve the the trajectories of individual atoms, two key ingredients are needed to : 1) a *force field*, that is, a parametrisation of the potential $U(\mathbf{r}_1, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_N)$ and 2) a numeric *integration scheme*. Next, we discuss these two in further detail, alongside other algorithms needed in order to guarantee a physically reasonable solution to the equations of motion in a bio-molecular system.

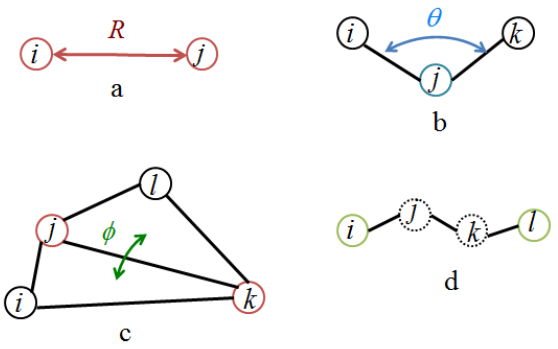


Figure 2.4: *Bonded interactions*: a) bond stretching, b) bond torsion, c) angle between connected bond planes (dihedral angle), d) angle between non-connected but bonded planes (improper dihedral angle).

2.2.2 Force fields of atomic interaction

In general, a *classical force field* refers to the form of the *potential function* $U(\mathbf{r}_i)$ and the respective set of parameters for the different atom types involved in a system. For most all-atom MD simulations of bio-molecules $U(\mathbf{r}_i)$ is usually decomposed as the sum of $U_{non-bonded}$ and U_{bonded} functions approximating non-bonded and bonded interactions between atoms,

$$U(\mathbf{r}_i) = U_{non-bonded} + U_{bonded} \tag{2.3}$$

Overall, $U(\mathbf{r}_i)$ is a many-body potential function, that is, it encompasses interactions between two to N bodies. On the one hand, *non-bonded interactions* are pairwise-additive two-body interactions that can be between neutral and charged atoms. On the other hand, *bonded interactions* typically include selected interactions between two, three, and four bonded atoms that correspond to different bond bending movements, Figure 2.4.

Non-bonded potentials For both neutral and charged atoms, different interactions take place depending on their inter-atomic distance, $r = |\mathbf{r}_i - \mathbf{r}_i|$. Thus, the form of the potential of interaction function of non-bonded atoms is typically approximated as a combination of a short-range fast-decaying component and long-range slowly-decaying one.

At short inter-atomic distances, although here atoms are assumed to be classic particles, i.e., with simultaneously known localised positions and momenta, quantum mechanical effects lead to a dominant atomic repulsion as a consequence of Pauli's exclusion principle. In contrast, at long-range inter-atomic distances, electrostatic interactions dominate in the form of a *dipole-dipole potential*, and whenever atoms carry net charge a *Coulomb potential* is added. Hence, the functional form of non-bonded potential features first the interaction between neutral atoms i and j, approximated by a Lennard-Jones potential,

$$V_{LJ}(\mathbf{r}_i,\mathbf{r}_j) = 4\epsilon_{ij}\left(\sigma_{ij}^{12}/r_{ij}^{12} - \sigma_{ij}^{6}/r_{ij}^{6}\right) \tag{2.4}$$

where σ_{ij} is the distance at which $V_{LJ} = 0$ and ϵ_{ij} is the energy depth of the potential well at the equilibrium distance $2^{1/6}\sigma_{ij}$. It is important to mention that σ_{ij} and ϵ_{ij} depend on the interacting atom types, but if parameters for atoms of the same type are known, empirical combination rules can be used to determine the cross parameters

$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}) \quad \text{and} \quad \epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}$$

Additionally, a Coulomb potential is included, with atoms assigned partial charges q_i and q_j according to the parametrisation of the force field, which is a crucial step

$$V_C(\mathbf{r}_i,\mathbf{r}_j) = -\frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}} \tag{2.5}$$

where, ε_0 is the vacuum permeability constant.

Ultimately, the total non-bonded potential will be taken as

$$U_{non-bonded} = \sum_{i<j}^{N}\left[V_{LJ}(\mathbf{r}_i,\mathbf{r}_j) + V_C(\mathbf{r}_i,\mathbf{r}_j)\right] \tag{2.6}$$

Bonded potentials Whenever two atoms are chemically bonded, the energy of this link can be approximated as a harmonic potential for a small deviation around the equilibrium bond length. Thus, since bonded atoms can be thought as mass points connected through springs, a whole molecule can be seen as a network encompassing nodes and links, and hence its potential energy is usually approximated by a set of selected two to four-body bonded interactions.

Two-atom bonded interaction: For a pair of bonded atoms (i,j), the energy associated with their bond stretching will be given by

$$V_b(r_{ij}) = \frac{1}{2}k_{b,ij}\left(r_{ij} - r_{0,ij}\right)^2 \tag{2.7}$$

where, r_{ij} is the relative distance between atoms as before, $r_{0,ij}$ is the equilibrium bond length, and $k_{b,ij}$ is the spring constant for the bond connecting the atom pair, as in Figure 2.4a.

Three-atom bonded interaction: In a similar fashion, whenever three atoms (i,j,k) are linearly connected as in Fig., 2.4b, the energy associated to their harmonic angular vibrations due to the bending movements between bonds (i,j) and (j,k) will be approximated by

$$V_\theta(\theta_{ijk}) = \frac{1}{2}k_{\theta,ijk}(\theta_{ijk} - \theta_{0,ijk})^2 \tag{2.8}$$

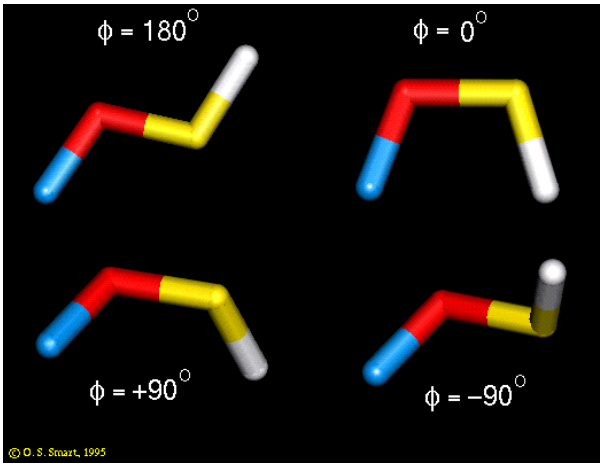


Figure 2.5: Various examples of dihedral angles (ϕ) adopted by a set of four linearly bonded atoms.

where the bending angle, θ_{ijk}, satisfies $cos(\theta_{ijk}) = \mathbf{r}_{ji} \cdot \mathbf{r}_{jk}/r_{ji}r_{jk}$, and similarly $\theta_{0,ijk}$ is its value at equilibrium, and $k_{\theta,ijk}$ the corresponding spring constant.

Four-atom bonded interaction: Finally, whenever four atoms (i,j,k,l) are consecutively bonded like in Figure 2.4c, the angle, ϕ_{ijkl}, between the two planes formed by atoms (i,j,k) and (j,k,l), is called the *dihedral angle*, often referred as *proper dihedral* or *torsion angle* too. Thus, ϕ_{ijkl} is usually taken from

$$cos(\phi_{ijkl}) = \mathbf{n}_{ijk} \cdot \mathbf{n}_{jkl}$$

where $\mathbf{n}_{ijk}$ and $\mathbf{n}_{jkl}$ are the unit normal vectors corresponding to planes (i,j,k) and (j,k,l), respectively defined by

$$\mathbf{n}_{ijk} = \frac{\mathbf{r}_{ji} \times \mathbf{r}_{jk}}{\mathbf{r}_{ji} \times \mathbf{r}_{jk}} \quad \text{and} \quad \mathbf{n}_{jkl} = \frac{\mathbf{r}_{kl} \times \mathbf{r}_{jk}}{\mathbf{r}_{kl} \times \mathbf{r}_{jk}}$$

so that conventionally, when $\phi_{ijkl} = 0$ degrees we say that i and l are in the *cis* conformation, and when $\phi_{ijkl} = 180$ degrees we say that i and l are in the *trans* conformation. See Figure 2.5.

Potentials for dihedral angles keep atomic groups like peptide bonds planar. Dihedral potentials play a crucial role in the local structure of biomolecules like proteins and nucleic acids, as energy constants of dihedral potentials are comparable to the energy of thermal motion k_BT and consequently these determine structural transitions. Although the origin of the dihedral potential is not completely well understood, repulsive interactions between overlapping bond orbitals and steric clashes [1] between atoms (such as C_1 and C_4 in butane) appear to be contributing factors. Thus overall, the standard functional form for representing a dihedral potential is given by

[1]Steric clashes are one of the artefacts prevalent in low-resolution structures and homology models. Steric clashes arise due to the unnatural overlap of any two non-bonding atoms in a protein structure

$$V_\phi = \sum_{i=1}^{3} \frac{V_n}{2} \left[1 + cos(n\phi - \gamma)\right]$$

which is a truncated Fourier series that allows accounting for complex angular variations of the potential energy, where V_n are the energy barriers associated to rotation; n the number of maxima in one full rotation; and γ a phase angle [6, 121]. Dihedral energy barriers are attributed to the exchange interaction of electronic clouds from adjacent bonds.

On the other hand, a separate class of angle called *improper dihedral* or *improper torsion*, ξ_{ijkl}, is defined in the same way as ϕ_{ijkl}, but whenever four bonded atoms can are part of a tetrahedral or a ring structure, like in Figures 2.4c and d). In this case, the potential for improper dihedral angles keep atomic groups like aromatic rings planar and also prevent molecules from flipping over to their mirror image, that is, it is used to select the correct *chirality* of atoms. The usual functional form of an improper angle potential is taken as harmonic

$$V_{id}(\xi_{ijkl}) = \frac{1}{2} k_{\xi,ijkl} \left(\xi_{ijkl} - \xi_{0,ijkl}\right)^2 \tag{2.9}$$

where, the featured parameters play the same role, as for other harmonic potentials above.

Overall, the bonded potential will be given by the addition of all the above bonded potential contributions from all atoms due to *bond stretching* (V_b), *bond bending* (V_θ), and both, proper (V_ϕ) and improper V_ξ torsion or *dihedral potentials*.

$$U_{bonded} = \sum_{i,j} V_b(r_{ij}) + \sum_{i,j,k} V_\theta(\theta_{ijk}) + \sum_{i,j,k,l} V_\phi(\phi_{ijkl}) + \sum_{i,j,k,l} V_\xi(\xi_{ijkl}) \tag{2.10}$$

Thus, the formula approximating the total potential (Equation 2.3) will be explicitly given by

$$U = \sum_{i<j}^{N} \left[V_{LJ}(\mathbf{r}_i, \mathbf{r}_j) + V_C(\mathbf{r}_i, \mathbf{r}_j)\right] + \sum_{i,j} V_b(r_{ij}) + \sum_{i,j,k} V_\theta(\theta_{ijk}) + \sum_{i,j,k,l} V_\phi(\phi_{ijkl}) + \sum_{i,j,k,l} V_\xi(\xi_{ijkl}) \tag{2.11}$$

Protein-lipid interactions The set of force-field parameters for protein-lipid interactions are usually determined from a combination of lipid force-field parameters; either from the CHARMM distribution [19, 62] or the Berger-parameters [15], and a choice of force-field employed for protein-solvent interactions; popular force-field families include GROMOS [111], AMBER [119], OPLS [48, 121], and CHARMM. Early studies proposed that an optimal combination of protein-lipid interactions is provided by the all-atom OPLS (OPLS-AA) and Berger force-fields. The Berger force-field is 'atom-united' - i.e., hydrogens and carbons in lipid tails are merged as pseudo-atoms, that results in faster calculations in contrast to CHARMM, which is all-atom. Despite being 'united-atom', the use of Berger lipids was previously found to lead to accurate hydration energies, reasonable reproduction of the strength of lipid-protein interactions [15, 106], and area-per-lipid values [122].

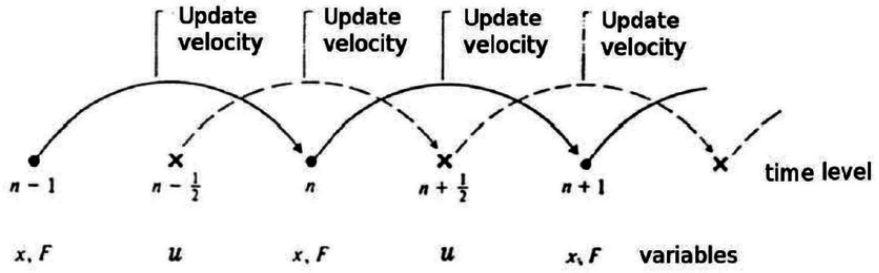


Figure 2.6: Schematic representation of the update rules of the leapfrog algorithm

Recent advancements have made it possible to offer a wider range of choice in terms of protein-lipid interaction parameters. These are provided within the above-mentioned force-field families; for instance, GROMOS54A7 [91], AMBER-Lipid16 [27], CHARMM36 [16]. See references [82, 90] for a critical review on these. However, there is not a clear cut to define the 'best' choice of force-field parameters, given the variety of physical properties tested for reproduction. At best, specific combinations should be simply avoided or considered with caution [90**?**].

In this work, we adopt the combination of OPLS-AA and Berger force-field parameters. The more contentious issue would be the choice of the lipid parameters, given its coarse-grained nature. However, this choice is justified by a number of considerations: simulation speed and storage limit; hundred replicas of (100 ns) MD simulations will be performed, reasonable reproduction of physical properties; as mentioned above, these do not show a substantial difference in contrast to simulations performed with an all-atom choice - i.e., CHARMM; atomic resolution of protein-assemblies is more relevant, as backbone and specific side-chain atomic interactions are assumed to play the main role in conformational stability during dynamics; finally, we get further encouragement by the previous simulation of a model of a cWza *Cys*-mutant structure, cWza-Y373C [64], in which this model was found to be stable after 100 ns of MD simulation, with OPLS-AA and Berger chose. Thus, these reasons overall make us believe that the choice of lipid force-field is sensible enough to observe the stability of modelled protein conformational groups, which are the main focus of our work.

2.2.3 Numeric integration

Integration of equations of motions After introducing the *force field* formulation (Equation 2.11), in order to resolve the trajectory of each individual atom in time, which is the collection of values of $\mathbf{r}_i$ for a given discrete of observational time values, $\{t_j; j = 1,\ldots,M\}$, the set of N-equations 2.2 are solved numerically via the *Verlet algorithm* [112],

$$\mathbf{r}(t+\Delta t) \approx 2\mathbf{r}(t) - \mathbf{r}(t-\Delta t) + \frac{\mathbf{F}}{m}\Delta t \tag{2.12}$$

This algorithm is a standard integration scheme that is simple but robust, as it has long-lasting numerical stability (unlike *primitive Euler* or *Runge-Kutta schemes*) and allows preservation of crucial physical properties, that is, inherent properties of a Hamiltonian system such as *i)* energy conservation, *ii)* time-reversal symmetry, and preservation of *iv)* volume of phase-space flow and *v)* the sum of areas of phase-space elements [24].

Derivation of this algorithm is easily obtained from the third-order Taylor expansions of $\mathbf{r}(t+\Delta t)$ and $\mathbf{r}(t-\Delta t)$, so that

$$\begin{aligned}\mathbf{r}(t+\Delta t) &= \mathbf{r}(t) + \frac{d\mathbf{r}(t)}{dt}\Delta t + \frac{1}{2}\frac{d^2\mathbf{r}(t)}{dt^2}\Delta t^2 + \frac{1}{6}\frac{d^3\mathbf{r}(t)}{dt^3}\Delta t^3 + \mathcal{O}(\Delta t^4)\\ \mathbf{r}(t-\Delta t) &= \mathbf{r}(t) - \frac{d\mathbf{r}(t)}{dt}\Delta t + \frac{1}{2}\frac{d^2\mathbf{r}(t)}{dt^2}\Delta t^2 - \frac{1}{6}\frac{d^3\mathbf{r}(t)}{dt^3}\Delta t^3 + \mathcal{O}(\Delta t^4)\end{aligned}$$

that when added together, it leads to

$$\mathbf{r}(t+\Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\Delta t) + \frac{d^2\mathbf{r}(t)}{dt^2}\Delta t^2 + \mathcal{O}(\Delta t^4) \tag{2.13}$$

where dependency on velocities, $\mathbf{v}(t) = \dot{\mathbf{r}}(t) = d\mathbf{r}(t)/dt$ is removed, solely depending on the acceleration $\mathbf{a}(t) = \ddot{\mathbf{r}}(t) = d^2\mathbf{r}(t)/dt^2$, which from Equation 2.2, is computed as

$$\mathbf{a}(t) = -\frac{1}{m}\nabla U \tag{2.14}$$

where U is approximated by Equation 2.11.

An important issue of the Verlet algorithm is that velocities are not generated. Since the kinetic energy of the system is computed as $K = 1/2\sum_{i=1}^{N} m_i\mathbf{v}_i^2$, hence velocities are needed to ultimately evaluate conservation of the *total energy* of the system, $E = K + U$, which is crucial to verify correctness of MD simulations. An intuitive solution to this issue might be to assume

$$\mathbf{v}(t) = \frac{1}{2\Delta t}[\mathbf{r}(t+\Delta t) - \mathbf{r}(t+\Delta t)]$$

However, the associated error is of order Δt^2, which can be an issue for accurate integration with temperature and/or pressure coupling. A variant of the Verlet algorithm, which can be proven to be equivalent to this, is the so-called *leap-frog* algorithm, which is implemented in GROMACS [42]; which leads to identical trajectories (Equation 2.13), but with the advantage that numerical errors in velocities are smaller [24, 65]. Leap-frog defines velocities at half-integer time-steps

$$\mathbf{v}(t+\frac{1}{2}\Delta t) = \frac{1}{\Delta t}[\mathbf{r}(t+\Delta t) - \mathbf{r}(t)] \quad \text{and} \quad \mathbf{v}(t-\frac{1}{2}\Delta t) = \frac{1}{\Delta t}[\mathbf{r}(t) - \mathbf{r}(t-\Delta t)]$$

From these, update rules for velocities and positions, are given by

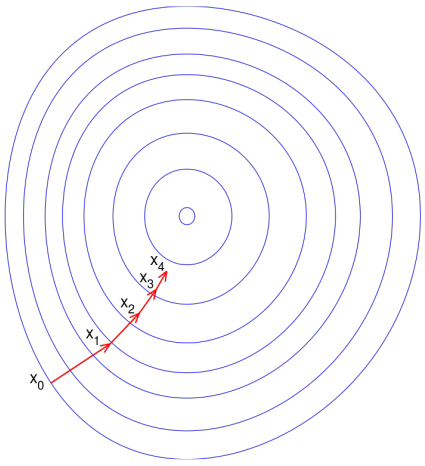


Figure 2.7: Schematic representation of the *gradient descent algorithm*, showing the update of a variable x towards the local minimum of a function F, on a series of level sets in the direction of the negative gradient, $-\nabla F$.

$$\mathbf{v}(t+\frac{1}{2}\Delta t) = \mathbf{v}(t-\frac{1}{2}\Delta t) + \frac{\mathbf{F}(t)}{m}\Delta t \tag{2.15}$$

$$\mathbf{r}(t+t) = \mathbf{r}(t) + \mathbf{v}(t+\frac{1}{2}\Delta t)\Delta t \tag{2.16}$$

Since leap-frog and Verlet are equivalent algorithms, all symplectic properties of the latter are inherent in the former.

2.2.4 Pre-MD preparation

Determination of a sensible starting configuration - the set of initial positions and velocities - is crucial to start any integration scheme (Equations 2.15) to solve equations of motions 2.1. Due to the singular nature of non-bonded potentials (i.e., if $r \to 0 \Longrightarrow |V| \to +\infty$), these can lead to initial prohibitively large forces in a starting configuration, consequently making integration fail due to numerical blow up. Thus, in practice, two of the most recurrent techniques to achieve the above are *Energy Minimization* and *Equilibration*, which we will discuss next.

Energy Minimization As an initial approach to remove large forces from a starting configuration of N particles, with given coordinates $\mathbf{r}_0$, defined as

$$\begin{aligned}\mathbf{r}_0 &= (\mathbf{r}_{1,0}, \mathbf{r}_{2,0}, \ldots, \mathbf{r}_{N,0}) \\ &= (x_{1,0}, y_{1,0}, z_{1,0}, \ldots, x_{N,0}, y_{N,0}, z_{N,0})\end{aligned}$$

one of the most robust and easy to implement algorithms is the *steepest descent* method, also called *gradient descent*. This attempts to drive the system to the nearest local minimum of the total potential energy $U(\mathbf{r}_1,\ldots,\mathbf{r}_N)$ within the neighbourhood of $\mathbf{r}_0$ in the direction of fastest decrease.

The algorithm is started by first taking all the forces defined as

$$\begin{aligned}\mathbf{F} &= (\mathbf{F}_1,\ldots,\mathbf{F}_N)\\ &= (F_{x,1},F_{y,1},F_{z,1},\ldots,F_{x,N},F_{y,N},F_{z,N})\end{aligned}$$

where $\mathbf{F}_i = -\nabla_{\mathbf{r}_i} U(\mathbf{r}_1,\ldots,\mathbf{r}_N)$ and the updating particles positions according to the following rule

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \frac{\mathbf{F}_n}{max(|\mathbf{F}_n|)} h_n \tag{2.17}$$

where $h_n > 0$ is the maximum displacement and $\mathbf{F}_n$ is the force at the *n-th* iteration, with $n > 0$. Thus, if h_n values are small enough, the total potential decreases monotonically with each iteration, that is, $U(\mathbf{r}_n) \geq U(\mathbf{r}_{n+1})$ for $n \geq 0$.

In practice, as implemented in GROMACS, h_n are selected and/or adjusted at each iteration in order to decrease $U(\mathbf{r}_n)$, so that energies, forces, and positions are computed according to the following rules

- If $U(\mathbf{r}_{n+1}) < U(\mathbf{r}_n)$, $\mathbf{r}_{n+1}$ are accepted, and $h_{n+1} = 1.2h_n$
- Otherwise, $\mathbf{r}_{n+1}$ are rejected, and $h_n = 0.2h_n$

Also, as convergence to a local minimum is not guaranteed by the algorithm, this is stopped either after a finite number of specified steps or whenever $max(|\mathbf{F}_n|) < \varepsilon$, with ε a threshold value, which is estimated from the root-mean-square force f that a harmonic oscillator would exhibit at temperature T, given by

$$f = 2\pi\nu\sqrt{2mk_BT}$$

where ν is the frequency of the oscillator, m the reduced mass, and k_B the Boltzmann's constant.

2.2.5 NVT- and NPT-Equilibration

Molecular dynamics simulations of solids and liquids, need to implement a form of pressure and temperature control to allow their results to be comparable with experiments [24]. Although initial attempts such as velocity rescaling [125], allow some form of temperature control in

simulations, the dynamics is constrained to a microcanonical description (NVE ensemble); i.e., in phase-space $(\mathbf{q};\dot{\mathbf{q}})$ all the microstates of the simulated systems lie on a constant energy shell, which is physically incompatible with a system under laboratory conditions, where controlled temperature(T) and pressure(P_{ext}) introduce fluctuations to the system's energy, so that the phase-space is sampled according to an exponential distribution [7, 24, 73]

The Nosé-Hoover algorithm provides an improved pressure (and temperature) control algorithm in contrast to other ones, since this allows simulations to include anisotropic effects, like in the study of phase transitions of solids, where pressure (and temperature) control needs to be modified to account for heterogeneities and fluctuations in size and shape of the simulation-cell containing the particles. However, to appreciate the authors' contribution, prior understanding of the main ideas is vital.

The foundation of the methods relies on *extending* the dynamic description of the system, originally derived from a Euler-Lagrange equations with $\mathscr{L}_0(\mathbf{q};\dot{\mathbf{q}})$; by featuring new generalised coordinates and momenta $(V,\dot{V},s,\dot{s})$, that account for fluctuations in position ($\mathbf{q}' = \mathbf{q}/V$) and motion ($\dot{\mathbf{q}}' = s\cdot\dot{\mathbf{q}}$) of particles, as a consequence of coupling the system with an external barostat [7] and thermostat [73] respectively. Hence, a new Lagrangian, $\mathscr{L}'$, including contributions of "kinetic" and "potential"-like terms, as below, is proposed

$$\mathscr{L}'((\mathbf{q}',V,s);(\dot{\mathbf{q}}',\dot{V},\dot{s})) = \mathscr{L}_0(\mathbf{q}',\dot{\mathbf{q}}') + K(V) + K(s) - U(V) - U(s)$$

so that, the *pseudo-hamiltonian* $\mathscr{H}'((\mathbf{q}',V,s);(\dot{\mathbf{q}}',\dot{V},\dot{s}))$ derived from the above *extended-lagrangian*, is conserved in the extended-phase-space $((\mathbf{q}',V,s);(\dot{\mathbf{q}}',\dot{V},\dot{s}))$, which then if ergodicity is assumed, can lead to an corresponding probability distributions back in the original phase-space $(\mathbf{q};\dot{\mathbf{q}})$, that approximate the *isobaric-isothermal* distribution (NPT), save terms of order $\mathscr{O}(1/N^2)$ [7].

These ideas constitute the foundation of subsequent improvements, addressing issues regarding time rescaling (as a result of introducing s)[43], ergodicity [65], pressure-anisotropy inclusion [79], and other side problems, all of these discussed in the paper, so that incorporation of these lead to a newly proposed set of equations of motion tested on a harmonic oscillator (known for not being ergodic under previous algorithms) and the solid-fluid phase transition of C60 interacting particles, under anisotropic conditions of pressure [65].

2.2.6 Electrostatics in Biomolecular Simulations

Bulk systems such as liquids, crystals, and macromolecules are generally simulated employing Periodic Boundary Conditions (PBC), where a central simulation box (say $[0,L]^3 \subset \mathbb{R}^3$) that contains the particles whose dynamics we want to simulate (solving Newton's equations: $m_i\ddot{\mathbf{r}}_i = \mathbf{F}_i$, $i = 1,2,\cdots,N$), is thought as surrounded by multiple periodic copies of it, which approximate its infinite bulk environment, and thus avoiding the intrinsic disadvantages of finitely simulated systems [24].

In systems were charged particles are present, the total electrostatic potential energy of the simulated particles, $U(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)$, is needed to solve their dynamics, however, whenever PBC are imposed, the Coulomb interaction energy (normalised by $4\pi\varepsilon_0$), $\phi = qq'/|\mathbf{r} - \mathbf{r}'|$, due to infinite copies of main simulation box, makes U an infnite series (*direct summation method*)

$$U = \frac{1}{2} \sum_{i=1}^{N} \sum_{j}^{N} \sum_{\mathbf{n} \in \mathbb{Z}^3}^{\prime} \frac{q_i q_j}{|r_i - \mathbf{r}_j - \mathbf{n}L|}$$

which although convergent, it slowly tends towards its limit, resulting in huge computational cost unfeasible for practical purposes of many simulations.

Ewald summation method provides a solution to this problem by splitting ϕ into a *short-range* and a *long-range* contributions by a choice of a *window function*, $f(r)$, such that this can be rewritten as (with $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$)

$$\phi(\mathbf{r}_{ij}) = \frac{q_i q_j}{r_{ij}} = \frac{q_i q_j}{r_{ij}} f(r_{ij}) + \frac{q_i q_j}{r_{ij}} [1 - f(r_{ij})] = \phi_{short}(r_{ij}) + \phi_{long}(r_{ij})$$

and by choosing $f(r)$ as the complementary error function, i.e, $f(r) = erfc(\beta r)$ with a tuning parameter (Ewald parameter) $\beta > 0$, one can split U into two series which turn out to be exponentially convergent: one in real space and the other in discrete reciprocal or Fourier space, i.e., $U = U_{local}(\phi_{short}) + U_{distant}(\phi_{long})$, where the former term accounts for contributions by nearby image boxes and the latter due to distant ones. Further manipulations lead to a more simplified form which takes the form of different contributions [26]:

$$U = U_{local} + U_{distant} = U_{local} + U_{dipole} + U_{self} + U_{rec}$$

where the U_{local} converges exponentially in real space, and U_{dipole} and U_{self} are both constant energies due to total dipole moment, $\mathbf{D} = \sum_{i=1}^{N} q_i \mathbf{r}_i$, of the charges within the central simulation box, and their interactions between them and their respective periodic images $\left(\sim \sum_{i=1}^{N} q_i^2\right)$, i.e., *self-energy,* and with U_{rec} a term that features a series, ϕ_{rec}, in the discrete reciprocal space ($\mathbf{m} \in \mathbb{Z}^3$)

$$U_{rec} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} q_i q_j \cdot \phi_{rec}(\beta \mathbf{r}_{ij})$$

Although calculation of U_{rec} is of complexity $\mathcal{O}(N^2)$, or even $\mathcal{O}(N^{3/2})$ at best, if optimizing β, however, if one considers several particles $N > 10^4$, which is the case of simulations of systems such as macromolecules like DNA, for example. Thus, Ewald sum breaks due to high computational cost when estimating ϕ_{rec} (although it represents a huge improvement in contrast to the *direct summation method*).

PME method comes to overcome the limitations of Ewald summation by approximating ϕ_{rec} (and hence U_{rec} too) using a *Lagrangian interpolation*, $\tilde{\phi}_{rec}$, for a regular mesh of the central

simulation box, i.e., $\Lambda = \Lambda_X \times \Lambda_Y \times \Lambda_Z$ (with Λ_α a regular partition of the interval, $[0, L]$ along the α-axis, with $\alpha = X, Y, Z$), alongside the set of mesh images, i.e., $\phi_{rec}(\Lambda)$. [25]

The obtained interpolation polynomial can be written in the form of a discrete convolution which involves ϕ_{rec}, that can be calculated using *Fast Fourier Transform* (FFT), via the *convolution theorem* and then transformed back into real space [120]. The use of FFT comes to decrease the computational complexity to $\mathcal{O}(NlogN)$, since multiple values in $\phi_{rec}(\Lambda)$ can be computed simultaneously via the properties of the FFT.

CHAPTER 3

Docking of Octameric cWza Cys-Mutant Barrels and Structure-Function Relationship

Here, we will show how computational prediction of alternate and/or single conformations of cWza Cys*-mutant ion channels is possible via a proposed framework for geometric analysis applied to structures modelled by symmetric docking of peptides. Additionally, we will provide some understanding of what atomic interactions may take place to hold peptides together as an assembly with a particular conformation, and we will also argue why swapping Tyrosines (*Tyr*) with Cysteines (*Cys*) may be causing peptide structures to collapse to a single conformation. Docked structures per identified confirmation will be tested using equilibrium MD simulations to study the persistence of a conformation splitting amongst end simulated structures. Besides, as a first approach to establishing a structure-function relationship for cWza* Cys*-mutant channels, we will study the conductance states from docked structures via the HOLE programme; which estimates a channel's conductance based on an empirically corrected calculation of the channel's ohmic conductance, from the numerically approximate geometry of the hole passing through a channel. Finally, we will introduce results comparing the HOLE conductance values and energies of assembly from octameric docked channels against their oligomeric docked versions; with 4-11 assembling peptides, to contrast* oligomerisation *as an alternate mechanism to explain alternate low and high conductance states in experiments with cWza* Cys*-mutant lipid-assembling peptides.*

Sequence	Suspected conformations
cWza	*Narrow* and *Wide*
cWza-K375C	*Narrow* and *Wide*
cWza-S355C	*Narrow* and *Wide*
cWza-Y373C	*Single*

Figure 3.1: Summary of suspected conformations adopted by peptide-assembled ion channels with cWza *Cys*-mutant sequences, as previously indicated in Reference [64].

3.1 Modelling the Structure of Peptide-Assembled Barrels

Finding out experimentally the existence of alternate conformations for cWza-mutant channels is a real challenge. As introduced in Chapter 1.1 , *conformational changes* is the leading hypothesis to explain the *conductive-activity change* observed in single-channel recordings for cWza mutant peptide-assembled channels. However, as previously discussed too, obtaining high-resolution crystal structures of the cWza mutant channels seems unfeasible, as crystallization is in general a challenge for membrane proteins since they require a suitable hydrophobic environment for their assembly [21]. So, this limitation makes impossible to validate the conformational-change hypothesis via experimental means, and hence, no clear relationship between conductance and conformation can be established.

Computational modelling techniques have been developed as a workaround to overcome the limitations of experimental means to resolve protein structures. As discussed in Section 2.1, a range of techniques aim to predict detailed atomic-structures of proteins at different levels, such as *secondary structure* or *folding* of single peptides - provided their amino-acid sequence - and the *quaternary structure* of peptide-assembled proteins; that is, the geometry of preferential arrangements in a *peptide-peptide assembly* forming a protein complex; in our case a protein barrel channel.

Symmetric peptide-peptide docking is a technique able to model membrane protein complexes like the cWza-mutant channels while considering the effect on the energy of assembly due to a membrane environment. Symmetric docking can identify membrane assemblies with the strongest affinity under a symmetry constrain of assembly, which seems to play a key role in protein function (Section 2.1). So, next, we introduce results obtained from analysing the pore-inner geometry of peptide-peptide complexes modelled using membrane symmetric docking, assuming peptides with cWza Cys-mutant sequences, for which channels are suspected to adopt either a single or dual conformations as a consequence of their mutated sequence (Sec 1.1). See Table 3.1.

3.2 Methods

Some of the aspects of the methods used next have been previously introduced generically in Section 2.1. Here we introduce very specific aspects regarding the application of these methods step by step.

3.2.1 Membrane Symmetric Docking

Modelling of protein structures was performed via the *Membrane Symmetric Docking* application *MPSymDock*, which is part of the software suite RosettaMP, which provides a framework facilitating membrane protein modelling and design, combining knowledge-based and physics-based principles. However, before using *MPSymDock* some preparation stages described below are needed to guarantee the correct execution of the application.

Preparation of starting structures Performing docking requires an initial starting structure featuring all spatial-coordinates of all atoms from all peptides making up the assembly, which at the same time, is made of a fixed number of peptides units with some initial arrangement. As previously suggested in [64], we assume this number of units to be eight (i.e., octameric assembly), and the sequence of each unit to be identical, that is, the protein complex is assumed to be *homo-octameric*.

Thus, by using PyMOL (Version 1.7.0.0), initial structures with mutated cWza sequences were obtained first by taking the coordinates corresponding to residues 345 to 373 in the PDB (*2J58* ID) structure of *Escherichia coli* Wza, which is *homo-octameric*. Then, mutant structures were obtained via the PyMOL *Mutagenesis* application, first, by mutating residues in all units in order to match the consensus sequence cWza; where atomic clashes (overlaps) - either intrinsic to the crystal structure[1] or introduced by mutation - were reduced via PyMOL *Sculpt*, over 5000 cycles, and with protected backbone coordinates around mutated residues; and next, using the obtained cWza structure as starting point, further mutants were obtained by subsequent applications of *Mutagenesis* - in combination with *Sculpt* - to match cWza Cysteine-mutated sequences shown in Figure 1.2.

Refinement of starting structures Once all mutant structures are obtained, we energy minimize these to remove existing clashes that can lead to a numerical blowup and subsequent failure of *MPSymDock*. Additionally, as discussed in Section 2.1, since membrane docking assumes an implicit membrane model, a reasonable position and orientation of the structure in the membrane frame of reference, i.e., embedding, is also needed for a correct energy assessment. Thus, before docking, we performed a high-resolution refinement via the *FastRelax* application [51, 108], so that 1000 energy-minimized or refined output models were obtained per cWza

[1] Structures derived from the PDB often feature clashes and side-chain rotameric conformations that may lead to unreasonable energies when estimated via force field

mutant sequence. Next, out of this pool of models, only the ones with the lowest total-score were chosen and subsequently optimized for embedding [12], so that output models were chosen for the subsequent docking stage; we refer to these models as the *refined models*.

Performing docking After *refined models* were obtained per sequence, to start docking simulations, we first generated *symmetry definition files* which describe the arrangement of peptide units around a symmetry axis of assembly according to an 8-fold cyclic symmetry (homo-octameric assembly), where chain A in each refined structure was taken as the *master unit* for assembly and the direction of assembly was taken from chain A to chain B, that is, in clockwise direction seeing the protein assembly from the C-terminal side of the units. Thus, *MPSymDock* was subsequently executed so that a set of 1000 output structures were obtained per cWza *Cys*-sequence (Figure 1.2).

3.2.2 Numerical Conductance Estimation

Here, the HOLE programme was employed to estimate conductance of all generated docked models provided their structure coordinates [96–98]. HOLE estimates the conductance, G_{pred}, of a docked channel by first determining the internal pore dimensions of the hole running through its molecular structure, then integrating its conductance according to the pore geometry; G_{macro}, and finally correcting it according to empirical factors [2], γ, with the pore minimum radius, R_{min}^{HOLE}, as the preferential factor for use, as this provides a linear scaling for correction, that is,

$$G_{pred} = \gamma \cdot G_{macro} \quad \text{with} \quad \gamma = \gamma\left(R_{min}^{HOLE}\right) \tag{3.1}$$

As a note, HOLE estimates the geometry of a pore via a Monte Carlo routine employing annealing to search for the best pathway for a sphere of variable radius to pass through a channel. Van der Waal radial values for atoms facing the pore are taken from AMBER force field parameters. See Reference [98] for a more in-depth report of the methodology details.

From HOLE pore dimensions, we used estimates of the pore minimum radius R_{min}^{HOLE} and pore length L^{HOLE}, to screen for indications of conformational differences that could relate to conductance estimates. Results will be shown and discussed in Subsection 3.3.6.

$$\frac{G_{macro}}{G_{exptl}} \sim \gamma = a \cdot R_{min}^{HOLE} + b$$

3.2.3 Structure Geometry Analysis

As a first approach, HOLE was used to approximate the pore dimensions of the holes running through molecular structures of docked channels, to relate these to conductance estimates, as presented in Subsection 3.3.6. However, HOLE estimated pore dimensions lose details regarding the identity of preferential backbone arrangements and/or residue sidechain rotameric configurations associated with single or alternate conductance estimated values.

Thus, here we introduce a framework that can provide more in-depth information as to how peptide-units assemble geometry when forming docked ion channels. Firstly, we deal with the question of how peptides orient in an assembly via a definition of peptide-backbone *Euler angles*, and secondly, we deal with how the interior volume in an assembly is affected by the radial positions of peptide-backbone and residue sidechain atoms relative to the symmetry axis of the assembly.

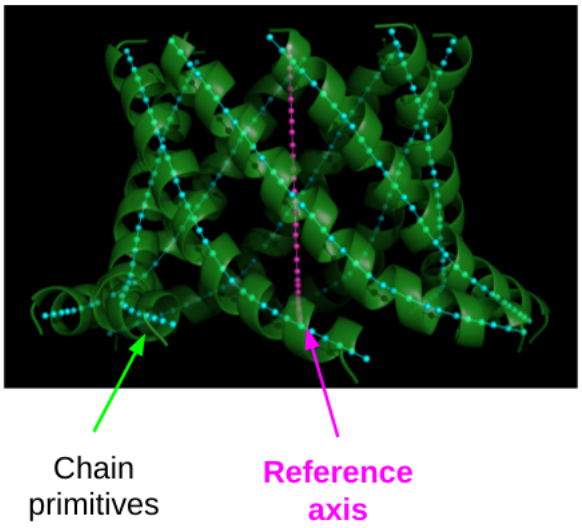


Figure 3.2: Screenshot of the secondary structure of the D4 domain of Wza, featuring all the *primitive curves* passing through each of the eight peptide chains assembling the protein complex. The *reference axis* passing through the peptide assembly is also shown.

3.2.3.1 The Frame of Reference of A Symmetric Peptide-Assembly

To describe positions and orientations of any object in space we first need to define a *frame of reference* for measurement. Here, we employ a frame of reference that is easily constructed via the *AMPAL* framework within the *ISAMBARD* software package, a Python-based framework for structural analysis and rational design of biomolecules, with a particular focus on parametric modelling of proteins [124]. AMPAL stands for Atom, Monomer, Polymer, Assembly and Ligand and is a simple, intuitive framework for representing biomolecular structures.

For a docked model, we construct an *intrinsic coordinate system* to it with an orthonormal base of vectors

$$\mathscr{B} = \{\hat{\mathbf{e}}_{x'}, \hat{\mathbf{e}}_{y'}, \hat{\mathbf{e}}_{z'}\}$$

and centred at the *assembly centre of mass*, $\mathbf{R}_0$. We call this here, the *Assembly Intrinsic Frame of Reference* (Assembly-IFR). Thus, *atomic coordinates* in the PDB of a model can be transformed accordingly so that for a protein with N atoms, the *i-th* atom with *absolute coordinates* $\mathbf{r}_i = (x_i, y_i, z_i)$ shown in the PDB are transformed as

[2]HOLE employs a wide range of geometrical and physical quantities as correcting factors, such as pore minimum radius, pore length, pore aspect ration, average electrostatic potential, among others. See HOLE references [98] for an extensive report of these.

$$x_i \rightarrow x'_i = (\mathbf{r}_i - \mathbf{R}_0) \cdot \hat{\mathbf{e}}_{x'}$$
$$y_i \rightarrow y'_i = (\mathbf{r}_i - \mathbf{R}_0) \cdot \hat{\mathbf{e}}_{y'}$$
$$z_i \rightarrow z'_i = (\mathbf{r}_i - \mathbf{R}_0) \cdot \hat{\mathbf{e}}_{z'}$$

Base vectors $\{\hat{\mathbf{e}}_{x'}, \hat{\mathbf{e}}_{y'}, \hat{\mathbf{e}}_{z'}\}$ can be easily constructed using the AMPAL-defined objects *Backbone Primitives* and *Reference Axis*. Next, we introduce their definitions.

Backbone Primitives are a collection of curves defining a smooth path that crosses the backbone of each peptide chain making up a docked model

$$\{\mathscr{C}^{\alpha}_{primitive} : \alpha = 1, 2, \ldots, N_{chains}\}$$

Each primitive curve is defined as a pathway of pseudo-atoms passing through the interior of each peptide helix, with many points as residues there are in the sequence of the peptide chain.

Thus, for a single peptide chain, say α, with $N_{residues}$, its backbone primitive will be the *ordered set* of $\mathbf{R}^{\alpha}_k$ coordinates, with $k = 1, 2, \ldots, N_{residues}$, of the pseudo-monomers, that is,

$$\mathscr{C}^{\alpha}_{primitive} = \left\{\mathbf{R}^{\alpha}_k : k = 1, 2, \ldots, N_{residues}\right\} \tag{3.2}$$

See Figure 3.2.

The Reference Axis of a protein assembly is defined as the average curve, C, over all backbone primitives, that is,

$$C = \left\{\bar{\mathbf{R}}_k : k = 1, 2, \ldots, N_{residues}\right\} \quad \text{with} \quad \bar{\mathbf{R}}_k = \frac{1}{N_{chains}} \sum_{\alpha=1}^{N_{chains}} \mathbf{R}^{\alpha}_k \tag{3.3}$$

which will be essentially a straight line whose direction $\hat{\mathbf{u}}_C$ is taken upwards the residue sequence, that is,

$$\hat{\mathbf{u}}_C = \frac{\bar{\mathbf{R}}_{N_{residues}} - \bar{\mathbf{R}}_1}{|\bar{\mathbf{R}}_{N_{residues}} - \bar{\mathbf{R}}_1|} \tag{3.4}$$

Thus, N-terminal residues will be downwards C, whereas C-terminal ones will be upwards C.

It is crucial to note that due to cyclic symmetry *i)* the *assembly centre of mass* $\mathbf{R}_0$ will lie along the *reference axis*, C; and that *ii)* the unit vector, $\hat{\mathbf{u}}_{\alpha}$, pointing from the *assembly centre of mass*, $\mathbf{R}_0$, towards the *backbone centre of mass*, $\mathbf{R}^{\alpha}_{com}$, of any chain α, will be perpendicular to the direction of C, that is, $\hat{\mathbf{u}}_{\alpha} \perp \hat{\mathbf{u}}_C$. Moreover, all unit vectors $\hat{\mathbf{u}}_{\alpha}$ will lie in a common plane, and any of these vectors can be obtained from any other, via a rotation around the reference axis C.

Therefore, from all the above the base of unit vectors for the Assembly-IFR can be defined according to

$$\hat{\mathbf{e}}_{x'} = \hat{\mathbf{u}}_\alpha \tag{3.5}$$

$$\hat{\mathbf{e}}_{y'} = \hat{\mathbf{u}}_C \times \hat{\mathbf{u}}_\alpha \tag{3.6}$$

$$\hat{\mathbf{e}}_{z'} = \hat{\mathbf{u}}_C \tag{3.7}$$

where,

$$\hat{\mathbf{u}}_\alpha = \frac{\mathbf{R}^\alpha_{com} - \mathbf{R}_0}{|\mathbf{R}^\alpha_{com} - \mathbf{R}_0|} \tag{3.8}$$

with α taken arbitrarily for a perfectly symmetric assembly. Here, for all docked models we take $\alpha = A$. Also, note that $\hat{\mathbf{u}}_C \times \hat{\mathbf{u}}_\alpha$ refers to the vector inner product; assuming a right-handed coordinate system.

To sum up, all PDB coordinates will be transformed within the Assembly-IFR according to

$$x_i \rightarrow x'_i = (\mathbf{r}_i - \mathbf{R}_0) \cdot \hat{\mathbf{u}}_A \tag{3.9}$$

$$y_i \rightarrow y'_i = (\mathbf{r}_i - \mathbf{R}_0) \cdot \hat{\mathbf{u}}_C \times \hat{\mathbf{u}}_A \tag{3.10}$$

$$z_i \rightarrow z'_i = (\mathbf{r}_i - \mathbf{R}_0) \cdot \hat{\mathbf{u}}_C \tag{3.11}$$

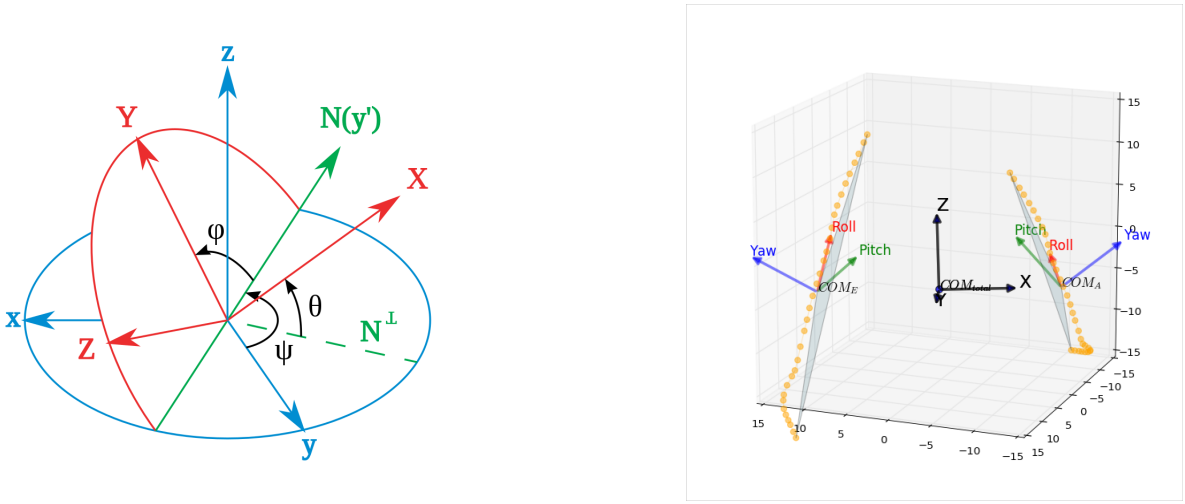


Figure 3.3: **Left** Illustration of a general definition of the Euler Angles (θ, ψ, φ) between the *extrinsic* (from an observer's viewpoint) *xyz* and the *intrinsic* (attached to the rotating body) *XYZ* coordinate systems. **Right** Illustration of the *extrinsic* (*XYZ*) and *intrinsic* (*Yaw-Roll-Pitch*) coordinate systems employed for definition of the Euler Angles (Equations 3.15) of a docked peptide-assembled channel. The *extrinsic* system is defined by the *Assembly Intrinsic Frame of Reference* with the origin at the assembly centre of mass, while the *intrinsic* system is defined by the *Peptide Intrinsic Frame of Reference*. *Yaw-Roll-Pitch* axes of two opposite peptide chains A and E of an octameric docked cWza structure are shown. Vertices of the green, transparent triangles used to define the *Yaw-Roll-Pitch* axes, coincide with the first and last point of the primitive curve of each chain, with the middle vertex at the backbone centre of mass of each chain.

3.2.3.2 Orientation of Assembling Peptides

The *Euler angles* are a set of three angles (θ, ψ, ϕ) that can describe the orientation of a rigid body for a fixed coordinate system xyz, also referred as *extrinsic*. To define these, first we need to define a coordinate system XYZ attached to the rigid body, also referred as *intrinsic*. Thus, angles between the axes of the intrinsic and extrinsic coordinate systems can be measured. Many of these angles will be redundant and therefore the specific choice of measured angles is usually a matter of convention and convenience. Nevertheless, the Euler angles will be the minimum set of (three) measured angles that can describe the orientation of one coordinate system to the other. See Figure 3.3.

For the particular case of a peptide-assembly, here we define Euler angles in a particular way to describe how the backbone of assembling units orient about the Assembly-IFR, defined in Subsection 3.2.3.1. Thus, next, we introduce a definition for a *Peptide Intrinsic Frame of Reference* (Peptide-IFR).

A Peptide Intrinsic Frame of Reference For a rigid body, in general, an intrinsic frame of reference can be always constructed by taking three fixed points in it. In the case of all symmetric docked models here, the backbone of a peptide α can be considered as a rigid body [3], and hence, a simple choice of these points will be the *centre of mass of the peptide backbone*, $\mathbf{R}^{\alpha}_{com}$, and the two endpoints of its backbone primitive $\mathscr{C}^{\alpha}_{primitive}$, that is, $\mathbf{R}^{\alpha}_{1}$ and $\mathbf{R}^{\alpha}_{N_{residues}}$.

To construct the coordinate axes of the Peptide-IFR, we first take the triangle $\triangle A_{\alpha}B_{\alpha}C_{\alpha}$ with vertices at

$$A_{\alpha} = \mathbf{R}^{\alpha}_{1}$$
$$B_{\alpha} = \mathbf{R}^{\alpha}_{com}$$
$$C_{\alpha} = \mathbf{R}^{\alpha}_{N_{residues}}$$

Then, we define the unit vectors along the edges $B_{\alpha}A_{\alpha}$ and $B_{\alpha}C_{\alpha}$ of the triangle so that B_{α}, the centre of mass of the backbone peptide is taken as a pivotal point, that is,

$$\hat{\mathbf{e}}_{B_{\alpha}A_{\alpha}} = \frac{A_{\alpha} - B_{\alpha}}{|A_{\alpha} - B_{\alpha}|}$$
$$\hat{\mathbf{e}}_{B_{\alpha}C_{\alpha}} = \frac{C_{\alpha} - B_{\alpha}}{|C_{\alpha} - B_{\alpha}|}$$

Thus, from the above a set of three orthonormal vectors $\{\hat{\mathbf{e}}^{\alpha}_{yaw}, \hat{\mathbf{e}}^{\alpha}_{pitch}, \hat{\mathbf{e}}^{\alpha}_{roll}\}$ can be easily constructed, defining what we call here as the *Yaw*, *Pitch*, and *Roll* rotation axes of a peptide backbone,

$$\hat{\mathbf{e}}^{\alpha}_{yaw} = \hat{\mathbf{e}}_{B_{\alpha}C_{\alpha}} \times \hat{\mathbf{e}}_{B_{\alpha}A_{\alpha}} \tag{3.12}$$

$$\hat{\mathbf{e}}^{\alpha}_{pitch} = \frac{\hat{\mathbf{e}}_{B_{\alpha}C_{\alpha}} + \hat{\mathbf{e}}_{B_{\alpha}A_{\alpha}}}{|\hat{\mathbf{e}}_{B_{\alpha}C_{\alpha}} + \hat{\mathbf{e}}_{B_{\alpha}A_{\alpha}}|} \tag{3.13}$$

$$\hat{\mathbf{e}}^{\alpha}_{roll} = \hat{\mathbf{e}}^{\alpha}_{yaw} \times \hat{\mathbf{e}}^{\alpha}_{pitch} \tag{3.14}$$

Note that axes names are adopted here as an analogy to the description of the rotation of an aircraft in flight about the fixed frame of reference of an observer from the ground. In this way, the definition of the above unit vectors becomes more intuitive to interpret.

Finally, we define next a set of Euler Angles between the axes of the Peptide-IFR (*intrinsic* frame of reference) and the Assembly-IFR (*extrinsic* frame of reference), given their orthonormal coordinate bases $\{\hat{\mathbf{e}}^{\alpha}_{yaw}, \hat{\mathbf{e}}^{\alpha}_{pitch}, \hat{\mathbf{e}}^{\alpha}_{roll}\}$ and $\{\hat{\mathbf{e}}_{x'}, \hat{\mathbf{e}}_{y'}, \hat{\mathbf{e}}_{z'}\}$ in Equations 3.5 and 3.12, with $\alpha = 1, 2, \ldots, N_{chains}$

[3] All docked models for a particular cWza Cys-mutant sequence have identical peptide backbones, because of the way RosettaMP performs docking and as seen in RMSD calculations from the peptide-peptide alignment between different models. However, the rigid body assumption can strictly fail for URMD simulations

A set of Euler angles to measure peptide backbone rotations (θ,ψ,ϕ) are introduced here as the *elevation angle* (θ), the *interchain angle* (ψ), and the *facing angle* (ϕ), defined per chains $\alpha = 1,2,\ldots,N_{chains}$ making up a docked assembly.

The elevation angle θ_α is defined as the angle between the *Roll* axis and its projection onto the plane X'Y' of the Assembly-IFR. This angle is intended to give us an idea of how a peptide backbone elevates in relation to the plane of the assembly (X'Y'), which should be perpendicular to the axis of symmetry of the assembly (*reference axis*), and parallel to the lipid plane where the peptide-assembly can sit. Thus, we have

$$cos\theta_\alpha = \hat{\mathbf{e}}^\alpha_{roll} \cdot \hat{\mathbf{e}}^\alpha_{roll-X'Y'}$$

where $\hat{\mathbf{e}}^\alpha_{roll-X'Y'}$ is the normalised vector of the projection of $\hat{\mathbf{e}}^\alpha_{roll}$ onto X'Y', that is,

$$\hat{\mathbf{e}}^\alpha_{roll-X'Y'} = \frac{(\hat{\mathbf{e}}^\alpha_{roll} \cdot \hat{\mathbf{e}}_{x'})\hat{\mathbf{e}}_{x'} + (\hat{\mathbf{e}}^\alpha_{roll} \cdot \hat{\mathbf{e}}_{y'})\hat{\mathbf{e}}_{y'}}{|(\hat{\mathbf{e}}^\alpha_{roll} \cdot \hat{\mathbf{e}}_{x'})\hat{\mathbf{e}}_{x'} + (\hat{\mathbf{e}}^\alpha_{roll} \cdot \hat{\mathbf{e}}_{y'})\hat{\mathbf{e}}_{y'}|}$$

The inter-chain angle $\psi_{\alpha\beta}$ is defined as an angle between the *Roll* axes of neighbouring backbones of peptides in the assembly, say α and β chains. This angle is intended to give us an idea of how backbones orient with respect to each other when docked together in an assembly. Thus, for two consecutive chains we have

$$cos\psi_{\alpha\beta} = \hat{\mathbf{e}}^\alpha_{roll} \cdot \hat{\mathbf{e}}^\beta_{roll}$$

The facing angle ϕ_α is defined as the angle between the *Pitch* axis and its projection onto the plane X'Y' of the Assembly-IFR. This angle is intended to give us an idea of how a peptide backbone is facing the interior (towards the *solvent environment*) or the exterior (towards the *lipid environment*) of the assembly.

$$cos\phi_\alpha = \hat{\mathbf{e}}_{x'} \cdot \hat{\mathbf{e}}^\alpha_{pitch-X'Y'}$$

where $\hat{\mathbf{e}}^\alpha_{pitch-X'Y'}$ is the normalised vector of the projection of $\hat{\mathbf{e}}^\alpha_{pitch}$ onto X'Y', that is,

$$\hat{\mathbf{e}}^\alpha_{pitch-X'Y'} = \frac{(\hat{\mathbf{e}}^\alpha_{pitch} \cdot \hat{\mathbf{e}}_{x'})\hat{\mathbf{e}}_{x'} + (\hat{\mathbf{e}}^\alpha_{pitch} \cdot \hat{\mathbf{e}}_{y'})\hat{\mathbf{e}}_{y'}}{|(\hat{\mathbf{e}}^\alpha_{pitch} \cdot \hat{\mathbf{e}}_{x'})\hat{\mathbf{e}}_{x'} + (\hat{\mathbf{e}}^\alpha_{pitch} \cdot \hat{\mathbf{e}}_{y'})\hat{\mathbf{e}}_{y'}|}$$

In summary, the collection of Euler angles defined here to describe peptide backbone rotations for all units $\alpha = 1,2,\ldots,N_{chains}$, are

$$\begin{aligned} cos\theta_\alpha &= \hat{\mathbf{e}}^\alpha_{roll} \cdot \hat{\mathbf{e}}^\alpha_{roll-X'Y'} \\ cos\psi_{\alpha\beta} &= \hat{\mathbf{e}}^\alpha_{roll} \cdot \hat{\mathbf{e}}^\beta_{roll} \\ cos\phi_\alpha &= \hat{\mathbf{e}}_{x'} \cdot \hat{\mathbf{e}}^\alpha_{pitch-X'Y'} \end{aligned} \tag{3.15}$$

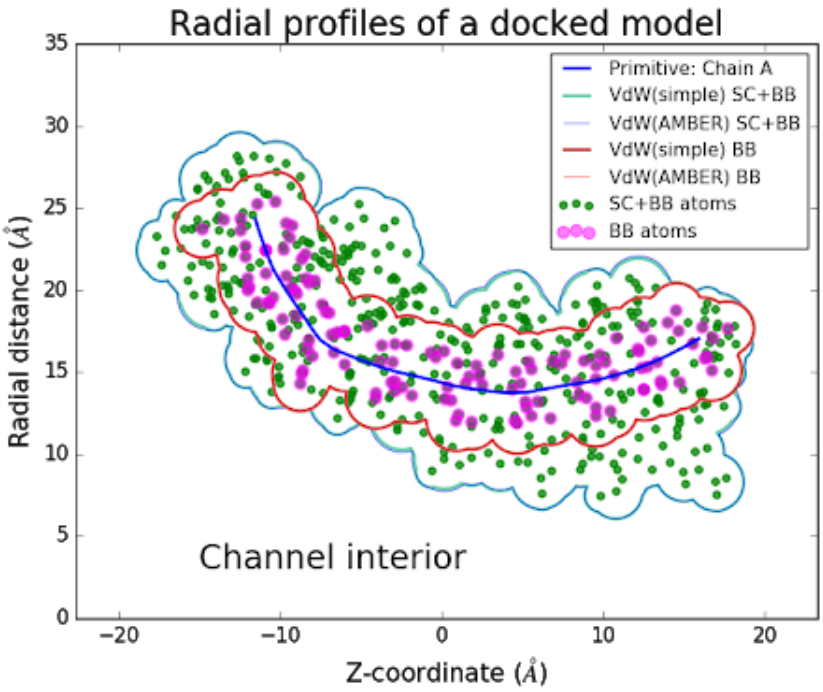


Figure 3.4: Illustration and comparison between primitive ($\Gamma^{\alpha}_{primitive}$) and Van der Waals (VdW) (Γ^{α}_{VdW}) radial profiles obtained from a single peptide chain ($\alpha = A$ chain). VdW profiles from *backbone* atoms (BB) and from *all atoms* (backbone and sidechain atoms, i.e., BB+SC) with different VdW radii parameters (*Simple* AMBER and AMBER force field) taken from the HOLE programme [98] are also compared, but without any noticeable difference. Scattered, coloured points show the projected position of backbone and sidechain atoms onto the *Peptide Radial Plane* of chain A.

3.2.3.3 Radial Profile Analysis

Now, we deal with the question of how the positioning of peptides in an assembly affects the dimensions of the internal volume within a channel. Here, we approach the problem by introducing the concept of *radial profile*, which allows us to determine how the internal radius within a channel varies as we move from one its ends to the other, either due to backbone atoms or backbone and sidechain atoms combined, as seen in Figure 3.4. Thus, we introduce some key concepts that will allow us to calculate different types of radial profiles which can give us a different type of information as to how the internal radial dimensions of a channel vary.

Key Concepts *The Peptide Radial Plane* S_α of a peptide α, is defined as the plane formed by the set unit vectors $\{\hat{\mathbf{e}}_{z'}, \hat{\mathbf{u}}_\alpha\}$, which were previously defined by Equations 3.4 and 3.8. In consequence, all S_α planes have the *reference axis* C as a common intersection (Subsection 3.2.3.1).

The projector Π_α is a function that maps any vector $\mathbf{r}'$ measured within the Assembly-IFR to

its projection onto the plane S_α corresponding to the assembling peptide α, that is,

$$\begin{aligned}\Pi_\alpha(\mathbf{r}') &= (\mathbf{r}'\cdot\hat{\mathbf{e}}_{z'})\hat{\mathbf{e}}_{z'} + (\mathbf{r}'\cdot\hat{\mathbf{u}}_\alpha)\hat{\mathbf{u}}_\alpha \\ &= z' + (\mathbf{r}'\cdot\hat{\mathbf{u}}_\alpha)\hat{\mathbf{u}}_\alpha \\ &= (z', \mathbf{r}'\cdot\hat{\mathbf{u}}_\alpha)\end{aligned}$$

where $\alpha = 1,2,\ldots,N_{chains}$.

The Primitive Radial Profile of a peptide α will be simply defined as the curve, $\Gamma^\alpha_{primitive}$, resulting from the projection of its primitive, $\mathscr{C}^\alpha_{primitive}$, onto its radial plane, S_α, after transformation into the Assembly-IFR; that is,

$$\Gamma^\alpha_{primitive} = \left\{\Pi_\alpha(\mathbf{R}^\alpha_k - \mathbf{R}_0) : k = 1,2,\ldots,N_{residues}\right\} \tag{3.16}$$

where $\mathbf{R}^\alpha_k$ and $\mathbf{R}_0$ are defined by previous Equations.

The Van der Waals Radial Profile of a *collection of N atoms* with positions $\{\mathbf{r}'_i : i = 1,2,\ldots,N\}$, which belong to the a peptide α, will be defined as the curve, Γ^α_{VdW}, resulting from taking *the outermost boundary* of the union of all circles with radii $\{r^{VdW}_i > 0 : i = 1,2,\ldots,N\}$ centred at the projection of each atomic position onto the radial plane, S_α. See Figure 3.4 for an illustration.

Formally, this can be defined as,

$$\Gamma^\alpha_{VdW} = \partial_{outermost}\left\{\bigcup_{i=1}^{N} S^{VdW}_i(\Pi_\alpha(\mathbf{r}'_i))\right\} \tag{3.17}$$

where S^{VdW}_i referes to the *i-th* circle centred at coordinates $\Pi_\alpha(\mathbf{r}'_i))$ in the S_α plane; with radius matching the VdW radius corresponding to the *i-th* atom, with $i = 1,2,\ldots,N$.

Here, Radii of the circles correspond to the *Van der Waals Radii* taken from the AMBER force field, which is implemented in the HOLE programme (See references in Subsection 3.2.2).

Hence, from the above we can determine VdW radial profiles for particular collections of atoms, such as *backbone* (BB) atoms and for all atoms (AA) - including both backbone and sidechain atoms - belonging to the peptide α, which will be referred here as *the VdW BB-profile* and the *the VdW AA-profile*, respectively.

Note that due to cyclic symmetry of docked models, the VdW profiles of a single chain, say $\alpha = 1$, is enough to analyse the geometry a single structure.

Metrics over Radial Profiles are employed here to differentiate between profiles from docked structures with the same cWza Cys-mutant sequence and identify any noticeable difference in their conformation; with the same idea in mind as intended by using Euler Angles. Some of the metrics explored in our work are defined next.

Geometric metrics can provide information into how the overall shape of the channel internal volume changes due to radial positioning of atoms away from the reference axis (axis of assembly) C. So, for a radial profile of any type, $\Gamma = \{(z', \gamma(z')) : z' \textit{ coordinate in } C\}$, we introduce the metrics below:

- *Profile Length*

$$L(\Gamma) = \mathbf{max}\{z' : \textit{for all z' of } \Gamma\} - \mathbf{min}\{z' : \textit{for all z' of } \Gamma\}$$

- *Profile Minimum Radius*

$$R_{min}(\Gamma) = \mathbf{min}\{\gamma(z') : \textit{for all z' of } \Gamma\}$$

- *Profile Maximum Radius*

$$R_{max}(\Gamma) = \mathbf{max}\{\gamma(z') : \textit{for all z' of } \Gamma\}$$

Statistical metrics are introduced here to complement the information from geometric metrics, as they provide information about how atomic positions away from C distribute and fluctuate on average. Here we employ the metrics defined below.

- *Profile Radial Average*

$$\mu(\Gamma) = \frac{1}{|\Gamma|} \sum_{z' \textit{ of } \Gamma} \gamma(z')$$

- *Profile Radial Standard Deviation*

$$\sigma(\Gamma) = \sqrt{\frac{1}{|\Gamma|} \sum_{z' \textit{ of } \Gamma} [\gamma(z') - \mu(\Gamma)]^2}$$

- *Profile Radial Coefficient of Variation*

$$CV(\Gamma) = \frac{\sigma(\Gamma)}{\mu(\Gamma)}$$

3.2.4 Peptide-Peptide Interaction Analysis

Here we introduced some of the most relevant *peptide-peptide* or *inter-chain* interactions known for contributing towards the structural stability of α-helical peptide assemblies. Also, we introduce a framework of analysis based on identification and quantification of the probability of appearance of these interactions, use to find distinguishing signs of cWza Cys-mutant modelled structures.

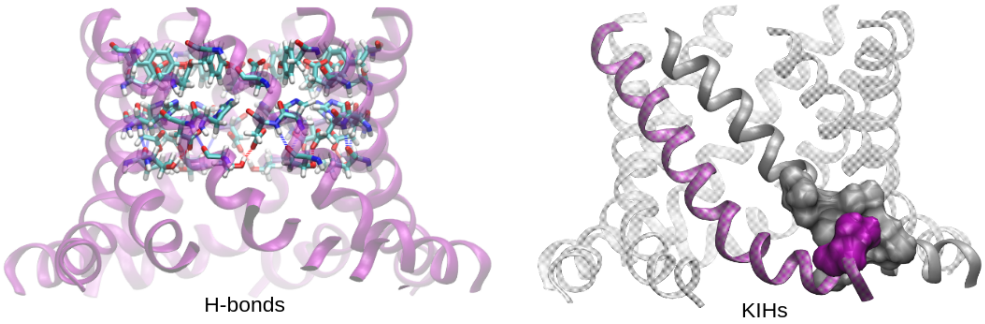


Figure 3.5: Screenshots of the secondary structure of a docked cWza ion channel highlighting some of its interchain Hydrogen bonds (*Left*) and Knobs-Into-Holes interacting atoms (*Right*). All interactions were identified using the *Isambard* Software Package [124].

Hydrogen Bonds or H-bonds are a form of attractive electrostatic interaction between a hydrogen bound to an electronegative molecule X (*donor*) and another electronegative atom Y, bound to a molecule Z (*acceptor*) [9].

$$X - H \; | \; | \; | \; Y - Z$$

H-bonds are known to be responsible for a range of peculiar physical and chemical properties of N, O, and F compounds; like the high boiling point of water, for instance. In particular, in proteins, H-bonds are known for stabilising the backbone helical shape (*secondary structure*) of α-helical peptides, and overall they contribute to the specificity of protein-protein and protein-ligand interactions.

H-bonds can be identified computationally provided the coordinates of a protein structure according to criteria based on *i)* the linear distance between the donor to the acceptor atoms, $d_{XY} = |\mathbf{r}_X - \mathbf{r}_Y|$, and the angle between

Different types of H-bonds can be formed between atoms belonging to different atomic groups in the the peptide-assembly, such as *backbones* (BB) and *sidechain* (SC) atoms in the same peptide (*Intra-Chain*) and/or from different peptides (*Inter-Chain*). Here, *Isambard* implements the above criteria to identify the whole set of interacting atoms according to this wide range of possible H-bonds. Filtering and statistics of relevant H-bond types and the identity of H-bond forming atoms will be described later.

Knobs-Into-Holes (KIHs) are another relevant form of peptide-peptide interactions mainly found in assemblies of α-helical peptides, condensing the notion that peptides interact via buried contacts between them. More specifically, KIHs are fundamentally a packing motif that takes place between sidechains of α-helical peptides, where the sidechain of a residue in a peptide forms a *knob* which fits into a *hole* formed by the space created by *four residues* on another

single α-helix. This model of packing has been widely confirmed at atomic resolution for a variety of *water soluble* coiled-coil arrangements of two to seven peptides [105, 114] . Also previously, studies observed that α-helix pairs in the transmembrane regions of the photosynthetic reaction centre (PRC), cytochrome C oxidase and bacteriorhodopsin interact by KIHs, although these were observed to be less compact and regular by contrast to water-soluble proteins [56]. Additionally, the structure of the membrane Wza D4 domain was observed too to show a propensity for KIHs (Supplement of Reference [64]).

Computational Identification of KIHs is possible via the SOCKET algorithm implemented within *Isambard*, which assumes a KIH to take place between two α-helices, X and Y, if the distances between the centre of a sidechain of a residue in X, k_X (*knob* residue), and all the centres of the four nearest residue-sidechains in Y , $h^1_Y, h^2_Y, h^3_Y, h^4_Y$ (*hole* residues), are within a distance cutoff of 7Å [114].

$$k_X >> h^1_Y, h^2_Y, h^3_Y, h^4_Y$$

See Figure 3.5, *Right*.

SOCKET assumes a simplified representation for residues, where each residue is represented by its sidechain *centre* and a *terminal end*. The *sidechain centre* is defined as the average position - i.e, geometric centre - of all sidechain atoms from the C_β onwards, but excluding hydrogens; for the case of *Glycine*, the C_α atom is taken as the centre. The *terminal end* is simply the terminal atom of the sidechain or the average position if two termini are present.

Salt bridges (SBs) are the result of the attractive electrostatic interaction between ionised sites in residues with net charge, that is, *Arg* (R), *Lys* (K), and *His* [4] (H), which carry a net *positive charge* (+); and *Asp* (D), *Glu*(E), which carry a *negative charge* (-). SBs are known to have a stabilising influence in the kinetics of folding of α-helices and are also suggested to promote the stability of coiled-coil assemblies [67, 86, 102].

Computational Identification of SBs is possible again via *Isambard*, which simply identifies these based on the inter-atomic distance between ionised sites of all charged residues with an opposite charge, within a distance range of $2.5 - 4$Å. In this work, *Histidines* are not considered when searching for SGs though, as determining the charged nature of them is usually extremely dependent on its surrounding environment, hence, making the computational search more complex than simply based on inter-atomic distances.

[4]Histidine can be uncharged or positively charged, as Histidine's pKa can easily be perturbed by its local environment, hence providing this with the ability to behave both as a polar or charged amino acid, as well as a hydrophobic residue [14]

3.2.4.1 Framework of Analysis

The framework introduced here is formulated in such a way that peptide-peptide interactions from docked models can be analysed regardless of their type, i.e., *H-bonds*, *KIHs*, or *SBs*. Thus, the analysis employed in our work comprehends the following aspects

- *Identity of Interacting Atomic-Groups*
  We identify the specific *identity* of interacting atoms and/or residues and the *direction* of the interaction between peptide pairs.

  In more detail, for a given structure s within a set of models S (i.e., $s \in S$), we identify its interactions described according to the following notation, reading the interaction from left to right,

$$\begin{aligned}
\textit{Hydrogen bond} \quad & r^s_{i,X} - H \quad ||| \quad O - r^s_{j,Y} \quad : \quad \delta(XY) \\
\textit{Knob-Into-Hole} \quad & r^s_{i,X} \quad >> \quad (r^s_{j,Y}, r^s_{k,Y}, r^s_{l,Y}, r^s_{m,Y}) \quad : \quad \delta(XY) \\
\textit{Salt bridge} \quad & r^s_{i,X} - a^+ \quad --- \quad a^- - r^s_{i,Y} \quad : \quad \delta(XY)
\end{aligned}$$

  where, $r^s_{i,X}$ is a residue in the peptide sequences of chain X of model s. Since models are docked from peptides with an identical sequence (*homomeric*), then i corresponds to some residue number in $1,2,\ldots,N_{residues}$. We say that the *direction of the interaction*, $\delta(XY)$, goes from chain X to chain Y, and this is *clockwise* if this is *parallel* to the direction of increment of the chains, that is, $X < Y$, with $X, Y = 1,2,\ldots,N_{chains}$; otherwise, we say it is *anti-clockwise*, with $Y > X$, or *intra-chain* if the interacting residues are in the same chain.

$$\delta(XY) = \begin{cases} +1 & \text{if } X < Y \textit{ clockwise} \\ -1 & \text{if } X > Y \textit{ anti-clockwise} \\ 0 & \text{if } X = Y \textit{ intra-chain} \end{cases}$$

  For example, for the interchain H-bond, ω, we have

$$\omega = [Tyr_{29,A} - H \quad ||| \quad -Thr_{26,B} \quad : \quad \delta(AB)]$$

  since, the interactions go from chain A to B, and clearly $A < B$ in the chain sequence. Then, $\delta(AB) = +1$, that is the interaction goes in the same direction of increment of the chain sequence.

  Similar notation and conventions apply here to KIHs and SBs; where for the latter we indicate the atoms corresponding to the *positively* and *negatively* charged sidechain sites as a^+ and a^-.

- *Propensity of Interchain Interaction*
  Once interactions are identified, then we quantify their probability of appearance (*propensity*) among given a set of structures with the same peptide sequence.

  In more detail, say we have a set of models S, and we want to identify the probability that an interaction, ω - for instance, an H-bond - is found among all structures in S, that is, the *propensity* of the interaction, $P_S(\omega)$, defined as

$$P_S(\omega) = \frac{\sum_{s\ in\ S} \text{ Times } \omega \text{ appears in } s}{N_{chains} \times \text{ Total number of structure in } S}$$

  where $0 \leq P_S(\omega) \leq 1$.

  Note that for a symmetric docked model, say s, with existent ω interaction, the number of times this will appear in s will match N_{chains}. However, any induced distortion of the symmetry can lower this number, as it may be in the case of MD simulations, due to thermal fluctuations.

- *Conformational Specificity*
  If conformational groups are identified within a set of docked models with the same peptide sequence, then we want to determine what interactions are distinguishing of each alternate conformation.

  In more detail, say that for a set of models S, docked out of N_{chains} peptide units with identical sequence, we identify two alternate conformations with distinctive geometries, say C_0 and C_1, so that the sets S_{C_0} and S_{C_1} (*conformation groups*) - which are subsets of S, contain these structures. Now, suppose that after analysing interchain interactions of type ω (e.g., H-bonds) for both sets, we find a set of N_ω *common interactions* for all models in S_{C_0} and S_{C_1}, for instance,

$$\omega_1, \omega_2, \ldots, \omega_{N_\omega}$$

  for which, we also determine their propensities according to conformation, that is,

$$P_{S_{C_0}}(\omega_1), P_{S_{C_0}}(\omega_2), \ldots, P_{S_{C_0}}(\omega_{N_\omega})$$
$$P_{S_{C_1}}(\omega_1), P_{S_{C_1}}(\omega_2), \ldots, P_{S_{C_1}}(\omega_{N_\omega})$$

  from which we can expect a difference in propensity values for some ω interactions, depending on the structures corresponding to a conformation group.

  Thus, we say that an interaction, say ω_k, will be *representative* of a conformation, say C_0, *if and only if*, the difference in propensities for that interaction, $\Delta P_{S_{C_0}/S_{C_1}}(\omega_k)$, is bigger than some threshold ($0 \leq \delta \leq 1$), that is

(3.18) $$0 < \delta < \Delta P_{S_{C_0}/S_{C_1}}(\omega_k) = P_{S_{C_0}}(\omega_k) - P_{S_{C_1}}(\omega_k)$$

meaning that ω_k has simply a higher propensity within S_{C_0} than in S_{C_1} structures. Hence, ω_k can be considered a marker of S_{C_0} structures. Thus, a similar criterion can be formulated to identify markers of S_{C_1} structures, based on interactions ω, of any type, i.e., H-bonds, KIHs, and/or SBs.

- *Robustness of Interactions*
  Finally, once identified and classified according to alternate conformations, if existent, interactions are assessed in terms of their survival after putting docked models through MD, where the disordering (*entropic*) effects due to a dynamic bilayer and solvent environment are simulated.

  In more detail, we draw two assess the robustness of interactions

  - We want to know how robust *common interactions* among a set of structures, S', are after MD.
  - We want to know how robust *distinguishing interactions* from *conformation groups* are after MD.

Suppose that we have a set of structures S', and that after analysing their interchain interactions of type ω, e.g., H-bonds, we determine their set of N_ω common interactions $\omega_1, \omega_2, \ldots, \omega_{N_\omega}$, as well as their propensities

$$P^{docked}_{S'}(\omega_1), P^{docked}_{S'}(\omega_2), \ldots, P^{docked}_{S'}(\omega_{N_\omega})$$

Next, suppose that we put these S' structures through MD and that we analyse their ω interactions to determine new values of propensity, either right after the end of the simulation or from the trajectory by taking average propensities, that is,

$$P^{MD}_{S'}(\omega_1), P^{MD}_{S'}(\omega_2), \ldots, P^{MD}_{S'}(\omega_{N_\omega})$$

Since we expect propensities to change after simulation, to determine how robust initially identified interactions are after MD, we will say that an interaction, ω_k is robust under MD, *if and only if*, the absolute difference in propensities, $\Delta P^{MD/docked}_{S'}(\omega_k)$, between docked and MD stages is lower than some threshold ($0 \leq \delta' \leq 1$), that is

(3.19) $$\left|\Delta P^{MD/docked}_{S'}(\omega_k)\right| = \left|P^{MD}_{S'}(\omega_k) - P^{docked}_{S'}(\omega_k)\right| < \delta'$$

Based on the above scheme, we can see that we can easily test for *robustness of distinguishing interactions*, by taking the absolute difference between values of docked and MD stages of $\Delta P_{S_{C_0}/S_{C_1}}(\omega_k)$, that is,

$$\left|\left(P^{MD}_{S_{C_0}}(\omega_k) - P^{MD}_{S_{C_1}}(\omega_k)\right) - \left(P^{docked}_{S_{C_0}}(\omega_k) - P^{docked}_{S_{C_1}}(\omega_k)\right)\right| < \delta' \tag{3.20}$$

provided that the following conditions are also simultaneously satisfied

$$\begin{aligned} 0 < \delta < & P^{MD}_{S_{C_0}}(\omega_k) - P^{MD}_{S_{C_1}}(\omega_k) \\ 0 < \delta < & P^{docked}_{S_{C_0}}(\omega_k) - P^{docked}_{S_{C_1}}(\omega_k) \end{aligned} \tag{3.21}$$

In short, a distinguishing interaction ω_k of the conformation group C_0 will be robust, if and only if, conditions 3.20 and 3.21 hold provided the range of propensity values $P^{docked}_{S_{C_0}}, P^{docked}_{S_{C_1}}, P^{MD}_{S_{C_0}}, P^{MD}_{S_{C_1}}$ and thresholds $0 < \delta, \delta' < 1$. Similar criteria can formulated for other distinguishing interactions or the other conformation group.

3.2.5 All-Atom Molecular Dynamics Assays

3.2.5.1 Simulation Stages

Setup Individual docked models were embedded in a patch of electrically neutral POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) bilayer of 128 lipids per leaflet; previously solvated and equilibrated in a simulation box of size $12 \times 12 \times 8$ nm [5]. Embedding was performed using the compression-insertion-expansion protocol `g_membed` of GROMACS [122]. Initial positions and orientations of individual proteins in the bilayer were determined before embedding according to lowest energy values of embedding, estimated by `memembed` via a knowledge-based statistical potential [78].

Once embedded, protein-bilayer complexes were re-solvated with SPC-water, within new simulation boxes of standard size $12 \times 12 \times 10$nm removing water molecules within the bilayer and inserting $K^+ - Cl^-$ ions into the solvent at 1M concentration. The *OPLS-AA* force field was employed to parametrise interactions among peptides, water molecules, and ions, while *Berger* parameters were employed for lipid-protein interactions [15, 106, 122]. Original hydrogens from all docked models were preserved during set up.

Energy Minimisation MD (EMMD)

To remove possible atomic clashes from embedding and re-solvation, 20,000 steps of EMMD were performed for all complexes via *stepest-descent* (Subsection 2.2.4); implemented in GROMACS.

Position-Restrained MD (PRMD)

After EMMD, all end complexes were *equilibrated* for Temperature and Pressure (NPT-ensemble) with common targets of temperature (Nosé-Hoover Chain thermostat) and pressure (Parrinello-Rahman barostat) of 310 K and 1 atm, respectively, for 5 ns of simulation with a timestep of 1 fs. Positions of protein heavy-atom (non-hydrogens) were restrained harmonically ($k_x = k_y = k_z = 1000$ kcal/mol) and bond-vibrations were constrained using the P-LINCS algorithm.

Un-Restrained MD (URMD)

After equilibration, final coordinates were taken as the starting point for MD production for 100ns, with a time step of 2fs; without restrains, with newly generated velocities, keeping identical targets of temperature and pressure.

[5] Structure downloaded from Peter Tieleman website: `http://wcm.ucalgary.ca/tieleman/downloads`

3.3 Results and Discussion

For modeling of the cWza-based channels, we employed *symmetric docking* for identification of various sets of (1000) outcome structures from independent transmembrane-protein-energy (score) minimisation attempts, per number of chains (n) assembling a channel with corresponding C_n-symmetry; $n = 4 - 11$ chains, for all sequences of the *first-type*; cWza-Y373C, and the *second-type*; cWza, cWza-K375C, and cWza-S355C.

The *master unit* to initiate docking per sequence was obtained from the first chain of the octameric X-ray crystal structure of the outer-membrane domain (*Alanine* 345 to *Threonine* 376) of natural Wza (PDB id: 2J58), mutated accordingly to match cWza-based sequences using the PyMOL software. Sampling algorithms and the score function for docking were implemented in a customised fashion via the RosettaMP framework, as part of the *Rosetta* protein modelling software [3].

As part of the analysis, for all generated structures their conductance is estimated numerically; using the HOLE programme [98], and then compared against their *interface energy score* values; the energy due to pairwise interactions between atoms from different chains, to assess the stability of the docked assemblies, and hence to identify any possible trend in the distribution of optimal models consistent with either a single or dual conductance case per selected sequence type. First, to examine *gating*, results for assemblies with 8 chains (*octamers*) are introduced and discussed in Subsec., **??**; since the natural structure of Wza is octameric, in addition to suggestive data from electrophysiology experiments (pore blocking), it is suspected that the cWza-based channels should be octameric too [64]. Next, to examine *oligomerisation*, remaining results for structures assembled with $n \neq 8$ (*oligomers*) are similarly analysed and discussed in Subsection 3.3.7.

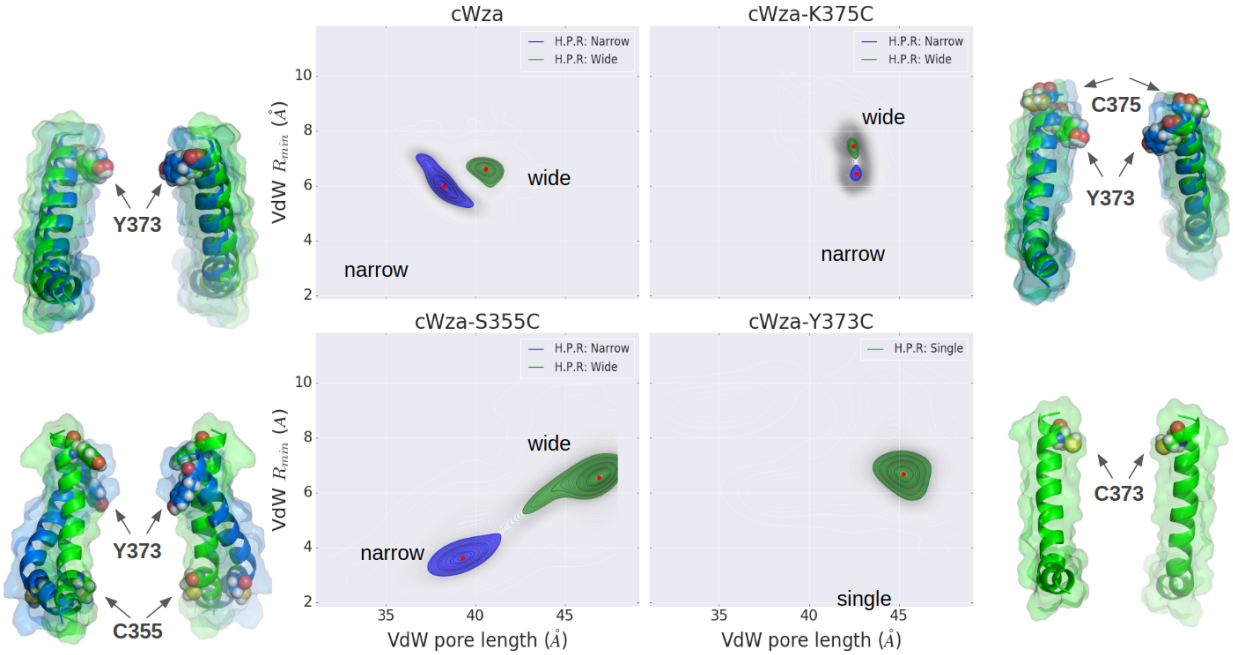


Figure 3.6: *Subplots*: High Probability Regions (HPR) in the Van der Waals (VdW) dimensional distinguish alternate *Narrow* / *Wide* or *Single* conformations of docked cWza *Cys*-mutant octameric channels; peaks of estimated probability density functions are marked with red dots. *Screenshots*: Two helices facing each other are taken from structures corresponding to Narrow/Wide or Single conformations around the probability peaks. Colours of the helices and regions highlighted in plots are matched. *Tyrosines* (Y373) and Cysteines atoms are shown in every mutant.

3.3.1 Identification of Alternate Conformations

Clear identification of dominant Narrow and Wide conformations for cWza, cWza-K375C, *and* cWza-S355C *docked models, was observed from estimated dimensions of the internal volume within docked channels via their all-atom Van der Waals profiles; the length of the profile (*VdW pore length*) and the minimum radial distance ($VdW - R_{min}$). Conformations can be linked to Low and High conductance states since conductance positively correlates positively with the $VdW - R_{min}$. Also, Euler angles and other radial profile metrics revealed details regarding the distinctive backbone and / or sidechain radial displacements and orientations that distinguish conformations and that affect the size and surface properties of bottlenecks of the channels; crucial for conductance.*

As a simple approach to determine existent conformations, after geometric analysis of docked structures we came up with a procedure, described as follows:

First, from all docked models with the same sequence, we took their *VdW All-Atom Profiles* for their first chain (Subsubsection 3.2.3.3); because of cyclic symmetry. Next, from each profile,

we took the *profile length* and the *minimal radial distance*, defining a pair of coordinates per model

$$\left(L(\Gamma^{A}_{AA,VdW}), R_{min}(\Gamma^{A}_{AA,VdW})\right) \tag{3.22}$$

to describe in a simple way the overall dimensions of the internal volume within a channel model. All 1000 docked models per sequence were taken to obtain collections of scattered (L, R_{min}) data.

Second, from the scattered (L, R_{min}) data per sequence, their associated *Probability Density Function* (PDF), $\rho = \rho(L, R_{min})$, was estimated via a Gaussian *Kernel Density Estimator* (KDE). See contour density maps in Appendix Figure A.1.

Third, we determined numerically the *local maxima* of PDFs over the 2D space of coordinates (L, R_{min}) [110],

$$(L^{i,max}, R^{i,max}_{min}) \quad with \quad i = 1, 2, \ldots, N_{maxima}(\rho)$$

for which their height $\rho(L^{i,max}, R^{i,max}_{min})$ was expected to be bigger than 0.01, in order to exclude shallow peaks.

Fourth, we determined *non-overlapping* High Probability Regions (HPR), S_i, over the (L, R_{min})-plane from the PDF around each maximum, $i = 1, 2, \ldots, N_{maxima}(\rho)$, using the criterion below

$$S_i = \left\{(L, R_{min}) : \rho(L, R_{min}) \geq 0.5 \cdot \rho(L^{i,max}, R^{i,max}_{min})\right\}$$

Contour data density maps were employed to easily identify the boundaries of S_i.

Fifth, and final, structures with coordinates (L, R_{min}) falling within the boundaries of each HPR centred at each maximum of the PDF, were filtered out models according to their HPR.

After applying the above procedure, in the 2D space of (L, R_{min}) VdW dimensions, we identified two HPRs for models with peptide sequences *cWza, cWza-K375C, cWza-S355C*, and a single HPR for *cWza-Y373C* models, as seen in Figure 3.6. Moreover, maxima of HPRs for the first three sequences, defined an order relation between them; for instance, for the first and second maximum: $L^{1,max} < L^{2,max}$ and $R^{1,max}_{min} < R^{2,max}_{min}$. Hence, models falling within the HPR around the first maximum were labelled as *narrow*, whereas those around the second maximum were labelled as *wide*. So, from here on, we will refer to the sets of classified structures as *conformation groups*.See Table 3.1, for details on coordinates of maxima of the PDFs, for all mutants.

To sum up, the number of HPRs or *conformation groups* found for all cWza *Cys*-mutant sequences matched the number of conductance states found in experiments for each sequence. Recall that for channels made of peptides with any of the sequences *cWza, cWza-K375C, cWza-S355C*, low and high experimental conductance states were observed, whereas for *cWza-Y373C* a

Sequence	Narrow	Wide	Single
cWza	38.24, 5.97 Å	40.60, 6.62 Å	-
cWza-K375C	42.63, 6.42 Å	42.45, 7.43 Å	-
cWza-S355C	39.28, 3.64 Å	46.85, 6.58 Å	-
cWza-Y373C	-	-	45.28, 6.96 Å

Table 3.1: Coordinates of local maxima of Probability Density Functions over the space of (L, R_{min}) *VdW*-dimension, per cWza *Cys*-mutant peptide sequence of docked models. Maxima are matched to their High Probability Regions, labelled according to their conformation group (Narrow, Wide, or Single).

single experimental conductance state was only seen. In addition, since for many natural channels, the minimum radius of their bottlenecks is a dominant factor which positively correlates with their conductance, we infer that identified conformations should be linked to conductance states of low and high conductance (Figures 1.2). Therefore, conformation groups identified from docked structures seem to relate to conductance states observed in experiments, at a qualitative level.

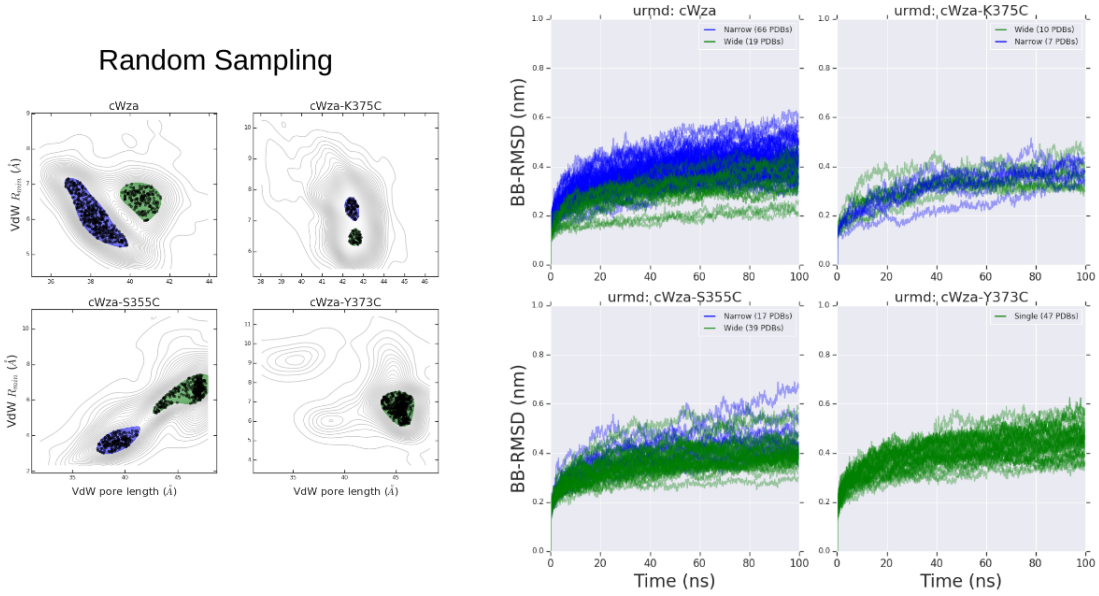


Figure 3.7: *Left*: Random sampling of previously identified conformation groups in VdW (L,R_{min}) dimension space. Scattered data points represent actual models falling within the boundaries of each conformation group, per mutant. *Right*: MD evolution of Backbone-Root-Mean-Square Displacement (BB-RMSD) with respect to starting structure (final structure at the end of PRMD stage), for sampled structures according to conformation, per *Cys*-mutant sequence.

3.3.2 Molecular Dynamics of Docked Conformations and Limit Conformations

*When simulated in Unrestrained Molecular Dynamics (URMD), randomly sampled docked models from previously found conformational groups (Figure 3.6) converged towards limit conformational-groups (90-100 ns) close to equilibrium. These were found via the Radius-of-gyration components around the Z-axis (*perpendicular to the membrane plane*) and the X/Y axes of each structure frame (*parallel to the membrane plane*). Moreover, these post-MD conformational-groups showed a dual, alternate, group separation for* cWza-K375C *and* cWza-S355C *mutant channels; whereas, for* cWza-Y373C *channels, a single limit conformational group predominated. This segregation/aggregation of structures matched the grouping initially found amongst docking via their pore dimensions. For* cWza *channels though, only one group was found by the end of simulations, that is, the initial conformational separation was erased. Nevertheless, simulations of docked structures showed that end segregated conformational-groups (*cWza-K375C *and* cWza-S355C*) cannot be explained via thermal fluctuations. Thus, overall, docked structures can reproduce features of the conformational states inferred from experiments of cWza* Cys*-mutant channels.*

To test the robustness of the classification of identified conformational groups we simulated docked structures using MD. We aimed to observe whether conformational segregation/aggregation according to pore internal dimensions were preserved after the dynamics of cWza *Cys*-mutant structures. Docked channels were sampled randomly taking 20% of models from each conformational group; corresponding to a total of 303 PDB structure files (Figure 3.7; *Left*). Next, using GROMACS (Version 5.0.6) and following the simulation protocol outlined in Subsection 3.2.5, each structure was embedded in lipid bilayer, energy minimised (EMMD stage), equilibrated for pressure and temperature (PRMD stage), to ultimately simulate dynamics (URMD stage) for 100 ns ($\approx 30\mu s$ in total), with time-step of 2 fs, saving trajectories frames every 100 ps. Velocities and forces were not saved during the simulation to avoid excessive use of storage. All simulations were performed independently and in parallel in an HPC facility (BlueGem[6]), each taking 32 CPUs (2 nodes) to keep computational time to roughly 3 days per simulation.

After performing MD, we attempted to determine the pore dimensions of the simulated docked structure per frame in its trajectory. However, we found that peptide assemblies experienced noticeable distortion of their initial cyclic symmetry - inherited from docking. More specifically, Backbone *Root Mean Square Fluctuations* (BB-RMSF) per residue, showed that backbone structural distortions were more pronounced for residues near the N- and C-terminal ends of the assembling peptide units (Appendix B). Thus, because of the inaccurate definition of the *assembly reference-axis* in such distorted structures the methodology described in Subsection 3.2.3.1 was found inadequate to accurately measure the pore minimum radius of structures. Nevertheless, provisionally to estimate how end simulated channels deviated with respect to their initial structures we employed a decomposition of the *backbone radius of gyration*, around the Z-axis, $R_{g,Z}$; to account for the distribution of assembling peptide backbone-atoms (N, O, and C_α) parallel to the membrane bilayer plane, that is,

$$R_{g,Z} = \frac{1}{M} \sum_{i=1}^{N_{BB}} [(x_i - x_{COM})^2 + (y_i - y_{COM})^2] m_i$$

and the mean radius of gyration around the X and Y axes, that is, $\frac{1}{2}(R_{g,X} + R_{g,Y})$, where

$$R_{g,X} = \frac{1}{M} \sum_{i=1}^{N_{BB}} [(y_i - y_{COM})^2 + (z_i - z_{COM})^2] m_i$$
$$R_{g,Y} = \frac{1}{M} \sum_{i=1}^{N_{BB}} [(x_i - x_{COM})^2 + (z_i - z_{COM})^2] m_i$$

to account for the distribution of atoms perpendicular to the membrane bilayer plane.

Although the radius of gyration is a lower-resolution metric to describe the geometry of channels, this seems suitable given the major changes experienced by the peptide assemblies

[6]BlueGem is an HPC facility provided by the Bristol Synthetic Biology Centre, a.k.a., *BrisSynBio*

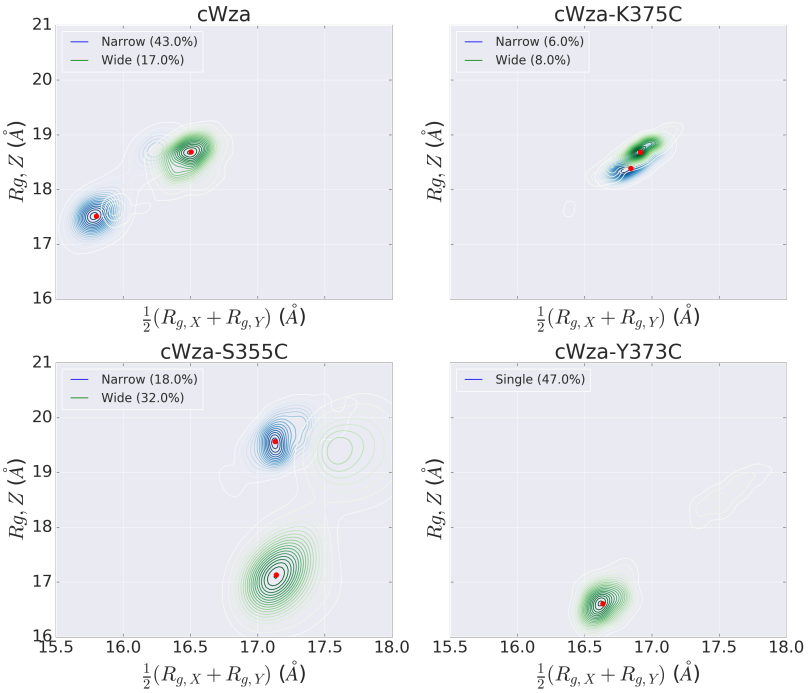
during dynamics. Moreover, the radius-of-gyration decomposition of previously classified docked models preserves a conformational segregation/aggregation similar to the one identified via VdW-pore-dimensions (high-resolution metric); as seen in the *Top Panel* of Figure 3.8. Hence, we subsequently employed the radius-of-gyration decomposition ($R_{g,XY}$,$R_{g,Z}$) for structural analysis of MD trajectories.

First, though, we aimed to determine whether 100 ns of simulation were enough to allow structures to relax. Thus, for each simulated structure, we extracted the joint time-series of their radius-of-gyration components: ($R_{g,XY}(t)$,$R_{g,Z}(t)$), for all data frames in the 100 ns MD trajectories. Then, we analysed the end trends (80-100 ns) via the rate-of-change of these individual time-series and determined that simulated structures tend to equilibrium overall for both R_g-components, for all conformational groups, all mutants. Thus, docking can provide peptide-assemblies that equilibrate in MD within the time-length of our simulations. See Appendix B for details of the methodology employed to determine equilibration from ($R_{g,XY}(t)$,$R_{g,Z}(t)$) time-series and for a further discussion on the atomic-details behind the symmetry-breaking observed from simulated structures.

Once found that most structures reached equilibrium, we analysed the last 10 ns of the dynamics (90-100 ns) to determine whether the initial conformational segregation/aggregation of the R_g-components was preserved. Then, all ($R_{g,XY}(t)$,$R_{g,Z}(t)$) time-series corresponding to their assigned initial conformational group were concatenated to form a single data-set, per docked conformation, per mutant. Ultimately, probability density functions (PDFs) were estimated per data-set via a Gaussian-Kernel Density-Estimator [7] to generate the contour plots shown in Figure 3.8.

End structures tend to cluster according to their radius-of-gyration decomposition. From the contour plots in Figure 3.8, we see that despite major rearrangements of assembling backbone-units, discrete groups or *limit conformations* can still be distinguished. However, the initial grouping of docked conformations (Top Panel) did not match always the newly identified limit groups of conformations revealed by the PDFs (Bottom panel). For *cWza* channels, a single conformational group can be only distinguished, meaning that the initial conformational distinction in docked structures was erased. For *cWza-K375C* channels, while Narrow structures define a single group, Wide structures split into two groups, with one of these matching the Narrow group. This may indicate that end Wide structures transitioned to end Narrow structures, however, overall two groups can still be observed. For *cWza-S355C* channels, a persistent and more clear separation is observed between end Narrow and Wide structures. Although Wide structures split into two main groups, one of these dominates as shown by the peak height in $R_{g,Z}$. Finally, for *cWza-Y373C* channels, their limit structures clustered into a main limit conformation, as initially found in docked structures.

[7] PDFs were determined over a 400×400 grid, taking the minimum and maximum of the R_g-components. Determination of the band-width was done according to Scott's rule, which is implemented by default in `SciPy` - Python library.



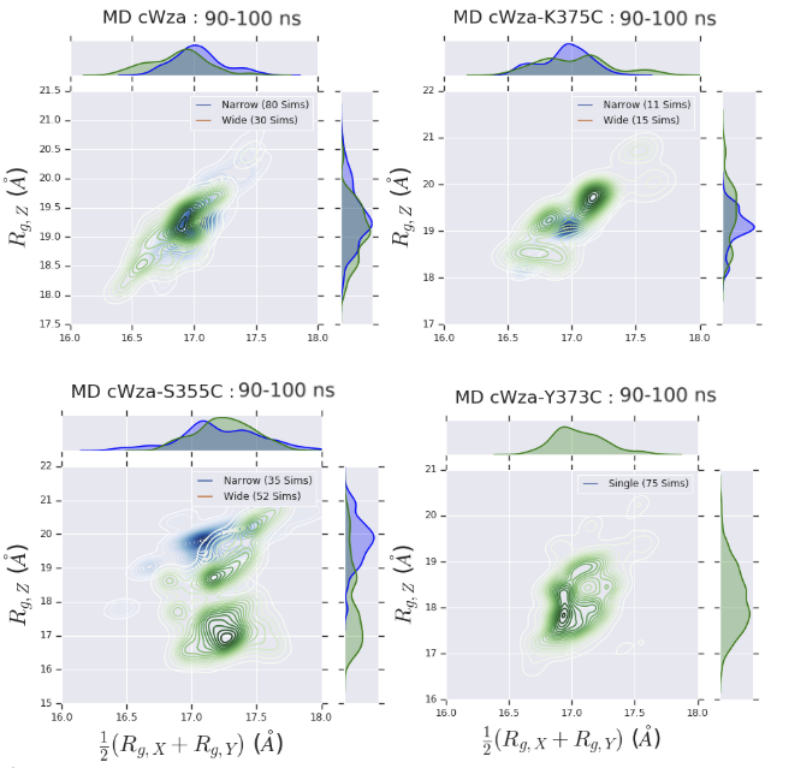


Figure 3.8: Probability Density Functions (PDFs) from radius-of-gyration decomposition data show a similar cluster segregation as when using VdW-dimensions (Figure 3.6) before (*Top*) and after MD simulation (*Bottom*). Merged data for simulated docked channels indicate the appearance of *limit conformations* by the end of 100ns of URMD. All cWza *Cys*-mutant sequences are featured. Blue and Green contour plots correspond to Narrow and Wide (Single; for cWza-Y373C structures) docked conformations. Red dots in the top plot indicate the peak position of each PDF.

Overall, an aggregation/segregation into conformational groups can still be observed in MD for end simulated structures; with some agreement with the originally classified groups from docked structures. However, as indicated before (Section 1.1), conformational changes between *Low-* and *High-conductance states* is driven by an input of energy provided by an increment in the applied voltage. Hence, thermal fluctuations of the channels should not be able to drive conformational transitions. To find out whether this feature is reproduced in our results, we determined the strength of the fluctuations (S) for the last 10 ns of each simulation, when the structures tend to equilibrium. The fluctuation values (S) per simulation were calculated as

$$S = \sqrt{\sigma^2_{R_{g,XY}} + \sigma^2_{R_{g,Z}}}$$

where $\sigma_{R_{g,XY}}$ and $\sigma_{R_{g,Z}}$ are respectively the standard-deviation values of the time-series of $R_{g,XY}$ and $R_{g,Z}$ values from the MD trajectory (90-100 ns) of a simulated structure.

Then, we compared their distribution, $P(S)$, with the *minimum distance between all maxima* (peaks) in the PDFs of merged ($R_{g,XY}$,$R_{g,Z}$)-datasets (*Bottom*, Figure B.8), i.e., d_{min}. So, if the orders of $P(S)$ and d_{min} are comparable, thermal fluctuations may be behind the conformational segregation in our simulations. Analysis shows that this not the case, as d_{min} is larger than the mean strength of the fluctuations up to the standard deviation of their distribution, that is, $|\mu(S) - d_{min}| < \sigma(S)$, for $P(S)$ of cWza-K375C and cWza-S355C - Figure B.9 (Appendix B); the mutants that show peaks in the PDFs of their radius-of-gyration decomposition values (Bottom, Figure B.8). No difference was found neither for distributions of fluctuations $P(S)$ from classified groups (Narrow and Wide) nor when merging all datasets per mutant. See Appendix B for supplementary figures to better understand this analysis.

Overall, we have found that our simulated docked-structures tend to equilibrium, can show separate limit conformational-groups, and that this grouping cannot be explained via thermal fluctuations. More importantly, limit conformational-groups identified by backbone radius-of-gyration show a degree of qualitative agreement with experimentally inferred alternate conformations for *cWza-K375C*, *cWza-S355C*, and a single conformation for *cWza-Y373C*. For the *cWza* channel a single limit conformational-group was observed only though. However, the origin of the observed clustering by the end of dynamical simulations seems unclear. We will deal with this issue next.

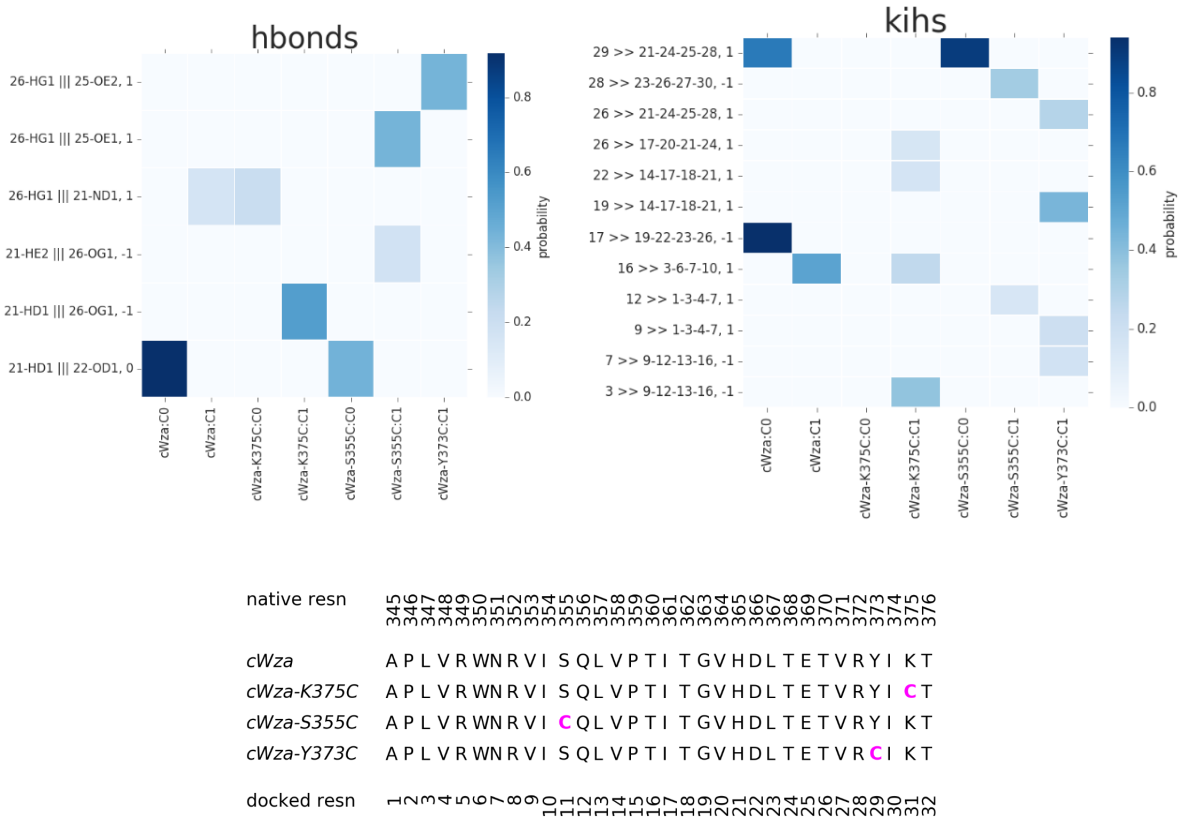


Figure 3.9: Set of H-bonds (*Left*) and KIHs (*Right*) identified to be distinguishing of docked conformations per sequence and that also survive after PRMD (initial conditions of URMD stage) of sampled models (Subsection 3.3.2). Original *interaction propensities* from docked models (pre-MD) are shown in all heat-maps. Interactions (heat-map rows) follow the notation in Subsubsection 3.2.4.1, and with residue numbers following the RosettaMP convention to start from 1 (`docked resn`); an equivalence with native residue numbers, `native resn`, (from the crystal structure of Wza) is shown in the *Bottom Panel*, with mutant Cysteines (*Cys*) highlighted to indicate their position in all mutated sequences.

3.3.3 Peptide-Peptide Interactions in Alternate Docked Conformations

Docked models were found to display minimal sets of distinguishing inter-chain H-bonds *and* KIHs *according to conformation, which were found to be robust after Position-Restrained Molecular Dynamics (*PRMD*). Thus, the identities of these interactions were employed to identify the initial condition of peptide-peptide interactions present in simulated docked structures right at the start of* URMD.

To understand the origin of previously found limit conformations of docked models by the end of their dynamic simulations, we analysed the main inter-chain interactions holding units together in docked peptide-assemblies prior to MD simulation; *Hydrogen(*H*)-bonds*, *Knobs-Into-Holes* (KIHs), and *Salt-bridges* (SBs) (Subsection 3.2.4). We determined both the identity and the

propensity of appearance per chain-pair for H-bonds, KIHs, and SBs, following the *framework of analysis* described by the end of Subsection 3.2.4.

First of all, Inter-chain Sidechain-Sidechain H-bonds (i.e., between consecutive peptides), Intra-chain Sidechain-Sidechain H-bonds (i.e., within the same peptide), and KIHs were found to be the most present interactions among peptide-docked structures. By contrast, Inter-chain and Intra-chain SBs [8], and other Inter-chain H-bond types such as Backbone-Sidechain and Backbone-Backbone were overall scarce. Hence, we only focused on the analysis of the former. As a note, from here on whenever we refer to H-bonds, it will be implied that we refer to Inter-chain Sidechain-Sidechain and/or Intra-chain Sidechain-Sidechain H-bonds.

Identities and Propensities of H-bonds and KIHs were determined for all models classified according to their docked-conformation per cWza *Cys*-mutant sequence. Similarly, we analysed interactions from structures from EMMD (end coordinates) and PRMD (average interactions over 100 frames from last 1 ns) stages, for randomly sampled models according to docked-conformations (Figure 3.7, *Left*). Then, we reduced the set of identified interactions to find a common set of the most significant ones [9], we merged all sets of identified interactions for docking and PRMD, and then filtered out those interactions whose probabilities were lower than 10%; per docked-conformation, per mutant, per stage. See Figures in Appendix B.

To relate peptide-peptide interactions at the start of unrestrained dynamics (Subsection 3.3.2) to interactions identified by symmetric docking according to conformation, we reduced all previous sets of identified interactions from all pre-URMD stages (docking, EMMD, and PRMD), to focus only on interactions that were both distinguishing of each docked conformation or *conformation specific* and/or *robust* up to the PRMD stage. Thus, we performed the *interaction assessment* outlined by the end of Subsection 3.3.2, so that end interactions were reduced to the set displayed in Figure 3.9, with their propensities relative to the size of the conformation sample set. Extended data from interaction assessment and threshold values used to filter out distinguishing and/or robust interactions per interaction type, per conformation, per mutant, can be found in Appendix B.

So far, we have defined a reduced set of interactions which are conformation-specific and define the initial peptide-peptide interactions present in sampled docked models prior to their simulated dynamics. Next, we will relate these interactions to the limit conformations mentioned in Subsection 3.3.2, in order to see whether it is possible to separate limit conformations based on their initial interactions, as their initial docked conformations (through VdW pore dimensions) was found not being robust by the end of simulated dynamics.

[8] Isambard implementation to identify salt-bridges ignores Histidine (*His*) residues as positively charged, given their ambiguous nature, briefly discussed in Subsection 3.2.4. So, it is likely that SBs were underrepresented in our analysis.

[9] Note that during simulated MD stages, interactions change due to perturbation of symmetry of docked structures, then leading to discrepancies in the identity of present interactions between docked structures and those corresponding to any MD stage.

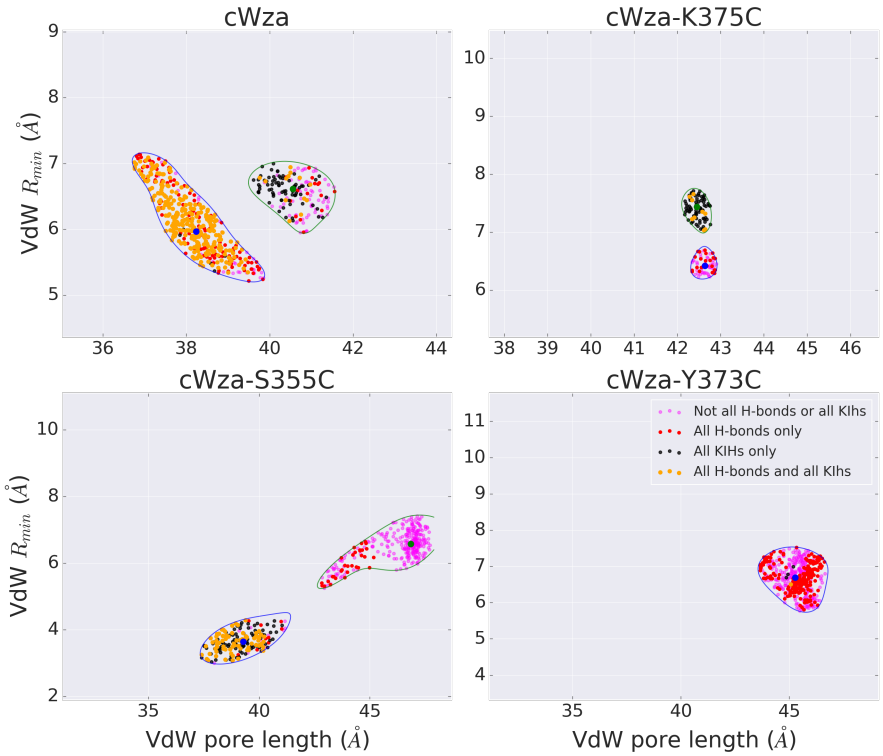


Figure 3.10: Docked conformation groups can be decomposed into subsets of structures including particular combinations of inter-chain H-bonds and KIHs, previously identified as distinguishing according to conformation.

3.3.4 Peptide-Peptide Interactions and Limit Conformations in Molecular Dynamics

Although initially identified docked conformations related poorly to previously found limit conformations in URMD, initial conditions of peptide-peptide H-bonds *and* KIHs *related strongly to these. Thus, starting peptide-peptide interactions can bias the end observation of determined limit conformations, more than initially docked conformations.*

Here, we will relate previously identified MD steady limit conformations of randomly sampled structures from docked conformations (Subsection 3.3.2) to the set H-bonds and KIHs found in structures right at the start of their dynamics (Subsection 3.3.3), in order to see whether these initial (interaction) conditions can account for the discrete clustering of limit structures, better than docked conformation. However, given the variety of possible combinations of H-bonds and KIHs that can be considered - despite the reduction performed - we need to define first how to group initial conditions of H-bonds/KIHs combinations that we can eventually relate to the limit conformation groups, per docked conformation, per mutant.

As a first approach to the above problem, we grouped initial interaction conditions according to the following conditions for H-bonds and/or KIHs, per docked conformation (C_0 and/or C_1), per cWza *Cys*-mutant sequence, shown in Figure 3.9.

- All H-bonds AND all KIHs
- All H-bonds, NOT all KIHs
- All KIHs, NOT all H-bonds
- NOT All H-bonds OR All H-bonds

these groups were proposed under the assumption that the more interactions present in a structure, the more the peptide arrangements making up docked assemblies will be held in place. Consequently, structures belonging to any docked conformation group per mutant sequence, will fall only within any of the above-defined non-overlapping *interaction groups*, so that we can think of the set of structures to be partitioned according to their interactions, as illustrated in Figure 3.10.

Now, in order to relate limit conformations to the above interaction groups defining initial interaction conditions, we first grouped available URMD trajectories for simulated docked models falling within each interaction group (Figure 3.10). Next, we merged their backbone radius of gyration data ($BB-R_{g,XY}$,$BB-R_{g,Z}$) for the last 10ns of their trajectories, per interaction group. Finally, their estimated PDFs - relative to the set of models filtered per interaction group, were estimated and their contour levels plotted as shown in Figure 3.12, for all interaction groups, for all conformations, for all cWza *Cys*-mutant sequences.

Hence, in Figure 3.12, we can see how emerging clusters from limit structures (*Left*) seem to be decomposed according to the identified initial interaction conditions at the beginning of simulated dynamics (*Right*). For a more clear break down of the backbone radius of gyration data per model, see Appendix D. Next, we describe and discuss these data according to mutant sequences and docked conformation.

cWza: C0 We can see that limit structures correspond to models that feature as initial interaction conditions, both all H-bonds and all KIHs in Figure 3.9, make up a large part of the limit conformation cluster. A significant contribution to defining the limit conformation cluster is seen due to models with all H-bonds but not all KIHs, in Figure 3.9. We note that for these models, initially there was a single H-bond, that took place between residues *His-365* and *Asp*-366 in the same chain (intra-chain). On the other hand, we see from the KIHs present at the start of the dynamics that these residues also take part in KIHs of opposite interchain direction (clockwise and anti-clockwise), as part of groups of residues making up the holes. Hence, this suggests that these residues taking simultaneous involvement in both types of interactions are the result of a coupling between H-bonds and KIHs. Thus, interactions are not independent.

cWza: C1 We can see that limit structures correspond mostly to models with initial interaction conditions excluding all H-bonds and all KIHs in Figure 3.9, which show a single H-bond and a single KIH. Other interaction groups are too small to judge their contribution to the overall conformation cluster. So, there is no clear information to judge key residues participating in starting interactions at the start of the dynamics, at least conformation-specific interactions shown in Figure 3.9.

cWza-K375C: C0 Limit structures correspond to models with initial interaction conditions featuring only a single inter-chain H-bond half-way between the N- and C-terminal ends, with Nitrogen being the acceptor atom between residues *Thr-370* (donor) and *His-365* (acceptor), per chain pair. Distinguishing KIHs are absent for this docked conformation. H-bonds with Oxygen acceptors are weaker in strength (5.0 kcal/mol) in comparison to H-bonds with Nitrogen acceptors (6.9 kcal/mol) - with a difference of 1.9 kcal/mol $\approx$ 3.2 k_BT. So, given the strength of initial H-bonds, this suggests to be consistent with observed limit models which are more contracted, than limit conformations for models from the Wide docked conformation (C_1); discussed below.

cWza-K375C: C1 We can see that limit structures correspond to models that feature all KIHs but not H-bonds in Figure 3.9, as initial interaction conditions. Thus, models featured only KIHs, four of them simultaneous at the start of simulated dynamics; two of them located near the bottom of the channels (N-terminal end) and the other two above the centre of the channels (before C-terminal end). Interestingly, the identity of bottom KIHs revealed that knob-residues of one KIH, correspond to a hole-residue in the other KIH, with both KIHs having opposite directions in the peptide assembly; this suggests that bottom KIHs are coupled at the start of the dynamics. On the other hand, the identity of residues for the upper pair of KIHs, point in the same direction, so that knob-residues belong to the same chain, but with their hole residues sharing residues, perhaps providing a joint hole. So, this suggests a very particular packing motif of combined KIHs that lock neighbouring peptides at their common interface.

cWza:S355C: C0 Limit structures for these models aggregate mainly in two clusters - separated from clusters from limit structures for a model with Wide docked conformations (C_1), with one of them shown as dominant. From Figure 3.12 (*Right*) we can see that the dominant cluster (black) will correspond to models with all KIHs as starting condition, however, from Figure 3.9 we see that there is a single distinguishing KIH, also featured by cWza models with equal Narrow docked conformation (C_0). However, the subdominant cluster (yellow) correspond to models with both H-bonds and KIHs as starting interactions, where the featured H-bond is the same intra-chain H-bond featured by cWza:C_0 models. Overall, this analysis suggests that *His-365* and *Asp*-366 interaction lead to coupling of H-bonds and KIHs between peptides. We see though that limit conformations actually are expanded, hence, initial Narrow docked conformation does not match a contracted limit conformation.

cWza-S355C: C1 Similarly, limit structures aggregate in two clusters - and separately from clusters for limit structures of a model with Narrow docked conformations (C_0), with one of them shown as dominant. From Figure 3.12 (*Right*) we can see that the dominant cluster (magenta) corresponds to models without all H-bonds or all KIHs in Figure 3.9. Thus, at this stage, it is ambiguous to tell what interactions prevail at the start of the dynamics, as combinations of 2 H-bonds and 2 KIHs can take place. However, note that *Thr-370* is involved in both inter-chain H-bond (either as donor or acceptor residue) and KIHs - like a hole residue. Thus, the coupling of H-bonds and KIHs is observed again, as for cWza:C0 but with different residues involved. On the other hand, the subdominant cluster will correspond to models with all inter-chain H-bonds, however, the number of simulated trajectories for these models is low, hence, discussion of its relevance will require further simulations. Despite this, we see that *Thr-370* seems a recurrent residue involved in H-bonds, playing both roles of donor and acceptor.

cWza-Y373C In Figure 3.12 limit structures aggregate in a single expanded cluster (*Left*), made of the overlap between limit structures with all H-bonds and limit structures without all H-bonds or all KIHs; which separately make two overlapping clusters at the same time. Thus, models from the single dominant docked conformation of cWza-Y373C channels are mapped to a single limit conformation, regardless of the differences in starting interaction conditions. In mode detail, these two conditions correspond to, first, structures featuring a single inter-chain H-bond per chain pair, involving *Glu-369*(donor) and *Thr-370* (acceptor), and second, structures featuring combinations of four different KIHs, two in the lower half of protein structures (near N-terminal end) and two in the upper half of the structure (near the C-terminal end). Lower KIHs show a similar pattern of arrangement, as for cWza-K375C: C1, where Lower KIHs point in opposite directions, and with the knob residue of one, playing as a hole-residue of the other, in different chains. In contrast, knob residues upper KIHs are found in the same chain, but their hole residues in the neighbouring chain are shared. Interestingly, starting KIHs for cWza-Y373C are not common for other docked structures with other cWza *Cys*-mutant sequences, which suggests that mutation of *Tyr-373* may promote interaction between residues as a result of possible side-chain rearrangements taking place locally in peptide-peptide interfaces.

Overall, the above analysis reveals what initial peptide-peptide interaction conditions can determine the limit conformations of docked conformations per mutant sequence, despite the lack of a one-to-one relationship between docked conformations and limit conformations. Moreover, mechanistic details of initial interactions were revealed, showing that residues such as *His-365*, *Thr-370*, and *Asp-366* couple H-bonds and KIHs - like in cWza:C0 and cWza-K375C:C0, or just there were recurrently observed in interaction conditions associated to limit conformations. The role of *Tyr-373* is suspected to promote common interactions observed among docked conformations of cWza *Cys*-mutants; however, when swapped with *Cys*, this seems to promote inter-chain residue-residue interactions that would not be observed otherwise. Further discussion on the role of *Tyrosine* in cWza *Cys*-mutant channels will be discussed later on.

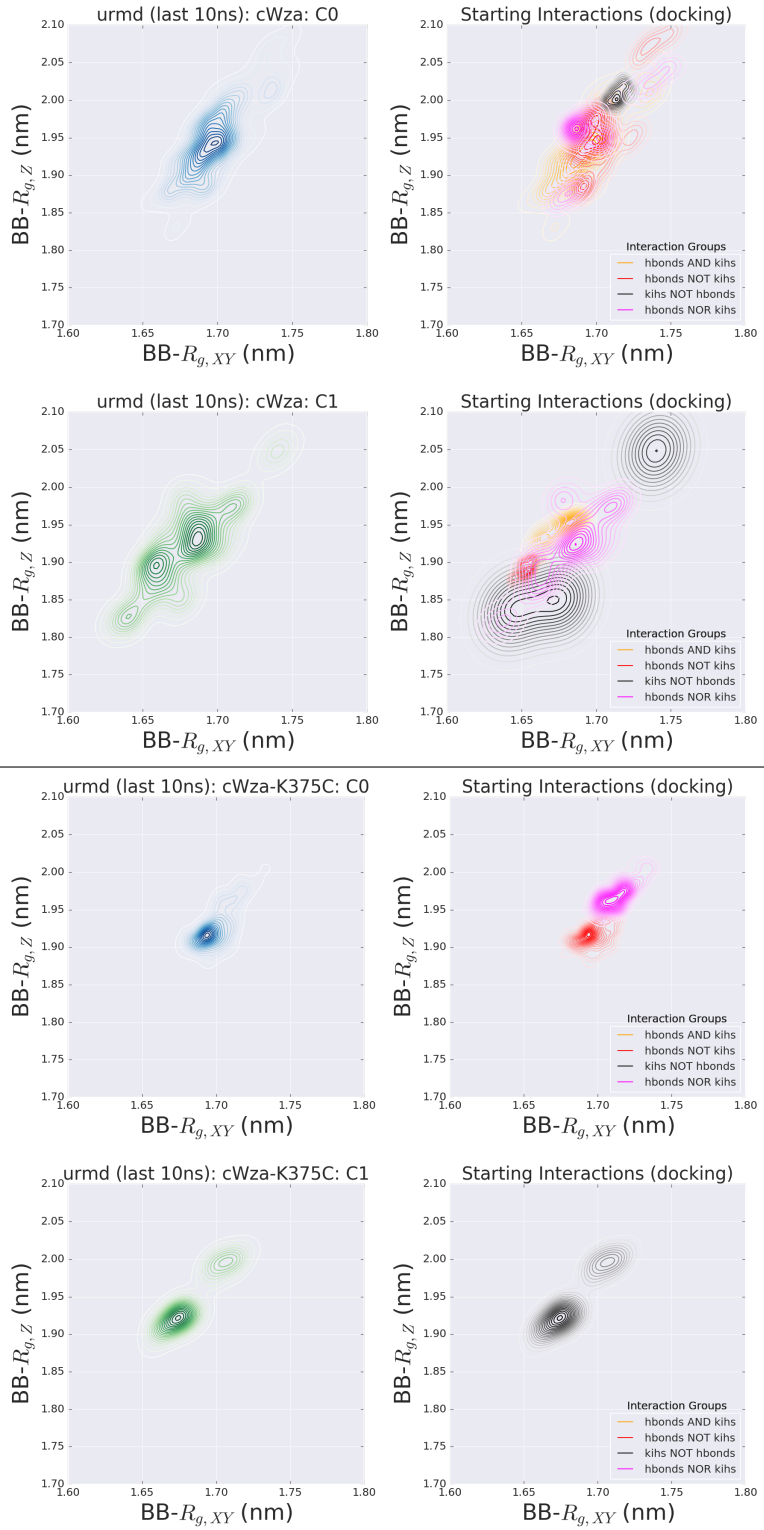


Figure 3.11: Breakdown of PDF contour-plots in Figure 3.8, according to identified interaction groups per docked conformation, per sequence. *Right column*: Individual PDFs from radius-of-gyration decomposition data per interaction group. *Right column*: PDFs for whole data per conformation. *Top Panels*: Data for **cWza** models with docked conformations C0 (Narrow) and C1 (Wide). *Bottom Panels*: Data for **cWza-K375C** models with docked conformations C0 (Narrow) and C1 (Wide).

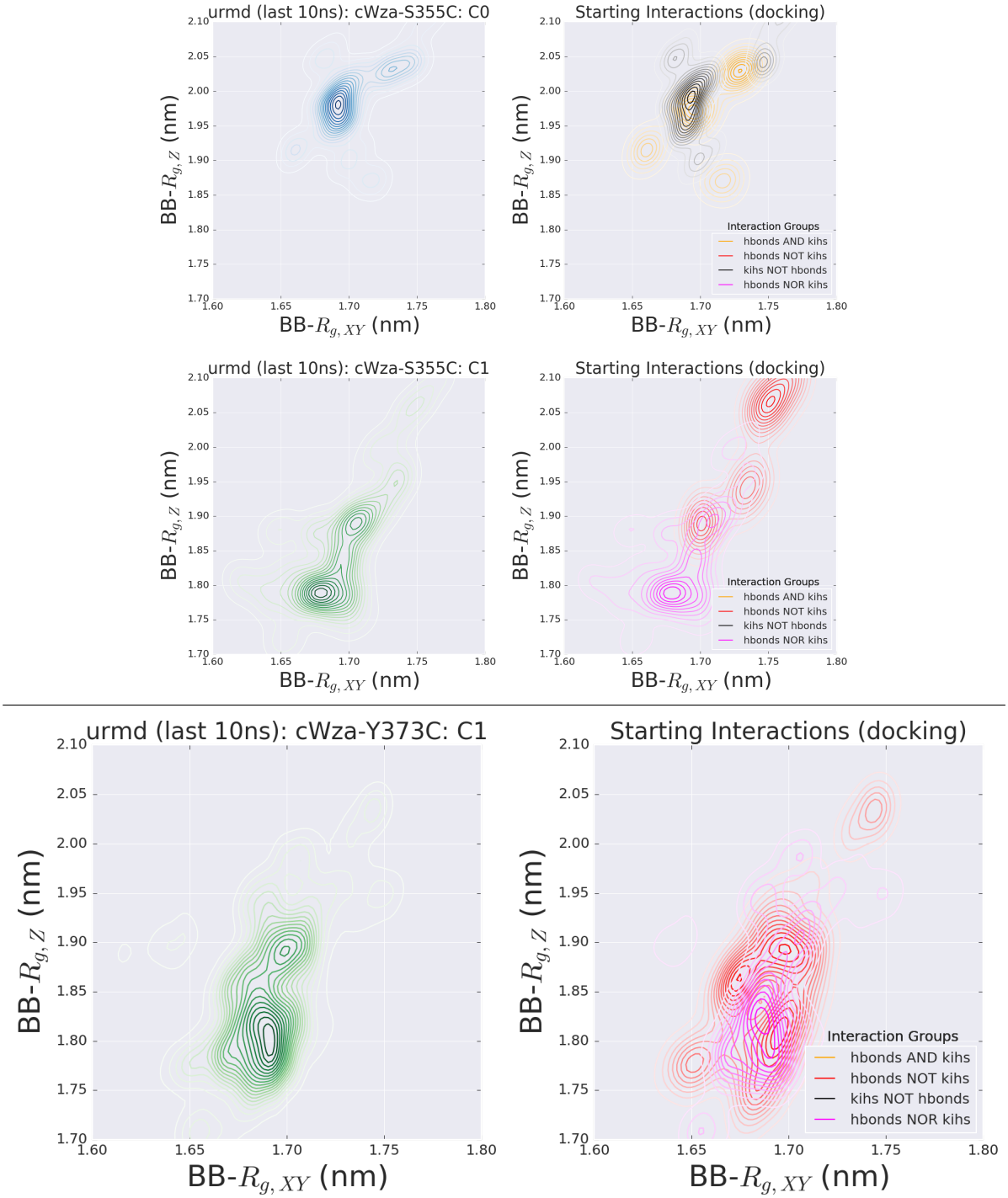


Figure 3.12: Breakdown of PDF contour-plots in Figure 3.8, according to identified interaction groups per docked conformation, per sequence. *Right column*: Individual PDFs from radius-of-gyration decomposition data per interaction group. *Right column*: PDFs for whole data per conformation. *Top Panels*: Data for **cWza-S355C** models with docked conformations C0 (Narrow) and C1 (Wide). *Bottom Panels*: Data for **cWza-Y373C** models with docked conformation C1 (Single).

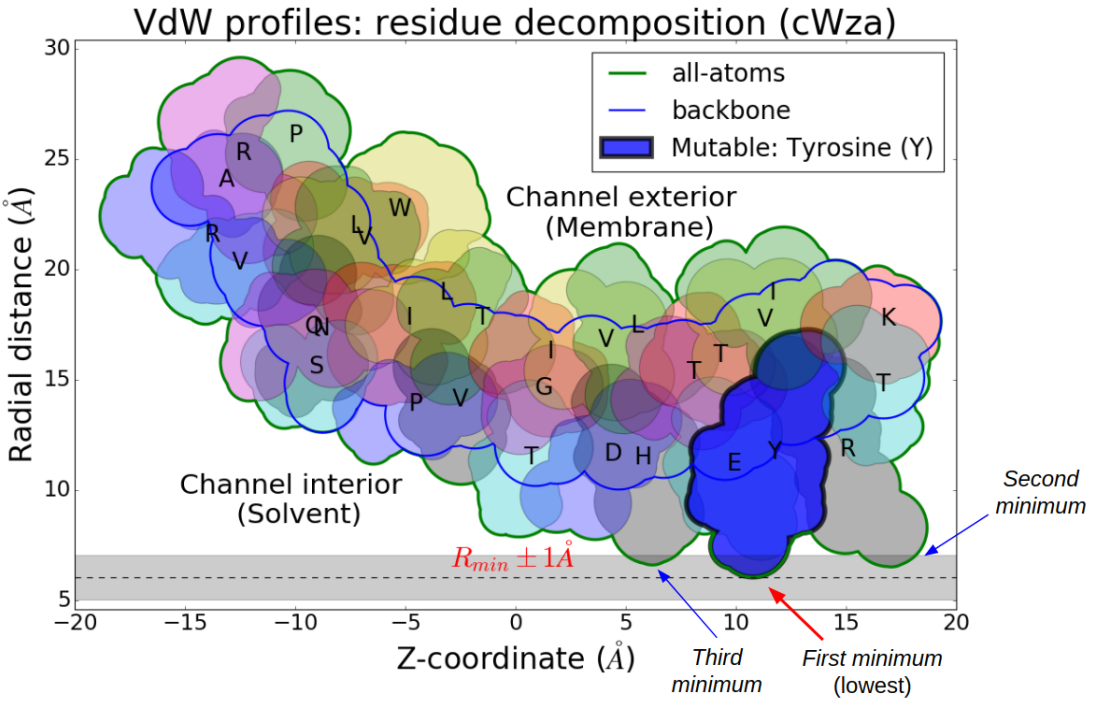


Figure 3.13: Minima of the Van der Waals (VdW) radial profiles can be used to identify residues located at the bottleneck of the interior of a channel; to be exposed to a solvent medium.

3.3.5 Tyrosine as a Key Residue at the Bottleneck

Tyrosine, a residue sitting at the bottleneck of docked channels with alternate conformations, seems to promote residue-residue interactions that induce correlations among H-bonds and KIHs participating residues, which when absent, uncorrelated H-bonds and KIHs take place in single conformation channels, which have a single conductance state in EXPTS. This observation will require further work to relate correlation between H-bonds and KIHs as promoters of alternate conformations, possibly due to enhanced flexibility between inter-chain sidechain-sidechain interfaces.

cWza : Bottleneck analysis For the majority of the docked models two main minima are found, with *Tyrosine* (Y, non-polar) and *Glumatic acid* (E, negatively charged) consistently and respectively identified as the first (~ 100% models) and the second minima (roughly 80% models). So, if Tyrosine is consistently the lowest minima in the VdW profile, and the distribution of $min(R_{VdW})$ shows a low-resolution bimodality, we can infer that fluctuations in Tyrosine rotameric conformations should be responsible for this difference for the distribution for $min(R_{prim})$. So, Tyrosine conformational noise spoils clear bimodality at the bottleneck.

Further, when looking at the mid-point position and length of the *bottleneck section*, we find that the position is unimodal, but the length of it, is bimodal, with extreme lengths of mostly ~ 1

or 10[$\mathring{A}$], that is, the outermost residues exposed to the solvent in the pore interior are either very close to each other or they set apart. Sidechain conformation can compensate for any backbone radial displacement or change in orientation that may narrow the VdW radial distance along the pore.

cWza-K375C : Bottleneck analysis Most models (~ 60%) show a single identified minima (within the 1[$\mathring{A}$] tolerance), which is mostly well localised along with the pore extension. *Tyrosine* (Y, non-polar) is again identified as the outermost residue within the pore interior (~ 100%). *Glutamic acid* (E, negatively charged) and *Arginine* (R, positively charged) show up as second minima again half of the time each. So, since $min(R_{VdW})$ and $\frac{\sigma}{\mu}$ imply that residues at the bottleneck section display preferential conformations and that its identity remains the same. This implies that separation in conformation arises from the restricted rotamers that *Tyr* is only allowed to adopt; in contrast to *Tyr* - and *Glu*- in cWza models.

cWza-S355C : Bottleneck analysis Like cWza-K375C, most models have a single minimum, however, the positioning of this disperses uniformly within a range of 5[$\mathring{A}$]; an effect due to the separation in unit orientation (tilt, interchain, and facing angles?). Interestingly, the identity of the outermost residue is not *Tyr* most of the time (only ~ 50%), in contrast to cWza and cWza-K375C. Also, nearly 30% and 20% fo the time *Arg* (R, positively charged) (next to *Tyr* in the sequence) and *His* (H, likely to be protonated), respectively, take over as the outermost exposed residues. So, the above suggests that the change in orientation may account mostly for the bimodality in the narrowness of the bottleneck and the fuzziness of the identity and hence physical properties of the bottleneck.

cWza-Y373C : Bottleneck analysis Unlike all previous mutants (cWza, cWza-K375C, and cWza-S355C), pores for these mutants show dominantly more than two minima, that is, usually multiple residues are equally exposed to the

- Arg (R, positively charged, ~ 60%)
- Variable identity; Arg (R, positive), Lys (K, positive), Asp (D, negative); all similar probability ~ 20 – 30%
- Cys (C, non-polar,~ 60%)
- Variable identity; Thr (T, polar), Asp (D, negative), Glu (E, negative), Cys (C, non-polar), Lys (K, positive); all similar probability ~ 20%.
- Glu (E, negative, ~ 70%)

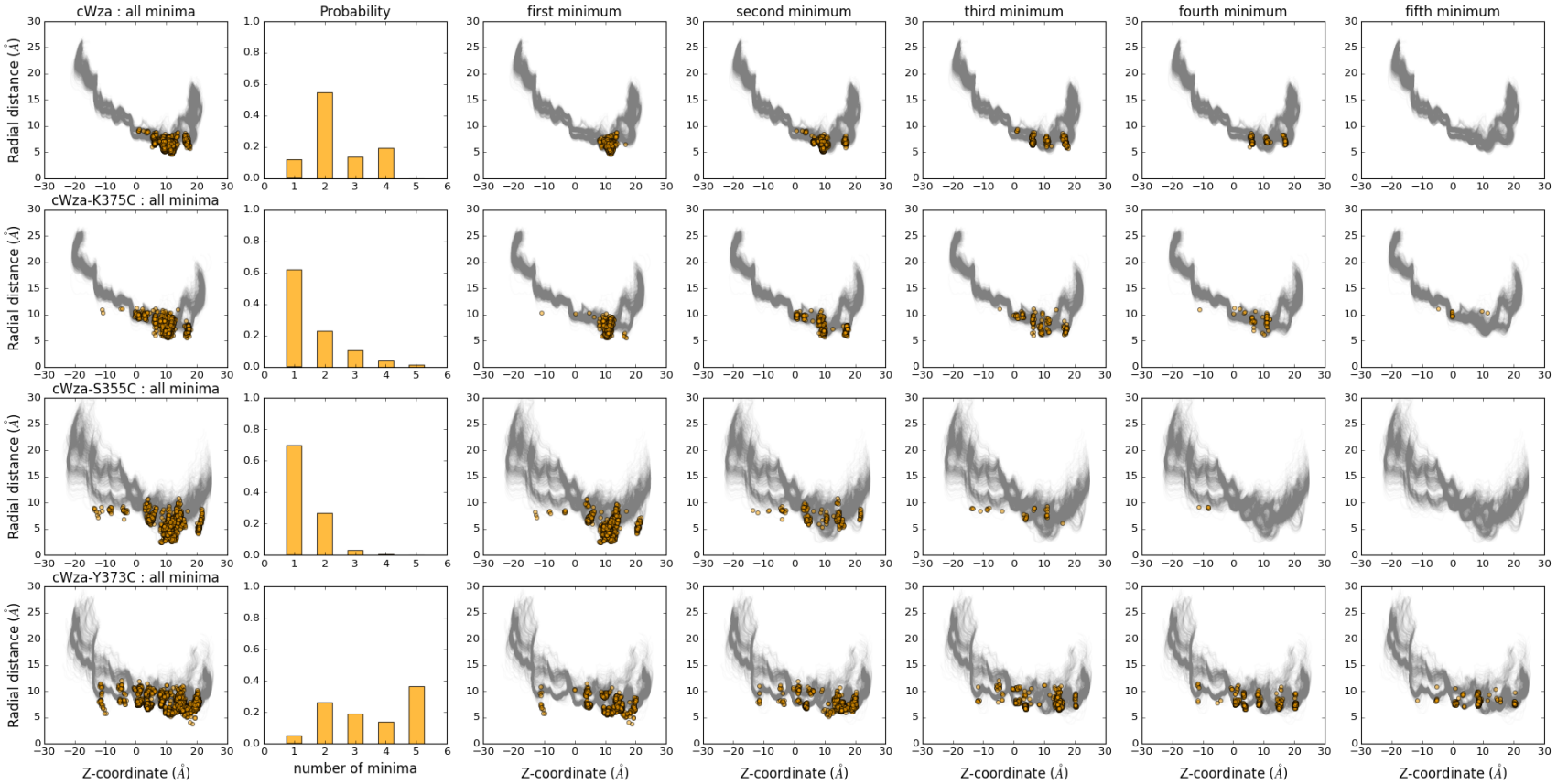


Figure 3.14: *First Column*: For all docked models per sequence (*Rows*), we show the overlapping of all positions (as yellow dots (·)) of all the bottleneck residues (*minima*) found within ±1Å around the lowest point of the VdW profile of every docked model. Note that from the VdW profiles more than one residue can be identified to sit at the bottleneck (*Second Column*). However, as some residues may be located deeper (or closer to the Axis of Assembly) in relation to the VdW profile of their chain, these can be ranked; with the deepest (at the narrowest point of a channel) as the *first minima* (*Third Column*), and with other residues ranked subsequently (*Fourth to Seventh Column*).

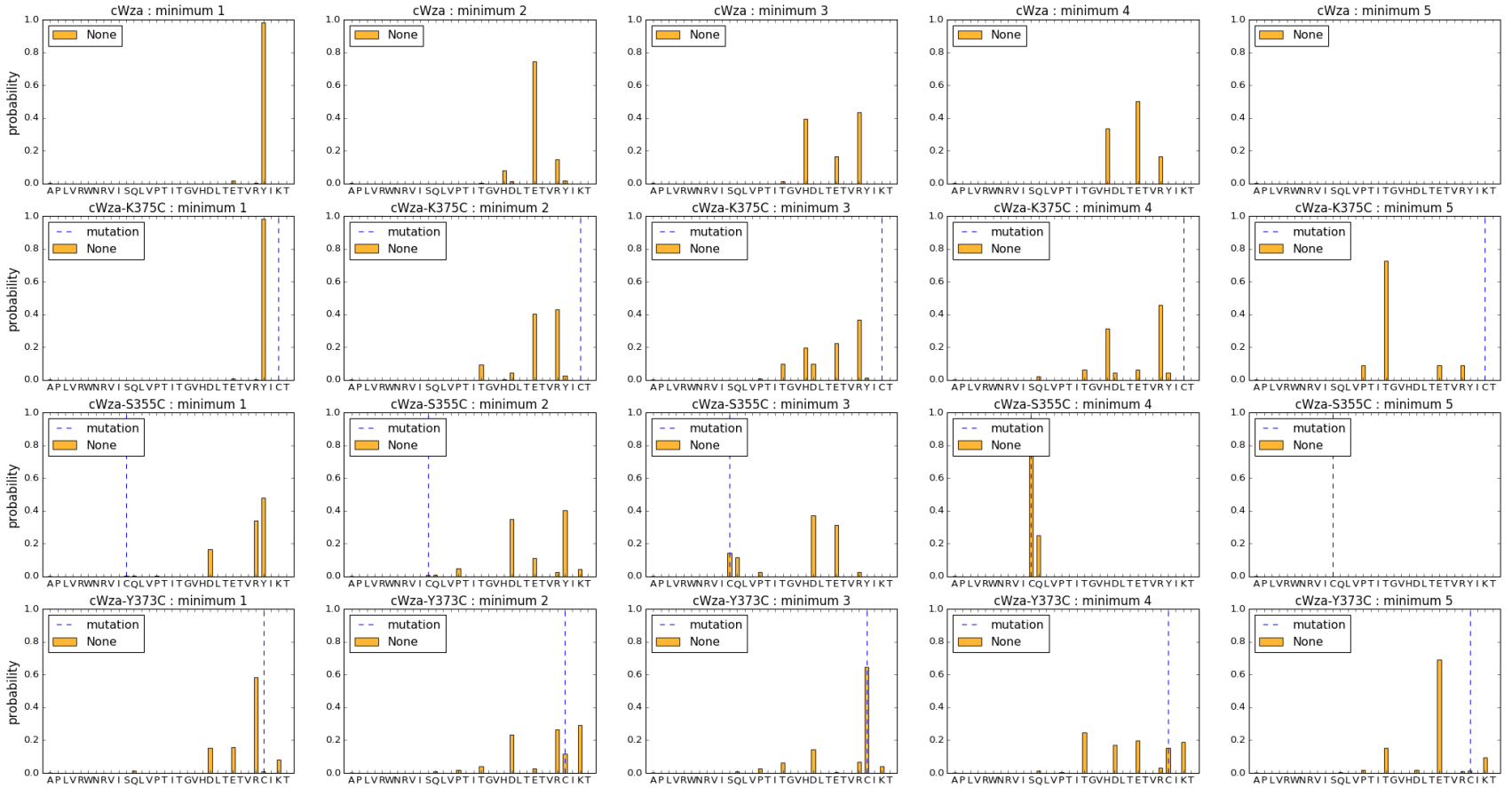


Figure 3.15: As the identity of bottleneck residues (*minima*) is variable, we related the previously ranked minima (Figure 3.3.5) to their residue identity in the peptide sequence of any cWza *Cys*-mutant channel (*Rows*). Here we show the propensity per residue to be at the bottleneck of any docked channel as any of the previously identified ranked minima (*Columns*). The positions of the mutated residues with Cysteine in each peptide sequence is indicated with a dotted line (- -).

3.3.6 Identification of Alternate Conductive States

Overall, cWza *and* cWza-Y373C *docked models respectively showed alternate (*Low *and* High*) and single conductance values in qualitative agreement with experiments. Also, estimated* solvent-accessible *pore dimensions provided indication of alternate and single conformations. Models with* cWza-K375C *and* cWza-S355C *sequences showed weak evidence of alternate (*Low *and* High*) conductance, unlike in experiments. Quantitative agreement between conductance values from models and from experiments was not observed. However, further tests suggested that cyclic symmetry of docked structures is key for observation of alternate conductance.*

As a first result, docked models corresponding to the peptide sequence *cWza* showed HOLE-estimated conductance values distributed according to a bimodal distribution. This means that cWza docked models show distinguishable populations of *Low-conductance* and *High-conductance* values. When fitted to a *Gaussian Mixture* with two modes, mean conductance values (G^{HOLE}) with included standard deviations are found to be

$$G^{HOLE}_{Low} = 0.93 \pm 0.097 \text{ nS}$$
$$G^{HOLE}_{High} = 1.267 \pm 0.076 \text{ nS}$$

Thus, models with conductance values falling under each of these modes were accordingly classified as either *Low-conductance* or *High-conductance*. In addition, when looking at the distribution of interface scores (RosettaMP) against the estimated conductance values for all models, Low-conductance and High-conductance models showed practically identical interface score distributions; Figure 3.16.

This result is qualitatively consistent with the experimental fact that channels formed by peptides with cWza sequence display low (G^{EXPT}_{Low}) and high (G^{EXPT}_{Low}) conductance values too (Sec., 1.1.2), although HOLE-estimated conductance values - and their ratio, are far from agreeing with experiments, that is,

$$G^{EXPT}_{Low} = 0.38 \pm 0.020 \text{ nS}$$
$$G^{EXPT}_{High} = 0.97 \pm 0.060 \text{ nS}$$
$$G^{EXPT}_{Low}/G^{EXPT}_{High} \approx 0.391 \neq 0.734 \approx G^{HOLE}_{Low}/G^{HOLE}_{High}$$

Further analysis of the docked cWza structures revealed that two dominant groups of structures can be distinguished by having HOLE-estimated narrow (≈ 8.75Å) and wide (≈ 10.25Å) internal *solvent-accessible* average radii, which positively correlate with the observed low and high conductance groups, respectively, but with a common pore length (≈ 46Å). See Fig., 3.18. This shows that docked structures with low and high conductance are indeed structures with

overall contracted and expanded inner volumes, which should be a consequence of alternate *conformations* that result from the different backbone and sidechain atomic arrangements.

By contrast, for the remaining sequences showing alternate channel conductance in experiments, docked structures with *cWza-K375C* sequence showed a weak dual separation in conductance by contrast to cWza-sequence models, with the interface score distribution showing no separation either, Figure E.1. A clear lack of separation between pores with narrow and wide internal radii was observed too (Figure E.2), although pore lengths displayed a preferential value like cWza models (Figure E.3). By contrast, *cWza-S355C* pores displayed a broad dispersion in their conductance values distribution, without clear dual conductance separation; Figure E.4. Similarly, the distribution of pore radii showed no clear dual separation between structures, but noticeably there was separation of structures in pore length, but all with similar conductance values. See Figures E.5 and E.6.

On the other hand, for models with *cWza-Y373C* sequence, the scatter plot of estimated conductance against interface score values, most models showed a single dominant conductance value, with a wide range of score energies though; Figure 3.17. Recall that assembled pores for this sequence show ohmic conductance in electrophysiology, which is a sign of a single steady structure (Subsection 1.1.2). Further analysis revealed that one group of structures with similar pore internal dimensions can be effectively identified; Figure 3.19.

Although docked structures with cWza-sequence show a clear dual conductance separation consistent with experiments, this separation will be a consequence of the imposed symmetry constraint. To test this, two independent post-docking assays revealed that under *symmetry perturbation*, that is, when atomic coordinates in all chains of a docked structure relax independently (RosettaMP Relax), the dual conductance separation vanished showing a collapse to a one preferential conductance value, both when relaxing 1) *sidechain atoms only* - which are the most external atoms in the interior of a protein exposed to solvent, and 2) both *backbone and sidechain atoms*; see Figures F.1. However, RosettaMP perturbations are performed under artificial atomic movements and also assuming an implicit lipid model for energy scoring. To introduce more realistic perturbations, we also performed assays with EMMD[10] assays with a sample of cWza docked models embedded in membrane and solvent, which showed that the initially observed difference in low and high conductance is considerably decreased after HOLE-conductance values of post-EMMD models were analysed. See Figures F.2.

Overall, the above tests suggest that symmetry is a key constraint in order to observe alternate conductance of cWza octameric models. This may not be surprising, as it has been indicated elsewhere that in the membrane environment protein complexes often assemble as symmetrically, where symmetry can enhance structural stability and also lead to added flexibility, which is crucial for conformational changes related to function [3, 29, 92, 123]

[10]Details of MD setup are identical to those described in Subsection 3.2.5.

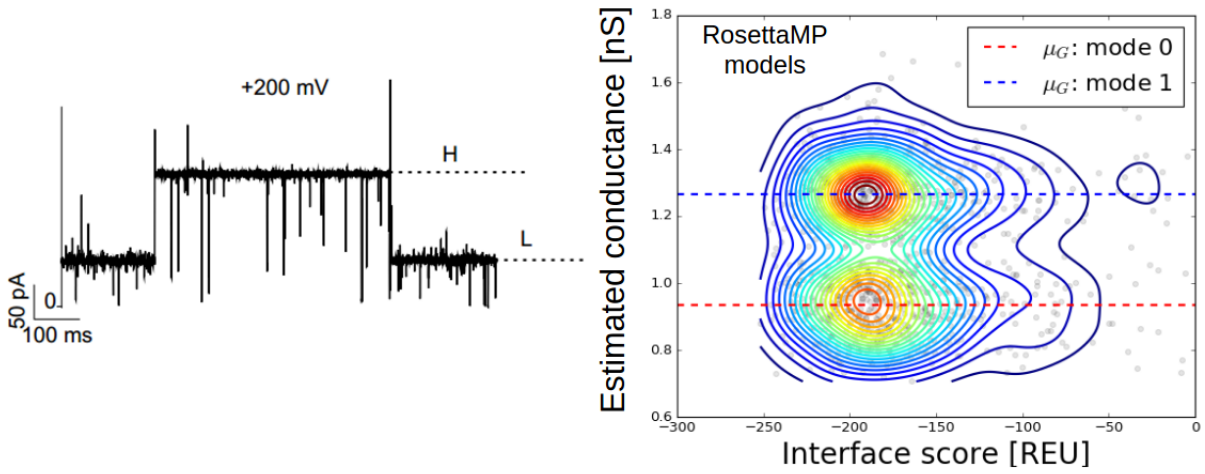


Figure 3.16: *Left*: Electric activity of cWza-sequence ion channel, recorded at *+200 mV* in electrophysiology; L and H represent respectively the Low- and High-conductance activity of the assembled channel (dual conductance) [64]. *Right:* Scatter plot of *interface scores* (strength of association) against *numerically estimated conductance* for a 1000 optimal 8-unit models with cWza-sequence, found using RosettaMP via a symmetry-constrained docking protocol. Contour plots show the density of scattered data, with red, the densest, and blue, the least dense accumulation of data-points. *Two conductance states with similar energy distributions are shown by most docked structures*

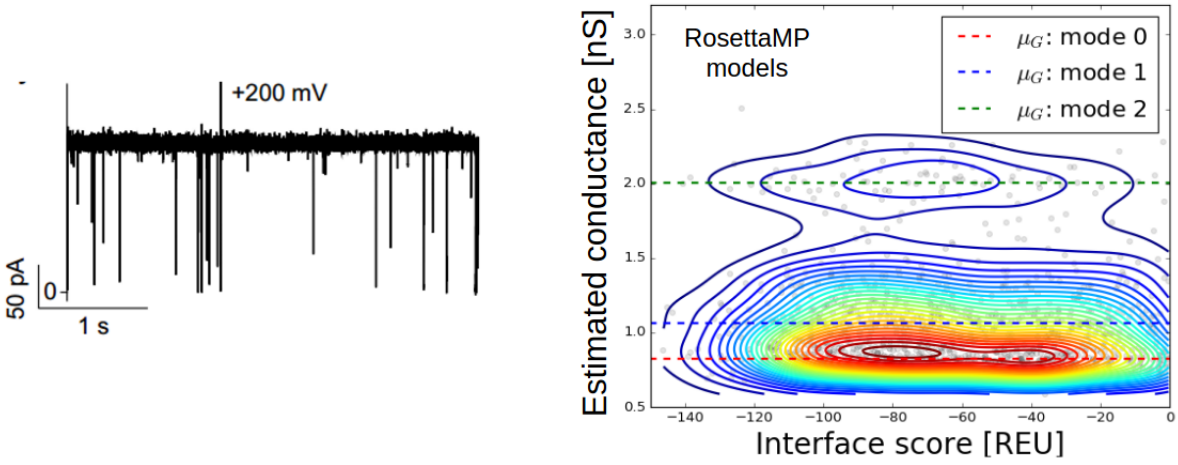


Figure 3.17: *Left*: Electric activity of cWza-Y373C-sequence ion channel, recorded at *+200 mV* in electrophysiology; only one single conductance is observed, even when varying the applied voltage; $|\Delta V| \leq +200mV$ [64]. *Right:* Scatter plot of *interface scores* (strength of association) against *numerically estimated conductance* for a 1000 optimal 8-unit models with cWza-Y373C-sequence. Similarly, contour plots show the density of scattered data, with red, the densest, and blue, the least dense accumulation of data-points. *This shows the existence of practically one single conductance state shown by most docked structures.*

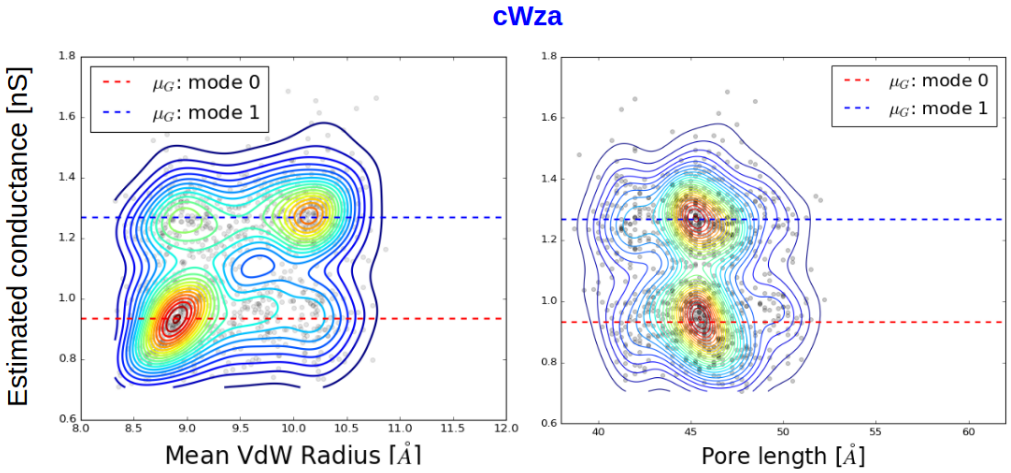


Figure 3.18: Scatter and contour plots from data comparing pore dimensions against estimated conductance for all cWza-sequence octameric docked models. *Left*: The *mean Van der Waals* radius of a channel's pore is compared against its estimated conductance for all docked models. *Right*: Similarly, for all pore length values. *Plots suggest that low and high conductances can be effectively explained by narrow and wide channel pores alone.*

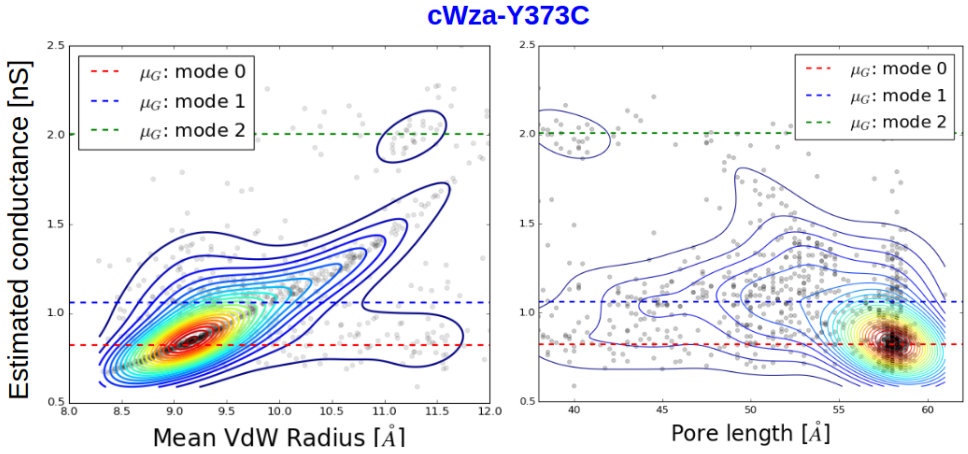


Figure 3.19: Scatter and contour plots from data comparing pore dimensions against estimated conductance for all cWza-Y373C sequence octameric docked models. Plots are analogue to those in Fig., 3.18. *We can see that for most models, their conductance will associated to a single set of effective pore dimensions.*

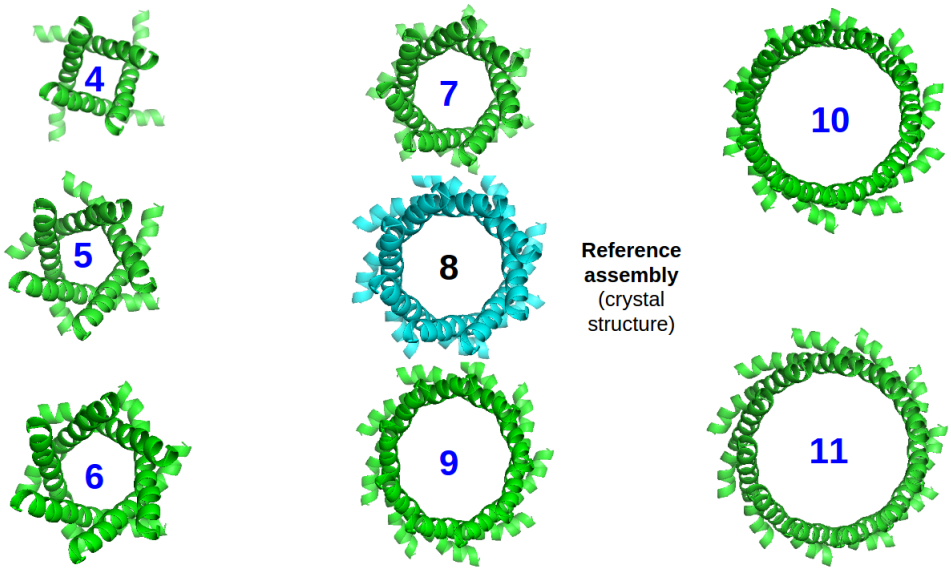


Figure 3.20: *cWza Oligomers*: Models docked out of 4 to 11 peptide units with identical *cWza* sequence. Note that 8 is the number of units in the crystal structure of native *Wza-D4*

3.3.7 *Oligomerisation* as an Alternative Conductance Changing Mechanism

> *Although oligomerisation could alternatively explain the appearance of* Low *and* High *conductive states for cWza* Cys*-mutant models, overall docked models with* $n = 4-11$ *units could not provide clear evidence for preferential oligomeric states with the distinctive association. Also, although oligomers with* $n = 5-8$ *units do not need to rely on the robust backbone and sidechain symmetry to display separable conductance states, no clear preference could be defined among oligomers to explain Low and High conductance, at least at this stage.*

Octameric docked structures with *cWza Cys*-sequence can to some extent reproduce experimental dual and single conductance states, however, these are sensitive to symmetry perturbation. As suggested in Section 1.1, dual conductance of channels can be alternatively explained by *oligomerisation*, which relates changes in conductance to changes in the number of associating units making up a channel. Also, although octameric peptide association is suggested by the native crystal structure of Wza-D4 and from *blocking experiments* with cWza Cys-mutant channels, these are not a clear cut given the lack of high-resolution structures [64].

Here we introduce results examining whether *olgomerisation* can alternatively explain experimental dual conductance; so that assemblies with two preferential numbers of units can be linked to low and high conductance states observed for some cWza Cys-mutant channels, thus, excluding the reliance on the backbone and sidechain symmetry, as for octameric models.

For all cWza *Cys*-mutant sequences, using our docking protocol once more (Section 3.2.1) we

generated 1000 docked models per sequence, per case of non-octameric assembly with $n = 4-11$ assembling peptide units - except for $n = 8$- and corresponding cyclic C_n-symmetries. Master units for docking were identical to the ones used for octameric models, the only changes in the setup were the corresponding symmetry files to initialise docking. HOLE-estimated conductance values were determined for all docked models. Also, *total energy scores per unit* were obtained, i.e., the total energy due to all pairwise interactions, normalised by the number of units, $n = 4-11$, making up an assembly.

As a first approach, we analysed the average *interface energy scores per unit* for all *oligomeric models* - according to their number of units n, to determine whether oligomeric states showed a distinctively strong association. First, for sequences showing experimental dual conductance; *cWza, cWza-K375C, cWza-S355C*, these did not show any clear distinction that favoured a single or two oligomeric states, which could be potentially associated to Low and High conductance (Figure G.5). But it was observed as a common trend that an increasing number of peptides makes scores increase; which implies that large chain assemblies are energetically less favourable. Similarly, results for the single conductance *cWza-Y373C* docked models did not show any clear sign of a preferential oligomeric state either (Figure G.5). Thus, supporting evidence towards preferential oligomeric states could not be seen alone in energy scores of docked models here, even for octameric assemblies ($n = 8$).

Next, despite the above, we subsequently attempted to identify candidate *oligomeric structures* that could explain experimental dual conductance of channels with *cWza, cWza-K375C, cWza-S355C* sequences based on conductance-energy relation. Models were first reduced to oligomers with $n = 5-8$ units alone. On the one hand, models with $n = 4$ units were discarded as non-conductive, since their surface-accessible internal minimum diameters were lower than the average diameter of a water molecule (~ 3.2Å [32]) (Figure G.6). On the other hand, models with $n = 9-11$ units were discarded too, as these showed a range of unfavourable scores - close to zero or positive (Figures G.1 and G.3); moreover, electrophysiology experiments suggest that cWza *Cys*-mutant channels cannot have more than 8 units [64]. Similarly, *cWza-Y373C* oligomers with $n = 5-8$ units were only considered. Then, contour plots for conductance and energy scores were obtained for docked and post-docking Rosetta relaxed models - like in Figures 3.16 and 3.17, to compare how conductance and energy shift after symmetry perturbation. The most relevant results are summarised below. Sequences *cWza, cWza-K375C, cWza-S355C* were labeled as *Type-2*, and *cWza-Y373C* was label as *Type-1* for the analysis below.

For docked models, as an important observation (Fig., G.7), Low and High conductance states are a reproducible feature of 7-unit oligomers with sequence type-2; like for octameric models. Single conductance was shown consistently for all oligomers with sequence type-1, however, dispersion in energy scores does not allow to clearly distinguish whether oligomers will have a stronger association per unit in comparison to octameric channels. Also, some oligomers with 7 units displayed conductance values and energy scores comparable to those for octameric models,

as shown by the overlapping between contour plots; sequence type-2. Thus, the value of low conductance by octameric models will be also displayed by 7-unit models with High conductance. However, a marginal difference in energy scores favours alternate conductance 7-unit structures over octameric ones.

As an attempt to disambiguate conductance between different oligomeric structures, non-symmetric relaxation assays were performed to test the robustness of dual conductance observed for oligomers with 7 units. Contour plots in Fig., 3.21 show results for all docked structures after *all-atom* (backbone and sidechains) relaxation; for comparison, results for octameric models are featured too. First, we note that for sequence type-2 models dual conductance vanished for 7-unit models, as well as for octameric ones - as pointed out before, moreover, their energies per unit were similar. Also, overall conductances values from 7- and 8-unit models were separable. For sequence-type-1 though, distinction by conductance for 6- to 8-unit relaxed models becomes unclear, likewise by their energy. A set of similar results is also shown when only non-symmetrically relaxing sidechain atoms while keeping backbone chains restrained; Appendix Fig., G.8. This suggests that it is enough to perturb the local symmetry of atoms in close contact with solvent in the interior of a pore to make dual conductance separation vanish for all docked models.

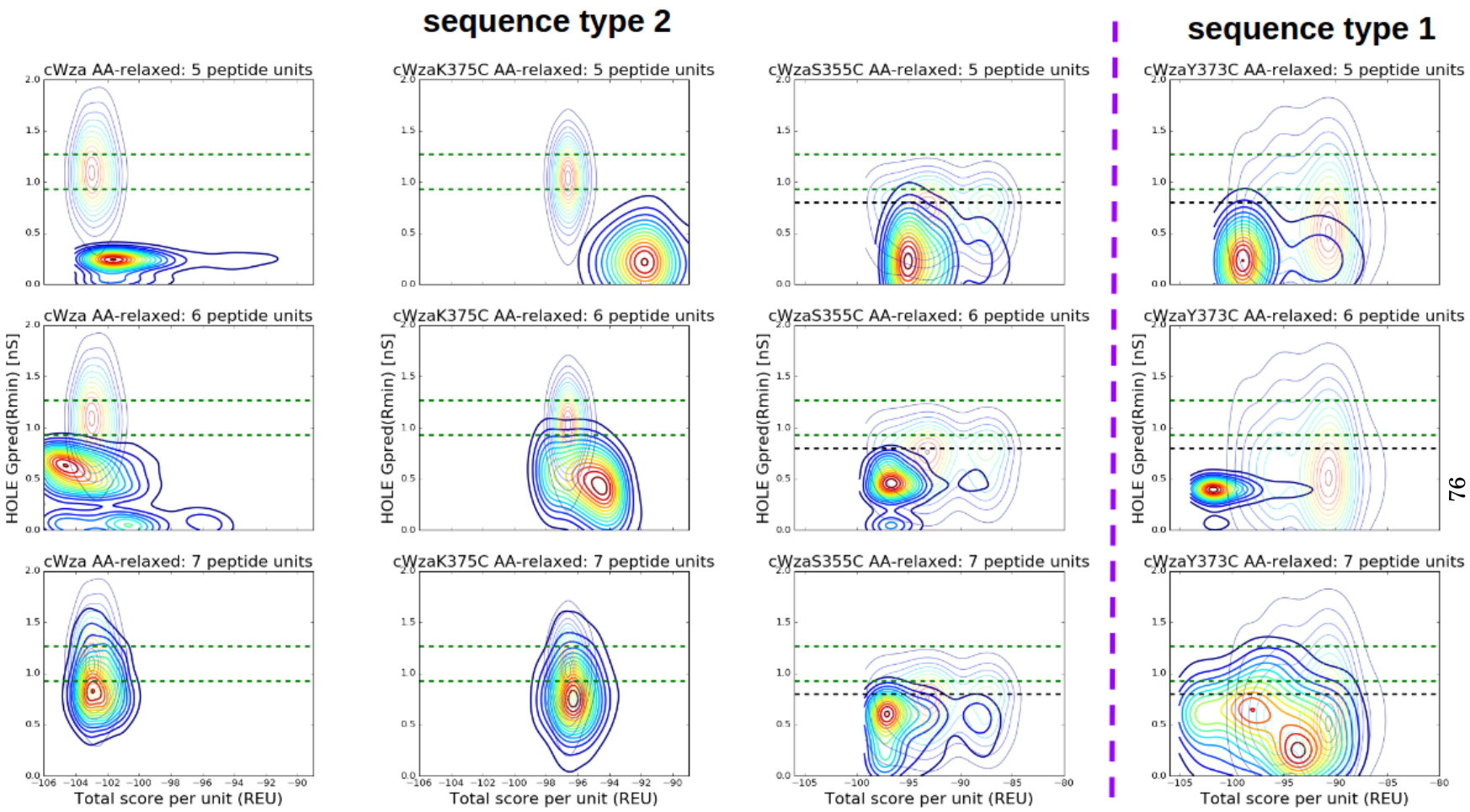


Figure 3.21: Contour plots from scatter plots of *total energy scores per unit* against estimated conductance values for all docked model after *all-atom(AA) non-symmetric relaxation*, per sequence, per number of chains; $n = 5-7$. Again, colours represent density of scattered data, with red, the densest, and blue, the least dense accumulation of data-points. All contour plots in transparent correspond to data from octameric models; included for comparison. *Note that dual conductance vanishes after perturbing symmetry for all atoms.*

3.4 Conclusions

So far, we have developed a framework for conformation prediction from pools of candidate structures generated employing Symmetric Docking according to peptide sequences, tested with cWza *Cys*-mutant ion channels. Conformations consistent with those expected from experiments were distinguished by finding the most likely inner pore dimensions of docked channels, via a geometric analysis of all docked structures per sequence. Thus, their dynamics were simulated for 100 ns for randomly selected modes per identified conformation, finding that end peptide assemblies did separate according to alternate conformations, but according to their inertia. Initially identified conformations from docked structures related poorly to end conformations in dynamic simulations, however, it was found that specific groups of starting peptide-peptide Hydrogen-bonds and Knob-Into-Holes identified from structures corresponding to initial conformations, related better to end conformations in the dynamics. Further analysis of docked conformations suggests that *Tyrosines*, identified to sit at the bottleneck of channels, when swapped with *Cysteines*, these seem to allow peptide-peptide interactions holding the structure fixed, which may not take place otherwise, as a consequence of side-chain rearrangements due to mutation.

Also, alternate conductive states were identified from docked models, with a numerical estimation of conductance values using the HOLE programme. Although estimated conductance values did not agree with experimentally determined ones, alternate conductance estates could be resolved from data. However, conductance estates were found to be dependent on the imposed cyclic symmetry, which when perturbed this led to a loss of distinction between conductance states. Last, for completion, we generated docked models with variable number of assembling units, with 4 to 11 peptides, to see how docked oligomeric models compared in estimated energy and conductance, to find out whether octameric assemblies were preferred over others, and whether conductance changes could be explained alternatively through oligomerisation instead of conformational changes. No clear evidence was found favouring an oligomeric state over others for all docked channels, however, interestingly dual conductance was found to be a feature for lower-order oligomers (< 7 units). Overall, no clear evidence favours oligomerisation over conformational changes. This will be however subject for future research.

CHAPTER 4

COMPUTATIONAL ELECTROPHYSIOLOGY OF PEPTIDE-ASSEMBLED CHANNELS AND CONDUCTANCE ESTIMATION

*As an alternative to HOLE, we explored the use of MD simulations to estimate the conductance of and to study in better detail the conduction mechanisms across individual peptide-assembled channels. Thus, in this chapter, we will introduce and discuss results showing the application of a proposed protocol for MD simulation to lipid-embedded peptide-assembled channels; as similarly done for docked structures (*Chapter 3*), but featuring an externally applied electric field to mimic the scenario in an electrophysiology experimental layout. Caveats of the protocol regarding artefacts that can limit accuracy and reproducibility of results will be discussed throughout the chapter. The list of tested membrane proteins includes a model of the* cWza-Y373C *channel, in addition to Hexameric (6 peptide units) and Heptameric (7 peptide units) models of* de novo *Coiled-Coil Trans-membrane (*CCTM*) and α-Hemolysin Coiled-Coil (*aHL-CC*) hybrid peptide-assembled channels.*

4.1 Simulating Ion Permeation in MD with a Constant Electric Field

Ion permeation through protein pores can be driven either by an existent *electrochemical gradient*, like in the cell native environment, or an externally controlled *electric potential difference*, like in experiments. Unsurprisingly, in MD simulations, conductance properties of protein channels are generally studied under conditions that approximate these two scenarios.

Simulation of an electrochemical gradient requires an existent *charge imbalance* -or difference in net charge- between opposite sides of the membrane. In practice, periodic boundary conditions

(PBCs) prevent this, since opposite sides of a single membrane are actually connected, and hence charge imbalance is not possible. A solution to this issue is the simulation *twin parallel bilayers*; where charge imbalance is imposed by discrete ion pair swaps between bilayer compartments. This has proven to lead to biologically realistic transmembrane potentials, and when embedded pores in each bilayer are present, ion flux is driven by the controlled electrochemical gradient, and hence transport can be studied under native-like conditions [88]. However, this approach generally demands considerable simulation time and storage in comparison to single bilayer simulations. An alternative solution is the use of a *combined cell*, made up of a vacuum cell stacked on top of another cell containing the single solvated membrane. This slightly less demanding option preserves imposition of charge imbalance since conductive media are separated by a highly non-conductive one. Likewise, this approach has been used in MD studies of protein pores [61].

Unlike simulations subject to charge imbalance, another widely used approach relies on the imposition of an external potential difference or a *target voltage*, ΔV_0, between opposite ends of a simulation cell. This approach only requires one single bilayer, which leads to reduced simulation time and storage. Moreover, since ΔV_0 can be controlled like in experiments, simulation results are easy to compare with single-channel recording data. This not easily possible when imposing charge imbalance, since voltage can vary considerably by hundreds of millivolts upon one single ion pair exchange. Also as ion flux becomes unsustainable when charge balance is restored due to permeation. To establish the electric potential difference parallel to the membrane to drive permeation, say along the Z-axis, a linear potential gradient, $V(z) = z \cdot \Delta V_0 / L_z$, is imposed across the whole simulated system, hence leading to a constant *electric field* $\mathbf{E}_0 = (0, 0, -V_0/L_z)$ perpendicular to the bilayer plane. Thus, the net force exerted on any single charged particle is simply increased by a constant force $q\mathbf{E}_0$, which is continuous across the periodic boundaries. The simplicity and advantages of this approach have allowed this to be used not only in studies of natural and synthetic ion channels, but also in peptide insertion, electroporation, and translocation of DNA and large biomolecules through solid-state nanopores [63]. Moreover, a rigorous mathematical framework showing the equivalence of this approach to conditions found in experimental settings is already established [87].

In this work, we adopt the *constant electric* field approach to simulate the transport of ions and characterise their net rate of transport (*conductance*) across *de novo* protein models currently available, described in (SEC). However, before any attempt, there are a number of practical considerations relevant to ion flux production and measurement that should be seriously considered to ensure the reliability of simulation results. Next, we summarise some of the most relevant ones.

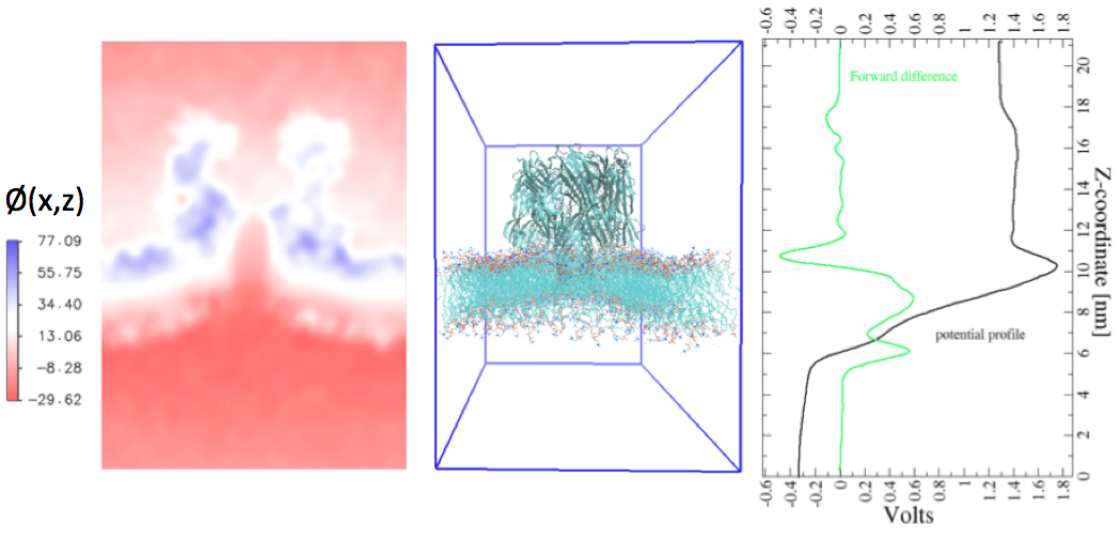


Figure 4.1: *Left*: Electrostatic potential map of a plane XZ crossing the centre of α-hemolysin in DPPC bilayer (*Middle*). Scale bar shows dimensionless potential values (normalised by k_BT/e; where k_B is Boltzmann constant, e fundamental charge, and temperature T = 300 K). *Right*: Average potential profile along Z-axis. Note how the potential tends to a constant value near the ends of the cell (*black curve*), which implies a zero electric field away from the membrane and the protein (*green curve*).

4.2 Methods

4.2.1 Electric potential maps and profiles

As mentioned previously, determination of the *total electrostatic potential*, $\phi(\mathbf{r},t)$, throughout a simulated systems is necessary to verify stationarity and also to understand how different parts of our system contribute to the local electrostatics. Thus, in order to map the total electrostatic potential at any position within our simulation cell, $\mathbf{r} \in L_x \times L_y \times L_z \subset \mathbb{R}^3$, at any time, t, of simulated dynamics, in principle one requires to solve Poisson equations in the vacuum, featuring point densities due to all charged particles (q_i, $i = 1,2,\cdots,N$) within the cell ($\mathbf{r}'_i \in L_x \times L_y \times L_z$, $i = 1,2,\cdots,N$), plus their *periodic images* when assuming Periodic Boundary Conditions (PBCs)

$$\nabla^2\phi(\mathbf{r},t) = 4\pi \sum_{\mathbf{n}\epsilon\mathbb{Z}^3} \sum_{i=1}^{N} q_i \delta\left(\mathbf{r} - (\mathbf{r}'_i(t) + \mathbf{nL})\right) \tag{4.1}$$

where $\mathbf{nL} = (n_xL_x, n_yL_y, n_zL_z)$, with n_x, n_y, and n_z integers, and L_x, L_y, and L_z the respective lengths of the cell edges along the axes, X, Y, and Z.

We approximate $\phi(\mathbf{r},t)$ using the *PMEPot* plugin of VMD (Version 1.9.2) [44], which solves Eq., 4.1, through Fast Fourier Transform (FFT) on a grid that partitions the central simulation cell ($\mathbf{n} = (0,0,0)$), but approximating point charge densities, $q_i\delta(\mathbf{r} - \mathbf{r}')$, as spherical Gaussians,

$q_i(\beta/\sqrt{\pi})^3 exp(-\beta^2|\mathbf{r}-\mathbf{r}'|^2)$; which return point charge values, q_i, upon integration, and for which the *Ewald factor*, β, controls their sharpness. Thus, the computed potential $\phi_{PMEPot}(\mathbf{r},t)$ reasonably approximates $\phi(\mathbf{r},t)$. Despite actually representing the *long-range contribution* to the total potential, differing just by a constant given by the *short-range term*, as given in the Particle-Mesh-Ewald algorithm [25].

Provided coordinates and charge values of all particles (PQR format file), we can get either *instant* (provided a single simulation frame) or *average maps* (provided a trajectory) of the electrostatic potential across the simulation cell. For instance, in Fig., 4.1 *(left)*, we show the YZ view of an average electrostatic potential map, obtained from averaging over the last 1 ns (100 frames) of 20 ns of dynamics of *α-hemolysin* embedded in DPPC (1,2-dipalmitoyl-sn-glycero-3-phosphocholine) bilayer (*Fig.,* 4.1, *middle*) and solvated in 1M KCl, at $\Delta V_0 = +1.2V$, with the applied *electric field* $\mathbf{E_0} = (0,0,E_0 = -\Delta V_0/L_z)$, perpendicular to the membrane plane.

From Fig., 4.1 (*Left*), we can see, qualitatively that the electric potential main drop happens across the membrane region, which may be a signature of solvent relaxation. However, to observe this more clearly, we computed the *electric potential profile*, $\bar{\phi}(z)$, which has been introduced previously, as a way to observe, on the one hand, the correct reproduction of the total electric potential, and on the other hand, to estimate stationarity of this as well [2]. Here, the potential profiles represents a spatial average across the XY plane, with varying position z, that is

$$\bar{\phi}(z) = \int_{L_x \times L_y} \phi(x,y,z)/(L_x L_y) dx dy \tag{4.2}$$

which in its discrete form, is just a sum over grid points in a plane parallel to XY plane.

To estimate stationarity of $\phi(\mathbf{r},t)$, one looks at the profile given by $\{(z,\bar{\phi}(z)) : 0 \le z \le L_z\}$, which in steady-state we expect to satisfy simultaneously the limiting conditions

$$\bar{\phi}(z = L_z) - \bar{\phi}(z = 0) \to \Delta V_0 \tag{4.3}$$

$$\frac{d\bar{\phi}}{dz}(z \to 0^-);\ \frac{d\bar{\phi}}{dz}(z \to L_z^+) \to 0 \tag{4.4}$$

On the *right* of Fig., 4.1, we show the corresponding average profile, $\bar{\phi}(z)$, from the previously determined electric potential map of *α-hemolysin* on the *left*. Note that both, the average electric potential long Z ($\bar{\phi}(z)$) and the average total electric field ($E = -d\bar{\phi}/dz$) satisfy the above conditions Eq., 4.3 and 4.4.

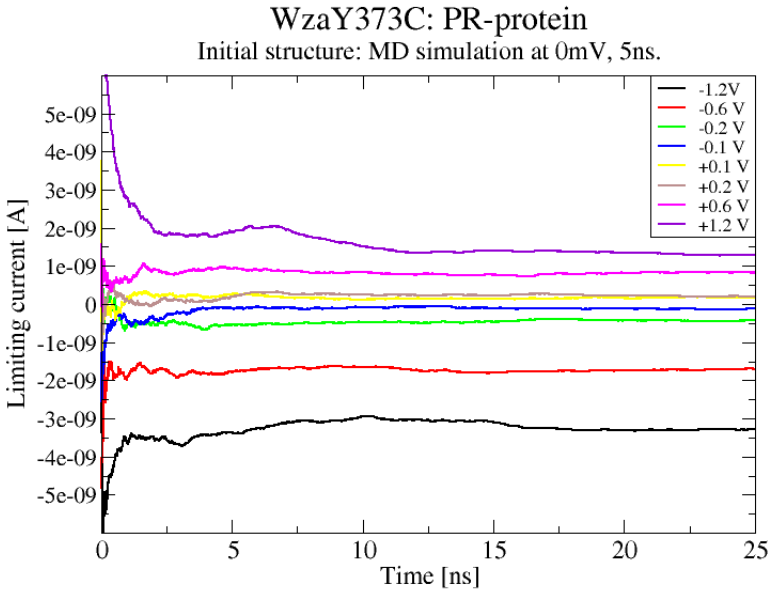


Figure 4.2: Limiting ion current, $I_{lim}(t)$, for various simulations of position-restrained cWzaY373C. Simulations at target voltages $0 < \Delta V_0 < +1.2V$, are obtained by taking the last state of the simulation at $+1.2V$, as their common initial state, that is, the high applied voltage is suddenly dropped to ΔV_0. Similarly, for simulations at $\Delta V_0 < 0$. Note how the limiting current approaches an asymptotic value in less than 5-10 ns of dynamics. Beyond that, the ion current can be considered as steady.

4.2.2 Ion current calculation

Determination of ionic current is key to keep track of how current settles down to become steady, so that membrane pore conductance (G) can be computed reliably in simulations; potentially be compared against experimental estimates. In MD studies, an ionic current is usually computed instantaneously at some time $t + \Delta t$, either by counting permeations of ions crossing the pore volume per unit of time [87] or based on their average velocity within the pore [2]. Here we adopt the velocity-based approach for ease of coding. Thus, the *instant charge* transported across the pore between times t and $t + \Delta t$, can be defined by

$$Q(t+\Delta t) = \frac{\Delta t}{d} \sum_{channel} q_i \dot{z}_i(t+\Delta t) \tag{4.5}$$

with q_i the charge of the *i-th* ion within the protein during the observed period $[t, t+\Delta t]$; $\dot{z}_i(t+\Delta t)$ the velocity of the ion along the direction of permeation (Z-axis), estimated by $\dot{z}_i(t+\Delta t) = [z_i(t+\Delta t) - z_i(t)]/\Delta t$; and d the length of the protein pore volume along the direction of permeation, i.e., the distance along Z, that an ion needs to travel to permeate across the pore.

Usually quantifying the *cumulative current* and the *limiting instant current* turns out to be more useful, as these turn data provided by Q into a method to estimate average ionic current, and hence conductance, but also into a diagnostic tool of flux stationarity. These are respectively defined by

$$I_{cum}(t_{obs}) = \sum_{t_0 \leq t' \leq t_{obs}} Q(t')/\Delta t \tag{4.6}$$

$$I_{lim}(t_{obs}) = \frac{\sum_{t_0 \leq t' \leq t_{obs}} Q(t')}{t_{obs}} \tag{4.7}$$

The cumulative current is just a a sum of the *instant current* values ($Q(t')/\Delta t$) at every time point, up to some *observation time* t_{obs} of the dynamics, starting at some time t_0. $I_{cum}(t_{obs})$ can be used to estimate the *average ionic current*, $\bar{I}([t_0, t_{obs}])$, within the observed time period; by simple linear fitting, provided the ion current is stationary. However, stability of this average becomes unclear if slow trends develop in $I_{cum}(t_{obs})$ data as t_{obs} increases. So, $I_{lim}(t_{obs})$ becomes useful, since this represents the temporal evolution of the arithmetic average of instant charge, $Q(t')$, up to t_{obs}. Consequently, if the current is stationary, we would expect $I_{lim}(t_{obs})$ to converge towards some expected mean value (limiting current), that should also match $\bar{I}([t_0, t_{obs}])$, as $t_{obs} \to +\infty$ in the steady-state regime. Figs., 4.2 illustrate how $I_{lim}(t_{obs})$ looks like when computing ion current data for the protein pore cWzaY373C at various applied voltages.

4.2.3 Analysis of protein conformational stability

As mentioned in Sec., **??**, monitoring stability of the peptide pore conformations is crucial to know to what extent and how fast pores expand before reaching a stable conformation, which affects the rate of ion transport observed, due to the influence of an applied constant electric field. Here, we summarise some of the initial tests carried out to characterise protein conformational changes.

Root-Mean-Square Displacement (RMSD) One of the most recurrent tests to check for *overall* protein conformational stability is RMSD. This quantifies to what extent the overall *geometrical* structure at time t, deviates from the initial one at some time t_0(= 0, usually), by looking at the *standard deviation* of the displacement between pairwise relative positions of all atoms in the protein structure. Mathematically, this is expressed as

$$RMSD(t,t_0) = \sqrt{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}|\mathbf{r}_{ij}(t)-\mathbf{r}_{ij}(t_0)|^2} \tag{4.8}$$

where, $\mathbf{r}_{ij}(t) = \mathbf{r}_i(t) - \mathbf{r}_j(t)$; is the relative position of the *i-th* atom with respect to the*j-th* atom, at positions $\mathbf{r}_i$ and $\mathbf{r}_j$, respectively, at time t. Thus, conformational stability is usually regarded as being characterised by an increasing RMSD profile that eventually reaches a plateau for a stable protein.

Radius of gyration and moments of inertia RMSD provides some information about how stable, the protein conformation is in time as a whole. However, this by definition hinders anisotropic structural deformation if existent, moreover, RMSD neglects mass distribution. As an initial attempt to reveal any anisotropy in the mass distribution of the protein, we compute the corresponding *moments of inertia* from the distribution of C_α-atoms of the protein, about every coordinate axis, $X = x, y, z$, given by

$$I_X = \sum_{i=1}^{N} m_i \left(\mathbf{r}_i^2 - [\mathbf{r}_i \cdot \hat{\mathbf{e}}_X]^2\right) \tag{4.9}$$

assuming all atomic positions $\mathbf{r}_i(t)$, measured from the *centre of mass* of the protein at some time $t > 0$; where $\hat{\mathbf{e}}_X$ are the unit vectors of the respective cartesian axes. However, to give a more intuitive idea as to how mass concentrates around one particular axis, we look at the *radius of gyration*, $R_{g,X}$, about some X-axis, defined by the relation

$$R_{g,X} = \sqrt{I_X/M} \tag{4.10}$$

where, M is the total mass of the set of atoms, i.e., $M = \sum_{i=1}^{N} m_i$. Since $I_X = MR_{g,X}^2$ corresponds to the moment of inertia of a ring of mass M and radius $R_{g,X}$, the latter gives us an estimate of how far the mass is concentrated relative to the axis X, or the *compactness* of the structure about some specific axis X. Thus, the total *radius of gyration*, R_g, defined by

$$R_g^2 = \frac{1}{2}(R_{g,x}^2 + R_{g,y}^2 + R_{g,z}^2) \tag{4.11}$$

is similarly regarded, as an overall estimate of the *compactness of folding* of a protein [**?**]. Hence, the temporal evolution of either $R_{g,X}(t)$ or $R_g(t)$, may provide us additional information about any anisotropic axis-specific or general change of protein mass distribution, respectively.

Minimum Radius and Length As a final test, we also looked at the temporal evolution of the *pore length*, $\ell_{pore}(t)$, and the *minimum radius*, $R_{min}(t)$, of the pore interior, to have a more intuitive (but perhaps less precise) view of how the pore structure elongates and contracts as the protein evolves in conformation while sitting in the membrane. Calculations of ℓ_{pore} and R_{min}, from crystal structures, have been found both to linearly correlate with experimentally determined values of conductance for various membrane protein channels. So, it is worth considering the evolution of these two observables, also as a proxy to how pore conductance can change in time.

On the one hand, ℓ_{pore} is simply estimated by

$$\ell_{pore} = max(\mathbf{z}^{C_\alpha}) - min(\mathbf{z}^{C_\alpha}) \tag{4.12}$$

where $\mathbf{z}^{C_\alpha} = \{z_1, z_2, \cdots, z_{N_{C_\alpha}}\}$, is the set of Z-coordinates of all C_α-atoms in the protein structure. On the other hand, R_{min} is calculated as the minimum radius of the *pore inner surface*; which is approximated as the outer surface of a series of equally spaced, overlapping spheres that fit within the pore, along a pathway that crosses the pore from the entrance to exit [98].

4.3 Artefacts in Simulation of Ion Current

4.3.1 Pressure Control and External Field Interaction

It has been mentioned previously that artificial heating or spurious dynamics is produced in membrane protein simulations whenever an external electric field coexists with pressure control. Known attempts of simulation of ion permeation are generally done under *constant volume* conditions (NVT-ensemble), although it is not clear what the effect of this artefact is on the observed dynamics [2, 100]. However, observational evidence from an independent run of MD simulation of α-hemolysin shows that while keeping pressure control at +1.2 and +0.6[V] , the simulation cell elongates along the direction of the applied electric field, leading to uncontrolled deformation of the central cell, after 20 ns. This was avoided by simply keeping the cell volume constant, as suggested early.

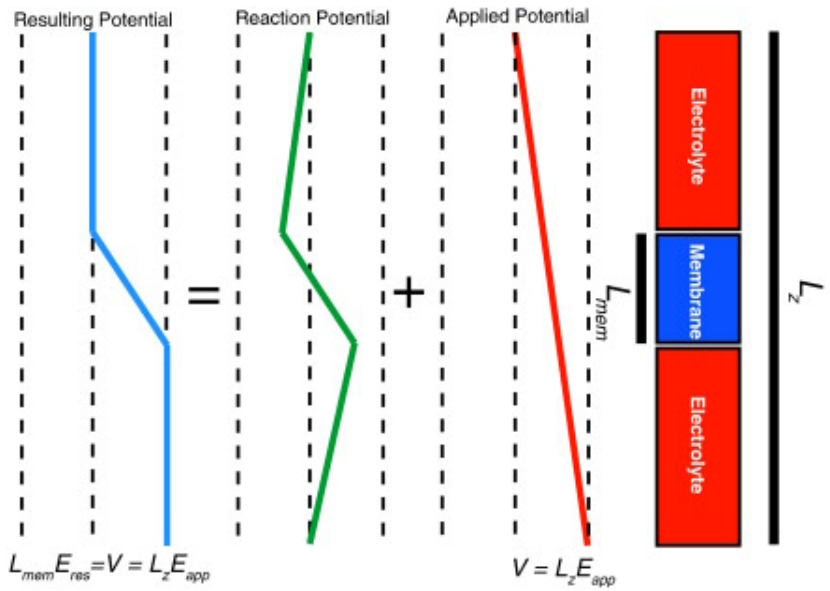


Figure 4.3: Picture taken from [36], shows a schematic representation of how the total electric potential across a membrane electrolyte system is made of the addition of the externally *applied potential*, generating a constant electric field E_{app}, plus a *reaction potential* that is the resultant potential due to all charged particles in the system. Note how when the electrostatics is correctly reproduced [36], away from the membrane the *resulting potential* remains constant. Note that the hight of the drop of the resulting potential corresponds to $\Delta V_0 = L_z E_{app}$.

4.3.2 Solvent Relaxation

Setting up initial conditions of the electrolyte medium before simulation and measurement of ion flux is a crucial but sometimes neglected step affecting the simulation of membrane pores [36, 87]. This stage is key to produce dynamics that can correctly reproduce *i)* electrostatics throughout the simulated system and *ii)* ion flux, both in a close to a steady or relaxed state. Evidence suggests that failure to achieve this *solvent relaxed state*, can lead to artificially large ion current across a simulated pore in comparison to experimental estimates. Thus, we discuss what is necessary to avoid some of these artefacts affecting solvent relaxation.

Electrostatics is vital in driving permeation of ions across membrane pores, then a misrepresentation of it in simulations is expected to lead to unphysical dynamics. In particular, when introducing an external field in simulations, finite-size effects can be easily overlooked. Previous MD assays of solid carbon pores (20Å diameter and thickness) with varying cell size ($Lz = 8 and 16$ nm) in the direction of the electric field ($E_z = -\Delta V_0 / L_z$), provide evidence that in small cells, larger ionic current can be measured in comparison to large ones; as large as *1.3 times* at $\Delta V_0 = +1V$, despite the current being steady. Further analysis showed that a reduction in the effective friction upon permeant ion is observed for the small cell system, then enhancing ion current [36]. Theoretical studies argue that since ions and water are mobile, these experience a rearrangement in response to the external field to avoid charge imbalance in the bulk region, that

might lead to prohibitively large energies. This rearrangement generates a *reaction electric field* that cancels out the external one, hence leading to a drop to zero of the *total electric field* away from the membrane, characterised by a flat total average electrostatic potential near the cell ends, with a voltage difference approximating the target value ΔV_0 (See Fig., 4). Thus, the main drop in the total electric potential takes places in the vicinity of the membrane, as a consequence of a slight charge imbalance away from the bulk [36, 87]. If a small simulation cell is used, these crucial features of the electrostatics will be not reproduced. Evidence from cWzaY373C, clearly shows that for a cell of length $L_z \approx 7nm$, the total electric field at the cell ends, does not drop to zero, moreover large fluctuations of the total electric potential are observed too. Fortunately, this issue is easily solved by setting up a cell with larger L_z, but at the expense of increasing water and ions on both sides of the membrane 4.3, shows a comparison of how the total average electrostatics look like for different values of L_z, after 5 ns of simulation, to allow the electrostatics to reach a relaxed state. So, electrostatics require the potential to relax and being correctly reproduced away from the membrane, before considering any ion flux measurement to determine reliable pore conductance (Section 4.3).

Yet, ion flux generally reaches stationary state after several nanoseconds of simulation, hence requiring large computational time. Tests show that if simulating ion permeation across the cWza-Y373C channel, longer than 150 ns of dynamics are needed to get steady ion current at low voltages within the physiological range, i.e., $|V_0| < +0.2V$, despite electrostatics being correctly reproduced with the right cell size; see Figure 4.6. Previously, use of large target voltages ($> +1.2$ V) has been proven to accelerate ion flux relaxation in membrane pore simulation, so that once steady ionic flux is reached in a few of nanoseconds ($\lessapprox 5$ ns, i.e., 30 times shorter), the applied voltage can be suddenly lowered to a desired target value, $|V_0| < +1.2$ V, correspondingly producing a lower ionic flux, but that similarly reaches steadiness in a few nanoseconds of dynamics. Pioneering MD studies of natural protein pores, *α-hemolysin* and *mechano-sensitive channel* of small conductance (MscS), show that this practical mean to accelerate solvent relaxation, can lead to conductance results of the order of experimental estimates, alongside qualitative observations of rectified and selective behaviours [2, 100].

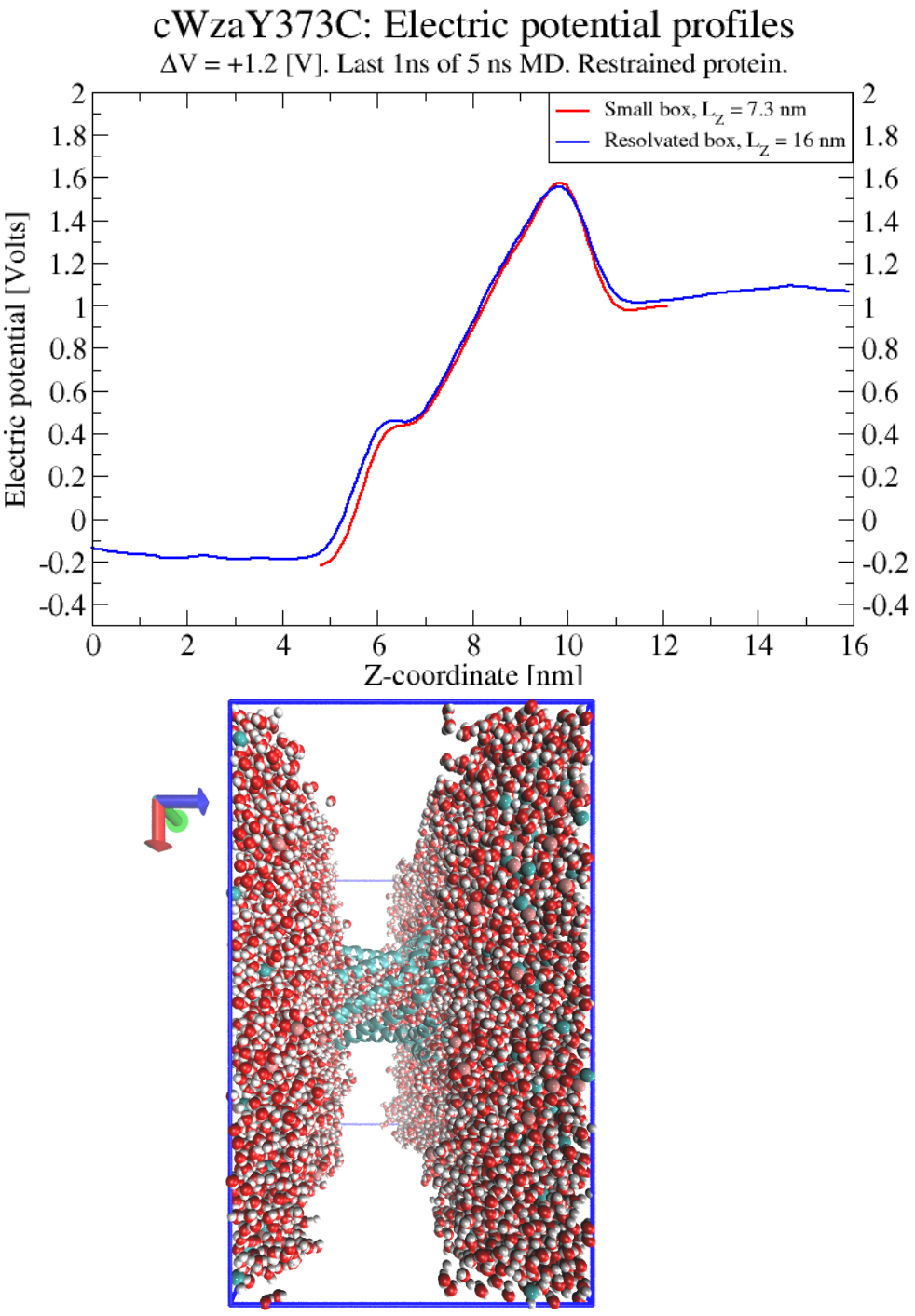


Figure 4.4: *Top*: Total electrostatic potential curves as a function of z-coordinate averaged over MD trajectories of two simulations of cWzaY373C in membrane with different cell lengths, $L_z \approx 7$ (red) and 16 nm (blue). *Bottom*: Plane XZ of small simulation cell; membrane was excluded to allow visibility of the protein. For both systems, motion of cWzaY373C atoms was limited by restraints, hence pore dimensions remained practically constant. Also, a hyper-voltage voltage has been applied, $+1.2V$, to ensure ion current reaches the steady-state quickly.

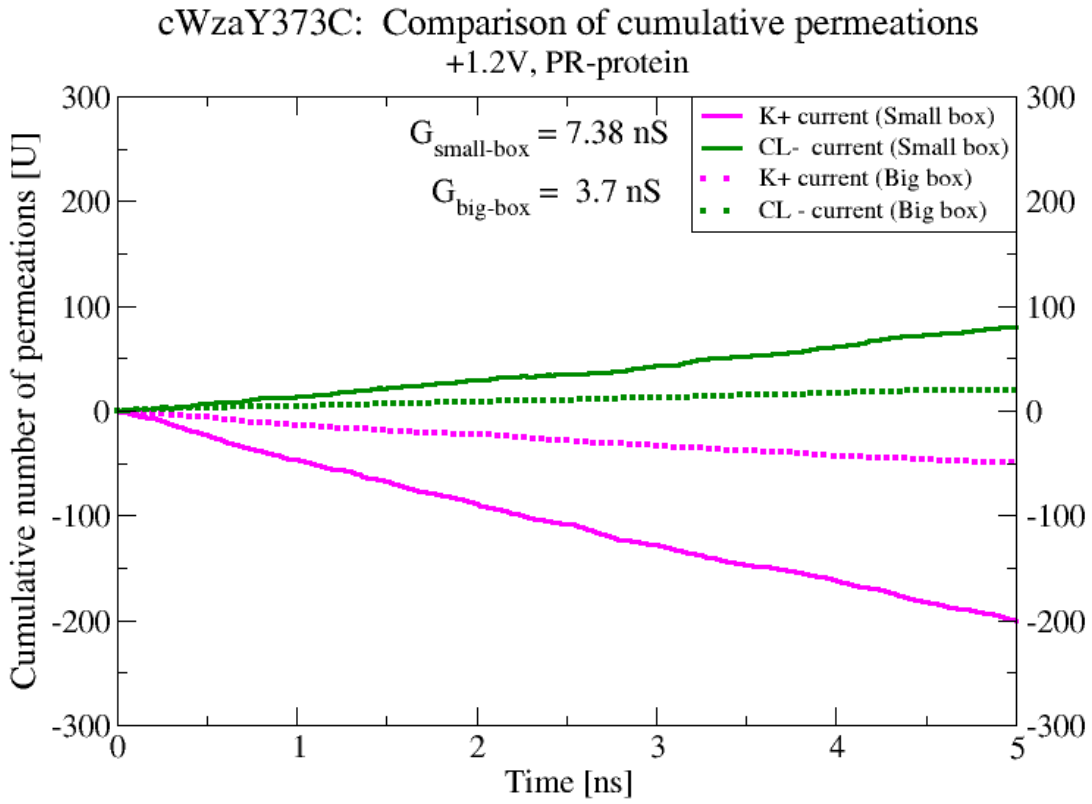


Figure 4.5: Test results showing the difference in the amount of permeation events for a small and a large cell system. Other sources contributing to ion current variability, were suppressed.

4.3.2.1 Finite-size effects on ion flux: Small box vs. Large box

To test to what extent finite-size effects can solely affect ion current calculations for cell lengths of $L_z = 80$ and 160Å, we carried out 5 ns of MD, while position restraining atomic coordinates of cWzaY373C; to avoid variability of the measured current due to conformational instability, and at an applied hyper-voltage of$+1.2V$; to accelerate ion current stationarity which might also introduce variability in ion current calculations. Potential profiles for both cells (Fig., 4.4, *Top*) averaged over the last 1ns of dynamics seem stationary, with the profile of the small system satisfying potential relaxation conditions (Eqs., 4.3 and 4.4). Nevertheless, cumulative ion currents show significant differences, as shown in Figure 4.5

Note that the number of permeation events per ion species for the small cell system increase between 3 and 4-fold in comparison to the large cell system. Implying approximate conductances of $7.38nS$ and $3.7nS$, respectively. Although both, deviate respectively 7 and 5-fold from the experimental estimate of cWzaY373C ($0.75 \pm 0.04nS$), it is clear that a small cell size leads to larger artificial current.

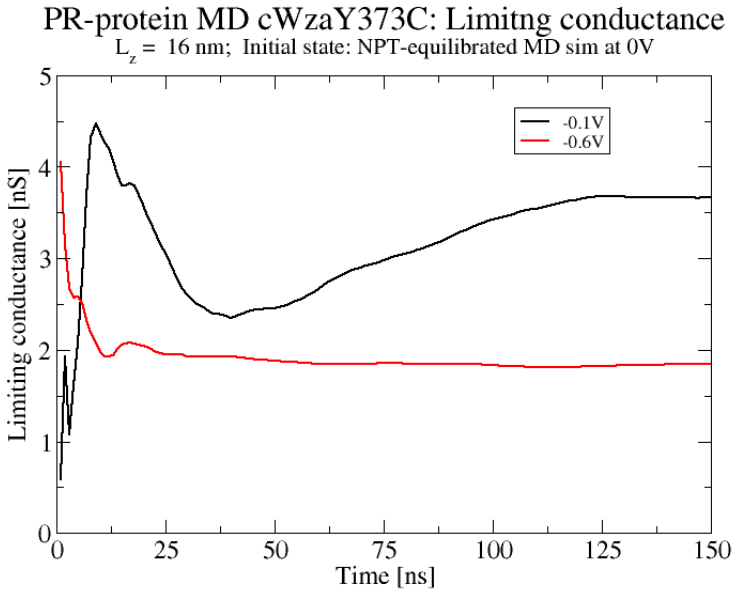


Figure 4.6: Limiting behaviour of the conductance as the ion current attempts to reach steadiness at initially applied low and medium-large voltages.

4.3.2.2 Slow relaxation of ion flux

To find out how slowly relaxation of the solvent was reached if starting dynamics at low applied voltages, $\Delta V_0 = -0.1$ and $-0.6V$, we carried out 150 ns of MD while position restraining atomic coordinates of cWzaY373C; to avoid variability of the measured current due to conformational instability. To show more clearly how conductance estimation is affected by slow relaxation, we computed a limiting conductance, G_{lim}, defined by simply dividing the limiting current (Eq., 4.7) by the corresponding applied voltage, i.e., $G_{lim} = |I_{lim}|/|\Delta V_0|$. Results are shown in Figure 4.6.

Note how large fluctuations in the conductance are observed at low voltage, $-0.1V$, with the G_{lim} clearly not steady, reaching a final value of $\approx 3.7\,nS$, which deviates nearly 2.5-fold from the value found using accelerated solvent relaxation by hyper-voltages, i.e, $G_{PRMD} \approx 1.5\,nS$. Moreover, the final $G_{lim}(-0.1V)$ is even larger than at $G_{lim}(-0.6V) \approx 2.8\,nS$. Although $G_{lim}(-0.6V)$ still deviates 1.9-fold from G_{PRMD}, it is clear that large fluctuations are at least suppressed after nearly 25 ns of dynamics, showing a decreasing but steady decrease. Hence, it is clear how ion current relaxation can be a slow process if starting from low applied voltages, in comparison to initiation with hyper-voltages for which steadiness is reached in less than 10 ns. See Fig., 4.2 for comparison.

4.3.3 Protein assembly instability

In addition to solvent relaxation, a further instability affecting reliable conductance determination arises from the fact that when protein motion is unrestrained, its structure undergoes an expansion while sitting in the membrane. So, if no stationary flux is reached while the pore expands monotonically - with the pore size becoming eventually stable, once stationarity is reached, an ion current increment is observed. For instance, cWzaY373C simulation at hyper-voltage +1.2 V , results in unstable pore expansion throughout 25 ns of dynamics. Then, if taking the final state of this simulation as the initial state of another at lower voltage - as done to accelerate solvent relaxation (SEC), then conductance will increase because of the increased pore size. But, if dynamics was initiated at a lower voltage from the beginning, then avoiding unstable expansion at +1.2 V, the pore conductance will be lower as a consequence of a narrower pore structure (Appendix 5.2.3).

As mentioned before, protein conformation plays a key role in measurements of ion flux, since pore expansion may lead to larger conductance estimates. This is evident in Fig., 4.13 in the Results section of cWzaY373C, for which after removing restraints on protein atoms, I-V data reveal ≈ 1.5-fold increase in calculated ion current (for $|\Delta V_0| \leq 0.2$ V), in comparison to protein-restrained results. Thus, over a 25 ns period of unrestrained MD, assessment of the overall protein conformational stability was first performed by RMSD (Eq., 4.8) and then by analysis of the radius of gyration, $R_{g,X}$, relative to each cartesian coordinate axis (Eq., 4.10).

RMSD results in Fig., 4.3.3(*Top*) expose, on the one hand, that an overall expansion of the protein geometry takes places as a voltage is applied, with results for target voltages $|\Delta V_0| < 1.2$ V fluctuating the least. At hyper-voltages a notorious overall expansion is observed, but at $+1.2V$ the largest. RMSD trends of the cWzaY373C-restrained, at some applied voltage, and unrestrained MD without applied voltage were also computed as comparison controls. These results show that at low voltages overall expansion does not greatly differ from *intrinsic pore expansion* (at 0V), but at hyper-voltages, the structure suddenly responds to the applied field.

A more refined analysis by $R_{g,X}$, revealed the slow expansion of the cWzaY373C structure around the Z-axis, with the most pronounced at hyper-voltage -1.2 V than any other voltage. Results for the remaining axes, characterised by $R^2_{g,y} - R^2_{g,x}$, showed fluctuating trends with no clear increase. See Fig., 4.3.3 (*Bottom*).

Finally, given the notorious difference in results at hyper-voltages, we examined MD trajectories of protein atoms visually to observe the extent of the pore expansion. In Fig., 4.7, we show the final snapshots (at 25 ns) of unrestrained MD at +1.2 V (*Top*) and -1.2 V (*Bottom*). These make evident that at $+1.2V$ the structure of cWzaY373C expands as peptide atoms are pulled away from each other with a degraded helicity, by contrast to the starting structure (at 0 ns); in grey in the background. By contrast, at -1.2 V, expansion seems marginal. Final minimum diameters increased from ≈ 12Å (0 ns) to ≈ 20 and 14Å (at 25ns), respectively. Lengths monitored in time also showed slowly increasing trends, but highly fluctuating between 48 and 58 Å, even

for low voltages and in the absence of it, except when position-restrained atomic coordinates.

This issue nonetheless is easily prevented just by restraining the motion of heavy atoms of the protein. Later on, we will introduce specific indications as to how to implement restraints with care at different simulation stages. Also, although we cannot guarantee whether a significant number of ion permeation events can be observed through the narrowest pores - CCTM and aHL-CC. We will follow this approach provisionally to test this initial idea suggested from cWzaY373C simulations.

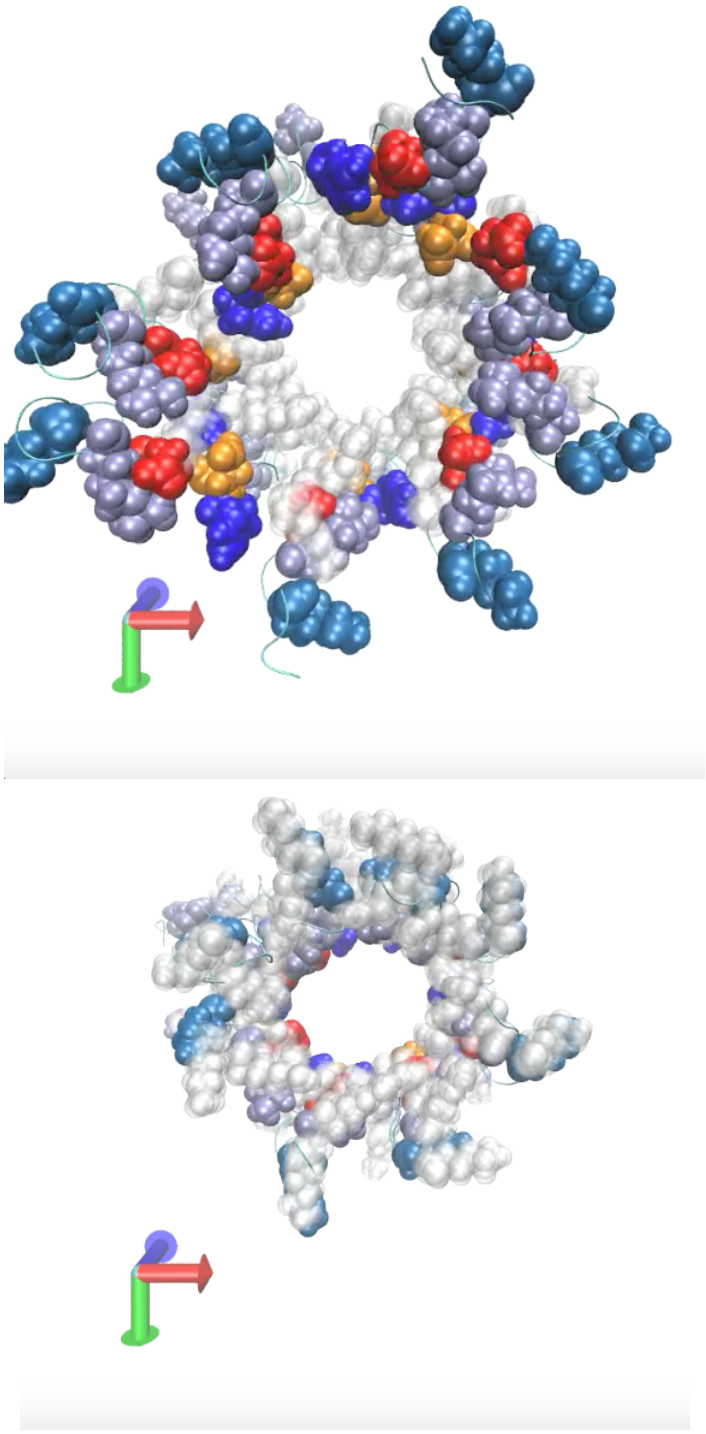


Figure 4.7: Bottom View of the end structures (t = 25 ns) of (Un-Restrained) MD simulated cWza-Y373C channel under applied voltage conditions at +1.2 V (*Top*) and −1.2 V (*Bottom*), highlighting with positively (Blue tones) and negatively (Red and Orange) charged residues as VdW spheres. The atoms of the starting structure (t = 0 ns) are shown as white spheres overlapping the coloured ones of the end structure for comparison; structures were backbone aligned using VMD.

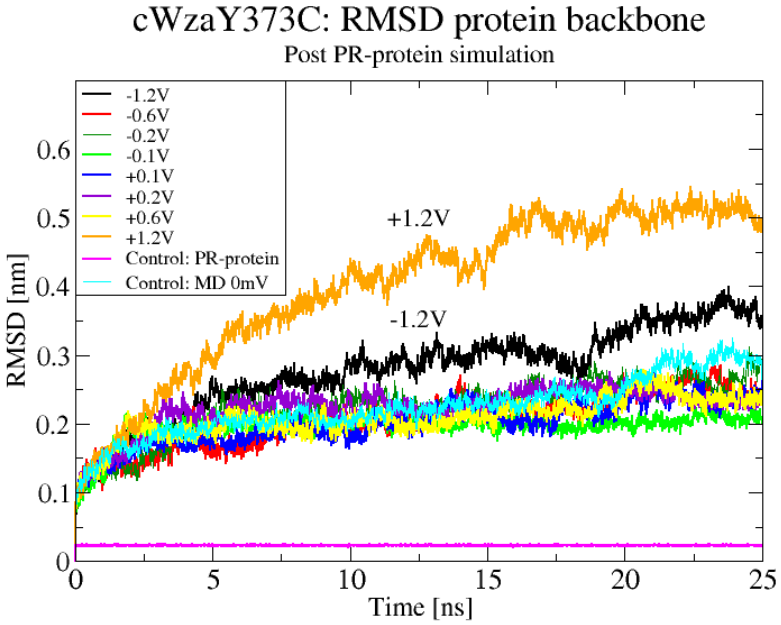


Figure 4.8: Temporal evolution of the Backbone-RMSD of the cWza-Y373C channel from all 25-ns MD simulations with applied target voltages; after flux relaxation (PRMD phase). Data from the PRMD phase (25-ns long) with +0.2 V and from a simulation without applied voltage (0 V) are featured for comparison. Note that the largest fluctuations in RMSD data happen at *hyper-voltage* values (±1.2 V); from which structural instability was suspected. See Figure 4.7. All Backbone-RMSD values at time $t > 0$ were calculated with respect to the starting structure ($t = 0$ ns).

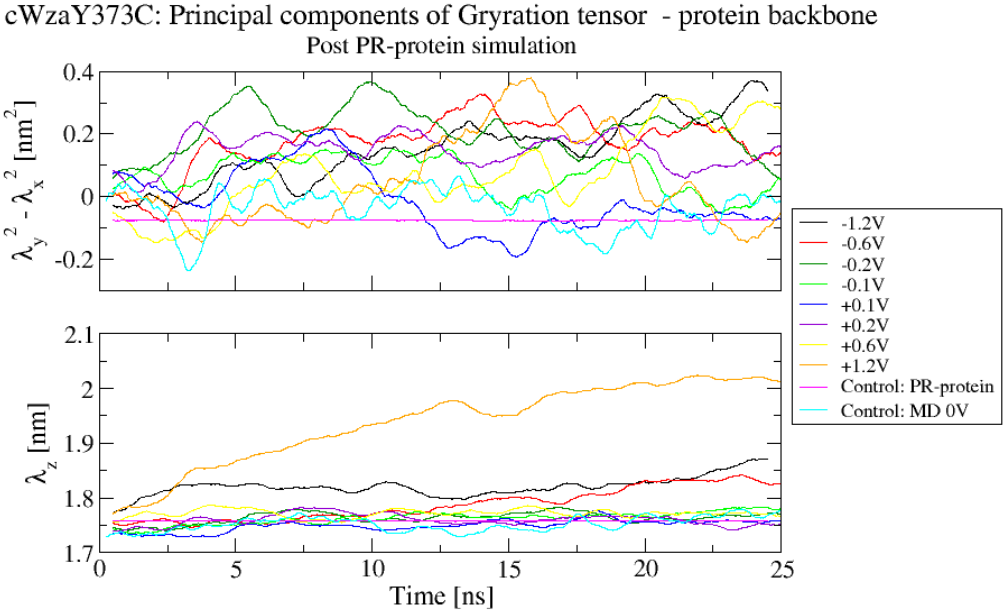


Figure 4.9: Temporal evolution of the *acylindricity* ($\lambda_y^2 - \lambda_x^2$) - which becomes zero for cylindrically symmetric distribution of particles; and the Backbone Radius-of-gyration Z-component (λ_z); where $\lambda_{x(y)}$ is the X(Y)-component. Note that for ±1.2 and −0.6 V, λ_z indicates that particles increasingly spread over the XY plane by the end of every simulation.

4.4 Benchmarking

To ensure the reliability of all the above-described ideas and methods, we benchmarked our MD simulation protocol and data analysis code, against systems for which experimental and MD simulation data are well known. Benchmark systems in our work included: *1)* Bulk KCl electrolyte and *2)* the α-hemolysin protein pore. In summary, we tested implementation of specific methodologies such as *i)* ion current calculation and *ii)* electric potential maps and profile (Subsection 4.2.1), in addition to assumptions such as *iii)* coexistence of pressure control and external fields, *iv)* accelerated solvent relaxation through high voltages (Section 4.3), as well as *v)* the influence of the water and force field choices in our simulations. Correction and calibration of our simulation procedure were carried out to some extent through this useful exercise. Testing of our MD simulation procedure is key to ensure the reliability of its results. Thus, we validated our current protocol against two well-known systems in simulations: KCl electrolyte system and the α-hemolysin protein pore [2]. Details of our simulations are found below, whereas test with these are justified and their results summarised in following sections.

KCl electrolyte Using GROMACS, we first set up a cubic box of pre-equilibrated water (SPC, TIP3P, TIP4P, and TIP5P models) of equal length $L = 74$Å, with a predefined force field (CHARMM27 and OPLS-AA). Next, replaced water with K and Cl ions at neutral 1M concentration, then energy minimised (2000 steps) and equilibrated the mixture at 310 K (Nosé-Hoover) and 1 bar (Parrinello-Rahman) for 5 ns. After this, we started production of dynamics for 30 ns, restarting velocities, at different applied voltages $\Delta V_0 (= -E_0 \times L > 0)$ along the Z-axis, restarting velocities, with the final simulation state of equilibration being the initial stage of production.

α-hemolysin We first downloaded GROMACS MD data for an equilibrated α-hemolysin protein already embedded in DPPC bilayer, from the Memprot Database [101]. Given the pre-existent conditions of the system, we had to remove and resolvate the system, with 1M KCl electrolyte, using SPC water model. Plus, topologies were modified to enforce OPLS-AA and Berger force fields. α-hemolysing and lipid coordinates were not modified to preserve protein embedding. Next, the newly setup system was energy minimised (40,000 steps) and equilibrated at 310 K(Nosé-Hoover) and 1 bar (Parrinello-Rahman) for 5 ns. After this, we started MD production without applied voltage at fixed volume (NVT-ensemble) for 5 ns, restarting velocities, to verify that the protein remained perfectly embedded. Finally, production was continued at applied voltages ΔV_0= +1.2 and $0.6V$, with the field perpendicular to the bilayer plane, for 20 ns, but with the final state at $+1.2V$ being the initial one at $+0.6V$.

Correctness of ion current calculations

To test reliability of our Python (Version 2.7.10) coded implementation of *instant charge* calculation -apart from code inspection, $Q(t+\Delta t) = \frac{\Delta t}{d} \sum_{channel} q_i \dot{z}_i(t+\Delta t)$ (Eq., 4.5), and related

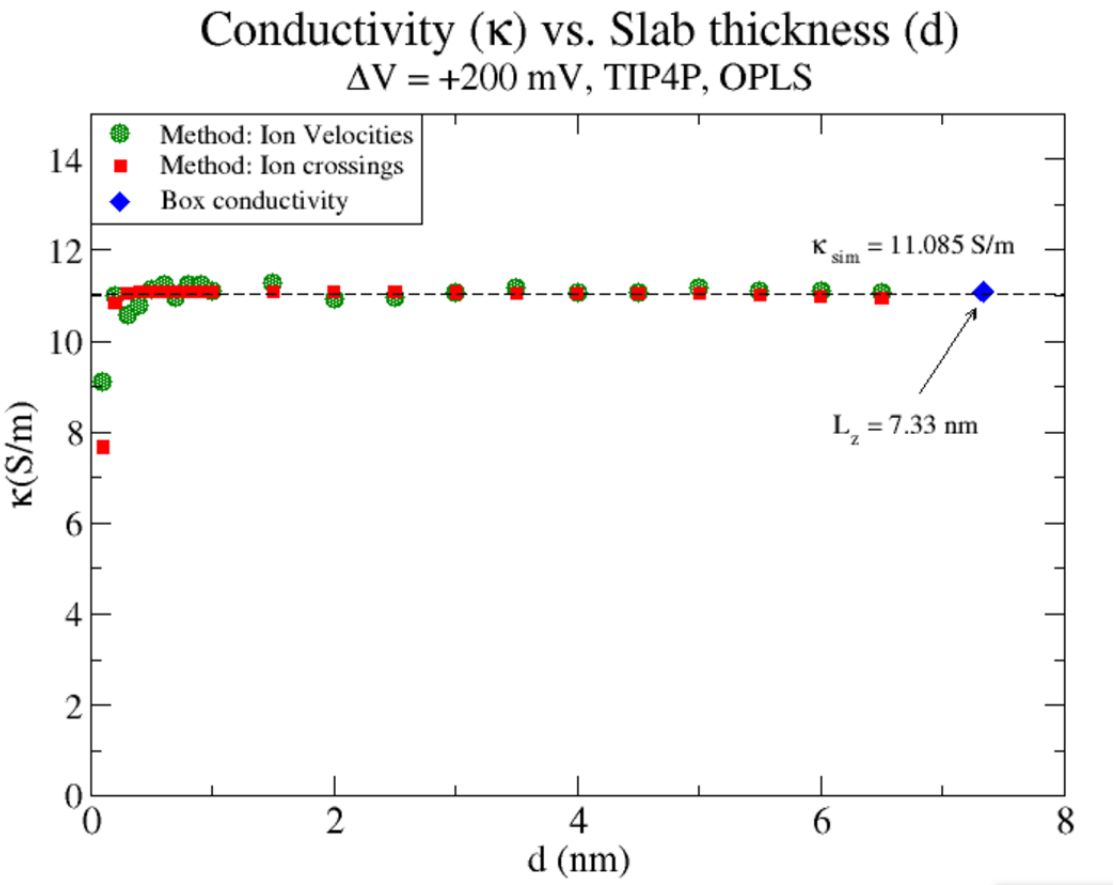


Figure 4.10: Conductivity values obtained from ion-currents across slabs of variable thickness (d) along the Z-axis and centred in the middle of the box of simulated 1 M K-Cl electrolyte. For comparison, ion currents were estimated via two different methods: 1) based on the average Z-velocities of the ions within a slab (●), or 2) based on the rate of discrete ion crossings through the slab's faces in Z (■), over the course of the simulation. The direction of the applied voltage difference (ΔV = +200 mV) was parallel to the Z-axis, and the Z-axis length of the simulation box (L_z = 7.33 nm), as well as the water-model/force-field choice - TIP4P/OPLS-AA, were fixed. Overall, no difference was observed between estimated values form both methods, and conductivity estimates were always very close to the one of the whole box (♦), even for slabs with < 1 nm of thickness.

equations to estimate conductance, we first used our electrolyte MD trajectories, introduced above, to compute: *i) bulk conductivity* κ_{bulk} of a sufficiently large fixed volume, and ii) *bulk conductivity* κ_{bulk} of a slab of variable thickness.

Recall that the *bulk conductance*, G_{bulk}, of a rectangular volume of electrolyte of length, ℓ, and cross-sectional area, a, can be calculated as $G_{bulk} = \kappa_{bulk} \times a/\ell$. Then, from Ohm's law, $I = G_{bulk}\Delta V_0$, for our MD simulations κ_{bulk} can be simply determined by

$$\kappa_{bulk} = \frac{I}{\Delta V_0} \times \frac{a}{\ell} \tag{4.13}$$

We first tested whether our code can produce a κ_{bulk} value close to the one observed previously in simulations out of current-voltage data. Results produced a value of κ_{bulk} = 11.085 [nS/nm], from our electrolyte simulations (OPLS, TIP4P water model) at voltages $0 < \Delta V_0 < 0.6V$. This is value compares to κ_{MD} = 12.3 [nS/nm], found in previous MD simulations, both being close to the experimental κ_{exptl} = 11.0 [nS/nm] [2]).

The second tests how our code performs against another one; implementing ion crossing counting to work out ionic current, given an identical task of computing κ_{bulk} at fixed voltage ($\Delta V_0 = 0.2V$), for a slab section of variable thickness, d - in the Eq., for $Q(t+\Delta t)$, of the electrolyte simulation cell, i.e., $0 < d \leq L = 74$Å. Thus, assuming that both codes and methods are correct, results for κ_{bulk} are expected to coincide up to some extremely small thickness d. Plus, κ_{bulk} is expected to be nearly constant, for a homogenous and isotropic electrolyte. Corrections to our code where needed, to find matching results as shown in Fig., 4.10.

Effect of force field and water model choice on conductance

Ion-solvent and ion-pore interactions play a key role in transport. Thus, since these interactions are generally parametrised differently given different water models and force field, as a first test to assess the effect of our initial force field, OPLS-AA, and water model, SPC, choices on observed ionic currents for cWzaY373C and other channels, we compared current-voltage data for different combinations of force fields and water models, as mentioned earlier.

Results summarised in Fig., 4.11 show that average estimates of κ_{sim} (Eq., 4.13) roughly range between 10 to16nS/nm, per force-field-water-model combination, leading to an average percentage error of 17%, relative to κ_{exptl}. So, force field and water model choices do not account for any increase in observed MD ion current greater than 17%, in comparison to experiments.

Although the above results suggest that ionic current calculations, and hence conductance, are fairly sensitive to force field and water model choices, this suggestion might not be necessarily extended to simulations with protein pores, where further ion-protein interaction takes place. Next, we analysed ion flux data from α-hemolysin MD simulation, described early, which already enforces our initial choice of force-field-water-model combination. Thus, after using our previously benchmarked ion-current code, we found an average pore conductance of $G_{\alpha-hemolysin} \approx 1.4nS$, at ΔV_0= +1.2 and $0.6V$. This result is of the same order as previous estimates by other MD studies for KCl electrolyte at 1M concentration, $G_{MD} \sim 1nS$ [2, 17], as well as in comparison to experimental estimates, $G_{exptl} \sim 1nS$ [66], under similar electrolyte conditions.

4.5 Simulation protocol

Next, we introduce a general simulation protocol, that puts into practice methodologies and observations discussed in all previous sections. We formulate this in a general way at this stage of

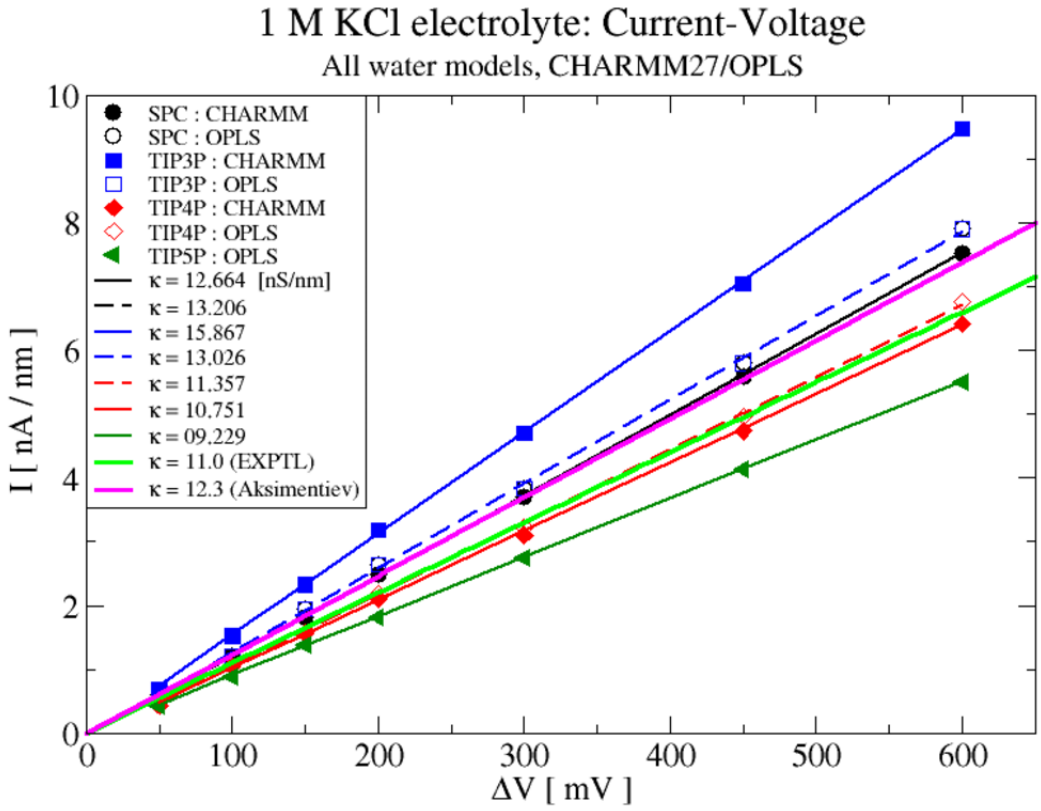


Figure 4.11: Current-Voltage curves from 1M K-Cl electrolyte simulations with different force-fields and water-models combinations. All straight lines were least-square fitted to the simulation data-points, with slopes corresponding to estimated *conductivity* values, κ. The experimentally determined value of conductivity, κ_{exptl} = 11 nS/nm, is featured for comparison, alongside a previously reported value from similar simulations by Aksimentiev and Schulten [2]. The SPC/OPLS-AA water-model/force-field choice leads to a 20% error relative to κ_{exptl}.

our research, since we are interested in testing whether or not, we can apply a similar simulation procedure to any synthetic protein pores of the types described in 4.6.1.

Setup Initial atomic coordinates in PDB format of each pore construct (Subsec., 4.6.1) were provided by the Woolfson group. Every pore was embedded in a patch of POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) bilayer of 128 lipids per leaflet; using the compression-insertion-expansion protocol of *g_membed* of GROMACS [122]. The *Hippo* software-package was used to set initial positioning and orientation of the protein in the bilayer via Monte Carlo movements to avoid unfavourable energetics for the embedding process. Solvation of the protein-lipid complex was done at 1M KCl electrolyte concentration with SPC water model, within periodic ($L_{x,} \times L_y \times L_z$) simulation cells with $12 \times 12 \times 7$nm for cWzaY373C, and $12 \times 12 \times 11$nm for all CCTM and aHL-hybrid peptide pores. Peptides were parametrised with the *OPLS-AA*

force field, while *Berger parameters* were employed for lipid-protein interaction [15]. To finalise embedding, we equilibrated lipid-protein complexes for 5 ns, at 310 K (Nosé-Hoover Chain) and 1 atm (Parrinello-Rahman), applying harmonic-restraints on protein heavy-atom positions, i.e., non-hydrogens, and constraining bond-vibrations (P-LINCS).

Note that for this and any other subsequent stages of equilibration and simulation of dynamics, integration was done with *leap-frog* algorithm, over a 2 fs timestep, while saving coordinates every 10 ps.

Resolvation of simulation box To avoid finite-size effects on the total electric potential (Section 4.3), by removing all previous and adding more water and ions at 1M KCl, we resized every periodic simulation cell to new dimensions of $12 \times 12 \times 16$nm for cWzaY373C and $12 \times 12 \times 20$nm for all CCTM and aHL-hybrid pores, while keeping the centre of mass of each bilayer centred within the box. Newly solvated systems were energy minimised (40,000 steps), and equilibrated again at 310 K and 1 atm for 5 ns, likewise keeping restraints and constraints on protein and bonds.

Solvent relaxation First, to ensure equilibrated electric potential and steady ion flux conditions (Section 4.3), we simulate dynamics of the system for 30 ns, at large applied voltages of $\pm 1.2V$ initially, and fixed volume (NVT-ensemble), while keeping the system at 310 K and the protein restrained. Both, electric potential and ion flux should be monitored throughout and in particular by the end of the 30 ns period, to verify relaxation according to procedures in Subsection 4.2.1.

Once achieved solvent relaxation at $\pm 1.2V$, the last simulation state (coordinates and velocities) is taken as the common initial state of the subsequent simulation at low target voltages (ΔV_0), i.e., take final state at $+1.2V$, as initial state of simulation at $+\Delta V_0 < +1.2V$; similarly at $-\Delta V_0 > -1.2V$. Thus, we produce another 30 ns of dynamics per target value ΔV_0, without changing temperature, fixed volume, and protein restraining conditions. Again, electric potential and ion flux relaxation should be monitored throughout and by the end of the simulation, hence, when relaxed at the new target voltage ΔV_0, the system should be ready for production of dynamics.

Production of dynamics So far, protein coordinates have been position restrained from the start, keeping the narrowest possible protein conformation. Recall that early evidence suggested that during solvent relaxation, protein peptides move apart in response to initially applied large voltages ($\pm 1.2V$), and if not restrained the pore conformation is ultimately expanded, leading to severely biased estimates of conductance. However, having followed the above-described stages, now as both relaxed electric potential and steady ion flux behaviour are obtained, we can release protein coordinates.

As the final stage of our simulation protocol, we remove protein position restraints, initiating dynamics by taking the final state of the solvent-relaxed runs corresponding to each target

voltage ΔV_0 , with each new run being 25-30-ns long again; keeping the temperature ($310K$) and fixed volume conditions. Unlike previous runs, were electric potential and ion flux were monitored to ensure stationarity, we now also monitor protein conformational stability; using RMSD (Equation 4.8) as a first test. Thus, after both solvent and protein conformation are stationary, we then estimate the average ion current, $\bar{I}$, by fitting data from Equation 4.6; plotting this against the corresponding target voltage values ΔV_0, to finally compare and characterise simulated conductance.

4.6 Results and Discussion

Here we introduce the results from the application of our simulation protocol, described in Section 4.5. But first, we describe the protein structures that we simulated employing this.

4.6.1 Structures of Simulated Peptide-Assembled Ion Channels

cWzaY373C: An engineered octameric α-helical barrel

cWzaY373C is made of eight α-helical peptides of 35 amino acid residues in length each (Figure 4.12, *Top left*), and a net charged of $+2e$ per peptide. The pore shows an hourglass-like shape with its narrowest diameter $\phi \approx 12$Å, and a total length ≈ 45Å, with the most inner lining of amino acids made of hydrophilic residues, with positively charged ones (2×*Arg, His*, and *Lys*) distributed almost evenly at both ends of each peptide, unlike negatively charged residues (*Glu* and *Asp*) which are mostly placed near the lower end (Figure 4.12, *Middle left*). By contrast, the outer surface is lined with hydrophilic residues that allow the pore to sit in the membrane environment (Figure 4.12, *Bottom left*). Further details about the computational construction of the atomic coordinates of cWzaY373C can be found in Reference [64].

Coiled-coil Trans-Membrane (CCTM) Channels

Electrophysiology assays with synthesised *de novo* α-helical peptides by Woolfson group suggest that -similarly to the cWza pore mutants- these new peptides can also insert themselves and assemble to form open pores in the membrane and able to conduct ions. These synthetic pores are thought to assemble in groups of 5 to 7 α-helical peptides with a net charge of $+4e$ per peptide, which are joint together forming a *coiled-coil transmembrane* (CCTM) structural pattern (Figure 4.12, *Top central*). In the absence of X-ray crystallography data, two computational constructs -based on ongoing research by the Woolfson group- are tested first: CCTM-Pent-KLLW and CCTM-Hept. The structures are respectively made of 5 and 7 peptides, featuring a common assembly pattern of a hydrophilic (polar) cylindrical core (Figure 4.12, *Middle central*), concentrical to an outer, cylindrical, hydrophobic shell, which at the upper end joins 4 consecutive positively-charged (Lys) Lysine amino-acid residues lining the cylinder-like edge (Figure 4.12, *Bottom central*). Inner diameters of CCTM-Pent-KLLW and CCTM-Hept are respectively, $\phi \approx 3$ and 6Å, and a total length of ≈ 70Å.

Despite having atomic coordinates for only these two constructs, these are useful to complement MD simulation assays for narrower pores than the cWzaY373C. In natural narrow protein channels such as *Gramicidin-A*, which has an internal diameter of $\phi_{internal} \sim 4-5$Å, conduction is not restricted by ion size like in wide pores, such as *porins* for which $\phi_{internal} \sim 10-20$Å, instead a more complex dynamics is observed as a consequence of a lower number ions and water molecules fitting in the pore interior [41].

α-Hemolysin Coiled-Coil (aHL-CC) Hybrid Channels

Based on similar experimental efforts for cWza mutants and CCTM pores, the Woolfson group has also worked on the construction of *hybrid pores*. Hybrid peptides are made out of individual β-strands which are linked to *de novo* water-soluble α-helices through engineered amino acid *linkers*. The association of multiple peptides are thought to lead to an assembly pattern like in Figure 4.12(*Top right*), where a β and a α-helical barrel are shown in yellow and purple, respectively, should correspond to the membrane and solvent-exposed parts of the designed hybrid pores.

Among all the possible experimental hybrid pores, we currently restrict our research to two particular designs: aHL-CCHex2 and aHL-CCHept, thought to be made of 6 and 7 peptides respectively, with a net charge of $+e$ per peptide. Present computational structures of these pores by the Woolfson group are based on artificial combinations of X-ray crystallography data from α-hemolysin β-strands (PDB id: 7AHL) with α-helices of CCHex2 (PDB id: 4PN9) and CCHept (PDB id: 4PNA)*de novo* proteins [105]. As shown in Figure 4.12 (*Middle and Bottom, Right*), the membrane part of both constructs is made of a hydrophilic core, patchily covered by hydrophobic residues in a cylindrical manner. The part exposed to the solvent has a hydrophobic core in contrast; with unevenly hydrophilic residues covering its outer surface. Note that the outer surface of the α -helical part is decorated by positive-negative charged pairs of *Glu-Lys* residues (Glutamate and Lysin), with its upper and lower ends decorated with annular arrangements of Lysin (*Lys*) and Arginine (*Arg*), positively-charged residues, respectively. Also, the β-part on its lower edge has a ring of positive-negative residue pairs of *Lys-Asp* (Lysine and Aspartate) (Figure 4.12), *Bottom right*). However, the α-helical component of aHL-CCHept is more densely covered by hydrophilic residues than aHL-CCHex2. Minimum inner diameters of aHL-CCHex2 and aHL-CCHept are respectively, $\phi \approx 3$ and 5Å, both with a membrane length $\approx$ 45Å, and a total length $\approx$ 90Å. Note that the membrane thickness in our simulations is roughly 40Å at 310K.

4.6.2 Simulation of ion permeation: cWzaY373C pore

At all stages of MD simulation cWzaY373C with applied voltages, a number of the order of 100 to 1000 ion permeation events were observed. However, as mentioned in Section 4.3, artefacts affecting reliable ion current calculations are removed in a sequential manner according to the *present version* of our simulation protocol (Subsection 4.5). We introduce and discuss the sequence of corrections to our average ion current values, $\bar{I}$ (Subsec., 4.2.2), at different target voltages, ΔV_0, and for different simulation stages, to illustrate how the *final conductance estimate*, G_{pred}, of cWzaY373C, compares for non-corrected stages and experimental observations; as condensed in Figure 4.13. A series of observation are made:

- *Initial set*: Small simulation cell, non-relaxed solvent, and unrestrained protein, at low voltage,

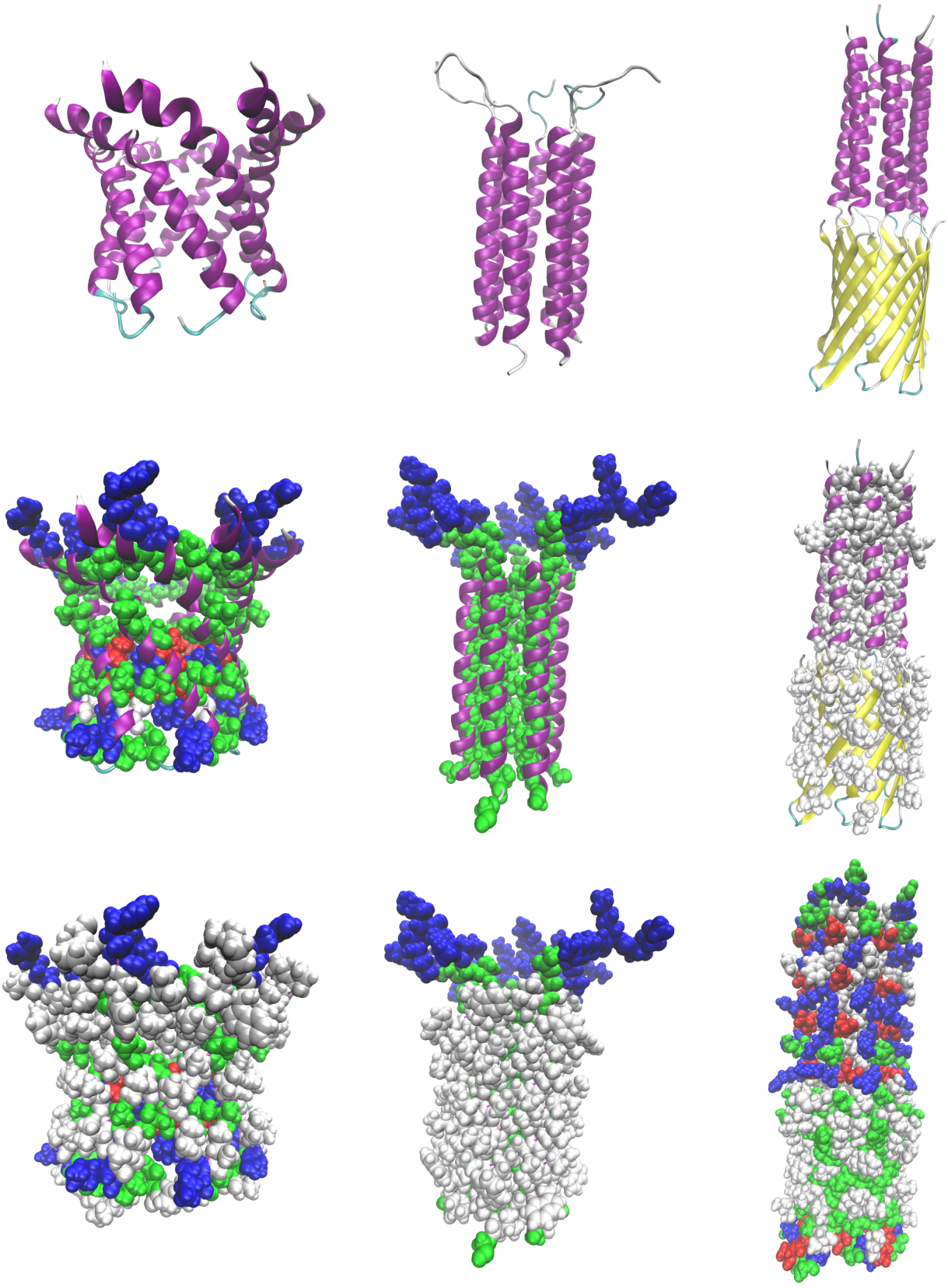


Figure 4.12: **Columns.** *Left*: cWzaY373C barrel; *Centre*: CCTM-Hept pore; *Right*: aHL-CCHept. **Rows.** *Top*: Secondary structure; *Middle*: Secondary structure, plus most internal lining of amino acids; *Bottom*: internal and outer lining superimposed (all amino acids). **Colours.** *Green*: Hydrophilic; *Gray*: Hydrophobic; *Red and Blue*: Negatively and Positively charged amino acid residues.

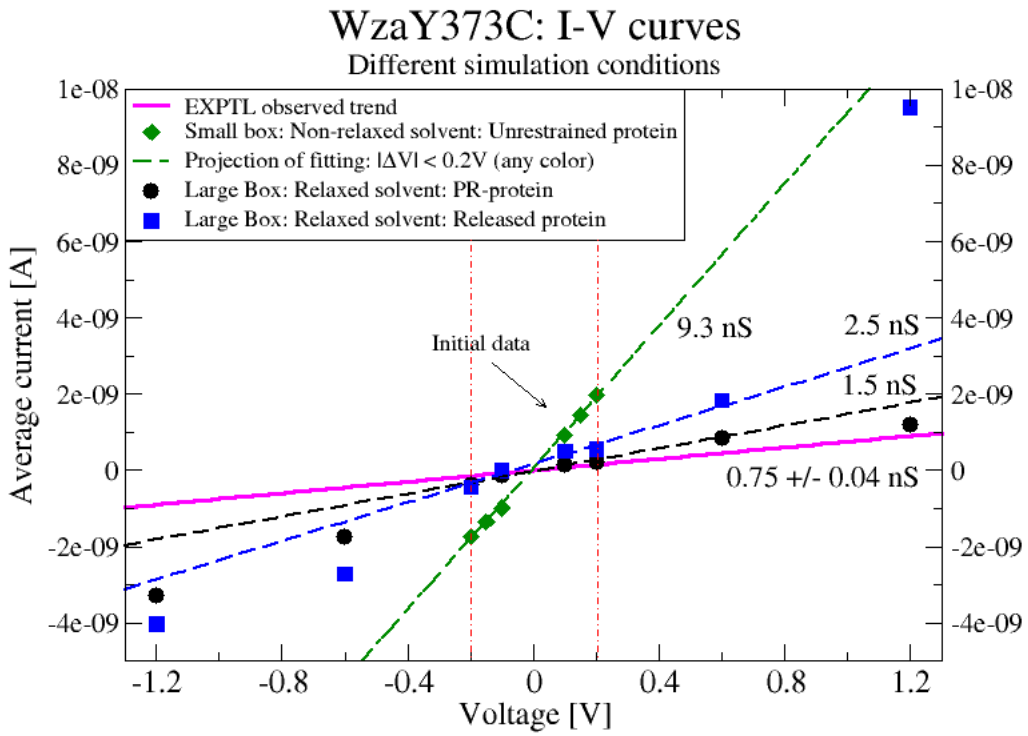


Figure 4.13: Summary of current-voltage (I-V) curves for various stages of the MD simulation protocol (Subsec., **??**) applied of cWzaY373C in POPC membrane. Experimental I-V trend is also included for comparison. Vertical dashed lines in red delimit the voltage interval, $-0.2 \leq \Delta V_0 \leq +0.2$, for which linear fitting was performed.

Since cWzaY373C is initially embedded in POPC, within a small simulation box. After thermal equilibration, production of unrestrained dynamics for 30 ns, at low applied voltages, ($|\Delta V_0| < 0.2[V]$), leads to average current calculations (*green diamond data points*, Figure 4.13) that imply a conductance estimate of $G_{setup} \approx 9.3\,nS$, which massively deviates nearly 12 times in comparison to the experimental estimate, $G_{exptl} = 0.75 \pm 0.04\,nS$, based on current-voltage trends (*magenta line*). Solvent was found unrelaxed, as expected for a small simulation box and at low voltages, in agreement with observation in Section 4.3.

- *Intermediate set*: Resolvated cell, position restrained (PR) protein, accelerated solvent relaxation by hyper-voltages ($\pm 1.2V$),

Once resolvation of the simulation cell was done, starting with the the narrowest conformation of cWzaY373C, at hyper-voltages and then lowering to target voltages ($|\Delta V_0| < 1.2V$), 30 ns of protein-restrained dynamics produced average ion current values (*black circle data points*), gathered after the solvent relaxed state was verified for each simulation after the first 10 ns (See Figs., 4.4 and 4.2). Note that, although ΔV_0 values extend beyond the experimental regime

$|\Delta V_0| < 0.2[V]$, we only linearly fit data points within this small interval (indicated by vertical red dashed lines), since larger values cannot be compared against experimental data. So, the estimated conductance for this stage is $G_{PRMD} \approx 1.5$ nS (*black dashed line*), that is 2-fold G_{exptl}. Despite the proximity of this result to experimental estimates, we remark that since protein coordinates are restrained throughout dynamics production, we need to quantify how pore conformational expansion -as discussed in Section 4.3- may change conductance results.

- *Final set*: Resolvated cell, relaxed solvent, protein released from restraints

This final set of results (*blue square data points*) shows that despite having a large cell that avoids finite-size effects and a relaxed solvent, after removing PR on cWzaY373C, pore expansion accounts for a 1.7-fold increase of the estimated conductance in comparison to G_{PRMD}, this is $G_{MD} \approx 2.5 nS$ (*blue dashed line*), which represents a 3.3-fold deviation with respect to G_{exptl}.

It is striking to notice that in comparison to PR conditions, pore expansion is significantly asymmetrical at large voltages ($\pm 1.2V$) after removing PR. Further inspection of MD trajectories reveals that at $+1.2V$, lipid heads interact with charged groups of the pore, in such a way that the asymmetric distribution of charges in the pore (See Figure 4.3.3), makes expansion more pronounced at one end of the pore, as a number of salt bridges are formed between atoms in the lipid head and charged groups in the protein. These results suggest that maybe protein charge-residue distribution may affect the stability of the structure in simulations, at least at large voltages, which might not be necessarily observed at low voltages. This possibility will be evaluated in the future.

Despite the deviation of our final conductance result, G_{MD}, for cWzaY373C, future reports will address whether it is possible to work out further corrections or effects not yet considered in our present protocol that may affect conductance predictions. Also, future MD assays with cWza mutants -for which also experimental work is in progress by the Woolfson group- will be necessary to test whether this 3.3-fold deviation with respect to experiments, actually consistently overestimates the conductance for other channels.

4.6.3 Simulation of ion permeation: CCTM pores

Following a similar procedure to cWzaY373C, we simulated dynamics of both CCTM-Pent-KLLW and CCTM-Hept constructs, at high voltages $\pm 1.2V$, while keeping positions of their atoms restrained for 25 ns of dynamics. Here, we summarise preliminary results for this simulation stage for "one single simulation" per voltage value. Detailed analysis of these will be performed later as well.

Unlike cWzaY373C, analysis of the trajectories did not reveal any observable permeations across CCTM-Pent-KLLW throughout the 25 ns interval at $\pm 1.2V$. In contrast, results of CCTM-Hept at both $\pm 1.2V$, show that observed ion permeation is the outcome of a process, which in its simplest form, requires that:

- Initially, a salt bridge between a K^+ and a Cl^- ion forms. The resulting dipole is aligned in the direction of the external electric field. Figure 4.14(I) shows the case at $-1.2V$, with the electric field pointing vertically upwards and the dipole aligned in its directions.

- When the dipole is near the lower channel end, with the Cl^- ion close to the end and the K^+ ion on top, an external K^+ ion from the bulk approaches the lower end of the channel, gets attracted to the Cl^- ion and then forms another salt bridge with the $Cl^-$of the dipole (Figure 4.14, II).

- Subsequently, the dipole becomes unstable owing to mutual repulsion between K^+ ions. Once the top K^+ ion is weakly attracted to the central Cl^- ion, water molecules cage and separate the K^+ ion. Then, this K^+ moves upwards towards the upper channel end, away from the remaining ions, but without permeating. Now, the Cl^- ion and the previously recruited K^+, form a single dipole pointing downwards, placed at the lower end of the channel and pointing in opposite direction to the external field (Figure 4.14, III).

- As a misaligned dipole creates an unfavourable energetic condition, the dipole flips to align with the external field. And once alignment takes place, we end up with a situation like at the beginning (Figure 4.14, *fourth left-to-right*). Then, a new K^+ ion from the bulk is recruited, destabilising the new dipole, once this happens the K^+ ion previously left near the upper channel end (shown in Figure 4.14, IV) is permeated shortly.

On the other hand, an additional assay of CCTM-Pent-KLLW at$+2V$ of 5 ns of dynamics, shows a scenario somewhat similar to the formation of ion salt-bridges, mediating permeation, but without observed recruitment of bulk ions. However, this initial trajectory is too short to draw any conclusion. We must acknowledge that the above-described permeation mechanisms while seems to be dominant, other scenarios are also possible. For instance, two aligned dipoles can be found simultaneously inside the channel. So, translocation of a K^+ ion becomes a competitive scenario between the two dipoles. Moreover, we should keep in mind that the current evidence is merely based on one single simulation, and with the channel being position-restrained; this will

require verification in future repetitions of the simulation, at same voltages, and also at lower voltages closer to experiments. The effect of removing position restraining of the protein will be addressed in the future as well.

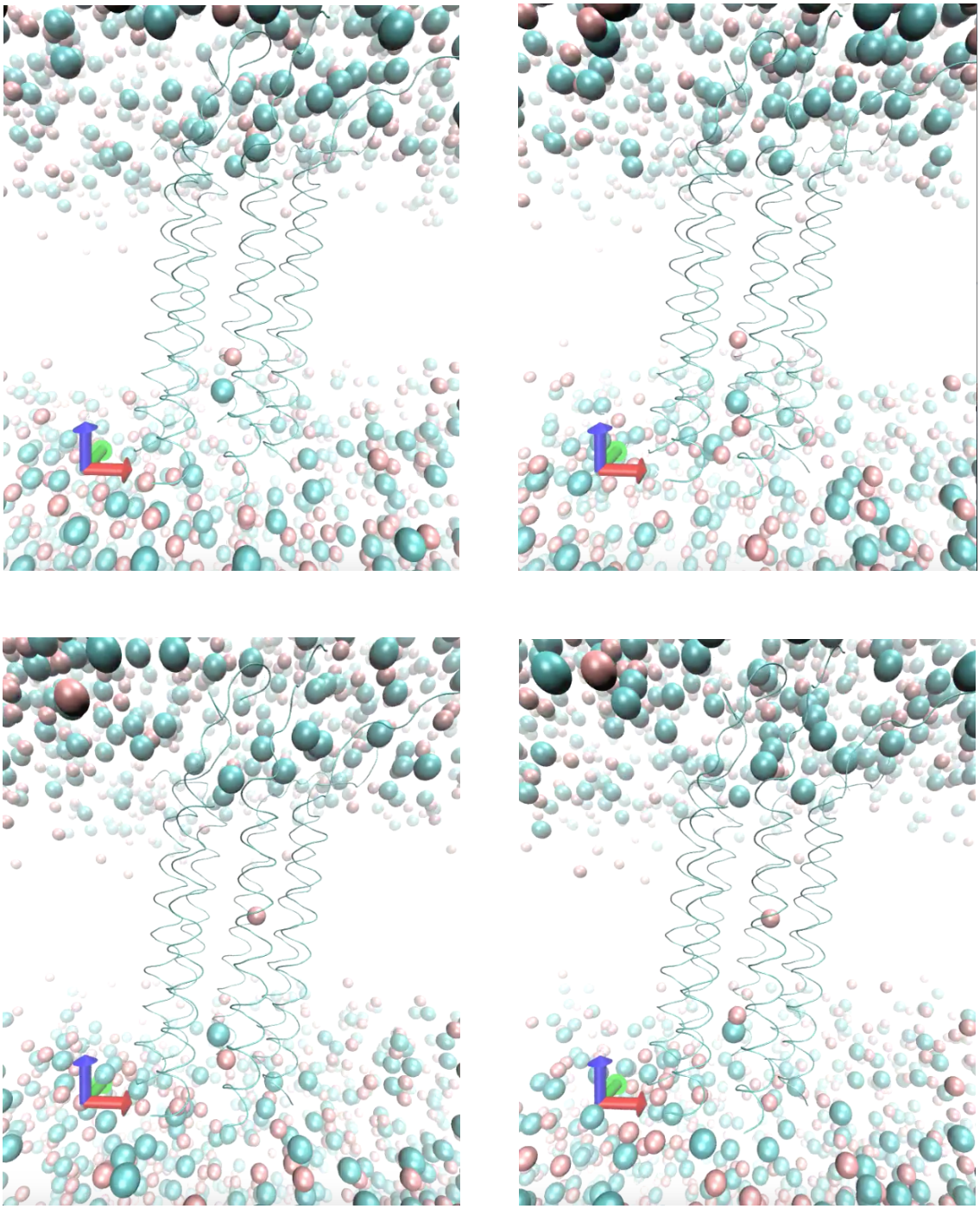


Figure 4.14: Series of snapshots of the *single-dipole mediated gating* mechanism found in PR-simulation of CCTM-Hept at voltage −1.2 V. Key stages are presented: **I)** dipole formation, **II)** dipole destabilisation by external K^+, **III)** new dipole formation, and **IV)** dipole alignment. Water molecules and lipids were removed for clarity, also CCTM-Hept ribbon representation was made the thinnest possible to allow visibility of ions within the pore. Effective ionic radii of depicted K^+(pink) and Cl^-(green) are 1.38 and 1.81Å.

4.6.4 Simulation of ion permeation: aHL-CC hybrid pores

Now, we look at the outcome of simulating the aHL-hybrid constructs. Similarly to CCTM results, here we only summarise results at high voltages of $\pm 1.2V$ and under PR-protein conditions, for 25 ns of dynamics. Like CCTM, simulated trajectories revealed a poor number of permeation events across the whole length of the channel, with one single ion crossing for the entire 25 ns period at $+1.2V$ across aHL-CCHept, and none at $-1.2V$ across the same channel, and for aHL-CCHex2 at both $\pm 1.2V$.

Despite having "one single permeation" event, this gives us insight into what a possible mechanism of permeation may take place across hybrid $\alpha\beta$-channels. A series of observations are made:

- First, we note that in all present simulations of aHL-CCHex2 and aHL-CCHept, regardless of the applied voltage, most of the time the interior of the α-helical part stays in a state of vacuum or *gas phase*; expected due to the hydrophobic nature of its core as well as their small diameter $\phi < 7$Å [11]. By contrast, the membrane part remains filled with electrolyte all the time (Figure 4.15, *Top left*).

- However, at $+1.2[V]$ across aHL-CCHept, the single permeation event of a Cl^- ion reveals that once this is inside the membrane part of the channel, to permeate across the α -helical barrel this should first reach the water-vacuum interface, at about the hight of the linker region. Then, at the interface, the ion becomes exposed to the vacuum so that this is "pushed", while pulling neighbouring water molecules with it, forming a "water jet" that moves up and down. See Figure 4.15, *Top right* and *Bottom left* for a close-up.

- Eventually, as the ion is persistently moved upwards the vacuum region, the "water jet" exceeds a critical hight after which the ion alongside a few water molecules detach themselves from the water jet coming from the bottom (Figure 4.15, *Bottom right*). After this, the ion and the water surrounding it, move towards the upper end of the hydrophobic core of the α-barrel. Once there, the ion and its water shell incorporate to the electrolyte bulk.

Further inspection of the MD trajectories reveals that as the permeant ion in the water jet moves up and down in the hydrophobic region, accumulation of Cl^-ions near the water-vacuum interface takes place, while K^+ ions accumulate at the upper end of the α-barrel. This suggests that permeation may be the result of an electric potential gradient across the vacuum region, as ions of opposite sign accumulate at opposite ends of it. Similarly, for aHL-CCHex2 at $+1.2V$, a water jet is also observed, but without successful permeation.

When looking at the trajectories of aHL-CCHept at $-1.2V$, unlike the case at $+1.2V$, a rather different scenario with no effective accumulation of ions of one species within the membrane takes place. Instead, transport of Cl^- ions towards the interior of the membrane is observed,

which makes us suspect that the motion of Cl^- should be due to the dominant local electric field rather than to the externally applied one. Simulation of aHL-CCHex2 at $-1.2V$ reveals a similar behaviour.

Finally, closer look at all MD trajectories (both hybrid pores, at $\pm 1.2V$) reveals permeation of ions and water takes place on the side of the channel through the linker region. It is unclear whether this is the consequence of either i) applying large voltages while PR the protein -and hence not reproducible in both simulations at lower voltages and without PR and experiments at $\pm 1.2V$ due to membrane breakdown- or ii) an artificial structural defect of the computational design of the pore -since atomic coordinates were produced from combined crystallography data (Sec., 4.6.1), which might not necessarily exist for the real pore structure.

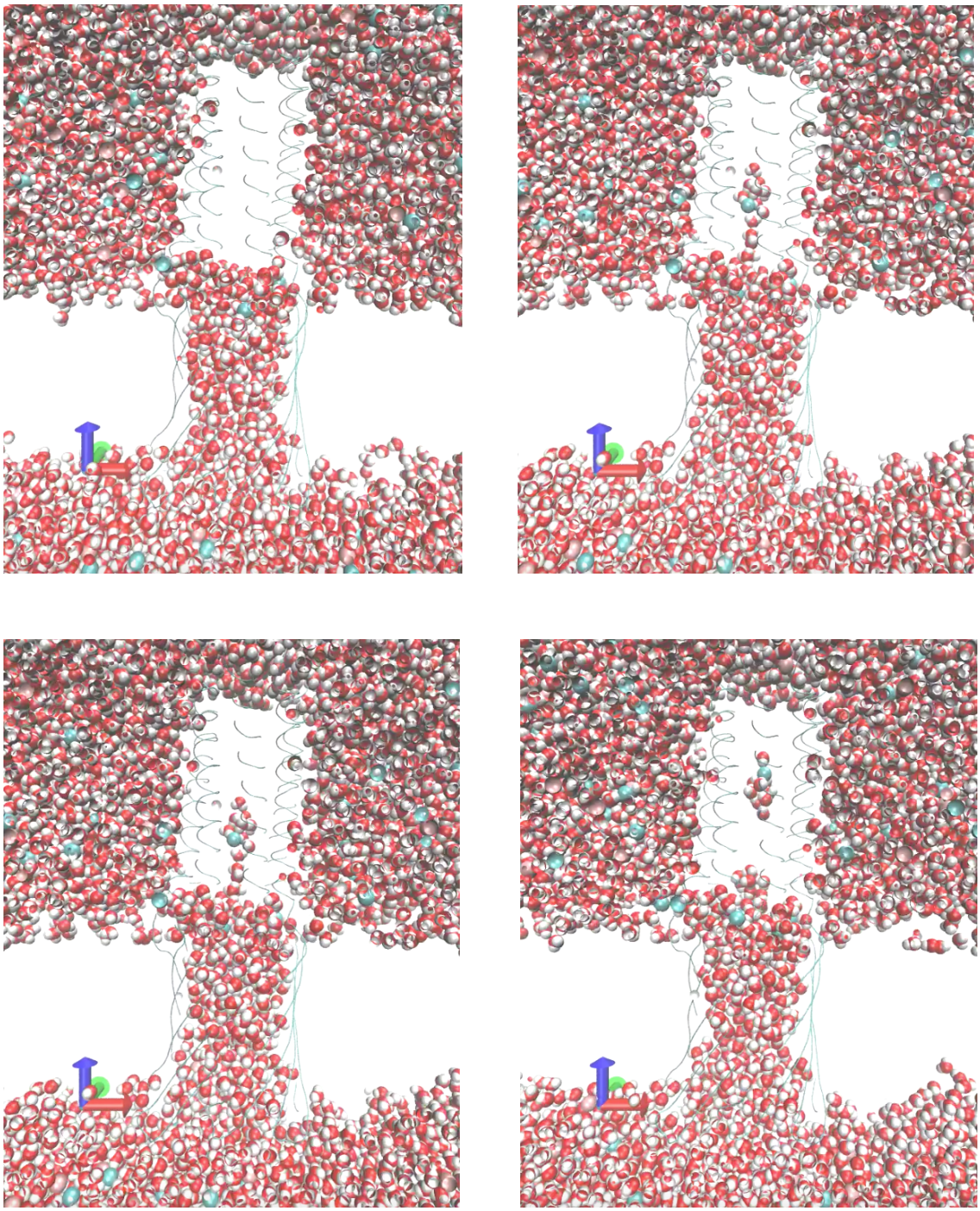


Figure 4.15: Series of snapshots of *hydrophobic gating* mechanism found in PR-simulation of aHL-CCHept at hyper-voltage +1.2 V. Key stages are presented: **I)** hydrophobic core in gas phase, **II)** formation of water jet containing Cl^- permeant ion, **III)** close-up of water jet, with Cl^-ion in blue, and **IV)** breakdown of water jet and ion permeation. Lipids were removed for clarity, ribbon representation of the protein was made the thinnest possible to allow visibility of ions and water within the pore. Water molecules are shown with the Van der Walls representation; Oxygen (*red*), Hydrogen (*white*). Effective ionic radii of depicted K^+(*pink*) and Cl^-(*green*) are 1.38 and 1.81Å.

4.7 Conclusions

As main the main outcome, employing a model of the cWza-Y373C channel; known to be stable in simulations, we defined an MD protocol to estimate a channel's conductance by determining the amount of ion flux that passes through it due to an applied voltage difference. The best-estimated average conductance (2.5 nS) of this channel was 3.3- fold the experimental one (0.75 nS); also nearly 2-fold with respect to HOLE (1.2 nS). Although not close enough to the experimental value; in contrast to similarly reported calculations [2, 99], ion flux was found stationary and conductance to be constant for tested voltages ($< \pm 200$ mV), which was consistent with observations from electrophysiology experiments with cWza-Y373C channels (Section 1.1).

Although similar calculations can be attempted with similar small peptide-assembled channels, issues limiting the accuracy of our conductance calculations per se must be considered. Finite-size effects, slow current relaxation, and peptide-assembly instability due to peptide-lipid binding were revealed as artefacts leading to ion current enhancement and non-stationarity. These were tested and fixed respectively by *1)* choosing an appropriate box simulation length in the direction of the applied electric field; *2)* using hyper-voltages (± 1.2 V) to accelerate ion current relaxation prior to target voltage values ($< \pm 200$ mV), and *3)* by position-restraining lipids for all stages prior to simulations with target voltage values. Additionally, conductance estimates were found to be subject to fluctuations too as a result of the imposed secondary structure (folding) of residues narrowing the channel's C-terminal entrance (*Trp-374*, *Pro-375*, *Asn-376* - missing in the structure of Wza); protocol variations, due to velocity generation between stages; and intrinsic fluctuations among replicas, while keeping a fixed folding of terminal residues and a common protocol.

Nevertheless, further assays with CCTM and aHL-CC hybrid channels revealed that for very narrow channels, conductance calculations are limited due to a lack of permeation events as a consequence of their size; even at extremely high-voltages (above ± 1.2 V). Despite this, our simulations were proven to be informative about the possible permeation mechanisms taking place within such narrow channels, such as *hydro-phobic gating* [10]. Alternative techniques such as *coarse-grained* MD simulations can be potentially employed to enhance ion permeation [63]. Also, *biased-sampling algorithms* such as umbrella sampling [107], meta-dynamics [55], and non-equilibrium pulling [45], are techniques that are likely to be used to enhance ion permeation rates, as the movement of ions is forced to a pre-defined path along with a reaction coordinate. These will be subject of future work though.

CHAPTER 5

CONCLUSIONS AND OUTLOOK

cWza is a synthetic amino-acid sequence that is 35-residues long. This sequence was engineered, via the redesign of the natural sequence of the outer-membrane part (Ala-345 to Asn-379) of Wza; the polysaccharide exporter in E. coli. Peptides synthesised with the cWza sequence fold as alpha-helices. Also, even when the cWza sequence is modified with punctual Cysteine (C) mutations (*aka*, cWza Cys-mutant sequences) [64].

Single-channel recording experiments with these synthesised peptides have been already conducted [64]. Peptides are typically released in a chamber with a planar lipid bilayer dividing an electrolyte medium into two. Identical peptides with a cWza *Cys*-mutant sequence were reported to be able to assemble in parallel and form hollow pores spanning lipid bilayer, hence allowing ions to pass through, i.e., *ion channels*. For some channels (cWza, cWza-K375C, cWza-S355C sequences), their individually measured electric currents transition alternately between low and high conductance under the same applied voltage (+100mV). But channels made of peptides where Tyrosines (Y-373) are replaced with Cysteines (C) - *aka* cWza-Y373C - display a single conductance regardless of voltage ($|\Delta V| < \pm 200$ mV). A hypothesis is that the transition between low and high conductance states of these channels is the result of a structural transition between narrow and wide conformations. While for cWza-Y373C channels, a single conformation was hypothesised [64]. However, experimental structural data to support these hypotheses remains unavailable, which is often difficult to obtain for membrane proteins overall. Nevertheless, various recent advancements have allowed computational methods to become a powerful alternative to overcome this limitation [29].

In our research, we aimed to reveal via computation whether alternate conformations are the mechanism explaining observed changes in conductance for some of the cWza *Cys*-mutant channels. While for cWza-Y373C channels, whether a single conformation can only exist. Thus,

we employed Peptide-Peptide Docking to predict three-dimensional structures of the cWza *Cys*-mutant ion channels. Then, we examined whether cWza, cWza-K375C, and cWza-S355C predicted structures could be classified according to narrow and wide conformation groups, as suggested in experiments. While for cWza-Y373C structures, whether a single group of conformations could be distinguished. Additionally, we studied the conductance of predicted structures to distinguish their conductance states, and whether these could be associated with distinguishable conformations. Finally, we developed a Molecular Dynamics simulation procedure to study ion permeation in atomic detail, across small peptide-assembled channels under applied voltage conditions; like in experiments for cWza *Cys*-mutant channels. The main conclusions of our research are described below in further detail.

5.1 Classification of Conformations of cWza *Cys*-mutant Ion Channels

After employing docking to obtain candidate structures of cWza *Cys*-mutant channels, we aimed to find out whether narrow and wide structural conformations could be distinguished for predicted cWza, cWza-K375C, and cWza-S355C channels. While for cWza-Y373C structures, whether a single conformation existed.

Based on the minimum radius and the length of all-atom VdW radial profiles from predicted structures, it was revealed that cWza, cWza-K375C, and cWza-S355C channels were likely to fall within two separate groups, that we could relate to narrow and wide conformations. While for cWza-Y373C channels, a single conformation group was only found. However, we also validated this conformation classification under dynamic conditions. Thus, we performed equilibrium Molecular Dynamics simulations (100 ns) of membrane-embedded structures sampled from previously found docked conformations according to their sequence. By the end of every simulation, the dimensions of every structure from the last snapshots (10 ns) were determined according to a decomposition of their backbone radius of gyration around axes parallel and perpendicular to the membrane plane. Again, structures with sequences cWza, cWza-K375C, and cWza-S355C were found to fall within separate clusters according to the components of their radius-of-gyration. While for cWza-Y373C channels, an overall single cluster was identified under the same procedure.

Overall, our computational results are in good agreement with hypothesised conformations from single-channel recording experiments. The combined outcomes before and after simulations show the existence of narrow and wide conformational groups for cWza, cWza-K375C, and cWza-S355C channels, while showing a single one for cWza-Y373C channels.

Additionally, Hydrogen-bonds and Knobs-Into-Holes were found to play an important role in the differentiation of conformations of cWza *Cys*-mutant channels (Subsection 3.2.4), as these interactions resulted to be linked to the cluster separation by the end of dynamic simulations

(Subsection 3.3.4). Also, insight was gained into the inner surface properties within channel structures, as well as how mutation of *Tyrosines* (Y) with *Cysteines* (C) in peptide sequences may promote peptide-peptide interactions that are key to make alternate conformations to collapse to a single one (Subsection 3.3.5).

Future research is needed to address some of the possible limitations of our current methodology. Accuracy of our conclusions could be enhanced by **1)** an *extended sampling* of docking parameters - regarding the selection of the peptide master unit for assembly or the symmetry type; **2)** use of *alternate metrics* from VdW radial profiles for classification of docked models - e.g., the mean or the coefficient of variation of the profiles; **3)** use of a *lipid model* matching the composition of the bilayer in electrophysiology experiments, DPhPC - known to have an increased stiffness in relation to POPC in experiments and simulations, hence possibly affecting the dimensions of simulated structures [93, 117]; **4)** by a systematic modelling of the secondary structure (as α-helical, β-strand, or linear coil) of *missing residues* at C-terminal peptide ends - *Trp*-377, *Pro*-378, and *Asp*-379; although these may be relevant for conductance estimation, as they have a significant effect in ion permeation simulations, their absence in docked models did not affect the stability of end conformations; and finally **5)**, *fixed ionisation states* were only considered - this may limit the extent of our results, since Histidine was persistently found to participate in peptide-peptide interactions. Histidine can be uncharged or positively charged, as Histidine's pKa[1] values can easily be perturbed by its local environment [14]. Thus, Histidine has been suggested to act as a molecular switch triggering large conformational changes in protein complexes [38, 39, 109]. Calculations to determine the most probable ionisation state of amino-acids like Histidine and others can be performed via available software that deals with pKa calculations; see Reference [63] for a comprehensive list.

5.2 HOLE Cannot Distinguish Alternate Conformations for Some cWza *Cys*-mutant Ion Channels

As a comparison, we employed the HOLE programme to find out whether this could also distinguish between narrow and wide structural conformations from predicted structures of the cWza, cWza-K375C, and cWza-S355C channels. And similarly, whether a single conformation was only found for cWza-Y373C structures. HOLE is a widely used programme for numerical conductance and pore geometry estimation of ion channels provided a three-dimensional structure [96, 98].

For all predicted structures, we used the HOLE programme to estimate their conductance values. Their pore dimensions were determined too, as the mean-radius and the length of their HOLE solvent-accessible radial profiles. Thus, cWza channels were found likely to fall within two separate groups of low and high average HOLE conductance. These groups respectively matched

[1]It is the equilibrium constant for acid dissociation - when hydrogens bind or unbind to a molecule, e.g., an amino-acid.

two groups with narrow and wide mean-radii, but with a common average length. While for cWza-Y373C structures, a single high-probability group of conductance matched structures with a common mean-radius and length, on average.

However, cWza-K375C and cWza-S355C structures showed mixed results. cWza-K375C pore dimensions displayed a similar split into narrow and wide mean-radii, with a common length, but with poor resolution of low and high HOLE conductances. While for cWza-S355C structures, neither their HOLE conductance nor their mean-radii displayed a bimodal split that could be resolved; only their length (Appendix E).

These results show that using the HOLE methodology to judge conformational separation is limited, as seen in predicted structures of cWza *Cys*-mutant channels. We suspect this a result from the over-smoothened HOLE radial profiles, which lose molecular detail, in contrast to our previous results from VdW radial profiles. Similarly, its ability to resolve conductance states was found limited. Perhaps, as a result of the intrinsic conductance estimation criteria employed by HOLE.

Based on our results, the limitations of HOLE conductance and conformation predictions should be considered before its application. The assumed linear correlation between experimental conductance and HOLE estimated minimum radius does not hold for many channels (See Subsection 2.4 of Reference [96]). Also, as cWza conductance estimates did not agree with the experimentally determined ones (2-fold), neither their Low-to-High conductance ratios, future research could address the issue of improving the HOLE methodology for conductance prediction.

5.3 Study of Ion Permeation with Molecular Detail Across Small Peptide-Assembled Ion Channels

At present, for the cWza *Cys*-mutant channels, the details of how ions permeate across these are simply absent via experiments; usually challenging to resolve. Knowledge of the ion permeation mechanisms is crucial to better understand the structure-to-function relationship for small peptide-assembled ion channels overall. Thus, we aimed to develop a procedure for Molecular Dynamics simulations to observe with atomic resolution ions passing through small-peptide ion channels, to identify channel-specific permeation mechanisms.

Based on previous work on Molecular Dynamics of ion channels [2, 100] and testing with benchmark systems, we developed a procedure for MD simulation of ion permeation across small peptide-assembled channels embedded in POPC lipid bilayer. Our procedure mainly deals with the protein-bilayer-solvent set up before the simulation of unrestrained dynamics with an applied target voltage ($|V| \leq \pm 200$mV). Thus, artefacts affecting the rate of ion permeation across channels and the structural stability of the embedded channels are minimised during the simulation setup (Subsection 4.3).

We applied our procedure to a model of the cWza-Y373C ion channel, found stable in an MD

simulation (200 ns) without applied voltage ($\Delta V = 0$ mV) [64]. In our simulations (30 ns) we applied voltage ($|V| \leq \pm 200$mV). Thus, ion flux through the cWza-Y373C channel was measured and found stationary, for all applied voltages. While the channel's structure remained stable too (Subsection 4.6.2). As for the permeation mechanism, ion size was the limiting factor for permeation, since the channel's pore cavity was wide enough (≈ 6Å, minimum radius).

Using the same procedure, we simulated two Coiled-Coil Trans-Membrane (CCTM) channels and two α-Hemolysin-Coiled-Coil (aHL-CC) hybrid channels. CCTM channels are assembled from *alpha*-helical peptides with *de novo* (i.e., not observed in nature) amino-acid sequences. While aHL-CC hybrid channels are made of peptides with one half of their sequence taken from the membrane-spanning part of the natural α-hemolysin channel, linked to another half corresponding to the sequence of a *de novo* water-soluble α-helical barrel. Models were built and provided by members of the Woolfson group (University of Bristol); see Subsection 4.6.1. MD trajectories (30 ns, $\Delta V = \pm 1.2$ V) revealed details of two different mechanisms driving ion permeation through these two types of channels. For CCTM channels, ions permeated provided the presence of a K-Cl dipole within the channel, that when aligned to the external field, was found to facilitate permeation; we called this *dipole-mediated* permeation (Subsections 4.6.3). While for aHL-CC hybrid channels, Cl ions were seen to selectively permeate through the hydrophobic pore cavity of the part of the channel protruding the bilayer; aka *hydrophobic gating* (Subsection 4.6.4). However, ion permeation events were scarce for all simulation. Overall, our results provide atomic-level insight into different permeation mechanisms of simulated peptide-assembled channels; simply unknown from single-channel recording experiments.

Further research will be needed to address issues found in the sampling of ion permeation events. For the cWza-Y373C channel, ion flux measurements were found overestimated in comparison to experiments (3.3-fold) [64]. The flux was sensitive to the secondary structure of the last three C-terminal residues in all peptides, as these residues limit the dimensions of the channel's pore at one end (Appendix H). These residues are originally absent in the crystal structure of Wza; used to build the cWza-Y373C model (Supplement, Reference [64]). Future studies could focus on a systematic modelling procedure of these missing residues and their effects in the measured in flux. Thus, simulations could be potentially used for conductance calculation predictions. For CCTM and hybrid aHL-CC channels, poor observation of permeation events was noticeable, even under high applied voltages ($\Delta V = 1.2, 2$ V). The narrow pore dimensions of CCTMs and the high hydrophobicity within the soluble pores of hybrid aHL-CCs, both pose large energy barriers for permeation [11, 46]. Extended simulations could enhance permeation events, however, this is computationally demanding. As an alternative, coarse-grained MD and enhanced sampling techniques could be employed (Subsection 4.7). Also, the use of a lipid model matching the one in experiments (DPhPc [28, 93]), the use of a more sophisticated force field choice (e.g., CHARMM36 [52, 90]), and prolonged simulations (beyond 30 ns), are common issues that may require attention in future research.

5.4 Implications For Membrane Protein Engineering and Design

Here, we have focused only on *synthetic* barrel-like channels assembled from α-helical peptides with cWza Cysteine-mutants sequences by the *Woolfson group* (University of Bristol) [64]. However, these are just one - although unique in its type - amongst many types of potential synthetic channels made of membrane-spanning α-helices, with either sequences engineered from natural ones via mutation (*redesign*) or designed sequences unseen in nature (*de novo*) [13, 33, 72, 94]

Consequently, provided further advances in structure prediction of synthetic ion channels, our work fits in the immediate next step of providing insight into whether modelled structure separate in conformation (Subsection 3.3.1), and if so, how these conformations relate to function; here, conductive properties. Also, as a potential side application, our framework for analysis of peptide-peptide interactions (Subsection 3.2.4), may be found useful to automatically identify and classify residues involved in peptide-peptide packing with the advantage of integrating MD simulation data, and hence, perhaps to identify sequences motifs involved in folding and assembly; key to establishing *sequence-to-structure* relationships useful for *de novo* design [47, 115, 129].

APPENDIX A

Extended Geometric Analysis of Octameric Docked Structures

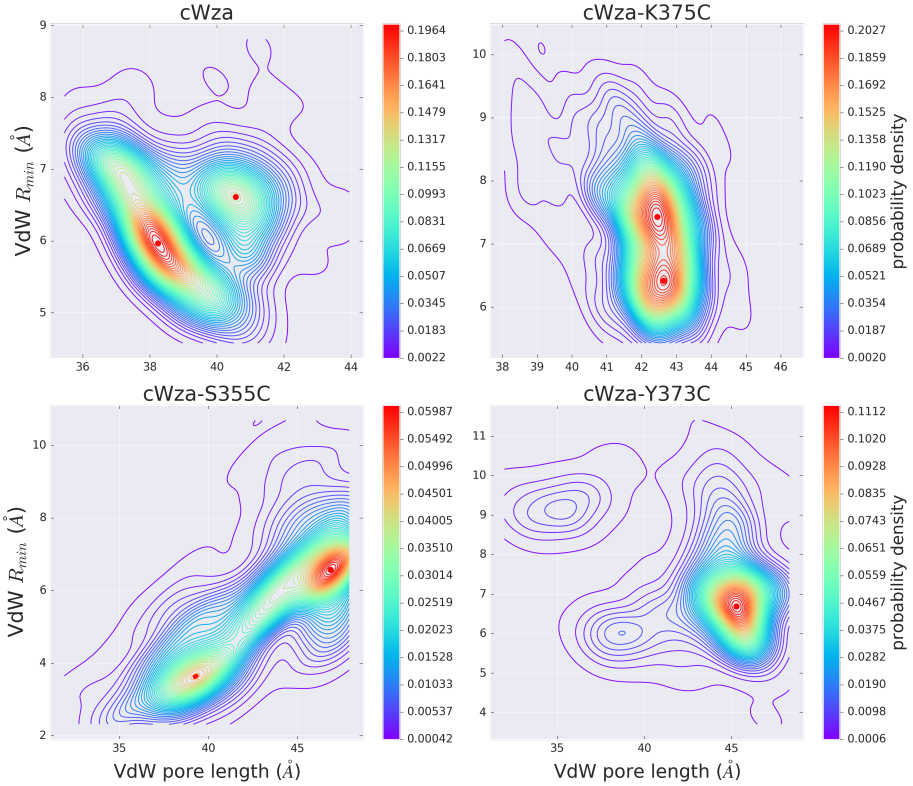


Figure A.1: Contour plots of all Probability Density Functions (PDFs) of VdW dimensions (L,R_{min}) determined from all generated docked models (unclassified), per cWza *Cys*-mutant sequence. Red dots (·) indicate the position of the highest maxima of each PDF.

Geometric Details of Single and Alternate Conformations

Extended geometric analysis of models from identified conformation groups for all cWza Cys*-mutant docked channels revealed the variable nature of structural differences for models within or between conformation groups. More precisely,* cWza *channel conformations differed overall due to whole radial displacement of backbone units and conformation of sidechains near terminal ends;* cWza-K375C *channel conformations differed overall due to conformation of* Tyrosines *(Y) sitting at the bottleneck of channels;* cWza-S355C *channel conformations differed overall due to alternate peptide backbone-backbone angles. Finally, single conformations of* cWza-Y373C *channels showed alternate peptide backbone-backbone angles, however, sidechain conformations seemed to compensate for this difference, resulting in channels with identical bottleneck radius, hence a single conformation is only seen. Also, cWza-Y373C channels seem to have distinctively flatter inner surfaces in contrast to channels with other sequences. See screenshots in Figure 3.6 for some illustration.*

Results above showed that either single or alternate conformations with distinctive pore VdW-dimensions (L,R_{min}) were identifiable, however, fine details of their pore geometry were ignored. Radial profiles condense a range of geometrical information of a pore, regarding displacement, orientation, and arrangements of backbone and sidechain atoms distinguishing structures within and between conformation groups. Moreover, this information can provide clues of how a channel structure may change as it transitions between alternate conformations, as well as distinctions in their internal surface properties.

Here, we introduce and discuss the most relevant results from the analysis of *Euler angles* (Subsection 3.2.3.2); regarding the orientation of backbone units, and *radial profiles* (Subsection 3.2.3.3); regarding backbone and sidechain atomic displacements and positioning ,to have a more accurate description of the distinctive geometry of single and alternate conformations of the cWza *Cys*-mutant docked models. Also, we introduce an analysis regarding the geometry and identity of residues sitting at the *bottleneck* of docked channels; from which inference about their internal solvent-exposed surface properties can be drawn. All the data described and discussed below can be found in the pages ahead of this section.

cWza : Euler angles (θ,ψ,ϕ)

Distributions for all Euler angles are unimodal. Hence, a preferential backbone orientation is adopted by most docked models. So, no conformation separation can be related to backbone chain orientation. Figure A.3.

cWza : Primitive radial profiles (z',γ_{prim})

The profile length (along Z-axis) distribution is unimodal, which is consistent with the fact that there is a single preferential unit orientation. However, a clear dual conformational separation

is observed for $\mu(R_{prim})$, $min(R_{prim})$, and $max(R_{prim})$ (although with lower resolution between modes), which is consistent with the idea that units separate in backbone conformation due to radial displacement. σ and $\frac{\sigma}{\mu}$ for R_{prim} show similar consistency.

cWza : VdW radial profiles $(z', \gamma_{AA,VdW})$

Although the length of the primitive profile is unimodal in distribution, this is unimodal for the VdW profiles, hence, revealing that sidechain conformations of residues on both ends of the peptide sequence change the pore length.

On the other hand, $\mu(R_{VdW})$ is also bimodally distributed like $\mu(R_{prim})$. $\sigma(R_{VdW})$ distribution remains unimodal, but $\frac{\sigma}{\mu}, min(R_{VdW})$ turn unclear in its bimodality, so they tend to unimodal. Hence, we can interpret that sidechain conformation compensates for any backbone radial movement to reduce the width of the pore in its bottleneck. For $max(R_{VdW})$ a dual separation can still be acknowledged, so sidechain conformation does not affect the pore width at the pore ends as much as at the bottleneck.

cWza-K375C : Euler angles (θ, ϕ, ψ)

Like cWza docked models, these mutant models also show unimodal distributions for all of their *Euler angles*, indicating that backbone peptide units have got a single preferential orientation.

cWza-K375C : Primitive radial profiles (z, R_{prim})

However, in contrast to cWza models, all radial metrics for primitive profiles show unimodal distributions, revealing that backbone radial displacement does not show a conformation separation.

cWza-K375C : VdW radial profiles (z, R_{VdW})

Both lengths, from primitive and VdW radial profiles show unimodal distribution. Unlike cWza, where sidechain conformation of residues on both ends affected the length of the models. So, no variable sidechain conformation takes places at the end of the pore. Similarly, $\mu(R_{VdW})$ does not show any conformation separation, given that backbone mean radial positions are unimodally distributed. Likewise, $max(R_{VdW})$. However, $\sigma, \frac{\sigma}{\mu}$ and $min(R_{VdW})$ show bimodality of variable resolution, indicating a conformation separation. Clearly, for the latter, this indicates that preferential conformations take place for the outermost residues exposed to the solvent. On the other hand, $\frac{\sigma}{\mu}$ reveals that dispersion in the VdW profiles is caused by sidechain conformation.

cWza-S355C : Euler angles (θ, ϕ, ψ)

Unlike cWza and cWza-K375C models, these mutants display bimodal distributions for all *Euler angles*. So, a clear conformation separation is seen. So, this indicates that peptide backbone

units open and close by tilting up and down, as well as narrowing and widening the angle between neighbouring peptides, which also relates to how the peptide curvature faces the pore interior.

cWza-S355C : Primitive radial profiles (z, R_{prim})

Bimodal distributions are displayed for both, lenght, $\mu(R_{prim})$, and $max(R_{prim})$ which seems to be related to the preferential peptide orientations. In contrast, $\sigma, \frac{\sigma}{\mu}$, and $min(R_{prim})$ show unimodal distributions, but with wide dispersion though.

cWza-S355C : VdW radial profiles (z, R_{VdW})

Although for backbone unit lengths are bimodal, this is partly erased as sidechain conformations are taken into account in VdW-profiles. A multimodal distribution is seen but with a dominant length value. So, sidechain conformations on both ends compensate for preferential unit orientations. For $\mu(R_{VdW})$, although similarly bimodal to $\mu(R_{prim})$, sidechain conformations reduce the resolution (separation) between modes. On the other hand, dispersion for σ and $\frac{\sigma}{\mu}$ of R_{VdW} increases, in contrast to R_{prim}. Interestingly, bimodality arises for $min(R_{VdW})$ and remains for $max(R_{VdW})$. So the effect of sidechain conformations seems to be more dramatic near the bottleneck region than for the N-terminal one (very left of the profile).

cWza-Y373C : Euler angles (θ, ϕ, ψ)

All Euler angles show a degree of multimodality, however, a single dominant mode arises for all angles. Three modes can be identified for the facing angle.

cWza-Y373C : Primitive radial profiles (z, R_{prim})

Qualitatively, these profiles show consistency with the existence of three preferential facing angles. However, the unit length is dominantly unimodal, so facing orientation does not affect this. By contrast, the mean radius shows three modes, with two modes of comparable dominance. So unit radial displacement is affected by the unit orientation, but not the length. $\sigma, \frac{\sigma}{\mu}$, and $min(R_{prim})$ are unimodal only. $min(R_{prim})$ is unimodal, but with a wide dispersion.

cWza-Y373C : VdW radial profiles (z, R_{VdW})

A single pore length remains dominant, so sidechain conformation on the pore ends does not lead to pore enlargement, perhaps due to sidechain constriction. Multimodality of $\mu(R_{VdW})$ remains, although resolution between the two dominant peaks is reduced slightly, ie., peaks get closer. σ remains unimodal. However, $\frac{\sigma}{\mu}$ shows a dispersed peak, suggesting the presence of a new mode, but with poor resolution. Thus, overall sidechain conformation leads to potential group separation. $min(R_{VdW})$ remains with a dominant mode, with a higher dispersion in contrast to its R_{prim}-version. However, sidechain conformation does not lead to effective conformation separation. Neither for $max(R_{VdW})$.

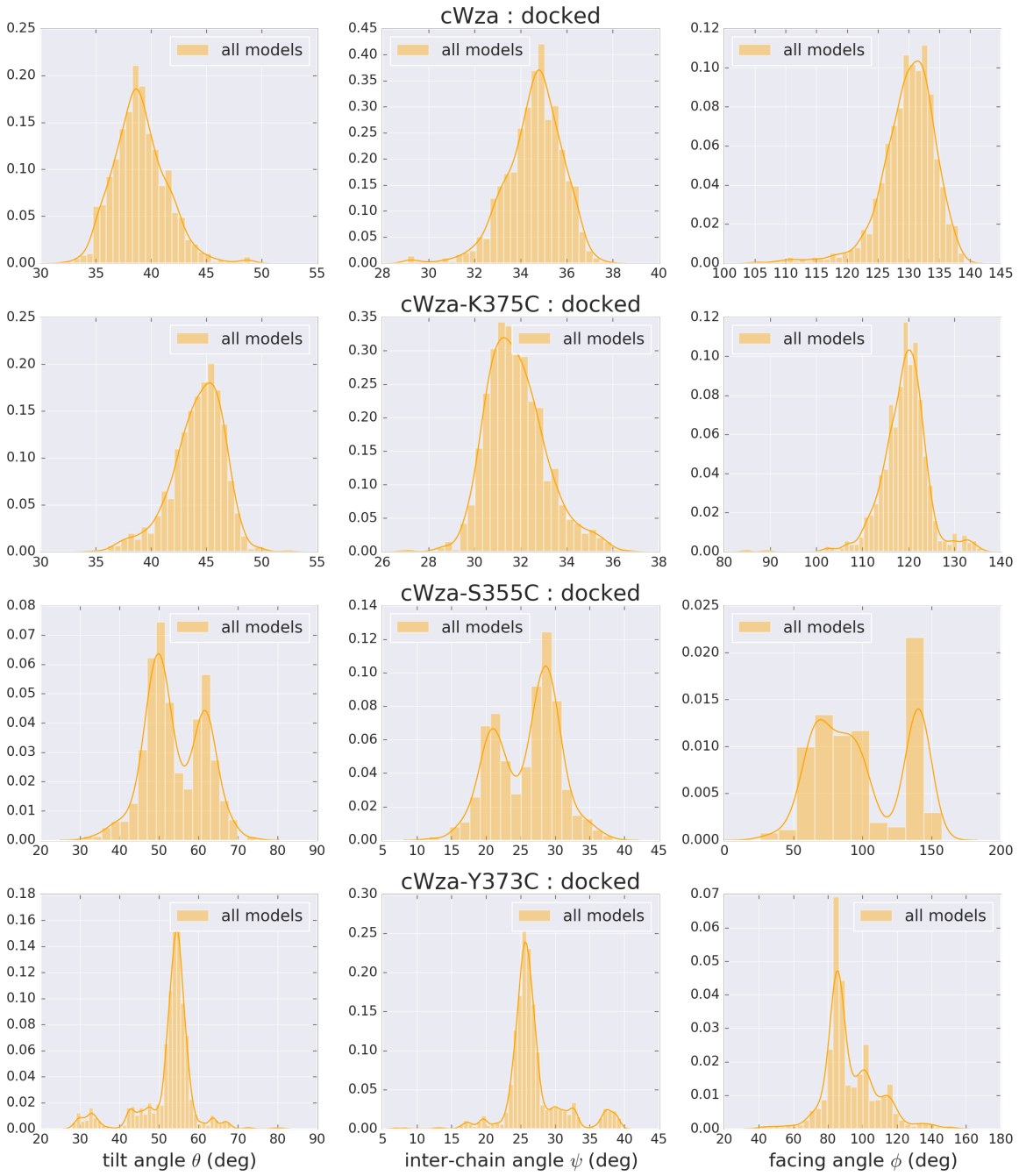


Figure A.2: Probability Density Functions (PDFs) for all **Euler Angles** (θ, ψ, φ) from all 1000 docked models (**unclassified** , i.e., without conformational classification) per cWza *Cys*-mutant sequence. These data show how peptide backbones orient in space in relation to the Intrinsic Frame of Reference of every peptide-assembly.

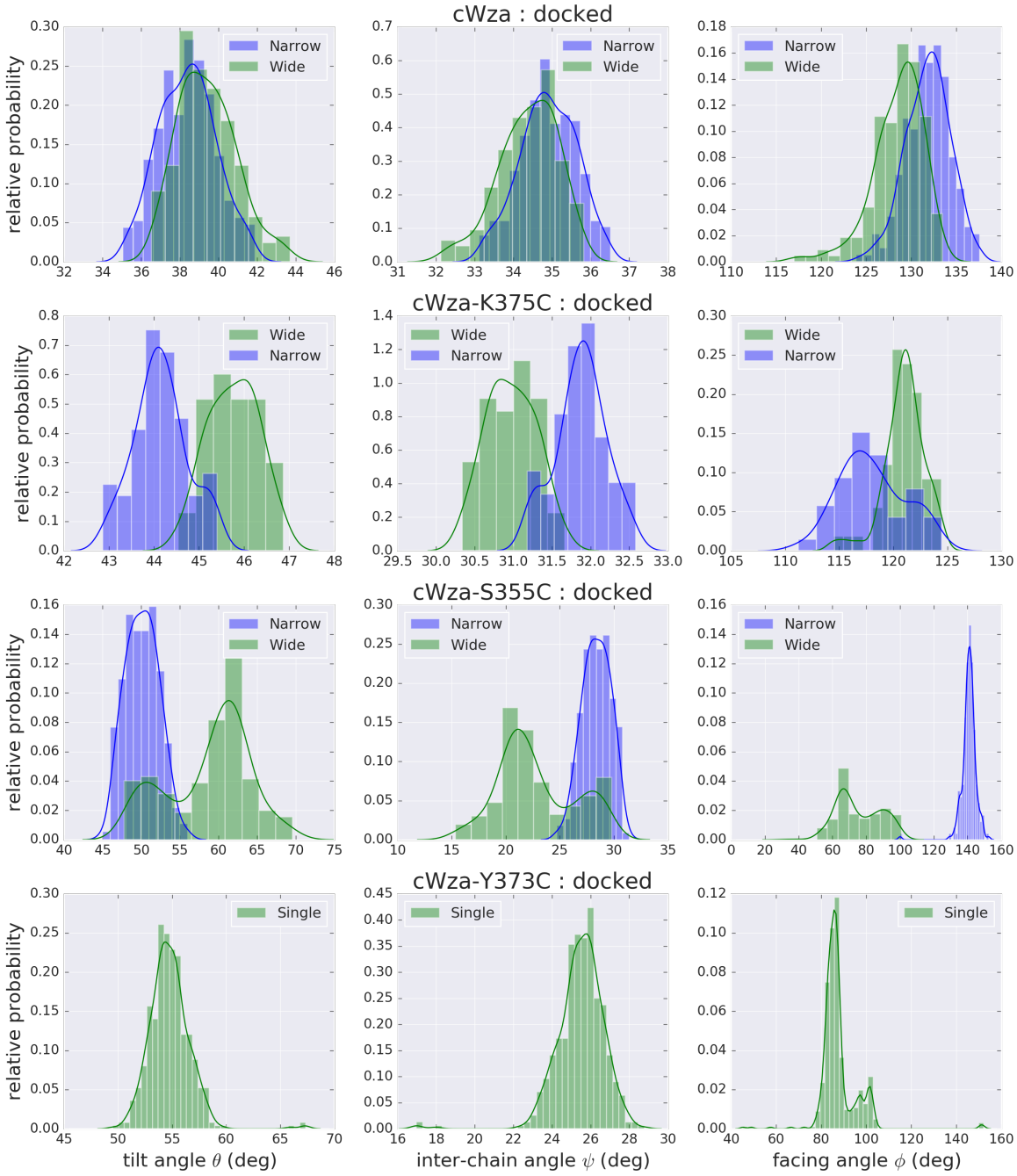


Figure A.3: Probability Density Functions (PDFs) for all **Euler Angles** (θ, ψ, φ) from all **classified** (i.e., according to conformation. See Figure 3.6 in Subsection 3.3.1) docked models per cWza *Cys*-mutant sequence.

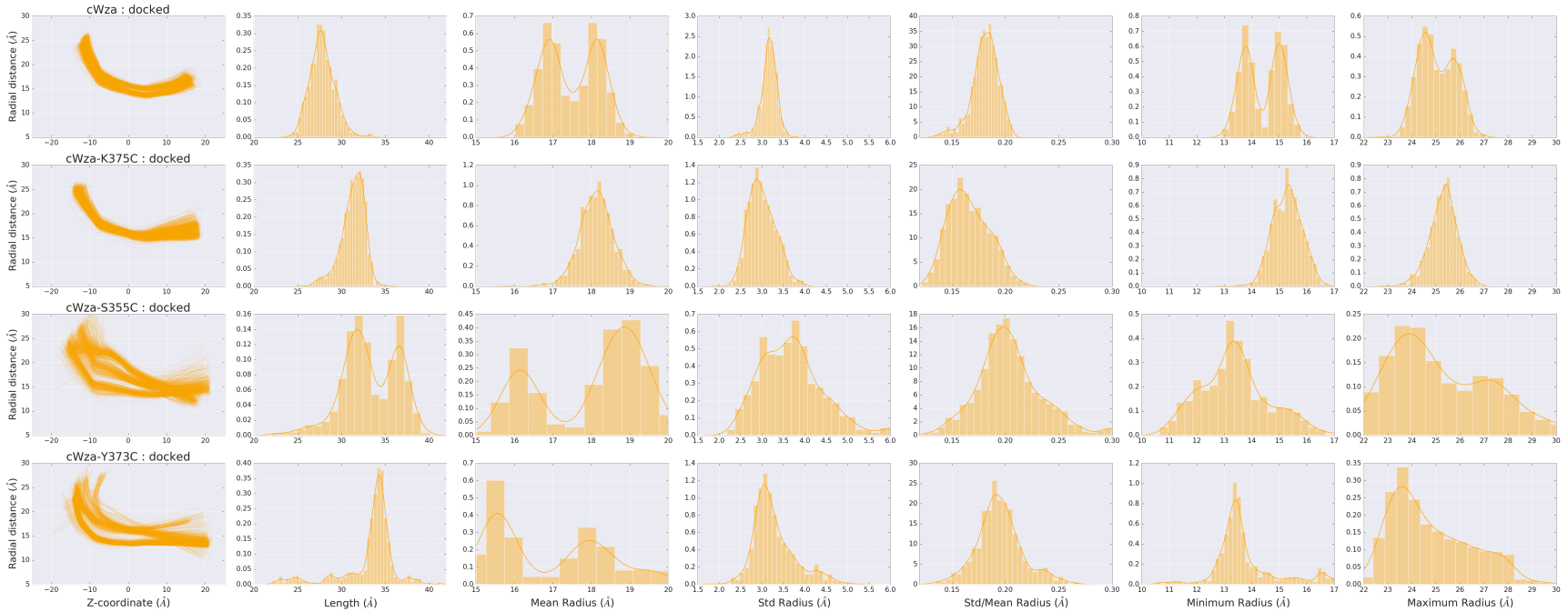


Figure A.4: **Primitive Radial Profiles** from all 1000 docked models (**unclassified**) per cWza *Cys*-mutant sequence. These profiles show how backbone units position in space in relation to the axis of assembly. Note the presence of dual modes for various radial metrics.

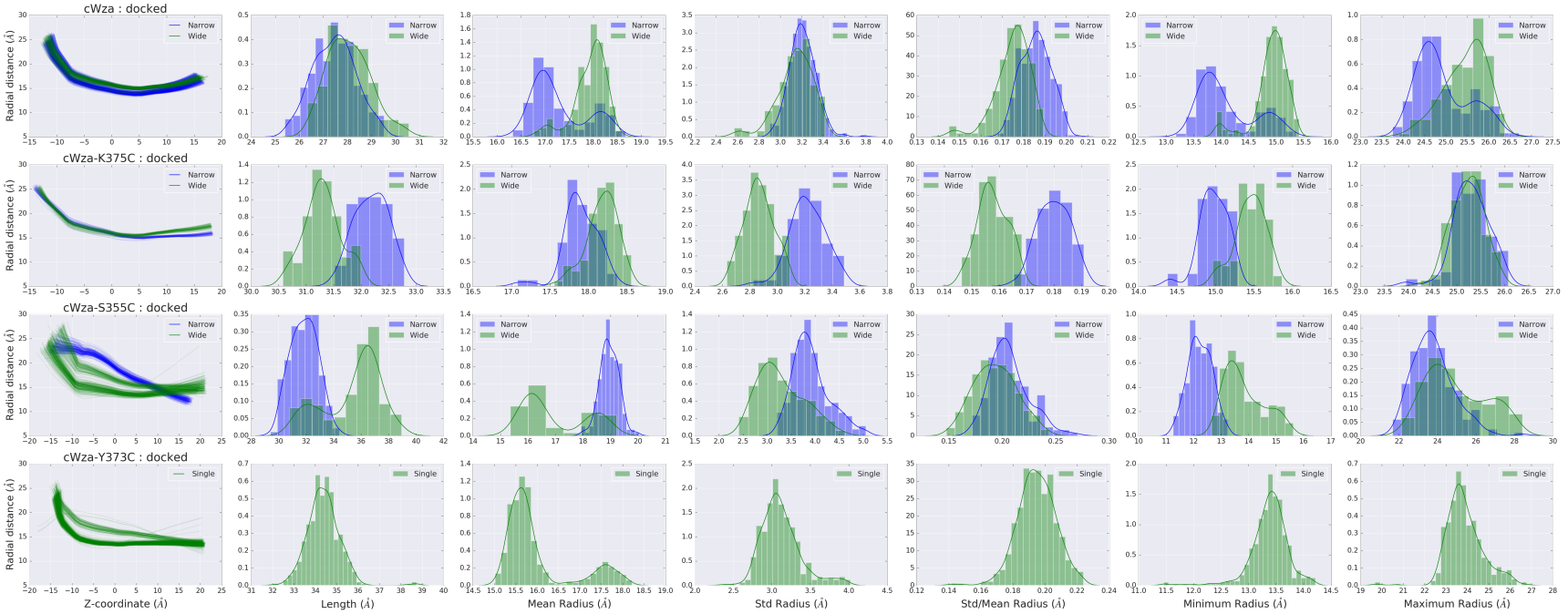


Figure A.5: **Primitive Radial Profiles** from all **classified** (i.e., according to conformation. See Figure 3.6 in Subsection 3.3.1) docked models per cWza *Cys*-mutant sequence. Note the mode decomposition induced once the conformations classification via VdW (L,R_{min}) dimensions is taken into account.

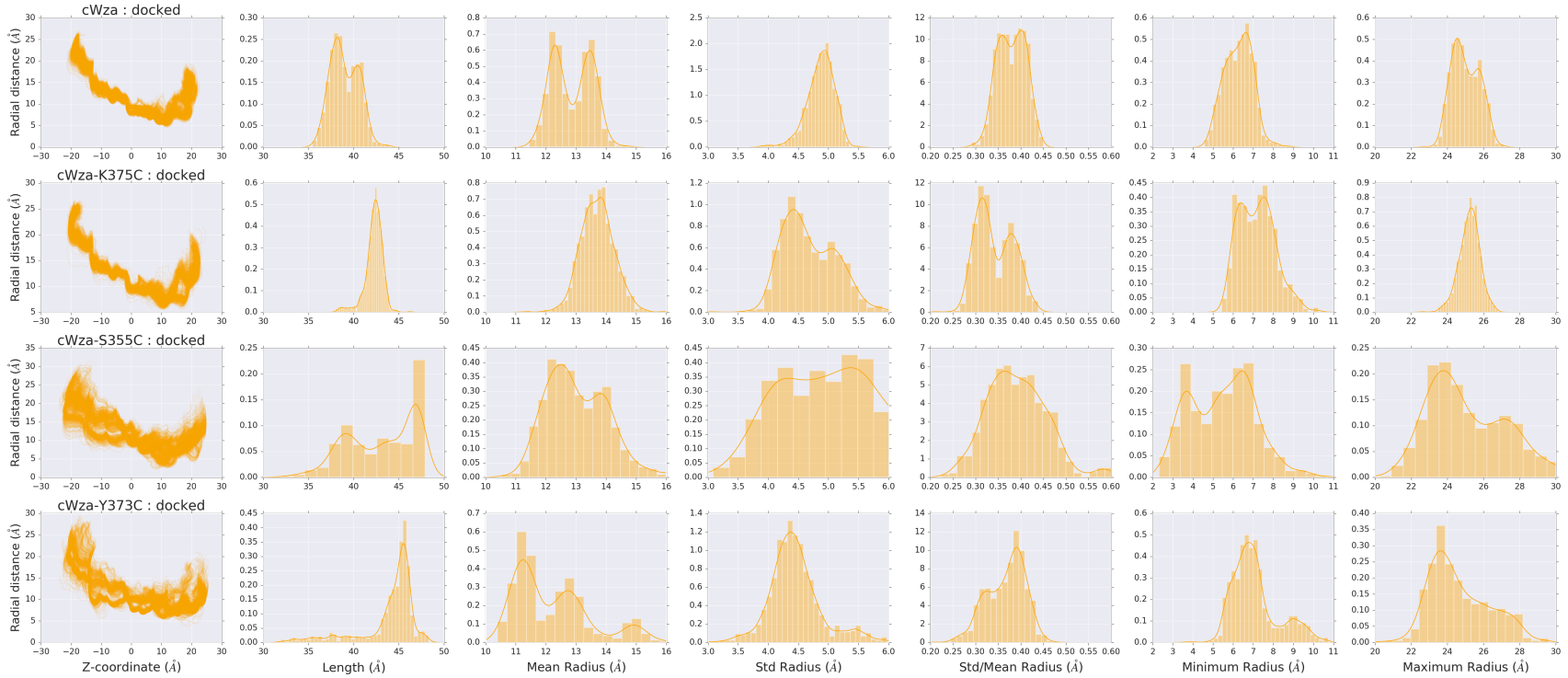


Figure A.6: **VdW Radial Profiles** from all 1000 docked models (**unclassified**) per cWza *Cys*-mutant sequence. These profiles show how side-chain atoms position in space in relation to the axis of assembly. Note the presence of dual modes for various radial metrics, as seen for primitive profiles in Figure A.

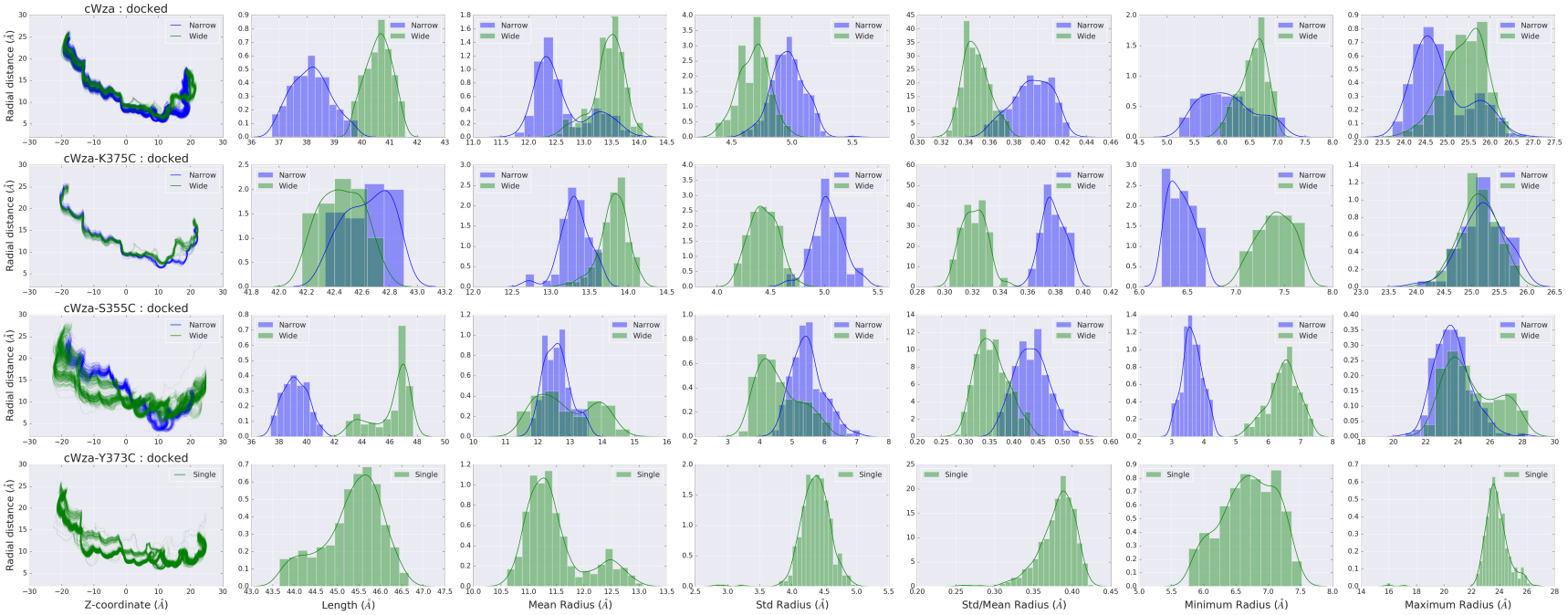


Figure A.7: **VdW Radial Profiles** from all **classified** (i.e., according to conformation. See Figure 3.6 in Subsection 3.3.1) docked models per cWza *Cys*-mutant sequence. Note the mode decomposition induced once the conformations classification via VdW (L,R_{min}) dimensions is taken into account.

APPENDIX B

SYMMETRY BREAKING AND EQUILIBRATION IN MD OF DOCKED CHANNELS

B.1 Atomic Details of Symmetry Breaking

To provide insight into symmetry breaking of docked channels we inspected the interface between lipids and limit structures carefully. Thus, we found that positively charged residues *Arg*-349 and *Arg*-352; near N-terminal peptide ends, and, *Arg*-372 and *Lys*-375; near C-terminal peptide ends, formed salt bridges [1] with the partly negatively charged *Oxygens* of Phosphates in the heads of the lipids surrounding embedded proteins, which seem to pull peptide units apart. See Figure B.1. Moreover, salt-bridges were also found present from early stages of the dynamics, that is, within the first 10 ns; and even be traced back to the end coordinates of PRMD.

It could be argued that observed lipid-peptide binding is an artefact of not including missing C-terminal residues in modelled cWza *Cys*-mutant peptides; as terminal Tryptophans (*Trp*), for instance, have been suggested to provide transmembrane α-helical peptides with anchoring to the bilayer, hence promoting structural stability [95]; additionally, added residue may screen residues from exposure to lipid heads. However, as suggested in simulation of with cWza *Cys*-mutant docked structures with added missing C-terminal residues; *Trp*-377, *Pro*-378, and *Asn*-379, salt bridge formation was still observed, moreover, structural fluctuations near terminal ends were still persistently observed with a similar order of magnitude. Additionally, assays with restrained charged residues near the end peptides in all chains showed that backbone structural fluctuations per residue (RMSF) were reduced by these localised restraints, hence, reinforcing the idea that peptides may be pulling assembling units apart. See RMSF data in Figure B.1.

In addition, it could also be argued that using a lipid model that matches the one employed in

[1] Salt-bridges were identified using the VMD Plugin *Salt Bridges*

electrophysiology, DPhPC, might affect the observation of salt-bridges and possibly symmetry disruption too, as DPhPC lipids lateral diffusion is reduced in bilayer phase, so that DPhPC bilayers can be considered as stiffer than POPC bilayers [93] [117]. However, both experimental DPhPC lipids and those employed in our simulations, POPC, share the same Phosphate groups, hence also featuring a partly negatively charged Oxygen in the heads of lipids; a common feature among *Phosphatidylcholine* (PC) lipids. So, this may imply that lipid-binding could still be observed, however, this would require testing outside the scope of this work. Salt-bridge binding of lipid head groups to *Arg* and *Lys* residues in transmembrane ion channels, has been previously inferred in experiments and simulations of natural channels such as the voltage-gated channel Kv1.2, the KscA potassium channel, and Cys-loop receptors. Also, lipid-peptide binding has been suggested to be a key modulator of gating (opening-closing) of ion channels [85] [49].

Following the above, lipid-peptide binding in our simulations seems to be an intrinsic property of the protein-membrane complex, rather than a simulation artefact. Thus, although we suspect that simulation with a stiffer bilayer model such as DPhPC may lead to a less pronounced symmetry breaking, provisionally, limit conformations were found to be steady, despite these caveats, as it will be discussed in the Section B.2 of this Appendix. Hence, we conclude that docking can effectively provide starting structures for URMD which converge to steady conformations over time.

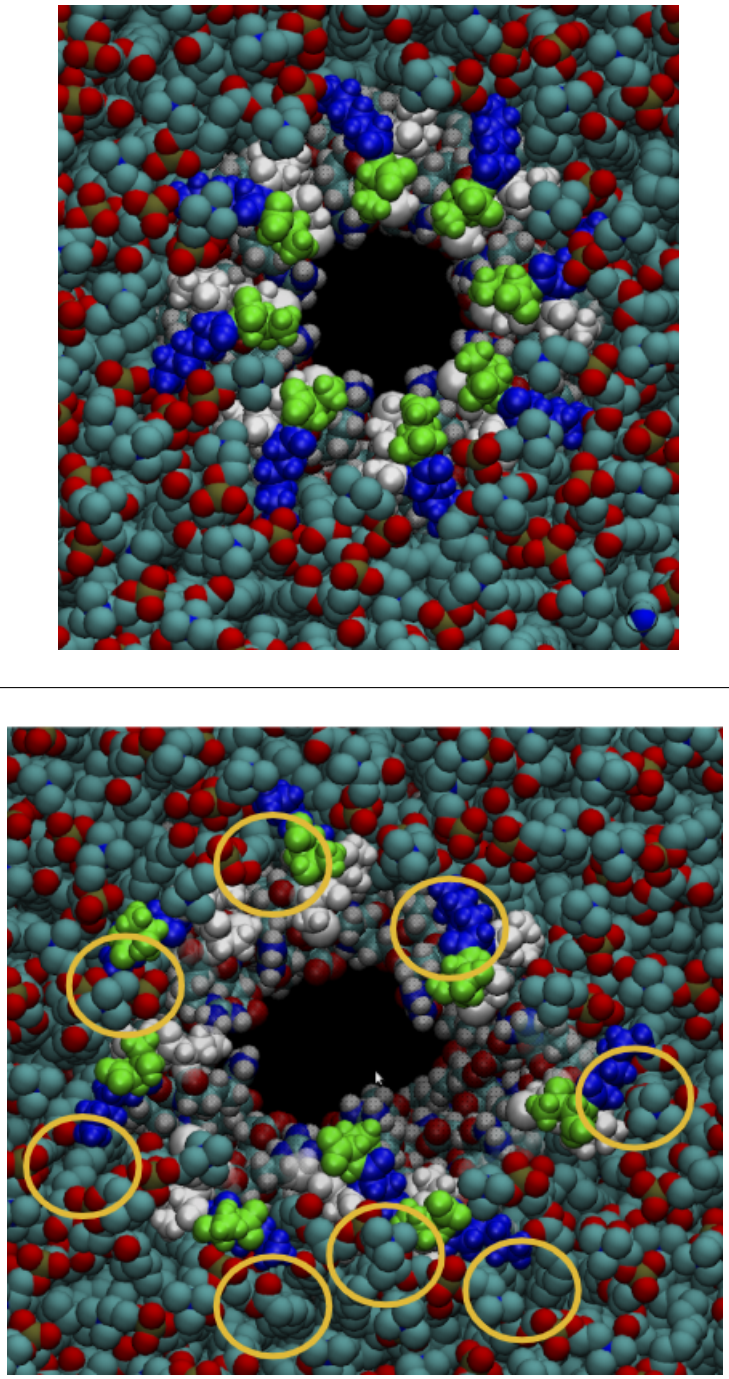


Figure B.1: Screenshots of the top-view of a cWza docked-structure embedded in lipid bilayer before (*Top*) - at the end of PRMD stage - and after (*Bottom*) 100 ns of URMD. Yellow circles show the location where salt-bridge binding between Lys-residues and lipi-head Oxygens occurred.

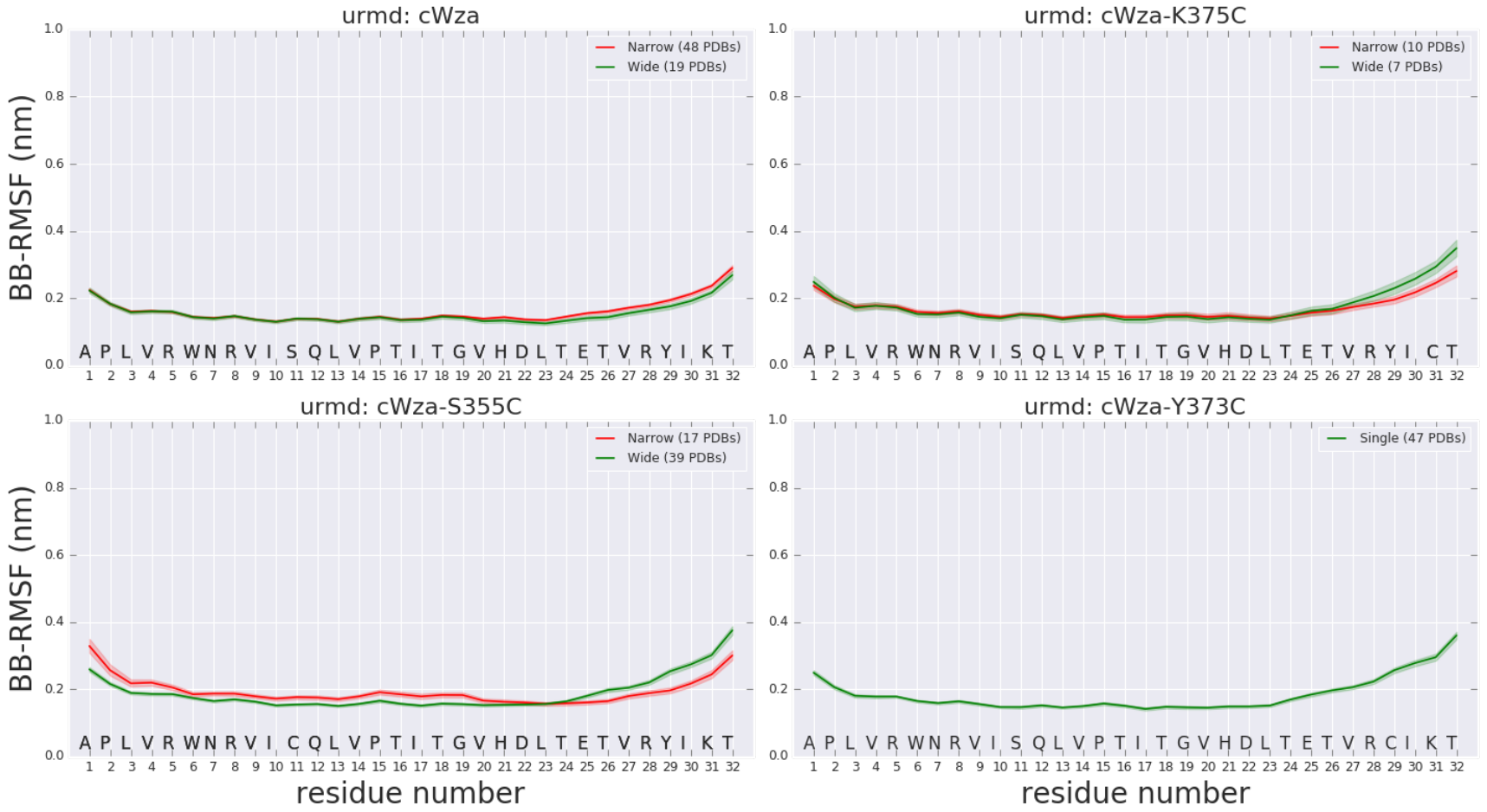


Figure B.2: Average Root-Mean-Squared-Fluctuation (*RMSF*) values were determined per residue for its backbone-atoms (BB), for the full length (100 ns) of MD trajectories of docked structures per conformational group, per mutant. Values were averaged over all peptides.

B.2 Equilibration of Simulated Structures

In MD, equilibration refers to a state reached by the simulated system for which the magnitude of observable variables does not vary in time. In particular, for simulated peptide-assemblies their structures are said to be equilibrated by the end of a simulation if metrics describing their geometry, X, do not change in time, usually indicated by the presence of a plateau in the temporal evolution of $X(t)$. In other words, when the rate-of-change of X is close enough to zero from some observable time $T_{obs} > 0$ during the simulation, that is, $dX(t)/dt \rightarrow 0$ for $T_{obs} < t < T_{sim}$. See Figure B.3, *Top* Panel.

For a single MD simulation, structural equilibration can be simply judged from visualising the evolution of X in time. However, when several simulations with different starting condition types have to be compared, even taking a small sample of trajectories to visualise X can be a time-consuming task, and possibly not representative. To deal with this challenge, we next describe a proposed methodology based on metrics extracting geometrical features from the time-series of any X observable. Thus, the task of judging the tendency towards equilibration for a set of several MD trajectories can be facilitated via automation of this analysis.

Methodology

Time-series extraction Suppose that we have a set of N different MD trajectories of the same time-length T_{sim}. Then, from each trajectory we first extract the evolution of a variable of interest, say X. Hence, we extract the time-series $X_n(t)$, with $0 < t \leq T_{sim}$, for all simulations $n = 1, 2, \ldots, N$. Then, we wonder whether most time-series $X_n(t)$ have reached equilibration by the end of their simulations ($T_{obs} < t < T_{sim}$).

Rate-of-change of time-series Once all $X_n(t)$ time-series have been extracted, then determine their forward-difference, defined by $V_n(t) = [X_n(t + \delta t) - X_n(t)]/\delta t$. Assuming that all time-series have identical time-resolution, then we can take $\delta t = 1$. Thus, the time-series can be considered discrete, and V_n will be calculated from the difference values between consecutive points X_n in time.

Smoothing of rate-of-change time-series As seen in Figure B.3 (Second, Top-to-Bottom), the time-series $V(t)$, is highly variable and with very small values, which as a result contribute to the rugged appearance of the time-series. No clear trends can be usually observed. Thus, we smoothed all time-series $V_n(t)$ via a moving average of the series with a *window* of common size $w << T_{obs}$ that slides backwards in time. As illustrated in Figure B.3 (Third, Top-to-Bottom).

$$\bar{V}_w^n(t) = \frac{\delta t}{w} \sum_{t' \in [t-w,t]} V_n(t') \quad \text{with } t \leq T_{sim} \tag{B.1}$$

The window size was set to $w = 1$ ns, for all analysed time-series. New smoothed time-series are denoted as $\bar{V}_w^n(t)$, with $n = 1,2,\ldots,N$.

Zero-crossing unitary time-series As seen in Figure B.3 (Third, Top-to-Bottom), the running-average time-series $\bar{V}_w^n(t)$ are able to display trends in the rate-of-change of X. However, although $\bar{V}_w(t)$ values can be small, that is, close to zero, these can be mostly either positive or negative, which is an indication that X has predominantly an increasing or decreasing trend, respectively. As way to visualise this more clearly, we extract the *unitary time-series* of every $\bar{V}_w^n(t)$ series, with $n = 1,2,\ldots,N$, defined at every time-point as the time-series

$$\frac{\bar{V}_w^n(t)}{|\bar{V}_w^n(t)|} \quad \text{with } t \leq T_{sim} \tag{B.2}$$

for every trajectory $n = 1,2,\ldots,N$. We assume that due to numerical variability of $\bar{V}_w^n(t)$ is always different to zero. Thus, the unitary time-series only takes +1 or -1 values, which indicate whether the smoothed series is above or below zero, respectively. See Figure B.3 (Bottom) for an illustration.

Numerical indicators of equilibrium Ultimately, for every single time-series, $X_n(t)$, we will employ two metrics as indicators of equilibration. First, we define what we call the *Effective Average Rate*, which we compute as

$$\gamma(X_n) = \frac{1}{T_{sim} - T_{obs}} \int_{T_{obs}}^{T_{sim}} \frac{\bar{V}_w^n(t)}{< max\{|\bar{V}_w^n(t)|\} >_N} dt \tag{B.3}$$

that in essence, represents the mean value of the smoothed rate-of-change $\bar{V}_w^n(t)$ within the observation time-interval $[T_{obs}, T_{sim}]$ for which we want to assess the tendency towards equilibrium. Also, this is normalised by the average over all maximum-values, $max\{|\bar{V}_w^n|\}$, of all the smoothed-series $\bar{V}_w^n(t)$ from all simulations, $n = 1,2,\ldots,N$, for the full simulation length $(0 < t < T_{sim})$.

Secondly, we define what we call the *Effective Average Unit-Rate*, which we compute as

$$\hat{\gamma}(X_n) = \frac{1}{T_{sim} - T_{obs}} \int_{T_{obs}}^{T_{sim}} \frac{\bar{V}_w^n(t)}{|\bar{V}_w^n(t)|} dt \tag{B.4}$$

that, in essence, represents the average difference between the number of times the smoothed series $\bar{V}_w^n(t)$ is above or below zero within the observation time-interval $[T_{obs}, T_{sim}]$ for which we want to assess the tendency towards equilibrium.

Thus, overall γ can gives an indication how close to zero the rate-of-change of X is, while $\hat{\gamma}$ indicates whether this is also equally oscillating around zero within the time-interval $[T_{obs}, T_{sim}]$, which for an equilibrated time-series we would expect both of these values to get as close to zero as possible, simultaneously.

Equilibrium of a group of simulated structures Finally, once all the values $\gamma(X_n)$ and $\hat{\gamma}(X_n)$ are computed for every X time-series, with $n = 1,2,\ldots,N$, we determine their respective distributions, $P(\gamma)$ and $P(\hat{\gamma})$. Both distributions are useful since from their distance and symmetry for their zero-mean values we can judge whether or not most of the N simulations tend to equilibrium. Thus, the more centred at zero and symmetric $P(\gamma)$ and $P(\hat{\gamma})$ are, the more likely it is that most simulated structures approximate towards equilibrium.

Application to simulations of docked structures

The above approach was applied to the temporal evolution of the radius-of-gyration components, $R_{g,Z}$ and $R_{g,XY}$, to assess whether simulated docked structures reached an equilibrated state by the end of MD. See details of the simulation setup in Section 3.3.2. Thus, the analysis was applied to simulations of structures from the same docked conformational-group (Section 3.3.1), for all cWza *Cys*-mutants. Figures B.4, B.5, B.6, and B.7 show the corresponding distributions of the *Effective Average Rate* (γ) and *Effective Average Unit-Rate* ($\hat{\gamma}$), per R_g-component, at the beginning (0-20 ns) and by the end (80-100 ns) of MD.

Overall, we see that there is a common trend in the distributions of all γ- and $\hat{\gamma}$-values for both R_g-components to move towards their zero-mean values (red, dotted-lines in plots) as we compare data from the beginning (*Top* panels) and the end (*Bottom* panels) of MD. Also, the development of a symmetric shape in these distributions is observed too. Hence, we conclude from these trends that most simulated docked-structures tend towards an equilibrated state within their simulated time (100 ns), as the rates of change of their R_g-components seem to be close-to-zero and equally fluctuating around zero-rate-values. Interestingly though, we note that $P(\gamma)$ and $P(\hat{\gamma})$ distributions for $R_{g,Z}$ values seem to experience the major shift towards a zero-mean when comparing beginning and end MD-data. This seems an indication that backbone units of simulated structures tend to undergo faster and broader fluctuations parallel to the bilayer plane, by contrast to distributions for $R_{g,XY}$. The latter may be also an indication that most peptide-assemblies are firmly embedded to the bilayer, and hence fluctuations perpendicular to the bilayer are slower and more restricted.

Further rigour may be needed to validate our proposed methodology in terms of the definition of stationarity, within the broad field of *Analysis of Time-Series*. Nevertheless, our methodology intended to match the intuitive judgement that most MD practitioners apply when assessing whether an observable quantity $X(t)$ in a simulation reaches equilibrium. Thus, the results employing the geometrical features extracted the analysed signals according to our methodology seem to indicate this is the case.

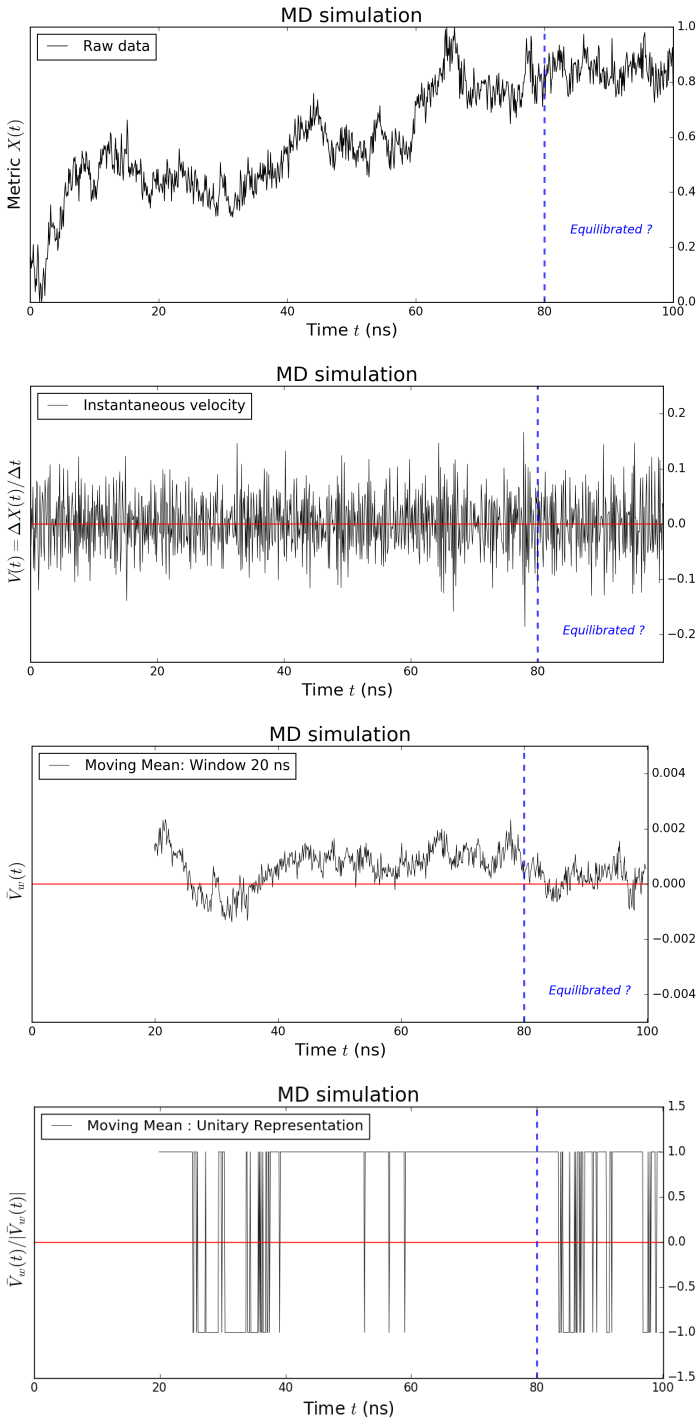


Figure B.3: Different data-sets extracted from the raw time-series (*Top*) of an observable variable X from an MD simulation. *Top-to-Bottom*: Raw data (X), rate-of-change (V), smoothed average of rate-of-change ($\bar{V}_w$), and the unitary rate-of-change ($\bar{V}_w/|\bar{V}_w|$). The blue dotted-line (- -) indicates the time-interval $[T_{obs}, T_{sim}]$ were equilibration of $X(t)$ is suspected to take place before the end-time of the simulation (T_{sim}).

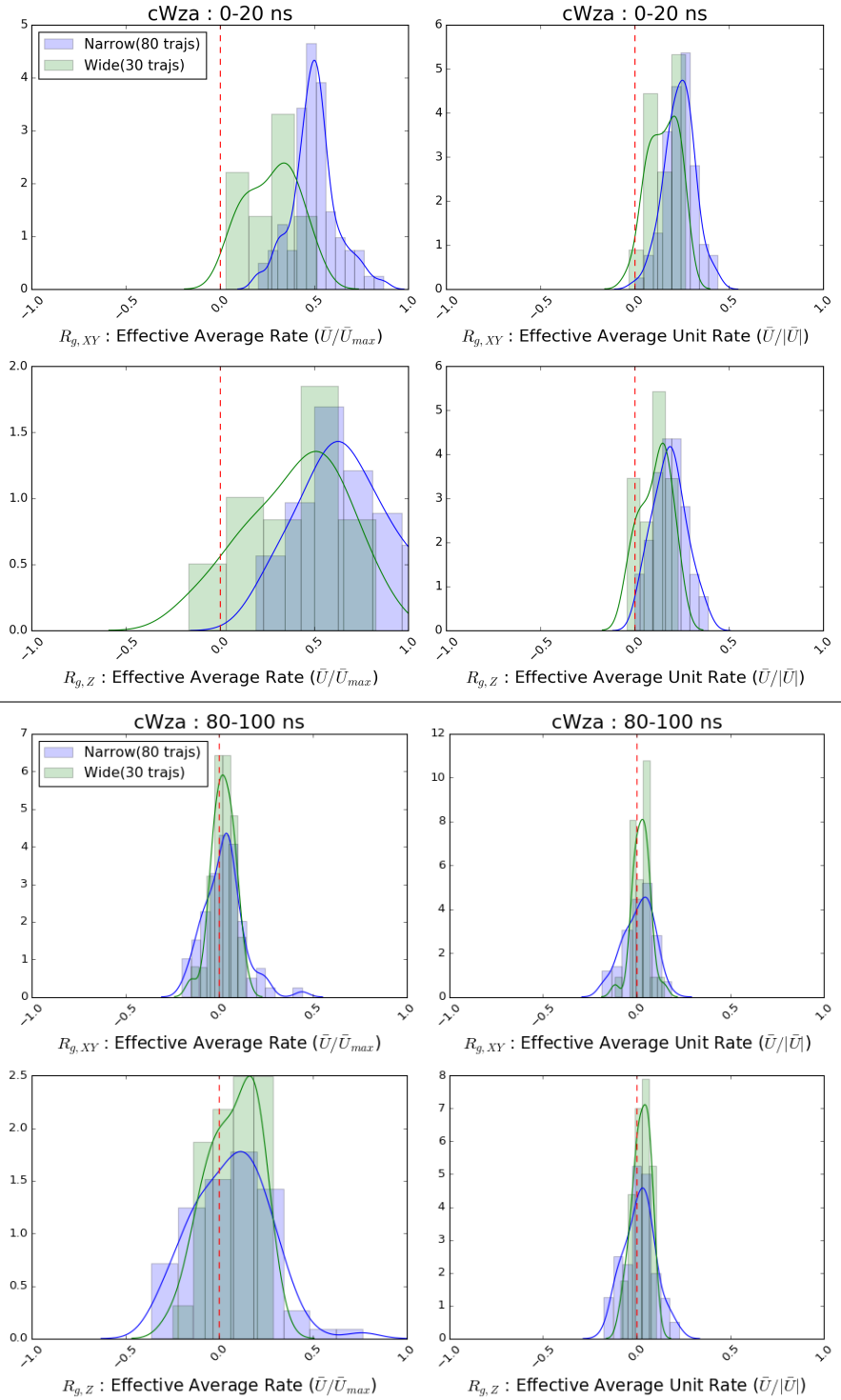


Figure B.4: **cWza**: Normalised distributions of the dimensionless *Effective Average Rate* (γ) and *Effective Average Unit-Rate* ($\hat{\gamma}$), per R_g-component, at the beginning (*Top* panel) and by the end (*Bottom* panel) of MD. Coloured distributions correspond to data extracted for simulation of docked-structures from the same conformational-group of the mutant sequence.

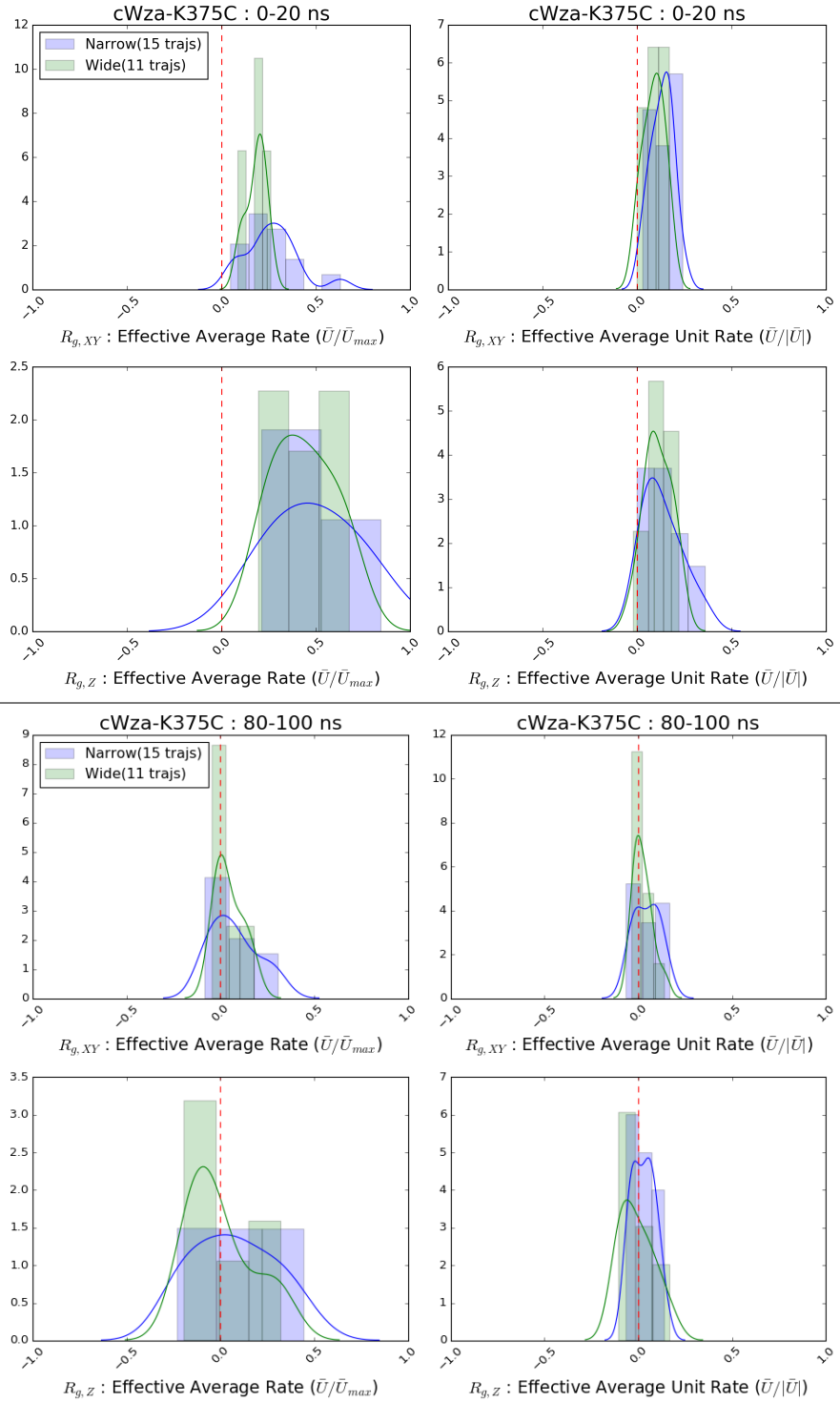


Figure B.5: **cWza-K375C**: Normalised distributions of the dimensionless *Effective Average Rate* (γ) and *Effective Average Unit-Rate* ($\hat{\gamma}$), per R_g-component, at the beginning (*Top* panel) and by the end (*Bottom* panel) of MD. Coloured distributions correspond to data extracted for simulation of docked-structures from the same conformational-group of the mutant sequence.

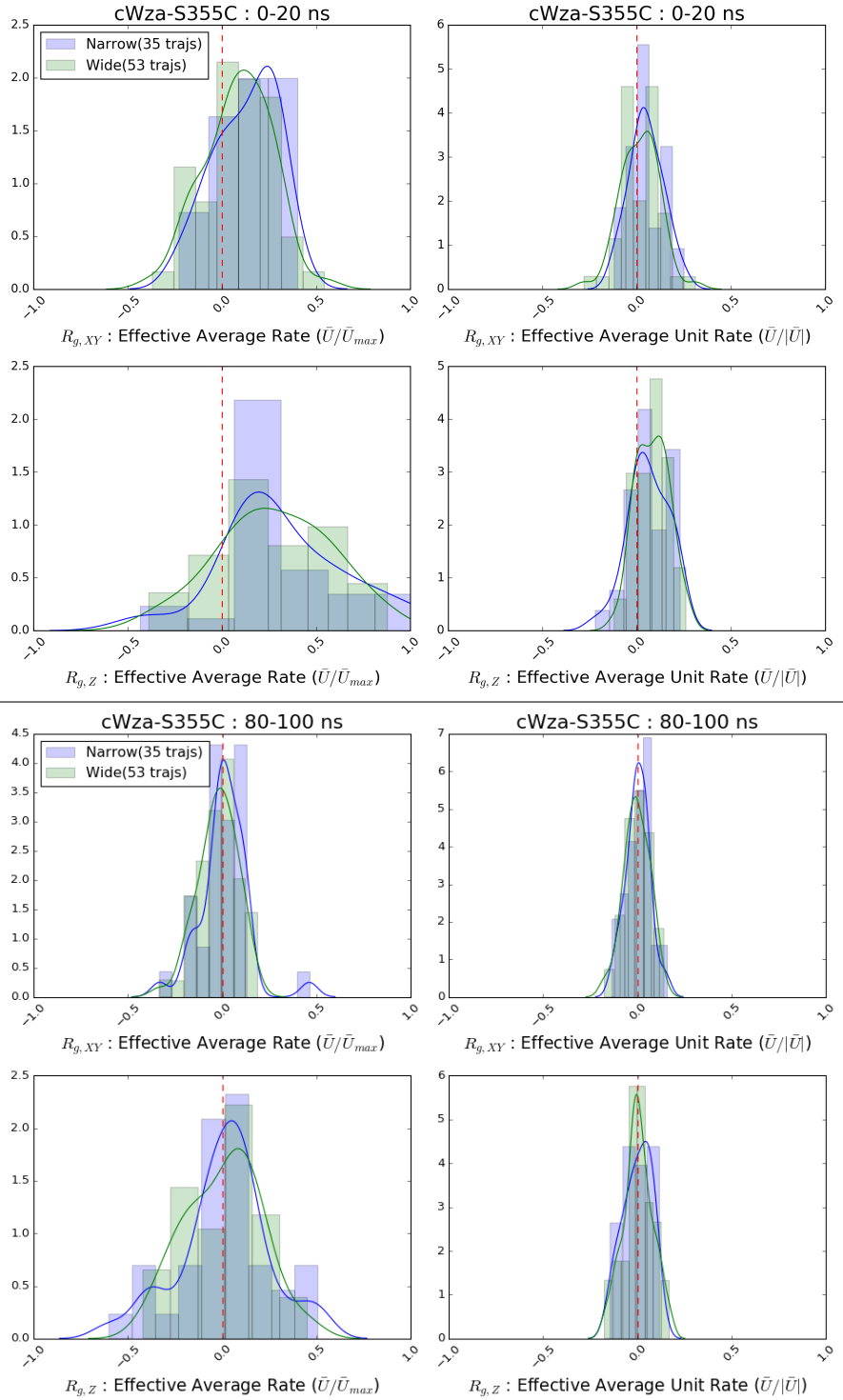


Figure B.6: **cWza-S355C**: Normalised distributions of the dimensionless *Effective Average Rate* (γ) and *Effective Average Unit-Rate* ($\hat{\gamma}$), per R_g-component, at the beginning (*Top* panel) and by the end (*Bottom* panel) of MD. Coloured distributions correspond to data extracted for simulation of docked-structures from the same conformational-group of the mutant sequence.

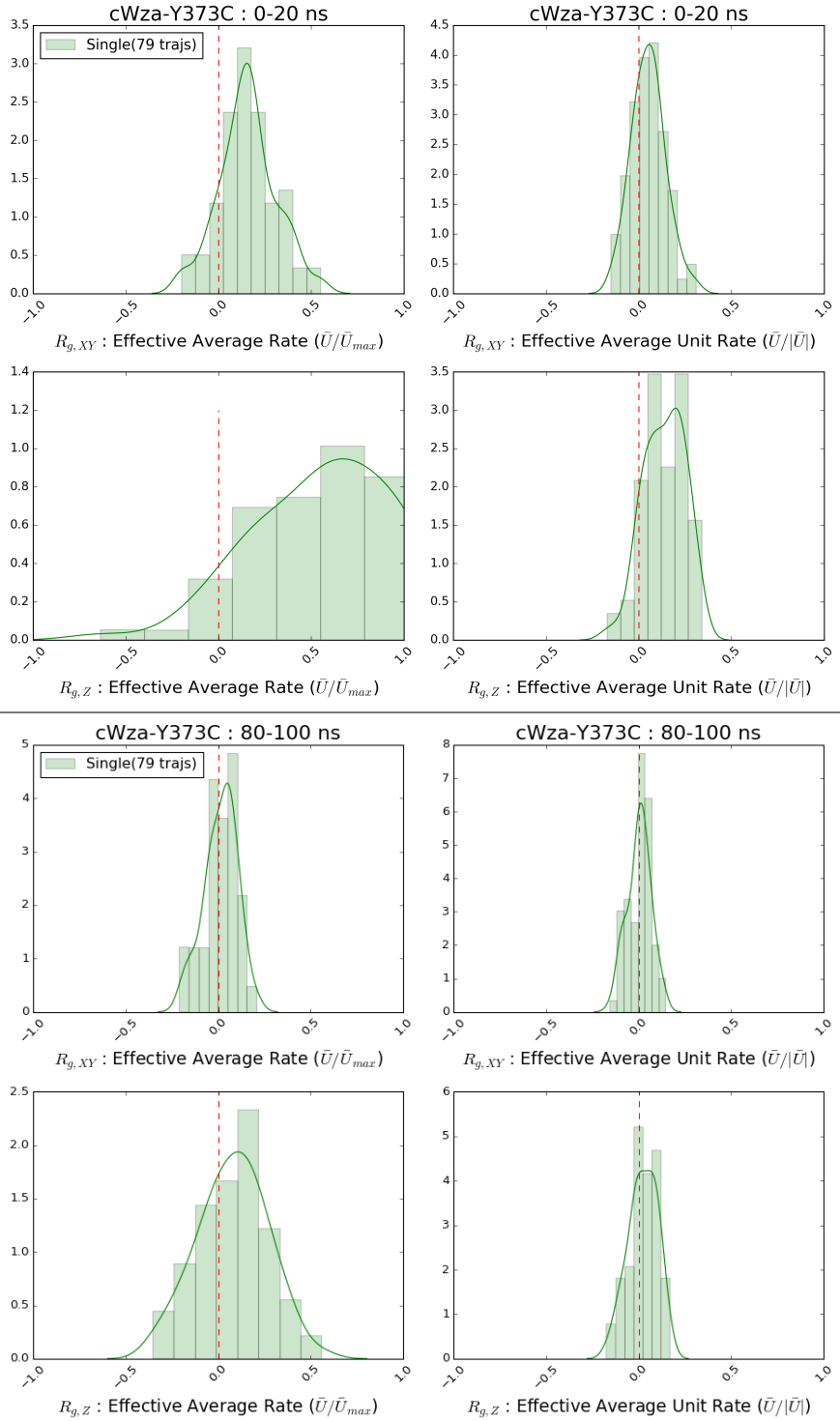


Figure B.7: **cWza-Y373C**: Normalised distributions of the dimensionless *Effective Average Rate* (γ) and *Effective Average Unit-Rate* ($\hat{\gamma}$), per R_g-component, at the beginning (*Top* panel) and by the end (*Bottom* panel) of MD. Coloured distributions correspond to data extracted for simulation of docked-structures from the same conformational-group of the mutant sequence.

B.3 Supplement of Analysis of Thermal Structural-Fluctuations in Simulations

Here, we introduce the data first described in Subsection 3.3.2, regarding the analysis of the strength of the fluctuations, S, for the R_g-components of cWza *Cys*-mutant docked structures by the end of their MD simulation.

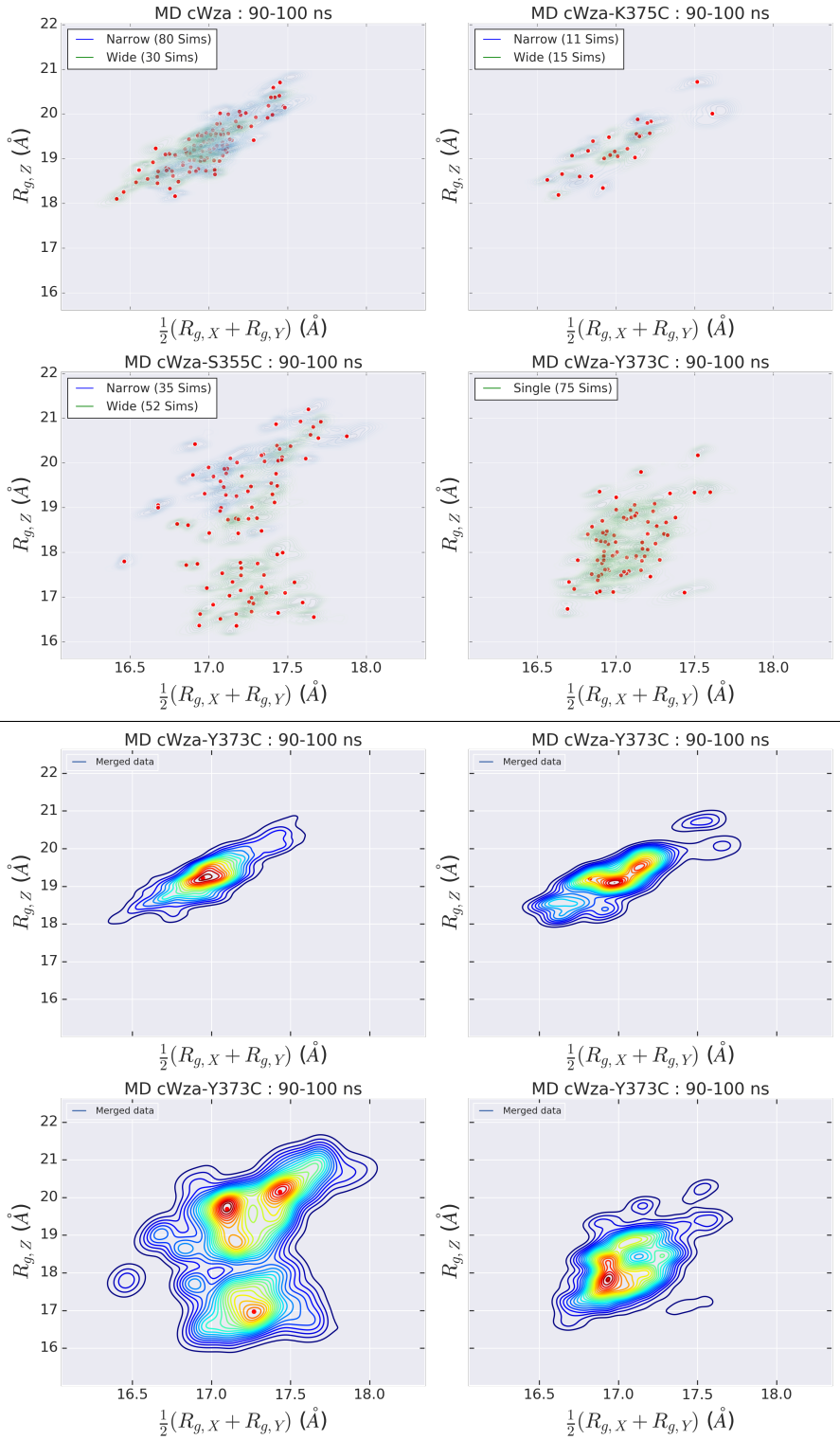


Figure B.8: *Top Panel*: Transparent contour-plots show the extent of the variability of $R_{g,XY}$ and $R_{g,Z}$ values for the last 20 ns of every individual MD simulations of docked structures. Red dots show the mean-values of each R_g-component for the same time interval. *Bottom Panel*: Contour plots showing the estimated Probability Density Functions per cWza *Cys*-mutant for the merged R_g-decomposition data from all MD simulation trajectories (80-100 ns) of docked structures and all conformational-groups.

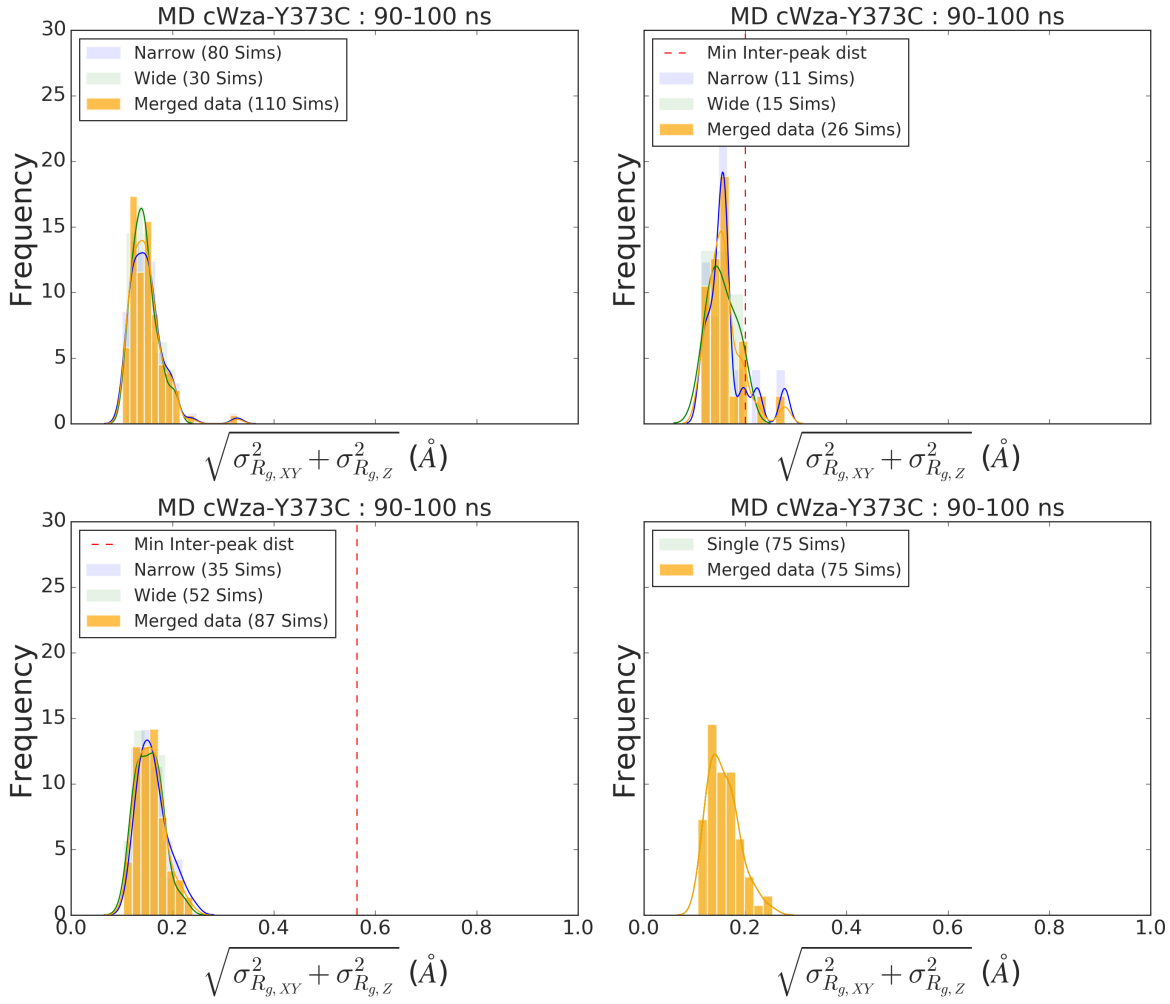


Figure B.9: Distribution of fluctuation-strength values (S) for all individual MD trajectories, 90-100 ns, of docked structures per conformational group (Blue and Green histograms) and for merged data (Yellow histograms) - regardless of conformational group. Red dotted-lines (- -) show the minimum inter-peak distance for the peaks (i.e., maxima) in the PDFs of merged R_g-decomposition data in Figure B.8, *Bottom* panels. Lines are absent for those maps where a single peak was only identified in the PDF of a mutant sequence.

APPENDIX C

EXTENDED INTERACTION ASSESSMENT DATA: H-BONDS AND KIHS

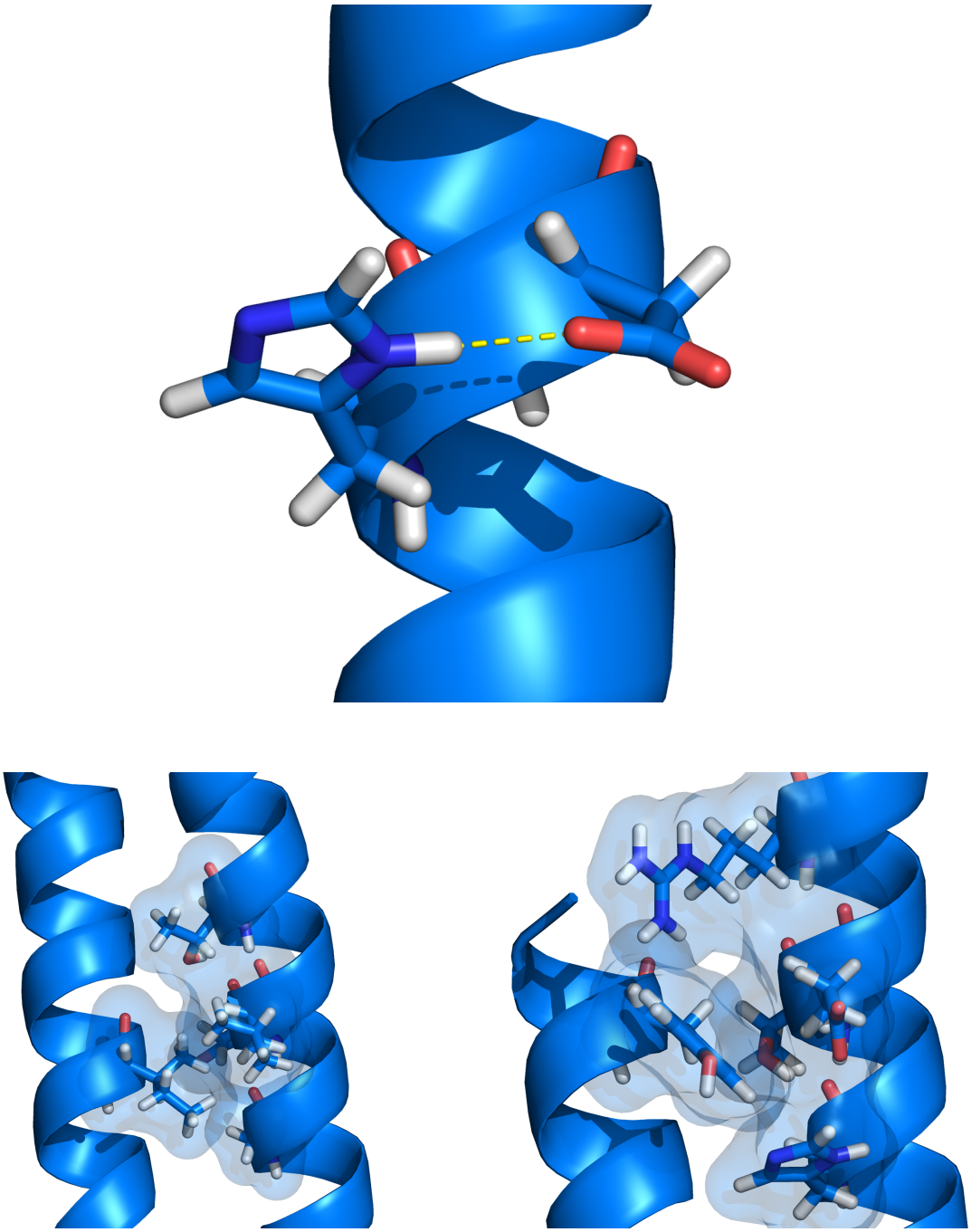


Figure C.1: Screenshots of atomic inter-chain interactions form a docked structure corresponding to cWza: C0 conformation. H-bonds and KIHs

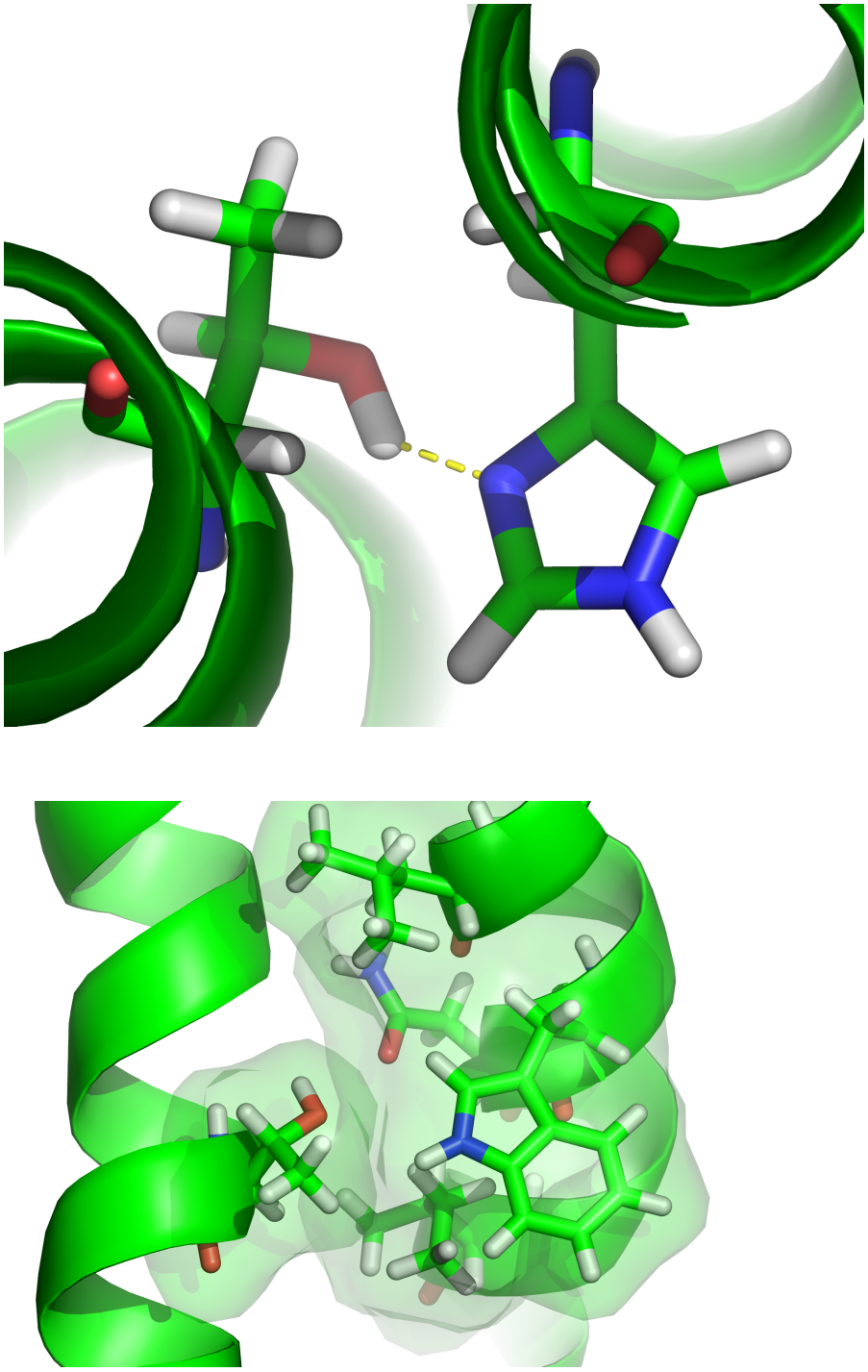


Figure C.2: Screenshots of atomic inter-chain interactions form a docked structure corresponding to cWza: C1. H-bonds and KIHs

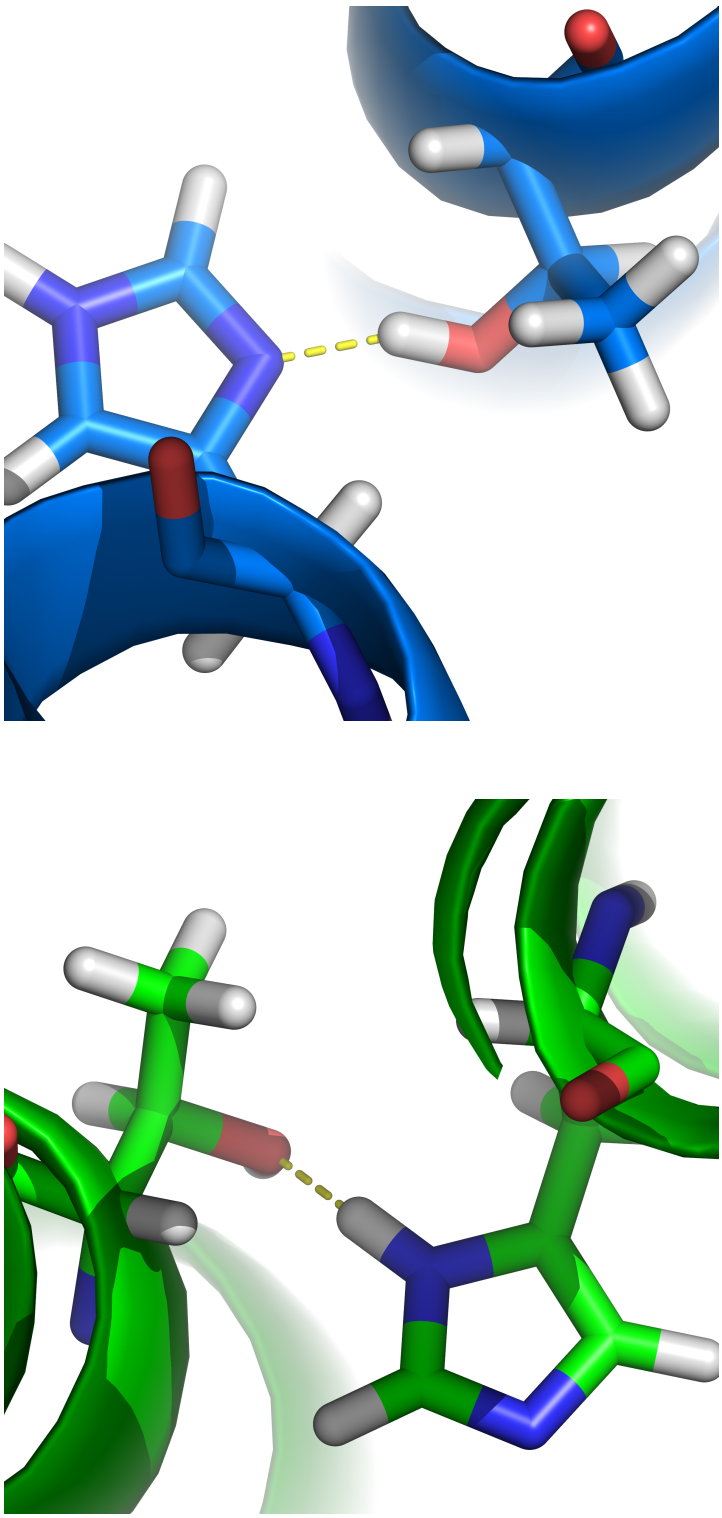


Figure C.3: Screenshots of atomic inter-chain interactions form a docked structure corresponding to cWza-K375C: C0

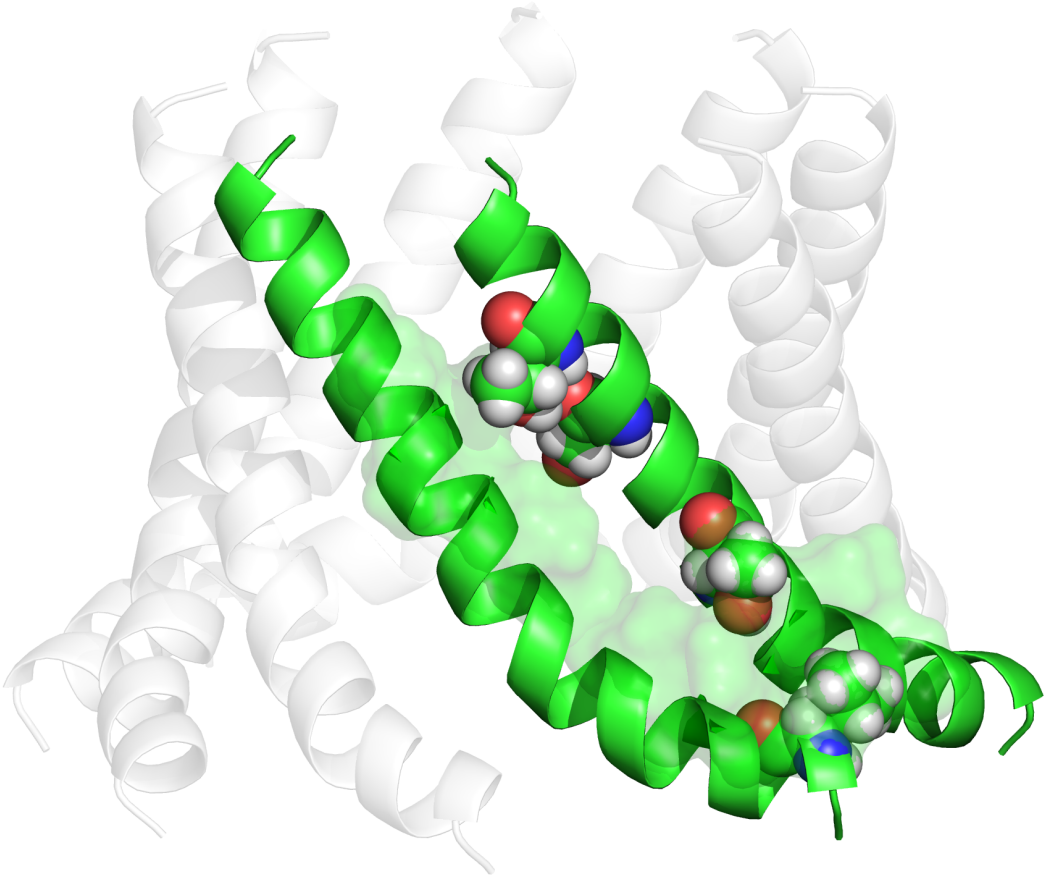


Figure C.4: Screenshots of atomic inter-chain interactions form a docked structure corresponding to cWza-K375C: C1, all KIHs

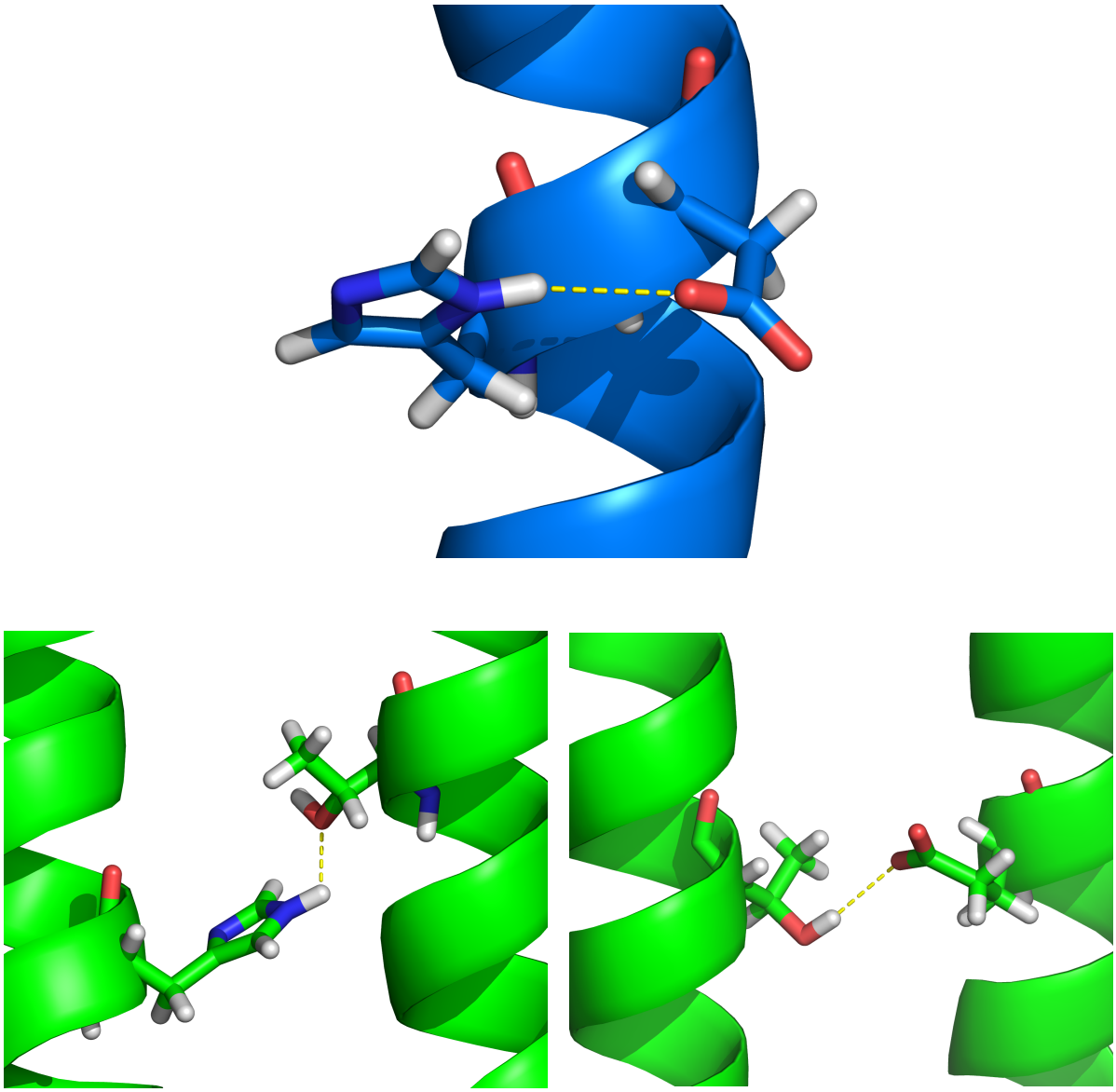


Figure C.5: Screenshots of atomic inter-chain interactions form a docked structure corresponding to cWza-S355C: C0 and C1, H-bonds.

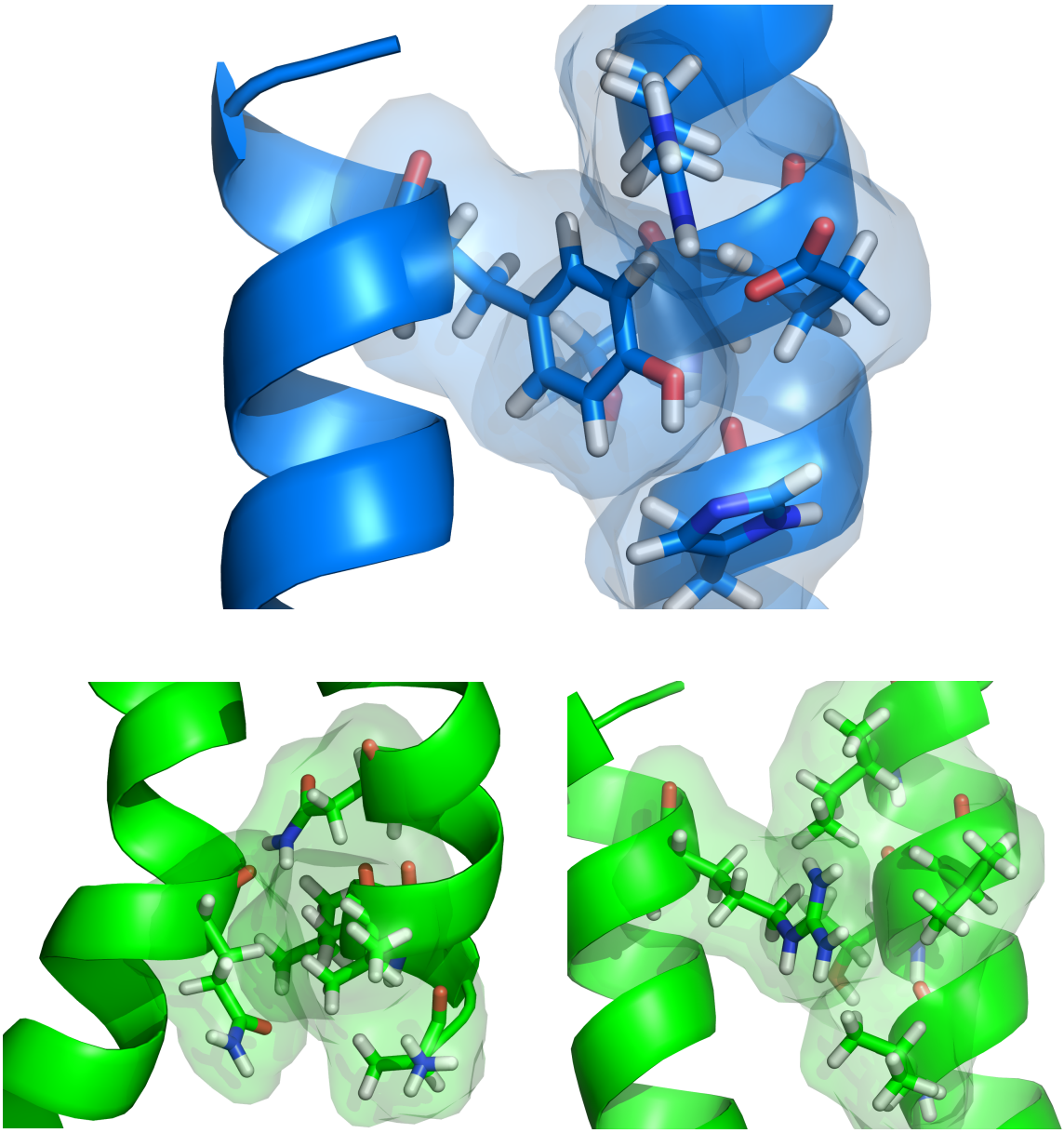


Figure C.6: Screenshots of atomic inter-chain interactions form a docked structure corresponding to cWza-S355C: C0 and C1 conformations, KIHs.

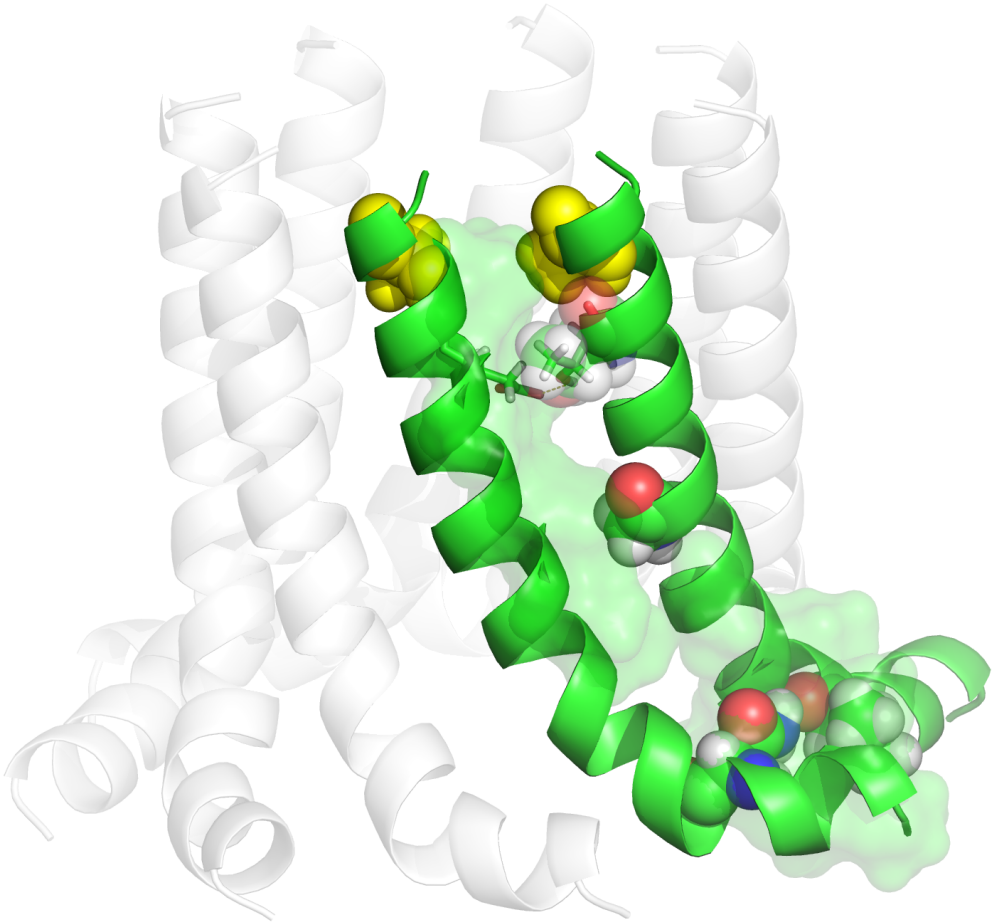


Figure C.7: Screenshots of atomic inter-chain interactions form a docked structure corresponding to cWza-Y373C. All H-bonds and all KIHs

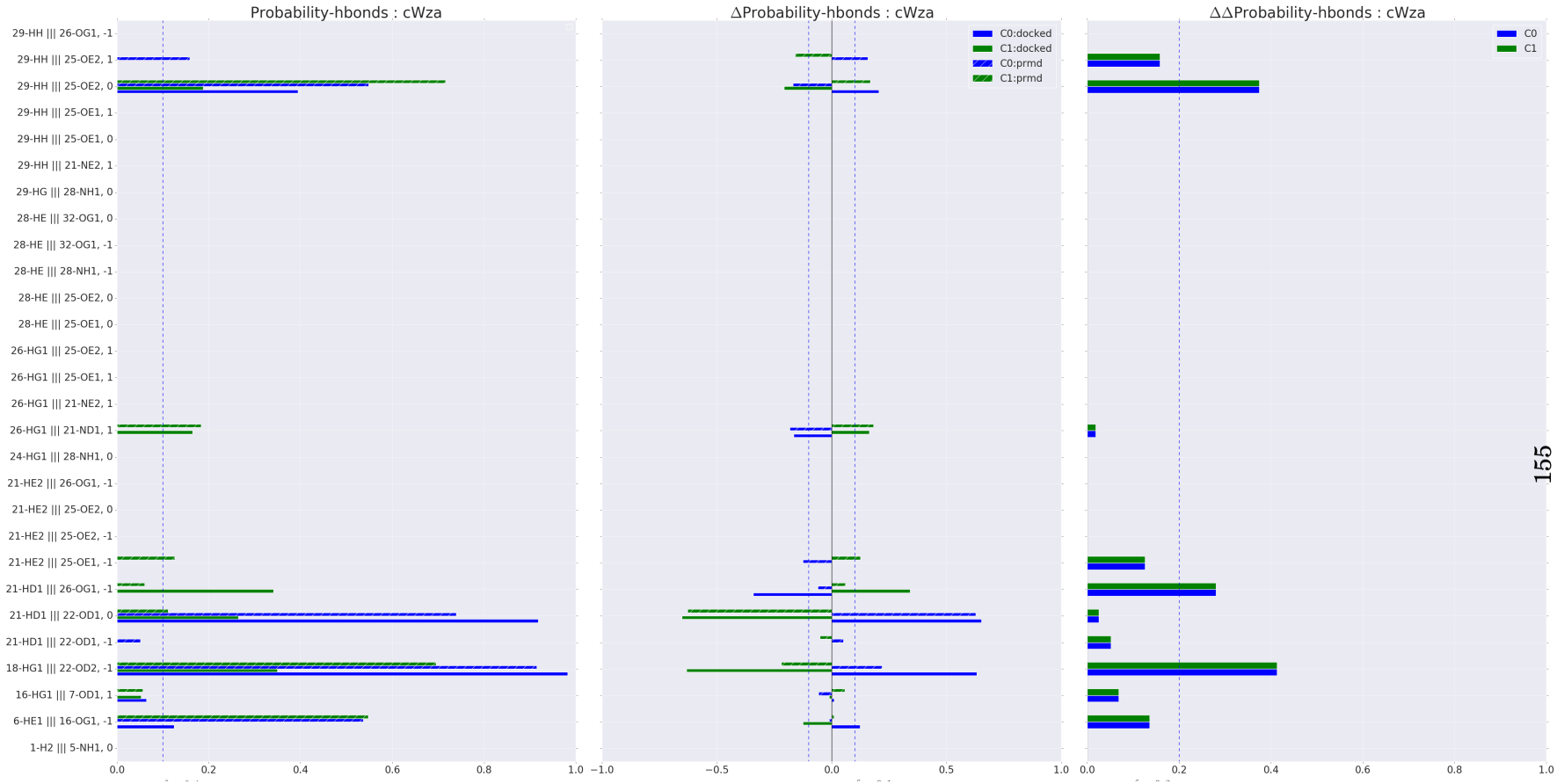


Figure C.8: *Interaction-Assessment Raw Data*: **H-bonds**: **cWza** docked channels. All interaction probabilities (P) (or *propensities*) and their finite differences (ΔP and $\Delta\Delta P$) for Narrow (**C0**) and Wide (**C1**) conformation structures were assessed, according to the methodology referenced in Subsection 3.2.4. Threshold values used: $\delta_0 = 0.1$ for P values; $\delta_1 = 0.1$ for ΔP values; and $\delta_2 = 0.2$ for $\Delta\Delta P$ values. Only structures from *Docked* and *PRMD* stages were considered.

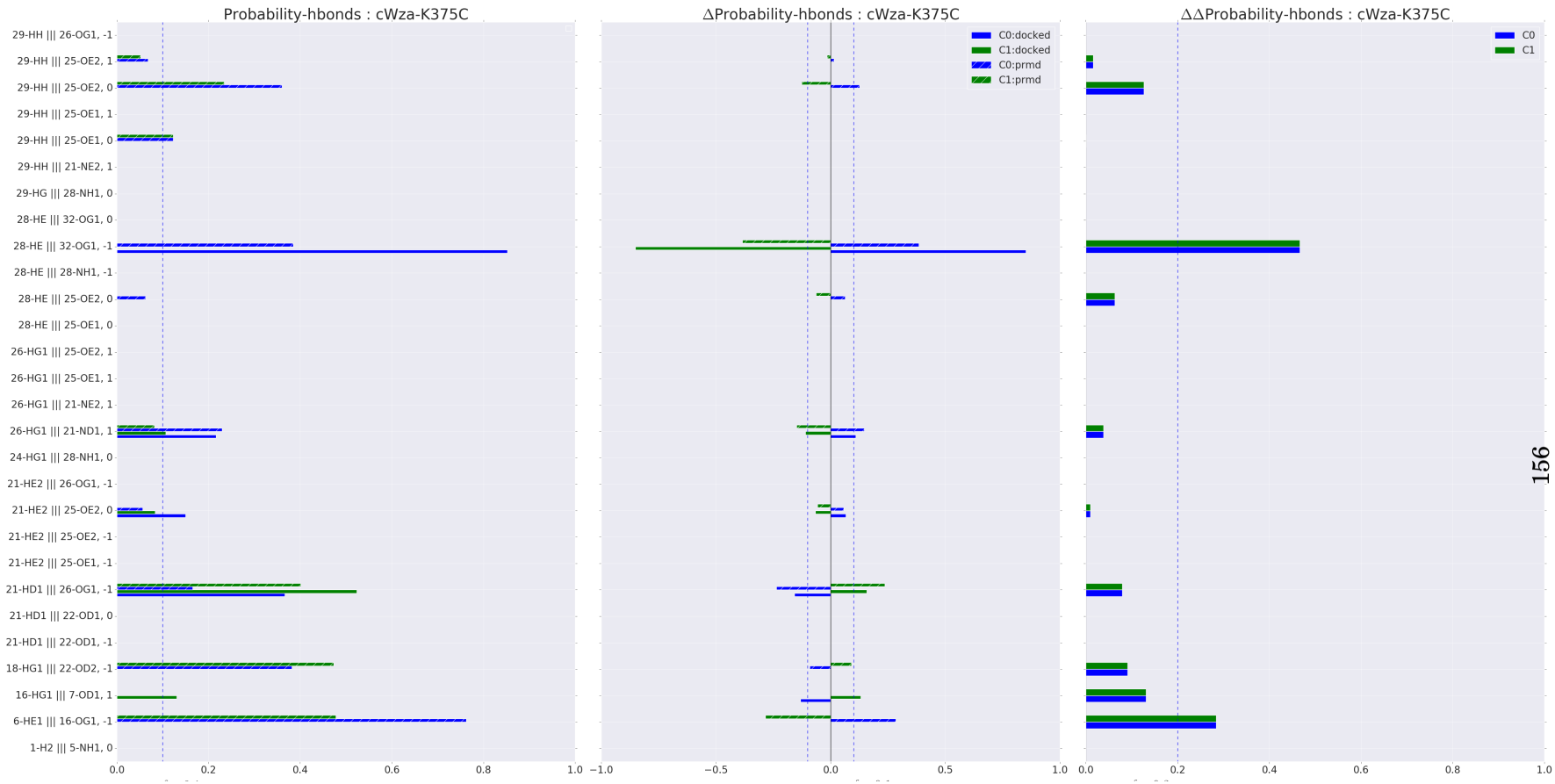


Figure C.9: *Interaction-Assessment Raw Data*: **H-bonds**: **cWza-K375C** docked channels. All interaction probabilities (P) (or *propensities*) and their finite differences (ΔP and $\Delta\Delta P$) for Narrow (**C0**) and Wide (**C1**) conformation structures were assessed, according to the methodology referenced in Subsection 3.2.4. Threshold values used: $\delta_0 = 0.1$ for P values; $\delta_1 = 0.1$ for ΔP values; and $\delta_2 = 0.2$ for $\Delta\Delta P$ values. Only structures from *Docked* and *PRMD* stages were considered.

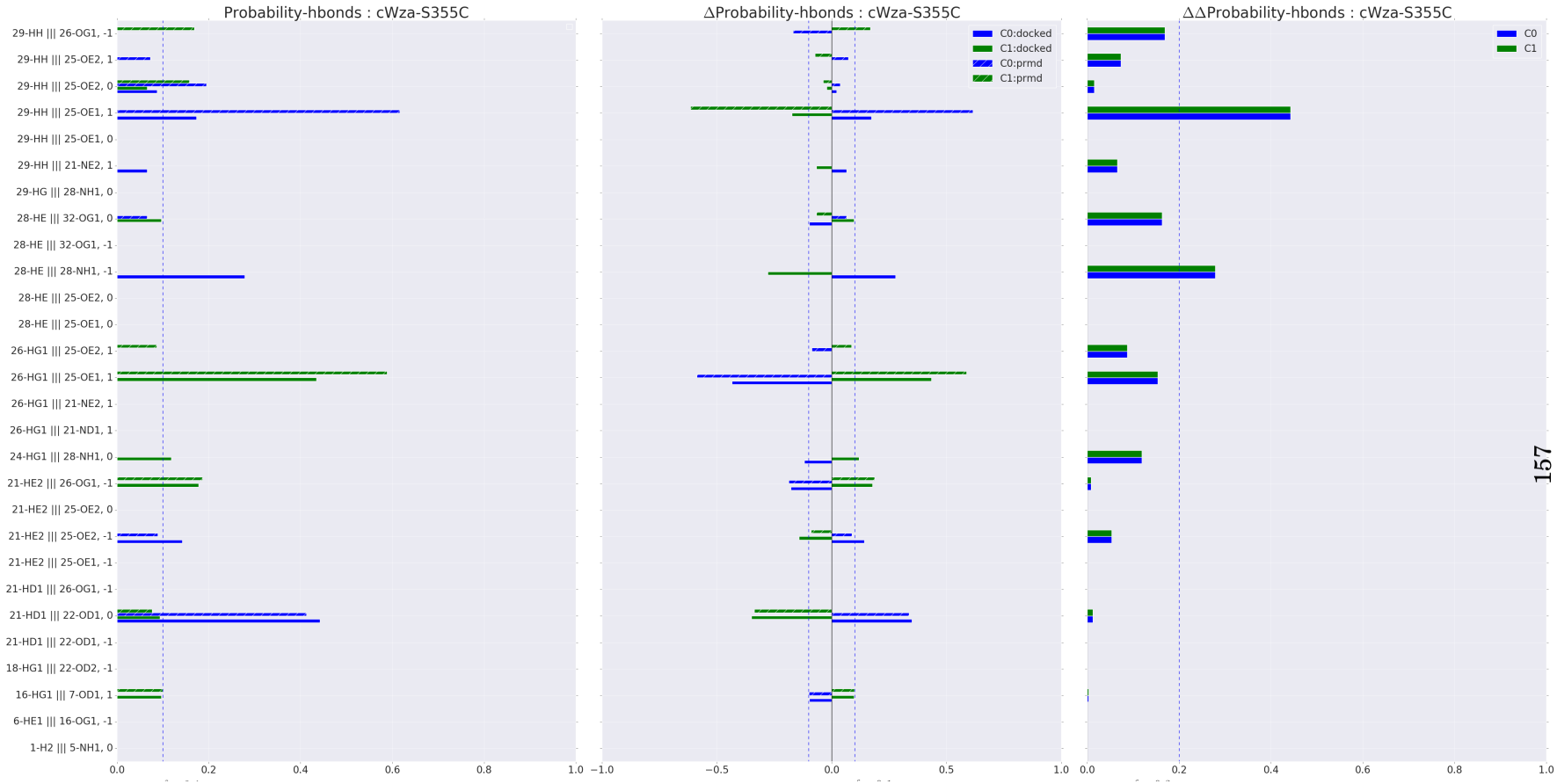


Figure C.10: *Interaction-Assessment Raw Data*: **H-bonds**: **cWza-S355C** docked channels. All interaction probabilities (P) (or *propensities*) and their finite differences (ΔP and $\Delta\Delta P$) for Narrow (**C0**) and Wide (**C1**) conformation structures were assessed, according to the methodology referenced in Subsection 3.2.4. Threshold values used: $\delta_0 = 0.1$ for P values; $\delta_1 = 0.1$ for ΔP values; and $\delta_2 = 0.2$ for $\Delta\Delta P$ values. Only structures from *Docked* and *PRMD* stages were considered.

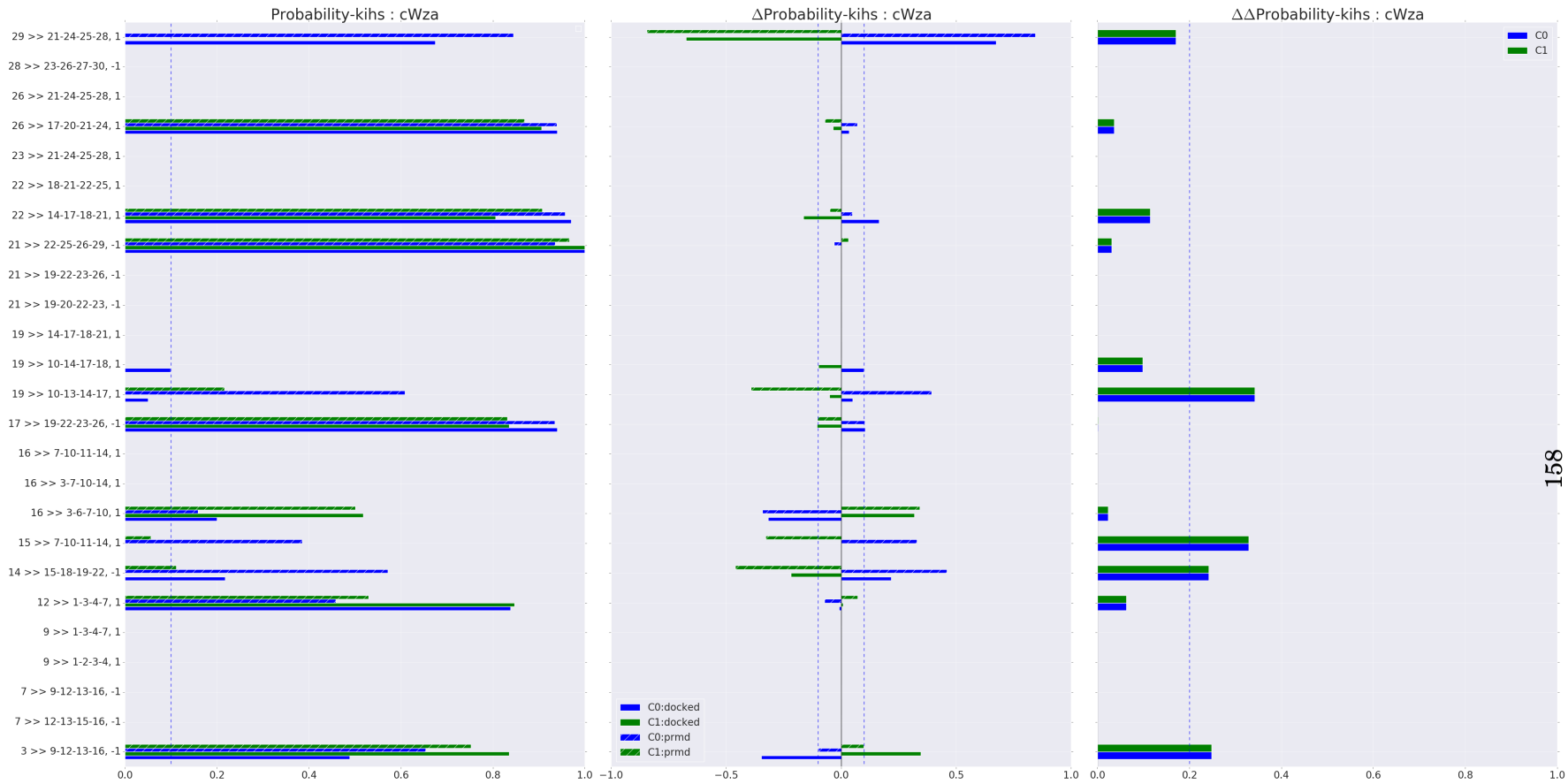


Figure C.11: *Interaction-Assessment Raw Data*: **Knobs-Into-Holes**: **cWza** docked channels. All interaction probabilities (P) (or *propensities*) and their finite differences (ΔP and $\Delta\Delta P$) for Narrow (**C0**) and Wide (**C1**) conformation structures were assessed, according to the methodology referenced in Subsection 3.2.4. Threshold values used: $\delta_0 = 0.1$ for P values; $\delta_1 = 0.1$ for ΔP values; and $\delta_2 = 0.2$ for $\Delta\Delta P$ values. Only structures from *Docked* and *PRMD* stages were considered.

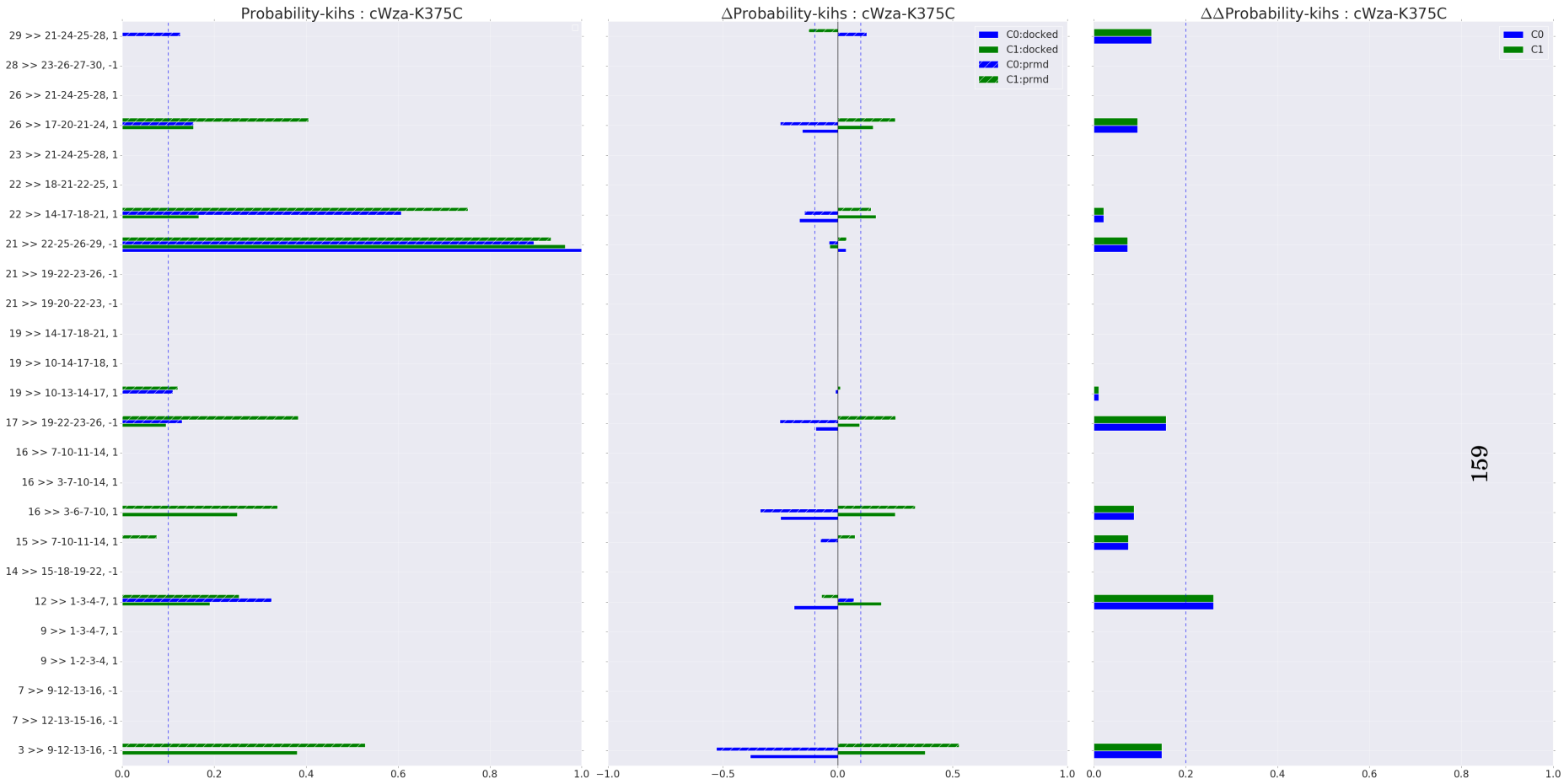


Figure C.12: *Interaction-Assessment Raw Data*: **Knobs-Into-Holes**: **cWza-K375C** docked channels. All interaction probabilities (P) (or *propensities*) and their finite differences (ΔP and $\Delta\Delta P$) for Narrow (**C0**) and Wide (**C1**) conformation structures were assessed, according to the methodology referenced in Subsection 3.2.4. Threshold values used: $\delta_0 = 0.1$ for P values; $\delta_1 = 0.1$ for ΔP values; and $\delta_2 = 0.2$ for $\Delta\Delta P$ values. Only structures from *Docked* and *PRMD* stages were considered.

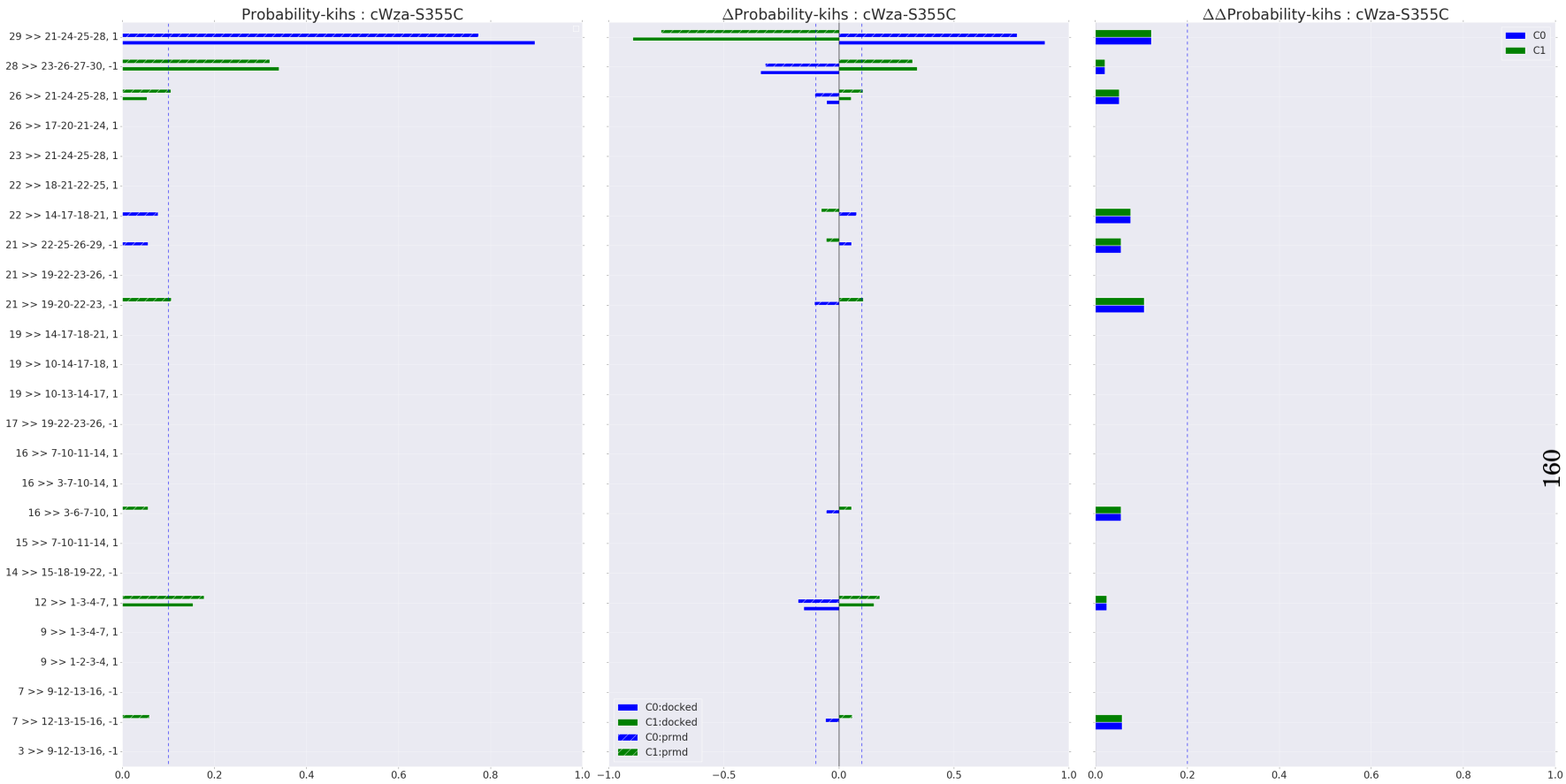


Figure C.13: *Interaction-Assessment Raw Data*: **Knobs-Into-Holes**: **cWza-S355C** docked channels. All interaction probabilities (P) (or *propensities*) and their finite differences (ΔP and $\Delta\Delta P$) for Narrow (**C0**) and Wide (**C1**) conformation structures were assessed, according to the methodology referenced in Subsection 3.2.4. Threshold values used: $\delta_0 = 0.1$ for P values; $\delta_1 = 0.1$ for ΔP values; and $\delta_2 = 0.2$ for $\Delta\Delta P$ values. Only structures from *Docked* and *PRMD* stages were considered.

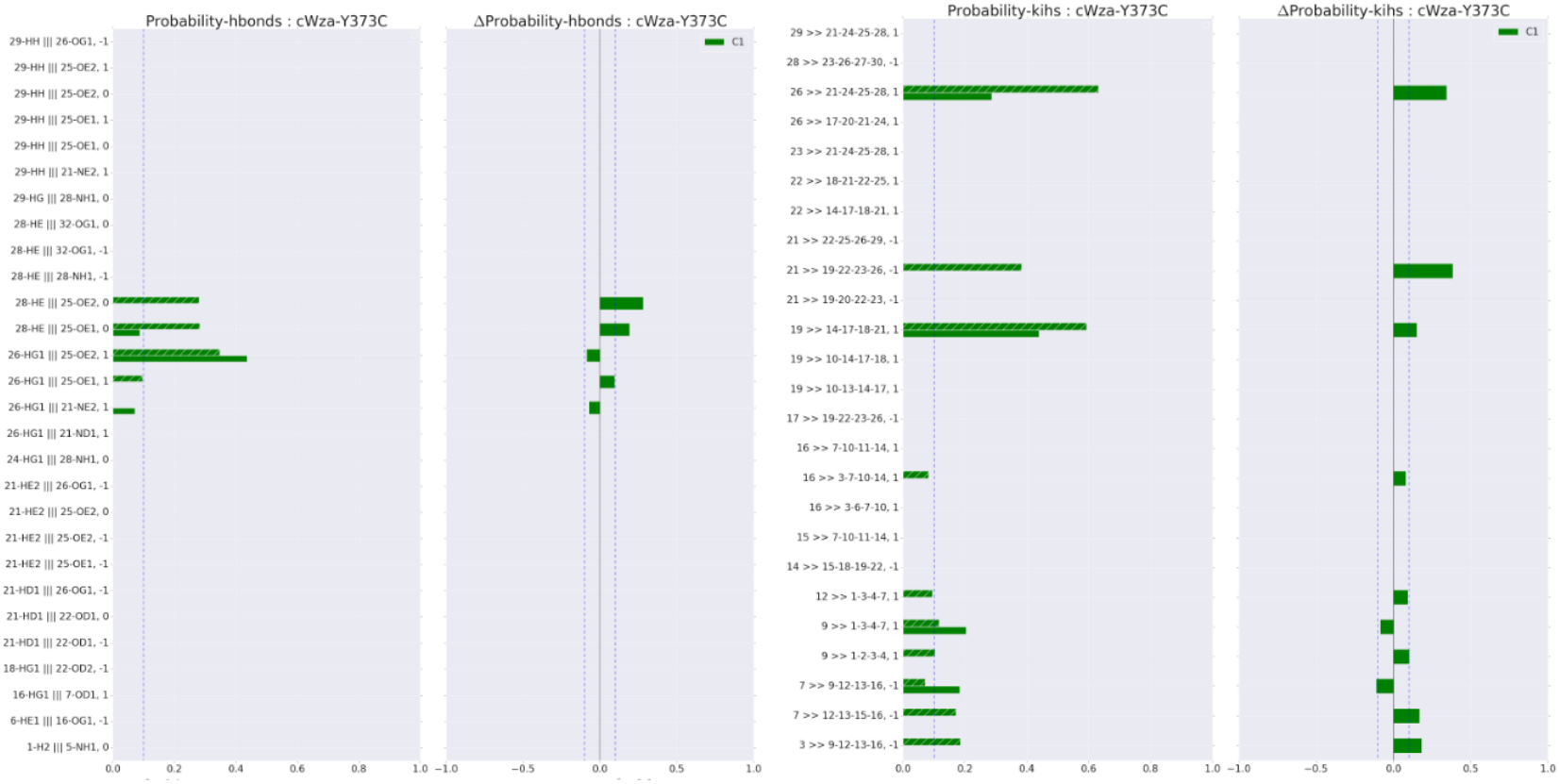


Figure C.14: *Interaction-Assessment Raw Data*: **H-bonds** (*Left*) and **Knobs-Into-Holes** (*Right*): **cWza-Y373C** docked channels. All interaction probabilities (P) (or *propensities*) and their finite differences (ΔP) for the Single (**C1**) conformation structures were assessed, according to the methodology referenced in Subsection 3.2.4. Threshold values used: $\delta_0 = 0.1$ for P values; $\delta_1 = 0.1$ for ΔP values; and $\delta_2 = 0.2$ for $\Delta\Delta P$ values. Only structures from *Docked* and *PRMD* stages were considered. **NOTE**: ΔP here just takes the difference between docked and prmd propensities, since no alternate conformation is considered for comparison, unlike other cWza *Cys*-mutant sequences.

APPENDIX D

Radius of Gyration Data Break-Down Per Docked Conformation, Per Model

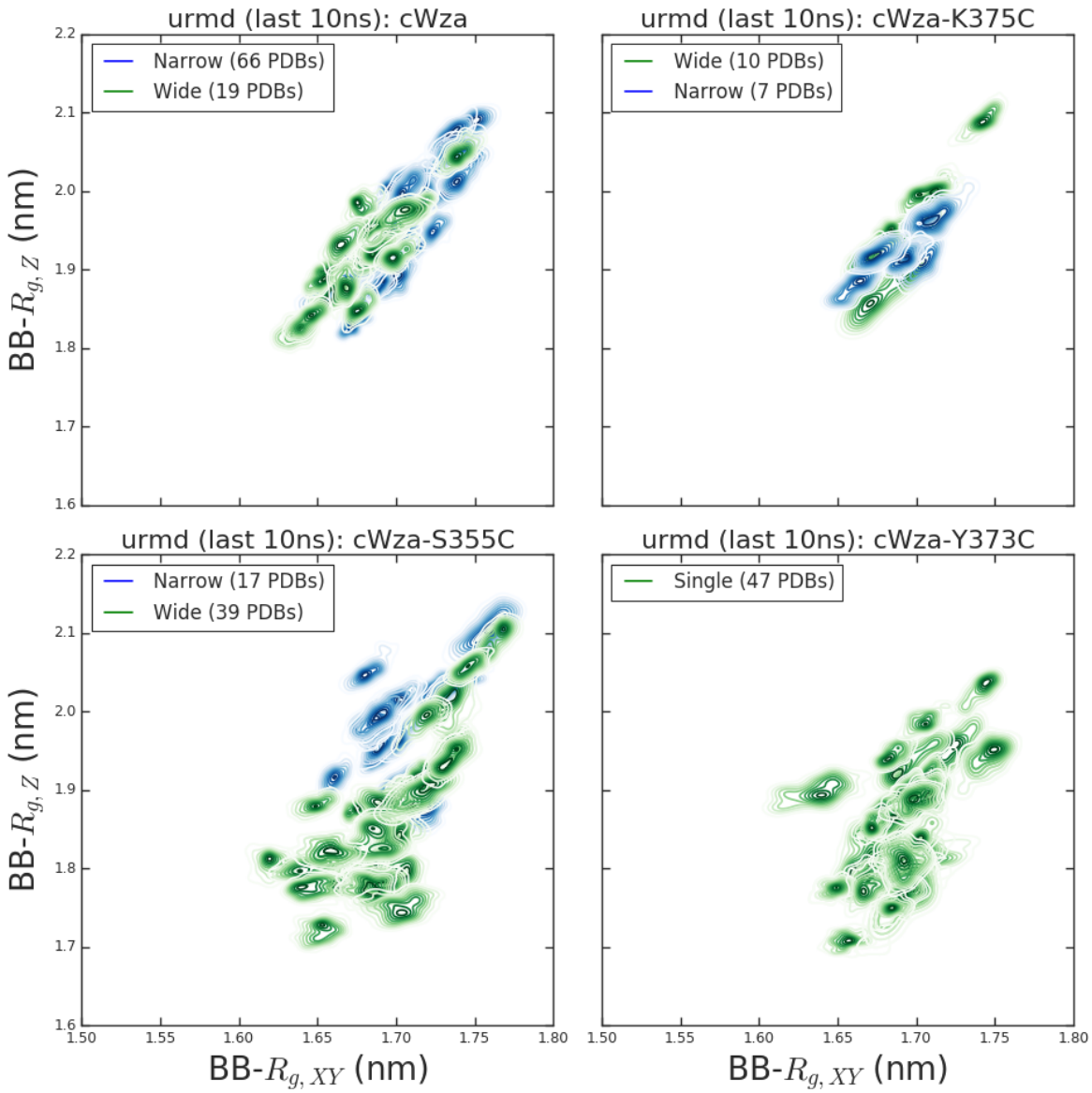


Figure D.1: Each subplot features clusters of contour plots; each of them associated to a docked channel according to a docked conformation, and cWza *Cys*-mutant peptide sequence (Subsection 3.3.1). Individual contour plots visualise the distribution of data (Gaussian *Kernel Density Estimation*) from the radius-of-gyration decomposition of an individual protein structure from all frames of the last 10 ns (out of 100 ns) of its *Un-Restrained Molecular Dynamics* (URMD) trajectory. Thus, contour plots in Figure 3.8 (Subsection 3.3.2) are the overall contour plots obtained from gathering all data-points of these contour plots, per docked conformation, per sequence.

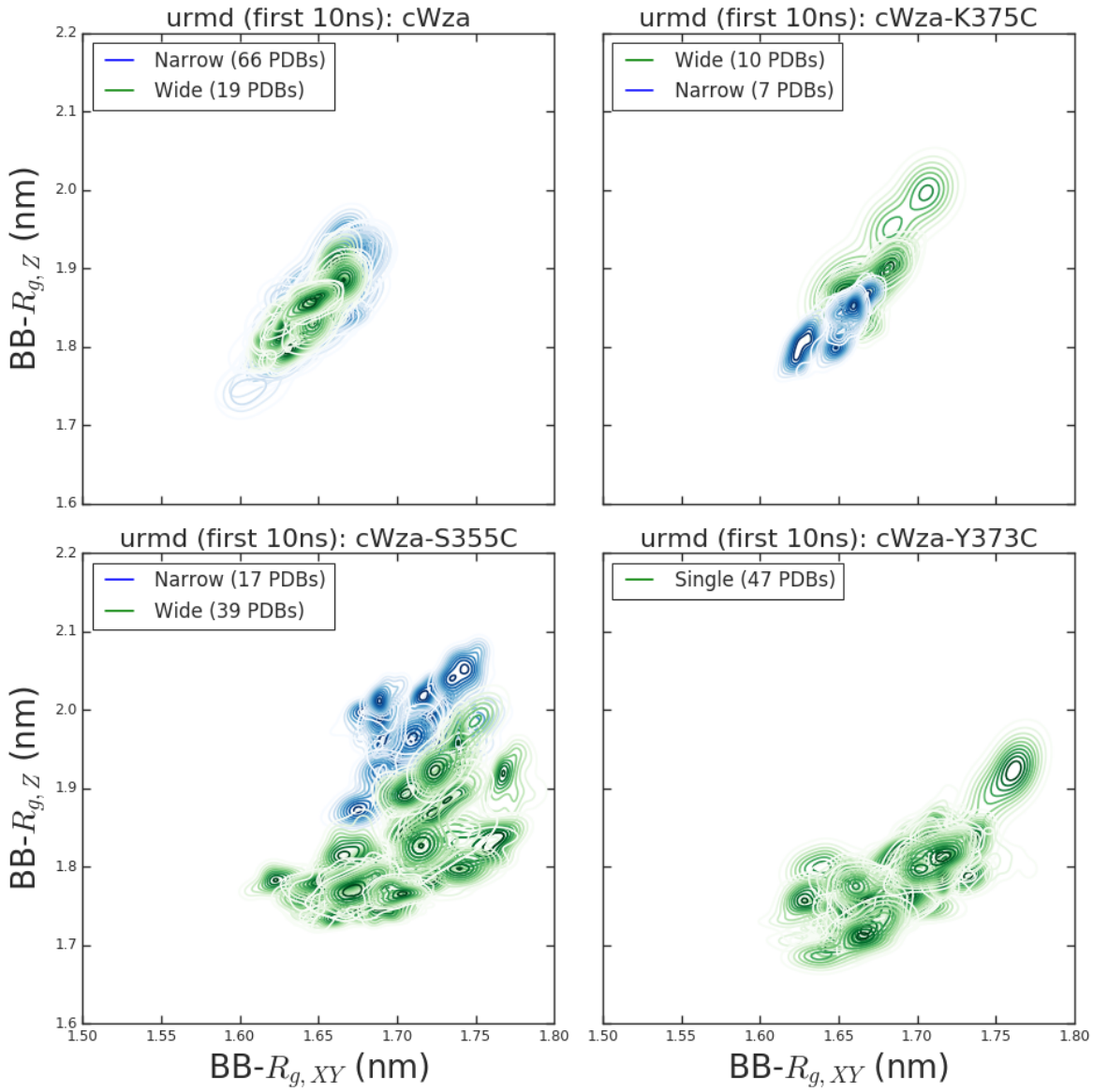


Figure D.2: Same meaning as in Figure D.1, but for the first 10 ns (out of 100 ns) of URMD trajectories.

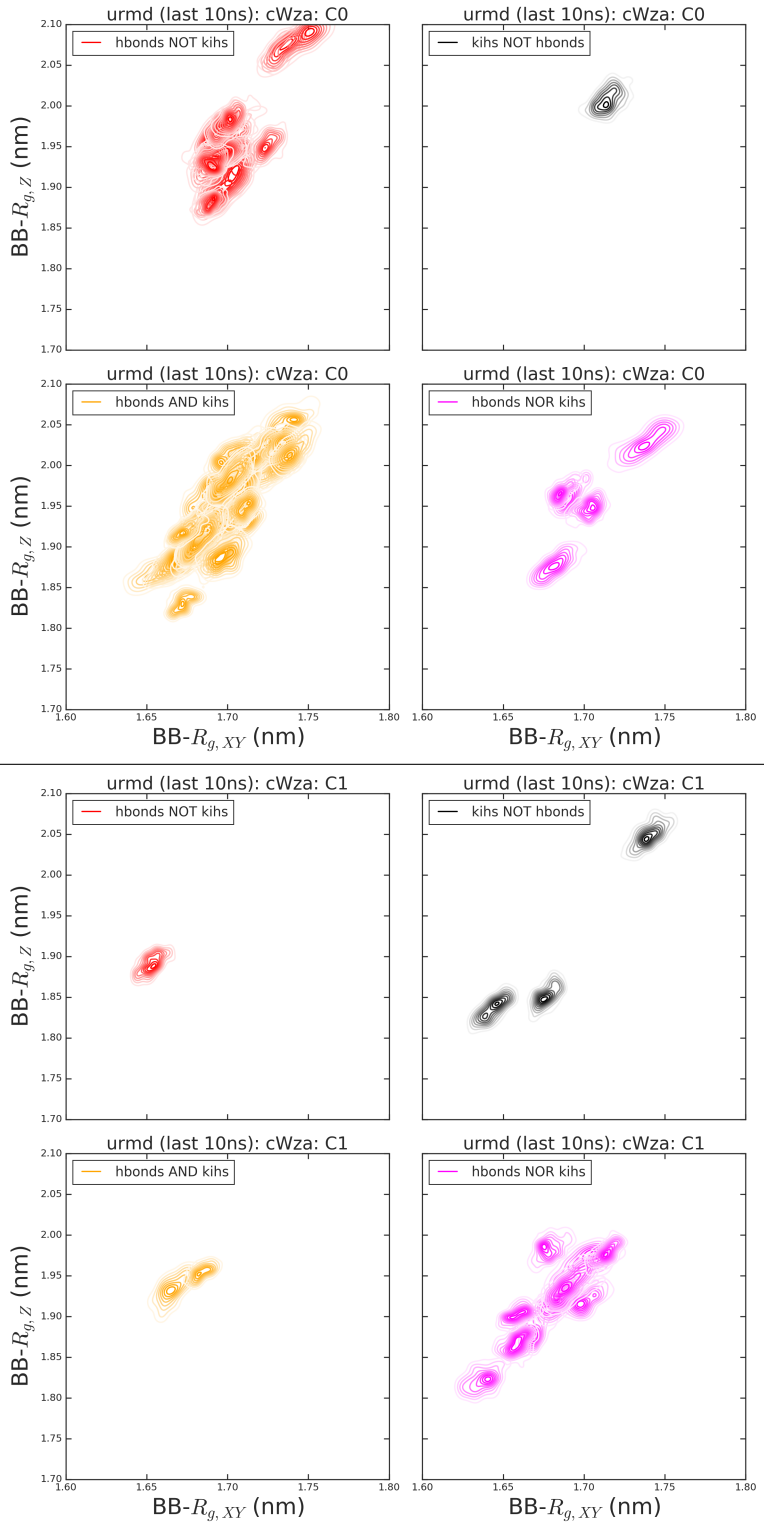


Figure D.3: **cWza** channels: Breakdown of contour-plot clusters in Figure D.1, according to identified *Interaction Partition* groups per docked conformation (Narrow/Wide) of models with selected peptide sequence. See Figure 3.10. **Top Panel**, Narrow docked-conformation structures (**C0**). **Bottom**, Wide docked-conformation structures (**C1**). **Note**: Contour plots in Figure 3.12 are obtained from gathering all data-points of contour plots here, per docked interaction partition, per docked conformation, per sequence.

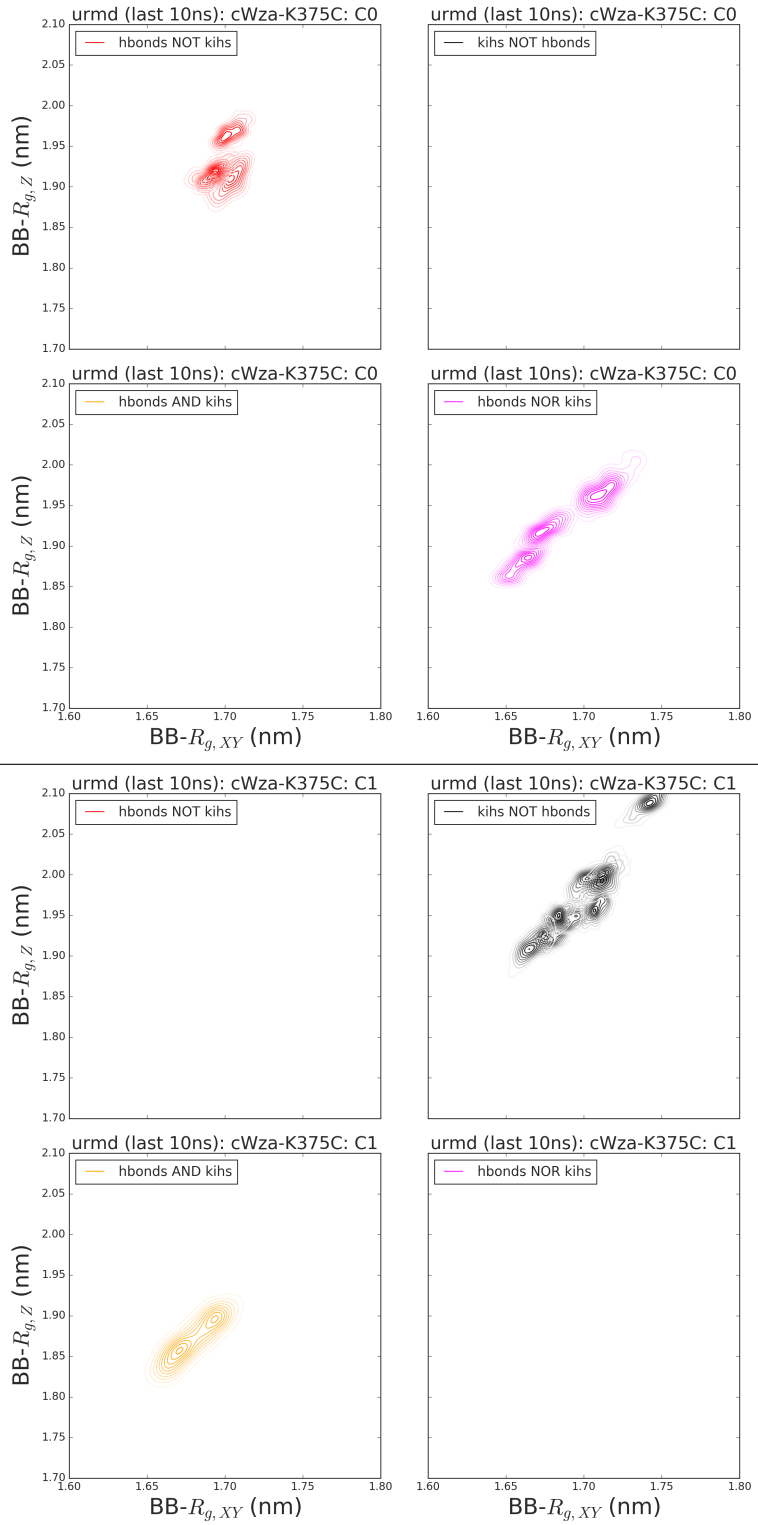


Figure D.4: **cWza-K375C** channels: Breakdown of contour-plot clusters in Figure D.1, according to identified *Interaction Partition* groups per docked conformation (Narrow/Wide) of models with selected peptide sequence. See Figure 3.10. **Top Panel**, Narrow docked-conformation structures (**C0**). **Bottom**, Wide docked-conformation structures (**C1**). **Note**: Contour plots in Figure 3.12 are obtained from gathering all data-points of contour plots here, per docked interaction partition, per docked conformation, per sequence.

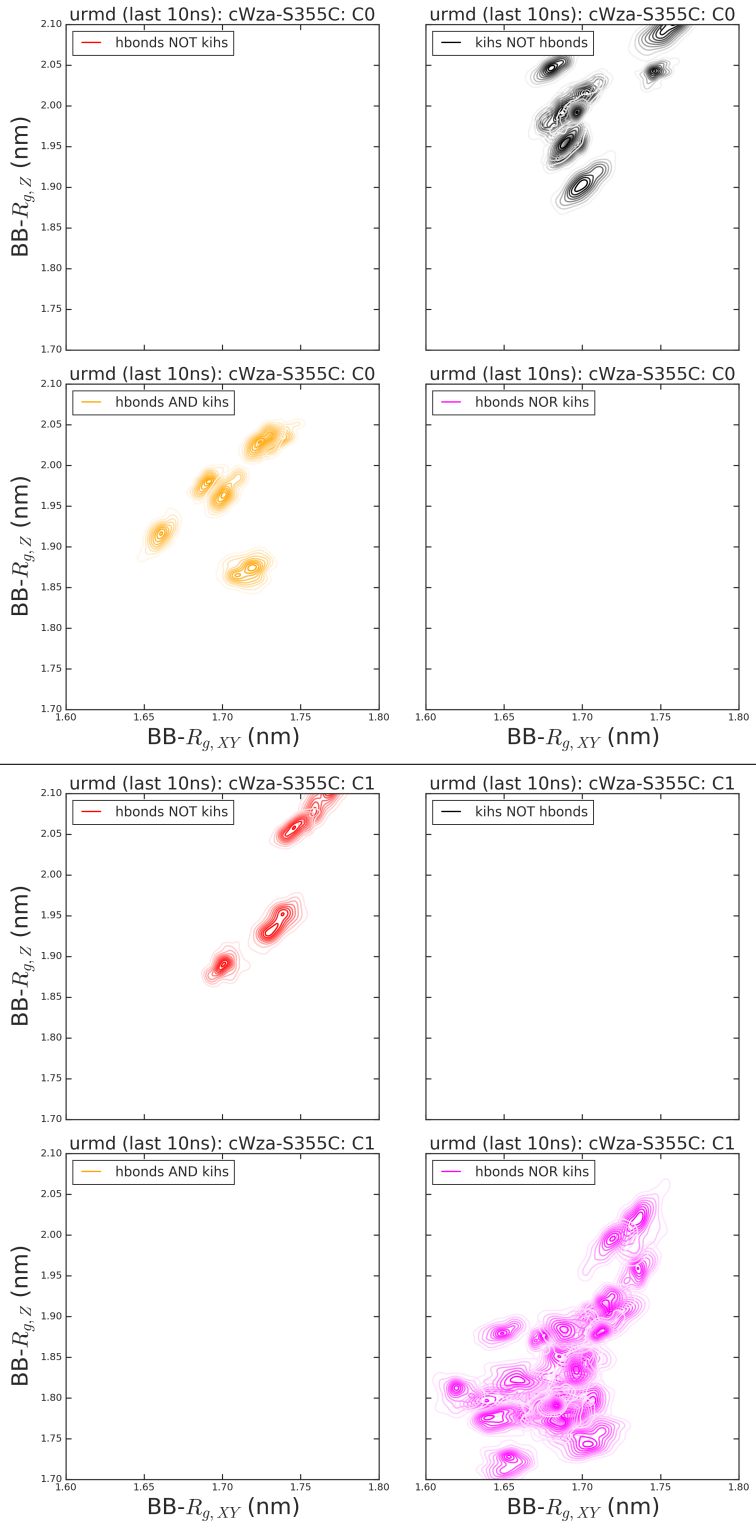


Figure D.5: **cWza-S355C** channels: Breakdown of contour-plot clusters in Figure D.1, according to identified *Interaction Partition* groups per docked conformation (Narrow/Wide) of models with selected peptide sequence. See Figure 3.10. **Top Panel**, Narrow docked-conformation structures (**C0**). **Bottom**, Wide docked-conformation structures (**C1**). **Note**: Contour plots in Figure 3.12 are obtained from gathering all data-points of contour plots here, per docked interaction partition, per docked conformation, per sequence.

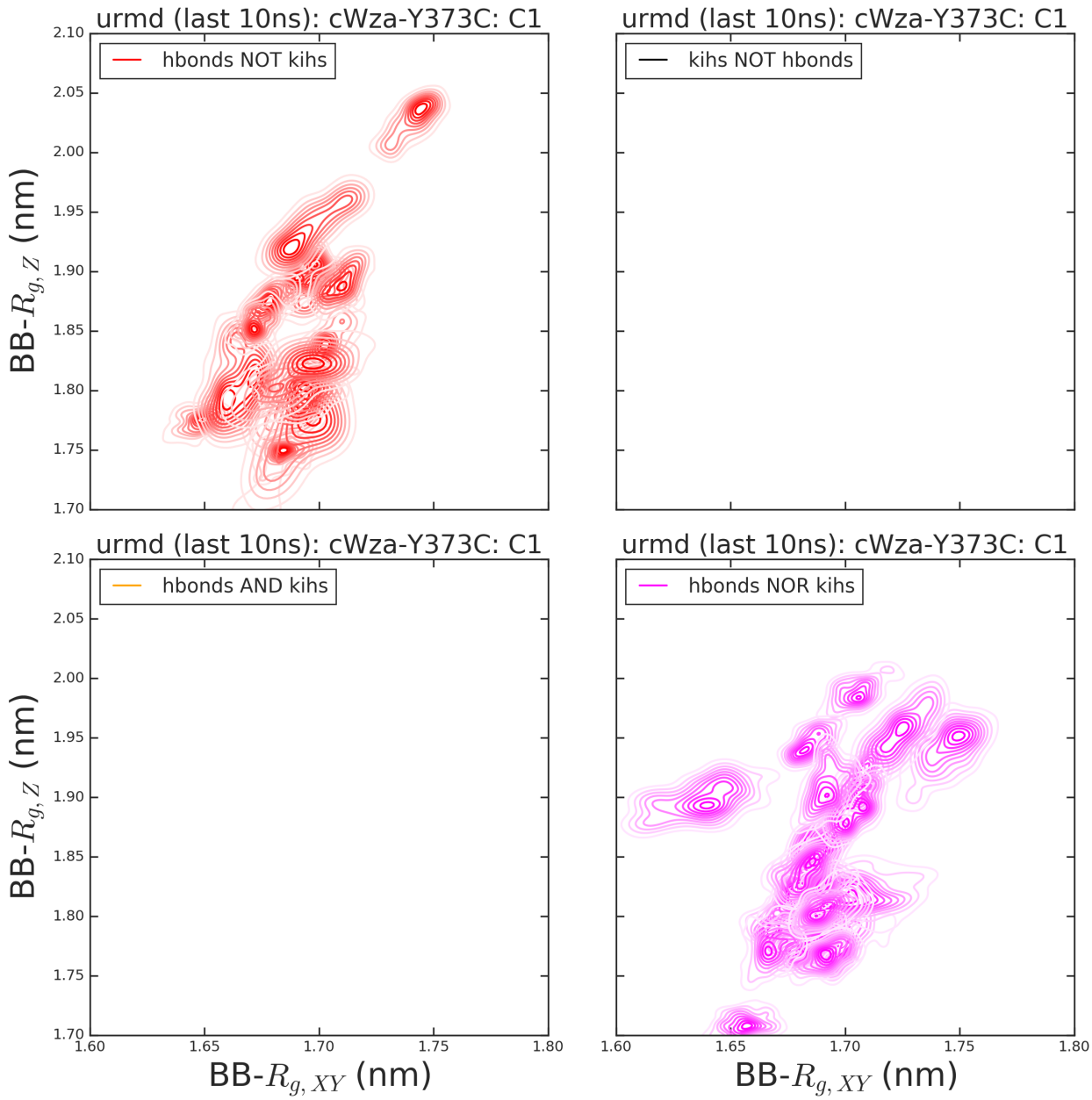


Figure D.6: **cWza-Y373C** channels: Breakdown of contour-plot cluster in Figure D.1, according to identified *Interaction Partition* groups per *Single* (**C1**) docked conformation. See Figure 3.10. **Note**: Contour plots in Figure 3.12 are obtained from gathering all data-points of contour plots here, per docked interaction partition, per docked conformation, per sequence.

APPENDIX E

COMPLEMENTARY HOLE ESTIMATED CONDUCTANCE AND PORE DIMENSIONS: CWZA-K375C AND CWZA-S355C

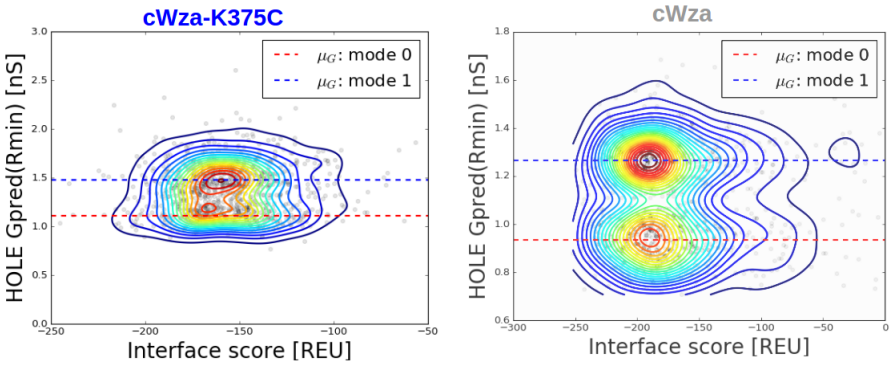


Figure E.1: *Left*: Scatter plot of *interface scores* (strength of association) against *numerically estimated conductance* for a 1000 optimal 8-unit models with cWza-K375C sequence, found using RosettaMP via a symmetry-constrained docking protocol. *Right*: Results for cWza-channels are shown again (Fig., 3.16) for comparison only. *Two conductance states with similar energy distributions can be distinguished for cWza-K375C docked structures, but with lower resolution in contrast to cWza-models.*

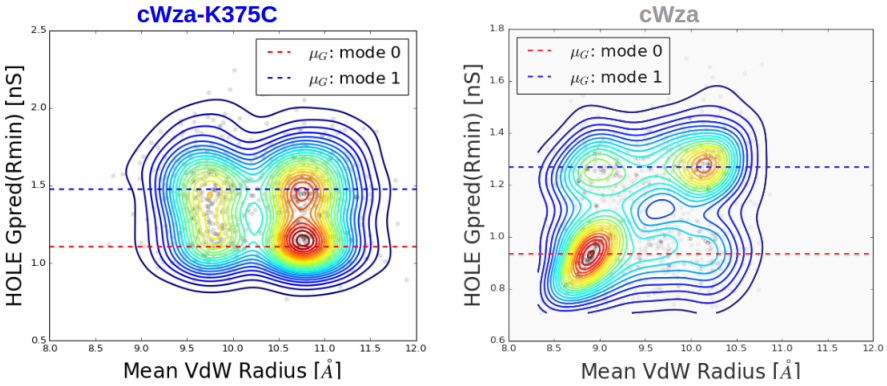


Figure E.2: Similar plot to Fig., 3.18, *Left*. *In contrast to cWza-sequence results, association between separate narrow and wide pore dimensions and low and high conductance less pronounced for cWza-K375C channels.*

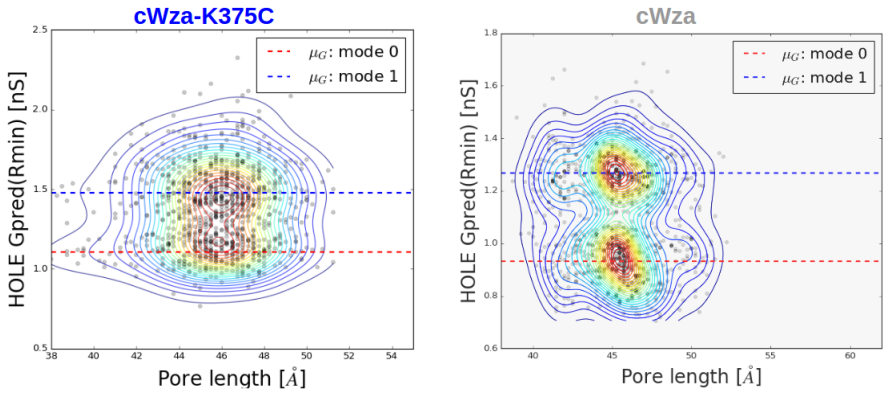


Figure E.3: Similar plot to Fig., 3.18, *Right*. *Similar to cWza-results, pore length cannot explain dual conductance separation for cWza-K375C channels either.*

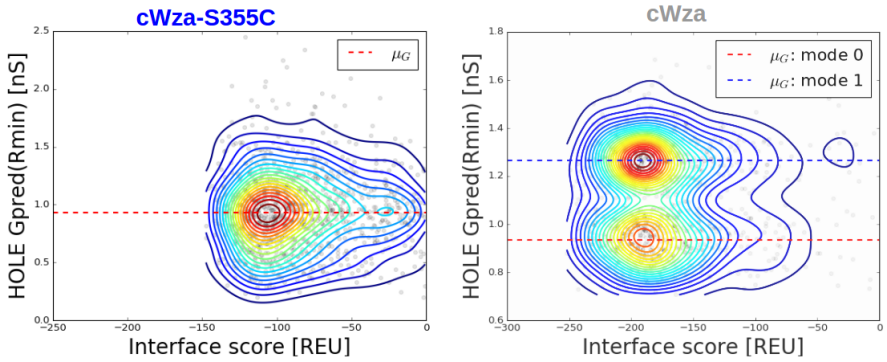


Figure E.4: *Left*: Scatter plot of *interface scores* (strength of association) against *numerically estimated conductance* for a 1000 optimal 8-unit models with cWza-S355C sequence, found using RosettaMP via a symmetry-constrained docking protocol. *Right*: Results for cWza-channels are shown again (Fig., 3.16) for comparison only. *In contrast to cWza-models, neither a conductance nor energy separation can be distinguished for cWza-S355C docked structures.*

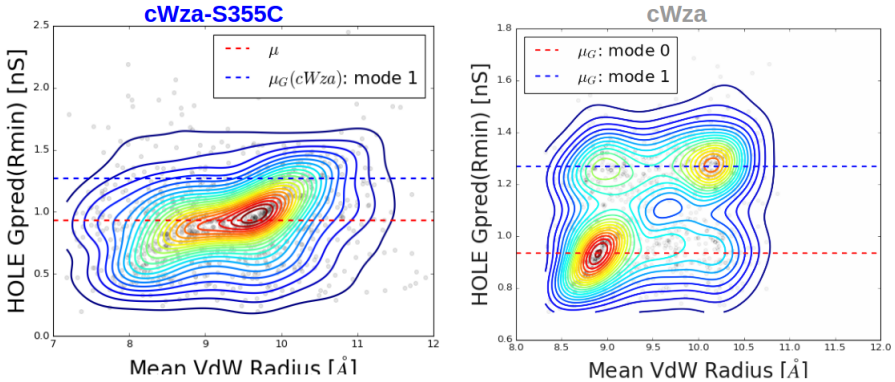


Figure E.5: Similar plot to Fig., 3.18, *Left*. *In contrast to cWza-sequence results, for cWza-S355C channels we see that although most models will show a preferential mean VdW radius, there will not be a*

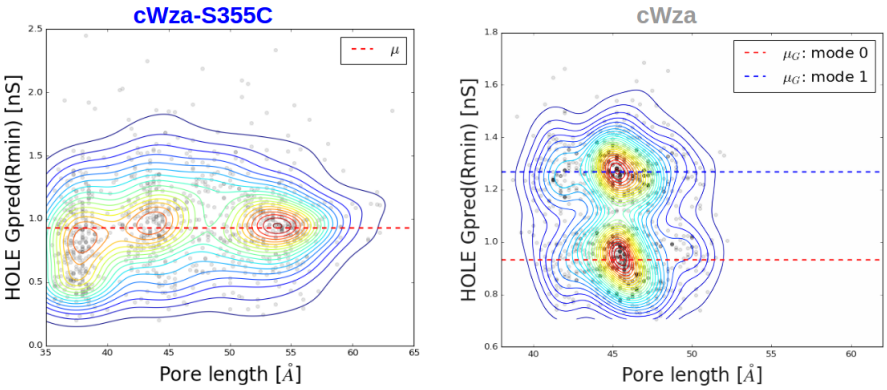


Figure E.6: Similar plot to Fig., 3.18, *Right*. *Unlike results for cWza (*right*) and cWza-K375C, pore lengths for cWza-S355C models show three separate preferential values. However, none of these relate to a clear change in conductance.*

APPENDIX F

LOW AND HIGH CONDUCTANCE *cWza* MODELS UNDER SYMMETRY PERTURBATION

NOTE: Figure F.1
All docked structures were relaxed using the `FastRelax` application of RosettaMP; obtaining 10 output structures per input structure. The output structure with the lowest total RosettaMP energy score was only chosen to estimate its conductance (G_{pred}) via HOLE; with empirical correction by HOLE-determined VdW minimum radius (R_{min}).

NOTE: Figure F.2
Prior to MDEM, docked structures with *Low* (*High*) conductance were sampled (100 models per state) provided their conductance (G) satisfied: $|G - \mu(G)_{L(H)}| \leq \frac{1}{2} \cdot \sigma(G)_{L(H)}$; with $\mu(G)_{L(H)}$ and $\sigma(G)_{L(H)}$ the mean and the standard deviation of the *Low* (*High*) conductance mode of the conductance distribution in Figure F.1 (*Left*); mode parameters were found via a *Gaussian-Mixture* fitting of the bi-modal distribution.

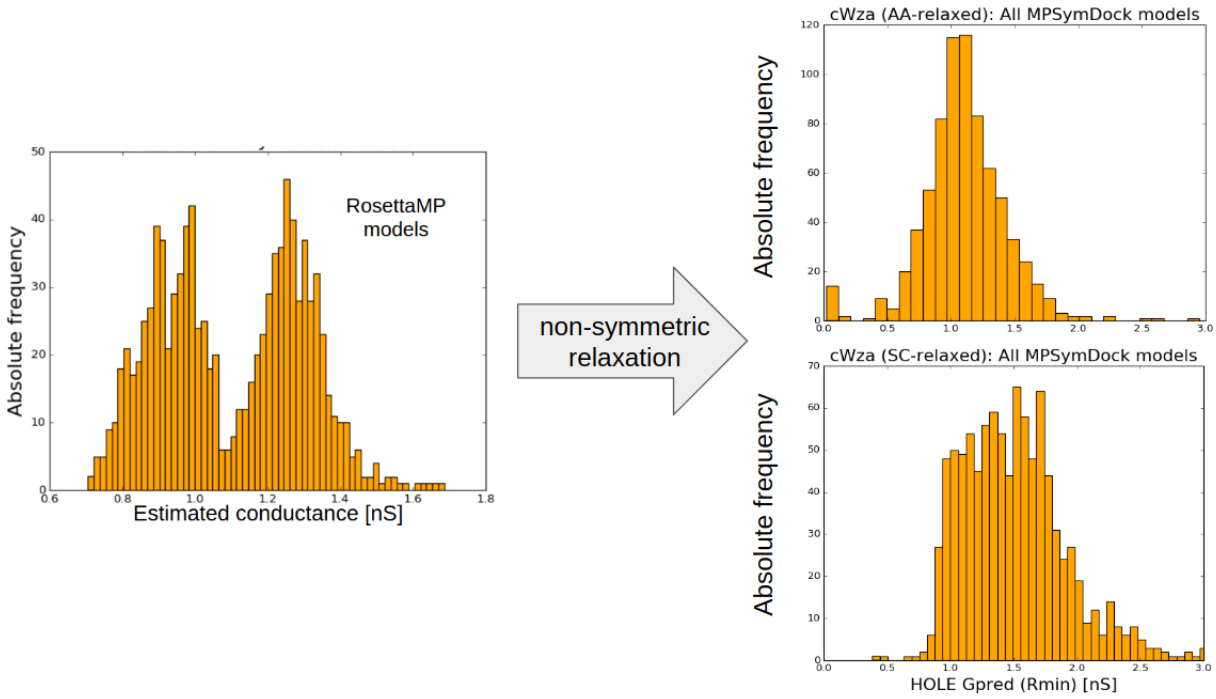


Figure F.1: Non-symmetric relaxation of docked cWza channels led to loss of separate conductance states; **Low** and *High* conductance distribution modes (*Left*), when either the coordinates of side-chain atoms (**Right - Bottom**) or all atoms (**Right - Top**) are perturbed towards nearest equilibrium state.

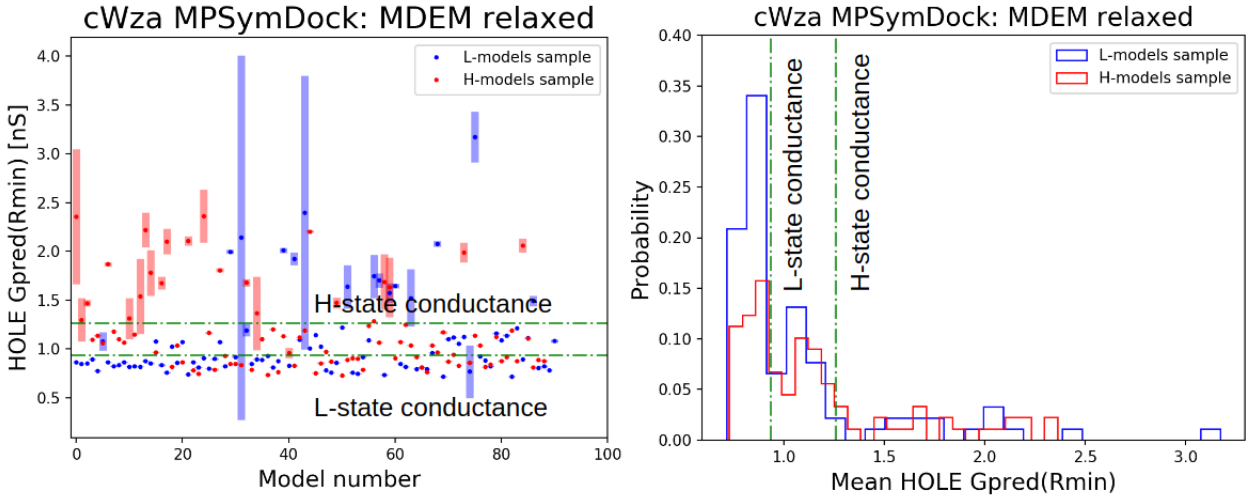


Figure F.2: Relaxation of docked cWza channels via MD Energy Minimisation (MDEM) led to loss of separate conductance states. **Left**: Estimated conductance of peptide-assembled channels at the end of MDEM; each data-point is the average over 10 attempts of a HOLE conductance calculation of the same structure, with bars representing the error. **Right**: Distribution of post-MDEM conductance values. $\mu(G)_{L(H)}$ and $\sigma(G)_{L(H)}$ are the mean and the standard deviation of the *Low* (*High*) conductance mode of the conductance distribution in Figure F.1 (*Left*). Green dotted lines (-.-) mark the positions of the $\mu(G)_{L(H)}$ values in all plots.

APPENDIX G

ENERGETICS AND CONDUCTANCE ESTIMATES OF DOCKED OLIGOMERIC MODELS

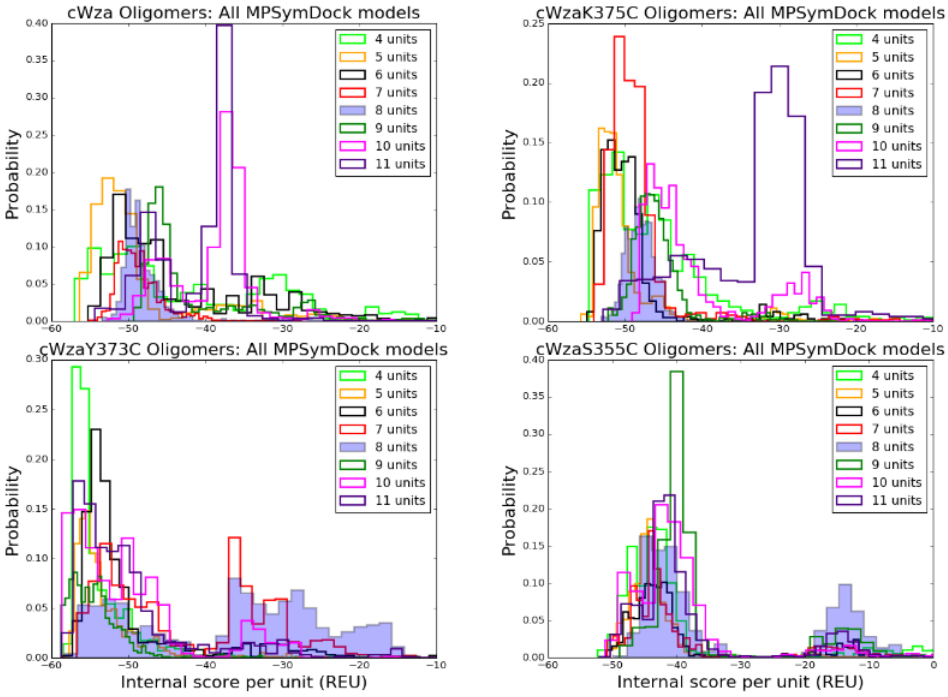


Figure G.1: Distributions of RosettaMP *internal energy* score values per peptide-unit; in Rosetta Energy Units (REU), for all docked oligomers with identical cWza *Cys*-mutant peptide sequences but with variable number of assembling peptides ($n = 4 - 11$ units). Distributions for octameric models ($n = 8$) are shaded.

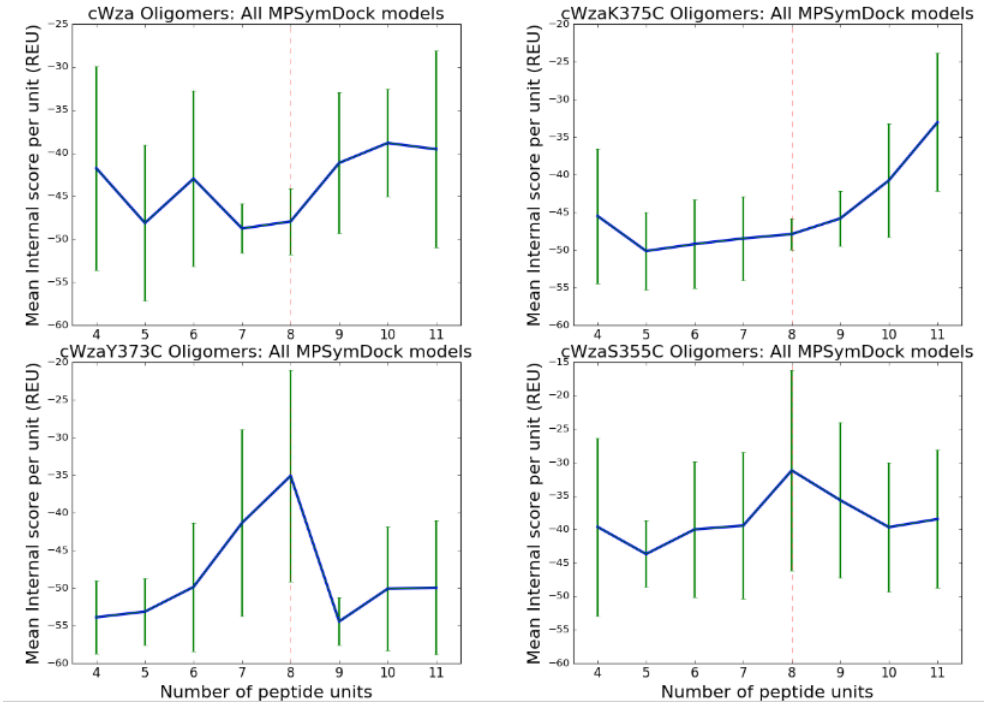


Figure G.2: Mean (μ) and Standard Deviation (σ) values per peptide-unit for the corresponding distributions of *internal energy* score values in Figure G.1 are plotted for all oligomers, per sequence. Dotted lines (- -) indicate the position of the data for octameric assemblies in all subplots.

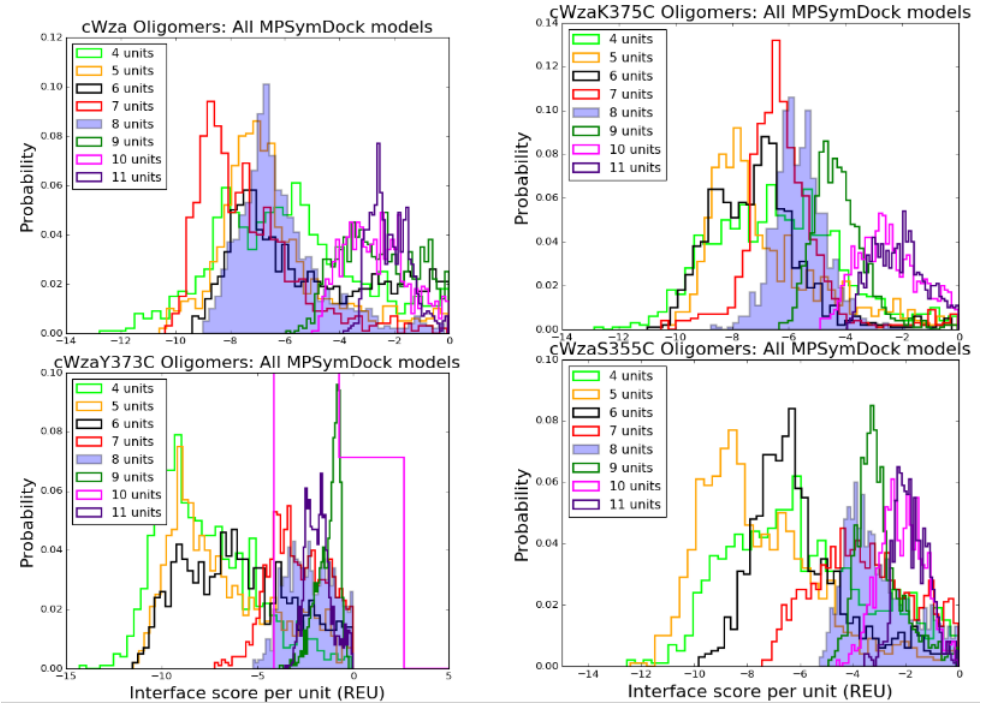


Figure G.3: Distributions of RosettaMP *interface energy* score values per peptide-unit; in Rosetta Energy Units (REU), for all docked oligomers with identical cWza *Cys*-mutant peptide sequences but with variable number of assembling peptides ($n = 4 - 11$ units). Distributions for octameric models ($n = 8$) are shaded.

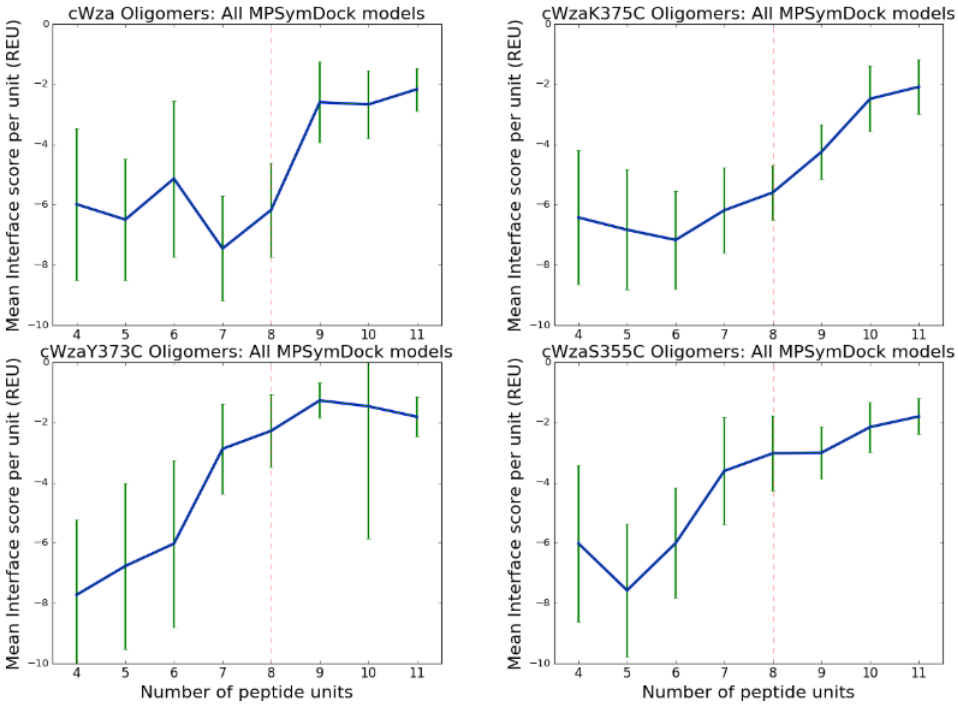


Figure G.4: Mean (μ) and Standard Deviation (σ) values for the corresponding distributions of *interface energy* score values per peptide-unit in Figure G.1 are plotted for all oligomers, per sequence. Dotted lines (- -) indicate the position of the data for octameric assemblies in all subplots. *Octameric assemblies do not have a preferentially lower energy than other oligomers.*

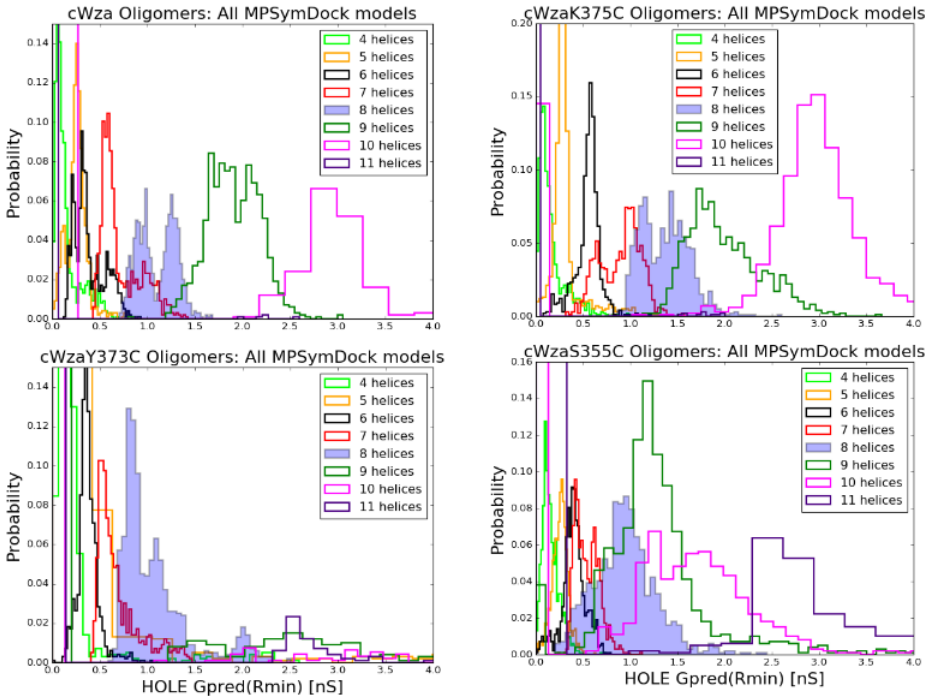


Figure G.5: Distributions of HOLE estimated conductance values (G_{pred}) for all docked oligomers with identical cWza *Cys*-mutant peptide sequences but with variable number of assembling peptides ($n = 4-11$ units). Distributions for octameric models ($n = 8$) are shaded. *Note that alternate conductive states also appear for non-octameric oligomers*.

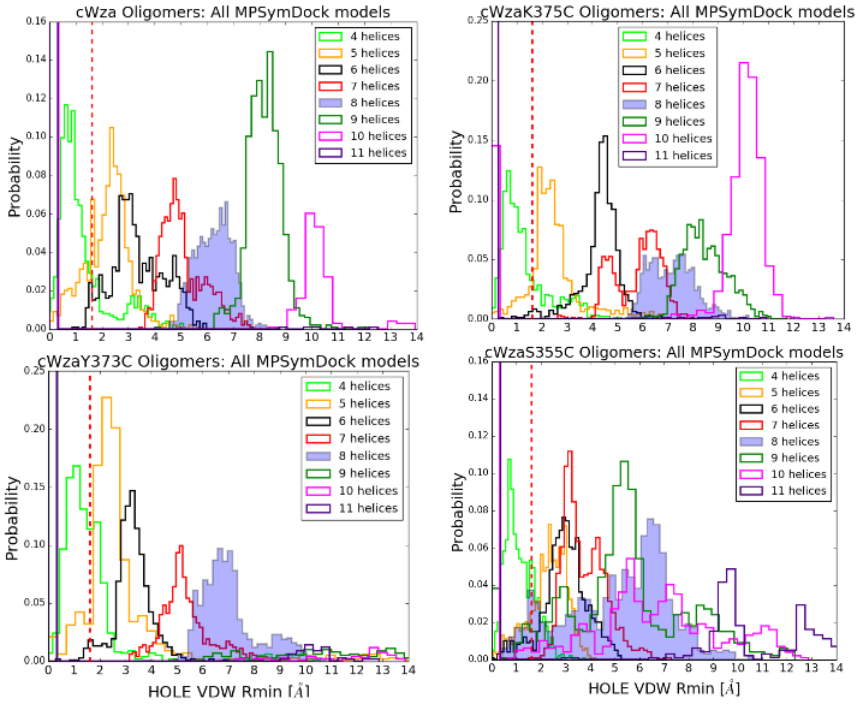


Figure G.6: Distributions of HOLE estimated VdW minimum radii values (R_{min}) for all docked oligomers with identical cWza *Cys*-mutant peptide sequences but with variable number of assembling peptides ($n = 4-11$ units). Distributions for octameric models ($n = 8$) are shaded. Also, dotted lines (- -) indicate the value of the mean VdW radius for a water molecule (1.37Å) in all subplots. *Note that pores of tetramers ($n = 4$) are not wide enough for solvent permeation*.

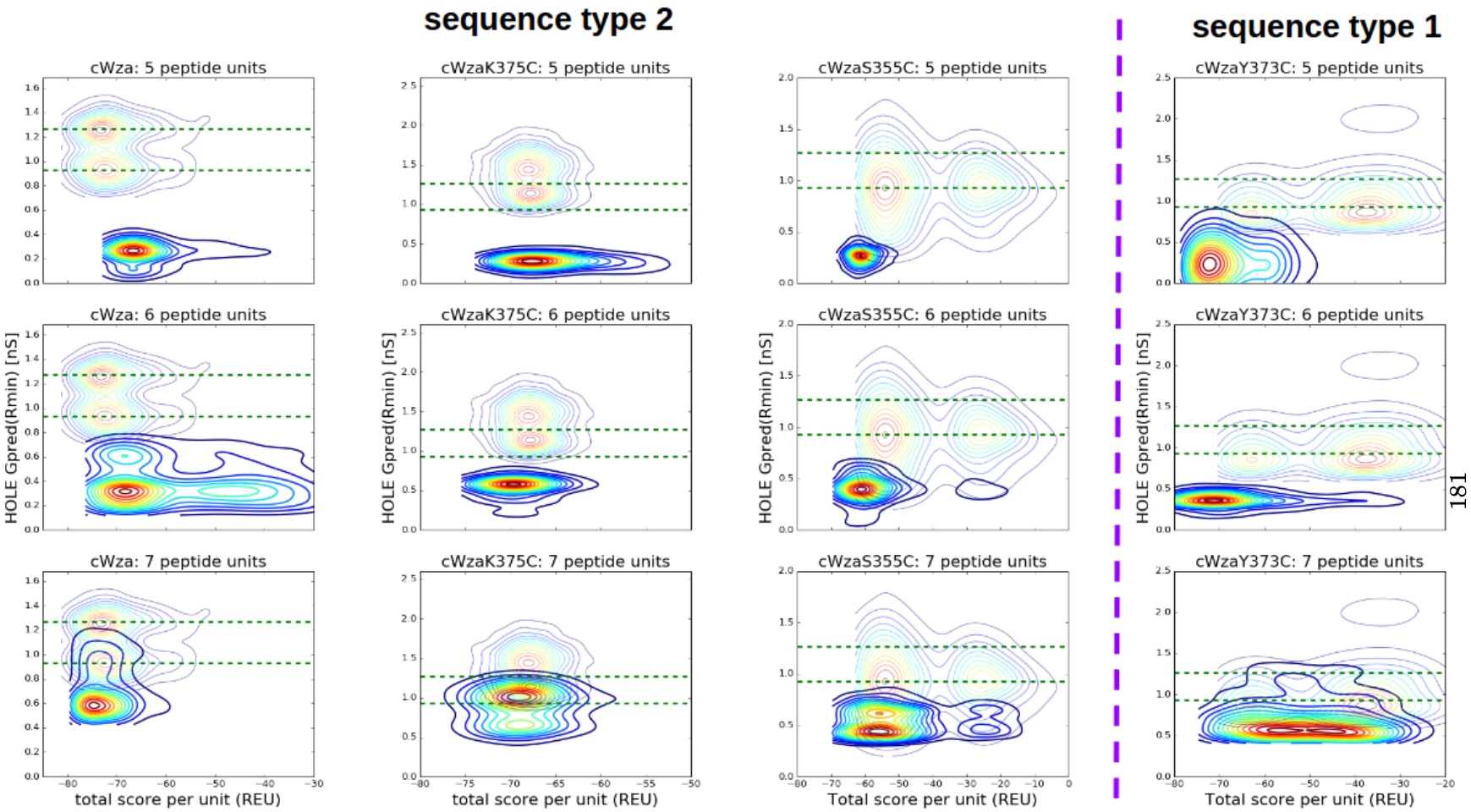


Figure G.7: Contour plots from scatter plots of *total energy scores per unit* against estimated conductance values for all docked model, per sequence, per number of chains; $n = 5 - 7$. Again, colours represent the density of scattered data, with red, the densest, and blue, the least dense accumulation of data-points. All contour plots in transparent correspond to data from octameric models; included for comparison. *Note that dual conductance is reproduced by oligomeric structures with* $n \neq 8$ *too.*

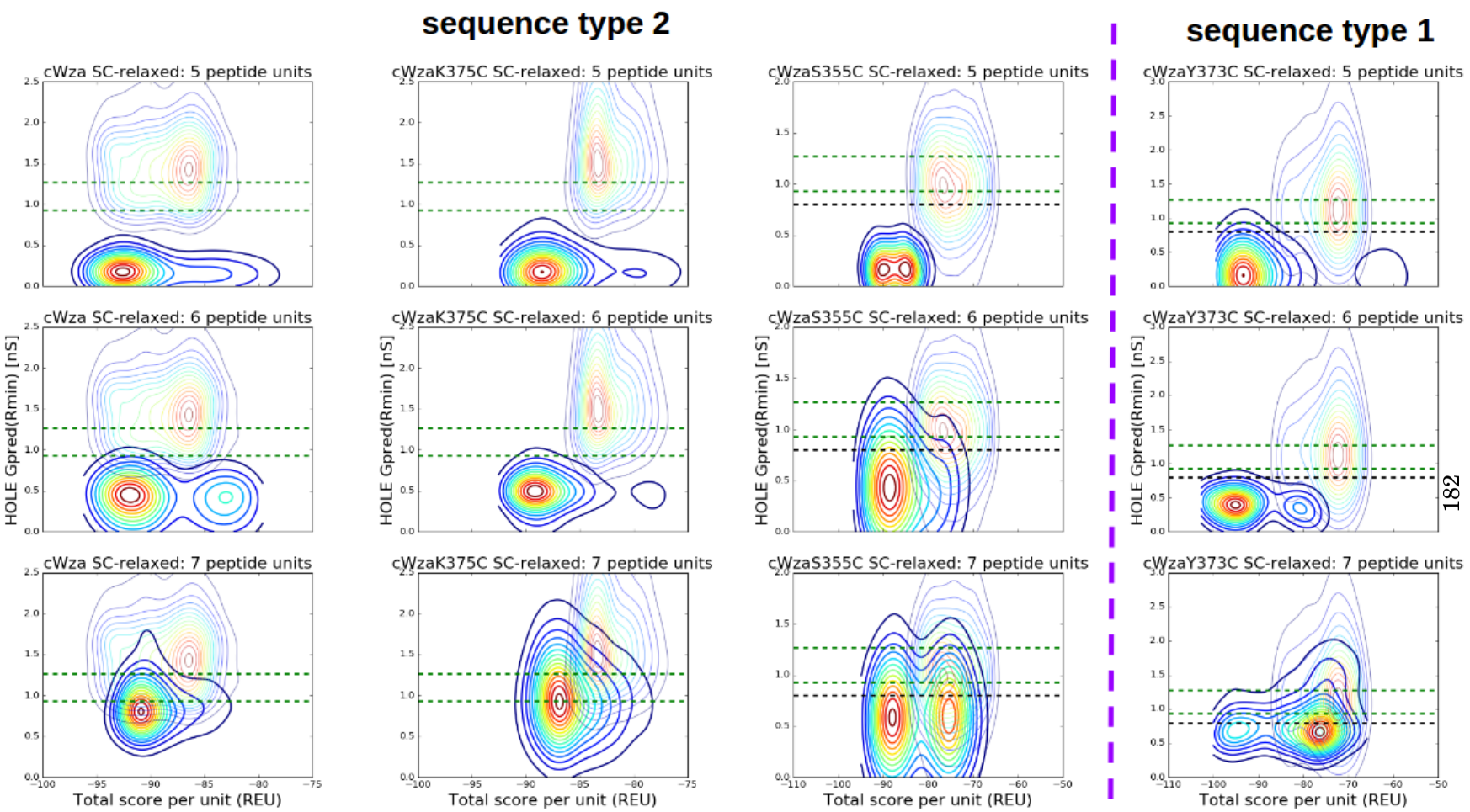


Figure G.8: Contour plots from scatter plots of *total energy scores per unit* against estimated conductance values for all docked model after *sidechain(SC) non-symmetric relaxation*, per sequence, per number of chains; $n = 5 - 7$. Again, colours represent density of scattered data, with red, the densest, and blue, the least dense accumulation of data-points. All contour plots in transparent correspond to data from octameric models; included for comparison. *Note that dual conductance vanishes after perturbing symmetry for sidechain atoms, similarly to when all-atoms are relaxed; Fig., 3.21.*

APPENDIX H

Reproducibility Assays for MD Conductance Calculations

Feature	PRMD (0V,5ns)	MD (0V,5ns)	PRMD (±1.2V)	PRMD (±0.2V)	MD (±0.2V)
NVT-ensemble	NPT	Yes	Yes	Yes	Yes
Time-step (dt) [fs]	1 fs	1 fs	2 fs	2 fs	2 fs
Velocity generation	Yes	Yes	No	Yes	No
Starting conditions	*Coords*	*Coords*	*Coords* / *Vels*	*Coords* / *Vels*	*Coords* / *Vels*

Table H.1: **Protocol A**

Feature	PRMD (0V,5ns)	MD (0V,5ns)	PRMD (±1.2V)	PRMD (±0.2V)	MD (±0.2V)
NVT-ensemble	NPT	Yes	Yes	Yes	Yes
Time-step (dt) [fs]	1 fs	2 fs	2 fs	2 fs	2 fs
Velocity generation	Yes	No	No	No	No
Starting conditions	*Coords*	*Coords*	*Coords*	*Coords*	*Coords*

Table H.2: **Protocol B**

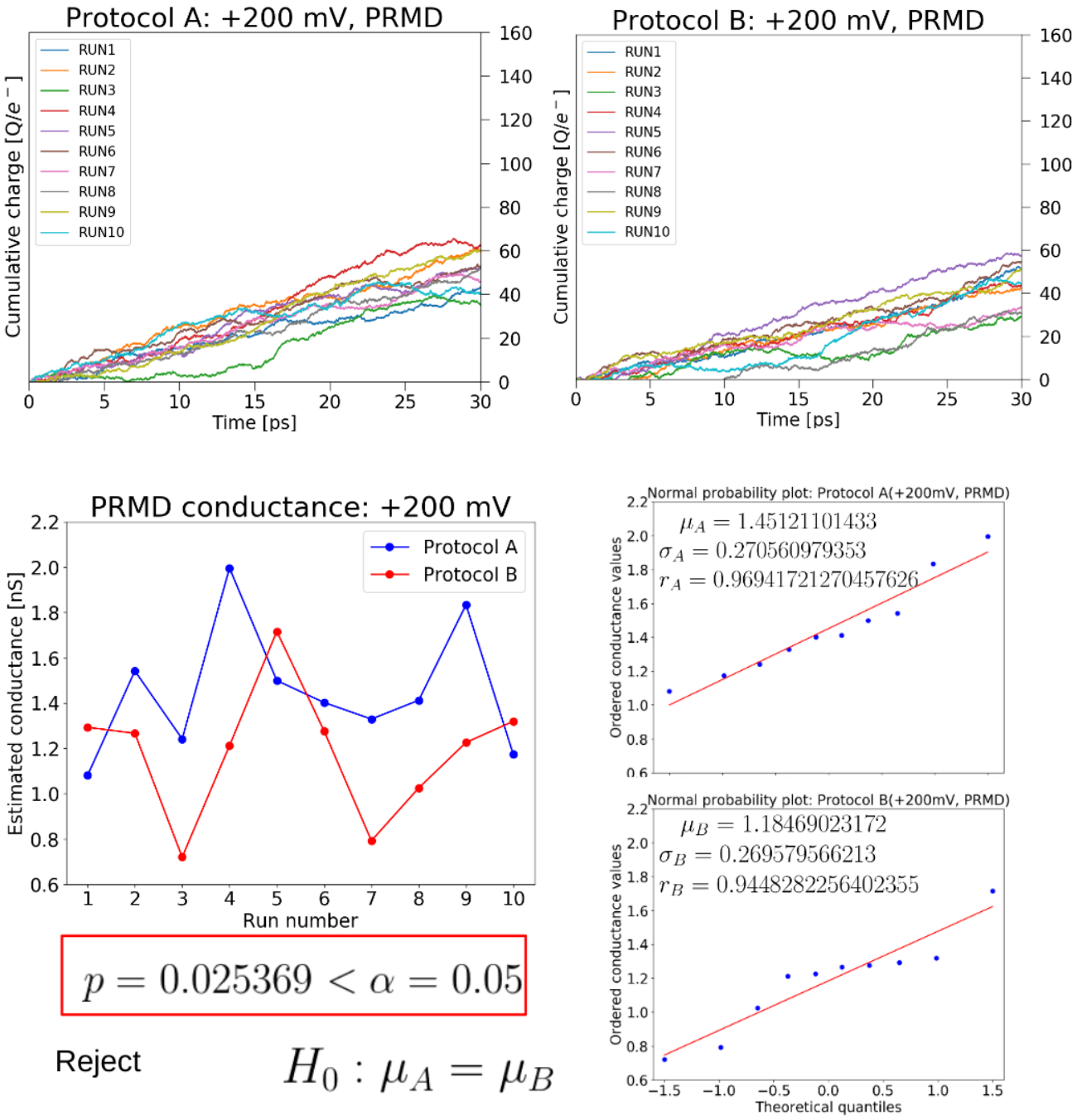


Figure H.1: Welch's T-test to determine the significance of the effect of the protocol selection on mean conductance calculations. PRMD stage with applied voltage.

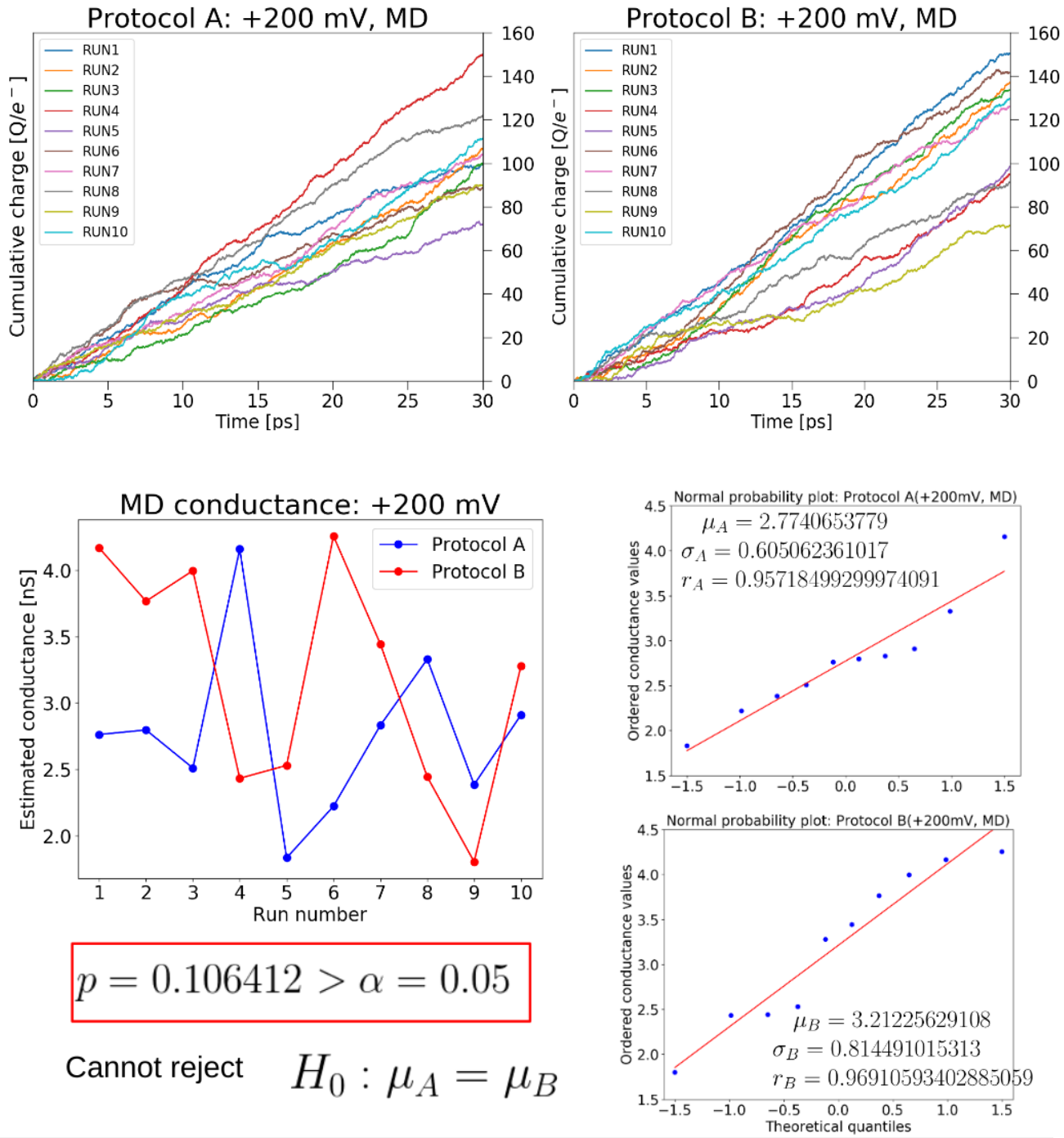
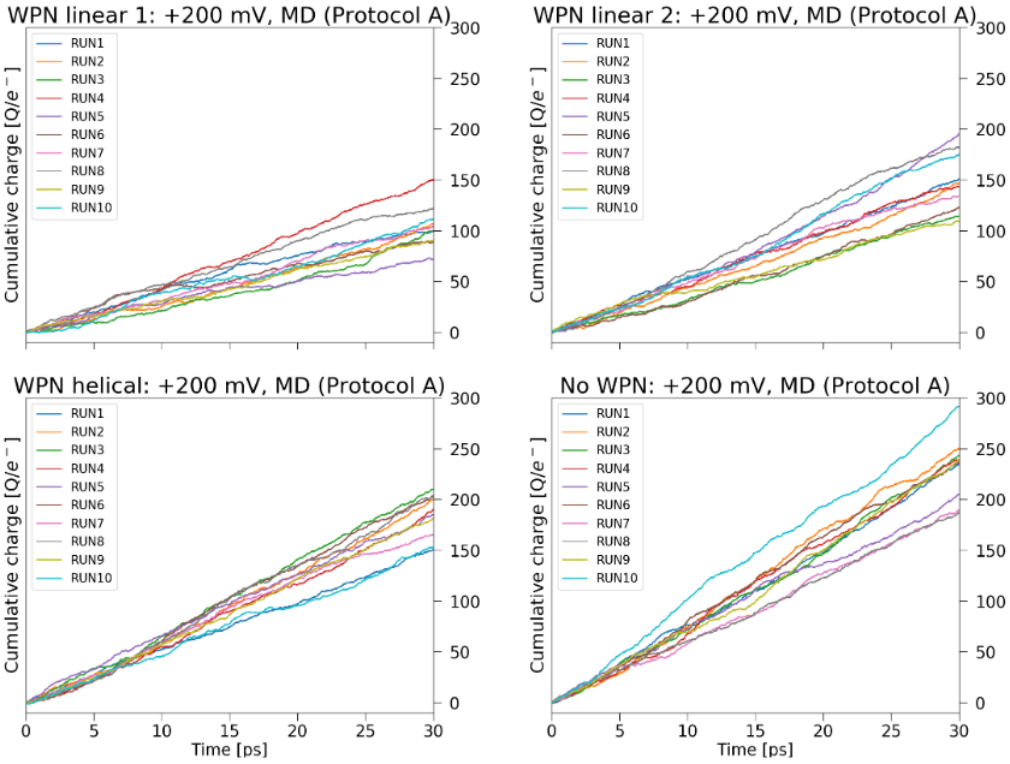


Figure H.2: Welch's T-test to determine the significance of the effect of the protocol selection on mean conductance calculations. Unrestrained-MD stage with applied voltage.



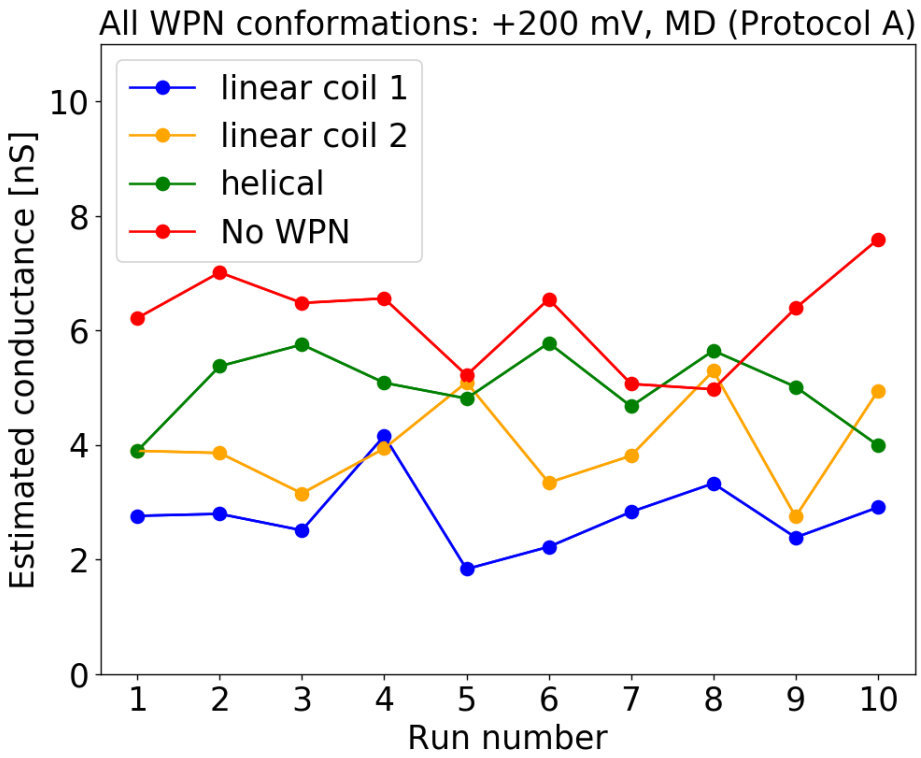


Figure H.3: Effect on the conductance calculation due to the secondary structure of added C-Terminal missing residues in the crystal structure of Wza

BIBLIOGRAPHY

[1] S. AHMADI, L. BARRIOS HERRERA, M. CHEHELAMIRANI, J. HOSTAŠ, S. JALIFE, AND D. R. SALAHUB, *Multiscale modeling of enzymes: QM-cluster, QM/MM, and QM/MM/MD: A tutorial review*, International Journal of Quantum Chemistry, 118 (2018), p. e25558.

[2] A. AKSIMENTIEV AND K. SCHULTEN, *Imaging α-Hemolysin with Molecular Dynamics: Ionic Conductance, Osmotic Permeability, and the Electrostatic Potential Map*, Biophysical Journal, 88 (2005), pp. 3745–3761.

[3] R. F. ALFORD, J. K. LEMAN, B. D. WEITZNER, A. M. DURAN, D. C. TILLEY, A. ELAZAR, AND J. J. GRAY, *An Integrated Framework Advancing Membrane Protein Modeling and Design*, PLOS Computational Biology, 11 (2015), p. e1004398.

[4] M. P. ALLEN AND D. J. TILDESLEY, *Computer simulation of liquids*.

[5] J. G. ALMEIDA, A. J. PRETO, P. I. KOUKOS, A. M. BONVIN, AND I. S. MOREIRA, *Membrane proteins structures: A review on computational modeling tools*, Biochimica et Biophysica Acta (BBA) - Biomembranes, 1859 (2017), pp. 2021–2039.

[6] G. A. K. AND, R. A. FRIESNER*, J. T.-R. AND, AND W. L. JORGENSEN, *Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides†*, (2001).

[7] H. C. ANDERSEN, *Molecular dynamics simulations at constant pressure and/or temperature*, The Journal of Chemical Physics, 72 (1980), p. 2384.

[8] I. ANDRÉ, P. BRADLEY, C. WANG, AND D. BAKER, *Prediction of the structure of symmetrical protein assemblies.*, Proceedings of the National Academy of Sciences of the United States of America, 104 (2007), pp. 17656–61.

[9] E. ARUNAN, G. R. DESIRAJU, R. A. KLEIN, J. SADLEJ, S. SCHEINER, I. ALKORTA, D. C. CLARY, R. H. CRABTREE, J. J. DANNENBERG, P. HOBZA, H. G. KJAERGAARD, A. C. LEGON, B. MENNUCCI, AND D. J. NESBITT, *Definition of the hydrogen bond (IUPAC Recommendations 2011)*, Pure and Applied Chemistry, 83 (2011), pp. 1637–1641.

[10] P. ARYAL, M. S. SANSOM, AND S. J. TUCKER, *Hydrophobic Gating in Ion Channels*, Journal of Molecular Biology, 427 (2015), pp. 121–130.

[11] P. ARYAL, M. S. P. SANSOM, AND S. J. TUCKER, *Hydrophobic gating in ion channels.*, Journal of molecular biology, 427 (2015), pp. 121–30.

[12] P. BARTH, J. SCHONBRUN, AND D. BAKER, *Toward high-resolution prediction and design of transmembrane helical protein structures*, Proceedings of the National Academy of Sciences, 104 (2007), pp. 15682–15687.

[13] H. BAYLEY AND L. JAYASINGHE, *Functional engineered channels and pores (Review)*, Molecular Membrane Biology, 21 (2004), pp. 209–220.

[14] J. M. J. M. BERG, J. L. TYMOCZKO, L. STRYER, AND L. STRYER, *Biochemistry*, W.H. Freeman, 2007.

[15] O. BERGER, O. EDHOLM, AND F. JÄHNIG, *Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature*, Biophysical Journal, 72 (1997), pp. 2002–2013.

[16] R. B. BEST, X. ZHU, J. SHIM, P. E. M. LOPES, J. MITTAL, M. FEIG, AND A. D. MACKERELL, *Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone , ψ and Side-Chain χ ₁ and χ ₂ Dihedral Angles*, Journal of Chemical Theory and Computation, 8 (2012), pp. 3257–3273.

[17] S. BHATTACHARYA, J. MUZARD, L. PAYET, J. MATHÉ, U. BOCKELMANN, A. AKSIMENTIEV, AND V. VIASNOFF, *Rectification of the Current in α-Hemolysin Pore Depends on the Cation Type: The Alkali Series Probed by Molecular Dynamics Simulations and Experiments*, The Journal of Physical Chemistry C, 115 (2011), pp. 4255–4264.

[18] M. BORN AND R. OPPENHEIMER, *Zur Quantentheorie der Molekeln*, Annalen der Physik, 389 (1927), pp. 457–484.

[19] B. R. BROOKS, R. E. BRUCCOLERI, B. D. OLAFSON, D. J. STATES, S. SWAMINATHAN, AND M. KARPLUS, *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations*, Journal of Computational Chemistry, 4 (1983), pp. 187–217.

[20] C. E. CAPENER, I. H. SHRIVASTAVA, K. M. RANATUNGA, L. R. FORREST, G. R. SMITH, AND M. S. SANSOM, *Homology Modeling and Molecular Dynamics Simulation Studies of an Inward Rectifier Potassium Channel*, Biophysical Journal, 78 (2000), pp. 2929–2942.

[21] E. P. CARPENTER, K. BEIS, A. D. CAMERON, AND S. IWATA, *Overcoming the challenges of membrane protein crystallography.*, Current opinion in structural biology, 18 (2008), pp. 581–6.

[22] S. CHAUDHURY, M. BERRONDO, B. D. WEITZNER, P. MUTHU, H. BERGMAN, AND J. J. GRAY, *Benchmarking and Analysis of Protein Docking Performance in Rosetta v3.2*, PLoS ONE, 6 (2011), p. e22477.

[23] K.-Y. M. CHEN, J. SUN, J. S. SALVO, D. BAKER, AND P. BARTH, *High-Resolution Modeling of Transmembrane Helical Protein Structures from Distant Homologues*, PLoS Computational Biology, 10 (2014), p. e1003636.

[24] DAAN FRENKEL AND BEREND SMIT, *Understanding Molecular Simulation*, Elsevier, 2002.

[25] T. DARDEN, D. YORK, AND L. PEDERSEN, *Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems*, The Journal of Chemical Physics, 98 (1993), p. 10089.

[26] S. W. DE LEEUW, J. W. PERRAM, AND E. R. SMITH, *Simulation of Electrostatic Systems in Periodic Boundary Conditions. I. Lattice Sums and Dielectric Constants*, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 373 (1980), pp. 27–56.

[27] C. J. DICKSON, B. D. MADEJ, Å. A. SKJEVIK, R. M. BETZ, K. TEIGEN, I. R. GOULD, AND R. C. WALKER, *Lipid14: The Amber Lipid Force Field*, Journal of Chemical Theory and Computation, 10 (2014), pp. 865–879.

[28] D. J. DIEUDONNÉ, P. A. GURNEV, A. L. ROGOZEA, B. D. RAY, AND H. I. PETRACHE, *Physical Properties of the Lipid Diphytanoyl Phosphatidylcholine (DPhPc) used for Ion Channel Measurements*, Biophysical Journal, 98 (2010), pp. 667a–668a.

[29] F. DIMAIO, A. LEAVER-FAY, P. BRADLEY, D. BAKER, AND I. ANDRÉ, *Modeling Symmetric Macromolecular Structures in Rosetta3*, PLoS ONE, 6 (2011), p. e20450.

[30] C. DONG, K. BEIS, J. NESPER, A. L. BRUNKAN-LAMONTAGNE, B. R. CLARKE, C. WHITFIELD, AND J. H. NAISMITH, *Wza the translocon for E. coli capsular polysaccharides defines a new class of membrane protein*, Nature, 444 (2006), pp. 226–229.

[31] R. FRIEDMAN, S. KHALID, C. APONTE-SANTAMARÍA, E. ARUTYUNOVA, M. BECKER, K. J. BOYD, M. CHRISTENSEN, J. T. S. COIMBRA, S. CONCILIO, C. DADAY, F. J. VAN EERDEN, P. A. FERNANDES, F. GRÄTER, D. HAKOBYAN, A. HEUER, K. KARATHANOU, F. KELLER, M. J. LEMIEUX, S. J. MARRINK, E. R. MAY, A. MAZUMDAR, R. NAFTALIN, M. PICKHOLZ, S. PIOTTO, P. POHL, P. QUINN, M. J. RAMOS, B. SCHIØTT, D. SENGUPTA, L. SESSA, S. VANNI, T. ZEPPELIN, V. ZONI, A.-N. BONDAR, AND C. DOMENE, *Understanding Conformational Dynamics of Complex Lipid Mixtures Relevant to Biology*, The Journal of Membrane Biology, 251 (2018), pp. 609–631.

[32] L. FU, A. BIENENSTOCK, AND S. BRENNAN, *X-ray study of the structure of liquid water*, The Journal of Chemical Physics, 131 (2009), p. 234702.

[33] R. J. C. GILBERT, H. BAYLEY, AND G. ANDERLUH, *Membrane pores: from structure and assembly, to medicine and technology*, Philosophical Transactions of the Royal Society B: Biological Sciences, 372 (2017), p. 20160208.

[34] D. S. GOODSELL AND A. J. OLSON, *Structural Symmetry and Protein Function*, Annual Review of Biophysics and Biomolecular Structure, 29 (2000), pp. 105–153.

[35] J. J. GRAY, S. MOUGHON, C. WANG, O. SCHUELER-FURMAN, B. KUHLMAN, C. A. ROHL, AND D. BAKER, *Protein–Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations*, Journal of Molecular Biology, 331 (2003), pp. 281–299.

[36] J. GUMBART, F. KHALILI-ARAGHI, M. SOTOMAYOR, AND B. ROUX, *Constant electric field simulations of the membrane potential illustrated with simple systems.*, Biochimica et biophysica acta, 1818 (2012), pp. 294–302.

[37] I. HALPERIN, B. MA, H. WOLFSON, AND R. NUSSINOV, *Principles of docking: An overview of search algorithms and a guide to scoring functions*, Proteins: Structure, Function, and Genetics, 47 (2002), pp. 409–443.

[38] J. HARRISON, C. HIGGINS, M. O'MEARA, J. KOELLHOFFER, B. KUHLMAN, AND J. LAI, *Role of Electrostatic Repulsion in Controlling pH-Dependent Conformational Changes of Viral Fusion Proteins*, Structure, 21 (2013), pp. 1085–1096.

[39] S. C. HARRISON, *The pH sensor for flavivirus membrane fusion.*, The Journal of cell biology, 183 (2008), pp. 177–9.

[40] K. HE, S. LUDTKE, D. WORCESTER, AND H. HUANG, *Neutron scattering in the plane of membranes: structure of alamethicin pores*, Biophysical Journal, 70 (1996), pp. 2659–2666.

[41] B. HILLE, *Ion channels of excitable membranes*, Sinauer, 2001.

[42] R. HOCKNEY, S. GOEL, AND J. EASTWOOD, *Quiet high-resolution computer models of a plasma*, Journal of Computational Physics, 14 (1974), pp. 148–158.

[43] W. G. HOOVER, *Canonical dynamics: Equilibrium phase-space distributions*, Physical Review A, 31 (1985), pp. 1695–1697.

[44] W. HUMPHREY, A. DALKE, AND K. SCHULTEN, *VMD: Visual molecular dynamics*, Journal of Molecular Graphics, 14 (1996), pp. 33–38.

[45] C. JARZYNSKI, *Nonequilibrium Equality for Free Energy Differences*, Physical Review Letters, 78 (1997), pp. 2690–2693.

[46] M. Ø. JENSEN, D. W. BORHANI, K. LINDORFF-LARSEN, P. MARAGAKIS, V. JOGINI, M. P. EASTWOOD, R. O. DROR, AND D. E. SHAW, *Principles of conduction and hydrophobic gating in K+ channels.*, Proceedings of the National Academy of Sciences of the United States of America, 107 (2010), pp. 5833–8.

[47] N. H. JOH, G. GRIGORYAN, Y. WU, AND W. F. DEGRADO, *Design of self-assembling transmembrane helical bundles to elucidate principles required for membrane protein folding and ion transport*, Philosophical Transactions of the Royal Society B: Biological Sciences, 372 (2017), p. 20160214.

[48] W. L. JORGENSEN AND J. TIRADO-RIVES, *The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin*, Journal of the American Chemical Society, 110 (1988), pp. 1657–1666.

[49] M. A. KASIMOVA, M. TAREK, A. K. SHAYTAN, K. V. SHAITAN, AND L. DELEMOTTE, *Voltage-gated ion channel modulation by lipids: Insights from molecular dynamics simulations*, Biochimica et Biophysica Acta (BBA) - Biomembranes, 1838 (2014), pp. 1322–1331.

[50] S. KELM, J. SHI, AND C. M. DEANE, *MEDELLER: homology-based coordinate generation for membrane proteins.*, Bioinformatics (Oxford, England), 26 (2010), pp. 2833–40.

[51] F. KHATIB, S. COOPER, M. D. TYKA, K. XU, I. MAKEDON, Z. POPOVIĆ, D. BAKER, AND F. PLAYERS, *Algorithm discovery by protein folding game players*, Proceedings of the National Academy of Sciences, 108 (2011), pp. 18949–18953.

[52] J. B. KLAUDA, R. M. VENABLE, J. A. FREITES, J. W. O'CONNOR, D. J. TOBIAS, C. MONDRAGON-RAMIREZ, I. VOROBYOV, A. D. MACKERELL, AND R. W. PASTOR, *Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types*, The Journal of Physical Chemistry B, 114 (2010), pp. 7830–7843.

[53] L. KONG, L. HARRINGTON, Q. LI, S. CHELEY, B. G. DAVIS, AND H. BAYLEY, *Single-molecule interrogation of a bacterial sugar transporter allows the discovery of an extracellular inhibitor.*, Nature chemistry, 5 (2013), pp. 651–9.

[54] B. KUHLMAN, D. BAKER, C. WANG, AND D. BAKER, *Native protein sequences are close to optimal for their structures.*, Proceedings of the National Academy of Sciences of the United States of America, 97 (2000), pp. 10383–8.

[55] A. LAIO AND M. PARRINELLO, *Escaping free-energy minima.*, Proceedings of the National Academy of Sciences of the United States of America, 99 (2002), pp. 12562–6.

[56] D. LANGOSCH AND J. HERINGA, *Interaction of transmembrane helices by a knobs-into-holes packing characteristic of soluble coiled coils*, Proteins: Structure, Function, and Genetics, 31 (1998), pp. 150–159.

[57] R. J. LAW, R. H. HENCHMAN, AND J. A. MCCAMMON, *A gating mechanism proposed from a simulation of a human alpha7 nicotinic acetylcholine receptor.*, Proceedings of the National Academy of Sciences of the United States of America, 102 (2005), pp. 6813–8.

[58] T. LAZARIDIS, *Effective energy function for proteins in lipid membranes*, Proteins: Structure, Function, and Genetics, 52 (2003), pp. 176–192.

[59] A. LEAVER-FAY, M. TYKA, S. M. LEWIS, O. F. LANGE, J. THOMPSON, R. JACAK, K. W. KAUFMAN, P. D. RENFREW, C. A. SMITH, W. SHEFFLER, I. W. DAVIS, S. COOPER, A. TREUILLE, D. J. MANDELL, F. RICHTER, Y.-E. A. BAN, S. J. FLEISHMAN, J. E. CORN, D. E. KIM, S. LYSKOV, M. BERRONDO, S. MENTZER, Z. POPOVIĆ, J. J. HAVRANEK, J. KARANICOLAS, R. DAS, J. MEILER, T. KORTEMME, J. J. GRAY, B. KUHLMAN, D. BAKER, AND P. BRADLEY, *Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules*, Methods in Enzymology, 487 (2011), pp. 545–574.

[60] R. LONSDALE, J. N. HARVEY, AND A. J. MULHOLLAND, *A practical guide to modelling enzyme-catalysed reactions*, Chemical Society Reviews, 41 (2012), p. 3025.

[61] LUCIE DELEMOTTE, FRANÇOIS DEHEZ, , WERNER TREPTOW, AND M. TAREK*, *Modeling Membranes under a Transmembrane Potential*, (2008).

[62] A. D. MACKERELL, D. BASHFORD, M. BELLOTT, R. L. DUNBRACK, J. D. EVANSECK, M. J. FIELD, S. FISCHER, J. GAO, H. GUO, S. HA, D. JOSEPH-MCCARTHY, L. KUCHNIR, K. KUCZERA, F. T. K. LAU, C. MATTOS, S. MICHNICK, T. NGO, D. T. NGUYEN, B. PRODHOM, W. E. REIHER, B. ROUX, M. SCHLENKRICH, J. C. SMITH, R. STOTE, J. STRAUB, M. WATANABE, J. WIÓRKIEWICZ-KUCZERA, D. YIN, AND M. KARPLUS, *All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins [†]*, The Journal of Physical Chemistry B, 102 (1998), pp. 3586–3616.

[63] C. MAFFEO, S. BHATTACHARYA, J. YOO, D. WELLS, AND A. AKSIMENTIEV, *Modeling and simulation of ion channels.*, Chemical reviews, 112 (2012), pp. 6250–84.

[64] K. R. MAHENDRAN, A. NIITSU, L. KONG, A. R. THOMSON, R. B. SESSIONS, D. N. WOOLFSON, AND H. BAYLEY, *A monodisperse transmembrane α-helical peptide barrel*, Nature Chemistry, 9 (2016), pp. 411–419.

[65] G. J. MARTYNA, D. J. TOBIAS, AND M. L. KLEIN, *Constant pressure molecular dynamics algorithms*, The Journal of Chemical Physics, 101 (1994), p. 4177.

[66] A. MELLER AND D. BRANTON, *Single molecule measurements of DNA transport through a nanopore*, ELECTROPHORESIS, 23 (2002), pp. 2583–2591.

[67] H. MEUZELAAR, J. VREEDE, AND S. WOUTERSEN, *Influence of Glu/Arg, Asp/Arg, and Glu/Lys Salt Bridges on α-Helical Stability and Folding Kinetics*, Biophysical Journal, 110 (2016), pp. 2328–2341.

[68] I. MORAES, G. EVANS, J. SANCHEZ-WEATHERBY, S. NEWSTEAD, AND P. D. S. STEWART, *Membrane protein structure determination — The next generation*, Biochimica et Biophysica Acta (BBA) - Biomembranes, 1838 (2014), pp. 78–87.

[69] I. S. MOREIRA, P. A. FERNANDES, AND M. J. RAMOS, *Protein-protein docking dealing with the unknown*, Journal of Computational Chemistry, 31 (2009), pp. NA–NA.

[70] J.-P. MORNON, B. HOFFMANN, S. JONIC, P. LEHN, AND I. CALLEBAUT, *Full-open and closed CFTR channels, with lateral tunnels from the cytoplasm and an alternative position of the F508 region, as revealed by molecular dynamics*, Cellular and Molecular Life Sciences, 72 (2015), pp. 1377–1403.

[71] A. J. MULHOLLAND, *Dispelling the effects of a sorceress in enzyme catalysis.*, Proceedings of the National Academy of Sciences of the United States of America, 113 (2016), pp. 2328–30.

[72] A. NIITSU, J. W. HEAL, K. FAULAND, A. R. THOMSON, AND D. N. WOOLFSON, *Membrane-spanning α-helical barrels as tractable protein-design targets.*, Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 372 (2017), p. 20160213.

[73] S. NOSE, *A unified formulation of the constant temperature molecular dynamics methods*, The Journal of Chemical Physics, 81 (1984), p. 511.

[74] T. NUGENT, *De Novo Membrane Protein Structure Prediction*, Humana Press, New York, NY, 2015, pp. 331–350.

[75] T. NUGENT AND D. T. JONES, *Predicting Transmembrane Helix Packing Arrangements using Residue Contacts and a Force-Directed Algorithm*, PLoS Computational Biology, 6 (2010), p. e1000714.

[76] T. NUGENT AND D. T. JONES, *Accurate de novo structure prediction of large transmembrane protein domains using fragment-assembly and correlated mutation analysis.*, Proceedings of the National Academy of Sciences of the United States of America, 109 (2012), pp. E1540–7.

[77] T. NUGENT AND D. T. JONES, *Membrane protein structural bioinformatics*, Journal of Structural Biology, 179 (2012), pp. 327–337.

[78] T. NUGENT AND D. T. JONES, *Membrane protein orientation and refinement using a knowledge-based statistical potential*, BMC Bioinformatics, 14 (2013), p. 276.

[79] M. PARRINELLO, A. RAHMAN, AND P. VASHISHTA, *Structural Transitions in Superionic Conductors*, Physical Review Letters, 50 (1983), pp. 1073–1076.

[80] E. PEROZO, D. M. CORTES, P. SOMPORNPISUT, A. KLODA, AND B. MARTINAC, *Open channel structure of MscL and the gating mechanism ofmechanosensitive channels*, Nature, 418 (2002), pp. 942–948.

[81] N. A. PIERCE AND E. WINFREE, *Protein Design is NP-hard*, Protein Engineering, Design and Selection, 15 (2002), pp. 779–782.

[82] T. J. PIGGOT, Á. PIÑEIRO, AND S. KHALID, *Molecular Dynamics Simulations of Phosphatidylcholine Membranes: A Comparative Force Field Study*, Journal of Chemical Theory and Computation, 8 (2012), pp. 4593–4609.

[83] D. POGER, B. CARON, AND A. E. MARK, *Validating lipid force fields against experimental data: Progress, challenges and perspectives*, Biochimica et Biophysica Acta (BBA) - Biomembranes, 1858 (2016), pp. 1556–1565.

[84] J. W. PONDER AND F. M. RICHARDS, *Tertiary templates for proteins: Use of packing criteria in the enumeration of allowed sequences for different structural classes*, Journal of Molecular Biology, 193 (1987), pp. 775–791.

[85] J. A. POVEDA, A. MARCELA GIUDICI, M. LOURDES RENART, A. MORALES, AND J. M. GONZÁLEZ-ROS, *Towards understanding the molecular basis of ion channel modulation by lipids: Mechanistic models and current paradigms*, Biochimica et Biophysica Acta (BBA) - Biomembranes, 1859 (2017), pp. 1507–1516.

[86] K. RÖDER AND D. J. WALES, *Transforming the Energy Landscape of a Coiled-Coil Peptide via Point Mutations*, Journal of Chemical Theory and Computation, 13 (2017), pp. 1468–1477.

[87] B. ROUX, *The Membrane Potential and its Representation by a Constant Electric Field in Computer Simulations*, Biophysical Journal, 95 (2008), pp. 4205–4216.

[88] J. N. SACHS, P. S. CROZIER, AND T. B. WOOLF, *Atomistic simulations of biologically realistic transmembrane potential gradients*, The Journal of Chemical Physics, 121 (2004), p. 10847.

[89] A. ŠALI AND T. L. BLUNDELL, *Comparative Protein Modelling by Satisfaction of Spatial Restraints*, Journal of Molecular Biology, 234 (1993), pp. 779–815.

[90] A. SANDOVAL-PEREZ, K. PLUHACKOVA, AND R. A. BÖCKMANN, *Critical Comparison of Biomembrane Force Fields: Protein–Lipid Interactions at the Membrane Interface*, Journal of Chemical Theory and Computation, 13 (2017), pp. 2310–2321.

[91] N. SCHMID, A. P. EICHENBERGER, A. CHOUTKO, S. RINIKER, M. WINGER, A. E. MARK, AND W. F. VAN GUNSTEREN, *Definition and testing of the GROMOS force-field versions 54A7 and 54B7*, European Biophysics Journal, 40 (2011), pp. 843–856.

[92] B. SCHULZE, A. SLJOKA, AND W. WHITELEY, *How does symmetry impact the flexibility of proteins?*, Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 372 (2014), p. 20120041.

[93] W. SHINODA, M. MIKAMI, T. BABA, AND M. HATO, *Molecular Dynamics Study on the Effect of Chain Branching on the Physical Properties of Lipid Bilayers: Structural Stability*, (2003).

[94] J. SIMMS AND P. J. BOOTH, *Membrane proteins by accident or design*, Current Opinion in Chemical Biology, 17 (2013), pp. 976–981.

[95] A. J. SITU, S.-M. KANG, B. B. FREY, W. AN, C. KIM, AND T. S. ULMER, *Membrane Anchoring of α-Helical Proteins: Role of Tryptophan*, The Journal of Physical Chemistry B, 122 (2018), pp. 1185–1194.

[96] O. S. SMART, J. BREED, G. R. SMITH, AND M. S. SANSOM, *A novel method for structure-based prediction of ion channel conductance properties.*, Biophysical journal, 72 (1997), pp. 1109–26.

[97] O. S. SMART, J. M. GOODFELLOW, AND B. A. WALLACE, *The pore dimensions of gramicidin A.*, Biophysical journal, 65 (1993), pp. 2455–60.

[98] O. S. SMART, J. G. NEDUVELIL, X. WANG, B. WALLACE, AND M. S. SANSOM, *HOLE: A program for the analysis of the pore dimensions of ion channel structural models*, Journal of Molecular Graphics, 14 (1996), pp. 354–360.

[99] M. SOTOMAYOR, T. A. VAN DER STRAATEN, U. RAVAIOLI, AND K. SCHULTEN, *Electrostatic properties of the mechanosensitive channel of small conductance MscS.*, Biophysical journal, 90 (2006), pp. 3496–510.

[100] M. SOTOMAYOR, V. VÁSQUEZ, E. PEROZO, AND K. SCHULTEN, *Ion conduction through MscS as determined by electrophysiology and simulation.*, Biophysical journal, 92 (2007), pp. 886–902.

[101] P. STANSFELD, J. GOOSE, M. CAFFREY, E. CARPENTER, J. PARKER, S. NEWSTEAD, AND M. SANSOM, *MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes*, Structure, 23 (2015), pp. 1350–1361.

[102] M. O. STEINMETZ, I. JELESAROV, W. M. MATOUSEK, S. HONNAPPA, W. JAHNKE, J. H. MISSIMER, S. FRANK, A. T. ALEXANDRESCU, AND R. A. KAMMERER, *Molecular basis of coiled-coil formation.*, Proceedings of the National Academy of Sciences of the United States of America, 104 (2007), pp. 7062–7.

[103] K. M. STRICKLAND, G. STOCK, G. CUI, H. HWANG, D. T. INFIELD, I. SCHMIDT-KREY, N. A. MCCARTY, AND J. C. GUMBART, *ATP-Dependent Signaling in Simulations of a Revised Model of Cystic Fibrosis Transmembrane Conductance Regulator (CFTR)*, The Journal of Physical Chemistry B, 123 (2019), pp. 3177–3188.

[104] S. SUKHAREV, S. R. DURELL, AND H. R. GUY, *Structural Models of the MscL Gating Mechanism*, Biophysical Journal, 81 (2001), pp. 917–936.

[105] A. R. THOMSON, C. W. WOOD, A. J. BURTON, G. J. BARTLETT, R. B. SESSIONS, R. L. BRADY, AND D. N. WOOLFSON, *Computational design of water-soluble -helical barrels*, Science, 346 (2014), pp. 485–488.

[106] D. P. TIELEMAN, J. L. MACCALLUM, W. L. ASH, C. KANDT, Z. XU, AND L. MONTICELLI, *Membrane protein simulations with a united-atom lipid and all-atom protein model: lipid–protein interactions, side chain transfer free energies and model proteins*, Journal of Physics: Condensed Matter, 18 (2006), pp. S1221–S1234.

[107] G. M. TORRIE AND J. P. VALLEAU, *Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid*, Chemical Physics Letters, 28 (1974), pp. 578–581.

[108] M. D. TYKA, D. A. KEEDY, I. ANDRÉ, F. DIMAIO, Y. SONG, D. C. RICHARDSON, J. S. RICHARDSON, AND D. BAKER, *Alternate States of Proteins Revealed by Detailed Energy Landscape Mapping*, Journal of Molecular Biology, 405 (2011), pp. 607–618.

[109] C. VALÉRY, S. DEVILLE-FOILLARD, C. LEFEBVRE, N. TABERNER, P. LEGRAND, F. MENEAU, C. MERIADEC, C. DELVAUX, T. BIZIEN, E. KASOTAKIS, C. LOPEZ-IGLESIAS, A. GALL, S. BRESSANELLI, M.-H. LE DU, M. PATERNOSTRE, AND F. ARTZNER, *Atomic view of the histidine environment stabilizing higher-pH conformations of pH-dependent proteins*, Nature Communications, 6 (2015), p. 7771.

[110] S. VAN DER WALT, J. L. SCHÖNBERGER, J. NUNEZ-IGLESIAS, F. BOULOGNE, J. D. WARNER, N. YAGER, E. GOUILLART, AND T. YU, *scikit-image: image processing in Python*, PeerJ, 2 (2014), p. e453.

[111] W. VAN GUNSTEREN AND H. BERENDSEN, *Algorithms for macromolecular dynamics and constraint dynamics*, Molecular Physics, 34 (1977), pp. 1311–1327.

[112] L. VERLET, *Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules*, Physical Review, 159 (1967), pp. 98–103.

[113] R. VIANELLO, C. DOMENE, AND J. MAVRI, *The Use of Multiscale Molecular Simulations in Understanding a Relationship between the Structure and Function of Biological Systems of the Brain: The Application to Monoamine Oxidase Enzymes*, Frontiers in Neuroscience, 10 (2016), p. 327.

[114] J. WALSHAW AND D. N. WOOLFSON, *SOCKET: a program for identifying and analysing coiled-coil motifs within protein structures*, Journal of Molecular Biology, 307 (2001), pp. 1427–1450.

[115] R. F. S. WALTERS AND W. F. DEGRADO, *Helix-packing motifs in membrane proteins.*, Proceedings of the National Academy of Sciences of the United States of America, 103 (2006), pp. 13658–63.

[116] A. WARSHEL AND R. P. BORA, *Perspective: Defining and quantifying the role of dynamics in enzyme catalysis*, The Journal of Chemical Physics, 144 (2016), p. 180901.

[117] WATARU SHINODA, *, MASUHIRO MIKAMI, , TERUHIKO BABA, AND M. HATO‡, *Molecular Dynamics Study on the Effects of Chain Branching on the Physical Properties of Lipid Bilayers: 2. Permeability*, (2004).

[118] B. WEINER, N. WOETZEL, M. KARAKAŞ, N. ALEXANDER, AND J. MEILER, *BCL::MP-Fold: Folding Membrane Proteins through Assembly of Transmembrane Helices*, Structure, 21 (2013), pp. 1107–1117.

[119] S. J. WEINER, P. A. KOLLMAN, D. T. NGUYEN, AND D. A. CASE, *An all atom force field for simulations of proteins and nucleic acids*, Journal of Computational Chemistry, 7 (1986), pp. 230–252.

[120] E. W. WEISSTEIN, *Fast Fourier Transform*.

[121] . WILLIAM L. JORGENSEN, , DAVID S. MAXWELL, AND J. TIRADO-RIVES, *Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids*, (1996).

[122] M. G. WOLF, M. HOEFLING, C. APONTE-SANTAMARÍA, H. GRUBMÜLLER, AND G. GROENHOF, *g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation*, Journal of Computational Chemistry, 31 (2010), pp. 2169–2174.

[123] P. G. WOLYNES, *Symmetry and the energy landscapes of biomolecules*, Proceedings of the National Academy of Sciences, 93 (1996).

[124] C. W. WOOD, J. W. HEAL, A. R. THOMSON, G. J. BARTLETT, A. IBARRA, R. L. BRADY, R. B. SESSIONS, AND D. N. WOOLFSON, *ISAMBARD: an open-source computational environment for biomolecular analysis, modelling and design*, Bioinformatics, 33 (2017), pp. 3043–3050.

[125] L. V. WOODCOCK, *Glass transition in the hard-sphere model*, Journal of the Chemical Society, Faraday Transactions 2, 72 (1976), p. 1667.

[126] V. YAROV-YAROVOY, J. SCHONBRUN, AND D. BAKER, *Multipass membrane protein structure prediction using Rosetta*, Proteins: Structure, Function, and Bioinformatics, 62 (2005), pp. 1010–1025.

[127] G. YELLEN, *The voltage-gated potassium channels and their relatives*, Nature, 419 (2002), pp. 35–42.

[128] L. ZHANG, L. WANG, Y.-T. KAO, W. QIU, Y. YANG, O. OKOBIAH, AND D. ZHONG, *Mapping hydration dynamics around a protein surface.*, Proceedings of the National Academy of Sciences of the United States of America, 104 (2007), pp. 18461–6.

[129] S.-Q. ZHANG, D. KULP, C. SCHRAMM, M. MRAVIC, I. SAMISH, AND W. DEGRADO, *The Membrane- and Soluble-Protein Helix-Helix Interactome: Similar Geometry via Different Interactions*, Structure, 23 (2015), pp. 527–541.