CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL

Unidad Mérida

DEPARTAMENTO DE FÍSICA APLICADA

"Biomolecular simulations: from peptide design to ion channels"

a tesis submitted by:

Luz América Chi Uluac

In partial fulfillment of the requirements for the degree of

Doctor of Sciences

in

Physical Chemistry

Thesis Advisors:

Dr. María Cristina Vargas González

Mérida, Yucatán, México November 2020

CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL

Unidad Mérida

DEPARTAMENTO DE FÍSICA APLICADA

"Simulaciones biomoleculares: del diseño de péptidos a los canales iónicos"

TESIS

Que presenta

Luz América Chi Uluac

Para obtener el grado de

Doctora en Ciencias

en

Fisicoquímica

Director(es) de Tesis:

Dr. María Cristina Vargas González

Mérida, Yucatán, México Noviembre de 2020

Molecular Simulations

Biomolecular simulations: from peptide design to ion channels.

Luz America Chi Uluac

Luz America Chi Uluac

Biomolecular simulations: from peptide design to ion channels.

Molecular Simulations

Abstract

Resistin is a protein of biological interest due to its connection with several diseases of worldwide concern. The first part of this work aims to design a series of cyclic peptides as potential ligands to resistin. To this end, we propose an approach based on a peptide design algorithm plus a two-stage selection which accounts for selectivity. Following this approach, we are able to establish a methodological route for the development of strong candidates to target resistin.

Moving into higher molecular complexity, we study the Chloride Channel family proteins (ClCs), which function as either anion channels or anion/proton exchangers. The ClC exchangers have diverse abilities to transport anions. Specifically, in case of nitrate, such diverse ability has been related to the presence of a conserved serine (Ser) or proline (Pro) in the so-called "selectivity filter" motif, but the molecular basis for this fact is unknown. In the second part of this work, we used Molecular Dynamics (MD) classical simulations and Metadynamics (MTD) to evaluate the effect of replacement of serine by proline and chloride (Cl^-) by nitrate (NO_3^-) in the selectivity filter motif of the bacterial ClCec-1 homolog.

Resumen

La resistina es una proteína de interés biológico debido a su conexión con varias enfermedades de importancia mundial. El objetivo de la primera parte de este trabajo es diseñar una serie de péptidos cíclicos como potenciales moléculas de reconocimiento para la resistina. Para este fin, proponemos un enfoque basado en un algoritmo de diseño de péptidos, más una selección post-diseño de dos etapas, que toma en cuenta la selectividad. Siguiendo este enfoque, establecemos una ruta metodológica para el desarrollo de candidatos fuertes para el diseño de ligandos de la resistina.

Incrementando la complejidad molecular, estudiamos las proteínas de la familia del Canal de Cloruro (ClC), que funcionan como canales de aniones o intercambiadores de aniones/protones. Los intercambiadores de ClC tienen diversas habilidades para transportar aniones. Específicamente en el caso del nitrato, dicha capacidad diversa se ha relacionado con la presencia de una serina o prolina conservada en el llamado "filtro de selectividad". Dado que se desconocen las bases moleculares de este hecho, en la segunda parte de este trabajo, utilizamos simulaciones clásicas de Dinámica Molecular (MD) y Metadynamica (MTD) para evaluar el efecto de la sustitución de serina por prolina y cloruro por nitrato en el filtro selectivo del homólogo bacteriano ClCec-1.

Acknowledgement

I thank my supervisor Dra. María Cristina Vargas González for her support and guidance during all my PhD program. To my reviewers: Dra. Mercedes Alfonso Prieto, Dra. María Antonieta Fernández Herrera, Dr. Gabriel Pérez Ángel, Dr. Emanuel Hernández Núñez and Dr. Humberto Saint Martin Posada; for their time and constructive revisions of the thesis. I thank CONACYT for the financial support during these four years of study. Part of this work was carried out with resources provided by CINVESTAV in its ABACUS supercomputing clusters in Toluca and KUKULKAN in Mérida. I also thank Forschungszentrum Jülich for the resources provided, Dr. Paolo Carloni for his scientific support, and the DAAD program for the financial support during my research stay in Germany. I thank my family for its patience and love.

Contents

1 Introduction

> *Everything that living things do can be understood in terms of the jigglings and wigglings of atoms*
>
> **— Richard Feynman, 1963**

Biomolecular modeling is a fertile and growing area, with exciting opportunities and multiple promising applications. Nowadays, giant amounts of data are being provided by genomics, proteomics, glycomics and structural biology. The challenge for biomolecular modeling is to assist in efforts to draw on these diverse data to develop new drugs, therapies, catalysts, biologically based nanotechnology and to shed light into the never ending source of surprises that represents the biological world. Increasingly, computer simulations of biological macromolecules are helping to meet this challenge. In particular, molecular simulations are being employed more and more to investigate features not directly accessible to experiments. Whereas it is indeed possible to take "snapshots" of crystal structures and probe features of the motion of molecules through single crystal X-ray diffraction (SC-XRD), nuclear magnetic resonance (NMR) and cryo-electron microscopy (Cryo-EM), no current experimental technique allows access to all the time scales of motion with atomic resolution. Simulations based on fundamental physics offer the potential of filling-in these crucial "gaps", modeling how proteins and other biomolecules move, fluctuate, interact, react and function [1, 2]. Applications include studies of protein folding and conformational changes, association of proteins with small molecules or other proteins, structure-based drug design, computation of binding free energies for ligands, modeling the dynamics of ion channels and transport across membranes. Molecular recognition, which is the process of biological macromolecules interacting with each other or various small molecules (peptide, ion, etc.) with high specificity and affinity to form a specific complex, constitutes the basis of all processes in living organisms. Here we address two of those central topics: peptide design to target a biomarker and ion transport and selectivity.

2 Peptide design to target resistin

2.1 Introduction

Peptide design is essential in drug and biosensor design approaches. However, experimental procedures involved in their design and screening are extremely expensive and time-consuming taking into account the number of possibilities when varying the 20 natural amino acids in “De novo” design. Thus, in silico approaches are commonly used in the early stages of the design in order to save time and money. Laio et al. [3] were able to design short peptides with high affinity for organic molecules by combining Molecular Dynamics (MD), Monte Carlo (MC), and Molecular Docking methods. In this approach, the peptide sequences and structures were optimized to maximize the binding affinity to a target molecule by sampling simultaneously in conformation and sequence spaces, taking fully into account the flexibility of the peptide during the sequence optimization. Their designed cyclic peptides were capable to bind the anticancer drug irinotecan with experimental validation [4]. Ebrahimi et al. designed peptides with high affinity to Epidermal Growth Factor Receptor which plays an essential role in anticancer therapy [5]. Similar algorithms have been used to design linear and cyclic peptides for well-known model systems: maltose binding protein [6] and beta-2-microglobulin [7], reaching experimentally the low micromolar affinity in both cases. We now extend the reach of the method to design cyclic peptides with high affinity for a system of medical interest related to diabetes mellitus type 2 (DMT2). In this work (alredy published [8]), we present a protocol to design cyclic peptides of 12 amino acid residues with high affinity to the hormone resistin in trimer form. The algorithm is based on a combination of Replica Exchange Molecular Dynamics (RE-MD) and Replica Exchange Monte Carlo (RE-MC). Since this protocol has already been tested experimentally in several systems [3, 5, 6, 7, 9], we are confident that it might serve in identifying good “lead compounds” for further molecular investigations. After the design algorithm, we follow a selection procedure wich accounts for selectivity against albumin and includes a more rigorous calculation of binding energy based

on the Alchemical Absolute Binding Free Energy (AABFE) method. Finally, we added an interaction analysis of the best designed peptides.

Due to the fact that during drug or peptide design, ligands need to be ranked by their binding affinity to a target protein, different computational methods have been developed to assess it. Molecular docking calculations can quickly predict the most favorable pose of one ligand in a complex and assess the binding affinity using a scoring function; these calculations have proved to be good enough for the search stage of the design, although they could be inaccurate when compared with experiments [10]. More accurate estimations can be obtained through newer endpoint methods like Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) which are based on the post-processing in an implicit solvent of a number of frames extracted from a MD simulation in water. Both, docking and MMPBSA calculations, are good options if one is primarily interested in a relative ranking of affinities [11]. We emphasize that during the design process, we decided to use a scoring function since it allows us to make the repeated estimation of the peptide-protein affinity more efficient. Alchemical Absolute Binding Free Energy (AABFE) calculations are at the highest level of theoretical rigor but they have a high computational cost. AABFE calculations are based on a nonphysical thermodynamic cycle, where the binding free energy (BFE) is computed as the sum of multiple steps during which the ligand is "inserted" or "removed" from different environments, such as a bound or an unbound state [12]. Unlike relative free energy calculations where the binding of related ligands to a receptor are made, absolute free energy measures the free energy of binding of a single ligand to a single receptor. AABFE has been applied in a standard way to the affinity study of different systems, being the engineered binding pocket of T4 lysozyme one of the most studied macromolecular systems [13]. In this work, rigorous AABFE calculations were performed on two of the peptides selected after the design process.

The Resistin-Like Molecules (RELMs) are proteins rich in cysteine that are secreted in mammals. Resistin and RELMβ members of this family are highly conserved, especially in the cysteine rich C terminus [14]. The RELMs have been related to a wide variety of biological processes like breast cancer, liver diseases, lung diseases, atherosclerotic plaques, inflamed joints, kidney diseases, cardiovascular diseases, autoimmune diseases, malignancy, asthma, bowel diseases and DMT2, among others [15, 16]. The relationship of the RELMs with this great variety of pathological environments highlights the important physiological role they play. One of the first results that attracted attention towards the RELMs was the suggestion of resistin as a potential link between obesity and insulin resistance, both related to DMT2. This result was supported by the following facts in rodents: i) Resistin level is increased in

genetic and diet-induced obesity, ii) The inhibition or genetic suppression of resistin, produced an increase in insulin sensitivity and glucose homeostasis and iii) Inverse administration of exogenous or transgenic resistin, promoted insulin resistance [17]. In contrast to the role of resistin in rodents, the main function of this hormone in relation to obesity and insulin resistance in DMT2 is controversial in humans. Some studies have confirmed the association [18, 19], but others have challenged this link [20, 21]. The importance of starting *in silico* studies on resistin and its potential ligands is warranted by the necessity of standardizing and improving methods for measuring resistin concentration and eventually solve the controversy surrounding resistin molecular mechanism and its association to DMT2 in humans, the eventual search of therapeutic procedures to modulate resistin's levels, the poor knowledge about resistin's receptors and the high costs for obtaining and working on this protein. Summarizing, our principal purpose is the design of peptides to be potentially used in the detection of resistin as a biomarker of many diseases like DMT2.

2.2 Background and state of the art

2.2.1 Molecular Docking

Docking is a method which predicts the preferred orientation of one molecule (ligand) with respect to a second (protein) when forming a stable complex. As docking is typically used to screen extensive small-molecule chemical libraries, the pose generation and the pose quality evaluation must be carried out by fast methods i.e., the computational cost should be low. To fulfill this, several simplifications are needed in the overall docking process. The first simplification in docking is related to water, as this solvent is neglected by most docking programs. The second is that this method generally assumes the receptor rigid, the covalent lengths and angles constant, while considering a chosen set of covalent bonds freely rotatable in the ligand [22].

There are two independent stages in the docking process: the search algorithm and the scoring function. The first refers to the methods which are used to create different ligand and protein conformations and the aligning of different ligand conformations within the binding site of the protein. The latter, the scoring, is required for a quantitative estimation of the pose quality. The scoring function can be seen as an attempt to approximate the binding free energy of the system. The optimization algorithm attempts to find the conformation corresponding to the global minimum

of this scoring function and other low-scoring conformations, which are then ranked. In particular, the Autodock Vina program uses the Iterated Local Search global optimizer as a global optimization algorithm where a succession of steps consisting of a mutation and a local optimization (Broyden-Fletcher-Goldfarb-Shanno method) are taken, with each step being accepted according to the Metropolis criterion. The empirically-weighted scoring function used by Autodock Vina takes into account parameters like Van der Waals interactions, hydrogen bonds and torsion penalty and the predicted free energy of binding is calculated from the intermolecular part of the lowest-scoring conformation[23, 24].

2.2.2 Replica Exchange

The general idea of parallel tempering is to simulate several replicas of the original system of interest, each replica at a different temperature. The high temperature systems are generally able to sample large volumes of phase space, whereas low temperature systems, whilst having precise sampling in a local region of phase space, may become trapped in local energy minima during the timescale of a typical computer simulation. Parallel tempering achieves good sampling by allowing the systems at different (but close) temperatures to exchange complete configurations. Thus, the inclusion of higher temperature systems ensures that the lower temperature systems can access a representative set of low-temperature regions of phase space[25].

The general idea of parallel tempering is not limited to exchanges or swaps between systems at different temperatures. Investigators have developed a number of methods based on swapping alternative parameters in order to minimize barriers that inhibit correct sampling. Fukunishi et al. developed a Hamiltonian parallel tempering method that they applied to biomolecular systems; in their approach, only part of the interaction energy between particles is scaled between the different replicas[25].

2.2.3 RELMs

There exist four types of Resistin-Like Molecules (RELMs): i) resistin, ii) RELMα, iii) RELMβ and iv) RELMγ. The RELMs share a similar primary sequence, even between species share $\sim$60$\%$ of sequence identity. But, meanwhile in mice and rats, there have been found the four types of RELMs, only resistin and RELMβ have been found in humans [26]. The RELMs primary structure contains three domains: i) an N-terminal signal sequence, ii) a variable middle portion, and iii)

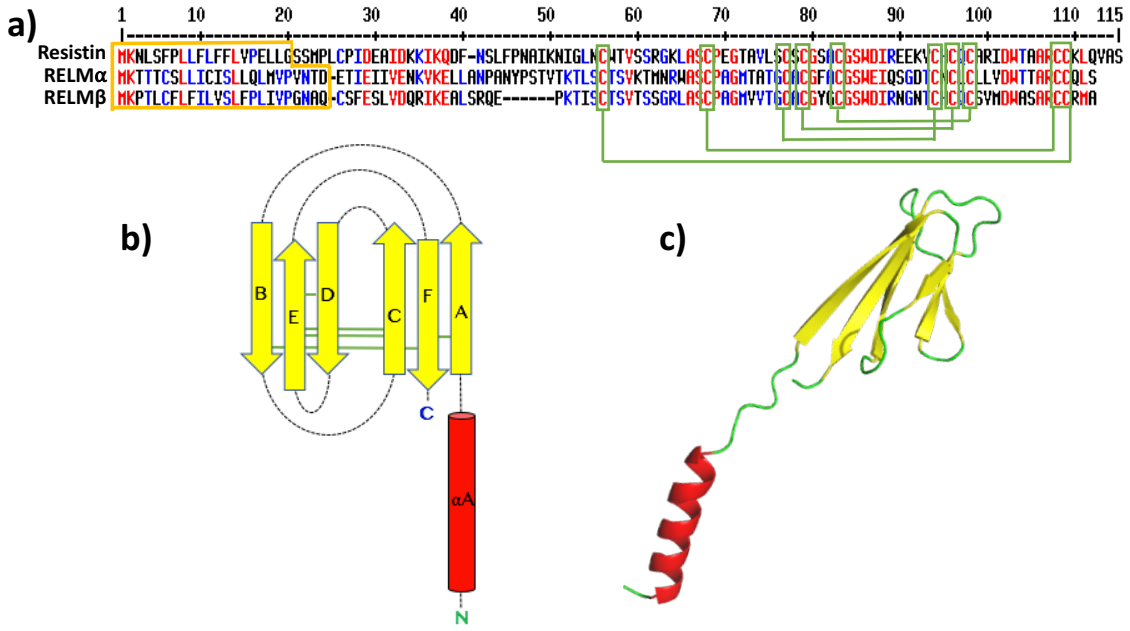


Fig. 2.1. – Structure of mouse RELMs. a) Primary sequence alignement of different RELMs (orange box represents the signal peptide, green box the conserved cysteines, green lines represents disulfide bonds and amino acids are represented with standard one letter code), b) folding topology (the yellow arrows represent beta-sheet secondary structures, red cilinders represent helical structures and green lines represents disulfide bonds) and c) tertiary structure.

a highly conserved C-terminal signature sequence (Cys-X_{11}-Cys-X_8-Cys-X-Cys-X_3-Cys-X_{10}-Cys-X-Cys-X-Cys-X_9-Cys-Cys-X_{3-6}), where X states for any other amino acid different from cysteine (Cys) and the subindices represent the number of amino acids between cysteines [14], this is displayed in Figure 2.1. This signature sequence has been proposed to be critical for disulfide bond formation and protein folding. The crystal structures of mouse resistin and RELMβ have been solved, revealing that they form trimers linked together via disulfide bonds to form hexameric assemblies. Dimerization of RELMβ and resistin was dependent on a cysteine in the N-terminal. This cysteine is lacking in RELMα and RELMγ, suggesting that they may exist as monomers, however their crystal structure has not been solved.

Regarding the RELMs receptors, tyrosine-protein kinase transmembrane receptor (ROR-1) [27] and decorin [28] are suggested receptors to mouse resistin. Toll-like receptor 4 (TLR4) [29], adenylyl cyclase-associated protein 1 (CAP1) [30] and human decorin [31] are suggested receptors to human resistin. In contrast with RELMβ that has identified receptor neither in humans nor in rodents [26]. Then, the information about binding domains in RELMs is poor as well. A study suggest that tryptophan amino acid in 82-position of chain A (Trp82A), Trp82C, serine amino acid

in 64 position of chain A (Ser64A) and Ser60C are key residues in the interaction between human resistin trimer and CAP1 SH3 domain model [30]. Another study agrees that TL4 receptor binds to resistin at the globular domain [29]. Therefore, a better understanding of the RELMs structure and their preferred interactions with other molecules may provide important information for identification of new receptors.

2.3 Objective

We aimed to design a series of cyclic peptides as potential ligands to resistin.

2.4 Methods

2.4.1 Peptide design

Binding sites

As a starting point, we chose a cyclic polyalanine as ligand, formed of ten continuous Alanines (Ala) and two Cys amino acids making a disulfide bridge between them. We chose two resistin putative binding sites, one located in the globular and highly conserved zone of the protein, Head (H) site, and the other in the helical part of the protein, Tail (T) site, both are shown in Figure 2.2. The H site was chosen using Raptor X [32], a protein structure prediction server, which predicts the binding sites of a protein given its sequence. Additionally, we found that the binding site predicted by Raptor X contained three of the major conserved amino acids among 6 species of resistin aligned (Appendix, section A.1) by CLUSTALW [33] tool. The T site was chosen because it was reported as a potential bioactive site in resistins which form hexamers through disulfide intertrimer bonds at the coiled-coil helical domain. Patel et al [34] found that a resistin mutant, lacking the intertrimer disulfide bonds, has more potent effects on hepatic insulin sensitivity than wild-type resistin. They suggest that a possible receptor binding geometry in which the coiled-coil tails are free, might be involved in receptor interactions - these tails are occluded in the hexamer form and free in the trimer form. Thus, it would be interesting to study a peptide that could inhibit this bioactivity.

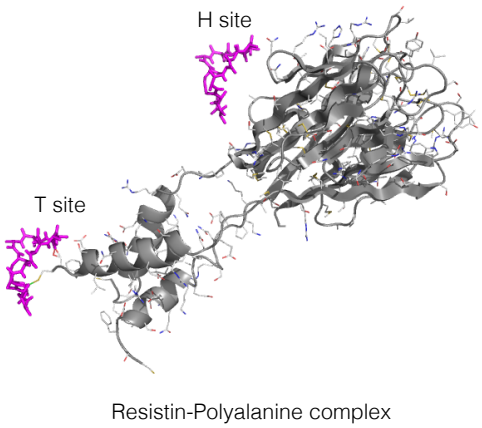


Fig. 2.2. – Polyalanine configuration in two resistin trimer binding sites: H site and T site.

Once the binding site has been selected, the optimization of the peptide sequence and configuration is the following step. This optimization procedure is based on Laio's et al algorithm [3, 7] that was adapted to resistin. The algorithm consists of several steps schematized in Figure 2.3, which are described below.

Preparation

In order to choose the best binding mode of the polyalanine ligand towards the protein binding site, a molecular docking is performed, using Autodock Vina [35]. The binding energy, Vina energy, is obtained using the scoring function of Autodock Vina and the mode of union corresponding to the lowest energy is always the chosen one.

The initial geometry obtained with docking was subjected to a conformation search (CS) process, which consists of: a relaxation by an MD energy minimization ($E_{min,1}$) in vacuum, followed by a replica exchange MD simulation in vacuum (RE-MD_1) for 8 different temperatures during 1 ns. Next, the replica's trajectories are put together and clustered by similar conformations ($Clust_1$). Then, a comparative process (Vina scoring) between peptide configurations according to its affinity to resistin is performed, in order to choose the configuration with the lowest energy as a starting point to begin the iterative process.

The design iterative process includes two stages: mutation and replica exchange. Mutation stage consists of: mutation, conformation search (CS) and mutation acceptance decision according to the Metropolis criterion (steps 3, 4 and 5 in Figure 2.3). The mutation stage is carried out simultaneously for three operating

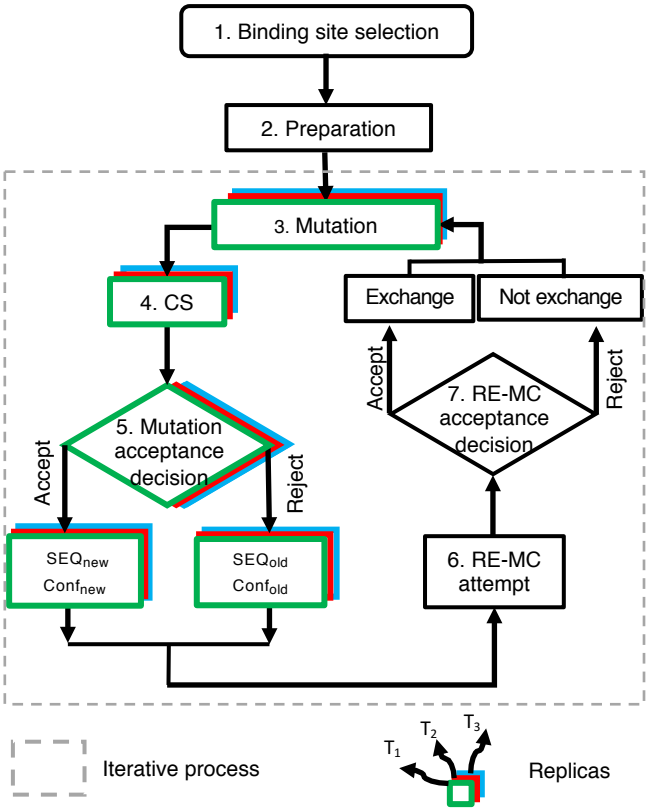


Fig. 2.3. – Steps of the peptide design process: 1) Binding site selection. 2) Preparation, that consists of molecular docking, $E_{min,1}$, RE-MD_1, $Clust_1$ and Vina scoring. Mutation stage: 3) Mutation step which consists of random amino acid selection replaced by another randomly chosen amino acid; 4) Conformation search ($E_{min,2}$, RE-MD_2, $Clust_2$, Vina scoring) and 5) Mutation acceptance decision. Replica exchange stage: 6) RE-MC attempt between the two simultaneous replicas randomly selected from replicas at T_1^f, T_2^f and T_3^f and 7) RE-MC acceptance decision.

temperatures. Replica exchange stage consists of: replica exchange attempt and replica exchange acceptance decision according to the Metropolis criterion (steps 6 and 7 in Figure 2.3).

Mutation

The mutation goal is to generate a new sequence of amino acids. For that, we randomly select one amino acid from the peptide (the cysteine amino acids of the peptide do not enter in the mutation process) and replace it with another one, also randomly selected from the 20 natural amino acids. It is important to point out that some hydrophilic amino acids ((Lysine, Lys), (Arginine, Arg), (Aspartic acid,

Asp), (Glutamic acid, Glu)) are assigned with high probability (0.11), to enhance the possibility of designing water-soluble peptides, other amino acids are assigned with a lower probability of 0.03.

Conformation search

The mutation step is followed by a conformation search (CS) process: $E_{min,2}$, RE-MD_2, $Clust_2$, Vina scoring and selection of the structure with the lowest energy. In this case, $E_{min,2}$ consists of:

i. Energy minimization including only the side chain of the mutated residue.

ii. Energy minimization including the mutated amino acid and nearest neighboring residues.

iii. Global energy minimization.

Mutation acceptance decision

The new peptide sequence, SEQ_{new}, is accepted or rejected according to the Metropolis criterion, with a probability of $\min(1, \exp[(E_{old} - E_{new})/T^f])$; where T^f is a fictitious "temperature" parameter that controls the probability of acceptance for the new configuration; E_{new} and E_{old} are the Vina energies of the new structure and the old one, respectively. The process continues, using the new structure in case of acceptance of the mutation, or the old one in case of rejection.

RE-MC attempt

To enhance the exploration of the sequence and conformation spaces, three independent calculations at T^f = 0.3, 0.6 and 0.9 kcal/mol are performed simultaneously during steps 3, 4 and 5 of the process. At the end of each mutation process, an attempt to swap sequences from two replicas randomly selected is performed. This exchange helps the process to move out of local minima.

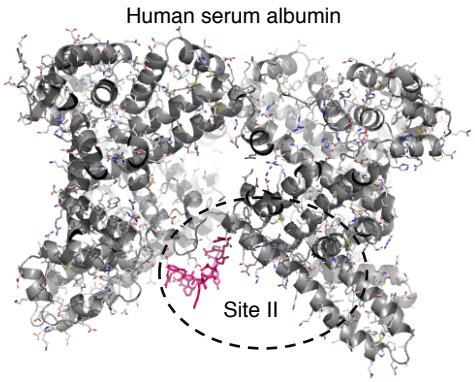


Fig. 2.4. – Albumin binding site II.

RE-MC acceptance decision

A swap between replica i at T_i^f, SEQ_i, E_i and replica j at T_j^f, SEQ_j, E_j are accepted with probability $\min(1, \exp[(E_i - E_j)(1/T_i^f - 1/T_j^f)])$.

In case of acceptance, the structure (sequence and configuration) of replica j and i are exchanged, in case of rejection, the structures remain the same. Regardless of whether the exchange of replicas is done, the iterative process begins again.

The entire design process was performed independently for the two binding sites selected. We report the results of 5 independent runs for each site, with 500 mutations per run. Performing more than 500 mutations does not appear to give better results because the system usually converges to a local minimum of energy.

2.4.2 Post-design peptide selection

By following the process described above, we designed a large number of peptides with very good affinities to resistin, the next step was to select those peptides that were also soluble and stable in water and selective to resistin when other proteins, such as albumin, were present. Albumin is the protein present in the blood in the highest concentration [36], it has binding sites in subdomains IIA and IIIA [37]; we chose site II to test peptide-albumin affinity as shown in Figure 2.4.

Peptide selection was made in two steps:

1. Selection of 20 peptides from each run containing the mutations with the highest affinity to resistin. A total of 200 peptides were chosen to be tested for the estimation of solubility in water and affinity to albumin. Vina energy of the peptide-albumin system was estimated after a molecular docking and a CS of the system, using a standard MD without replicas.

2. Selection of 2 peptides with good water solubility and low Vina affinity to albumin from each design run. We got 10 peptides for each binding site. From this set of 20 peptides (Appendix, section A.2), 4 were arbitrarily chosen to assess stability in water through a long MD simulation (150 ns) of the peptide-resistin systems.

During the long MD simulation in water, we analyzed the distance between the peptide and the resistin binding site ($D_{pep-res}$), the number of hydrogen bonds (N HBs) between the two molecules, the potential energy of peptide-resistin systems, Vina average energy over the last 100 ns of the simulation, gyration radius of the peptide ($R_{g,pep}$) or protein ($R_{g,res}$) in complexed state, the Root-Mean-Square Deviations (RMSD) of the peptide backbone atoms (in the complex state), the RMSD of the protein backbone atoms (in the complex) and the Root-Mean-Square Deviations of the interface atoms (IRMSD) between resistin and peptides. Based on the stability of the complexes in water according to the above criteria, two peptide-resistin systems were selected. The peptide-resistin affinity energy of the selected systems was estimated according to the following methods.

2.4.3 Binding affinity estimation: AABFE method

With the geometry of peptide-resistin obtained from the design algorithm, we performed an initialization process (IP) in water for each complex which consists of an MD energy minimization (EM), followed by a standard MD simulation in a canonical ensemble (MD_{NVT}), a standard isobaric-isothermic MD simulation (MD_{NPT}), a 2 ns unrestrained MD production run (MD_{prod}) and a clustering by conformations. A representative structure of the larger cluster was taken as the initial conformation for AABFE calculations.

AABFE calculations were performed using the thermodynamic cycle illustrated in Figure 2.5 [38, 39]. The initial state A represents the physical bound state while state F represents the physical unbound state. The free energy difference between the two states is what we consider as the binding free energy (BFE, ΔG_{bind}) of the complex. The direct calculation of ΔG_{bind} requires a simulation that starts in state F and then follows the binding process to state A. We can overcome the difficulties

with a suitably thermodynamic cycle of an alternative path that goes from state A through states B, C, D, E and finally to state F. Going from A to B means that an harmonic distance restraint between peptide and resistin is imposed in order to avoid the problem of the ligand leaving the binding pocket when protein-peptide interactions are being turned off in the next steps, giving ΔGrestr$_{on}$. From B to C, we turn off the electrostatic and Van der Waals (VdW) interactions between the peptide and its environment (water and resistin), giving ΔGcom$_{elec+VdW}$. States C and D are equivalent, i.e., resistin and peptide do not interact, giving ΔG$= 0$, but differ in the position of the ligand (inside or outside the binding site, respectively). Going from D to E means removing the ligand restraint imposed, giving ΔGrestr$_{off}$. Moving from state E to F represents turning on again the electrostatic and VdW interactions of the ligand with its environment (water), in order to obtain ΔGsolv$_{elec+VdW}$.

To calculate ΔG$_{bind}$, we sum all the free energy contributions, i.e.,

$$-(\Delta G_{bind}) = \Delta Grestr_{on} + \Delta Gcom_{elec+VdW} + \Delta Grestr_{off} + \Delta Gsolv_{elec+VdW}$$

Two types of simulations were carried out, some involving the complex, i.e., peptide-resistin (states A-C) and others only the ligand (E-F). To estimate ΔGrestr$_{on}$ and ΔGcomplex$_{elec+VdW}$ we performed complex simulations, to estimate ΔGrestr$_{off}$ we performed an analytical calculation with the expression described by Shirts [40], and for ΔGsolv$_{elec+VdW}$ we performed ligand simulations.

In order to have sufficient overlap in phase space between the physical states described previously, nonphysical or alchemical intermediate states were needed. As usual, each intermediate state is associated with a transition parameter [41] called λ. Each λ state consists of a unique set of λ values which control the different interaction types: electrostatic (λ_{coul}), Van der Waals (λ_{VdW}) and distance restraint (λ_{restr}). From A to B, 10 independently simulations were performed (States=0-9 in Table 2.1). From B to C, 20 independently simulations were carried out (States=10-29). From E to F, we used 20 separate simulations (States=10-29 in Table 2.1, without using λ_{restr} values). Lambda states and lambda values used are presented in Table 2.1.

For each simulation window or λ state, an IP was performed with 5 ns unrestrained production run for data collection with Hamiltonian Replica Exchange attempts between lambda states every 1000 steps [42]. From the simulations we got the potential energy difference ΔU$_{i,j}$ between states. In order to analyze the results we used Alchemical-Analysis.py script [41]. Through the variety of methods that

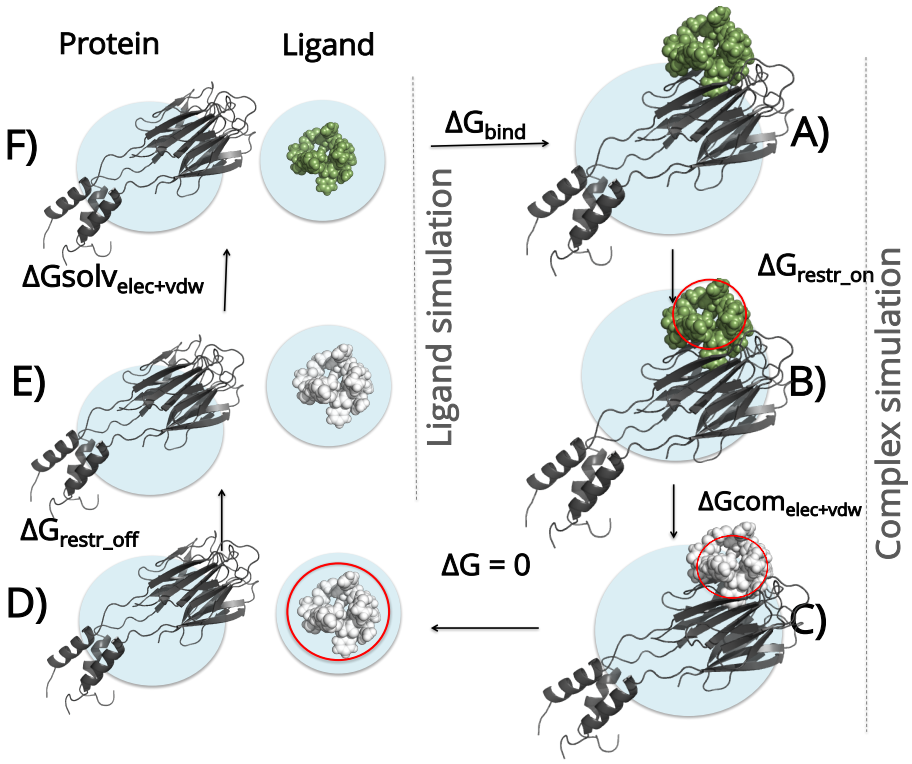


Fig. 2.5. – Thermodynamic cycle for BFE estimation in AABFE method. A) Bound state. B) Ligand restrained, indicated by a red circle and representing that the peptide is being confined to a certain volume. C) Ligand decoupled, represented in white because it is not interacting with its environment. D) Equivalent to state C. E) Turned off of the restraint. F) Ligand recoupled to water. Turquoise represents the water environment.

Tab. 2.1. – Lambda states (columns, numbered from 0 to 29) with their corresponding lambda values used during simulations. In λ_{restr}, 0 means non-restrained and 1 fully restrained. In λ_{coul} and λ_{VdW}, 0 means fully interacting and 1 non-interacting.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
λ_{restr}	0.0	0.01	0.025	0.05	0.075	0.1	0.2	0.35	0.5	0.75	1.0	1.00	1.00	1.00	1.0
λ_{coul}	0.0	0.00	0.000	0.00	0.000	0.0	0.0	0.00	0.0	0.00	0.0	0.25	0.50	0.75	1.0
λ_{VdW}	0.0	0.00	0.000	0.00	0.000	0.0	0.0	0.00	0.0	0.00	0.0	0.00	0.00	0.00	0.0
	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
λ_{restr}	1.00	1.0	1.0	1.0	1.0	1.0	1.0	1.00	1.0	1.00	1.0	1.00	1.0	1.00	1.0
λ_{coul}	1.00	1.0	1.0	1.0	1.0	1.0	1.0	1.00	1.0	1.00	1.0	1.00	1.0	1.00	1.0
λ_{VdW}	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1.0

Alchemical-Analysis.py script offers, we estimated free energy by Multistate Bennett Acceptance Ratio (MBAR) method using the last 2 ns of data. MBAR estimates ΔG by using the Zwanzing relationship [43], i.e.,

$$\Delta G_{ij} = -(1/\beta)\ln\langle\exp(-\beta\Delta U_{ij})\rangle_i$$

where ΔG_{ij} is the free energy between two adjacent states, β is the reciprocal of the temperature times Boltzmann constant, $\langle\rangle_i$ is an ensemble average over i. This considers that a free energy difference between two states can be computed via an appropriate exponential average of energy differences over an ensemble of configurations [41]. To avoid biased free energy estimates, Bennett Acceptance Ratio (BAR) method includes both forward ΔU_{ij} and reverse ΔU_{ji} potential energy differences between two adjacent states in the analysis, whereas the improved MBAR finds the best estimate of free energy changes between all states simultaneously by optimizing the matrix of the ΔG variances[41].

The uncertainty on ΔG was calculated as the root square of the square-sum of the uncertainty of the individual contributions, divided by the root square of the number of samples.

2.4.4 Per aminoacid contribution to the binding energy: MMPBSA estimation

Although AABFE estimations were found to be overall superior in obtaining correlation with experimental affinities for some systems [11], we used this method because it allowed us to insight about the amino acids contributions to the binding energy that are cumbersome to obtain with perturbation method, aditionally they are computationally cheaper and we already had the trajectories obtained in AABFE

calculations. Energy calculations with MMPBSA method were performed using g-mmpbsa package [44], a GROMACS tool for high-throughput MMPBSA calculations. MMPBSA calculations were performed using results from protein-ligand complex simulations in a single–trajectory (one-average) setup. Protein-ligand conformations were extracted from the MD simulations of the AABFE calculations in λ state = 0; the last 2 ns were used.

An extensive explanation of the terms involved in MMPBSA calculations can be found in Reference [44]. Briefly, MMPBSA approach considers three energetic terms for the free energy of binding estimation: $\Delta G_{bind} = G_{complex} - (G_{protein} + G_{ligand})$, where $G_{complex}$ is the protein-ligand free energy in solvent, $G_{protein}$ and G_{ligand} are the free energy of the isolated protein and ligand in solvent. Each of the three terms can be given by, $G_x = \langle E_{MM} \rangle - TS + \langle G_{solvation} \rangle$, where, $\langle E_{MM} \rangle$ is the average molecular mechanic potential energy in the vacuum. TS is the entropic contribution in the vacuum and $\langle G_{solvation} \rangle$ is the free energy of solvation. In this work, the entropy contribution was not included in the binding energy. E_{MM} is estimated based on molecular mechanics force field parameters, $E_{MM} = E_{bonded} + E_{nonbonded}$. In single trajectory approach, the conformation of protein and ligand in the bound and unbound forms are assumed to be identical, then ΔE_{bonded} is taken as zero. The solvation free energy is taken to be: $\Delta G_{solvation} = G_{polar} + G_{nonpolar}$; where G_{polar} and $G_{nonpolar}$ are the electrostatic and non-electrostatic contributions to the solvation free energy, respectively. G_{polar} is estimated by solving the Poisson-Boltzmann (PB) equation. $G_{nonpolar}$ includes repulsive and attractive forces between solute and solvent that are generated by cavity formation and van der Waals interactions and can be estimated using a variety of methods. SASA-SAV Non-polar Model used in this work, can be expressed as, $\Delta G_{nonpolar} = \gamma A + pV$, where γ is a coefficient related to surface tension of the solvent, A is the solvent accessible surface area (SASA) and V is the solvent accessible volume (SAV). In this method, the contributions due to the attractive solute-solvent interactions are not considered.

2.4.5 Computational details

The initial structure of the resistin was one trimer of the 1RH7 structure from PDB data base. Missing atoms in the crystal were modeled with Swiss PdbViewer and all molecules that were not in the protein were removed.

During molecular docking, we chose the coordinates of one of the atoms in the binding site as a center of a cubic docking grid.

All calculations: EM, MD_{NVT}, MD_{NPT} and MD_{prod} were performed using GROMACS with GPU support [45]. We used the Amber99SB-ILDN [46] force field for the protein and the ligand in all cases.

When simulations were in explicit solvent, we used TIP3P model for water molecules. The molecules were solvated in a box, with periodic boundary conditions and a minimum distance between the solute and the box of 1 nm, and sodium and chloride ions were added to neutralize the system.

EM steps were carried out using the steepest descent algorithm. In NVT ensemble simulations, harmonic position restraints were applied to the solute heavy atoms with a force constant of 1000 kJ $mol^{-1}nm^{-2}$. The MD production runs were carried out using a time step of 2 fs. Pressure coupling uses Berendesen [47] at 1 bar. The temperature was controlled by Langevin dynamics at 300K. Standard xyz periodic boundary conditions were adopted. A cut-off distance of 1.0 nm for Coulomb and Van der Waals neighbor list was updated according to the Verlet cut-off scheme. The long-range part of the Coulomb interactions was evaluated using PME [48] method with a relative tolerance of 10^{-5}, order 6 and Fourier spacing of 0.1. All bonds were constrained using LINCS [49], while SETTLE [50] was used for constraining the water molecules.

During RE-MD, the temperature values used were found through a tool called temperature generator for REMD simulations and they were 375, 382, 389, 396, 403, 410, 417 and 425 K [51]. In clustering, we used a cut-off in root-mean-square deviations (RMSD) range of 0.18 nm and the gromos method [52]. The Vina energy of the representative structure of each cluster with more than 10 elements was estimated. We considered the Vina error of 2.75 kcal/mol as reported in reference [35] and during the peptide design we use it only to score the relative quality of different ligands. We also performed the average of the Vina energy in long MD simulations using 25 frames extracted every 4 ns over the last 100 ns of the simulation.

During the mutation processes, we used Amber tools and tleap [53] for reconstruction of the mutated residue.

For the analysis of the distance between the peptide and the resistin binding site ($D_{pep-res}$) in the long MD simulations, we used the distance between one atom of the binding site and one atom of the peptide for complexes 453 and 203 (site H). In complexes 496 and 297 (site T), we used the distance between the center of mass of the three helices of the protein and the center of mass of the peptide. All the RMSD calculations where performed with respect of the first structure unless another

structure is specified. The IRMSD was calculated aligning firstly the trajectory to the backbone of the protein. The surface between resistin biding site and peptides was found using the InterfaceResidues.py script [54].

To reveal possible key residues contributing to binding free energy of peptide-resistin complexes, a free energy contribution per residue by using MMPBSA method was obtained. Energy calculations with MMPBSA method were performed using g-mmpbsa package [44], a GROMACS tool for high-throughput MMPBSA calculations. Protein-ligand conformations were extracted from the MD simulations of the AABFE calculations in λ state = 0; the last 2 ns were used. An extensive explanation of the terms involved in MMPBSA calculations can be found in the work of Kumari et al. [44].

In order to predict peptide solubility, we used "INNOVAGEN peptide solubility calculator" [55]. This web tool assigning each amino acid a numerical value (hydrophilicity value) based on experimental data [56].

2.5 Results

2.5.1 Peptide design

In general, the Vina energy of starting systems was estimated to be around -8 kcal/mol. During the design process, systems were built whose Vina energy reached -30 kcal/mol; this fact shows the power of the algorithm capable of achieving energy improvements of up to 22 kcal/mol compared to the initial system. Three of the optimizations runs for each binding site are shown in Figure 2.6, where the evolution of the Vina energy of the complex is presented. Vina energy of the systems changes a lot during the first 180 mutations, less drastic changes are observed afterward. When Vina energy does not change significantly, we consider that the algorithm has reached a local minimum in energy. Average Vina energy over the last 320 steps was calculated to be -24.12 ± 3.25, -24.00 ± 1.84 and -24.25 ± 4.51 kcal/mol, respectively for: all 10 runs, 5 of site H, and 5 of site T. The low values of Vina energy reached could imply that the binding sites were rightly chosen, given that the design procedure is strongly affected by the starting configuration according to Gladich's et al work [4].

The structure of the peptide-resistin systems with the maximum affinity after the design runs in H and T sites are presented in Figure 2.7a and Figure 2.7c, respectively. We can see an important N HBs presented in the structures, Figure 2.7b and

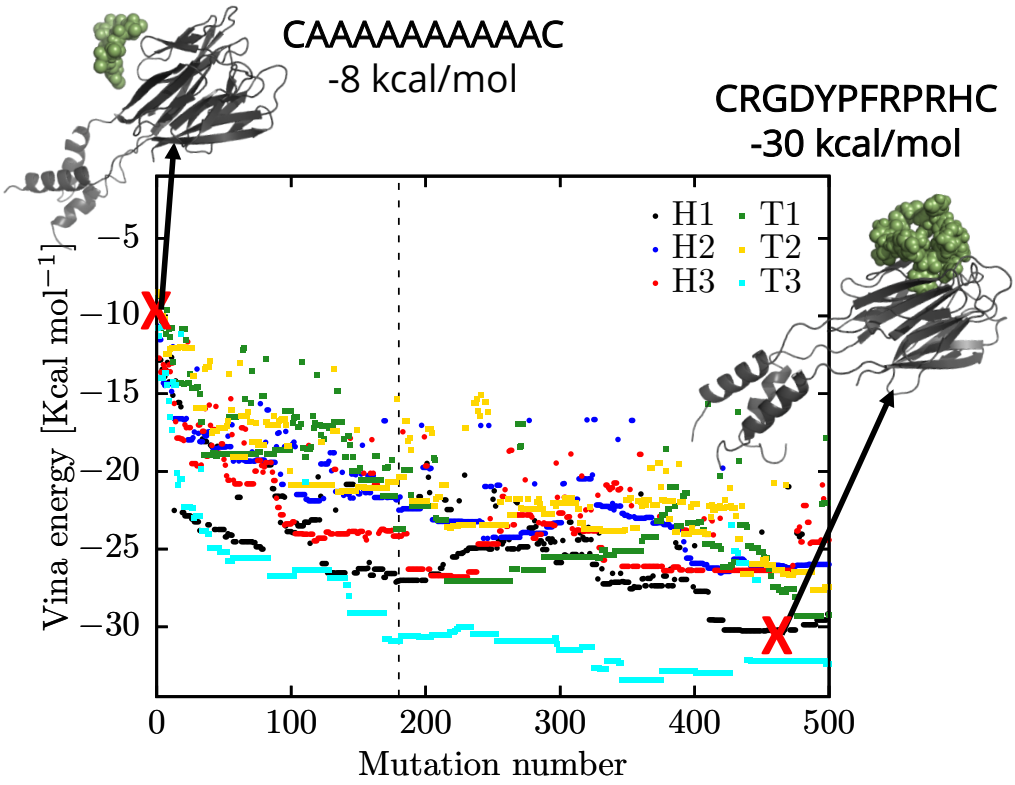


Fig. 2.6. – Evolution of the Vina energy for 6 runs, three for each binding site at operation temperature of T^f =0.3 kcal/mol. Legend indicates first the binding site and next the number of run.

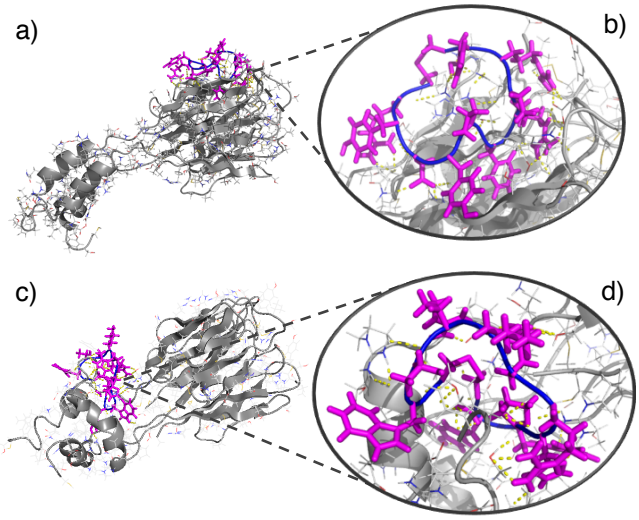


Fig. 2.7. – Complexes peptide-resistin. a) Structure of the peptide with the highest affinity to resistin in site H. b) Details of the HBs between the peptide and resistin in site H (complex 434). c) Structure of the peptide with the highest affinity to resistin in site T. d) Details of the HBs between the peptide and resistin in site T (complex 345).

Figure 2.7d. The peptide corresponding to the mutation 434 of one run on site H contains a hydrophobic phenylalanine highly buried inside the resistin globular core, Figure 2.7a and Figure 2.7b. The peptide corresponding to the mutation 345 in site T shows a conformation change in the helical tail of resistin, probably to expose less surface area to the solvent and stabilize the system, Figure 2.7c.

2.5.2 Post-design peptide selection

As a first selection phase, the only factor taken into account was the peptide affinity to resistin. We selected 100 peptides with high affinity to resistin for each binding site. In these sets of peptides, a multiple alignment of the sequences was performed. In site H, we obtained a consensus sequence of Cys-X-X-X-Arg-X-X-X-X-X-X-Cys in the 54% of the sequences. While in site T, we got a consensus sequence of Cys-Arg-X-X-Arg-X-X-X-X-X-X-Cys in the 48% of the sequences, where X states for any other amino acid different from cysteine (Cys) and arginine (Arg) amino acid. These patterns could be relevant to the new design of drugs or ligands for resistin.

In a second selection stage, we selected peptides with good water solubility, low affinity to albumin and not very similar sequences between them. We present an example set of 10 peptides from that selection in Table 2.2. As we can see, peptides from mutations 434 and 345 were the best binders in each binding site but also had

a very high affinity to albumin, so they were not considered further. Peptides from mutations 238 and 329 were discarded due to their poor water solubility. Only four of the 200 peptides had poor water solubility as expected due to the probabilities we set for amino acid selection in mutation process. When very similar sequences were found, such as the peptides of mutations 434 and 453, the one with the best other characteristics was selected; in this case, the low affinity to albumin was the discriminating factor.

Tab. 2.2. – Set of 10 peptides selected in the first phase. $E_{pep-res}$ is the Vina energy for peptide-resistin complexes in the design. $E_{pep-alb}$ is the Vina energy of peptide-albumin complexes after docking and MD simulation. All the Vina energy values are given in kcal/mol. WS is the water solubility.

Mutation	Site	Sequence	$E_{pep-res}$	$E_{pep-alb}$	WS
434	H	CRYDYPFRPRHC	-30.26	-28.04	good
453	H	CRGDYPFRPRHC	-30.22	-20.74	good
203	H	CQEWEPHFPWEC	-26.71	-15.78	good
492	H	CRRNREDEHRYC	-25.26	-19.58	good
238	H	CQKWNPQFPWEC	-24.66	-25.73	poor
345	T	CWTRKRRHPWYC	-33.42	-33.14	good
329	T	CWTRYRRHGWYC	-32.69	-33.94	poor
297	T	CSERWRRHGWYC	-31.11	-20.12	good
496	T	CMEKFKRKKEDC	-27.42	-16.43	good
456	T	CRRSRFHRWWRC	-26.72	-26.88	good

The final selection of four peptides is shown in Table 2.3: Vina energy of the peptides bound to resistin differs in at least 10 units from the Vina energy of the peptides bound to albumin. We selected sequences that did not share more than two amino acids in order to have variability.

Tab. 2.3. – Final selection of 4 peptides and their characteristics. $E_{pep-res,vac}$ is the Vina energy for peptide-resistin complexes in the design (vacuum). $E_{pep-alb,vac}$ is the Vina energy of peptide-albumin complexes after a docking and MD in vacuum. $E_{pep-res,sol}$ is the average Vina energy of the complexes from 150 ns of MD simulations in water. All the Vina energy values are given in kcal/mol.

Mutation	Sequence	$E_{pep-res,vac}$	$E_{pep-alb,vac}$	$E_{pep-res,sol}$	site
453	CRGDYPFRPRHC	-30.22	-20.74	-17.40 $\pm$ 2.1	H
203	CQEWEPHFPWEC	-26.71	-15.78	-17.80 $\pm$ 1.4	H
496	CMEKFKRKKEDC	-27.42	-16.43	-8.19 $\pm$ 1.3	T
297	CSERWRRHGWYC	-31.11	-20.12	-14.65 $\pm$ 3.1	T

For each of the four peptides selected, we performed a long MD simulation of complexes in water (150 ns), in order to select the two most stably bound peptides to perform free energy calculations. Through this long MD simulation, we analyzed a series of parameters: RMSD, $D_{pep-res}$, N HBs, $R_{g,pep}$, $R_{g,res}$, potential energy and

Vina average energy of peptide-resistin systems (as mentioned at the end of section 2.4.2).

The $D_{pep-res}$ becomes quite stable during the long MD simulation, as showed in Figure 2.8a. These results suggest that complexes remain bound up 150 ns of simulation. Complex 453 had the lowest fluctuation throughout the simulation, with $D_{pep-res}$ of about 0.9 nm. In contrast, complex 496 had the highest fluctuation in $D_{pep-res}$, showing a significant $D_{pep-res}$ change in the first 25 ns, after that it stabilizes around 1.8 nm. Complexes 297 and 203 showed little fluctuation around 0.5 and 1.2 nm, respectively. Peptide 453 was the one closest to the binding site compared to the other peptides. Peptides designed for H site showed the shortest $D_{pep-res}$ value.

Formation of inter-molecular hydrogen bonds is closely coupled to binding/unbinding events. In Figure 2.8b we present the variation of the N HBs during the long MD simulation. The N HBs oscillates between 8 and 20 at the beginning of the simulation and they reach stabilization after about 75 ns of MD. In complex 496, the N HBs decreases to 5 during the initial change in $D_{pep-res}$ observed in Figure 2.8-a. Peptides 203 and 453 designed for the site H, exhibited the least fluctuation in the N HBs formed with the resistin, while complex 297 maintains the highest N HBs during the simulation. These results were consistent with the $D_{pep-res}$ analysis.

To characterize the compactness of protein and ligands in all the complexes, the radius of gyration (R_g) of each system was calculated and the results are shown in Figure 2.8c and Figure 2.8d for protein and ligands, respectively. From Figure 2.8c, the fluctuations of resistin in the different systems are quite small around 2.2 nm, which means that the global protein structure in the complexes is well preserved and stable during simulations. From Figure 2.8d, the fluctuations of peptides in the complex are quite small around 0.6 nm, being the 496 and 203 peptides the ones that fluctuate the most. We selected systems containing peptides 453 and 297 although they were not the structures with the smallest value of R_g.

The plot of potential energy of the complexes through long MD simulation showed stabilization since the beginning of the simulation with small relative energy fluctuation, Figure 2.9a. This result makes us conclude that the structure is stable. Complexes 453 and 297 showed the lowest potential energy.

Protein backbone RMSD values were estimated with respect to a reference structure in the equilibrated stage of the simulation (frame at 40 ns). It can be seen in Figure 2.9b that proteins in the complexed state remain relatively stable in the last 100 ns with RMSD around 0.6 nm. In Figure 2.9c we can see that the RMSD of

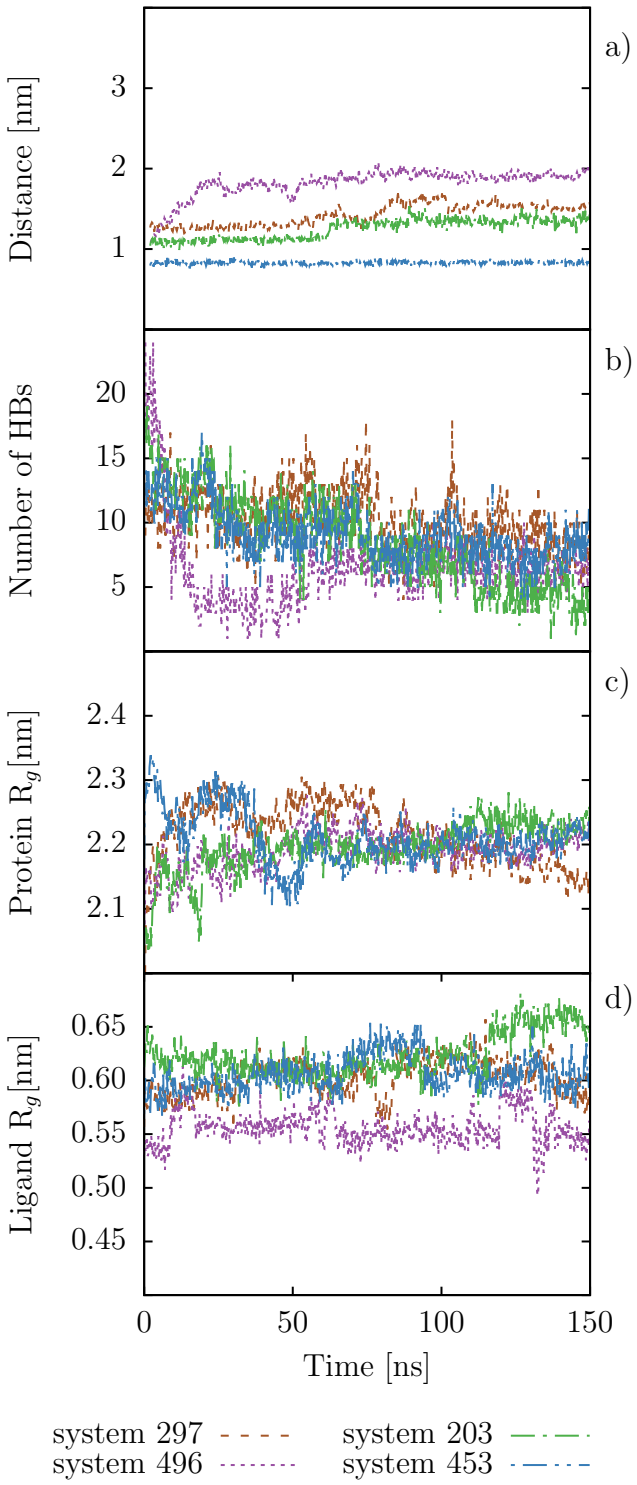


Fig. 2.8. – Time-evolution of a) distance between peptide and resistin binding site($D_{pep-res}$), b) number of Hydrogen Bonds between protein and ligand, c) gyration radius of the protein in the complex and d) gyration radius of the ligand in the complex.

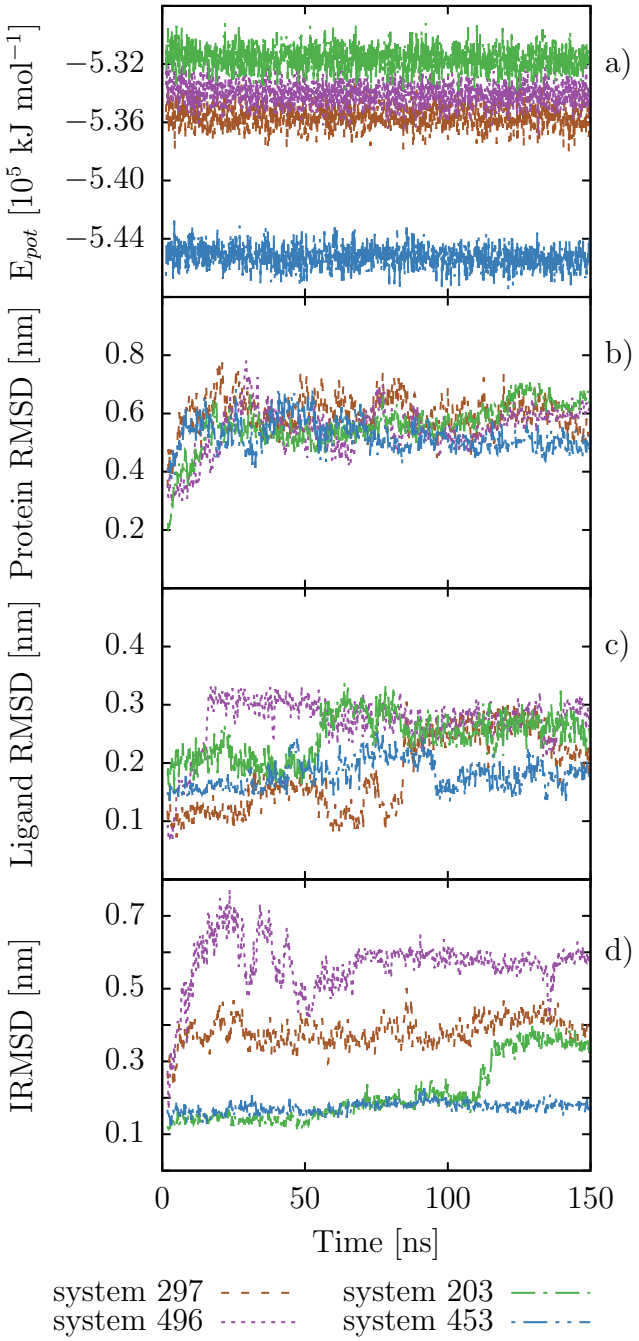


Fig. 2.9. – Time-evolution of a) potential energy of the complexes and b) protein RMSD in complex, c) ligand RMSD in complex and d) IRMSD between protein and ligand.

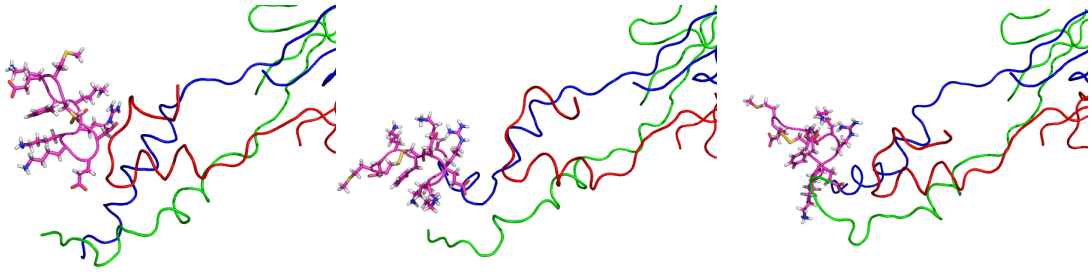


Fig. 2.10. – Frames extracted from the long MD simulation of complex 496. Chain A is presented in red, Chain B in blue and Chain C in green.

ligands remains around 0.2 nm, being the 453 complex the one that fluctuates the least. In Figure 2.9d we can see that the IRMSD between protein and ligand in complex 453 remain around 0.15 nm and fluctuates the least. Complex 496 had the biggest IRMSD and it fluctuates the most. This shows that the 453 complex has a very stable interaction surface.

Due to the fact that the peptides were designed in vacuum, we had to assess their affinity while solvated. The average Vina energy of the solvated complex during the last 100 ns of MD in water was estimated and compared with the Vina energy of the complex in vacuum (Table 2.3). There was a significant difference in the solvated complexes and the complexes in vacuum. This difference was more evident in system 496.

Peptide 496 seems to have the worst characteristics but we have found that during the simulation it moves from one of the helices to the other as showed in Figure 2.10 which is a very interesting behavior that should be explored in future work.

Based on all the previous analysis of the simulation, we selected complexes 453 (site H) and 297 (site T) for further studies.

2.5.3 Binding affinity estimation

Having selected complexes 453 and 297, we proceeded to calculate the binding free energy (BFE). We performed 5 ns of simulations but used only the last 2 ns for data collection. The convergence behavior of the BFE is shown in Figure 2.11, where ΔG_{bind} is calculated according to equation 2.4.3 and averaged over 3 independent AABFE calculating simulations; plotted every 100 ps for each complex. It can be seen in Figure 2.11 that 5 ns of simulations yield good convergence of AABFE calculations, however, this estimation can be improved using longer sampling and more lambdas.

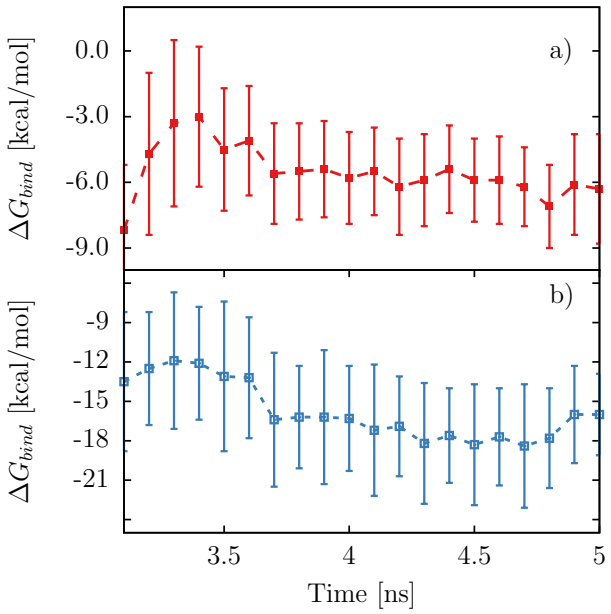


Fig. 2.11. – Calculated BFE vs time of simulation. a) complex 297 and b) complex 453.

The AABFE calculations confirm that peptide 453 has good affinity to resistin, while peptide 297 does not seem to be a good option. The ΔG_{bind} values obtained for complexes 297 and 453 were -6.26 $\pm$ 2.3 and -16.21 $\pm$ 3.03 kcal/mol, respectively. These results show that AABFE calculations could be important discriminators to select and confirm peptides for further study.

To compare the RMSD histogram distributions of the protein for the bound (complexed) and unbound (isolated) conformation ensembles visited during 50 ns of simulations in water, we present Figure 2.12. The RMSD distribution for isolated protein and proteins in complexed state spans from 0.3 to 0.9 nm with two peaks in the distribution, one between 0.5 and 0.6 nm and the other between 0.6 and 0.7 nm. The distributions overlap significantly during the 50 ns of simulation. These results confirmed the structural stability of the protein and give confidence in the results of AABFE calculations. In order to understand the interactions between receptor and ligands, we built distance matrices consisting of the smallest distance between residue pairs over the 5 ns simulations of AABFE calculations. Then, we linked such contact maps images to the residues by using the program PyMOL[57] to see how peptides keep in close contact with several residues of protein. Residues 1-243 belong to protein and residues 244-255 belong to the peptides. In complex 453, we

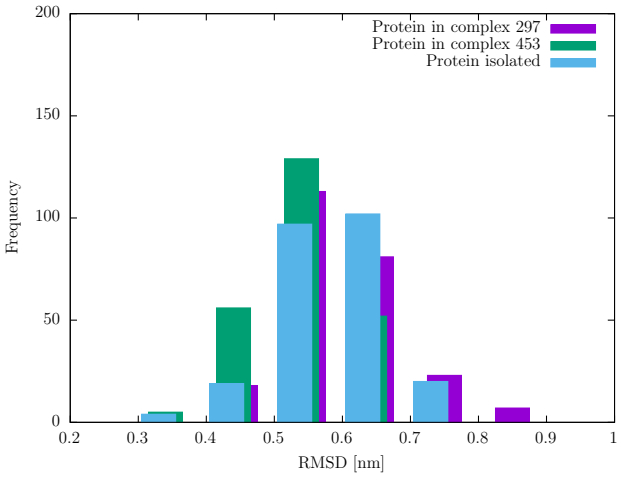


Fig. 2.12. – Histograms of the backbone RMSD of the Resistin conformation obtained from 50 ns of MD simulations of the protein in isolated and in complexed states for the systems 297 and 453.

found the closest average contacts between residues Thr29-Phe250, Asp71-Pro252, Trp136-Arg251, Ala47-Phe250 and Arg139-Asp247, showed in Figure 2.13a. In complex 297, the closest average contacts were found between residues Cys82-Trp253, Pro101-Tyr254, Gln18-Tyr253 and Cys176-Tyr254, showed in Figure 2.13b.

To reveal key residues contributing to binding free energy of peptide-resistin complexes, a free energy contribution per residue can be calculated by means of the MMPBSA method. Total MMPBSA BFE values, -35 $\pm$ 4.2 and -4 $\pm$ 0.16 kcal/mol for complex 453 and 297 respectively, agree with results from AABFE at confirming that peptide 453 has good affinity to resistin while peptide 297 does not seem to be a good option. Figure 2.14 shows that for complex 297, residues Ala243, Trp253, Tyr254 and Cys255 contribute favorably with more than 15 kcal/mol to the total binding free energy of complex 297, and residues Cys82 and Arg179 are unfavorable to the binding. The polar Cys82 shows a significant polar solvation energy (results not shown), which is unfavorable to ligand binding according to Reference [58]. For complex 453, residues Asp137, Arg139, Phe250 and Cys255 contribute more than 15 kcal/mol to the total binding free energy of complex 453, and residues Cys244, Arg253 and His254 contribute unfavorably to the binding. A comparison of the contribution per each amino acid to the BFE shows that more protein amino acids contribute stronger in complex 453 than in complex 297 (Appendix, section A.3). This could imply that the binding site is more favorable for complex 453 than for complex 297. We also analyzed the different types of interaction that contribute to total binding energy (Appendix, section A.3), and found that polar

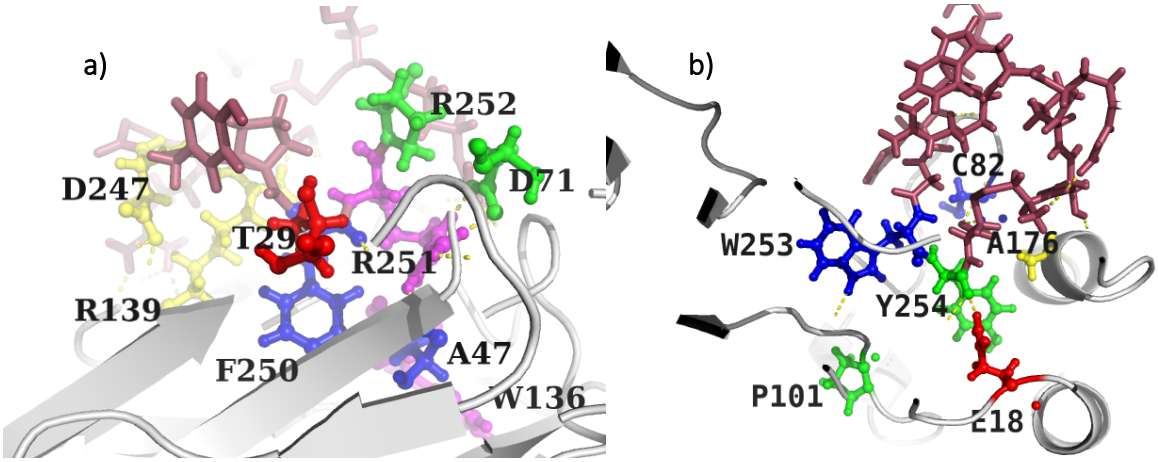


Fig. 2.13. – a) close contacts in complex 453, b) close contacts in complex 297. (protein in cartoon, the peptide in sticks, relevant residues for the association are represented in CPK representation and residues are presented in one-letter code).

solvation contributes unfavorably to binding and more strongly affects complex 297 (Appendix, section A.3).

A 2D interaction map of the average structure of the peptide-resistin complexes over the 5 ns of the MD trajectory was performed by using the program MOE [59]. In complex 297, Cys82 shows a reduction of solvent exposure induced by the ligand as we can see in Figure 2.15. Glu181- Cys244, Glu181-Arg247, Leu177-Ser245, Cys82-Cys255 and Gln18-Tyr254 form H-bonds. In complex 453, Arg114-Cys255, Gln147-Cys255, Asn140-Gly246, Arg76-Gly246, Ser74-Phe250, Asp71-Pro249 form H-bonds. Tyr50-Arg253 presents interaction between aromatic ring of Tyr and cations of Arg as we can see in Figure 2.15. We also counted the number and type of interactions (HBs, the cation-arene interactions and the number of amino acids with reduced solvent exposure due to the ligand presence; SI, Table 7) and found more of them in complex 453.

According to the distance map, the per residue free energy contribution and 2D interaction maps: Tyr 254, Gln18 and Cys 255 are key residues in the binding of complex 297; Arg139, Phe250 and Cys255 are key residues in the binding of complex 453.

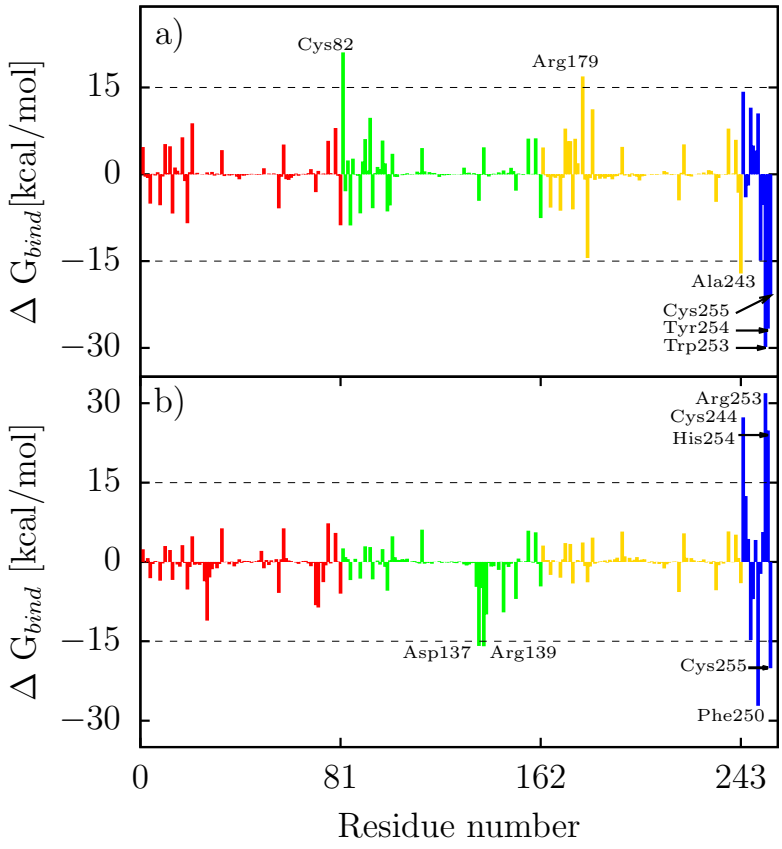


Fig. 2.14. – Contribution to the BFE per residue by MMPBSA method. a) interaction spectrum of complex 297 and b) interaction spectrum of complex 453. Residues 1-243 belong to the protein (chain A, red; chain B, green; chain C, yellow) and residues 244-255 belong to the peptide (blue).

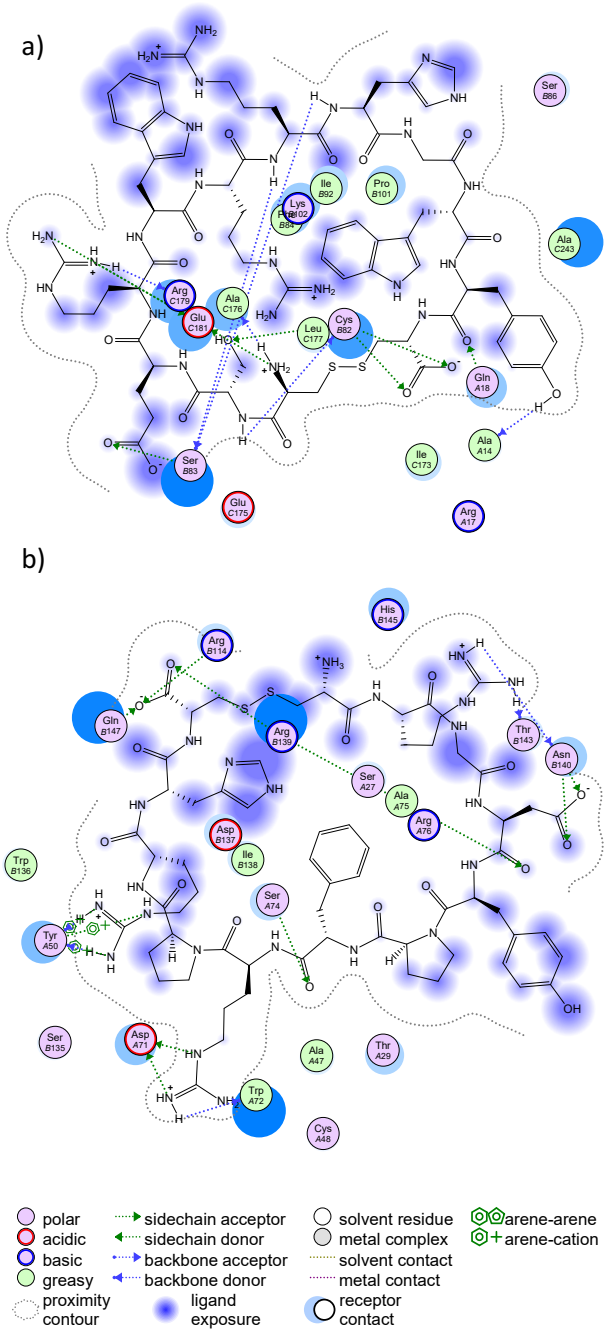


Fig. 2.15. – 2D diagram of interactions between resistin and peptides. a) Complex 297 and b) complex 453.

2.6 Conclusion

In this work, we were able to design cyclic dodecapeptides with high affinity to resistin in two binding sites, whose Vina energy reached up to -30 kcal/mol. We identified consensus sequences in peptides in at least 48% of the total sequences with high affinity to resistin. In site H, we obtained a consensus sequence of Cys-x-x-x-Arg-x-x-x-x-x-x-Cys, while in site T, we got a consensus sequence of Cys-Arg-x-x-Arg-x-x-x-x-x-x-Cys. This pattern could be relevant for the new design of drugs or ligands to resistin. However, peptides with this pattern were not further studied in this work because others peptides fulfilled better the demands imposed after the design.

We selected a subset of 4 peptides with good water solubility, low peptide-albumin affinity, good water stability over 150 ns of MD simulations and average binding energy in water up to -14 kcal/mol, excepting the peptide 496 that showed less average energy, but presented an interesting behavior that merits a special study: it moves from one of the helices to other during a long MD simulation. This behavior could affect the receptor binding geometry that promotes the resistin bioactivity, so it could be important to look for peptides with this kind of behavior and further study them.

The number of possible mutations when using 20 essential amino acids to design relatively small peptides is enormous. The design algorithm used in this work offers a rational pathway that shortens the number of possible options to explore experimentally. Furthermore, the set of ligands designed can be refined taking into account their water solubility and stability, their low selectivity for molecules like albumin and their affinity for resistin when solvated in water.

During the design process the affinity was assessed with Vina. In order to confirm the good affinity in the systems, binding affinity with AABFE method was estimated in complex 453 and complex 297. Ligand 453 showed good affinity while ligand 297 showed just a modest affinity. Due to the fact that AABFE is considered one of the methods with the highest level of theoretical rigor, it could give us more reliable information to select the best candidates among those obtained from the algorithm; thus, peptide 297 does not seem to be as good as peptide 453. It is important to explore with BFE calculations other candidates in order to have a good collection of peptides with good affinity, i.e., similar to that of peptide 453. Regarding the interactions, Arg139, Phe250 and Cys255 were key residues in the binding of peptide 453 to resistin.

We were able to establish a methodological route for the development of strong candidates as peptide 453 for the design of potential ligands to resistin. We suggest that peptide 453 should be further evaluated using suitable in vitro and/or in vivo assays. This study depicts a rational design of probable leads that can be used for other target molecules.

3 Firsts insights into the NO_3^- over Cl^- selectivity in ClC-ec1 transporter and its (S→P) mutant

3.1 Introduction

The Chloride Channel family proteins (ClCs) are found in all domains of life from bacteria to humans and either function as anion/proton exchangers (active transport) or as chloride channels (passive transport). ClC channels form part of plasma membranes of cells, playing a role determining the membrane potential in muscle fibers, balancing salt transport in kidney, inner ear, and other epithelial tissues [60]. ClC exchangers are localized in membranes of cellular compartments (endosomes, lysosomes, plant vacuoles, and synaptic vesicles), playing a key role in acidification [60], facilitating endocytosis [61] and lysosomal function [62]. When chloride transport fails due to mutations in ClCs genes, diverse disorders arise, including osteopetrosis, neurodegeneration, leukodystrophy, deafness, blindness, myotonia, hyperaldosteronism, renal salt loss, proteinuria, kidney stones, male infertility, and mental epilepsy [63]. This great variety of disorders highlights the physiological importance of this kind of proteins.

Transporters dissipate the electrochemical gradient of one substrate (anion/proton) to accumulate the other with substrate binding or dissociation being the primary driving force for turnover [64]. Secondary active transporters are called antiporters when anion/proton cotransport is in different directions. Actually, some intracellular mammalian ClCs (ClC-3 to ClC-7), one plant (AtClCa) and one bacterial homologue (ClC-ec1) are antiporters [63]. ClC-ec1 (Cl^-/H^+ exchanger from E. coli) is the most studied prototypical ClC transporter [65]. The ClC proteins like ClC-ec1 have a homodimeric structure and each monomer unit is able to carry out Cl^-/H^+ exchange independently from the other subunit [66]. ClC monomer contains three Cl^- binding sites (Figure 3.1): internal, central and external binding sites (S_{in}, S_{cen} and S_{ex}).

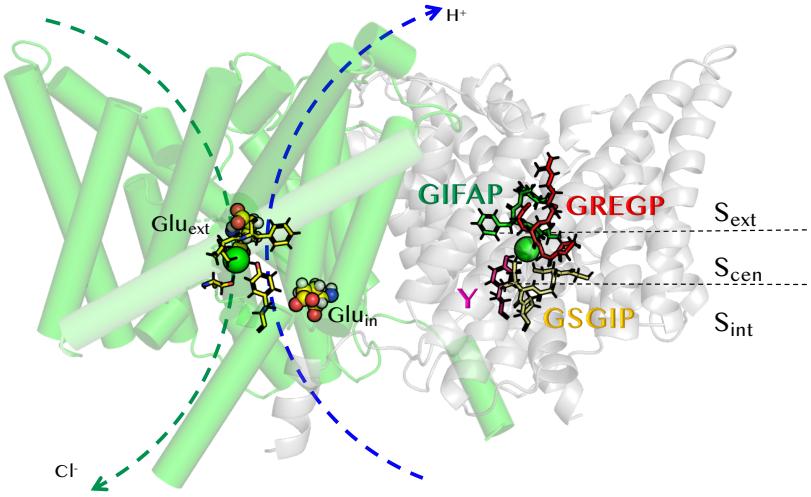


Fig. 3.1. – ClC-ec1 dimer. Anion/proton permeation pathways (left) and the ion binding sites formed by four conserved motifs (right): GIFAP (355-359), GSGIP (106-110) and GREGP (146-150), where amino acids are showed in standard one letter nomenclature and numbering is according ClC-ec1 PDB structure.

One exciting topic into the study of ion channels and transporters is ion selectivity [67]. Ion selectivity can have different meanings, depending on the molecular system that is being considered (ion channel or a transporter) and on the experimental conditions that are used to probe the system (equilibrium binding or nonequilibrium flux and ionic current measurements). In the case of a transporter, it is reasonable to expect that ion binding selectivity will be primarily governed thermodynamically by equilibrium binding, because the lifetime of the various conformational states of the protein is extremely long. In the case of ion channels, selectivity can mean different things, although ultimately it is understood that the wrong kind of ions should encounter more difficulty than the right kind of ions to permeate. Then, could be instructive to carry out two kind of studies, classical MD in order to study the equilibrium behaviour of the systems and metadynamics simulations (MTD), one computational technique that is sensitive for barriers and wells on the free energy landscape. The MTD technique involves enhanced sampling over collective variables using a biased potential to force the system to leave local minima and thus sample low-probability states in complex systems [68]. MTD was already used to estimate the conductance of different ion channels [69] [70] and the results were in satisfactory agreement with experimental observations. It has been applied to study the gating mechanism in the ClC Chloride Channel (*salmonella* homolog) [71] but no to study selectivity in ClCs.

The ClCs channels and the bacterial antiporter ClC-ec1 share an ion selectivity sequence of $SCN^- > Cl^- > Br^- > NO_3^- > F^- > I^-$ [72]. Hence, most ClC members transport chloride ions (Cl^-) that is one of the most abundant anions in extracellular and intracellular fluids [73]. An exception is the AtClCa plant protein that displays a preference for exchanging nitrate over chloride with a selectivity sequence of $NO_3^- \sim I^- > Br^- > Cl^- > SO_4^{2-} >$ glutamate$^-$. Indeed, nitrate (NO_3^-) is one of the main available forms of nitrogen for plants since they can not use nitrogen directly from the air. This is a fundamental step for plant biology; understanding molecular mechanisms implicated in nitrate transport is also a key step in improving nitrogen efficiency for sustainable agriculture [74].

Intriguingly, only one substitution (S107P) in the bacterial protein ClC-ec1 is able to change the protein from Cl^-/H^+ to a NO_3^-/H^+ exchanger [64]. This mutation is located at the central ion-binding S_{cen} site. An elegant molecular dynamics simulation study [75] has shown that, for the wild-type protein, nitrate displays two distinct binding coordinations at the S_{cen} site and is hydrated in a different manner compared to chloride, i.e., nitrate forms pseudo-water-wires instead of continuous water wires needed to support the connection of the two proton transfer sites. This results provide important structural details of nitrate binding in ClC-ec1.

However, the molecular basis of the effect of the S→P substitution on selectivity has not been given. Here we addressed more details related with this substitution by comparing between wild-type (WT) ClC-ec1 protein and mutated (S107P) ClC-ec1 protein with both ions (Cl^- and NO_3^-). Hereinafter we call the four systems WT-Cl^-, S107P-Cl^-, WT-NO_3^- and S107P-NO_3^- for WT ClC-ec1 with Cl^-, (S107P) ClC-ec1 mutant with Cl^-, WT ClC-ec1 with NO_3^- and (S107P) mutant ClC-ec1 with NO_3^-), respectively.

3.2 Background and State of the art

3.2.1 Selectivity

One of the first theories of ionic selectivity was provided by Eisenman [76]. Selectivity was explained as originating from the difference between the hydration free energy of the ion and the energy of the interaction between the ion and a charged/polar binding site within the channel. This theory predicted the existence of 11 selectivity sequences. Unfortunately, it has not been posible to identify charged or polar group responsible for the selectivity of Na^+ and K^+ monovalent cations. Later,

Laio and coworkers [77] showed that the selectivity properties of such monovalent cation channels can originate from geometrical properties of the inner core of the channel without any critical contribution from electrostatic interactions between the permeating ions and charged of polar groups. These cases make it clear that the study of selectivity in ionic channels or even transporters is still in its infancy but looking at the electrostatic or geometrical aspects could be a good starting point.

There are some general rules about selectivity in channels: a) the height of the energy barrier an ion needs to overcome will determine the ease for an ion to get into a channel, b) the energy barriers determine the selectivity sequences as one ion is subsituted by another with higher barriers for the least permeant ion, c) the stronger the attraction to the binding site, the deeper the well, the deeper the well, the slower are the rate constants for leaving it, and the slower the overall permeation process. Nevertheless there could be some paradoxical behaviours [78].

Experimentally, the selectivity of transport proteins can be measured by different methods like comparing the current amplitude in different ionic conditions or from comparing permeability ratios. Permeability ratio is taken from the reversal potential (zero net current) under bi-ionic conditions. Computational studies are able to simulate this situation and explain the origin of the reversal potential in an asymmetric ionic concentration with near-quantitative accuracy for OmpF porin and hemolysin. However, measurements of reversal potentials become impractical for highly selective channels such as the K^+ channels. In this case, alternate methods such as Ba^{2+} blockade relief or Na^+ punchthrough were preferred. Ba^{2+} block is more sensitive to the depth of the free energy minima of the binding sites (i.e., equilibrium binding), whereas Na^+ punchthrough is more sensitive to the height of free energy barriers (i.e., nonequilibrium rates) [67]. Fortunately, as mentioned before, MTD is a computational technique that is sensitive for barriers and wells on the free energy landscape.

3.2.2 Cl^- transport pathway

The Cl^- binding region exists in three different conformations which differ in the position of Glu_{ext}. In ClC-ec1 protein (wild type, WT), Glu_{ext} occupies S_{ext} (middle position) occluding the ion conduction pathway towards the outside (Figure 3.2). The E148Q mutant of ClC-ec1 (presumably equivalent to the protonated E148) represents the open state of the channel where Glu_{ext} moves out of the pathway (up position). A third state is only found in the structure of cmClC (algae homolog),

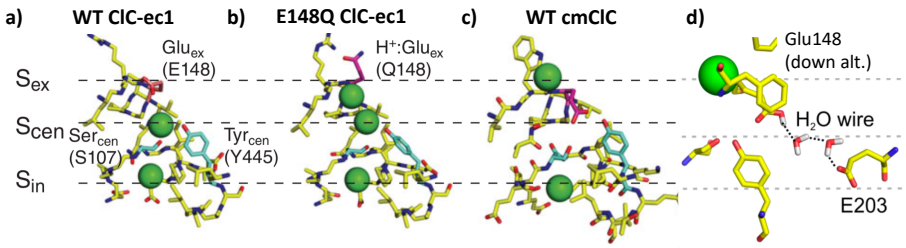


Fig. 3.2. – Glu148 positions. Views of the conformations of the ion binding region of a) WT ClC-ec1 (middle), b) E148Q ClC-ec1 (up), c)WT cmClC (down) and d) WT ClC-ec1 extracted from PMF calculation (down alt.). Slightly adapted from [72] and [79].

where the Glu_{ext} is in S_{cen} (down position) [79]. Recently, structural studies provide evidence that shows Glu_{ext} in a structural intermediate position (middle down position) [80]. Also, a combined molecular dynamics with biochemical and electrophysiological measurements suggested that Glu_{ext} does not move through the pore (down alternative position) [79, 81].

The ClC monomer contains three Cl^- binding sites (Figure 3.1): internal, central and external binding sites (S_{in}, S_{cen} and S_{ex}). S_{ext} is defined by the backbone nitrogen atoms of residues Phe357, Ala358, Arg147, Glu148, and Gly149. S_{cen} is identified by the backbone nitrogen atoms of Ile356 and Phe357 and by the side-chain oxygen atoms of Ser107 and Tyr445. S_{int} is defined by the backbone nitrogen atoms of Ser107 and Glu108. Such residues are relatively well conserved, for example, $Sser_{cen}$ (S107) and Glu_{ex} (E148) are found in 42% and 88% of the ClC sequences, respectively [72].

3.2.3 H^+Cl^- coupling

Different models of the transport coupling in ClC exchangers has been proposed. A first model (Figure 3.3, Model I) was suggested by Feng and MacKinnon [82]. This model consists of the following phases: an intracellular proton reaches the negatively charged side chain of Glu_{ex}, which is placed at S_{cen} (I–II); Glu_{ex} moves and the Cl^- at S_{ext} leaves its binding site towards the outside (II–III); two Cl^- enter from the extracellular side (III–IV); Glu_{ex} is deprotonated (IV–V); the side chain of Glu_{ex} moves from S_{ext} to S_{cen}, pushing the Cl^- bound in S_{cen} to S_{int} and the ion bound in S_{int} into the cytoplasm (V-VI) and an extracellular Cl^- rapidly binds into S_{ext} (VI–I). Steps V and VI correspond to the crystal structures of EcClC-1 E148Q mutant of EcClC-1 and Wild Type EcClC-1, respectively. This model also called the

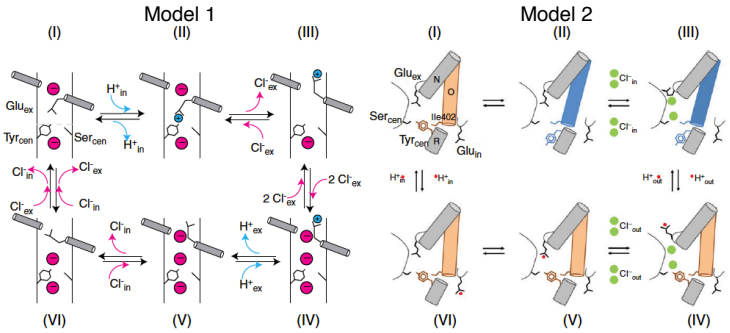


Fig. 3.3. – Models for the transport cycle of the ClC exchangers. Taken from [72]

kinetic gate, incorporates only the external gating glutamate as a gating element and does not involve conformational changes associated with transport.

A second model was suggested by Basilio et al [83] (Figure 3.3, Model II). It consists of the following phases: the protein in an occluded state opens the inner gate involving conformational changes of helix O (I–II); two Cl^- enter from the intracellular side displacing Glu_{ex} outside (II–III); Glu_{ex} is protonated from the outside and the inner gate closes (III–IV); protonated Glu_{ex} re-enters to the ion transport pathway (IV–V); the proton is transferred from Glu_{ex} to the intracellular solution via the internal glutamate (Glu203 in ClC-ec1) (V–VI) and transporter returns to its initial configuration (VI–I). This second model also called steric gates model, has been proposed to account for the finding that Tyr_{cen} act as the intracellular gate of the ClC exchangers and that a conformational change of helix O is required for transport.

Khantwal et al. [84] proposed an alternative mechanism (Figure 3.4) that highlights the relevance of global structural changes in ClC function. This mechanism is in agreement with the second model, except in steps IV-VI, which involve protein conformational changes. After protonation of Glu_{ex}, the helices N and P rearrange (IV) to open the extracellular vestibule and to allow the exit of the bound chloride anions (IV-V). Then, once the protonated Glu_{ex} occupies S_{cen} (state V), helices N and P rearrange again, closing the extracellular vestibule so that the proton transfer between Glu_{ex} and Glu_{in} is not exposed to either the extracellular or the intracellular media.

A modified version of the first model has been proposed too [63]. This version eliminates the modulatory Cl^- binding/unbinding step (II-III) by introducing a hypothetical transition during which the protonated side chain of Glu_{ex} and the Cl^- bound in S_{ext} swap their position and the side chain of the gating glutamate moves further out. This model partially remediates the inconsistency in the first model that

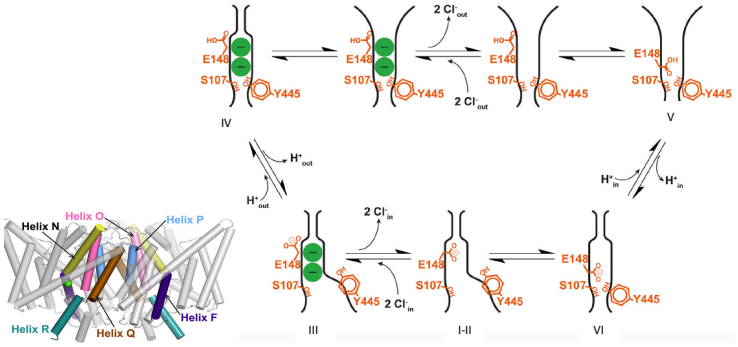


Fig. 3.4. – Helices names in ClCs (left) and Khantwal et al. mechanism for Cl ion transport (right). Taken from [84]

predicts an uncoupling of chloride proton exchange at acidic pH values. It states that the nonconductive state II is more highly populated at acidic pH. However, the proposed swap of the protonated glutamate side chain with a Cl^- is purely hypothetical and requires experimental verification.

Implicitly, it is assumed that the phenomenon of ion permeation and selectivity through an open conductive pore can be discussed independently from gating and inactivation. In reality, the situation is more complicated. For instance, the filter of some K^+ channels is unable to remain in a conductive conformation when exposed to an ionic solution with little or no K^+ ions. In case of ClCec1 protein, it has been suggested that gating and simultaneous occupancy of binding sites regulate affinity but not selectivity [64].

3.2.4 What we know about NO_3^- coupling and transport on ClCs

There are different behaviors against nitrate transport across the domains of life. The nitrate transport mechanism of the bacterial antiporter ClC-ec1 is partially uncoupled to protons. The human chloride transporter ClC-5 is more conductive for nitrate than for Cl^- but its nitrate transport mechanism is uncoupled to protons. In plants, AtClCa is able to transport nitrate more efficiently than Cl^- and to catalyze nitrate transport coupled to protons. AtClCa protein is localized at the tonoplast and functions as an antiporter, transporting $2NO_3^-$ versus $1H^+$ [85].

Although the molecular mechanism of nitrate exchange remains unknown, computational studies on bacterial ClC-ec1 transporter revealed that nitrate in the central anion-binding site S_{cen}, crucial for the chloride-coupled H^+ transport, has two binding modes (Modes I and II) in comparison with the single binding mode for Cl^-

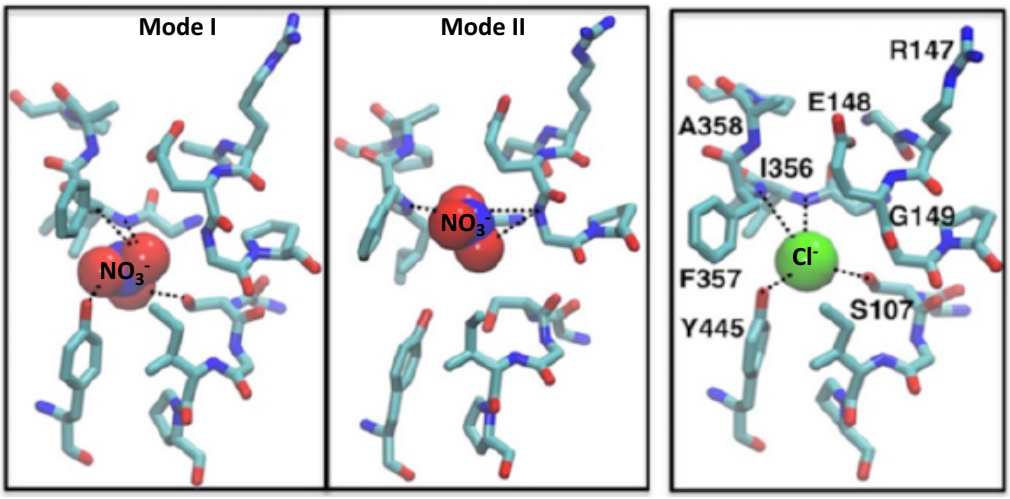


Fig. 3.5. – Anion binding modes. Two binding modes for nitrate (left) and one for chloride (right) Taken from [75]

[75] (Figure 3.5). In Mode I, nitrate is fitted into S_{cen} with one of its oxygen atoms coordinated by the backbone nitrogen atoms of I356 and F357, and the other two oxygen atoms coordinated by the side chain hydroxyl groups of S107 and Y445, respectively. Mode I is relatively more populated and is observed in both subunits. In Mode II, nitrate moves upward to a position between S_{cen} and E148. This binding mode involves hydrogen bond interactions between the oxygen atoms of NO_3^- and the backbone nitrogen atom of G149, instead of direct interaction with S107 and Y445 hydroxyl groups. However, binding Mode II only occurs in subunit I during 150 ns, before NO_3^- moves back to the original, S_{cen} binding site.

An interesting behavior has been observed in water that surrounds the ions [75] (Figure 3.6). In the case of Cl^- bound to S_{cen}, it stabilizes the formation of a wire of water molecules between Glu_{ext} and Glu_{in} after or without a conformational change of Glu_{in}. Nitrate, in the same site, leads instead to the formation of pseudo-water-wires odifferent from those formed with Cl^- anions. These different water rearrangements could potentially play a key role for the experimentally observed selective coupling between chloride/nitrate and protons.

On the other hand, it has been proposed that ClC selectivity is primarily determined by interactions of side chains in a well conserved filter motif (Figure 3.7) with the permeating ions. Specifically, a serine residue S_{cen} which participates in the central Cl^- binding site was shown to control the Cl^-/NO_3^- selectivity of ClC channels and transporters. In its conserved filter motif, AtClCa contains a proline (P160), while human ClC-5 and bacterial ClC-ec1 contain a serine (S168 and S107, respectively).

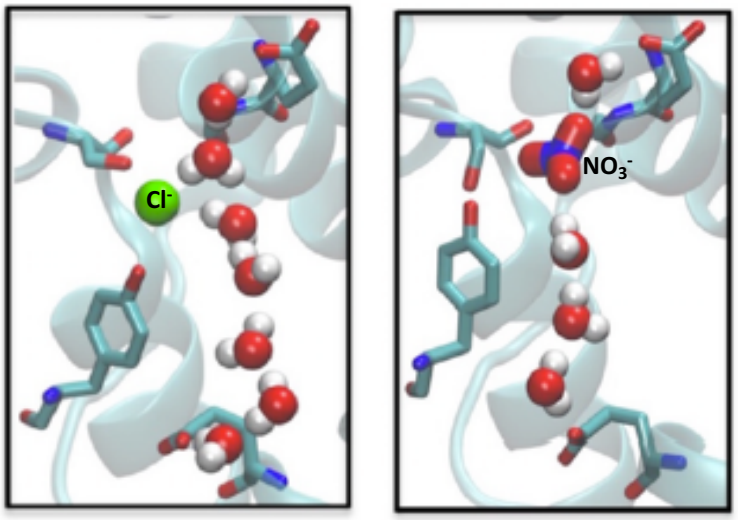


Fig. 3.6. – Effect of anion bound at S_{cen} on formation of water wires. Taken from [75]

```
AtClC-a 157 AAGPGIPEIKAYLN
AtClC-b 156 AAGPGIPEIKAYLN
AtClC-c 166 AAGSGIPEVKAYLN
AtClC-d 146 AAGSGIPEIKGYLN
ecClC-1 104 AGGSGIPEIEGALE
ClC-0   120 AVGSGIPELKTIIR
ClC-1   186 AVGSGIPEMKTILR
ClC-4   178 ACGSGIPEIKTILS
ClC-5   165 ACGSGIPEIKTILS
```

Fig. 3.7. – Sequence alignment of several ClC proteins in the region of the GXGIPE filter motif. Taken from [86]

Recent studies show that human ClC-5 mutant (S→P) enables coupling nitrate and proton transport, a (S→P) substitution confers NO_3^- selectivity upon the Cl^- specific ClC-ec1 transporter and ClC-0 channel [64], while, AtClCa mutant (P→S) loses nitrate selectivity.

Another important experimental result has shown that stronger binders (Cl^-) support tight proton coupling while weaker binders (NO_3^-) degrade coupling. It was also found that the binding selectivity of the E148A mutant is the same as WT, showing that E148 regulate affinity but not selectivity[64].

Recent studies in ClC-0 and ClC-K channels [87] suggests that the pore-lining backbone amides are the general determinants of inter-anion selectivity in ClCs and that side chains (Ser_{cen}) contribute to selectivity in only a subset of ClCs. All the

previous information highlights the necessity of studies to understand the mechanism of nitrate transport in ClCs and its mechanism of selectivity.

3.2.5 Metadynamics

The well known problem of time scale in MD simulations has given rise to the development of a group of methods known as "enhancing sampling methods". Enhanced sampling methods aim to accelerate sampling in simulations, allowing to explore processes that will otherwise not be tractable in reasonable computer times [88]. There is a subgroup of enhanced sampling methods where sampling is facilitated by the introduction of an aditional bias potential (or force) that acts on a selected number of degrees of freedom or collective variables (CVs), often referred to as "collective variables-based methods" [89]. They include the widely used umbrella sampling [90] and metadynamics [68]. In both methods, besides enhancing sampling, it is also possible to reconstruct the free energy surface (FES) as a function of the chosen CVs. But, at variance with umbrella sampling, in metadynamics no *a priori* knowledge of the landscape is required.

In metadynamics, an external history-dependent bias potential V_{bias}, which is a function of the CVs, is added to the Hamiltonian of the system H, i.e.,

$$H(x) = K(x) + V(x) + V_{bias}(x,t)$$

where x are configurational variables or the so called CVs, t is the time, K is the kinetic energy, V the potential energy and d the dimensionality or the number of collective variables. The bias potential can be written as a sum of gaussians deposited along the system trajectory in the CVs' space to discourage the system from revisiting configurations that have already been sampled, i.e.,

$$V_{bias}(x,t) = \int_0^t w \exp(-\sum_{i=1}^{d} \frac{(x_i(t) - x_i(t'))^2}{2\sigma_i^2}) dt'$$

where w is the gaussian height and σ_i is the width of the gaussian for the ith CV. After a transient time, the accumulated bias potential provides an unbiased estimate of the underlying free energy $F(x)$, i.e.,

$$F(x) \sim -V_{bias}(x, t \to \infty)$$

The general process is ilustrated in Figure 3.8.

Metadynamics presents two major drawbacks: i) in a single run, the bias potential does not converge to a static function that shows oscillations. This results in a possible overfilling of the underlying FES, and gives rise to the non trivial dilemma of when to stop a simulation; and ii) identifying a set of CVs appropriate for describing complex processes is far from trivial. The first issue has been overcome by developing well-tempered metadynamics, where the bias converges to its limiting value in a single run, thus avoiding overfilling, and it is possible to control the region of the FES that are physically meaningful to explore. [91]. Another improvement has been done by deriving a time-independent and locally convergent free energy estimator for metadynamics [92]. In case of the CVs choice, there is no a priori recipe for finding a suitable set of CVs, and in many cases it is necesary to proceed by trial and error, attempting several metadynamics simulations with different combinations of variables. Ideally the CVs should satisfy three properties: i) they should clearly distinguish between the initial state, the final state and the intermediates, ii) they should describe all the slow events that are relevant to the process of interest and iii) their number should not be too large, otherwise it will take a very long time to fill the free energy surface. Examples of CVs are distances, dihedral angles, coordination numbers, potential energy and box shape, between others [93].

The time interval between the addition of two gaussian functions, as well as the gaussian height and gaussian width, are tuned to optimize the ratio between accuracy and computational cost. By simply changing the size of the gaussian, metadynamics can be fitted to yield very quickly a rough map of the energy landscape by using large gaussians, or can be used to get a finer grained description by using smaller gaussians. Usually, the well-tempered metadynamics is used to change the gaussian size adaptively [91], i.e., the height of the gaussian is decreased with simulation time according to:

$$w = w_0 \exp(-\frac{V_{bias}}{k_B \Delta T})$$

where w_0 is an initial gaussian height, ΔT an input parameter with the dimension of a temperature, and k_B the Boltzmann constant. With this rescaling of the gaussian height, the bias potential smoothly converges in the long time limit, but it does not fully compensate the underlying free energy:

$$F(x) = -\frac{T + \Delta T}{\Delta T} V_{bias}(x, t \to \infty) + C$$

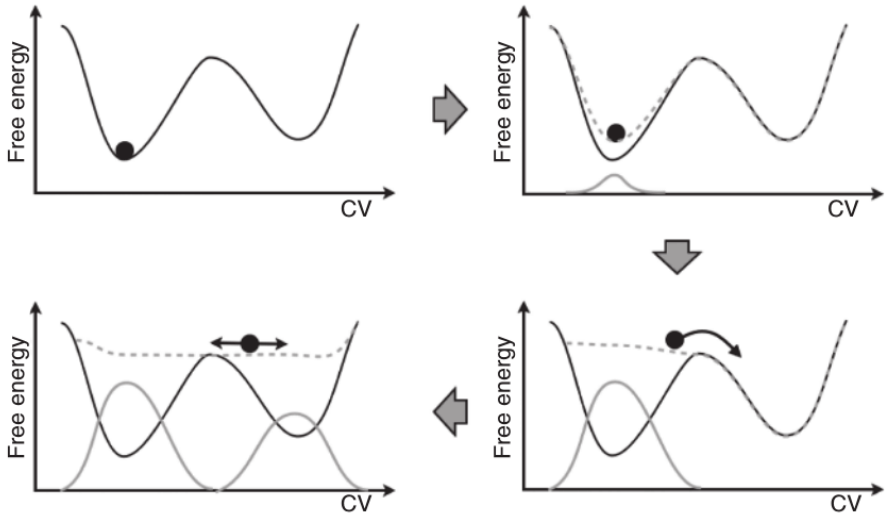


Fig. 3.8. – Representation of the metadynamics process. First the system evolves according to a normal dynamics, then a gaussian potential is deposited (solid gray line). This lifts the system and modifies the free energy landscape (dashed gray line) in which the system evolves. After a while the sum of Gaussian potentials fills up the first metastable state and the system moves into the second metastable basin. After this the second metastable basin is filled, the system evolves in a flat landscape. The summation of the deposited bias (solid gray profile) provides a first rough negative estimate of the free energy profile. Credits for the figure [94].

where T is the temperature of the system. In the long time limit, the CVs thus sample an ensemble at a temperature $T + \Delta T$ which is higher than the system temperature T. The parameter ΔT can be chosen to regulate the extent of free-energy exploration: $\Delta T = 0$ corresponds to standard molecular dynamics, $\Delta T \to \infty$ to standard metadynamics. In well-tempered metadynamics, the term "bias factor" is the ratio between the temperature of the CVs ($T + \Delta T$) and the system temperature (T), i.e.,

$$\gamma = \frac{T + \Delta T}{T}$$

The bias factor should thus be carefully chosen in order for the relevant free-energy barriers to be crossed efficiently in the time scale of the simulation.

3.3 Objective

- To provide insight on how ClC-ec1 protein and (S→P) mutant can manage to transport highly diverse anions such as Cl^- and nitrate.

3.4 Methods

The first part of our work is devoted to the Classical Molecular dynamics simulations of four systems: WT-Cl^-, WT-NO_3^-, S107P-Cl^-, and S107P-NO_3^-; corresponding to WT ClC-ec1 with Cl^-, (S107P) ClC-ec1 mutant with Cl^-, WT ClC-ec1 with NO_3^-, and (S107P) mutant ClC-ec1 with NO_3^- systems, respectively.

During the 800 ns MD simulation, we analyzed the radius of the pore, the distance between the ion's center of mass and selectivity filter's center of mass, the Root-Mean-Square Deviations (RMSD) of the protein backbone/SF and Glu_{ext} position and time-evolution of the rotameric state χ_1 of Phe357 residue.

The second part is dedicated to the construction of free energy surface (FES) of ion permeation events in order to quantify energetic barriers by running metadynamics simulations.

3.4.1 Metadynamics

We first tested our system using a spherical restraint to limit the volume in the solvated state and applying the bias in three collective variables (r, ϕ and θ), accordingly to the recently developed Volume-based metadynamics approach [95]. This approach has the advantage of permit an exhaustive exploration of all the binding pathways. Due to the lack of convergence of the simulation after 800ns, we decided to develop a cylindrical variant of the Volume-based metadynamics method, which is also based on the well-tempered metadynamics approach. In our variant, we took advantage of the fact that the flexibility of proteins (like ClC-ec1) embedded in a membrane is limited and thus its center of mass can be defined once and used for the rest of the simulation. Centered around this position we consider a cylinder of finite radius ρ_s, larger than the radius of the pore as ilustrated in Figure 3.9. Ion's positions are described in cylindrical coordinates (ρ, ϕ, z), and z coordinate is used as the CV, putting a MTD bias on it. A repulsive potential is added at the border

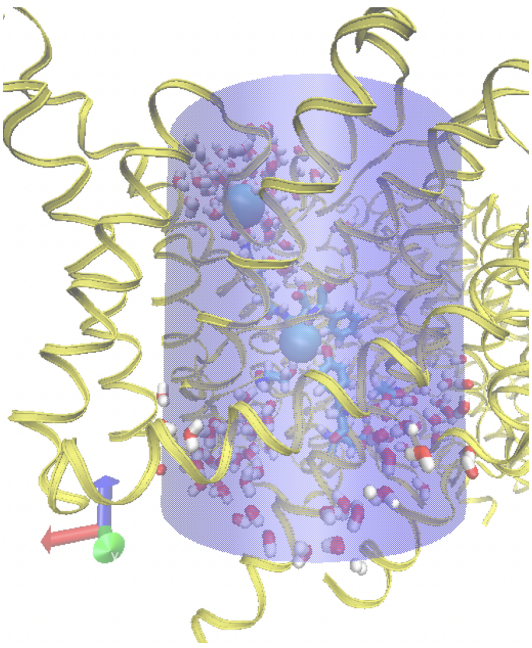


Fig. 3.9. – Representation of the cylindrical restraining potential added to the system.

of the cylinder, increasing the probability of a subsequent recrossing event. The restraining potential is in the form:

$$U_s(\rho(t), z(t)) = \begin{cases} \frac{1}{2}k(\rho(t) - \rho_s)^2 & \text{if } \rho(t) > \rho_s \\ \frac{1}{2}k(|z(t)| - z_s)^2 & \text{if } |z(t)| > z_s \\ 0 & \text{otherwise} \end{cases}$$

where k has to be large enough to prevent the ligand escaping from the confining volume, $\rho(t)$ is the xy projection distance of the ion from the center of mass of the target protein, ρ_s is the radius of the cylinder, $z(t)$ is the value of the z coordinate from the center of mass of the target protein and z_s is half of the cylinder's height.

3.5 Computational details

Simulations were performed on the well-know ClC-ec1 bacterial model (PDB ID: 1OTS, resolution: 2.5 Å), taken from OPM database [96]. The protein (WT and mutant)/membrane systems were built using the CHARMM-GUI web tool [97]. PROPKA [98] was used to predict the pKa of protein residues and determine the protonation states at pH = 3 (physiologycal pH for this transporter [99]). Results of the pka prediction suggested to keep residue Glu_{int} (E203) deprotonated, while

GLU_{ext} (E148) protonated. We also implemented E148 unprotonated due to its importance in the gating mechanism of the channel and in the proton coupling necessary for transport (at least for chloride).

Protein was embedded in 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer (100 lipid molecules for the monomer and 400 for the dinmer) and solvated with TIP3P[100] water with 150 mM NaCl salt concentration. Proteins and lipids were modeled with CHARMM36m[101] and LIPID17 force fields, respectively. Parameters for nitrate were taken from [102], while parameters for Cl^{-1} were the default given by CHARMM-GUI program [97]. The homodimer simulation box contains a total of (168,000) atoms, including crystalographic waters, with box length of 130 Åx 130 Åx 98 Å. All the simulations were performed using GROMACS-2018.3 [45] source code patched with PLUMED version 2.5.2 [103, 104] for the well-tempered metadynamics simulations. Besides simulations with the dimeric form of ClC-ec1, simulations with a single monomer were also performed.

Long range electrostatic interactions were evaluated using particle-mesh Ewald (PME) summations [48] and Lennard-Jones interactions were truncated at 12 Åwith an atom based force switching function which is effective at 10 Å. After energy minimization, each system was equilibrated in NVT ensemble for 500 ps and 6 ns in NPT ensemble. Equilibration was diveden into 6 steps (2 for NVT and 4 for NPT equilibrations) as suggested by CHARMM-GUI, harmonic position restraints were applied to the solute heavy atoms with a decreasing force constant between 4000 and 50 kJ mol^{-1} nm^{-2}. After equilibration we continued the production run for 800 ns in NPT ensemble. The integration time step is set at 2 fs and the simulation is performed at 310 K and 1.0 atm pressure. Nosé-Hoover chain thermostat with time constant 1.0 ps was used to control the temperature. Semiisotropic Parrinello-Rahman barostat with reference pressure 1 atm, time constant 5.0 ps and isothermal compressibility of 4.5 x 10^{-5} bar^{-1} were used to maintain the pressure of the system.

In volume based metadynamics, 1 CV (Z axis) with cilindrical restraint was used. We used gaussian width = 0.1 nm, pi/16 rad, pi/8 rad, gaussian height =1.3 kJ/mol, deposition time between gaussians = 500 MD time steps, BIASFACTOR=20 and TEMP=310 K. Sigma values were chosen based on the fluctuations of the three coordinates during the classical MD. Bias factor was chosen based on the expected barrier heights. Convergence was assumed as suggested in [92] by monitoring c(t) behaviour and difusion of the collective variable.

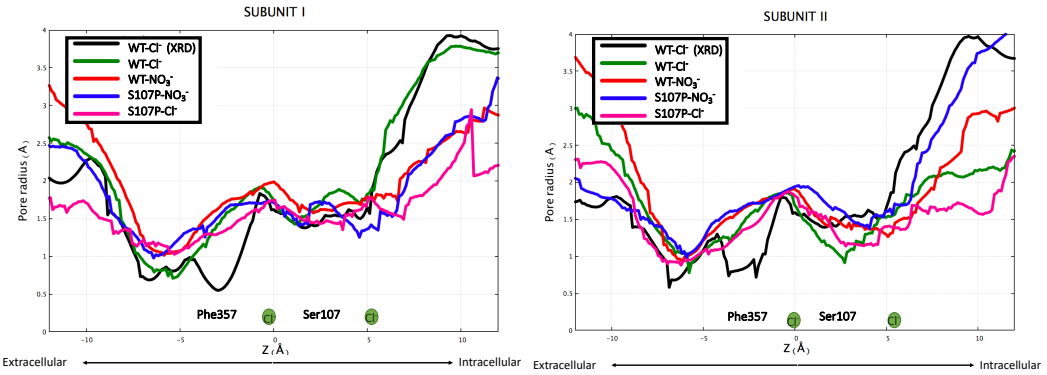


Fig. 3.10. – Comparision of pore radius profiles of the crystal structure and those of the MD simulation (average) with both anions (Cl^- and NO_3^-) bound at S_{cen} and both proteins (WT and S107P). The center of the bound anion at t=0, which correspond to the Cl^- position in the crystal structure, is set as the origin. Green circles represent the positions of the Cl^- ions in the crystal structure. Phe357 (F357) and Ser107/Pro107 (S107/P107) distance from the central Cl^- in the crystal structure are also indicated.

Energy calculations with MMPBSA method were performed using g-mmpbsa package [44], as in section 2.4.4. MMPBSA calculations were performed using results from MD clasical simulations in a single–trajectory (one-average) setup.

3.6 Results

During the 800 ns MD simulation, we analyzed the radius of the pore, the distance between the ion's center of mass and selectivity filter's center of mass, the Root-Mean-Square Deviations (RMSD) of the protein backbone/SF and Glu_{ext} position and time-evolution of the rotameric state χ_1 of Phe357 residue.

The average pore radius profile of the anion conduction pathway follows the trend observed for the crystal structure (Figure 3.10). The two main bottlenecks, corresponding to the constriction regions that occlude the anion from the aqueous solutions, are clearly preserved (around -5 and 3 Å), although the extracellular bottleneck shows a moderate expansion relative to the crystal structure in all the simulated systems other than S107P-Cl^- in subunit I. In general, the overall structure of the protein does not appear to be largerly affected by the chemical nature of the bound anion or the mutated aminoacid. Upon initial inspection, the selectivity properties of these anion transporters do not seem to arise simply from geometrical properties.

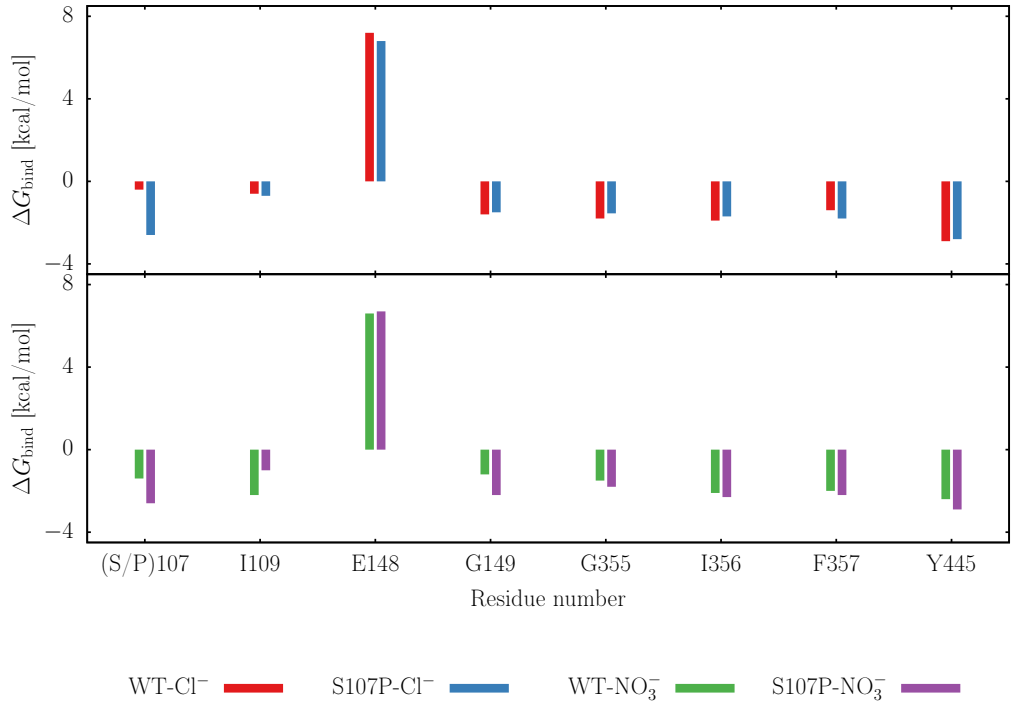


Fig. 3.11. – Interaction energy between the bound anion and the main highly conserved residues in the selectivity filter of the four simulated systems.

	WT		S107P mutant	
Anion	**Cl^-**	**NO^-_3**	**Cl^-**	**NO^-_3**
Permeability		Less permeant than Cl^- in WT [a] [64]		Slightly more permeant than Cl^- [c] [64]
Conductivity		Less conductive than Cl^- in WT[a, 64]		
Current			Barely above the background[d] [105], [e] [64]	Increase of outward current[d] [105]
Ion-H^+ Coupling		Weak[b,d] [105]	Cl^-/H^+ exchange is slowed ~8 fold[b] [64]	Stronger compared with WT [d] [105]
Binding			Not bind[b] [64]	Binding is enhaced ~4 fold[b] [64]
Affinity (kcal/mol)	N*=1, -4.3[64] N=2, -4.4[64]	N=1, -2.6 [64]	No heat[b] [64]	N=1, -3.3[64]
Stoiquiometry Anion : Proton	2:1	7-10:1[b]		

Fig. 3.12. – Summary of the main experimental features related to Cl^-/NO_3^- selectivity, binding and coupling through the ClC family (a.ClC-0 channel, b. ClCec1 transporter, c. ClC-0 and S123P/E166A, d.ClC-5 transporter, e. ClC-0 S123P channel, N means the number of sites considered for the measurement.

The average interaction energies between anions at S_{cen} and individual residues through the protein selectivity filter, presented in Figure 3.11 indicate that the anion binding is mainly supported by residues coordinating Cl^- at S_{cen} in the crystal structure, being Ser107 and Tyr445 the residues with the most favourable contributions and Glu148 the most unfavourable contribution for WT-Cl^- and WT-NO_3^- systems. S107P-Cl^- system shows a similar trend but the average interaction energy of residue Pro at position 107 decreases its value, showing that this residue could be partly responsably for the lack of experimental affinity for Cl^- at this position in the mutant [64] (Figure 3.12).

In Figure 3.13a, Backbone RMSD subunit I converge to a value around 0.2 nm for systems S107P-Cl^-, WT-NO_3^- and S107P-NO_3^-, but for WT-Cl^-, it takes a higher value around 0.3 nm. In Figure 3.13b, Selectivity Filter RMSD subunit I converge to a value around 0.1 nm for systems S107P-Cl^-, WT-NO_3^-, but for S107P-Cl^- and S107P-NO_3^-, it takes higher values around 0.2 nm and 0.3 nm, respectively. In Figure 3.13c, Backbone RMSD subunit II converges to a value of around 0.18 nm for all the systems. In Figure 3.13d, Selectivity Filter RMSD subunit II converges to a value of around 0.15 nm for systems S107P-Cl^-, WT-NO_3^- and S107P-NO_3^- but for S107P-NO_3^-, it takes a higher value around 0.25 nm. In order to know what are the residues responsible for the major changes in RMSD, we evaluated de Root Mean Square Fluctuations (RMSF) in the selectivity filter of WT-Cl^- chain I and S107P-

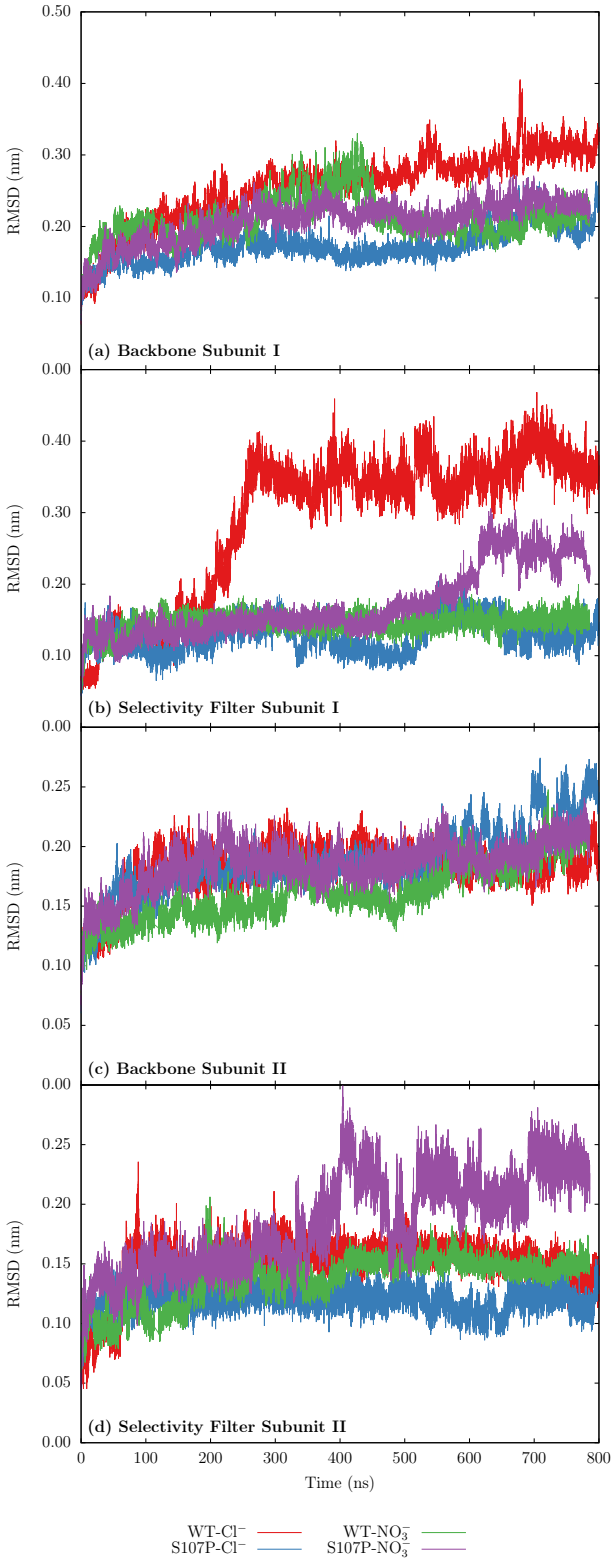


Fig. 3.13. – Time evolution of backbone protein RMSD (a,c) and Selectivity Filter RMSD (b,d).

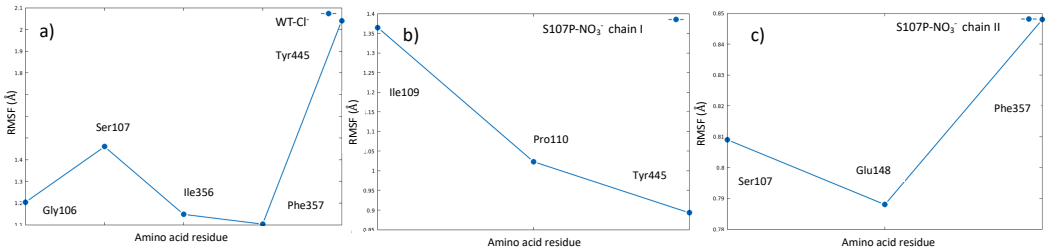


Fig. 3.14. – RMSF of a) SCL chain I; b) S107P-NO_3^- of chain I; and c) PNO3 chain II.

NO_3^- in both subunits (Figure 3.14). Results show that the most flexible residues are Ser107/Tyr445, Ile109/Pro110/Tyr445 and Ser107/Glu148/Phe357; for WT-Cl^- subunit I, S107P-NO_3^- subunit I and S107P-NO_3^- subunit II, respectively.

It can be observed in Figure 3.15a that the central ion distance in subunit I has major changes for WT-Cl^-, S107P-NO_3^- and S107P-Cl^-. In Figure 3.15b, the central ion distance in subunit II has major changes for S107P-NO_3^- system. After verification of trajectories, it was found that Cl^- in WT-Cl^- subunit I leaves the pore through the cytoplasmatic side, NO_3^- moves towards S_{ext} in S107P-NO_3^- subunit I and Cl^- moves towards S_{ext} in S107P-Cl^- subunit I. This results could suggest that the lower barriers are when translocating from S_{cen} to S_{int} for WT-Cl^- system, from S_{cen} to S_{ext} for S107P-NO_3^- and S107P-Cl^- systems, as also sugested by other computational studies [75]. In Figure 3.15c-d we present the distance between Glu_{ext} from S_{cen}. In S107P-NO_3^- system subunit I and WT-NO_3^- subunit II, Glu_{ext} changes from the "Up" to the "alternative down" position (Figure 3.16) that facilitates the formation of water wires. All the other systems remain with Glu_{ext} in the "Up" position. It is important to mention that this "alternative Down" position outside the ion pathway has been only recently reported for the WT with Cl^-[79, 81]. This result is important because it could suggest that the mutant with NO_3^- could share similar (compared to the WT ClC-ec1 with Cl^-) mechanisms for ion/proton coupling, anion transport and affinity, something that has not been reported so far. Unfortunately, we can not make many conclusions related to selectivity because according to experimental results [64], Glu_{ext} and simultaneous occupancy of S_{cen} and S_{ext} regulate affinity but not selectivity.

The behavior of Phe_{cen} was also followed during the simulation. This residue forms part of the Cl^- permeation pathway in ClCs by coordinating ions in S_{cen} and S_{ext} with its backbone amide in WT-Cl^- system. It was recently found that in WT-Cl^- system, Phe357 side chain exists in equilibrium between two rotameric states with angles of $-160°$ "up" and $-70°$ "down" and that the transition between this rotamers is critical for ion permeation [79]. We found the same equilibrium between the two states in all the four systems; this observation supports the idea that this rotameric

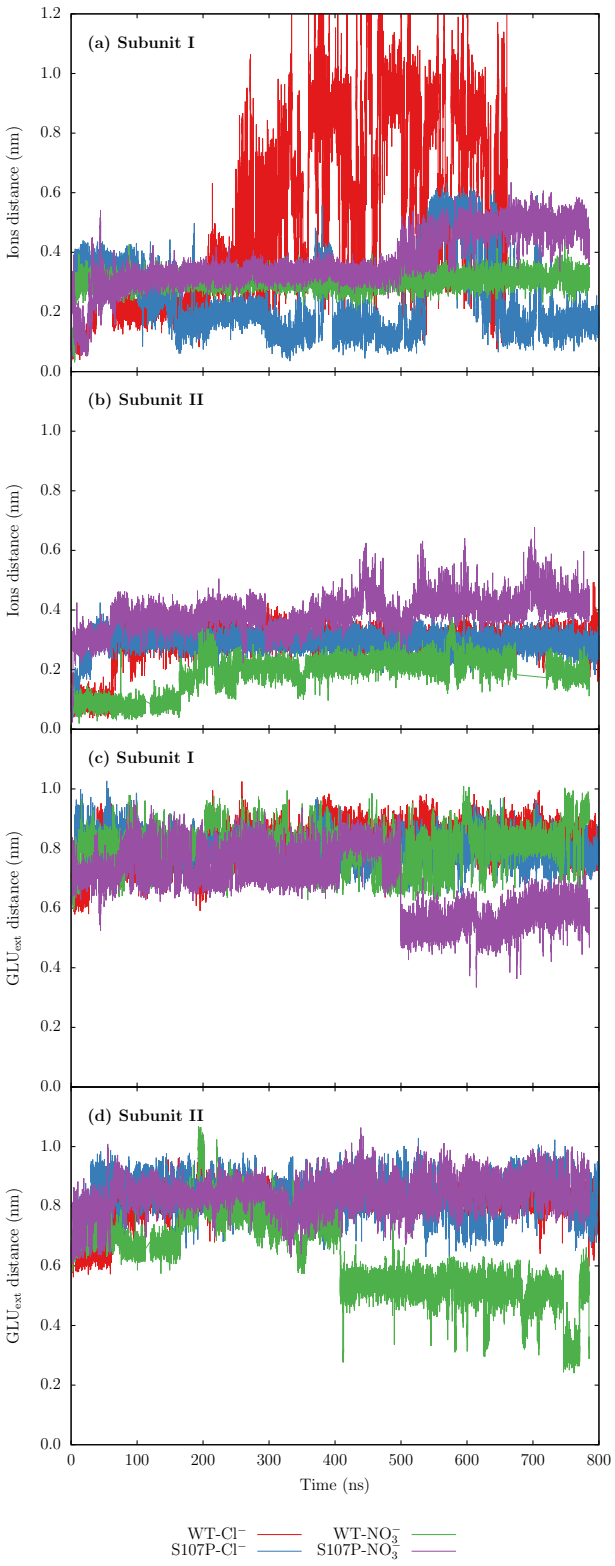


Fig. 3.15. – Time evolution of the distance of central ion from S_{cen} (a,b) and time evolution of the distance of Glu_{ext} from S_{cen} (c,d).

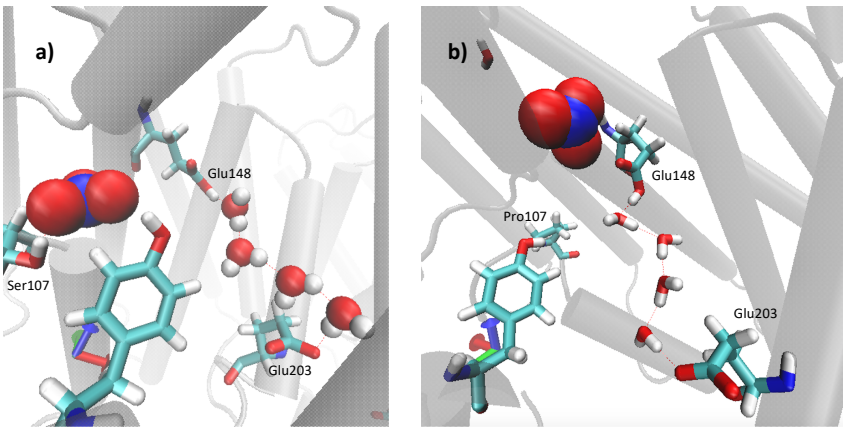


Fig. 3.16. – Glu_{ext} position and presence of water wires observed in a) WT-NO_3^- and b) S107P-NO_3^- systems.

transition could play a role in WT-NO_3^- and S107P-NO_3^- systems. However, for WT-Cl^-, S107P-NO_3^- and WT-NO_3^-, the prefered state is the "up" state, while for S107P-Cl^- is the down state (Figure 3.17). Wheter this different preferred state ("up") in the mutant system (S107P-Cl^-)(compared with the WT-Cl^- system) is related to the following facts (summarized in Figure 3.12): i) in S168P ClC-5 mutant Cl^- current magnitud is barely above the background and ii) Cl^- does not bind to the S107P ClC-ec1 mutant at the S_{cen} position, needs further investigation.

Due to the fact that the only system where the ion leaves the pore is the WT-Cl^-, we investigated a little bit more the details of the ion unbinding. We found a rearrangement in Ser107, that is unstabilizing Cl_{cen} and lately leading to Cl^- movement towards the cytoplasmatic solvent. The Ser107 movement is accompanied by the movement of the coil segment (residues 99 to 110) stabilized by the formation of hydrogen bond between the amide backbone of Ser107 and Glu111 side chain; the latter is initially forming salt bridge with Arg120 (Figure 3.18). It is important to mention that this rearrangement of the inner-gate backbone along residues Ser168-Gly169-Ile170 (equivalent to Ser107-Glu108-Ile109) has been recently reported in ClC-2 channel and is strikingly similar to that observed in the cryo-EM structure of the bovine ClC-K channel [105], paving the way for newer ClC-ec1 gating models.

The converged FES for the WT-Cl^- is shown in Figure 3.19 as obtained with the well-temperature metadynamics simulations explained above. In Figure 3.19a we present the behavior of the CV, which is diffusing over the allowed phase space. In Figure 3.19b we can observe the logarithmic behavior for the c(t) values which, together with the diffusivity of the CV, gives us an idea of the convergence of the system [92]. In Figure 3.20a-b, we could identify the main binding sites reported in the literature (S_{int}, S_{cen}, S_{ext} and S_{out}). Free energy barriers for Cl^- translocation

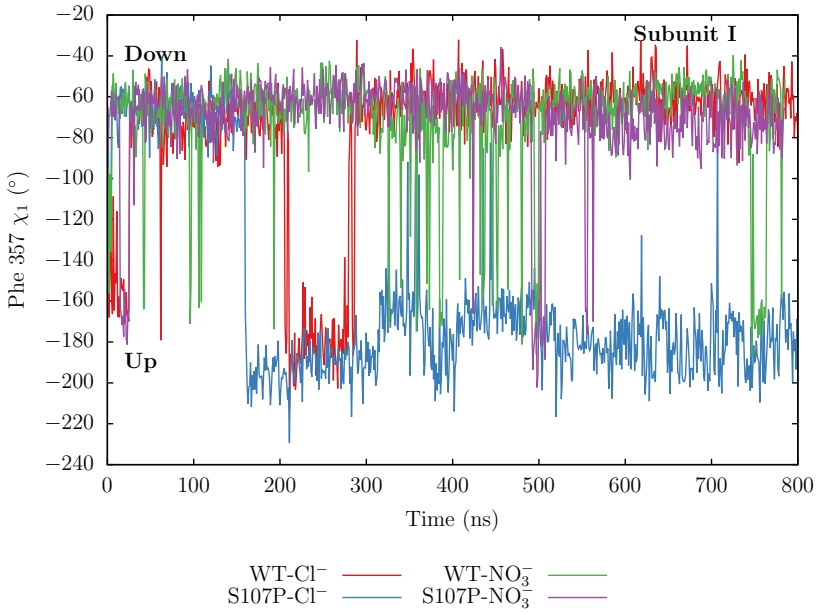


Fig. 3.17. – Time-evolution of the rotameric state χ_1 of Phe357 side chain.

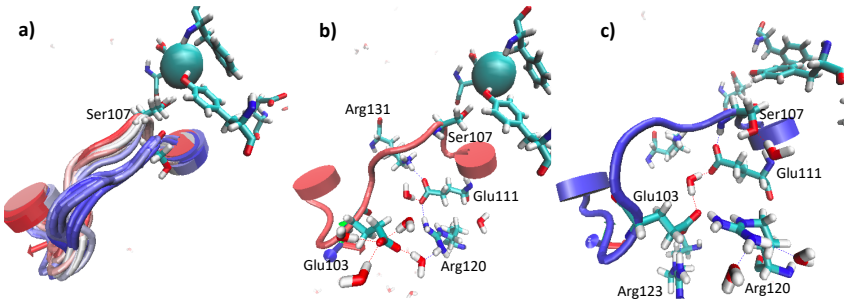


Fig. 3.18. – a) Ser107 rearrangement, b) an initial configuration (red) c) compared with an older configuration (blue). This rearrangement occured around 150 ns of the simulation.

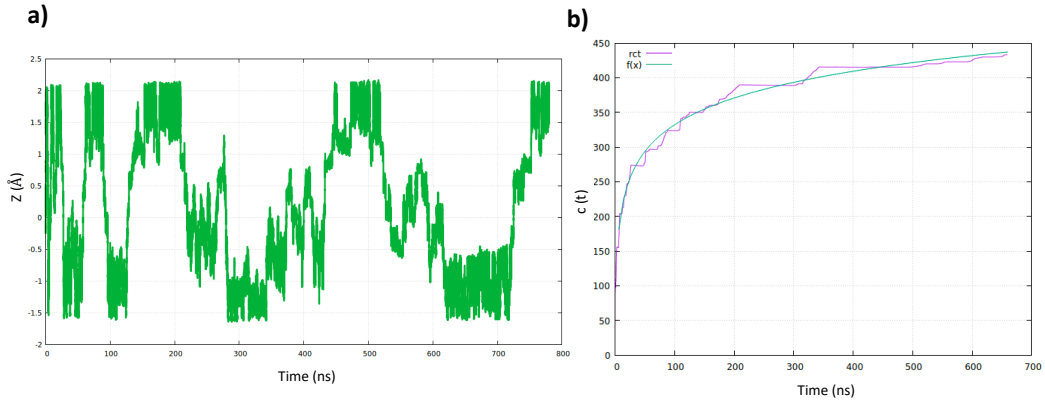


Fig. 3.19. – a) Time evolution of the CV (Z), b) c(t) fitted to a logaritmic model.

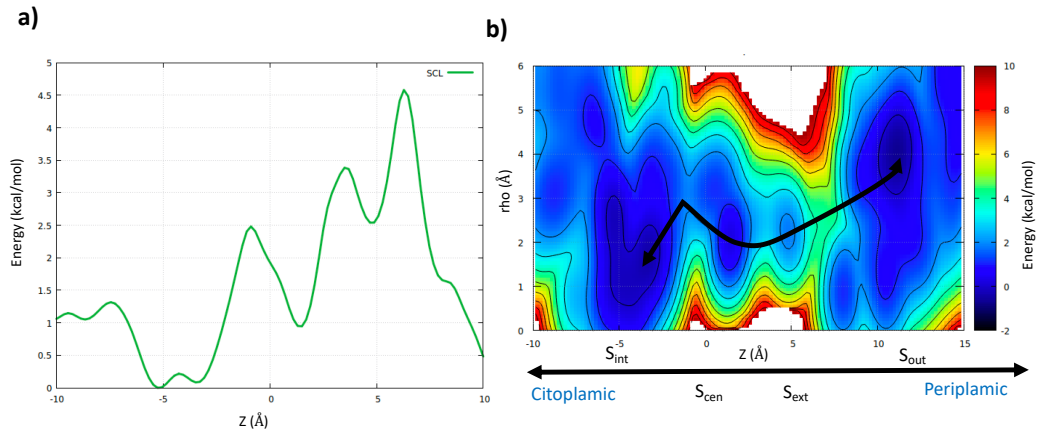


Fig. 3.20. – a) estimate of the free energy as a function of Z coordinate and b) estimate of the free energy as a function of Z coordinate and rho cilindrical coordinate, using 1 CV (Z coordinate).

are in agreement with those reported in the literature, between 3 and 8 kcal/mol [71, 106]. The lower internal barrier is in agreement with our classical non biased MD where the ion leaves the central binding site towards the internal one. Once validated the WT system, our next step is to reach convergence in the nitrate and mutant system to compare the energy barriers and minima between systems.

3.7 Conclusions

According to our classical simulations, the selectivity properties of ClC-ec1 WT and mutant do not seem to arise from geometrical properties.

Interaction energies in S107P-Cl^- system shows that the favourable energetic contribution of Pro at position 107 decreases its value compared to WT, showing that this residue could be partly responsably for the lack of experimental affinity for Cl^- at this position in the mutant.

We could identify, in S107P-NO_3^- system subunit I and WT-NO_3^- subunit II, that Glu_{ext} changes from the "Up" to the "alternative down" position which shows that the uncoupling of WT-NO_3^- system is not necessarily caused by the pseudo water wires as proposed in reference [75], because formal water wires can appear with this GLUext configuration without affecting the NO_3^- binding or the other way around.

For WT-Cl^-, S107P-NO_3^- and WT-NO_3^-, the preferred Glu_{ext} configuration is the "up" state while for S107P-Cl^- is the "down" state. This different behaviour in S107P-Cl^- system could be related to the fact that this is less permeant and with no affinity at the S_{cen} position for Cl^-.

We have been able to identify the main binding sites (S_{int}, S_{cen}, S_{ext} and S_{out}). Our results for the free energy barriers for Cl^- translocation are in agreement with the reported in the literature, which are between 3 and 8 kcal/mol [71, 106].

Bibliography

[1] Scott A Hollingsworth and Ron O Dror. Molecular dynamics simulation for all. *Neuron*, 99(6):1129–1143, 2018.

[2] Eric H Lee, Jen Hsin, Marcos Sotomayor, Gemma Comellas, and Klaus Schulten. Discovery through the computational microscope. *Structure*, 17(10):1295–1306, 2009.

[3] Rolando P Hong Enriquez, Silvia Pavan, Fabio Benedetti, Alessandro Tossi, Adriano Savoini, Federico Berti, and Alessandro Laio. Designing short peptides with high affinity for organic molecules: a combined docking, molecular dynamics, and monte carlo approach. *Journal of chemical theory and computation*, 8(3):1121–1128, 2012.

[4] Ivan Gladich, Alex Rodriguez, Rolando P Hong Enriquez, Filomena Guida, Federico Berti, and Alessandro Laio. Designing high-affinity peptides for organic molecules by explicit solvent molecular dynamics. *The Journal of Physical Chemistry B*, 119(41):12963–12969, 2015.

[5] Malihe Ebrahimi, Ahmad Mani-Varnosfaderani, Taghi Khayamian, and Sajjad Gharaghani. An in silico approach to design peptide mimetics based on docking and molecular dynamics simulation of EGFR–matuzumab complex. *Journal of the Iranian Chemical Society*, 13(10):1805–1817, 2016.

[6] Anna Russo, Pasqualina Liana Scognamiglio, Rolando P Hong Enriquez, Carlo Santambrogio, Rita Grandori, Daniela Marasco, Antonio Giordano, Giacinto Scoles, and Sara Fortuna. In Silico generation of peptides by replica exchange Monte Carlo: docking-based optimization of Maltose-binding-protein ligands. *PLoS One*, 10(8):e0133571, 2015.

[7] Miguel A Soler, Alex Rodriguez, Anna Russo, Abimbola Feyisara Adedeji, Cedrix J Dongmo Foumthuim, Cristina Cantarutti, Elena Ambrosetti, Loredana Casalis, Alessandra Corazza, and Giacinto Scoles. Computational design of cyclic peptides for the customized oriented immobilization of globular proteins. *Physical Chemistry Chemical Physics*, 19(4):2740–2748, 2017.

[8] L América Chi and M Cristina Vargas. In silico design of peptides as potential ligands to resistin. *Journal of Molecular Modeling*, 26:1–14, 2020.

[9] Filomena Guida, Anna Battisti, Ivan Gladich, Mauro Buzzo, Elena Marangon, Luciana Giodini, Giuseppe Toffoli, Alessandro Laio, and Federico Berti. Peptide biosensors for anticancer drugs: design in silico to work in denaturizing environment. *Biosensors and Bioelectronics*, 100:298–303, 2018.

[10] Xing Du, Yi Li, Yuan-Ling Xia, Shi-Meng Ai, Jing Liang, Peng Sang, Xing-Lai Ji, and Shu-Qun Liu. Insights into protein–ligand interactions: mechanisms, models, and methods. *International Journal of Molecular Sciences*, 17(2):144, 2016.

[11] Matteo Aldeghi, Michael J Bodkin, Stefan Knapp, and Philip C Biggin. Statistical analysis on the performance of Molecular Mechanics Poisson–Boltzmann Surface Area versus absolute binding free energy calculations: Bromodomains as a case study. *Journal of Chemical Information and Modeling*, 57(9):2203–2221, 2017.

[12] Michael R Shirts, David L Mobley, and John D Chodera. Alchemical free energy calculations: ready for prime time? *Annual Reports in Computational Chemistry*, 3:41–59, 2007.

[13] Sarah E. Boyce, David L. Mobley, Gabriel J. Rocklin, Alan P. Graves, Ken A. Dill, and Brian K. Shoichet. Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. *Journal of Molecular Biology*, 394(4):747–763, 2009.

[14] Claire M Steppan, Elizabeth J Brown, Christopher M Wright, Savitha Bhat, Ronadip R Banerjee, Charlotte Y Dai, Gregory H Enders, Debra G Silberg, Xiaoming Wen, Gary D Wu, et al. A family of tissue-specific resistin-like molecules. *Proceedings of the National Academy of Sciences*, 98(2):502–506, 2001.

[15] Mária Filková, Martin Haluzík, Steffen Gay, and Ladislav Šenolt. The role of resistin as a regulator of inflammation: implications for various human pathologies. *Clinical Immunology*, 133(2):157–170, 2009.

[16] Miguel Patrício, José Pereira, Joana Crisóstomo, Paulo Matafome, Manuel Gomes, Raquel Seiça, and Francisco Caramelo. Using resistin, glucose, age and BMI to predict the presence of breast cancer. *BMC Cancer*, 18(1):29, 2018.

[17] Claire M Steppan, Shannon T Bailey, Savitha Bhat, Elizabeth J Brown, Ronadip R Banerjee, Christopher M Wright, Hiralben R Patel, Rexford S Ahima, and Mitchell A Lazar. The hormone resistin links obesity to diabetes. *Nature*, 409(6818):307, 2001.

[18] Syeda Ijlal Zehra Zaidi and Tanvir Ali Khan Shirwany. Relationship of serum resistin with insulin resistance and obesity. *Journal of Ayub Medical College Abbottabad*, 27(3):552–555, 2015.

[19] Koichiro Azuma, Fuminori Katsukawa, Shuji Oguchi, Mitsuru Murata, Hajime Yamazaki, Akira Shimada, and Takao Saruta. Correlation between serum resistin level and adiposity in obese individuals. *Obesity Research*, 11(8):997–1001, 2003.

[20] KM Utzschneider, DB Carr, J Tong, TM Wallace, RL Hull, S Zraika, Q Xiao, JS Mistry, BM Retzlaff, RH Knopp, et al. Resistin is not associated with insulin sensitivity or the metabolic syndrome in humans. *Diabetologia*, 48(11):2330–2333, 2005.

[21] Ivan Nagaev and Ulf Smith. Insulin resistance and type 2 diabetes are not related to resistin expression in human fat cells or skeletal muscle. *Biochemical and Biophysical Research Communications*, 285(2):561–564, 2001.

[22] Tatu Pantsar and Antti Poso. Binding affinity via docking: fact and fiction. *Molecules*, 23(8):1899, 2018.

[23] Oleg Trott and Arthur J Olson. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2):455–461, 2010.

[24] Max W Chang, Christian Ayeni, Sebastian Breuer, and Bruce E Torbett. Virtual screening for HIV protease inhibitors: a comparison of AutoDock 4 and Vina. *PLoS One*, 5(8):e11955, 2010.

[25] David J Earl and Michael W Deem. Parallel tempering: theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910–3916, 2005.

[26] Gabrielle M Pine, Hashini M Batugedara, and Meera G Nair. Here, there and everywhere: resistin-like molecules in infection, inflammation, and metabolic disorders. *Cytokine*, 110:442–451, 2018.

[27] Beatriz Sánchez-Solana, Jorge Laborda, and Victoriano Baladrón. Mouse resistin modulates adipogenesis and glucose uptake in 3T3-L1 preadipocytes through the ROR1 receptor. *Molecular Endocrinology*, 26(1):110–127, 2012.

[28] Alexes C Daquinag, Yan Zhang, Felipe Amaya-Manzanares, Paul J Simmons, and Mikhail G Kolonin. An isoform of decorin is a resistin receptor on the surface of adipose progenitor cells. *Cell Stem Cell*, 9(1):74–86, 2011.

[29] Andrej Tarkowski, Jan Bjersing, Andrey Shestakov, and Maria I Bokarewa. Resistin competes with lipopolysaccharide for binding to toll-like receptor 4. *Journal of Cellular and Molecular Medicine*, 14(6b):1419–1431, 2010.

[30] Sahmin Lee, Hyun-Chae Lee, Yoo-Wook Kwon, Sang Eun Lee, Youngjin Cho, Joonoh Kim, Soobeom Lee, Ju-Young Kim, Jaewon Lee, Han-Mo Yang, et al. Adenylyl cyclase-associated protein 1 is a receptor for human resistin and mediates inflammatory actions of human monocytes. *Cell Metabolism*, 19(3):484–497, 2014.

[31] Hiroshi Onuma, Yasuharu Tabara, Ryoichi Kawamura, Jun Ohashi, Wataru Nishida, Yasunori Takata, Masaaki Ochi, Tatsuya Nishimiya, Ryuichi Kawamoto, Katsuhiko Kohara, et al. Plasma resistin is associated with single nucleotide polymorphisms of a possible resistin receptor, the decorin gene, in the general japanese population. *Diabetes*, 62(2):649–652, 2013.

[32] Morten Källberg, Haipeng Wang, Sheng Wang, Jian Peng, Zhiyong Wang, Hui Lu, and Jinbo Xu. Template-based protein structure modeling using the RaptorX web server. *Nature Protocols*, 7(8):1511–1522, 2012.

[33] Julie D Thompson, Toby Gibson, and Des G Higgins. Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics*, (1):2–3, 2002.

[34] Saurabh D Patel, Michael W Rajala, Luciano Rossetti, Philipp E Scherer, and Lawrence Shapiro. Disulfide-dependent multimeric assembly of resistin family hormones. *Science*, 304(5674):1154–1158, 2004.

[35] Oleg Trott and Arthur J Olson. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2):455–461, 2010.

[36] Stefania Evoli, David L Mobley, Rita Guzzi, and Bruno Rizzuti. Multiple binding modes of ibuprofen in human serum albumin identified by absolute binding free energy calculations. *Physical Chemistry Chemical Physics*, 18(47):32358–32368, 2016.

[37] Nastaran Moradi, Mohammad Reza Ashrafi-Kooshk, Sirous Ghobadi, Mohsen Shahlaei, and Reza Khodarahmi. Spectroscopic study of drug-binding characteristics of unmodified and pNPA-based acetylated human serum albumin: does esterase activity affect microenvironment of drug binding sites on the protein? *Journal of Luminescence*, 160:351–361, 2015.

[38] Stefan Boresch, Franz Tettinger, Martin Leitgeb, and Martin Karplus. Absolute binding free energies: a quantitative approach for their calculation. *The Journal of Physical Chemistry B*, 107(35):9535–9551, 2003.

[39] David L Mobley, John D Chodera, and Ken A Dill. Confine-and-release method: obtaining correct binding free energies in the presence of protein conformational change. *Journal of Chemical Theory and Computation*, 3(4):1231–1235, 2007.

[40] Michael R Shirts. *Calculating precise and accurate free energies in biomolecular systems*. PhD thesis, Stanford University, 2004.

[41] Pavel V Klimovich, Michael R Shirts, and David L Mobley. Guidelines for the analysis of free energy calculations. *Journal of Computer-Aided Molecular Design*, 29(5):397–411, 2015.

[42] Yilin Meng, Danial Sabri Dashti, and Adrian E Roitberg. Computing alchemical free energy differences with hamiltonian replica exchange molecular dynamics (H-REMD) simulations. *Journal of Chemical Theory and Computation*, 7(9):2721–2727, 2011.

[43] Robert W Zwanzig. High-temperature equation of state by a perturbation method. i. nonpolar gases. *The Journal of Chemical Physics*, 22(8):1420–1426, 1954.

[44] Rashmi Kumari, Rajendra Kumar, Open Source Drug Discovery Consortium, and Andrew Lynn. g-mmpbsa a GROMACS tool for high-throughput MM-PBSA calculations. *Journal of Chemical Information and Modeling*, 54(7):1951–1962, 2014.

[45] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman JC Berendsen. Gromacs: fast, flexible, and free. *Journal of Computational Chemistry*, 26(16):1701–1718, 2005.

[46] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1950–1958, 2010.

[47] Herman JC Berendsen, JPM van Postma, Wilfred F van Gunsteren, ARHJ DiNola, and JR Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690, 1984.

[48] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An nlog (n) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[49] Berk Hess, Henk Bekker, Herman JC Berendsen, and Johannes GEM Fraaije. LINCS: a linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.

[50] MSKP Settle. An analytical version of the shake and rattle algorithm for rigid water molecules. *Journal Computational Chemistry*, 13:952–962, 1992.

[51] Alexandra Patriksson and David van der Spoel. A temperature predictor for parallel tempering simulations. *Physical Chemistry Chemical Physics*, 10(15):2073–2077, 2008.

[52] Xavier Daura, Karl Gademann, Bernhard Jaun, Dieter Seebach, Wilfred F van Gunsteren, and Alan E Mark. Peptide folding: when simulation meets experiment. *Angewandte Chemie International Edition*, 38(1-2):236–240, 1999.

[53] David A Case, Thomas E Cheatham, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M Merz, Alexey Onufriev, Carlos Simmerling, Bing Wang, and Robert J Woods. The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26(16):1668–1688, 2005.

[54] Pymol tutorial: Interaction interface. `http://www.protein.osaka-u.ac.jp/rcsfp/supracryst/suzuki/jpxtal/Katsutani/en/interface.php`. Accessed: 2020-10-05.

[55] Peptide property calculator. `https://pepcalc.com/`. Accessed: 2020-10-05.

[56] Thomas P Hopp and Kenneth R Woods. A computer program for predicting protein antigenic determinants. *Molecular Immunology*, 20(4):483–489, 1983.

[57] Warren L DeLano. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*, 40:82–92, 2002.

[58] Jiyuan Liu, Zhen Tian, and Yalin Zhang. Structure-based discovery of potentially active semiochemicals for Cydia pomonella (L.). *Scientific Reports*, 6:34600, 2016.

[59] Alex M Clark and Paul Labute. 2D depiction of protein-ligand complexes. *Journal of Chemical Information and Modeling*, 47(5):1933–1944, 2007.

[60] Alessio Accardi and Alessandra Picollo. ClC channels and transporters: proteins with borderline personalities. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1798(8):1457–1464, 2010.

[61] Nils Piwon, Willy Günther, Michael Schwake, Michael R Bösl, and Thomas J Jentsch. ClC-5 Cl^--channel disruption impairs endocytosis in a mouse model for Dent's disease. *Nature*, 408(6810):369, 2000.

[62] Uwe Kornak, Dagmar Kasper, Michael R Bösl, Edelgard Kaiser, Michaela Schweizer, Ansgar Schulz, Wilhelm Friedrich, Günter Delling, and Thomas J Jentsch. Loss of the ClC-7 chloride channel leads to osteopetrosis in mice and man. *Cell*, 104(2):205–215, 2001.

[63] Thomas J Jentsch and Michael Pusch. ClC chloride channels and transporters: structure, function, physiology, and disease. *Physiological Reviews*, 98(3):1493–1590, 2018.

[64] Alessandra Picollo, Mattia Malvezzi, Jon CD Houtman, and Alessio Accardi. Basis of substrate binding and conservation of selectivity in the ClC family of channels and transporters. *Nature Structural & Molecular Biology*, 16(12):1294, 2009.

[65] G Zifarelli and M Pusch. ClC chloride channels and transporters: a biophysical and physiological perspective. In *Reviews of Physiology, Biochemistry and Pharmacology*, pages 23–76. Springer, 2007.

[66] Janice L Robertson, Ludmila Kolmakova-Partensky, and Christopher Miller. Design, function and structure of a monomeric ClC transporter. *Nature*, 468(7325):844–847, 2010.

[67] Benoît Roux, Simon Bernèche, Bernhard Egwolf, Bogdan Lev, Sergei Y Noskov, Christopher N Rowley, and Haibo Yu. Ion selectivity in channels and transporters. *Journal of General Physiology*, 137(5):415–426, 2011.

[68] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.

[69] Toby W Allen, Olaf S Andersen, and Benoît Roux. Energetics of ion conduction through the gramicidin channel. *Proceedings of the National Academy of Sciences*, 101(1):117–122, 2004.

[70] Simone Furini and Carmen Domene. Computational studies of transport in ion channels using metadynamics. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1858(7):1733–1740, 2016.

[71] Francesco Luigi Gervasio, Michele Parrinello, Matteo Ceccarelli, and Michael L Klein. Exploring the gating mechanism in the ClC chloride channel via metadynamics. *Journal of Molecular Biology*, 361(2):390–398, 2006.

[72] Alessio Accardi. Structure and gating of ClC channels and exchangers. *The Journal of Physiology*, 593(18):4129–4138, 2015.

[73] K Diem and C Lentner. Blood–inorganic substances. *Documents Geigy Scientific Tables. Ciba-Geigy Limited, Basle Switzerland*, 565, 1970.

[74] José A O'Brien, Andrea Vega, Eléonore Bouguyon, Gabriel Krouk, Alain Gojon, Gloria Coruzzi, and Rodrigo A Gutiérrez. Nitrate transport, sensing, and responses in plants. *Molecular Plant*, 9(6):837–856, 2016.

[75] Tao Jiang, Wei Han, Merritt Maduke, and Emad Tajkhorshid. Molecular basis for differential anion binding and proton coupling in the Cl^--/H^+ Exchanger ClC-ec1. *Journal of the American Chemical Society*, 138(9):3066–3075, 2016.

[76] George Eisenman and Richard Horn. Ionic selectivity revisited: the role of kinetic and equilibrium processes in ion permeation through channels. *The Journal of Membrane Biology*, 76(3):197–225, 1983.

[77] Alessandro Laio and Vincent Torre. Physical origin of selectivity in ionic channels of biological membranes. *Biophysical Journal*, 76(1):129–148, 1999.

[78] David J Aidley and Peter R Stanfield. *Ion channels: molecules in action*. Cambridge University Press, 1996.

[79] Lilia Leisle, Yanyan Xu, Eva Fortea, Sangyun Lee, Jason D Galpin, Malvin Vien, Christopher A Ahern, Alessio Accardi, and Simon Bernèche. Divergent Cl^- and H^+ pathways underlie transport coupling and gating in ClC exchangers and channels. *Elife*, 9:e51224, 2020.

[80] Kunwoong Park, Byoung-Cheol Lee, and Hyun-Ho Lim. Mutation of external glutamate residue reveals a new intermediate transport state and anion binding site in a ClC Cl^-/H^+ antiporter. *Proceedings of the National Academy of Sciences*, 116(35):17345–17354, 2019.

[81] Tanmay S Chavan, Ricky C Cheng, Tao Jiang, Irimpan I Mathews, Richard A Stein, Antoine Koehl, Hassane S Mchaourab, Emad Tajkhorshid, and Merritt Maduke. Structural characterization of an intermediate reveals a unified mechanism for the ClC Cl^-/H^+ transport cycle. *BioRxiv*, page 857136, 2019.

[82] Liang Feng, Ernest B Campbell, Yichun Hsiung, and Roderick MacKinnon. Structure of a eukaryotic ClC transporter defines an intermediate state in the transport cycle. *Science*, 330(6004):635–641, 2010.

[83] Daniel Basilio, Kristin Noack, Alessandra Picollo, and Alessio Accardi. Conformational changes required for H^+/Cl^- exchange mediated by a ClC transporter. *Nature Structural & Molecular Biology*, 21(5):456, 2014.

[84] Chandra M Khantwal, Sherwin J Abraham, Wei Han, Tao Jiang, Tanmay S Chavan, Ricky C Cheng, Shelley M Elvington, Corey W Liu, Irimpan I Mathews, Richard A Stein, et al. Revealing an outward-facing open conformational state in a ClC Cl^--$/H^+$ exchange transporter. *Elife*, 5:e11189, 2016.

[85] Stefanie Wege, Mathieu Jossier, Sophie Filleur, Sébastien Thomine, Hélène Barbier-Brygoo, Franco Gambale, and Alexis De Angeli. The proline 160 in the selectivity filter of the arabidopsis NO_3^-/H^+ exchanger AtClCa is essential for nitrate accumulation in planta. *The Plant Journal*, 63(5):861–869, 2010.

[86] Eun-Yeong Bergsdorf, Anselm A Zdebik, and Thomas J Jentsch. Residues important for nitrate/proton coupling in plant and mammalian ClC transporters. *Journal of Biological Chemistry*, 284(17):11184–11193, 2009.

[87] Lilia Leisle, Eva Fortea, Jason Galpin, Christopher Ahern, and Alessio Accardi. Backbone amide nitrogen atoms are key determinants of inter-anion discrimination in ClCs. *Biophysical Journal*, 114(3):493a, 2018.

[88] Workalemahu Berhanu, Ping Jiang, and Ulrich HE Hansmann. Enhanced sampling for biomolecular simulations. In *Computational Methods to Study*

the Structure and Dynamics of Biomolecules and Biomolecular Processes, pages 249–267. Springer, 2014.

[89] Cameron Abrams and Giovanni Bussi. Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy*, 16(1):163–199, 2014.

[90] Johannes Kästner. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6):932–942, 2011.

[91] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Physical Review Letters*, 100(2):020603, 2008.

[92] Pratyush Tiwary and Michele Parrinello. A time-independent free energy estimator for metadynamics. *The Journal of Physical Chemistry B*, 119(3):736–742, 2015.

[93] Alessandro Laio and Francesco L Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics*, 71(12):126601, 2008.

[94] Giovanni Bussi, Davide Branduardi, et al. Free-energy calculations with metadynamics: Theory and practice. *Reviews in Computational Chemistry*, 28:1–49, 2015.

[95] Riccardo Capelli, Paolo Carloni, and Michele Parrinello. Exhaustive search of ligand binding pathways via volume-based metadynamics. *The Journal of Physical Chemistry Letters*, 10(12):3495–3499, 2019.

[96] Mikhail A Lomize, Irina D Pogozheva, Hyeon Joo, Henry I Mosberg, and Andrei L Lomize. Opm database and ppm web server: resources for positioning of proteins in membranes. *Nucleic Acids Research*, 40(D1):D370–D376, 2012.

[97] Emilia L Wu, Xi Cheng, Sunhwan Jo, Huan Rui, Kevin C Song, Eder M Dávila-Contreras, Yifei Qi, Jumin Lee, Viviana Monje-Galvan, Richard M Venable, et al. CHARMM-GUI membrane builder toward realistic biological membrane simulations. *Journal of Computational Chemistry*, 35(27):1997–2004, 2014.

[98] Michał Rostkowski, Mats HM Olsson, Chresten R Søndergaard, and Jan H Jensen. Graphical analysis of ph-dependent properties of proteins predicted using propka. *BMC structural biology*, 11(1):1–6, 2011.

[99] Wang Nguitragool and Christopher Miller. Uncoupling of a ClC Cl^-/H^+ exchange transporter by polyatomic anions. *Journal of Molecular Biology*, 362(4):682–690, 2006.

[100] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

[101] Jeffery B Klauda, Richard M Venable, J Alfredo Freites, Joseph W O'Connor, Douglas J Tobias, Carlos Mondragon-Ramirez, Igor Vorobyov, Alexander D MacKerell Jr, and Richard 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(23):7830–7843, 2010.

[102] M Baaden, M Burgard, and G Wipff. TBP at the water- oil interface: the effect of tbp concentration and water acidity investigated by molecular dynamics simulations. *The Journal of Physical Chemistry B*, 105(45):11131–11141, 2001.

[103] Massimiliano Bonomi, Giovanni Bussi, Carlo Camilloni, Gareth A Tribello, Pavel Banáš, Alessandro Barducci, Mattia Bernetti, Peter G Bolhuis, Sandro Bottaro, Davide Branduardi, et al. Promoting transparency and reproducibility in enhanced molecular simulations. *Nature Methods*, 16(8):670–673, 2019.

[104] Gareth A Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. PLUMED 2: New feathers for an old bird. *Computer Physics Communications*, 185(2):604–613, 2014.

[105] Keri A McKiernan, Anna K Koster, Merritt Maduke, and Vijay S Pande. Dynamical model of the clc-2 ion channel reveals conformational changes associated with selectivity-filter gating. *PLOS Computational Biology*, 16(3):e1007530, 2020.

[106] Jordi Cohen and Klaus Schulten. Mechanism of anionic conduction across ClC. *Biophysical Journal*, 86(2):836–845, 2004.

List of Figures

List of Tables

Glossary

amino acid is an organic compound that contain amine (–NH2) and carboxyl (–COOH) functional groups, along with a side chain (R group) specific to each different amino acid. There are 20 amino acids found in proteins. In addition, pyrrolysine, used in the biosynthesis of proteins in some archaea and bacteria but not present in humans, and selenocysteine, a cysteine analogue only found in some lineages. The charged ones are: Arginine (Arg, R), Lysine (Lys, K), Aspartic acid (Asp, D) and Glutamic acid (Glu, E). The polar ones (form hydrogen bonds as proton donors or acceptors) are: Glutamine (Gln, Q), Asparagine (Asn, N), Histidine (His, H), Serine (Ser, S), Threonine (Thr, T), Tyrosine (Tyr, Y) and Cysteine (Cys, C). The amphipathic (often found at the surface of proteins or lipid membranes, sometimes also classified as polar) are: Tryptophan (Trp, W), Tyrosine (Tyr, Y) and Methionine (Met, M). The hydrophobic (normally buried inside the protein core) are: Alanine (Ala, A), Isoleucine (Ile, I), Leucine (Leu, L), Methionine (Met, M), Phenylalanine (Phe, F), Valine (Val, V), Proline (Pro, P) and Glycine (Gly, G). . 3, 7, 8, 10, 11, 21, 79

arene These compounds are hydrocarbons that contain a benzene ring as a structural unit. In addition to benzene, other examples include toluene and naphthalene. They contains delocalized pi electrons between carbon atoms forming a circle.. 31

bi-ionic conditions two different salts with a common anion on each side of the channel. 38

C-terminal is the end of an amino acid chain (protein or polypeptide), terminated by a free carboxyl group (-COOH). 6

CAP1 is a protein involved in species-specific signalling pathways.. 7

cyclic peptides Cyclic peptides are polypeptide chains taking cyclic ring structure. The ring structure can be formed by linking one end of the peptide and the other with an amide bond, or other chemically stable bonds such as lactone, ether, thioether, disulfide, and so on. Peptide cyclization is a frequently

used strategy for the development of peptides with enhanced conformational stability (compared to their linear analogs).. iii

decorin is a proteoglycan associated with collagen and represents a common component of connective tissue and extracellular matrix. 7

gyration radius is defined as the root mean square distance from each atom of the protein to their centroid.. 13, 24

in silico is an expression meaning “performed on computer or via computer simulation” in reference to biological experiments.. 3

lead compounds are structures which serves as a starting point for chemical modifications in order to improve potency, selectivity, or pharmacokinetic parameters.. 3

Metropolis criterion consists in first determine the change in the energy of the trial state compared to the previous state, namely $\Delta E = E(q_{trial}) - E(q_j)$ and ii) If $\Delta E \leq 0$, accept the trial state, that is, $q_{j+1} = q_{trial}$). If $\Delta E > 0$, then accept the trial state with probability $e^{-\beta \Delta E}$. This is accomplished by using a call to the pseudorandom number generator. If rand() $< e^{-\beta \Delta E}$, then accept the trial state. Otherwise the trial state is rejected and then set the new state of the system equal to the previous state $q_{j+1} = q_j$).. 9–11

N-terminal is the start of a protein or polypeptide referring to the free amine group (-NH2) located at the end of a polypeptide. 6, 7

peptide is a short chain of amino acids (between 2 and 50 amino acids), linked by peptide bonds.. 3, 4, 7–11, 13–15, 18, 19, 21–24, 26–33

protein is a large biomolecule, consisting of one or more long chains of amino acid residues.. 3–5, 7–9, 12, 13, 17–19, 23–25, 27–29

RMSD is the measure of the average distance between the atoms (usually the backbone atoms) of superimposed proteins or conformations. RMSD $= \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2}$, where δ_i is the distance between atom i and either a reference structure or the mean position of the N equivalent atoms.. 13, 18, 22, 23, 25–28

ROR-1 is a membrane protein and belongs to the ROR subfamily of cell surface receptors, currently under investigation for its role in the metastasis of cancer cells. 7

SCN Thiocyanate (also known as rhodanide) is the anion SCN^-. It is the conjugate base of thiocyanic acid. Common derivatives include the colourless salts potassium thiocyanate and sodium thiocyanate. Organic compounds containing the functional group SCN are also called thiocyanates. . 37

selectivity filter forms the most constricted part of the conduction pathway in a channel or transporter. iii

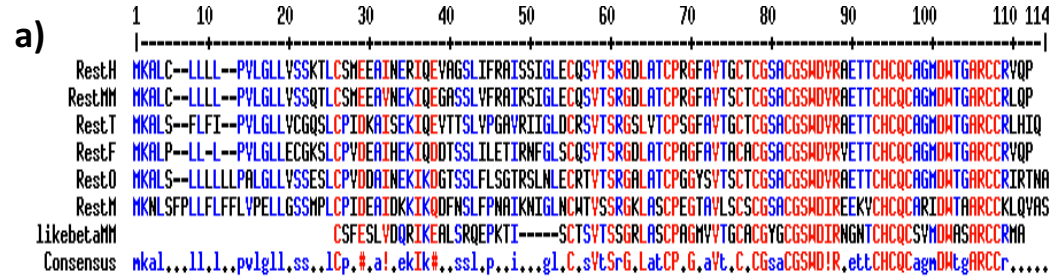
SH3 SRC Homology 3 Domain is a small protein domain of about 60 amino acid residues which is involved in the regulation of important cellular pathways, such as cell proliferation, migration and cytoskeletal modifications.. 8

TLR4 is a transmembrane protein, its activation leads to an intracellular signaling pathway and inflammatory cytokine production which is responsible for activating the innate immune system.. 7

Appendix A

A.1 Sequence alignement and binding site

a)



b)

CSFESLVDQRIKEALSRQEPKTISCTSVTSSGRLASCPAGMVVTGCACGYGCGSWDIRNGNTCHCQCSVMDWASARCCRMA

c)

Pocket	Binding Residues	Size (SA)
Site H	A39 G40 M41 V42 V43 T44 G45 I57 N59 G60 N61 R76	270.4

d)

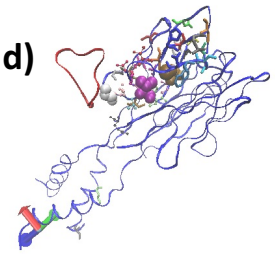


Fig. A.1. – Sequence alignement of resistin in diferent species and binding pocket prediction results. a) Primary sequence alignement of different RELMs. RestH (homo sapiens), RestMM (mus musculus), RestT (tupaia belangeri), restF (felis catus), restO (oryctolagus cuniculus), RestM (macata mulata) and likebetaMM (mus musculus, RELMβ). b) Sequence of the protein used during the simulation, red amino acids are the most conserved between all the species and brown amino acids are the predicted by RaptorX as part of the H binding site (in the globular area of the protein). c) List of all the amino acids predicted by Raptor X for H binding site. d) polyalanine-resistin starting structure where amino acids in "VDW" format correspond to the amino acid residues that coincide with the amino acids of the H binding site predicted and those mostly conserved between different species. Amino acids are represented with standard one letter code).

A.2 Post-design peptide selection

Mutation	Sequence	Site	$E_{pep-res}$	$E_{pep-alb}$	WS
453	CRGDYPFRPRHC	H	-30.22	-20.74	good
355	CLDSWMFDNRDC	H	-27.46	-17.68	good
441	CKNMRKENEDRC	H	-26.09	-16.23	good
482	CDNQRKPNEDRC	H	-26.03	-16.59	good
203	CQEWEPHFPWEC	H	-26.71	-15.78	good
485	CNFRKEDFRWEC	H	-24.63	-17.75	good
492	CRRNREDEHRYC	H	-25.26	-19.58	good
176	CRREDEYQWIIC	H	-24.15	-15.58	good
473	CHDHDDDRYMKC	H	-23.12	-13.05	good
170	CQHDDDHMGRHC	H	-22.79	-17.89	good
272	CREWKEHRWQKC	T	-26.32	-18.93	good
428	CRELEHHRWWDC	T	-25.83	-18.81	good
496	CMEKFKRKKEDC	T	-27.42	-16.43	good
407	CEFIDQGIKHDC	T	-23.99	-16.34	good
297	CSERWRRHGWYC	T	-31.11	-20.12	good
295	CIERWRRHGWYC	T	-30.69	-20.12	good
300	CRRDRLQRQDHC	T	-21.52	-17.91	good
238	CVKDEEFRKWEC	T	-20.19	-18.53	good
298	CQDWEWEYEWKC	T	-22.74	-16.39	good
280	CQHWEWEDEWKC	T	-22.28	-15.31	good

Fig. A.2. – Final set of peptides selected.

A.3 Protein-peptide interaction analysis

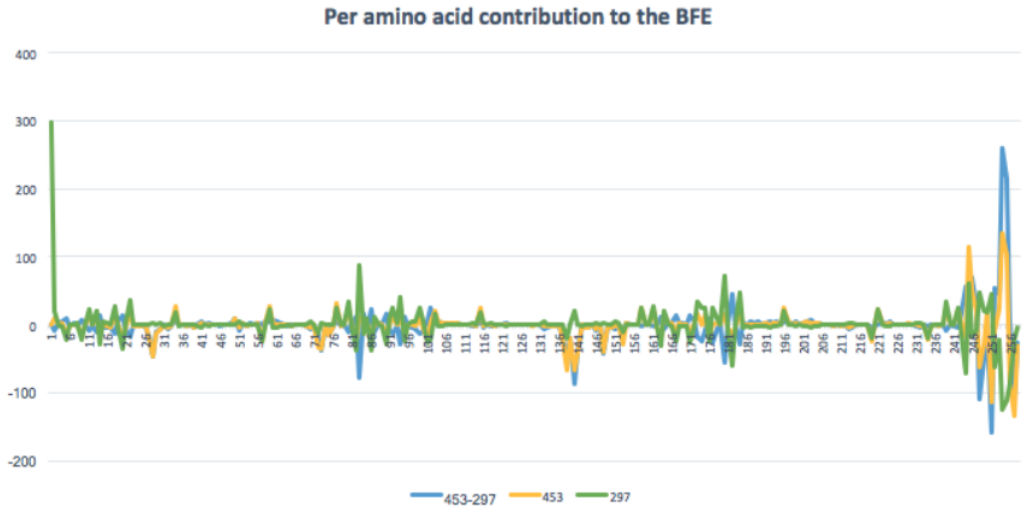


Fig. A.3. – Relative amino acids contributions to the BFE (kJ/mol). Energy for each amino acid was compared (453 – 297), a negative value implies a more favorable contribution of this amino acid for the system 453. If we consider only the relative binding energies with absolute values greater than 10 kJ/mol, there are 31 amino acids favorable to 453 and 15 unfavorable (most of the amino acids more favorable to 453 belong to the protein).

System	Interaction type	Number
453	Side chain HBs	9
297		6
453	Backbone HBs	5
297		6
453	Cation-arene	3
297		0
453	AA with reduced solvent exposure due to the ligand	15
297		12

Fig. A.4. – Number and type of interactions between protein-peptide for systems 453 and 297, according to MOE results. Hydrogen bonds (HBs) and amino acids (AA)

(Kcal/mol)	Van der Waals	Electrostatic	Polar solvation	SASA	SAV	BFE
453	-57.97	-50.27	129.29	-6.22	-50.089	-35.26
297	-75.74	-54.73	207.03	-9.5	-71.33	-4.28

Fig. A.5. – MMPBSA contributions to the BFE. Solvent Accessible Surface Area (SASA) and Solvent Accessible Volume (SAV).