Predicting Structural Determinants and Ligand Poses in Proteins Involved in Neurological Diseases: Bioinformatics and Molecular Simulation Studies

Thesis submitted for the degree of
Doctor Philosophiae

Candidate
Agata Kranjc

Supervisor
Prof. Paolo Carloni

February 2009

Contents

Introduction

Motivation of the present work

A disease is in a real sense a disordered physiology. Viruses, bacteria, mutations, etc. can trigger pathological events. Our body reacts with molecular, cellular and systemic responses that are the symptoms of disease. The ability to design drugs interfering with the progression of a disease comes with: (i) the knowledge of pathological, cellular and molecular mechanisms involved in the disease; (ii) the identification of the macromolecules (i.e. possible drug targets) involved in pathological pathways, their 3D structures and their functions. Drugs are able to bind strongly (and selectively) to target macromolecules involved in the pathological process, in order to interfere with their function. The biological activity of drug molecules is dependent on the three-dimensional arrangement of functional groups (the pharmacophores) which specifically bind to their target. Consequently, structural information of the target protein is crucial in design of new drugs.

NMR spectroscopy and X-ray crystallography have provided the structures of hundreds of proteins targeted in drug design [1, 2]. However, the gap between the number of proteins of known structures and the number of sequences is increasing [3]. Determining the structure of membrane proteins is especially difficult. In the case experimental structural information is not available, molecular modeling tools can be of great help. Particularly useful are comparative modeling approaches, that enable to construct protein 3D structures from their amino acids sequences, based on the 3D data of other similar proteins [4]. Since proteins are known to be in continuous motion between different conformations, it may be appropriate to combine such bioinformatics approaches with molecular dynamics (MD) simulations, which allow to identify a spectrum of protein conformations. With MD simulations important structural and functional features of the proteins, like ion binding or significant conformational changes can be observed at atomic detail.

Once drug target structures are known, molecular docking is used to locate drug candidates into proteins binding pockets. This consists in accommodating each candidate drug molecule from a large database in the binding pocket of the target, considered as rigid and without explicit treatment of water. The procedure is computationaly efficient. As binding of ligands may cause induced fit it is often necessary to study ligand-protein complexes by a more accurate and time-consuming technique. MD simulations allow for hydration and conformational fluctuations of the system. Using such simulations key ligand-target interactions, the ligand binding paths and its binding affinity can be elucidated. Therefore, the use of bioinformatics, docking and molecular dynamics may be combined in powerful way for structural prediction and drug design.

Following these approaches, in this thesis we focused on protein structural prediction and molecular docking. Most often such approaches have been compared with experimental data, such as site-directed mutagenesis as well as energetic or spectroscopic measurements. We have exploited the fact that in our

School a large neurobiology Sector is present. Therefore, the above computational approaches have been applied to systems of pharmacological relevance in neurobiological diseases. On one hand, (i) we have focused on structural predictions of Ca^{2+}-binding membrane proteins, for which at present, there is no structural information available. On the other hand, (ii) we have addressed challenging docking problems related to the prion protein, which is involved in transmissible spongiform encephalopathy or prion disease.

The studied systems are of pharmacological and neurobiological interest due to their involvement in severe degenerative diseases.

(i) Ca^{2+}-gated ionic channels

Ca^{2+}-gated channels, involved in the signaling in neurons and in the eye epithelium, are largely studied by neurobiologists. The malfunction of these channels causes different diseases. Here we focused on Ca^{2+}- and voltage-gated potassium channel (BK_{Ca}) [5] as well as on Ca^{2+}-gated Cl^- channel (bestrophin-1; Best1) [6]. To understand their functional properties, in particular Ca^{2+}binding, and consequently the pathological processes, structural information should be available. We have attempted at modeling the structural determinants of BK_{Ca} and Best1 Ca^{2+}-binding domains and to get an insight into Ca^{2+} binding.

BK_{Ca} are ubiquitously present in our body [7, 8]. They are responsible to return excited neuronal [9, 10] or muscular cells [11, 12] to the resting potential. The channels open upon an increase of intracellular Ca^{2+} ions along with depolarization [5]. The result is efflux of K^+ ions from the cell leading to a membrane hyperpolarization [13]. They are considered as putative targets for intervention in cardiovascular, respiratory, and urological diseases, along with epilepsy and ischemic reperfusion injury [7, 8]. Despite their function is well understood, the 3D structure and the details of Ca^{2+} binding, which leads to gating of the channel, are missing. To gain the structural information about domains involved in Ca^{2+} binding we modeled two BK_{Ca} cytoplasmic domains: RCK1 (Regulator of Conductance of K^+) and calcium bowl.

Best1, a recently discovered family of Cl^- channels, were suggested to be regulated by Ca^{2+} [6]. Mutations in Best1 leads to inherited macular dystrophy and consequently blindness known also as Best disease [14, 15]. Many functionally relevant mutations are located in the so called Asp-rich sequence at the C-terminal domain of Best1 [16, 6, 17, 18, 19]. Asp-rich domain shares significant sequence similarity with calcium bowl in BK_{Ca}[20, 21, 22]. Based on this analogy it has been suggested that the Asp-rich domain can bind Ca^{2+} ions [23]. We constructed structural models of the Asp-rich domain to understand its interactions with Ca^{2+} ions and to suggest mutational experiments, which can give an important knowledge of Best1 physiological function as well as of the pathophysiology of Best disease.

(ii) Prion protein

Prion diseases, or transmissible spongiform encephalopathies, are a group of fatal neurological diseases, that affect humans and animals [24]. Prion disease therapeutics can be targeted to the cellular form of prion protein (PrP^C), to its pathological form (PrP^{Sc}) or to the process of conversion between the two of them. Interfering with the conversion of PrP^C into PrP^{Sc} is a powerful strategy for pharmaceutical intervention [25]. Stabilizing the structure of PrP^C has the potential to remove the substrate for the pathogenesis and is applicable regardless of the disease etiology [26].

An excellent strategy to stabilize the PrP^C is based on using small fragments of antibodies. Recently, Fab D18 have been shown to permanently cure prion-infected cells by binding to the PrP^C [27, 28]. Unfortunately, Fab fragments cannot penetrate through the blood brain barrier what hampers their potential use in prion disease therapy. A promising solution for this problem are single-chain variable fragments (scFv) or mini-antibodies [29, 30, 31, 32]. Here, we investigated the anti-prion effects of D18scFv mini-antibody in collaboration with the experimental groups. We showed that anti-prion D18scFv is able to reduce PrP^{Sc} in infected cell culture. Then, we constructed a structural model of D18scFv. The PrP^C was docked to it and the key molecular interactions between the two proteins were defined. The identified interactions allowed us to propose a mechanism explaining how D18scFv reduced conversion and propagation of PrP^C.

Another strategy to fight prion diseases is to design ligands stabilizing the PrP^C. Unfortunately, however, *in silico* design of ligands targeting proteins undergoing fibrillation in neurodegenerative diseases, such as prion protein, is difficult due to the lack of deep binding pockets in these proteins and due to the paucity of 3D information of ligand-target complexes. Here, we established a computational approach that combines standard docking methods with molecular dynamics and free energy simulations in explicit solvent to address this issue in the context of prion protein. The protocol provides detailed information about the binding process, binding affinity and ligand-target interactions. Our approach emerges as a novel tool for studying ligands binding to the proteins with fibrillation properties.

Outline of the thesis

Part I presents the computational tools used in this work: the comparative modeling and molecular docking approaches along with molecular dynamics.

Part II presents structural predictions of Ca^{2+}-binding domains in Ca^{2+}-gated channels. A detailed description of the structure and function of these proteins can be found in the following Chapters. Chapter 4 focuses on human large conductance Ca^{2+}- and voltage-gated potassium channel (hBK_{Ca}). Bioinformatics approaches and MD simulations were used to construct models of two domains important for Ca^{2+} binding and channel gating, namely the Regulator of Conductance for K^+ (RCK1) and the so called calcium bowl. The relevance of these models for interpreting the available molecular biology data is then discussed.

Chapter 5 deals with bestrophins, a recently discovered family of Cl^- channels. Bestrophins feature a well conserved Asp-rich tract in their C-terminal part, which is homologous to Ca^{2+}-binding motifs in calcium bowl of hBK_{Ca}. Based on these considerations, we constructed homology models of human bestrophin-1 Asp-rich domain. MD simulations and free energy calculations were used to identify Asp and Glu residues binding Ca^{2+} and to predict effects of their mutations to Ala. My work, performed in collaboration with *C. Anselmi (SISSA/ISAS)*, was complemented by free energy calculations carried out by *F. Pietrucci (SISSA/ISAS*). Selected mutations were investigated by electrophysiological experiments performed by *Prof. A. Menini, J. Rievaj, F. W. Grillo, and A. Boccaccio* (*SISSA/ISAS*). The model of Asp-rich domain was then validated against experimental results.

Part III is devoted to the prion protein. In this Part, Chapter 6 presents *in vitro* studies of D18scFv anti-prion effects performed by *groups of Prof. C. Zurzolo (Institut Pasteur, Paris, France), Prof. G. Legname (SISSA/ISAS), L. Zentilin and M. Giacca (ICGEB, Trieste, Italy) and by Prof. S. B. Prusiner (Institute for Neurodegenerative Diseases, University of California San Francisco, U.S.A.)*

and structural prediction of a complex between the small antibody fragment (D18scFv) and PrP^C. The complex was modeled using bioinformatics approaches. Initially, the D18scFv fragment alone was modeled based on a similar antibody-fragment template and then docked with prion protein. Based on this, interactions relevant for the recognition between the two proteins and for the mechanism of action of D18scFv are discussed.

Chapter 7 describes a computational protocol for the design of ligands targeting cavity-less proteins, like most proteins involved in neurodegenerative diseases. Molecular docking methods are combined with MD simulations and free energy calculations using the metadynamics method [33, 34] to gain insights in ligand binding to such proteins, in our case to prion protein. We focused on a compound showing antiprion activity in vitro. Ligand-target interactions and ligand binding affinity as emerged by using our approach are compared with the available NMR data [35] and experimental constant of dissociation [35]. In this work, also other two students and one postdoc were involved beside myself, namely *S. Bongarzone, G. Rossetti and X. Biarnes (SISSA/ISAS).*

Finally, the conclusions are drawn in the last Chapter. The thesis closes with the List of publications and with the Acknowledgments.

Part I

Methods

Chapter 1

Comparative modeling

Comparative modeling, known also as homology modeling, is a computational procedure for generating 3D models for proteins of unknown structures (targets) based on sequence similarity to proteins of known structures (templates) [36]. It is based on the knowledge that protein structure is determined by its amino acid sequence [37] and that the protein topology is better conserved across evolution than the amino acid sequence. Therefore, the protein sequences that share detectable similarity with a template, even if they are distantly related, will fold into similar structures [38, 39] (Figure 1.1).

Even when the models are approximate because of low sequence identity (SI) with the template, different biologically useful information can be extracted. The models constructed from the templates with less than 30% of sequence identity might be used to refine NMR structures, to find binding/active sites by 3D motif searching or to predict approximate biochemical function [36]. Models based on 30-50% of SI allow to predict locations and volumes of binding sites as well as to provide insight on disease-linked mutations [36]. They are also useful in protein engineering. High quality models, based on more than 50% of SI can be used for ligand-protein or protein-protein docking.

There are four main steps in constructing models (for review see [36]): template identification, target template alignment, model construction and model evaluation (Figure 1.2).

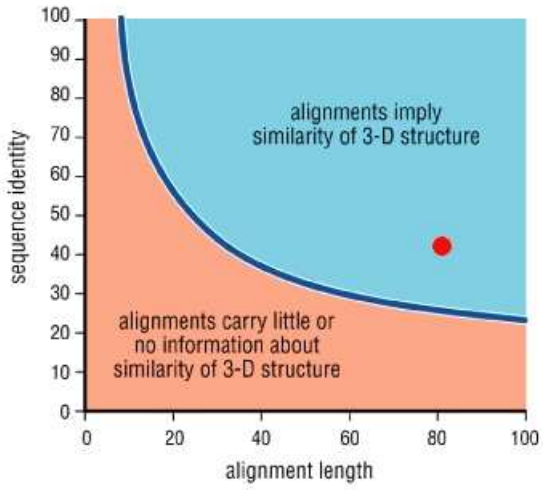

Figure 1.1: The two areas of sequence alignments. Two sequences are practically guaranteed to fold into the same structure if their length and percentage sequence identity fall into the region above blue line. For example, two sequences with 80 amino acids, 40% of which are identical, exhibit very high probability that they will fold into the same structure (red circle) (Figure taken from Protein Structure and Function by Gregory A. Petsko and Dagmar Ringe, New Science press, 1999-2004).

1.1 Template selection

The selection of the template is the initial step. The sequence of interest (the target sequence) is used as the query for search of proteins with known 3D structures (the templates) in the structure databases, like PDB [1, 2]. Different database search techniques are available on the web to facilitate identification of the templates [40, 41, 42]. The simplest method is based on serial pairwise sequence alignments that compares the target sequence with each of the database sequences independently and gives a set of possible homologs. Frequently used programs based on this method are BLAST [40, 43] and FASTA [42]. In the present work the BLAST method was used to search for templates in the PDB database [1, 2].

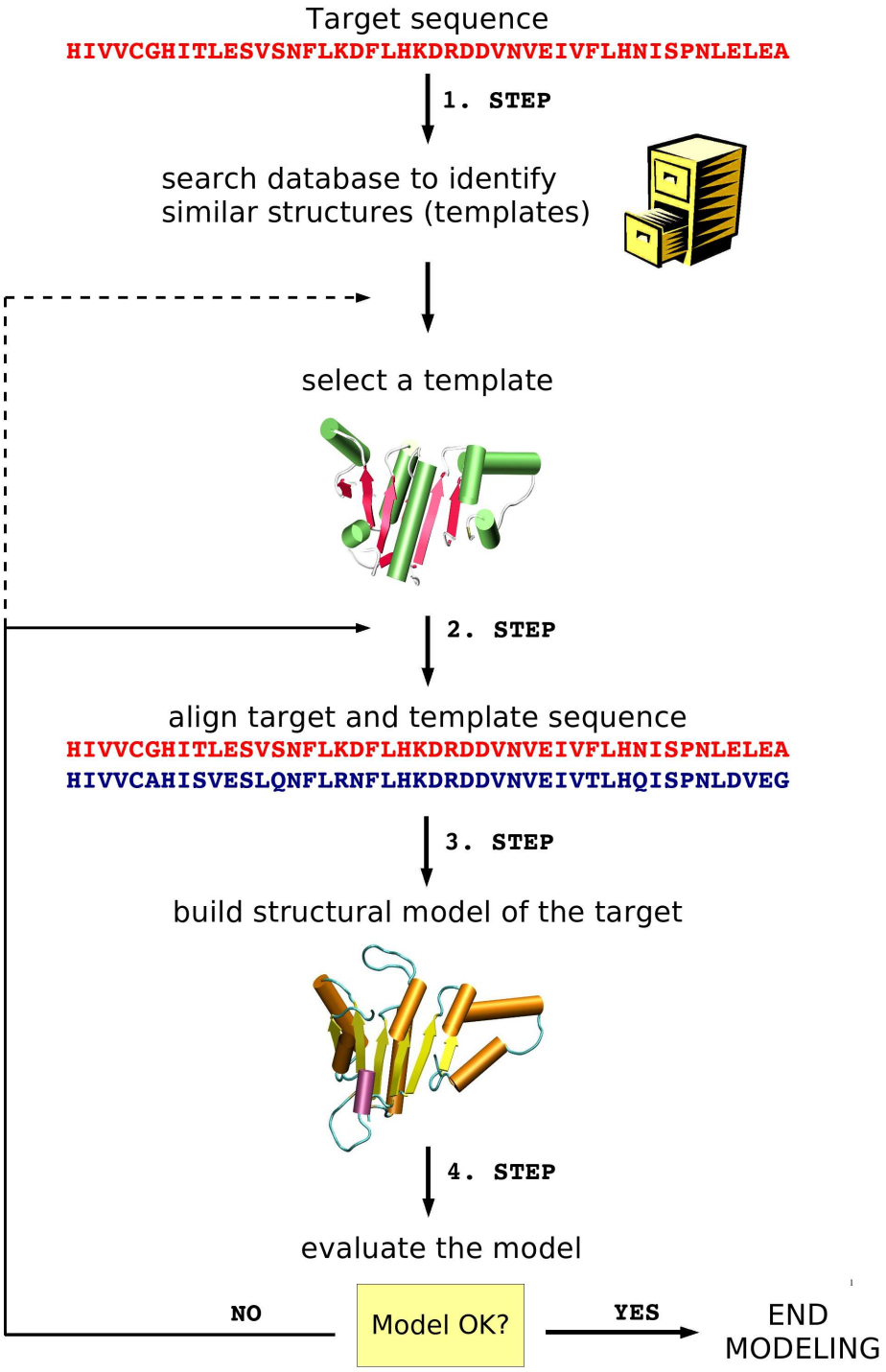

Figure 1.2: Steps in comparative protein structure modeling. See text for description.

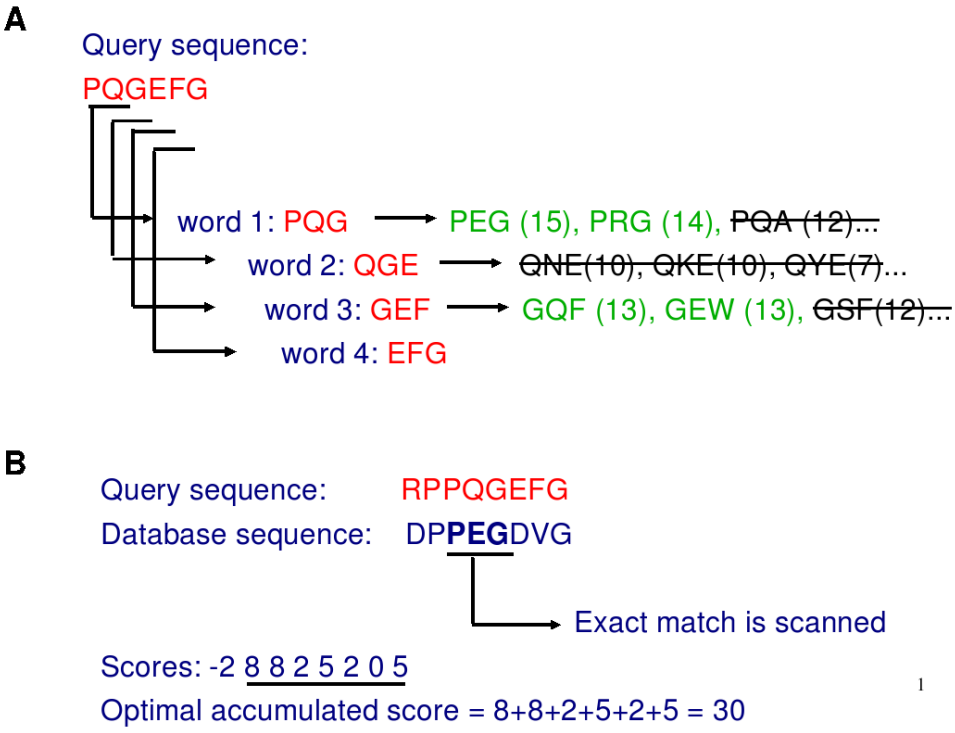

Figure 1.3: **A:** The BLAST algorithm parses protein sequences into 3 letter "words" and for every sequence in the query identifies database sequences with word matches that have a certain minimum score. **B:** When a high score match for a word is found, BLAST extends the sequence comparison between a query and the hit in both directions.

1.1.1 BLAST (Basic Local Alignment Search Tool)

The BLAST algorithm is a heuristic program, which means that it relies on some smart shortcuts to perform the search faster. BLAST performs "local" alignments. Most proteins in nature have functional domains often being repeated within the same protein as well as across different proteins from different species. The BLAST algorithm is tuned to find these domains or shorter stretches of sequence similarity. If instead BLAST would align the entire lengths of two sequences fewer similarities would be detected, especially with respect to domains and motifs.

The alignment in BLAST runs through different steps (Figure 1.3). First, all sequences with no relevance for the alignment process are filtered out, e.g. nucleotide sequences are hidden when protein sequence alignment is performed. Next, all possible 3-letter words in the query protein sequence are identified. Then, for each word from the query sequence a list of similar words is built and scored with substitution matrices (Figures 1.4). Words with score T$\geq$13 remain in the list of possible matching words, while those with lower scores are discarded (Figure 1.3A). The BLAST program scans the database sequences for the remaining high-scoring query words. All matches are scored again with the substitution matrices. If a high-score match is found in the database, it is used as seed for a possible ungapped alignment between the query and database sequences, by extending the seed in both directions (Figure 1.3B). The alignment is extending until the optimal accumulated score is lower than a cutoff score defined by the user.

The quality of a pairwise sequence alignment is evaluated by the "20x20 substitution matrix", which assigns a score for aligning any possible pair of residues. Pairs of amino acids which are identical or

physicochemically similar are scored higher than amino acids with very different properties (Figure 1.4). The two most used substitution matrices are PAM (Percentage of Acceptable point Mutations) [44] and BLOSUM (BLOck SUbstitution Matrices) [45], that are based on an analysis of the frequency with which a given amino acid is observed to be replaced by other amino acids among proteins with aligned sequences.

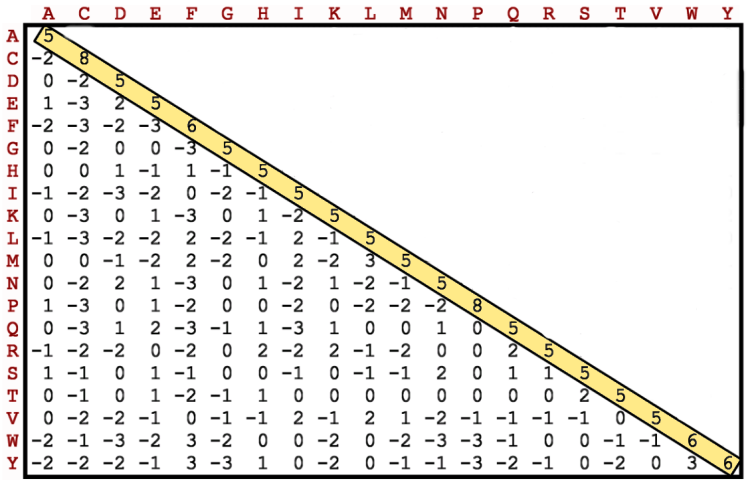

Figure 1.4: A symmetric substitution matrix used by alignment algorithms. The values along the diagonal are the highest, since they represent conserved residues. Positive scores are assigned also for the exchanges between residues with similar physicochemical properties (E.g. W –> F, Y; V –> I; K –> R), while scores for residues that differ in their chemical properties are negative.

Another widely used and more sensitive set of techniques (e.g. PSI-BLAST [43]) relies on multiple sequence comparison. PSI-BLAST iteratively expands the set of homologs of the target sequence. For a target sequence an initial set of homologs from a sequence database is collected, a weighted multiple alignment is made from the query sequence and its homologs, a position specific scoring matrix is constructed from the alignment, and the matrix is used to search the database for new homologs. These steps are repeated until no new homologs are found.

The advantage of multiple sequence comparison techniques is that they result in larger number of potential templates and they identify better templates for sequences that have only distant relationship to any solved structure. Once the list of template candidates is collected, it is necessary to choose the most appropriate template(s) for a given modeling problem. The following factors should be considered:

- higher sequence similarity between the target and the template leads to more accurate models
- multiple sequence comparison of subfamilies of the target and template protein family might identify the template closest to the target sequence
- take into account known features that determine protein structure, like conserved residues important for ligand binding, quaternary interactions, salt bridges etc.
- choose the template with the highest experimental quality (resolution, R-factor)

1.2 Target-template alignment

The accuracy of the models depends highly on the target-template sequence identity as well as on the alignment. When the target and template share more than 40% of sequence identity, the alignment is often correct and the models tend to be reliable. When the sequence identity between the target and template falls bellow 30%, the alignment is full of gaps and the predictive power of the modeling method is limited.

When a template is chosen a pairwise sequence alignment has to be performed using a specific method, because database searching methods do not give an optimal alignment. As a specialized alignment method often used is the ClustalW program [46, 47]. The basic multiple alignment algorithm consists of three main stages: (i) all pairs of sequences are aligned separately in order to calculate a distance matrix giving the divergence of each pair of sequences; (ii) a guide tree is calculated from the distance matrix; (iii) the sequences are progressively aligned according to the branching order in the guide tree.

The target-template alignment obtained as a result of ClustalW calculation is further improved manually based on structural data of the template. In particular all gaps are removed from secondary structure tracts, from buried regions and between two residues that are far away from each other. The multiple sequence alignment of the proteins belonging to the same family or subfamily can often improve the alignment due to the additional information about conserved residues important for structure and/or function of the protein. If the target and the template share very low sequence identity it is worth to build different models based on different alignments and then to choose best alignment according to the structural evaluation of the model. It is important to be aware that no current modeling procedure is able to recover from an incorrect alignment. Therefore it is worth to spend time trying to align the target and the template as good as possible. Multiple alignments are usually more reliable than pairwise alignments.

1.3 Model construction

Different programs [48, 49, 50] and web servers [51, 52] are nowadays available for building structural models of the proteins based on target-template alignment. All models presented in this work were modeled by program MODELLER [48]. It represents an automated approach to comparative protein modeling by satisfaction of spatial restraints. As input files the target-template alignment and spatial coordinates of the template are necessary. The target sequence is then aligned to the 3D structure of the template followed by the calculations of distance and dihedral angle constraints. The obtained spatial restraints are combined with stereochemical restraints on bond lengths, bond angles, dihedral angles and nonbonded contacts calculated by CHARMM22 [53] force field to an objective function. Finally the model is built by optimizing the objective function in Cartesian space. There is a possibility to add to the program-derived restraints, the other spatial restraints known from different experimental studies (NMR, fluorescence spectroscopy, site-directed mutagenesis...) what improves the accuracy of the models. Therefore modeling by satisfaction of spatial restraints is widely used and seems to be the most promising method in comparative modeling.

1.4 Evaluation of the models

The first step in the model evaluation is a check of the overall fold and structural features of the models predicted with homology modeling. The model is compared to the fold of the template. Obviously, there must be no knots in the topology, buried residues should be mostly hydrophobic. The fold of a model can be assessed by available experimental data and by conservation of structurally and/or functionally important residues.

Then, the Ramachandran plot is visualized and RMSD between the target an the template is calculated. The main chain conformations have to be in the acceptable regions of the Ramachandran plot. RMSD depend on the sequence identity, and can be is low as 1 Å if the SI is high and as high as 4-5 Å if the SI is less than 30%.

Furthermore, it is important that the model has a good stereochemistry. One of the programs that provide stereochemical analysis is PROCHECK [54, 55], that gives an output with the data of bond lengths, bond angles, peptide bond and side-chain conformations. PROCHECK program compares the values of these parameters in the modeled structure to the known parameters of the high resolution protein structure from which the model was constructed. Therefore, large deviations from most likely values are interpreted as strong indicators of errors in the model. If the errors are observed, the structure is introduced to the further refinement.

Chapter 2

Molecular docking

Molecular docking is a computational procedure that predicts binding mode of a ligand in its target protein. This is achieved by minimizing a scoring function which describes the interactions between ligand and target with respect to the atomic positions of the to moieties.

In this work, GOLD 3.0.1 [56, 57, 58] and AutoDock 3.0 [59] programs were used to predict ligand-protein interactions, while HADDOCK 2.0 program [60, 61] was utilized for building protein-protein complexes.

2.1 Ligand-protein docking

GOLD (Genetic Optimization for Ligand Docking) [57, 56, 62] is a program used for docking of small flexible compounds to the protein binding site. GOLD uses an island-based GA search strategy and includes rotational flexibility for selected receptor hydrogen along with full ligand flexibility. It has two implemented scoring functions, GoldScore and ChemScore, and a possibility for user defined scoring function. In this work GOLD 3.0.1 was used.

AutoDock [59] uses a genetic algorithm as a global optimizer combined with energy minimization as a local search method. The ligand is flexible, while the receptor is rigid.

2.1.1 Search Algorithms

The search of a pose of a ligand in a docking problem is addressed in two essentially different approaches: (1) a full solution space search (2) a gradual guided progression through solution space. The first scans the entire solution space in a predefined systematic manner [63]. In contrast, the second either scans only part of the solution space in a partially random and partially criteria-guided manner, or generates fitting solutions [63]. The second approach consists mainly of Monte Carlo (MC), simulated annealing, molecular dynamics (MD), and evolutionary algorithms such as genetic algorithms (GA) and Tabu search. The two docking programs used in this work, GOLD and AUTODOCK, search for optimal ligand binding modes inside a specific binding pocket using a genetic algorithm [64, 65].

2.1.1.1 Genetic Algorithms

Genetic algorithms (GA) [64, 65] are based on the language of natural genetics and biological evolution. They search for possible ligand binding modes by representing the ligand conformations in a modular way, using operations similar to mutations and crossover. The quality of the results is a function of the starting genes, the number of evolutionary events, i.e., the mutations and crossover, and the scoring function to pick the more favorable conformers.

The GA begins with an initial population, that is a set of chromosomes (ligand binding modes, chosen randomly). Chromosomes are defined by genes (variables) corresponding to the ligand translation (x, y, z coordinates of the center of mass), ligand orientation (rotation angles) and ligand conformation (torsion angle of each ligand rotatable bond). The population goes through a process of fitness evaluation: each chromosome is assigned a score based on a function which approximately estimate the free energy of binding (see Sections 2.1.2.1 and 2.1.2.2).

Once the initial population is evaluated, two chromosomes are selected as parents and starting from them a new population is built. A probability to become parent chromosome grows with the fitness score. The offspring chromosomes are obtained by a crossover and/or mutation processes on the parent chromosomes (Figure 2.1). In crossover the chromosomes of the parents are broken into two pieces at the same gene positions, then the first part of one chromosome is combined with the second part of the other chromosome, and *viceversa*, resulting in two children. Mutations randomly modify one or more selected gene(s) in the offspring chromosomes. The new children replace their parents in the population that enters a new run of algorithm. The whole cycle is repeated until some number of generations are completed or until some condition (RMSD, δG) is satisfied.

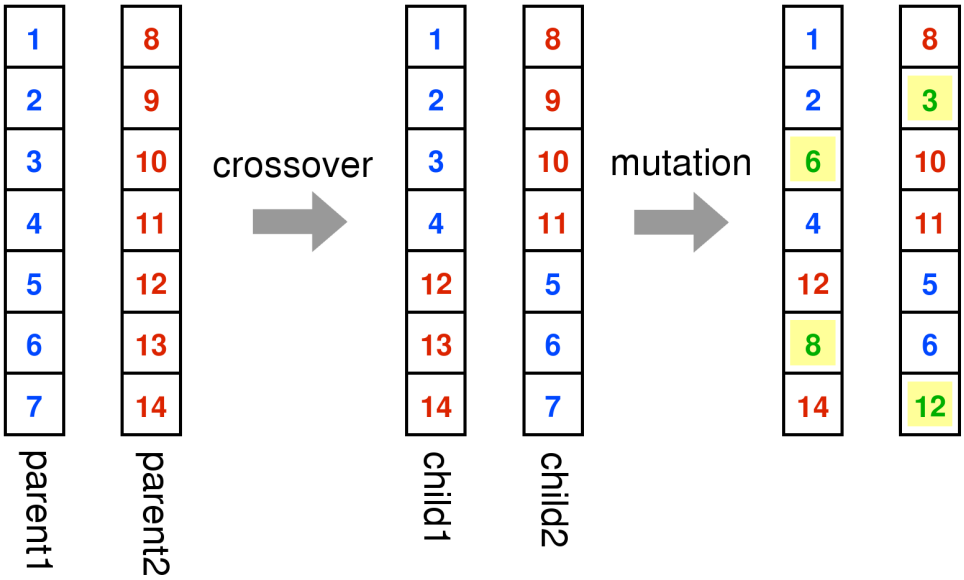

Figure 2.1: Scheme of crossover and mutation process. Let us assume two parent chromosomes: parent1 and parent2. After the crossover, the resulting children have first part of genes from parent1 and second part of genes from parent2. Mutations (highlighted in yellow) that may follow the crossover mutate a selected gene(s) in the offspring chromosome (e.g. in child1 genes 3 and 13 were mutated to genes 6 and 8).

2.1.1.2 Island-based genetic algorithm

In GOLD a so-called island-based genetic algorithm is employed (Figure 2.2). It means that rather than manipulate only one large population of chromosomes (as described in Section 2.1.1.1), several subpopulations (i.e. islands) are considered and individual chromosomes can migrate among them. This feature improves efficiency of the search. In addition, information about hydrogen bonds between ligand and receptor is also encoded in the chromosome. The ligand-receptor hydrogen bonds are matched with a least squares fitting protocol to maximize the number of inter-molecular hydrogen bonds.

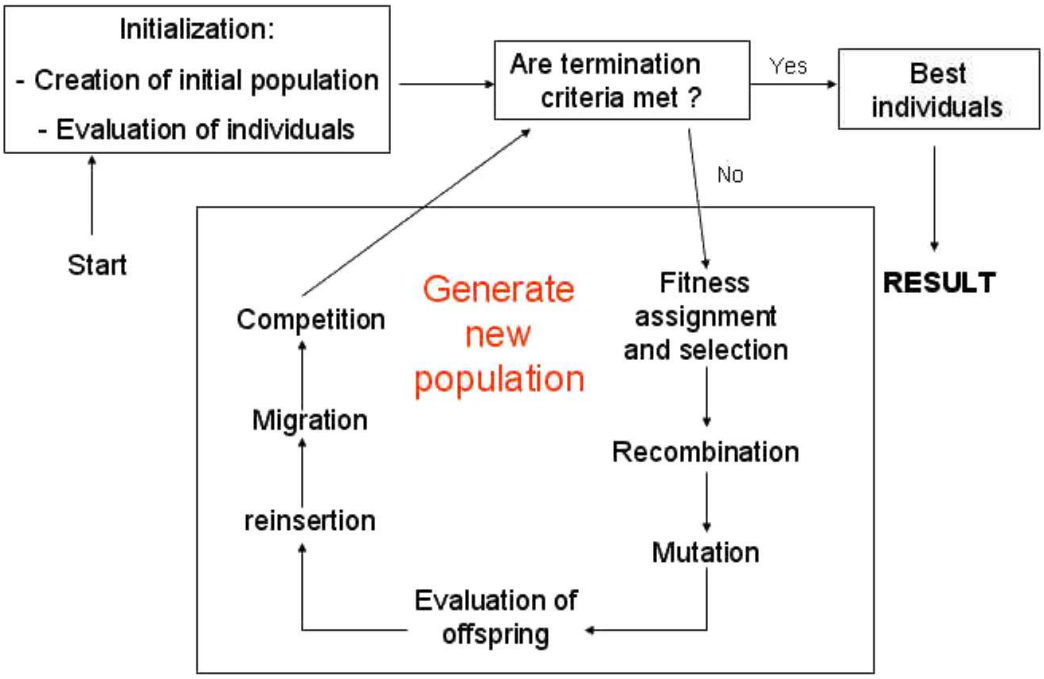

Figure 2.2: Scheme of genetic algorithm used in GOLD program

2.1.1.3 Lamarckian genetic algorithm

The Autodock uses a genetic algorithm (as described in section 2.1.1.1) in which is implemented a local search method that allows to minimize the scoring function of selected individuals. The optimized atomic coordinates (phenotype) are stored back in the chromosome. The new chromosome enters then into the new iteration of crossover and mutation of the genetic algorithm. Due to the transfer of information from phenotype to chromosome this algorithm is called Lamarckian.

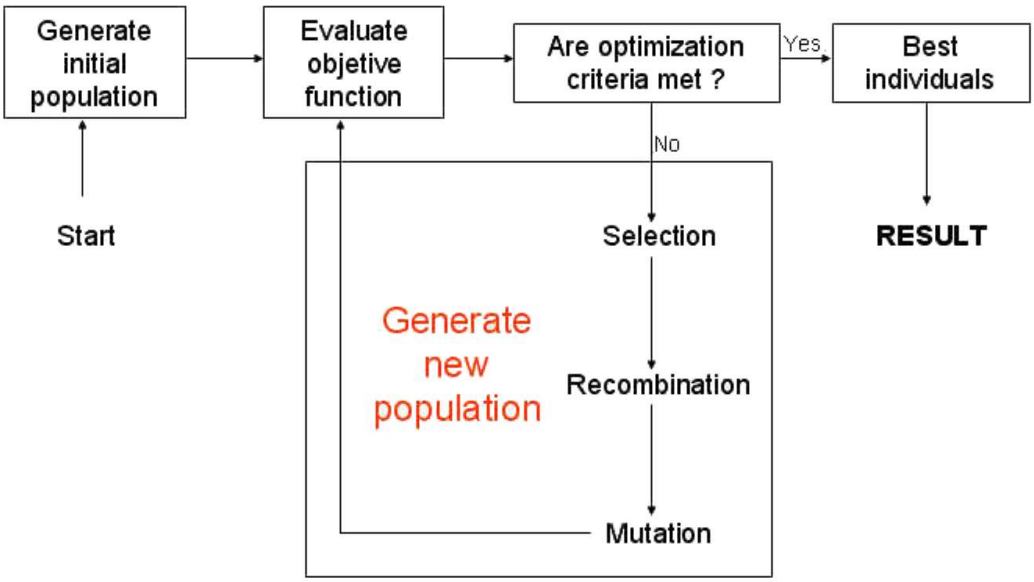

Figure 2.3: Scheme of genetic algorithm used in Autodock program.

2.1.2 Ranking of the solutions - Scoring functions

Scoring functions estimate the energetics associated with ligands binding to their target. The purpose of the scoring function is to discriminate between correct native solutions with low RMSD from the crystal and other docked complexes within a reasonable computation time. Although scoring functions may be formally defined as free energies, here they are only used to predict ligand binding poses.

2.1.2.1 GOLD scoring functions

The GOLD program has implemented two scoring functions, ChemScore [62] and GoldScore [57, 56]. The *ChemScore scoring function* was parametrized against experimental binding affinities data and gives as a result estimated binding affinity of the docked ligand. The ChemScore scoring function estimates the total binding free energy as a sum of different components (hydrogen bonding, metal and lipophilic interactions and loss of conformational entropy of the ligand upon binding):

$$\Delta G_{bind} = \Delta G_0 + \Delta G_{hbond} + \Delta G_{metal} + \Delta G_{lipo} + \Delta G_{rot}$$

The final ChemScore value is obtained by adding in a clash penalty and internal torsion terms, which militate against close contacts in docking and poor internal conformations. Covalent and constraint scores may also be included.

$$ChemScore = \Delta G_{bind} + P_{clash} + c_{internal} P_{internal} + (c_{covalent} P_{covalent} + P_{constraint})$$

The *GoldScore scoring function* is calculated as a sum of the protein-ligand hydrogen-bond energy, the protein-ligand van der Waals energy, the ligand internal van der Waals energy and ligand torsional strain energy. The contribution of ligand intramolecular hydrogen bonds can be added. ChemScore and GoldScore are about equally reliable. However, based on our experience, ChemScore performs better when lipophilic interaction between the ligand and receptor are prevalent, while GoldScore gives better results for the complexes between the polar ligand and receptor (hydrogen bond interactions).

2.1.2.2 Autodock scoring function

In AutoDock the implemented scoring function is defined as an empirical binding free energy function:

$$\Delta G = \Delta G_{vdW} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + \Delta G_{hbond} \sum_{i,j} E(\phi) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) \tag{2.1}$$

$$+\Delta G_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij})\, r_{ij}} + \Delta G_{tor} N_{tor} + \Delta G_{sol} \sum_{i,j} \left(S_i V_j + S_j V_i \right) e^{\left(-r_{ij}^2 / 2\sigma^2\right)}$$

The summations are performed over all pairs of ligand atoms, i, and protein atoms, j, in addition to all pairs of atoms in the ligand separated by three or more bonds. r_{ij} is the distance between the atoms, ϕ is the H-bond angle, and q_i is the electrostatic charge of atom i. All five ΔG terms on the right hand side are coefficients empirically determined using linear regression analysis from a set of thirty protein-ligand complexes with known binding constants. The first three terms are *in vacuo* interaction terms, namely Lennard-Jones dispersion, directional hydrogen bond and screened Coulomb electrostatic potential. ΔG_{tor} is a measure of the unfavorable entropy of ligand binding due to the restriction of conformational degrees of freedom, and N_{tor} is the number of sp^3 bonds in the ligand. The last term approximately accounts for the desolvation free energy upon ligand binding. For each atom in the ligand, fragmental volumes of surrounding protein atoms (V_j) are weighted by an exponential function and then summed, evaluating the percentage of volume around the ligand atom that is occupied by protein atoms. This percentage is then weighted by the atomic solvation parameter of the ligand atom (S_i) to give the desolvation energy.

2.2 Protein-protein docking

HADDOCK 2.0 (High Ambiguity Driven protein-protein Docking) [60, 61] is a docking algorithm which can handle protein-protein docking. The HADDOCK docking protocol has three steps:

(i) randomization of orientations and rigid body energy minimization

(ii) semirigid simulated annealing in torsion angle space

(iii) final refinement in Cartesian space with explicit solvent.

In the randomization step both molecules, A and B, are far apart and do randomly rotate around their center of mass. Then rigid body energy minimization is performed: first both macromolecules are allowed to rotate to minimize the intermolecular energy function. Then both translations and

rotations are allowed, and the two molecules are docked by rigid body energy minimization. The second step consists of simulated annealing refinements in which the side chains and the backbone at the interface are allowed to move to allow for some conformational rearrangements. The resulting structures undergo then steepest decent energy minimization. The final step consists of gentle MD simulation refinement in a shell of TIP3P [66] waters.

HADDOCK uses a physically-based interaction force fields to drive the docking process. The interaction energies are evaluated by full electrostatic and van der Waals energy terms with an 8.5 Å distance cutoff using the OPLS [67] force field. The final structures are clustered using the pair-wise backbone RMSD at the interface. Clusters are analyzed and ranked according to their average interaction energies (sum of E_{elec}, E_{vdW}, E_{AIR}) and their average buried surface area.

To get accurate docking predictions, HADDOCK takes as input also biochemical and/or biophysical experimental data, like mutagenesis data or information about chemical shift perturbation upon binding. Interface interactions are defined by Ambiguous Interaction Restraints (AIR) that involve the active and passive residues. Active residues are for example those detected by mutagenesis to be involved in protein-protein interactions or those showing a significant chemical shift perturbation upon complex formation. Passive residues are their closest solvent accessible neighbors. AIR are defined as an ambiguous intermolecular distance d_{iAB} between each active residue from molecule A and both active and passive residues from molecule B (and inversely for molecule B):

$$d_{iAB}^{eff} = (\sum_{m_iA=1}^{N_{atoms}} \sum_{k=1}^{N_{res}B} \sum_{n_kB=1}^{N_{atoms}} \frac{1}{d_{m_iAn_kB}^6})^{-\frac{1}{6}}$$

where N_{atoms} is the number of all atoms in selected residue and $N_{resB}$is the sum of active and passive residues in the molecule B. The AIR are satisfied every time the distance of an atom m from active residue i of molecule A and an atom n of passive or active residue k of molecule B is shorter than 2 Å.

Chapter 3

Molecular dynamics

Molecular dynamics (MD) is a computer simulation technique that enables the theoretical study of (biological) molecules. Given the initial positions of the atoms in the macromolecule and in the solvent, MD calculates the time evolution of the system. MD simulations provide detailed information on the fluctuations and conformational changes of proteins and nucleic acids. These methods are now routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes. They are also used in the determination of structures from x-ray crystallography and from NMR experiments.

MD is based on the Born-Oppenheimer approximation, i.e. the nuclei can be treated as classical particles and the electronic degrees of freedom can be integrated out. This approximation relies on the large difference between the mass of the nuclei (heavy) and that of the electrons (light). Under these greatly simplifying assumptions, atoms are treated as point masses and the dynamics of the system can be described by Newton's second law:

$$\vec{F}_i(t) = \frac{d^2\vec{r}_i(t)}{dt^2} \cdot m_i = -\frac{dU}{d\vec{r}_i} \tag{3.1}$$

where is $\vec{r}_i$ the position of the i^{th} atom, $\vec{F}_i$ the force acting on it, and U is the potential energy of the system (i.e. the force field). In this way, knowing the initial structure (by experiments or by computer modeling) and providing an initial velocity distribution consistent with the temperature simulated, it is possible to compute the time evolution of the system. Average values of several physico-chemical properties can be evaluated from the resulting trajectory [68].

The MD run consists of several steps:

1. Define the initial positions $\vec{r}_i$ and velocities $\vec{v}_i$ of all atoms, the force field U, and the boundary conditions P, T.
2. Compute the force $\vec{F}_i$ acting on each atom.
3. Solve Eq. 3.1 and update the configuration of the system.
4. Write the new positions, velocities, energies, ... , etc.
5. Back to point 2

In this work, the Gromacs simulation package [69] has been used. In the following are described the main MD algorithms and the form of the force field.

3.1 Equations of motion

MD simulations consist in the numerical, step-by-step solution of the classical equations of motion Eq. 3.1. The integration time step Δt is chosen at the beginning of the simulation and it remains unchanged during the run. The time step must be small enough to allow describing the fastest motions of the system: $\Delta t \leq 0.5$ fs is normally used when bonds involving hydrogens are allowed to stretch. Bond stretching is of little interest in most cases, therefore bonds are constrained to their equilibrium lengths using the LINCS algorithm [70]. In this way a larger$\Delta t \leq 1.5/2$ fs can be employed while still retaining a good accuracy. Nowadays it is common to simulate biological systems of ~ 50000 atoms for ~ 100 ns.

The most used algorithms for the integration of the equations of motion are the so called Verlet [71] and Leap-Frog algorithms [72]. In both algorithms the positions of each atom are expressed by Taylor expansions for the time increment Δt. The Leap-Frog algorithm however provides more accurate velocities [68] and it has been used in this thesis. Updated positions and velocities are computed in the following way:

$$\vec{v}_i(t + \Delta t/2) = \vec{v}_i(t - \Delta t/2) + \frac{\vec{F}_i(t)}{m_i} \cdot \Delta t \tag{3.2}$$

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i(t + \Delta t/2) \cdot \Delta t \tag{3.3}$$

which makes use of the positions and forces at time t and velocities at time $t + \Delta t/2$. The update of positions and that of velocities leaps each other: first, the velocities are calculated at half time step, then these are used to calculate the positions at one time step. A disadvantage of this algorithm is that velocities are not calculated at the same time as positions, but this can be solved by the following approximation:

$$\vec{v}_i(t) = \frac{\vec{v}_i(t - \Delta t/2) + \vec{v}_i(t + \Delta t/2)}{2} \tag{3.4}$$

In this way, the kinetic and potential energies can be summed at the same time t to compute the total energy.

3.2 Force field

The potential energy U in Eq. 3.1 is a function of the atomic coordinates of all atoms and determines the forces among them. In MD simulations of biological systems usually empirical force fields U are employed, whose parameters are obtained from gas-phase experiments and accurate quantum mechanical calculations [73]. In this work the Amber parm99 force field was used [74].

The force fields for biological systems have the following typical form:

$$U = E_{bonds} + E_{angles} + E_{dihedrals} + E_{vdw} + E_{elect} \tag{3.5}$$

Different terms are summed which correspond to bond distances stretching (E_{bonds}), bond angles bending (E_{angles}), bond dihedral or torsion angle ($E_{dihedrals}$), van der Waals potential (E_{vdw}) and electrostatic potential (E_{elect}). The first three terms are considered to be the intramolecular bonding interactions, as they involve multiplets of atoms connected by chemical bonds. The last two terms represent the non-bonded interactions between atoms. The specific form of each of these terms is shown below:

$$E_{bonds} = \sum_{bonds} \frac{1}{2} k_r (r - r_{eq})^2 \tag{3.6}$$

$$E_{angles} = \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_{eq})^2 \tag{3.7}$$

$$E_{dihedrals} = \sum_{dihedrals} \frac{1}{2} V_n (1 + \cos(n\varphi) - \gamma) \tag{3.8}$$

$$E_{vdw} = \sum_{i<j}^{atoms} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) \tag{3.9}$$

$$E_{elec} = \sum_{i<j}^{atoms} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{3.10}$$

In the bonding interactions, the bond stretching and angle bending terms are represented as harmonic energy functions where r_{eq} and θ_{eq} refer to equilibrium bond lengths and angles, k_r and k_θ being the vibrational constants. The equilibrium values of the bond and angle parameters are usually derived from structural databases, while the force constants are derived from infrared spectroscopy. In the third term, V_n is the torsional barrier corresponding to the n^{th} barrier of a given torsional angle with phase γ: dihedral parameters are calibrated on small model compounds, comparing the energies with those obtained by quantum chemical calculations. Improper dihedral angles are added to take into account quantum effects that are not present in U as, for example, to preserve planarity in aromatic rings.

In the non-bonded interactions, the van der Waals potential is described by a Lennard-Jones potential, containing an attractive and a repulsive term: the parameters are defined so as to reproduce chemico-physical properties (e.g., densities, enthalpies of vaporization, solvation free-energies) of organic liquids. The electrostatic energy is evaluated by assuming the dielectric constant ε equal to 1 (vacuum value), and using the restrained electrostatic potential model [75] to define partial atomic point charges: in this model, charges are assigned to the atom-centered points so as to fit the electrostatic potential derived from quantum chemistry calculations for a set of small representative molecules. Van der Waals and electrostatic interactions are calculated between atoms belonging to different molecules or for atoms in the same molecule separated by at least three bonds. In principle, the non-bonded interactions involve all pairs of atoms in the system, and they are the most expensive part of a MD calculation. In practical applications, however, the number of calculated interactions are limited by a predefined cutoff distance, so the non-bonded terms are calculated only between atoms separated by a distance not larger than the cutoff. For the van der Waals potential, this truncation introduce only a small error in the energy. This is not the case for the electrostatic potential, because the Coulomb interaction between charges q_i and q_j decays slowly with distance. Hence it can not be truncated, but when periodic boundary conditions are used, it is computed with efficient schemes such as Particle Mesh Ewald [76] (see Section 3.5) in conjunction with periodic boundary conditions (Section 3.3), which approximate the exact result to an acceptable error similar to the error in the van der Waals potential.

3.3 Periodic boundary conditions

MD simulations are usually performed under periodic boundary conditions (PBC), to minimize boundary effects and to mimic the presence of the "bulk" environment. In this approach, the system is surrounded with replicas of itself in all directions, to yield an infinite periodic lattice of identical cells.

When a particle moves in the central cell, its periodic image in every other cell moves accordingly (Figure 3.1). As one molecule leaves the central cell, its periodic image enters from the opposite side.

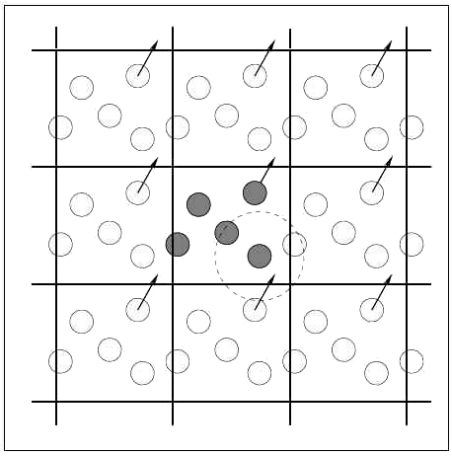

Figure 3.1: Periodic boundary conditions. As a particle moves out of the simulation box, an image particle moves in to replace it. In calculating particle interactions within the cutoff range, both real and image neighbors are included.

The PBC are taken into account only in the calculation of non-bonded interactions between atoms belonging to different molecules, and, if the potential range is not too long (the cutoff radius must not exceed half of the box side), the minimum image convention is adopted. This means that each atom interacts only with the nearest atom or image in the periodic array.

When a macromolecule is simulated in solution, this convention is not appropriate, since in principle the macromolecule should not interact with its periodic images. This means that the length of each box vector must exceed the length of the macromolecule in the direction of that edge plus twice the cutoff radius.

3.4 Neighbors list

Computing the non-bonded contribution to the interatomic forces in an MD simulation involves, in principle, a large number of pairwise calculations. Let us assume that the interaction potentials are short ranged, i.e. they vanish for $r_{ij} > r_{cut}$. When this happens, the program skips the expensive force calculation, and considers the next candidate pair of atoms. Nevertheless, the time to examine all pair separations is proportional to the number of distinct pairs and this still consumes a lot of computer time.

Some economies result from the use of lists of nearby pairs of atoms. The potential cutoff sphere of radius r_{cut}, around a particular atom is surrounded by a 'skin', to give a larger sphere of radius r_{list} as shown in Figure 3.2. At the first step in a simulation, a list is constructed of all the neighbors of each atom, for which the pair separation is within r_{list}. Over the next few MD time steps, only pairs appearing in the list are considered for force calculation. From time to time the list is reconstructed: it is important to do this before any unlisted pairs have crossed the safety zone and come within interaction range.

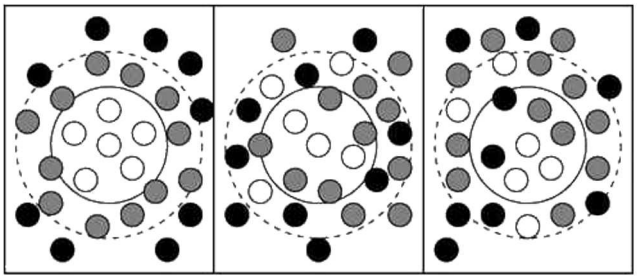

Figure 3.2: The potential cutoff range (r_{cut}, solid circle), and the list range (r_{list}, dashed circle), are indicated. The list must be reconstructed before particles originally outside the list range (black) have penetrated the potential cutoff sphere.

3.5 Long range forces

Long range forces are usually defined as those in which the spatial interaction falls off no faster than r^{-d} where d is the dimensionality of the system ($d = 3$ in common situations). In this category are the charge-charge interactions between ions ($V(r) \approx r^{-1}$). These interactions are problematic for the simulations, since their range is greater then half the box length. There are several ways to handle this problem.

3.5.1 Electrostatics

The Ewald sum [77] is a technique for efficiently summing the interactions among ions and all their periodic images. The potential energy can be written as

$$V = \frac{1}{2}\sum_{n}{}' \left(\frac{1}{4\pi\varepsilon_0} \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{q_i q_j}{r_{ij}, \mathrm{n}} \right) \tag{3.11}$$

Where q_i and q_j are the charges, n is the box index, r_{ij} is the distance between the charges and the prime indicates that we omit $i = j$ for $\mathrm{n} = 0$. Due to the long-range nature of the potential, this sum is conditionally convergent, and converges very slowly. Therefore, a very large number of images is required to achieve a reliable estimate of V. The idea behind the Ewald method is to surround every point charge by a charge distribution of equal magnitude and opposite sign ϱ_-, which spreads out radially from the charge. This distribution is conveniently taken to be Gaussian

$$\varrho_{\mathrm{i}}^{\mathrm{G}}(r) = q_i \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} \exp\left(-\alpha \left|r_i + nL\right|^2\right) \tag{3.12}$$

here α is an arbitrary parameter which does not determine the final result, but that can be adjusted to optimize the speed of convergence. In this way an efficient screening is performed, so that interactions rapidly go to 0 and direct summation is possible. This extra distribution acts like an ionic atmosphere, to screen the interaction between neighboring charges. The screened interactions are now short-ranged, and the total screened potential is calculated summing over all molecules. A charge distribution of the same sign as the original charge, and the same shape as the distribution $\varrho_i^G(r)$ is also added. This canceling distribution reduces the overall potential due to the original set of charges. In order to exclude self-interactions the contributions of these three charge densities should not be evaluated in r_i. However, it is convenient to keep self-interactions due to the canceling charge distribution ϱ_+,

since ϱ_+ is in this way periodic and can be represented as a rapidly converging Fourier sum. The spurious self-interaction can be easily subtracted separately. The canceling distribution is summed in the reciprocal space. In other words, the Fourier transforms of this distribution are added, and the total transformed back into real space. Thus, the total charge distribution of the system $\varrho(r)$ may be rewritten as:

$$\begin{aligned} \rho(r) &= \sum_i q_i \delta\left(r - r_i + nL\right) \\ \rho(r) &= \sum_i \left(q_i \delta\left(r - r_i + nL\right) - \rho_i^G(r)\right) + \sum_i \rho_i^G(r) \end{aligned} \tag{3.13}$$

Where the first sum (which only produce short ranged potentials) is calculated in the direct space, while the second is calculated in the reciprocal space:

$$V = \left(\varepsilon_{dir} - \varepsilon_{self} + \varepsilon_{rec}\right) \tag{3.14}$$

$$\varepsilon_{dir} = \frac{1}{2} \sum_n {}' \sum_{ij}^N \frac{q_i q_j}{|r_{ij}+nL|} erfc\left(\sqrt{\alpha}\,|r_{ij}+nL|\right) \tag{3.15}$$

$$\varepsilon_{self} = \sqrt{\frac{\alpha}{\pi}} \sum_{i=1}^N q_1^2 \tag{3.16}$$

$$\varepsilon_{rec} = \frac{2\pi}{V} \sum_{k\neq 0} \sum_{ij} \frac{\exp\left(\frac{k^2}{4\alpha}\right)}{k^2} q_i q_j \exp\left(-ik \cdot \left(r - r_j\right)\right) \tag{3.17}$$

where $erfc = \pi^{-\frac{1}{2}} \int_r^\infty \exp\left(-r^2\right) dx$ is the complementary error function.

ε_{dir} is very similar to Equation 3.11, although the long ranged $\frac{1}{r}$ function is here substituted by the short ranged $\frac{erfc(r)}{r}$: as a result, the interaction vanishes above a cutoff roughly equal to $\alpha^{-\frac{1}{2}}$, and for every i and j the interaction can be approximated by only one periodic image term. Typically, α is chosen such that such truncation error is of the order of $10^{-5}/10^{-6}$ of ε_{dir}. ε_{self} is the self interaction of the Gaussian charge distributions: it must be subtracted from the total, as the reciprocal space term ε_{rec} contains it, albeit it is a constant number, not depending on the atomic configuration. ε_{rec} is a sum over an infinite number of terms, but the factor $\frac{k^2}{4\alpha}$ ensures a fast convergence in the reciprocal space, and normally no more than $5/10$ wave vectors in each direction are required. Its calculation is however the most consuming part in the Ewald scheme.

The **Particle Mesh Ewald** (PME) algorithm [76, 78] allows a fast calculation of the electrostatic energy. Again, the interaction is split into a short-range part and a long-range part. The short-range part converges rapidly in real space, and it is evaluated simply as a sum of atom-atom interactions, like in the Ewald scheme. Instead, the long-range part is evaluated in a more efficient way as a fast Fourier transform, by representing the atom charges as a charge density field on a grid. The grid spacing is chosen compatibly with the periodic boundary conditions. Both the short-range and long-range parts can be safely approximated by truncating the summations to few terms. The PME algorithm is particularly suited to simulate periodic systems with a large number N of atoms, as the computational cost scales like $N \log(N)$.

3.6 Temperature and pressure control

In MD simulations is possible to adopt different types of thermodynamics ensembles, which are characterized by the control of certain thermodynamic quantities. The simpler is the microcanonical

ensemble (NVE), in which the number of particles, volume and total energy are constant. However the microcanonical ensemble does not correspond to the conditions under which most experiments are carried out. In experiments typically the pressure and/or the temperature are kept constant, therefore the canonical (NVT) or the isothermal-isobaric (NPT) ensembles are the most useful. To simulate the system in such ensembles, thermostat and barostat algorithms are required to control the temperature and pressure during the MD run. In this work the Berendsen thermostat and barostat were adopted [79], along with the Nose-Hoover thermostat.

3.6.1 Berendsen thermostat

The instantaneous value of the temperature $T(t)$ of a system with N_{df} degrees of freedom is related to the kinetic energy $Ekin$ via the particles velocities as follows [79]:

$$E_{kin}(t) = \sum_{i=1}^{N} \tfrac{1}{2} m_i v_i^2(t) = \tfrac{1}{2} N_{df} k_B T(t) \tag{3.18}$$

$$T(t) = \sum_{i=1}^{N} \frac{m_i v_i^2(t)}{N_{df} k_B} \tag{3.19}$$

A simple way to alter the temperature is to scale the velocities by a factor λ:

$$\Delta T = \sum_{i=1}^{N} \frac{m_i (\lambda v_i)^2}{N_{df} k_B} - \frac{m_i v_i^2}{N_{df} k_B} \tag{3.20}$$

$$\Delta T = (\lambda^2 - 1) T(t) \tag{3.21}$$

If T_0 is the desired temperature and $T(t)$ is the current temperature, the time variation of temperature is defined as:

$$\frac{dT(t)}{dt} = \frac{T_0 - T(t)}{\tau_T} \quad , \quad \Delta T = \Delta t \frac{T_0 - T(t)}{\tau_T} \tag{3.22}$$

It follows that:

$$\lambda(t) = \sqrt{1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T(t)} - 1 \right)} \tag{3.23}$$

The factor λ is used to scale the velocities at each time-step, in order to relax the temperature toward the desired temperature value T_0. The relaxation rate is controlled by the time coupling constant τ_t, which should be small enough to achieve the required temperature, but large enough to avoid disturbance of the physical properties of the system by coupling to the bath.

3.6.2 Berendsen barostat

The Berendsen algorithm to control the pressure of the system [79] is similar to the Berendsen thermostat. In this case, the algorithm scales the coordinates $\vec{r}_i$and the box vectors $h = (\vec{a}, \vec{b}, \vec{c})$ with a matrix $\mu_{\alpha\beta}$:

$$r_{i,\alpha} \rightarrow \sum_{\beta} \mu_{\alpha\beta} r_{i,\beta} \quad , \quad h_{\alpha\beta} \rightarrow \sum_{\gamma} \mu_{\alpha\gamma} \, h_{\gamma\beta} \; . \tag{3.24}$$

The effect is a first-order kinetic relaxation of the instantaneous pressure $P(t)$ towards a reference pressure P_0, with a time constant τ_P that can be specified as an input parameter. The basic equation

for the Berendsen algorithm barostat is:

$$\frac{dP(t)}{dt} = \frac{P_0 - P(t)}{\tau_P} \tag{3.25}$$

and the scaling factor $\mu(t)$ for the atoms positions and box vectors is

$$\mu_{\alpha\beta} = \delta_{\alpha\beta} - \beta_{\alpha\beta}\frac{\Delta t}{3\tau_P}\left[P_{0,\alpha\beta} - P_{\alpha\beta}(t)\right] \tag{3.26}$$

Where β is the isothermal compressibility, $P(t)$ is the instantaneous pressure, P_0 is the target pressure and τ_P is the pressure coupling constant.

3.6.3 Nose-Hoover thermostat

In the Nose-Hoover scheme [80, 81] the equation of motion of each atom is modified by a time-dependent frictional term $\xi(t)\vec{v}_i(t)$ as:

$$\vec{a_i}(t) = \frac{\partial \vec{v}_i(t)}{\partial t} = \frac{\vec{f_i}}{m_i} - \xi(t)\vec{v}_i(t) \quad , \tag{3.27}$$

$$\frac{\partial \xi(t)}{\partial t} = \frac{(T(t) - T_0)}{Q} \quad ,$$

where the strength of the coupling Q (often called the "mass" of the oscillator $\xi(t)$) is specified before the simulation, and it is linked to the time constant through the relation:

$$Q = \frac{\tau^2 T_0}{4\pi^2} \quad , \tag{3.28}$$

but the actual implementation is often written as:

$$\vec{v_i}(t + \Delta t/2) = \left(1 + \frac{1}{2}\xi(t)\Delta t\right)^{-1}\left(\left(1 - \frac{1}{2}\xi(t)\Delta t\right)\vec{v}_i(t - \Delta t/2) + \frac{\vec{f_i}(t)}{m_i}\Delta t\right) \quad , \tag{3.29}$$

where $(1 - (1/2)\xi(t)\Delta t)/(1 + (1/2)\xi(t)\Delta t)$ is used as an approximant of $(1 - \xi\Delta t)$ around $\xi\Delta t \simeq 0$, which prevents unphysical, divergent trajectories for $\xi(t)$ in situations when $|\xi\Delta t| > 1$.

3.7 Analysis of molecular dynamics trajectories

Several structural and fluctuation properties were calculated in this thesis starting from molecular dynamics trajectories.

3.7.1 Root Mean Square Displacement (RMSD)

The displacement of atoms or groups of atoms along the trajectory can be estimated by the Root Mean Square Displacement (RMSD). The RMSD of a set of N atoms at time t, with respect to a reference conformation (e.g. the initial conformation), reads:

$$RMSD(t) = \sqrt{\frac{\sum_{i=0}^{N} |\vec{r}_i(t) - \vec{r}_i^0|^2}{N}} \tag{3.30}$$

Where $\left|\vec{r}_i(t) - \vec{r}_i^0\right|$ is the displacement of the i^{th} atom at time t from the reference position r_i^0. An increase of the RMSD indicates that the protein moves to a conformation different from the initial structure and thus may suggest that a conformational change is occurring.

3.7.2 Root Mean Square Fluctuation (RMSF)

Similar to the previous analysis is the calculation of the fluctuations of an atom or a group of atoms from the average positions, according to the formula:

$$RMSF_i = \sqrt{\left\langle \left(\vec{r}_i - \langle \vec{r}_i \rangle\right)^2 \right\rangle} \tag{3.31}$$

where $\vec{r}_i$ is the position vector of the i^{th} atom and the brackets indicate average over the trajectory. This analysis gives clues about the most mobile regions of the protein, which will display a large value of $RMSF$.

Comment on free-energy calculations

In the projects presented in Chapters 5 and 7 part of the calculation was carried out by other members of the biomolecular simulation group of SISSA (see also Introduction and Outline of the thesis). These are the metadynamics free-energy calculations which have been performed by F. Pietrucci for bestrophin and X. Biarnes for the prion protein. For this reason the theory of free-energy calculations is not presented here but in the computational section of the corresponding chapters.

Part II

Calcium-gated ionic channels

Chapter 4

Molecular modeling of hBK Ca^{2+} binding domains

Abstract

Human big conductance Ca^{2+}- and voltage-gated K^+ channels (hBK_{Ca}) are putative drug targets for cardiovascular, respiratory and urological diseases. Here we have used molecular simulation and bioinformatics approaches to construct models of two domains important for Ca^{2+} binding and channel gating, namely the regulator of conductance for K^+ (RCK1) domain and the so-called calcium bowl (Ca^{2+} bowl). As templates for RCK1 were used the corresponding domains from a K^+ channel from *E. coli* and the K^+ channel from *Methanobacterium thermoautothropicum* (MthK). Ca^{2+} bowl was modeled upon the structure of the human thrombospondin-1 C-terminal fragment. The domain underwent 70-ns MD simulations in aqueous environment. The relevance of these models for interpreting the available molecular biology data is then discussed.

4.1 Introduction

Ca^{2+}- and voltage-gated big conductance K^+ channels (BK_{Ca} channels) are expressed by a variety of eukaryotic organisms. In mammals, they play an important role in muscle contractions, neuronal excitability, hormones and neurotransmitters release [7]. They are putative targets for pharmaceutical intervention for a variety of diseases, including cardiovascular, respiratory, urological diseases along with idiopathic epilepsy and ischemic reperfusion injury [8, 82].

BK_{Ca} channels open upon an increase of intracellular Ca^{2+} ions (up to a concentration spanning 0.5 μM - 50 mM [83]) along with an increase in the membrane electric potential [84]. Molecular biological experiments have established that these channels are homotetramers [85], composed of either α or α and β subunits along with four cytoplasmic domains: RCK1 and RCK2 (Regulate the Conductance of K^+), the Ca^{2+}-binding domain, the so-called "calcium-bowl" (Ca^{2+} bowl) and the serine proteinase-like domain (SPLD) (Figure 4.1). The transmembrane region of the α subunit includes the S4 helix as the voltage sensor responsible for voltage sensitivity [86], the pore region through which K^+ ions flow out of the cell [87] and the S0 helix (present only in this channel) that places the N-terminus outside the cell [88]. Ca^{2+} ions are bound to the cytoplasmic domains: Ca^{2+} binding to two putative Ca^{2+}-binding sites in the RCK1 domain causes the gating of the channel [89, 83, 90].

A linker then connects RCK1 to RCK2, which is also important for the gating [91, 92]. Finally, a Ca^{2+} bowl domain is present. Mutations in this domain reduce the ability of low concentration of Ca^{2+} to activate the channel [93].

Besides the α subunit, there is also a β subunit (Figure 4.1) [85], which exists in four different subtypes, $\beta 1$ - $\beta 4$ [94]. The most studied one is the $\beta 1$ subunit in smooth muscle BK_{Ca} channels [95, 96]. The $\beta 1$ subunit increases Ca^{2+} sensitivity and decreases voltage sensitivity of BK_{Ca} channels [95].

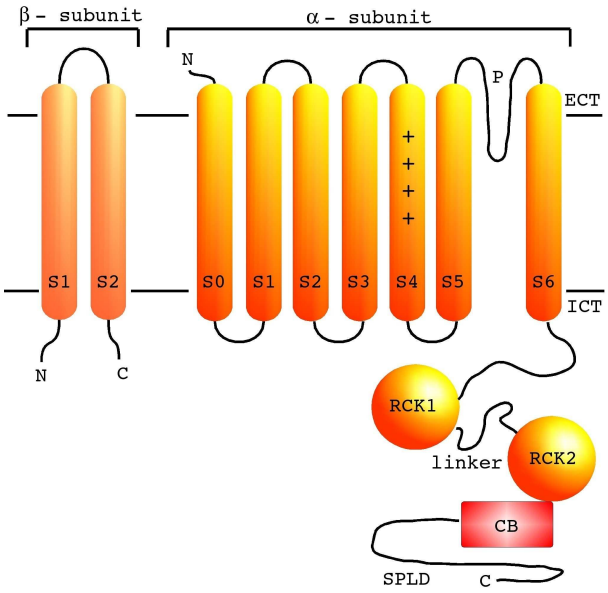

Figure 4.1: All BK_{Ca} channel monomers feature one subunit, which consists of a transmembrane domain (seven transmembrane-helix bundle and a pore helix), the RCK1, the linker and the RCK2 domains, the Ca^{2+} bowl (CB) and the SPLD domain [95]. Beside α subunit also β transmembrane subunit can be present, for which four different subtypes are known [94]. The most studied one ($\beta 1$), shown here, is present in BK_{Ca} channels expressed in smooth muscles [95, 96]

Most of the drugs targeting the protein are believed to bind to the α subunit cytoplasmic domain, affecting Ca^{2+} affinity and modulating the interaction with the β subunits [82]. Few ligands interact with the transmembrane domain [82]. Therefore, structural information of both the transmembrane and cytoplasmic domains of the α subunit would be of great help for the rational design of ligands with large affinity for the channel. Unfortunately, such information is still lacking.

Here we have used bioinformatics approaches, along with molecular dynamics (MD) simulations, to construct structural models for some of these domains. The available molecular biological data for the channel (see Table 4.1) have been used to validate models of two of the channel domains. Firstly, the RCK1 domain, employing as templates the correspondent domains of two structurally similar channels: a K^+channel from *E. coli* (PDB entry: 1ID1) [97] and the K^+channel from *Methanobacterium thermoautothropicum* (MthK) (PDB entry: 1LNQ) [91] (it is worth noting that the latter is Ca^{2+}-(and not voltage-) gated). Secondly, the Ca^{2+} bowl domain was modeled using a two stage strategy as the sequence analysis *per se* does not allow identifying the metal binding site(s). For this domain, we first modeled the Ca^{2+} bowl, using as the template one of the Ca^{2+}-binding domains from human thrombospondin-1 C-terminal fragment (PDB entry: 1UX6, from A865 to N897) [98], and we then attempt to identify the metal binding sites by performing MD simulations with Ca^{2+} ions placed in

five different locations, these are referred to for clarity as Ca1-Ca5. No attempts are made to construct tetrameric structures and only isolated RCK1 and Ca^{2+} bowl domains were considered. The way the subunits assemble is indeed not known and there are no experiments involving the subunit-subunit interface, which could help assisting the modeling of the quaternary structure of the protein.

In addition, for the transmembrane region and RCK2, although there are reasonably good templates (*e.g.* voltage-gated K^+ channel $K_v1.2$ [99], *E. coli* K^+ channel [97], MthK [91]), molecular biology data are not available. Therefore we limit the analysis for these domains to the sequence alignment. For the linker and SPLD, there are not reliable templates presently and they are not investigated here.

4.2 Computational details

4.2.1 Transmembrane region

hBK_{Ca} transmembrane region sequence was aligned to that of $K_v1.2$ (PDB entry: 2A79) [99] using the ClustalW program [47].

4.2.2 RCK domain

Two templates were found: the RCK domains in an *E. coli* K^+ channel (PDB entry: 1ID1) [97], which form a dimer, and those of the MthK channel (PDB entry: 1LNQ) [91], which construct a dimer of dimers. Both share with RCK1 18% of sequence identity (SI) and 40% of sequence similarity (SS).

We first performed a structural alignment between the RCK domains from MthK and *E. coli* by using Swiss-Pdb Viewer 3.7 [100]. Then we performed the sequence alignment of RCK1 from hBK_{Ca} against both templates, using the same program. Gaps were manually removed from the secondary structure element sequence tracts. 3D models were eventually built using MODELLER 6v2 [48]. A distance restraint between NZ@K513 and CG@D546 atoms was included in the model in order to form a salt bridge, whose presence has been suggested by the experiments [97].

4.2.3 Calcium bowl

The best template for Ca^{2+} bowl is the 35 residues long segment of the human thrombospondin-1 C-terminal fragment (PDB entry: 1UX6, from A865 to N897, SI=17%, SS=43%, Figure 4.5) [98], as shown by a Blast [40] search. It binds five Ca^{2+} ions; therefore five Ca^{2+} ions were added to the 3D model of Ca^{2+} bowl, built as above with MODELLER 6v2 [48], with similar coordination as in the template. The resulting adduct was inserted in a box of dimensions 46 x 46 x 46 $Å^3$ packed with 3955 water molecules. The system underwent MD simulations using the Gromacs-3.2.1 program [101]. The Amber parm99 [73] and TIP3P [66] force fields were used for the protein frame with the metal ions and the solvent, respectively. The model was initially minimized with positional restraints on the solute with a stepwise (each 100 ps) reduced force constant from 1000 kJ mol^{-1} nm^{-2} till 0 kJ mol^{-1} nm^{-2} (in steps of 200 kJ mol^{-1} nm^{-2}). After, an unconstrained MD was carried out for an overall length of 70 ns. A time step of 2 fs was used. All bond lengths were kept fixed applying the LINCS algorithm [70]. Temperature (300 K) and pressure (1 bar) were kept constant by coupling the system to a few Berendsen thermostats [102]. Periodic boundary conditions were applied treating long-range electrostatic interactions with the particle-mesh Ewald technique [76, 78]using a cut-off of 10 Å for

the real part of the electrostatic. The same cut-off was used for the van der Waals interactions. Pair lists were updated every 20 steps.

4.3 Results

4.3.1 Transmembrane region

The alignment for the transmembrane region S0-S3 with $K_v1.2$ turned out to exhibit large gaps, together with the two missing parts (36 and 12 amino acids) in the structure of $K_v1.2$; as a result, we restricted our analysis to the S4-S6 tract (Figure 4.2). The sequence identity for the region from S4-S6 between hBK_{Ca} and $K_v1.2$ is 21%, while the similarity for the same part is 50% (Figure 4.2).

The following features can be noticed: (i) as expected the GYG fingerprint, highly conserved signature among K^+ channels [103] is located in the selectivity filter. It is responsible for the high selectivity of K^+ over Na^+. (ii) In general, K_v channels have conserved PXP sequence as the gating hinge in the last part of S6 helix [104, 105, 106, 107, 108]. In hBK_{Ca} is conserved just one proline (P385), which could cause breaks or irregularities in the helix structure and can be therefore considered as putative gating hinge. However, also G376 in S6 helix, which is conserved among K^+channels and corresponds to G83 in MthK [91] and to G247 in $K_v1.2$ [99] could be a good candidate for the hinge. (iii) The S4 helix with its repeats of positively charged residues (Arg) followed by two hydrophobic residues represents the voltage sensor [86]. In comparison with $K_v1.2$, hBK_{Ca} lacks two positively charged residues. (iv) Two adjacent rings of four Glu residues are positioned at the entrance of the intracellular vestibule [87]. These eight Glu's play a key role for the high conductance of the channel by attracting intracellular K^+ ions, concentrated at the entrance of the intracellular vestibule. In fact, Glu to Gln mutation in either ring decreases single channel current by about 12 pA, whilst the mutation of all the eight Glu's in the ring cause the conversion to an inwardly rectifying channel (meaning that the channel current decreases with increasing the voltage) [87].

4.3.2 RCK1 domain

The RCK1 domain starts at the highly conserved ^{409}HIVVC [97] and finishes at FSMRS^{580} (Figure 4.3A). The fold of RCK1 model is the Rossmann fold with two additional helices (Figure 4.3B) [97]. The model includes the salt bridge formed by K513 and D546 (Figure 4.3C) as established by molecular biological experiments [97], which show that mutations K513D or D546K reduce open probability of the channel. This salt bridge has to be considered as an input in the model, as it does not emerge from the alignment procedure only.

Two putative Ca^{2+}-binding sites have been identified [89, 83, 90]. One is formed by D427 and D432 which binds Ca^{2+} in concentrations at around 10 μM. In fact, D432A produces a marked reduction in the ability of Ca^{2+} to shift gating, while D427A mutation exhibits the same behavior as the wild type channel [83]. In order to identify such binding sites, we scanned a set of Ca^{2+}-binding proteins from the PDB database. In general Ca^{2+}-binding sites involve three to four negatively charged residues (see Figure 4.4A), each contributing with either one or both oxygen atoms. In addition, as Ca^{2+} is usually coordinated with 6-8 oxygen atoms [109] (see Figure 4.4B), water and/or oxygen atoms from the amino acids backbone or the side chains of Gln and Asn could contribute to the binding. Therefore, in the high affinity binding site, beside the suggested D427 and D432, D434 and/or D435 may also bind Ca^{2+}. The model supports that D427 and D432 do not bind to the metal ion, although this has not been firmly established yet.

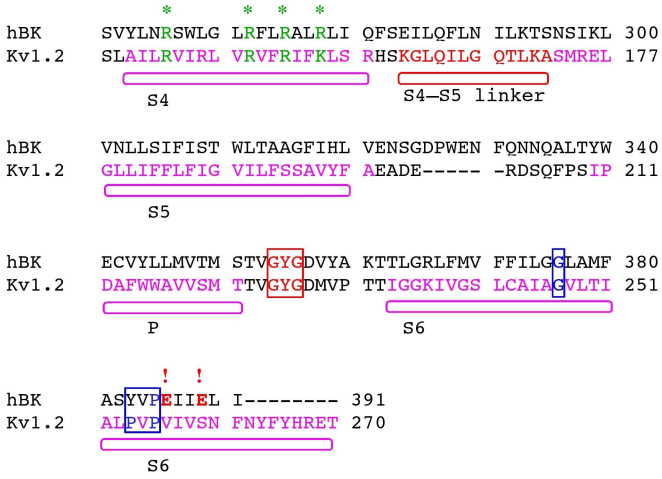

Figure 4.2: Alignment between S4-S6 region of hBK$_{Ca}$ and K$_v$1.2 channel. Structurally and functionally important features are shown: (i) positively charged residues in S4, which constitute the voltage sensor are labeled by stars (*); (ii) GYG as the fingerprint of selectivity filter is signed by a box [103]; (iii) G376 and P385 as putative gating hinges in hBK$_{Ca}$ are depicted by boxes (iv) E386 & E389 that form the ring of negatively charged residues [87]are denoted with the exclamation mark (!).

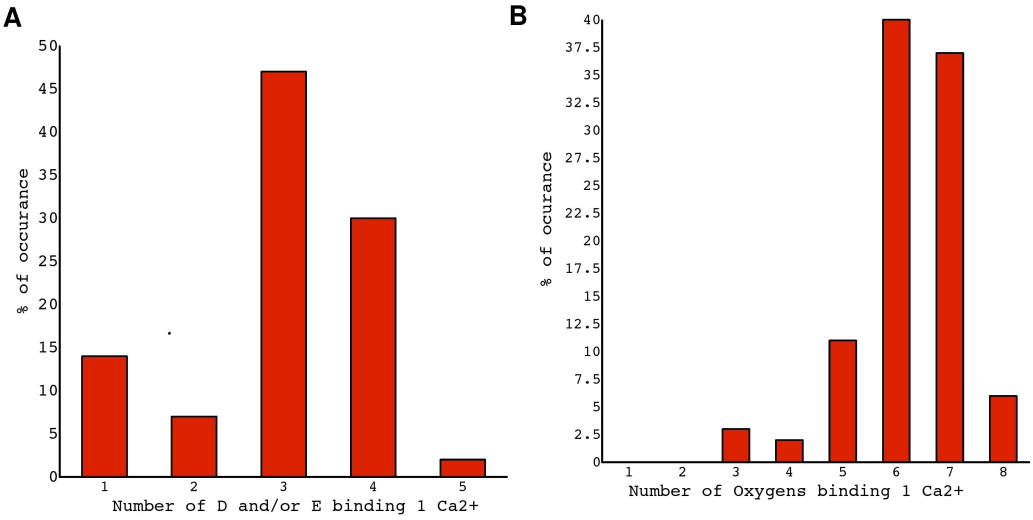

Figure 4.4: Number of negatively charged residues (Asp, Glu) **(A)** and oxygens **(B)** needed to bind one Ca^{2+} ion as determined by the analysis of 87 bonded Ca^{2+} ions in 20 Ca^{2+}-binding proteins from PDB database [1].

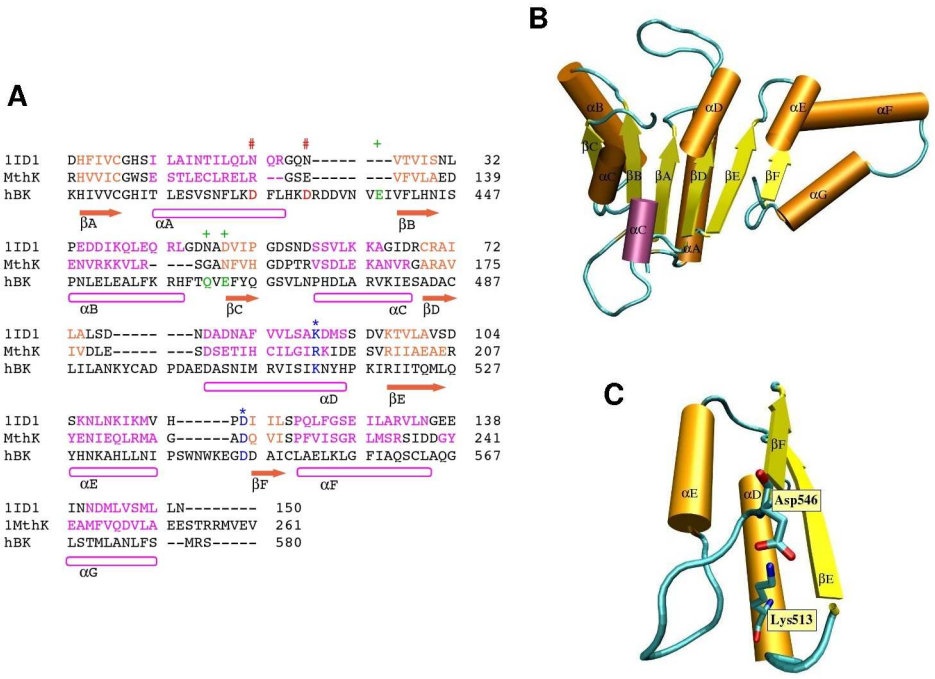

Figure 4.3: **A:** Alignment of the two RCK domains from *E. coli* and MthK with the corresponding sequence in hBK_{Ca}. The following features are shown using hBK_{Ca} notation: (i) salt bridge formed by K513 and D546 is labeled by stars (*); (ii) putative high affinity Ca^{2+}-binding site [83, 90] composed by D427 and D432 is depicted by hashes (#); (iii) putative low affinity Ca^{2+}-binding site [89, 83, 90], constructed by E439, Q462 and E464, is denoted with plus signs (+). **B:** The model of the RCK1 domain. **C:** Salt bridge formed by K513 and D546 was included in the model and is consistent with the molecular biology data [97].

4.3.3 Calcium bowl

The **Ca^{2+} bowl** is a crucial domain for Ca^{2+} binding [110, 111, 90]. Molecular biological experiments show that there are five adjacent Asp's residues (D959-D963 in the sequence: ^{945}TE...DQ DDDDDPD...TA^{979}), which most probably bind the metal ions [110, 111]. Beside them, there are other oxygen-containing residues, which can coordinate Ca^{2+} as well (Figure 4.5A) [110, 111]. The model can be formally divided in two segments, one formed by 17 (T945-D961) and the other by 18 (D962-A979) residues. An α-helix is present in the first segment, the rest forms a coil. Each segment bears two Ca^{2+}-binding sites (Ca1, Ca2 and Ca3, Ca4), whereas the fifth one (Ca5) is in between. Unfortunately, there are no templates sharing high sequence identity with Ca^{2+} bowl, furthermore the number of Ca^{2+} ions binding to it is unknown. The best template for this domain, the human thrombospondin-1 C-terminal fragment [98] exhibits SI=17% and SS=43% (Figure 4.5A). Therefore we decided to construct first a model based on such template (namely from T945 to A979 in hBK_{Ca}) and then to relax it by NPT MD simulations in aqueous solution. The five Ca^{2+} in the template were added in the corresponding binding sites of the model (Ca1-Ca5) in order to verify if they remained in these positions during the MD simulation. Obviously, the exact number and the coordination of Ca^{2+} bound to the Ca^{2+} bowl have to be validated against further biochemical experiments.

At the end of the MD run (70 ns), the model structure from T945 to E967 keeps the fold around

Ca^{2+} ions, while the part from L968 to A979 is unfolded. Ca^{2+}-binding sites are present in the structured part of the Ca^{2+} bowl model. Two Ca^{2+} ions, Ca1 and Ca3, remain well coordinated throughout the MD simulations forming well defined binding sites (Figure 4.5B and C), while the other three (Ca2, Ca4 and Ca5) get partially or completely solvated. Ca1 is bound by D959, D961, D963 and E967 carboxyl oxygen atoms along with D963, D965 backbone oxygen and one water molecule. Ca3 is coordinated by E946, D950, D957 and D962 carboxyl oxygen atoms beside the D959 backbone oxygen and a water molecule.

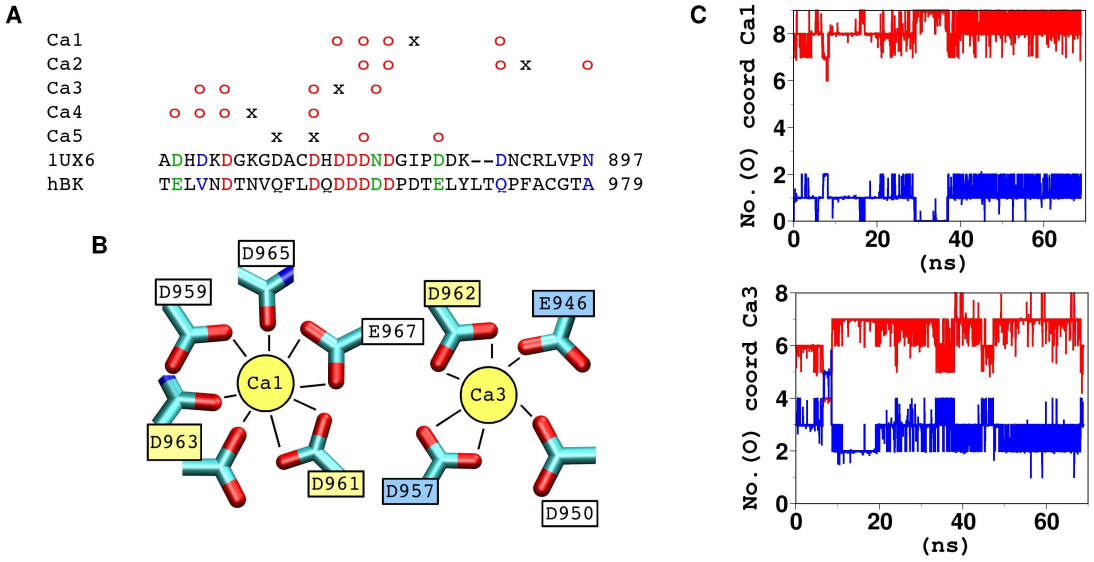

Figure 4.5: **A:** Alignment between human Thrombospondin-1 and hBK_{Ca} Ca^{2+}-binding domain (Ca^{2+} bowl). The template residues binding Ca^{2+} ions with the side chain are depicted with the circle 'o'; those binding Ca^{2+} with the backbone oxygen with 'x'. **B:** Ca^{2+}-binding sites (Ca1 and Ca3) in the Ca^{2+} bowl model after 70-ns MD simulations. Ala mutations of the residues depicted in blue boxes were experimentally shown to reduce Ca^{2+} binding for 30% while those in yellow boxes reduced Ca^{2+} binding for 50% (see also Table 4.1). Water molecules are not shown due to the clarity's sake. **C:** Ca^{2+} coordination number contributed by protein O-donors (red) and water molecules (blue) plotted as a function of MD simulation time. Ca^{2+} were defined as bound if their coordination number was at least 6 and not more than 2 water molecules contributed to their binding.

Ca4 remains coordinated by 5 peptide O-donors and three waters, what we consider as partially solvated. Ca2 is as well partially dissolved being coordinated by 4 peptide O-donors and four waters, while Ca5 is completely dissolved after ~ 9.6 ns. Ca5 solvation may be caused by the fact that in the initial model Ca5 is bonded to a smaller number of protein ligands than Ca1-Ca4.

Our MD model of Ca^{2+} bowl appears to be consistent with most experimental data (Table 4.1): the five adjacent Asp's (D959-D963) in the core of Ca^{2+} bowl are reported as important for Ca^{2+} sensitivity and/or binding [110] and in our model four of them (D959, D961, D962 and D963) coordinate Ca^{2+}. (ii) N949, N952 and Q954 are reported not to coordinate Ca^{2+} with their side chain oxygen atoms [110, 111], this is fully consistent with our model. (iii) Q958 and Q972 do not bind Ca^{2+}, which is consistent with the results obtained in the study by Bao *et al.* [110]. (iv) In addition, the finding that E946 and D957 bind Ca^{2+} is fully consistent with the data obtained by Sheng *et al.* [111], where they have shown that single or combined mutations of E946A and D957A reduces Ca^{2+} sensitivity and binding. However, it should be mentioned that it has been reported elsewhere [110] that both Glu's present in the Ca^{2+} bowl are not important for Ca^{2+} binding, in contrast to the previous study.

This issue therefore requires further studies. (v) The fact that D960 plays a role for Ca^{2+} sensitivity and binding [110] may be also consistent with our model. Although in our structure, Ca5 is not directly bound to D960, the negative electrostatic potential of the latter may play an important role for stabilizing the ion in the close proximity of the protein.

The model is not consistent with some experimental data. Firstly, mutation of D950 to Ala does not influence Ca^{2+} sensitivity or binding [111], yet in our model D950 binds Ca3. In addition, E967 should not bind Ca^{2+} [110], yet in our model it does bind Ca1. More refined models could take these pieces of information into account by performing constraint molecular dynamics simulations.

mutations	*experiment*	*model*	*model − 70ns*
E946A	no changes [110] reduces Ca^{2+} sensitivity and binding (30%) [111]	does not bind Ca^{2+}	Ca3 [1ox]
N949A	no change [110]	does not bind Ca^{2+}	does not bind Ca^{2+}
D950A	no change [111]	Ca3 [1ox]; Ca4 [1ox]	Ca3 [1ox]
N952A	no change [110, 111]	Ca4 [bb ox]	does not bind Ca^{2+}
Q954A	no change [110, 111]	Ca5 [bb ox]	does not bind Ca^{2+}
D957A	reduces Ca^{2+} sensitivity (50%) and Ca^{2+} binding (30%) [110, 111]	Ca3 [1ox]; Ca4 [1ox]; Ca5 [bb ox]	Ca3 [2ox]
Q958A	reduces Ca^{2+} sensitivity (50%) and binding (20%) [110]	does not bind Ca^{2+}	does not bind Ca^{2+}
D959A	reduces Ca^{2+} sensitivity (60%); no data for Ca^{2+} binding [110]	Ca1 [1ox]; Ca3 [bb ox]	Ca1 [1ox]
D960A	reduces Ca^{2+} sensitivity (90%) and binding (50%) [110]	Ca5 [2ox]	does not bind Ca^{2+}
D961A	no change in Ca^{2+} sensitivity; reduces Ca^{2+} binding (50%) [110]	Ca1 [1ox]; Ca2 [1ox]	Ca1 [2ox]
D962A	reduces Ca^{2+} sensitivity (90%) and binding (50%) [110]	Ca3 [1ox]	Ca3 [2ox]
D963A	reduces Ca^{2+} sensitivity (40%); no data about binding [110]	Ca1 [1ox]; Ca2 [1ox]	Ca1 [1ox,bb ox]
D965A	reduces Ca^{2+} binding (30%) [110]	Ca1 [bb ox]	Ca1 [bb ox]
E967A	no change [110]	does not bind Ca^{2+}	Ca1 [2ox]
Q972A	reduces Ca^{2+} sensitivity (40%); no data about binding [110]	does not bind Ca^{2+}	does not bind Ca^{2+}
E946A/D957A	reduces Ca^{2+} sensitivity and binding (60%) [111]		
D960A/D962A	reduces Ca^{2+} binding (80%) [110]		
D961A/D963A	reduces Ca^{2+} binding (75%) [110]		
K513D	cause upward shift in the Ca^{2+} activation profile [97]	forms salt bridge	
D546K	cause upward shift in the Ca^{2+} activation profile [97]	forms salt bridge	
K513D/D546K	shift in the Ca^{2+} activation profile similar to that of WT [97]	forms salt bridge	

Table 4.1: Available experimental data from which structural information can be extracted. In the columns MODEL are presented results from homology models followed by the results of MD simulations of Ca^{2+} binding to the Ca^{2+} bowl after 70 ns. Ca1-Ca5 identify position of the Ca^{2+} included in the model and represent which Ca^{2+} binds to the determined residue; "bb ox" stands for backbone oxygen.

4.4 Concluding remarks

We have used biocomputing techniques to provide structural insights into the human big conductance Ca^{2+}- and voltage-gated K^+ channel. We constructed a model of the RCK1 domain, which exhibits the template Rossmann fold with two additional helices.

A model of the Ca^{2+} bowl was constructed based on the thrombospondin-1 $T3_6$ Ca^{2+}-binding domain and studied by MD simulations. Two well formed Ca^{2+}-binding sites were identified that explain most of the available experimental data data and can be helpful for planning future molecular biology experiments in this pharmacologically relevant channels.

Chapter 5

Regulation of Bestrophins by Ca^{2+}

Abstract

Bestrophins are a recently discovered family of Cl^- channels, for which no structural information is available. Some family members are activated by increased intracellular Ca^{2+} concentration. Bestrophins feature a well conserved Asp-rich tract in their COOH terminus (Asp-rich domain), which is homologous to Ca^{2+}-binding motifs in human thrombospondins and in human big-conductance Ca^{2+}- and voltage-gated K^+ channels (BK_{Ca}). Consequently, the Asp-rich domain is also a candidate for Ca^{2+} binding in bestrophins.

Based on these considerations, we constructed homology models of human bestrophin-1 (Best1) Asp-rich domain using human thrombospondin-1 X-ray structure as a template. Molecular dynamics simulations were used to identify Asp and Glu residues binding Ca^{2+} and to predict the effects of their mutations to alanine. We then proceeded to test selected mutations in the Asp-rich domain of the highly homologous mouse bestrophin-2. The mutants expressed in HEK-293 cells were investigated by electrophysiological experiments using the whole-cell voltage-clamp technique.

Based on our molecular modeling results, we predicted that Asp-rich domain has two defined binding sites and that D301A and D304A mutations may impact the binding of the metal ions. The experiments confirmed that these mutations do actually affect the function of the protein causing a large decrease in the Ca^{2+}-activated Cl^- current, fully consistent with our predictions. In addition, other studied mutations (E306A, D312A) did not decrease Ca^{2+}-activated Cl^- current in agreement with modeling results.

The experimental work was done in collaboration with the group of Prof. A. Menini at Neurobiology sector of SISSA/ISAS: J. Rievaj, F. W. Grillo and A. Boccaccio. The computational work was done in collaboration with C. Anselmi and F. Pietrucci (SISSA/ISAS).

5.1 Introduction

In 1998, the gene VMD2, which encodes for the human bestrophin-1 (hBest1) protein, was found to be responsible for the inherited Best vitelliform macular dystrophy. This is an early-onset autosomal dominant maculopathy typically characterized by yellowish lesions in the central area of the retina [14, 15]. Since then, four human (hBest1-4) and three mouse (mBest1-3) genes were identified, together with other genes from different species (for review, see [16]). All human and mouse members of the

bestrophin family have been expressed in heterologous systems [6, 112, 113, 114, 115]. The proteins form Cl^- channels, likely to be composed by several subunits [6, 116]. Three of them (hBest1, hBest4 and mBest2) are regulated by Ca^{2+} [6, 112, 23, 117]. They have been proposed as the putative molecular counterpart of the Ca^{2+}-activated Cl^- current in some epithelial cells [118, 119, 120] and in olfactory sensory neurons (mBest2), where they were suggested to play a pivotal role for the olfactory transduction [117, 121, 122]. However, recent studies indicated that the TMEM16 family represents a new candidate molecular counterpart of Ca^{2+}-activated Cl^- channels [123, 124, 125].

Although the link between the protein and the Best disease has not been clarified yet [16, 126], it has been established that more than one hundred mutations, mostly located in hydrophobic domains, are associated with the disease [127, 128]. Some functionally highly relevant mutations (Q293K/H, L294V, ΔI295, G299E/R, E300D, D301E, T307I and D312N) [16, 6, 17, 18, 19] are instead located at the C-terminal domain inside the Asp-rich sequence from W287 to D323 (Asp-rich domain, hereafter). These mutations reduce whole-cell currents compared to wild-type (WT) hBest1 after their expression in heterologous systems [16, 6, 17, 18, 19]. Furthermore, inserting a stop codon in position L294 (but not in position F353) causes a strong decrease of the current [129].

Clearly, knowledge of the structural determinants of the protein is crucial to understand its function both in health and in disease conditions. However, no 3D structural information is available so far. Nevertheless, two topology models have been proposed [114, 130]. According to them, the N- and C-terminal domains of bestrophins would be located at the intracellular side of the membrane and would be connected to four [130] or five [114] hydrophobic domains forming the channel.

In an effort to characterize prominent structural and functional features of bestrophins, we noticed that the Asp-rich domain, which has been indicated as a possible Ca^{2+} sensor for bestrophins [23] is the most amenable region for molecular modeling studies. In fact, it has significant sequence similarity with other domains for which structural information is available. These include Ca^{2+}-binding domains as the cytoplasmic Ca^{2+} bowl motif [20, 131] of the big conductance Ca^{2+}- and voltage-gated K^+ channels (BK_{Ca}) [20, 131, 132, 133, 21] and the type 3 (T3) motifs in the C-terminal region of thrombospondins [134]. The first corresponds to a highly conserved 28 amino acid long peptide [20, 131] characterized by a net negative charge and five adjacent Asp's residues (D959-D963 in human BK_{Ca}). Site-directed mutagenesis experiments have confirmed the role of Ca^{2+} bowl in the binding of calcium ions [21, 22]. Although the experimental 3D structure is not available, a theoretical model of this domain in BK_{Ca} channels has been recently provided by us [135]. The model turned out to be consistent with most of the available experimental data [21, 22].

The second type of Ca^{2+}-binding domains, T3 repeats of thrombospondins, consists in tandems of Asp-rich motifs [136], which resembles EF hands in the spacing of acidic side chains [137, 138]. The X-ray structures of the C-terminal fragments of two human thrombospondins (TSP-1 and TSP-2) show that they comprise several T3 motifs bound to as much as twenty six Ca^{2+} ions [136, 139].

Because of the significant sequence similarity between the Asp-rich domain and the aforementioned Ca^{2+}-binding domains, it is plausible to hypothesize that Ca^{2+} activation of bestrophins could involve, at least in part, Ca^{2+} binding to the Asp-rich domain. This hypothesis is fully consistent with the experimental evidence from mutations of amino acids constituting putative Ca^{2+} ligands (E300D, D301E and D312N) in the domain, which affect dramatically the measured current in the HEK-293 cells [6, 18]. In addition, inserting a stop codon at position F353, located after the Asp-rich domain, does not cause a current reduction, while cutting the channel at the first amino acid in the Asp-rich domain (L294) significantly reduces the current [129]. Therefore we focused our attention on the Asp-rich domain, whereas other possible Ca^{2+}-binding domains were not explored here. On the other

hand, computer simulations of the entire channel, which would help to clarify how Ca^{2+} binding could affect channel gating, are not feasible due to the lack of structural information of the transmembrane region of the channel.

A straightforward approach to test if Asp-rich domain is involved in Ca^{2+} activation of bestrophins+ would be to mutate randomly its Asp/Glu residues, that have the highest affinity for Ca^{2+}, and then measure the functional properties of the channel. Here instead we followed a more rational approach, which uses computation to guide the experiments. We first constructed a structural model of hBest1 Asp-rich domain based on the TSP-1 $T3_6$ X-ray structure [136], which has the largest number of Ca^{2+}-binding residues conserved across bestrophins. The models of WT and selected alanine mutants were then studied by molecular simulations. Mutations with the largest predicted effect on Ca^{2+} binding turned out to cause a dramatic decrease in the experimentally measured mBest2 Ca^{2+}-activated current (more than 96% reduction with respect to the WT mBest2). The consistency between theory and experiments permits to have the first structural insights on the hBest1 protein.

5.2 Computational details

5.2.1 Bioinformatics

The sequences of the human TSP-1 $T3_{5-7}$ repeats (PDB code: 1UX6) [136] were aligned with the Ca^{2+} bowl from human and mouse BK_{Ca} channels [20, 131, 132, 133, 21, 140, 141] and the Asp-rich domains of hBest1 [14, 15], mBest1 [15], hBest2 [142], mBest2 [143], hBest3 [142], mBest3 [144], hBest4 [142], ceBest1 [6], dvBest1 [145], dmBest2 and dmBest4 [146] (Swiss-Prot entries and notations are reported in the legend of Figure 5.2). The alignment was initially performed with ClustalW [47] and then slightly refined to avoid apolar residues in Ca^{2+} binding positions. Sequence similarity was determined considering amino acid pairs having positive values in BLOSUM62 [147] substitution matrices. The homology model of hBest1 Asp-rich domain from W287 to D323 was then built with the MODELLER 6v2 program [4, 148], using the crystal structure of the TSP-1 T36 motif as a template (PDB code: 1UX6) [136]. In the best resulting model, five to three Ca^{2+} ions present in the template were kept. These are: (i) M5, with five Ca^{2+} ions coordinated. (ii) M4 is composed by four Ca^{2+} ions which are Ca_1, Ca_2, $Ca_4$and Ca_5. (iii) M3' and M3" are including three Ca^{2+} ions, Ca_1, Ca_2, Ca_4 or Ca_1, Ca_2, Ca_5, respectively.

5.2.2 MD simulations

Models of the Asp-rich domain were inserted in a box of edges 49, 50 and 53 Å filled with ~4,100 water molecules. Cl^- (three in M5, one in M4) or Na^+ (one in M3' and M3") counterions were added to ensure systems charge neutrality. The force fields for the protein, ions and water molecules were the Amber parm99 [149, 150] and TIP3P force fields, respectively [66].

For M5, energy minimization of water and counterions was performed with position restraints on the Asp-rich domain and Ca^{2+} ions, with a force constant of value 25.0 kcal mol^{-1} $Å^{-2}$. After the first 300 cycles of steepest descent minimization, conjugate gradient method was used until 50,000 cycles. Then steepest descent minimization of the system without restraints was performed for another 150 cycles, following by conjugate gradient minimization until 50,000 cycles. After that, constrained solute and Ca^{2+} ions underwent 60 ps of linear heating MD from 0 K till 300 K. Restraints on Asp-rich domain and Ca^{2+} ions corresponded to a force constant of 5.0 kcal mol^{-1} $Å^{-2}$. A time step of 2 fs was used.

Finally, the system underwent NPT MD simulations for 75 ps at 1 fs time step. The system was maintained at the reference pressure and temperature of 1 bar and 300 K, respectively, by coupling the system to a Berendsen thermostat and barostat [102]. A cut-off of 9 Å was used for calculating non-bonded interactions. Finally, unrestrained MD simulations were carried out for an overall length of 70 ns. A time step of 2 fs was used. All bond lengths were kept fixed applying the LINCS algorithm [151]. Temperature (300 K) and pressure (1 bar) were kept constant as aforementioned. Periodic boundary conditions were applied treating long-range electrostatic interactions with the particle-mesh Ewald technique using a short-range cutoff of 10 Å [152]. The same cutoff was used for the van der Waals interactions. Pair lists were updated every 20 steps.

Models with four or three Ca^{2+} ions were built starting from the MD snapshot of M5 at 9 ns (after the initial relaxation) by manually removing one or two ions (see above). In M4, M3' and M3" models Ca_3 was always removed since it is immediately solvated during the simulation of M5 (Figure 5.4). M4, M3' and M3" underwent then MD simulations for 20 ns (M4) and 35 ns (M3', M3"), respectively, using exactly the same protocol as that used for the M5 model.

The models of the Asp-rich domain mutants (D301A, D302A, D303A, D304A, E306A, E306A+D323A, D312A) were built, based on the M5 model, using the program Molden [153]. They underwent the same protocol of minimization and relaxation of the WT followed by 20-ns MD runs.

The systems underwent minimization using Amber 8.0 program [154] and then MD simulations using the Gromacs 3.2.1 program [155, 156]. Trajectory analysis was performed by using the VMD 1.8.6 program [157].

5.2.3 Test of the accuracy of the force field used in the MD simulations

The structures and energetics of Ca^{2+}-binding proteins are difficult to reproduce with current force fields [158, 159]. Factors which limit the accuracy include the lack of polarizability and charge transfer effects [160]. Here we tested the accuracy of three force fields, the Aqvist [161], the Bradbrook [162] and the AMBER one [163, 164]. We have performed 20-ns MD calculations of TSP-1 $T3_6$, for which the X-ray structure is available [165], and compared crystal and MD structures. TSP-1 $T3_6$ is a 39 amino acids long loop resembling EF hands motif in the spacing of acidic side chains [166, 167]. It binds five Ca^{2+} ions by eleven Asp and two Asn residues (Figure 5.1).

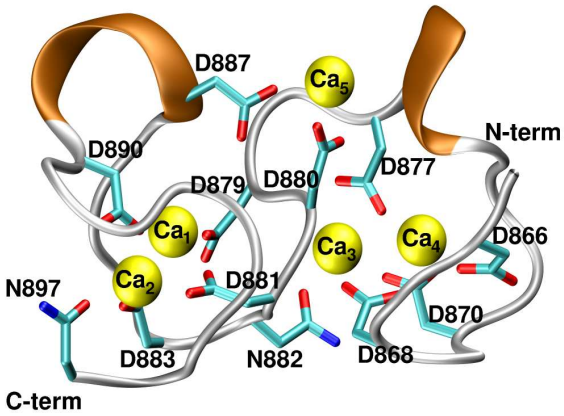

Figure 5.1: Crystal structure of the TSP-1 $T3_6$ domain. Five Ca^{2+} ions (indicated as Ca_{1-5} according to their positions) are bonded to the Asp/Asn residues of the domain.

Here we considered a structure in which all five (Ca_{1-5}) calcium ions were present. The TSP-1 $T3_6$ domain was inserted in a box of edges 46, 51 and 49 Å filled with $\sim$ 3700 water molecules and 2 Cl^- counterions. Computational details for the relaxation and unrestrained MD simulation are the same as reported in the section 5.2.2. In the three simulations, the final structure turned out to be rather similar to the X-ray structure: indeed the RMSD values calculated onto the N-terminal and C-terminal halves were as small as 3.2 Å and 2.4 Å, respectively (Table 5.1) and Ca_{1-4}(Figure 5.1) remained bound. However, the total backbone RMSD from the X-ray structure was relatively large (Table 5.1) because of conformational changes of Ca_5-binding groups (D880 and D887); in fact, Ca_5 was solvated after few ns ($\sim$ 2 ns for Aqvist- and Bradbrook-based MD; $\sim$ 1.3 ns for the AMBER-based MD).

	Aqvist [161]	Bradbrook [162]	AMBER [163, 164]
Total RMSD	3.5	5.2	5.5
RMSD of N-term	3.5	3.8	3.2
RMSD of C-term	2.6	4.3	2.4

Table 5.1: TSP-1 T36 backbone RMSD values (Å) relative to the X-ray structure after 20-ns MD simulations based on three different force fields. The N-term corresponds to the residues from A865 to H878, whereas the C-term corresponds to the residues from D879 to N897 (see also 5.2).

The loss of the Ca^{2+} ion might be caused by several factors, including the limitations of the accuracy of the force field describing the Ca^{2+} ions [159] and/or the absence of electrostatic packing forces in the simulation with respect to the crystal, along with a difference in temperature (the simulation was run at 300 K, the crystal structure was solved at 100 K). Force-matching methods, which were applied by us already for other metal-containing systems [168], might be used to further address this issue, which is however beyond the scope of this study. Because of the limitations of the approach used in our structural predictions, we used our molecular modeling only to identify alanine mutations in Asp-rich domain which may play a role for the Ca^{2+} binding. No attempts were made to use our computations to predict the structure of disease-linked mutants [169] and/or mutants for which electrophysiological experiments have been previously carried out [170]. In all of the other calculations presented here we have used the AMBER force field.

5.2.4 Metadynamics

A 60-ns long metadynamics simulation was performed by biasing two collective variables (CVs) with a history-dependent potential. These are the coordination numbers of Ca_2 and Ca_4 to Asp and Glu residues of the Asp-rich domain. They were defined as the number of contacts shorter than 4.3 Å among one Ca^{2+} ion and all carboxylate carbons of the studied domain:

$$S_i = \sum_j n_{ij} \quad , \quad n_{ij} = \frac{1 - \left(\frac{R_{ij}}{R_0}\right)^8}{1 - \left(\frac{R_{ij}}{R_0}\right)^{16}}$$

where i runs over Ca_2 and Ca_4, j runs over all carboxylate carbons, R_{ij} is the distance between the two atoms, R_0 = 4.3 Å is a cutoff distance chosen in order to discriminate between bonded and non-bonded Ca^{2+} ions (this cutoff distance ensures that Ca^{2+} ion is considered bonded when it is at most 3 Å away from carboxylate oxygen atoms (see e.g. [171]), and n_{ij} is a function switching smoothly

between 1 (bonded Ca^{2+} ion) and 0 (non-bonded Ca^{2+} ion). Similar CVs have been employed in a previous study of Ca^{2+}-binding polymers [172]. The 2-D Gaussians forming the metadynamics bias potential have width $\sigma_1 = \sigma_2 = 0.2$ (along S_1 and S_2 respectively) and a height of 0.2 kcal/mol. A new Gaussian is added along the trajectory every 1 ps. Metadynamics [173] is implemented in a modified version of the Gromacs 3.3.1 package [156].

5.2.5 Experimental studies

5.2.5.1 Site-specific mutations of mBest2 and heterologous expression

The clone of mBest2 in pCMV-Sport6 mammalian expression plasmid was obtained from the RZPD (Berlin, Germany) collection (Deutsches Ressourcenzentrum für Genomforschung). Site-specific mutations of mBest2 were made using a PCR-based site-directed mutagenesis kit (Quiagen). Mutations were confirmed by DNA sequencing.

mBest2 and its mutants were transfected into HEK-293 cells along with a vector that expressed EGFP (pEGFP; Invitrogen) at a 1:8 ratio using Fugene-6 transfection reagent (Roche). 2 days after transfection, cells were dissociated and replated. Transfected cells were identified by EGFP fluorescence and used for patch clamp experiments within 3 days after transfection.

5.2.5.2 Immunofluorescence and confocal microscopy

To test successful expression of proteins, as well as their spatial distribution in HEK-293 cells, we used co-staining of antibodies against mBest2 and labeled wheat germ agglutinin (WGA) as a marker of glycosylated surface-expressed proteins - a method previously used to show localization of proteins in cell membrane and/or close submembrane space [174, 175].

Transfected cells (without the EGFP expressing vector) were fixed with 4% paraphormaldehyde for 20 min, washed, incubated with 5 µg/ml WGA conjugated to Alexa Fluor 488 (Invitrogen) for 20-30 min, incubated for 30 min in blocking solution containing 1% BSA (Sigma), 0.1% Triton X-100 (Sigma) and 1% FCS (Sigma), then for 2 hours with an antibody against mBest2 [117] (diluted in 1:500 ratio with blocking solution), washed extensively and incubated with a A594-conjugated anti-rabbit secondary antibody (Invitrogen, diluted 1:500 in blocking solution) for 1 hour, washed and mounted with Vectashield (Vector Laboratories, Burlingame, CA). Immunoreactivity was analyzed by confocal microscopy with a Leica TCS SP2 confocal microscope. All incubations were made at room temperature; all substances were diluted in PBS, unless otherwise indicated.

5.2.5.3 Electrophysiological recordings

Currents were measured with an Axopatch 200B patch-clamp amplifier (Molecular Devices, CA, USA) in the whole-cell voltage-clamp mode. The amplifier was controlled via a Digidata 1440A (Molecular Devices, CA, USA).

Patch pipettes were made using borosilicate capillaries (WPI, Sarasota, Florida, USA) and pulled with a Narishige PP83 puller (Narishige, Tokyo, Japan), using a double stage pull. The pipette resistance was 2-4 MΩ. Data were sampled at 20 kHz and low-pass filtered at 10 kHz. Acquisition and storage of data were performed with the PClamp 10 software (Axon Instruments, CA, USA). Data analysis and figures were made with the Clampfit 10 software (Axon Instruments, CA, USA) and Igor 6.0 software (Wavemetrics, Lake Oswego, OR, USA).

The holding potential was 0 mV and a voltage protocol from -100 mV to +100 mV, with 20 mV steps of 200 ms duration was applied (starting 2 minutes after reaching the whole-cell configuration). Cell capacitance was estimated from a 10 mV step test protocol in the presence of the extracellular Cl^- current blocker SITS. Current densities were calculated normalizing the current recorded at +60 mV to the capacitance of each cell. Data are shown as mean ± standard error of the mean. N is the number of cells. Statistical significance was determined using un-paired t-tests. P values <0.05 were considered significant.

The standard extracellular solution contained (in mM): 140 NaCl, 5 KCl, 2 CaCl2, 1 MgCl2, and 10 HEPES, 10 Glucose, pH=7.4. The pipette solution contained (in mM): 130 CsCl, 2 MgCl2, 10 HEPES, 5 EGTA without added Ca^{2+} for the nominally 0 Ca^{2+} solution, or with 5 mM CaCl2 for the 22 µM Ca^{2+} free pipette solutions (free Ca^{2+} was calculated with the program WinMAXC, C. Patton, Stanford University, Palo Alto, CA, USA). pH was adjusted to 7.3 by adding HCl or CsOH. The osmolarity was adjusted with sucrose to 310 or 300 mOsm for the extracellular and intracellular solution, respectively.

To block Cl^- currents, the extracellular blocker 4-acetamido-4'-isothiocyanato-stilben-2,2'-disulfonate (SITS) was directly dissolved in the extracellular solution at 2 mM. All chemicals were purchased from Sigma. All experiments were carried out at room temperature (20-22°C).

5.3 Results

Our strategy to identify Asp and Glu amino acids important for the function of hBest1 was twofold. First, we attempted to identify residues that may bind the Ca^{2+} ions in the WT. Next, we predicted the effects of mutations to alanine of these residues on the structure and function of the Asp-rich domain. The effect is non-trivial, because residues, which bind Ca^{2+} ions in the WT, can be replaced by others when mutated to alanine.

5.3.1 Identification of acidic residues binding Ca^{2+} in hBest1 Asp-rich domain

The structural determinants of the Asp-rich domain were modeled by using the X-ray structure of TSP-1 $T3_6$ repeat [136] as a template, because it exhibits the largest number of Ca^{2+}-binding residues conserved across bestrophins family (Figure 5.2).

The number of Ca^{2+} ions bound to the Asp-rich domain is unknown. Typically, as observed from other Ca^{2+}-binding domains in the PDB database, each Ca^{2+} ion is usually coordinated by three Asp/Glu residues [171]. Since the template has five Ca^{2+} ions bound and 33% of negatively charged residues more than the Asp-rich domain, it is plausible to expect that the last would bind five or less Ca^{2+} ions. Therefore we considered models including five (M5), four (M4), and three (M3', M3") Ca^{2+} ions (Figure 5.3).

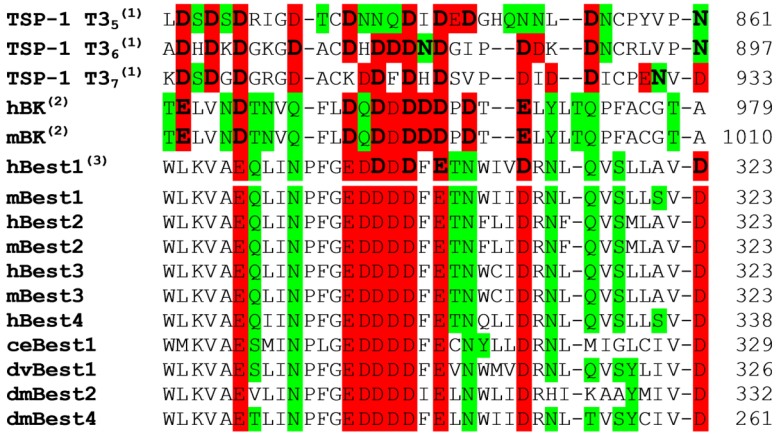

Figure 5.2: Alignment of the Asp-rich sequences from TSP-1, BK_{Ca} and bestrophins. Alignment between TSP-1 $T3_{5-7}$ repeats [136], human BK_{Ca} Ca^{2+} bowl (SwissProt entry: Q12791) [133], mouse BK_{Ca} Ca^{2+} bowl (SwissProt entry: Q08460) [132] and the Asp-rich domain at the C-terminal part of hBest1 (SwissProt entry: O76090) [14, 15]. hBest1 shares in this tract 33% to 36% of sequence similarity with TSP-1 $T3_{5-7}$ motifs and 42% with BK_{Ca} Ca^{2+} bowl [21]. Bold letters represent residues, which are involved in Ca^{2+} binding in TSP-1 (as found in the crystal structure (1) [136]), hBK_{Ca} and mBK_{Ca} (from the model (2) described in section 4.3.3) and hBest1 (this work (3)). O-donor residues are colored in red (negatively charged residues) or green (neutral residues). The alignments of selected tracts from the C-terminal domains of other bestrophins are also shown: mBest1 (mouse bestrophin-1, SwissProt entry: O88870), hBest2 (human bestrophin-2, SwissProt entry: Q8NFU1), mBest2 (mouse bestrophin-2, SwissProt entry: Q8BGM5), hBest3 (human bestrophin-3, SwissProt entry: Q8N1M1), mBest3 (mouse bestrophin-3, SwissProt entry: Q6H1V1), hBest4 (human bestrophin-4, SwissProt entry: Q8NFU0), ceBest1 (Caenorhabditis elegans bestrophin-1, SwissProt entry: Q21973), dvBest1 (Drosophila virilis bestrophin-1, SwissProt entry: Q20C62), dmBest2 (Drosophila melanogaster bestrophin-2, SwissProt entry: Q9VRW4), dmBest4 (Drosophila melanogaster bestrophin-4, SwissProt entry: Q9VUM6).

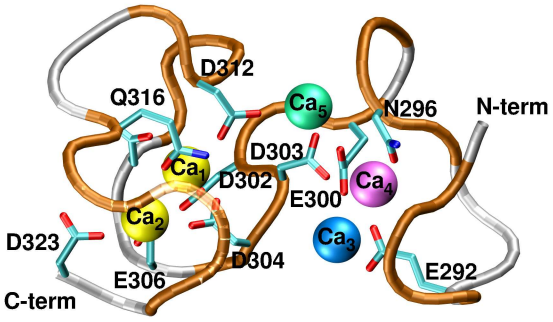

Figure 5.3: Initial model of the Asp-rich domain studied by the MD simulations. The Ca^{2+} ions and the residues coordinating them are shown as sphere and licorice, respectively. M5 comprises five Ca^{2+} ions, labeled for clarity as Ca_{1-5} according to their binding positions (The same labels are used throughout the text). M4 is the same as M5 except that it lacks Ca_3 (colored in blue). M3' and M3" are the same as M4 except that they lack Ca_5 (green), and Ca_4 (magenta), respectively.

In all molecular dynamics (MD) simulations, two ions (Ca_1 and Ca_2 in Figure 5.4) remained bound. Ca_1 was bound to D302, D304, E306 and D312, Ca_2 to D303, D304, E306, D312, Q316, S318 and D323 (Table 5.2).

Instead, the other ions were partially or completely solvated. The decrease in the number of bound Ca^{2+} ions with respect to TSP-1 $T3_6$ can be caused by a combination of factors: first of all, the lower number of negatively charged residues in Asp-rich domain with respect to $T3_6$ (C-type repeat is well conserved in hBest1, while N-type is not; Figure 5.2), but also the limitations associated with the force field (see section 5.2.3). MD simulations showed that residues D302, D304, E306, D312 and D323 (Table 5.2) play a major role for Ca^{2+} binding.

In order to identify other Ca^{2+}-binding residues, we used an enhanced sampling method, i.e. metadynamics [33]. In this approach, the MD simulation is altered by biasing the selected collective variables (CVs) with a history-dependent potential. The bias, constructed as a sum of Gaussians centered along the trajectory of the system, reduces the free-energy barriers and enhances the rate of transitions, in the present case the binding and unbinding of Ca^{2+} ions. The starting structure for the metadynamics simulation was selected among our energy-minimized models. We chose M3' model, because it kept the largest number of bound Ca^{2+} ions (three) during the MD simulations for as long as 25 ns (Figure 5.4).

	Ca1	**Ca2**	**Ca4**	**Ca5**
Glu292			56%	
Asn296			41%	
Glu300			100%	
Asp302	100%			
Asp303		31%	63%	88%
Asp304	100%	100%		
Asp304 (bb)	71%	—		
Glu306	100%	93%		87%
Asp312	100%	36%		71%
Asp312 (bb)	—	—		50%
Gln316		25%		
Ser318		44%		
Ser318 (bb)		73%		
Asp323		100%		

Table 5.2: Percentage of time during which specific residue binds Ca^{2+} ions. Percentages were averaged over all MD simulations of all the studied models. Ca_3 is not reported as it is never bound to the protein. "bb" indicates that the backbone carbonyl oxygen is involved in the binding.

The selected CVs measured the extent of binding of Ca_2 and Ca_4, respectively, since Ca_1 was always bound to a well-defined binding pocket in all previous simulations. The limitations of the force field along with the large number of conformational degrees of freedom of the system prevented us from calculating the free-energy profile associated to Ca^{2+} binding. Therefore, the simulation was only used to sample stable configurations, different from those studied in the MD.

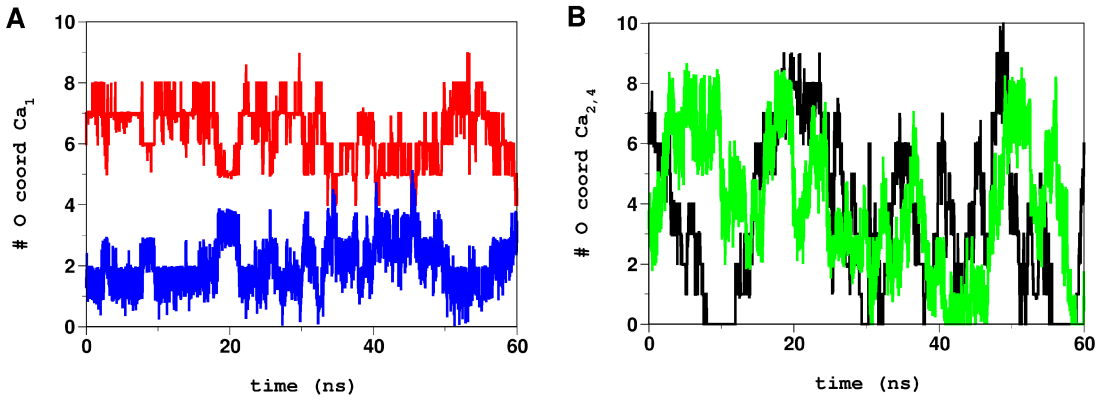

Figure 5.5: Metadynamics simulation of M3' model. - **A:** Number of peptide O-donors (red) and water molecules (blue) coordinating Ca_1 through the metadynamics simulations of the Asp-rich domain of hBest1. As in standard MD, Ca_1 is stably bound to the protein for most of the time. **B:** Number of peptide O-donors coordinating Ca_2 (black) and Ca_4 (green) plotted as a function of the metadynamics simulation time.

Throughout the metadynamics run, Ca_1 kept its coordination shell as in the previous MD simulations (Table 5.2 and Figure 5.5A). Due to the bias potential, Ca_2 and Ca_4 instead turned out to bind and unbind several times (Figure 5.5B) exploring different configurations.

Ca_2 was either bound to D301, D302, D303, D312 and D323 (Figure 5.6) or it was solvated. Analogously, the third Ca^{2+} ion (Ca_4 in Figure 5.6) was bound to D301, D302, D303, D304, E306, D312 and D323 in some of the stable conformations. Therefore, our metadynamics simulations confirmed that the residues identified by MD can play a role for Ca^{2+} binding. In addition, they further suggested a possible role of D301 and D303 for such binding.

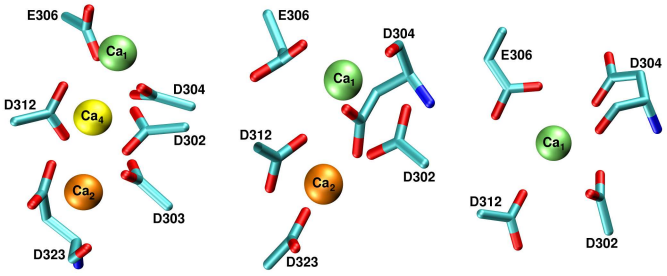

Figure 5.6: Metadynamics simulations of M3' model. Pictorial representation of the most populated structures with one, two or three coordinated Ca^{2+} ions emerging from the simulations. Notice that Ca_1 binding pocket formed by D302, D304, E306 and D312 is always occupied. In addition, the second Ca^{2+} ion always occupies Ca_2 binding pocket.

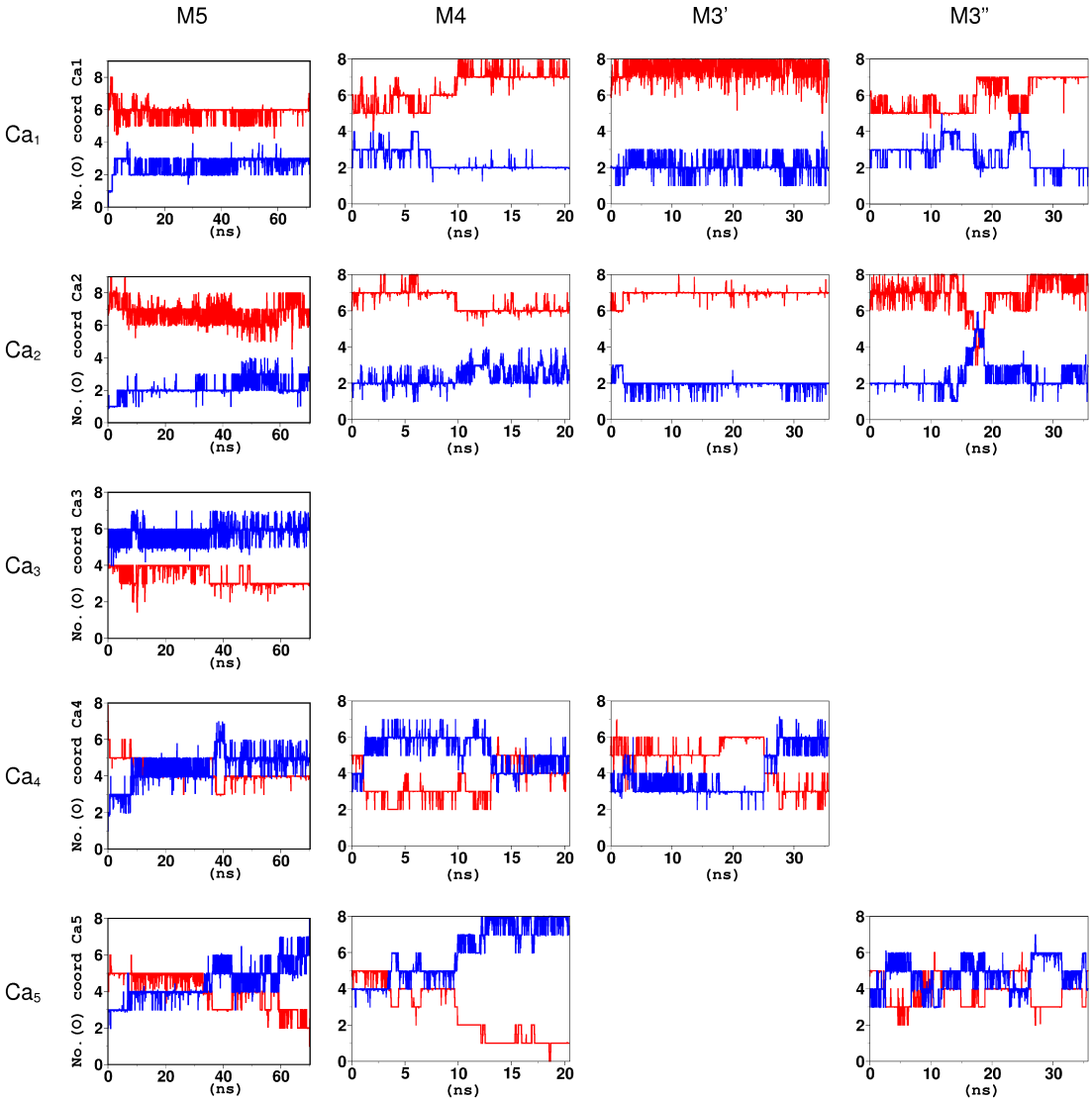

Figure 5.4: MD simulations of M3'-M5 models. Ca^{2+} coordination numbers contributed by protein O-donors (red) and water molecules (blue) plotted as a function of simulation time. Ca^{2+} ions were defined as stably bound if their coordination number was at least 6 and not more than 2 water molecules contributed to their binding.

5.3.2 Computer-aided design of mutants with low-affinity for Ca^{2+}

We attempted to identify mutations that could have an effect on Ca^{2+} binding. We constructed structural models in which putative key Asp and Glu residues among those identified in the above section were mutated to alanine. These are D301A, D302A, D303A, D304A, E306A, D312A and the double mutant E306A+D323A (see Table 5.3). Adducts with five Ca^{2+} ions underwent 20 ns of MD.

At the end of MD simulations, D301A bound only one Ca^{2+} ion (Table 5.3), while its 3-D structure remained similar to that of WT.

The Asp-rich domain bearing D304A mutation unfolded, possibly because D304 is the central residue of the domain and therefore responsible for its structural stability; all Ca^{2+} ions were partially or fully solvated (Table 5.3). Since the structural information of the channel is lacking, it is difficult to establish if unfolding of D304A would occur also in the real molecular surrounding, that is, when the Asp-rich domain is attached to the remaining bestrophin part.

Mutants D302A, D303A, E306A, and D312A resembled the WT as they bound at least two Ca^{2+} ions (Table 5.3). In fact, other residues contributed to Ca^{2+} coordination replacing the mutated ones.

Based on these results we concluded that D301A and D304A mutations have the largest effect on Ca^{2+} binding, while D302A, D303A, E306A, and D312A affect Ca^{2+} binding to a lesser extent. To verify this, the experiments of the WT and mutant mBest2 were performed. The putative Ca^{2+}-binding domain of mBest2 has almost the same sequence as that of hBest1 (Figure 5.2). To investigate whether the WT structures are affected by the differences, MD simulations were performed also on mBest2. As a result of MD simulations both models showed very similar behavior, with Ca_1 and Ca_2 that remained stably bound, while Ca_3 and Ca_5 were solvated. Ca_4 bound more peptide O-donors in mBest2 than in hBest1. However, this difference should not be overestimated, since even starting from the same initial system (e.g. hBest1), running two independent MD simulations of the same length could result in slightly different events.

	Ca1	Ca2	Ca3	Ca4	Ca5	EXP
WT (0 ns)	**6+0**	**6+1**	4+4	**6+1**	5+2	—-
WT (20 ns)	**6+1**	**7+1**	5+3	4+3	5+3	—-
D301A	5+3	**5+2**	5+3	5+3	3+4	no current
D302A	**6+2**	**6+2**	2+4	5+3	2+6	—-
D303A	**7+1**	**7+1**	3+5	5+3	2+6	—-
D304A	4+4	5+3	0+8	**3+5**	2+6	no current
E306A	**6+1**	**8+0**	4+4	**6+1**	4+4	current
D312A	**6+2**	**7+1**	2+4	4+4	0+7	current

Table 5.3: MD simulations of the M3' model and its alanine mutants. Ca^{2+} protein coordination numbers in WT and the investigated mutants of the Asp-rich domain, as observed after 20 ns of MD simulations. First number refers to the number of peptide O-donors, whereas the second refers to the number of coordinated waters. Ca^{2+} ions were considered bound if they were coordinated with at least 5 protein O-donors and not more than 2 water molecules. Bound ions are highlighted.

5.3.3 Experimental studies

5.3.3.1 Current recordings of wild-type and designed mutants

The changes of Ca^{2+} coordination in the aforementioned mutations may affect the amplitude of bestrophin Ca^{2+}-activated Cl^- currents, which can be experimentally measured. This was done here for mouse bestrophin2 (mBest2). WT and mutants currents were compared after their heterologous expression in the HEK-293 cell line (Figure 5.7).

To determine if mutant proteins are delivered to the plasma membrane as WT, we used immunostaining with antibodies against mBest2 [117]. The WT protein showed increased localization in plasma membrane and/or in close sub-membrane space in a significant portion of transfected cells, although some fraction of protein was localized also intracellulary (Figure 5.7A). These results are similar to data previously published for hBest2 [115] and xBest2 [174]. The localization of the proteins was further investigated by co-staining with fluorescently labeled wheat germ agglutinin (WGA), a selective marker of glycosylated surface-expressed proteins. Co-localization of both signals, which indicates membrane expression of the protein, was observed. The distribution of mBest2-related signal and its co-localization with WGA in mutants were very similar to WT, indicating that the proteins were expressed and delivered to the plasma membrane (Figure 5.7). Consequently, changes of the measured current were caused by a modification of the channel functionality and not by the inability of the mutant proteins to reach the plasma membrane.

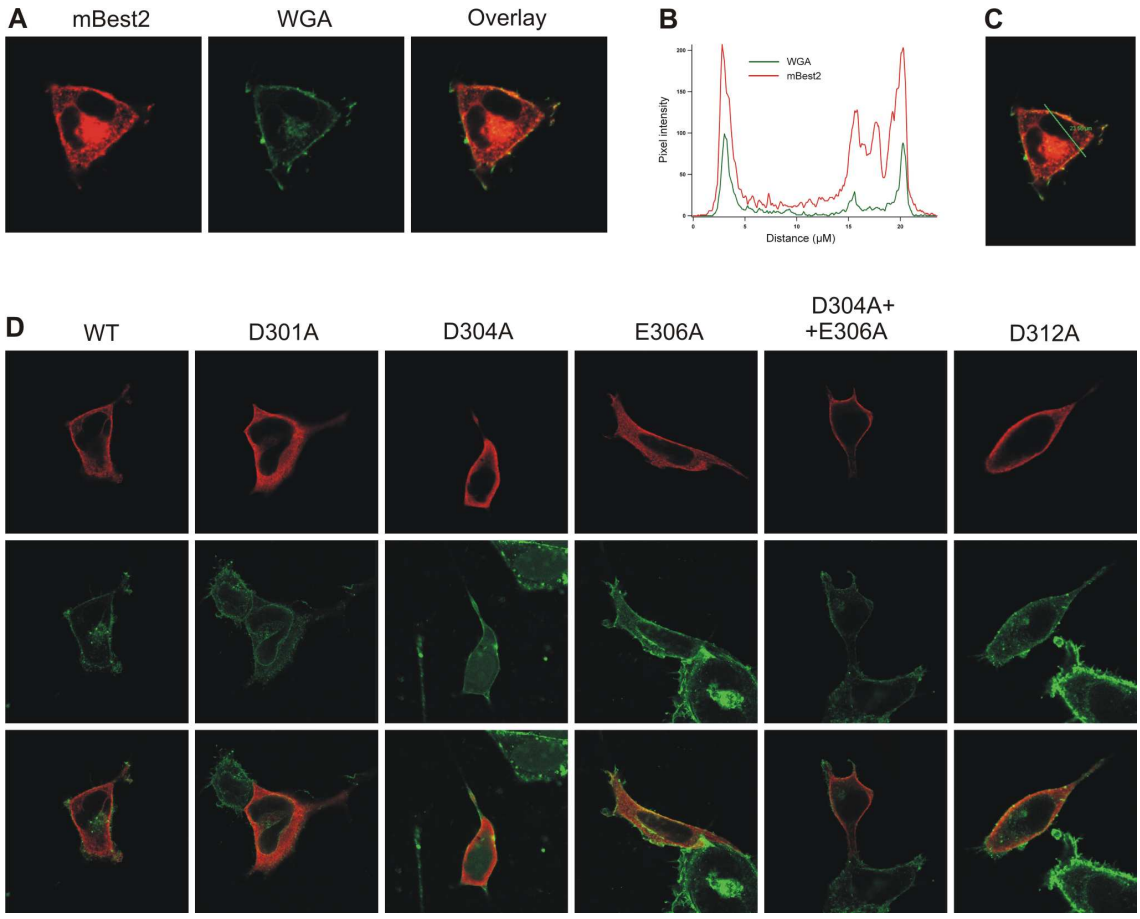

Figure 5.7: Localization of mBest2 and its mutants in HEK-293 cells. mBest2 was transfected into HEK-293 cells. 2-4 days after transfection, the cells were fixed, permeabilized and stained with antibody against mBest2, then visualized with Alexa-594-conjugated secondary antibody. Cells were also stained with Alexa-488 conjugated wheat germ agglutinin (WGA). Typical cell transfected with wild type mBest2 is shown in **(A)**. Signal from mBest2 staining (red) is shown in the left, signal from WGA (green) from the same cell is in the middle. Overlay of both signals (right) indicates the expression of protein in plasma membrane and/or in close submembrane space. The distribution of mBest2 at the membrane was determined by comparing red and green fluorescence along a line drawn across a cell. Distribution of WGA- and mBest2- associated fluorescence along the line across the cell is shown in **(B)**, line is shown in **(C)**. Similar procedure was repeated for all five investigated mutations. Results for wild type (WT) mBest2, D301A, D304A, E306A, D304A+E306A and D312A mutations are compared in **(D)**. Signal from staining against mBest2 (red) is shown in first row, signal from WGA (green) in the middle, overlay of both signals is in third row. Cells transfected with all investigated mutations showed similar distribution of signals, confirming increased localization of mutated proteins in plasma membrane and/or in close submembrane space. Non-transfected cells had negligible mBest2-associated fluorescence (data not shown).

We then measured currents in HEK-293 cells transfected with mBest2 WT or mutants both in nominally 0 free Ca^{2+} level and in 22 μM free Ca^{2+} intracellular concentration. Currents were measured in the whole-cell voltage-clamp configuration applying voltage-steps of 20 mV from -100 mV to +100 mV from a holding potential of 0 mV. The current values at +60 mV were normalized to the capacitance of each cell and averaged.

The WT current in nominally 0 Ca^{2+} was significantly smaller than that in 22 μM Ca^{2+} concentration (Figures 5.8-5.9).

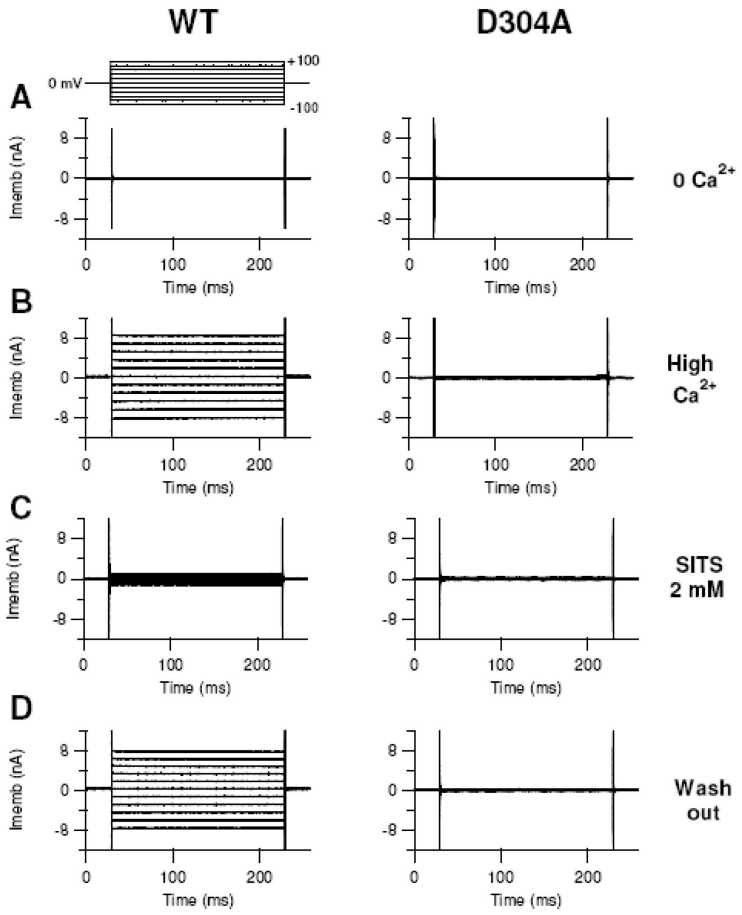

Figure 5.8: Currents in HEK-293 transfected cells. Currents in HEK-293 cells transfected with EGFP and WT mBest2 (left column) or D304A mutant (right column). Transfected cells were identified by the EGFP-expressed green fluorescence. Whole-cell voltage clamp recordings were obtained with pipette solutions containing nominally 0 **(A)** or 22 μM free Ca^{2+} **(B-D)**. For each type of channel, recordings in **B-D** were obtained from the same cell, while traces in **A** are from different cells. Voltage steps of 200 ms duration were given 2 min after the reaching of the whole-cell configuration from a holding potential of 0 mV to voltages between -100 and + 100 mV in 20 mV steps. In the left column, panels **C** and **D** show the reversible block of the WT current by 2 mM SITS, a Cl^- channel blocker. The mutant D304A did not show an appreciable current when Ca^{2+} was present in the intracellular solution (right column; see also Figure 5.9).

This shows that Ca^{2+} ions activated a current in agreement with previous results [112, 117]. Most of the Ca^{2+}-induced current was reversibly blocked by the Cl^- channel blocker SITS indicating the current is indeed carried by Cl^-. In contrast, D304A and D301A produced negligibly small currents in

both Ca^{2+} levels (Figure 5.9). Two other mutants, E306A and D312A, produced currents: whilst the amplitude of the current of E306A was similar to that of WT (Figure 5.9), D312A showed a somehow unexpected result. It produced a substantial amount of current in the nominally 0 Ca^{2+} concentration (Figure 5.9). In 22 μM Ca^{2+} concentration, the current significantly increased and was larger than that of WT. The current was also blocked by SITS in both Ca^{2+} levels, indicating that it is a Cl^- current (data not shown). Therefore, the D312A mutation not only allowed a Ca^{2+}-activated Cl^- current, but it also modified channel gating increasing its current also in the absence of Ca^{2+}. Finally, the double mutant D304A+E306A abolished the current, similarly to the single mutant D304A (Figure 5.9).

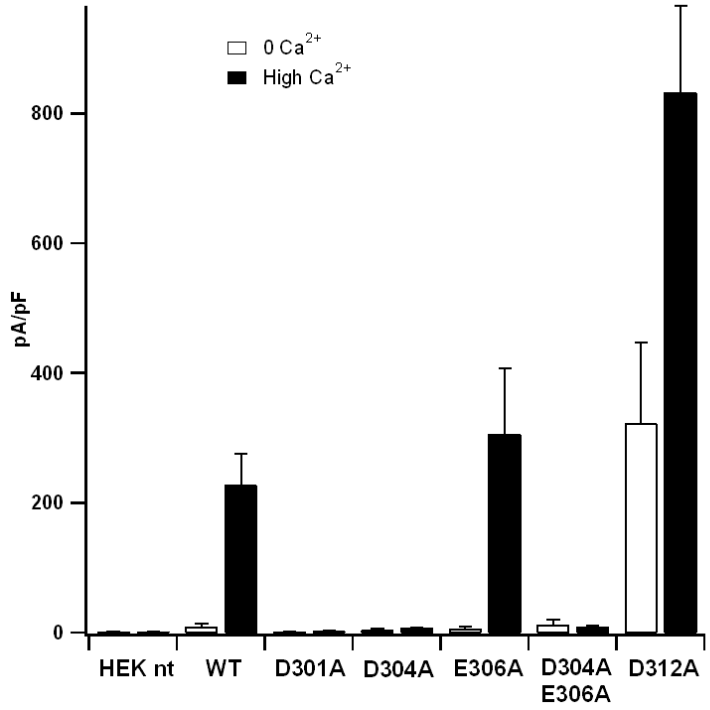

Figure 5.9: Comparison of current amplitudes from mutant and WT mBest2 Cl^- channels. Mean current densities from non-transfected cells (HEK nt) and cells transfected with EGFP and WT or mutated mBest2. The intracellular pipette solution contained nominally 0 (white bar) or 22 μM free Ca^{2+} (black bar). Data are shown as mean values ± standard error of the mean. Each data represents the mean of at least five cells. Control experiments with non-transfected HEK-293 cells did not show any appreciable current. Mean current densities in the presence of 22 μM free Ca^{2+} of WT, E306A and D312A were significantly higher than those in nominally 0 Ca^{2+} ($P<0.02$; N=5-8), whereas currents at the two Ca^{2+} levels were not statistically different for D301A, D304A and D304A+E306A ($P>0.05$, N=5-20). Currents in the presence of 22 μM free Ca^{2+} of D301A, D304A, and D304A+E306A were substantially decreased compared with WT ($P<0.002$ for each group, N=5-20). E306A had a mean value not significantly different from WT ($P=0.5$; N=7-8). D312A showed a current increase compared to WT both in the absence ($P=0.03$; N=6-8) and in the presence of Ca^{2+} ($P=0.003$; N=6-9).

Experiments showed that WT, E306A and D312A mutants allowed a substantial Ca^{2+}-activated current, whereas D301A and D304A mutants produced a negligibly small Ca^{2+}-activated current. These results were fully consistent with our theoretical predictions.

5.3.4 Concluding remarks

Understanding the interactions between the Ca^{2+} ions and the Asp-rich domain in bestrophins is an important step in the investigation of the physiological role of these proteins, because they are putatively involved in the Ca^{2+}-activated Cl^- current in some epithelial cells [118, 119, 120]. In addition, it seems to be also important for understanding the pathophysiology of the Best disease, since the Asp-rich domain is a "hot spot" for mutations causing the illness [127, 128]. At present nothing is known about the underlying structural features of the bestrophins.

We propose that the Asp-rich domain located at the C-terminal region of bestrophins can contribute to Ca^{2+} binding [23], due to its similarity with other Ca^{2+}-binding domains, as the Ca^{2+} bowl in BK_{Ca} [20, 21, 22] and T3 repeats of thrombospondins [134, 136, 139]. This statement is consistent with previous experimental data, which show that mutations of putative Ca^{2+} ligands in bestrophin Asp-rich domain affect dramatically the current [6, 18].

In order to identify negatively charged residues in the domain involved in Ca^{2+} binding, we adopted a multifaceted strategy involving homology modeling, MD simulations, immunostaining and electrophysiological experiments of WT and mutated bestrophins.

Theoretical data suggest that not all Asp and Glu residues have equal importance in Ca^{2+} binding. Simulations of D301A and D304A mutants show that they are crucial for the capacity of the domain to bind Ca^{2+}. Measures of Ca^{2+}-sensitive Cl^- current of WT and mutant bestrophins expressed in HEK-293 cells show that D301A and D304A mutations do actually affect the functionality of the channel, dramatically reducing the Ca^{2+}-activated current amplitude. These findings are fully consistent with our theoretical predictions.

Other residues, E306 and D312 bind Ca^{2+} in the MD simulations. In our experiments, the mutation of these residues to alanine does not cause reduction in the current in agreement with the simulations, since other O-donors replace the aforesaid residues in the Ca^{2+} coordination shell.

In conclusion, simulations suggest that at least two Ca^{2+}-binding sites could be present in the Asp-rich domain of bestrophins. Furthermore, electrophysiological experiments show that mutations predicted by our modeling have an impact on the function, decreasing the Ca^{2+}-activated current amplitude and confirming the importance of the bestrophin Asp-rich domain in Ca^{2+}- dependent activation of the channel.

Finally, the presence of Asp repeats also in thrombospondins and BK_{Ca} channels suggests that they could represent a general motif for Ca^{2+} binding in different classes of proteins.

5.3.4.1 Results of a recently published study by Xiao et al.

After the submission of this work for publication, a paper by Xiao et al. [176] has been published discussing the role of Ca^{2+} regulation in human bestrophin-1 (hBest1). They confirmed direct binding of Ca^{2+} to hBest1 Asp-rich domain by the autoradiography of the blot containing hBest1 C-terminus after its exposure to the $^{45}Ca^{2+}$. They studied different point mutations for each residue in the Asp-rich domain by measuring the currents after overexpression of the mutated proteins in HEK cells.

Based on the experimental results, they constructed a homology model of the Asp-rich domain using the third EF hand domain of human calmodulin (CaM, PDB code: 1CLL, [177]) as a template. They suggested one Ca^{2+}-binding site formed by the residues D312, N314, Q316, S318 and D323. However, they measured a reduced current only for D312 and D323, while the other residues were considered by analogy to the general EF-hand motif. In agreement with our simulations, they pointed out the role of D312 and D323 for Ca^{2+} binding. However, they did very limited theoretical

investigation: they estimated free energy of Ca^{2+} binding to a single configuration of WT and D312G mutant by implicit solvent Poisson-Boltzmann calculation. The results were not cross-checked by MD simulations in explicit solvent which is crucial for metal-ions based proteins.

Part III

Prion protein

Chapter 6

D18 scFv fragment in prion diseases therapy

Abstract

Prion is infectious protein responsible for a group of fatal neurodegenerative diseases called transmissible spongiform encephalopathies (TSEs), or prion diseases. In mammals, prions reproduce themselves by recruiting the normal cellular protein (PrP^{C}) and inducing its conversion to the disease-causing isoform denominated PrP^{Sc}. Recently, anti-prion antibodies have been shown to permanently cure prion-infected cells. However, the inability of full-length antibodies and proteins to cross the blood brain barrier (BBB) hampers their use in the therapy of TSEs *in vivo*. Alternatively, brain delivery of prion-specific single-chain variable fragment (scFv) by adeno-associated virus (AAV) transfer delays the onset of the disease in infected mice, although protection is not complete. We investigated the anti-prion effects of a recombinant anti-PrP (D18) scFv by direct addition to scrapie infected cell cultures or by infection with both lentivirus and adeno-associated virus (AAV) transducing vectors. We show that recombinant anti-PrP scFv is able to reduce PrP^{Sc} content in infected cells. In addition, we demonstrate that lentiviruses are more efficient than AAV in gene transferring of the anti-PrP scFv gene and in reducing PrP^{Sc} content in infected neuronal cell lines. Finally, we have used a bioinformatics approach to construct a structural model of D18scFv-PrP^{C} complex. Interestingly, according to the docking results, $Arg_{PrP}151$ (Arg151 from prion protein) is the key residue for the interactions with D18scFv, anchoring the PrP^{C} to the cavity of the antibody. Taken together, these results indicate that combined passive and active immunotherapy targeting PrP might be promising strategies for therapeutic intervention in prion diseases.

The experimental work was carried out by V. Campana, P. Casanova and Prof. C. Zurzolo at the Unité de Trafic Membranaire et Pathogénèse, Institut Pasteur, Paris, France; by L. Zentilin and M. Giacca at the International Centre for Genetic Engineering and Biotechnology (ICGEB), Trieste, Italy; by I. Mirabile and Prof. G. Legname at the Neurobiology sector of SISSA and by Prof. S. B. Prusiner from the Institute for Neurodegenerative Diseases, University of California San Francisco, U.S.A.

6.1 Introduction

Transmissible spongiform encephalopathies (TSEs) or prion diseases are fatal neurodegenerative disorders that affects humans and animals. The etiology of these pathologies can be infectious (like bovine spongiform encephalopathy or new variant Creutzfeldt-Jakob disease), sporadic or genetic. Industrial cannibalism has been responsible for one of the most infamous TSE epidemics, bovine spongiform encephalopathy (BSE), paralleled with an increasing number of human cases with a variant form of Creutzfeldt-Jakob disease from BSE-contaminated beef products [178]. All of TSEs are caused by a prion or PrP^{Sc} which is a conformational isoform of the cellular prion protein PrP^{C} [179].

PrP^{C} is a α-helical monomeric protein (Figure 6.1). The fold consists of a structured and an unstructured part [180]. The structured part consists of three α helices (residues D144-R156, N173-K194 and E200-Q227) and two β strands (residues M129-A133 and Q160-Y163) (PDB code: 1HJM [180]). The unstructured part consists of first 124 residues. Its structural features could not be determined so far.

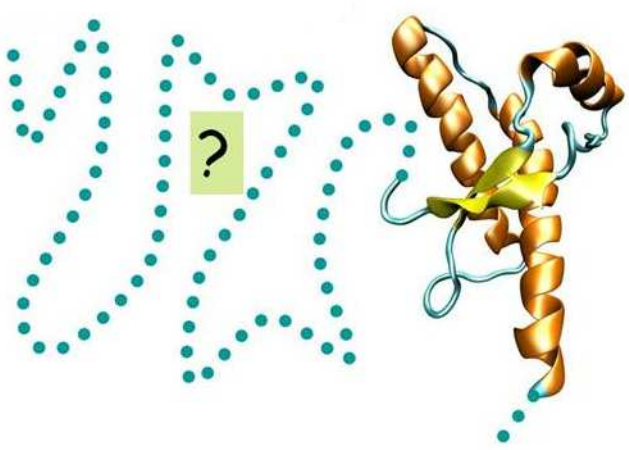

Figure 6.1: NMR structure of human PrP^{C} at pH=7. The dotted line represents the unstructured N-terminal part.

PrP^{C} is normally expressed in mammals, although its physiological role is not yet fully understood. It has been proposed to be involved in copper and/or zinc transport or metabolism [181, 182, 183], protection from oxidative stress [183], cellular signaling [184, 185], membrane excitability and synaptic transmission [186, 187], neuritogenesis [188] and apoptosis [184, 189]. In contrast, PrP^{Sc} is composed mostly by β-strands [179], and is therefore able to aggregate. In addition, it is partially resistant to the degradation by the proteases [179]. All these features concurrently result in a substantial difference, i.e. PrP^{Sc} is infectious, whereas PrP^{C} is not. The key element in pathogenesis of prion diseases is a direct interaction between the pathologic PrP^{Sc} and the endogenous PrP^{C} and the following conformational conversion of PrP^{C} into PrP^{Sc}. Conversion can be assisted with yet-unknown molecular chaperon (chaperon X) [190, 191]. PrP^{Sc} then agregates and forms fibrils that are deposited in neurons what leads to the neurodegeneration and rapid death [192].

Over the past years, many anti-prion compounds have been identified in *in vitro* models of prion replication, such as polysulphated polyanionic compounds, polyamine, tetrapyrroles, polyene antibiotics, peptides, tetracyclic and tryciclic compounds (reviewed in [193]). However most of these molecules were found to be toxic or ineffective in animal models of prion diseases (reviewed in [194, 195, 193]).

Alternatively, anti-prion protein antibodies have been shown to have an antiprion effect in cellular and animal models [28, 196, 197, 198]. Different publications report that antibodies directed against the middle portion of PrP (residues 91-110 and 132-156) can cure scrapie infected cells in culture and reduce PrP^{Sc} in spleens of infected mice (reviewed in [28, 196, 199]). Despite these encouraging results, development of effective immunotherapy presents several problems both in active and passive approaches. An important obstacle in the development of efficacious regimens for active immunization is host tolerance to endogenous PrP^{C}, which limited the therapeutic efficacy of this immunization approach [198, 200]. In comparison, passive immunization suffers from the intrinsic problem of poor antibody diffusion from vessels into tissues, especially in the nervous tissue: administration of monoclonal antibodies has been shown to prevent the pathogenesis only when applied simultaneously, or shortly after, peripheral prion infection [199]. Moreover, production of large amounts of monoclonal antibodies for therapy is technically challenging and expensive.

In 2001, Peretz and co-workers [28] analyzed the ability of some recombinant antibody antigen-binding fragments (Fabs; see Section 6.1.1) to inhibit prion propagation in a cultured mouse neuroblastoma cell line infected with PrP^{Sc}. The most effective Fab, D18, was found to abolish prion replication and to clear pre-exiting PrP^{Sc}, eliminating 50% of PrP^{Sc} from the cells within about 24 hours. The activity of D18 was attributed to its ability to specifically recognize the total population of PrP^{C} molecules on the cell surface. In PrP^{C}, D18 epitope spans from residues 132-156 and incorporates helix A. This sequence lays within the region of the protein thought to bind PrP^{Sc}, an essential step for prion propagation: therefore, it can be argued that D18 operates mechanistically by directly blocking or modifying interaction of PrP^{C} with PrP^{Sc} [28]. Compared to full-length antibodies, Fabs are smaller molecules but their transport via blood-brain barrier *in vivo* is still negligible. A promising solution to this problem is offered by the use of single chain variable fragments or scFv (see Section 6.1.1). ScFv are monovalent mini-antibodies, that maintain antigen specificity and can be engineered for intracellular expression or secretion.

So far, two groups have investigated the use of scFv in prion-infected cell culture system. One group showed the paracrine inhibition of prion replication by RD-4 cells expressing and secreting scFv(6H4) on co-cultured ScN2a (chronically prion infected neuroblastoma cells) [201]. Another interesting study showed the retention of PrP^{C} in the Endoplasmic Reticulum of HEK 293 and PC12 cells after the expression of a scFv from antibody 8H4 and 8F9, containing the ER retention sequence KDEL [29]. Starting from these results a subsequent set of *in vivo* analysis were carried on showing that mice intracerebrally injected with KDEL-8H4-NGF-differentiated PC12 cells infected with scrapie did not develop scrapie clinical signs or show any brain damage [30].

Therefore, these data would support the use of scFv as therapeutic approach. Additional support to the potential therapeutic value of intracerebral antibody production has recently been given by two studies showing that brain delivery of either prion-specific scFv by adeno-associated virus (AAV) transfer [202] or of a soluble prion antagonist (PrP-Fc2) by lentivirus transfer [203] delay the onset of the disease in infected mice, although protection in these studies was not complete. Because infectious prion replication occurs peripherally within the lymphoreticular system organs, the incomplete protection given by gene transferring of anti-prion molecules in the brain may be consequent to the fact that the peripheral pool of PrP^{Sc} continues to replicate and then migrate to the CNS by neuroinvasion [202].

In order to develop a system that could have therapeutic and not only prophylactic utility, in this study we combined active and passive immunotherapeutic approaches. First, we have produced a monovalent version of the anti-prion D18Fab, D18scFv, and we have tested it in prion-infected cells.

The big improvement of our strategy compared to the previous gene transferring is that our molecule might have effects both peripherically and into the brain. We show here that this recombinant anti-prion scFv is able to reduce PrP^{Sc} accumulation in different infected cell models of prion diseases.

Since it was argued that D18Fab operates mechanistically by directly blocking or modifying interaction of PrP^{C} with PrP^{Sc}, we have used bioinformatics approaches to construct a structural model of D18scFv-PrP^{C} complex and define important interactions, which prevent prion propagation. Based on the model of D18scFv-PrP^{C} complex we propose the antigenic determinant of PrP^{C} and the mechanism through which D18scFv prevents interactions of PrP^{C} with PrP^{Sc} or fibrillation of PrP^{Sc}.

Furthermore, in order to provide a localized supply of D18scFv (that could obviate the needs of repetitive injection in therapy), we have evaluated the *in vitro* efficiency of two virus-mediated gene transfer systems. We produced two viral vectors, one based on recombinant AAV serotype 2 (AAV2) and the other on HIV-1, carrying the D18scFv gene and showed that the latter one is more efficient than the AAV in reducing prion replication, at least in cultured infected cells *in vitro*. These results provide additional support to the use of small antibody fragments for active and passive immunotherapy in prion diseases and encourage the *in vivo* evaluation of an anti-prion scFv-based treatment for the therapy of prion disorders.

6.1.1 Antibody fragments

Antibodies are immune system-related proteins called immunoglobulins. They are "Y" shaped molecules composed of two identical long polypeptides (Heavy or H chains) and two identical short polypeptides (Light or L chains) (Figure 6.2). The light chains have variable (VL) and constant (CL) domains, while the heavy chains have a variable (VH) domain and three constant domains (CH1-3) (Figure 6.2).

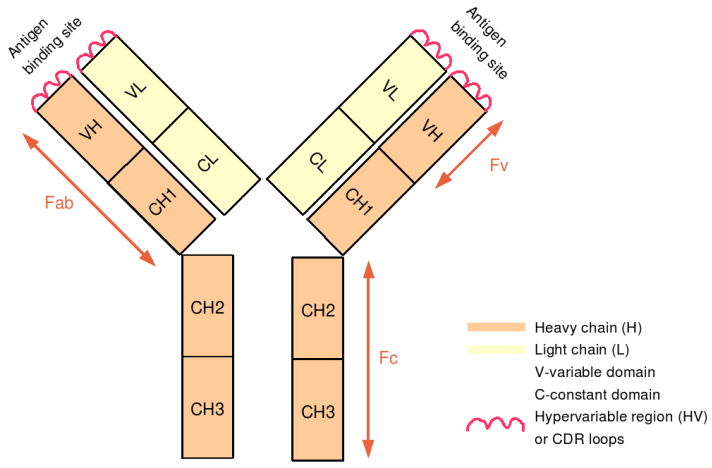

Figure 6.2: Immunoglobulin molecules are composed of heavy chains (orange) and light chains (yellow) joined by disulphide bonds so that each heavy chain is linked to a light chain and the two heavy chains are linked together. The immunoglobulin molecule can be dissected by partial digestion with proteases into three pieces, two Fab fragments and one Fc fragment. The Fab fragment contains the V regions and binds antigen. The Fc fragment is crystallizable and contains C regions.

All immunoglobulin domains are made up of two antiparallel β-sheet structures. The sheets are linked by a disulphide bridge and packed to a β barrel or a β sandwich. The distinctive folded structure of the antibody domains is known as the immunoglobulin fold. The variable domains, one from the heavy and another from the light chain, form the two antigen binding sites, one at the tip of each arm of the "Y" (Figure 6.2). The variable domains consist of hypervariable (HV) and framework (FR) regions. The HV regions are represented as loops most closely involved in binding to the antigen; called also complementarity determining regions (CDRs). Within light and heavy chains exist three CDRs, CDRL1-3 and CDRH1-3, respectively, that differ in length and sequence among different antibodies and are mainly responsible for the specificity (recognition) and affinity (binding) to their antigens. Four framework regions composed from β-sheets serve as a scaffold to hold CDRs in position to contact an antigen. The constant domains determine the mechanisms for antigen destruction.

Treating antibodies with a proteases release different fragments termed (Figure 6.2):

1. **Fab** (Fragment antigen binding) - they correspond to the two identical arms of the antibody molecule, which contain the complete light chains paired with the VH and CH1 domains of the heavy chains. Fab fragments contain antigen-binding activity.

2. **Fc** (Fragment crystallizable) - was originally observed to crystallize readily, and for this reason named the Fc. It corresponds to the paired CH2 and CH3 domains and is the part of the antibody molecule that interacts with effector molecules and cells. It contains no antigen-binding activity.

3. **Fv fragment** (fragment variable) - a truncated Fab comprising only the VH domain linked by a stretch of synthetic peptide to a VL domain (see Figure 6.4A). This is called single-chain variable fragment (**scFv**). scFv molecules may become valuable therapeutic agents because of their small size, which allows them to penetrate tissues readily.

6.2 Computational details

The sequence of the D18scFv [27] was compared to the sequences of known 3D structures collected in Protein Data Bank [1, 2] in order to find templates for homology modeling. Three templates were identified: the anti-(carcinoembryonic antigen) single chain Fv antibody MFE-23 (PDB code: 1QOK) [204], the anti-DNA binding antibody (PDB code: 2GKI) [205] and the anti-prion protein P scFv fragment (PDB code: 2HH0) [206]. The anti-prion protein P scFv fragment was further used as the template for modeling a 3D structure of D18scFv. 3D models were built using MODELLER 6v2 [48]. Amino acids of a signal sequence and poly-glycine parts of the sequences were neglected in the modeling. The structure of the D18scFv model was then used to dock the prion protein (PrP; PDB code: 1HJM) [180] by means of the HADDOCK2.0 program [207, 208]. Residues determined to be important for protein-protein recognition between PrP and D18scFv fragment were reported in the literature [27]. For the purpose of docking they were filtered according to the solvent accessibility with the NACCESS program [209]. The following residues were used as interactive residues in docking process: (i) PrP - Ala133, Ser135, His140, Gly142, Ser143, Asp144, Tyr145, Arg148, Arg151, Glu152, His155 (ii) D18scFv fragment - Tyr100, Tyr101, Gly102, Ser207, Asn208; Thr28, Asn54, Thr55, Val57, Gly59, Ser141, Ser142, Ser143, Asn145, Thr146, Ser166. All His residues were deprotonated. The remaining docking parameters were kept as default. Cluster analysis was then performed with programs available in HADDOCK2.0.

All details about experimental can be found in ref. [210].

Figure 6.3: Schema of the D18scFv in pET-22b(+). **A:** Sequence of the D18scFv. The different regions of the sequence are indicated with different colors: the pelB leader is light gray, the VH and VL sequences are red, the 20-AA linker is grey and the 6- histidine tag is yellow. **B:** Cloning of the D18scFv in pET-22b(+). The D18scFv was obtained by PCR ligation of the variable regions of the light [42] and heavy (VH) chains of FabD18, by inserting a 20-AA linker. The sequence corresponding to the scFv (A) was inserted between BamHI and NotI restriction enzyme sites present in the multiple cloning site of pET-22b(+) expression vector by keeping the pelB leader and the 6-histidine tag already present in the vector.

6.3 Results and discussion

It has been recently shown that anti-prion scFvs could be delivered to the central nervous system of mice by adeno-associated virus (AAV) and are able to delay the onset of the disease when used in prophylaxis but not when used after prion-exposure [202]. Contribution to the incomplete protection could derive from down-regulation of the scFv expression by PrP^{Sc} neuroinvasion. In addition, because infectious prions replicate peripherally within the lymphoreticular system organs, the incomplete protection may be consequent to the fact that the peripheral pool of PrP^{Sc} continues to replicate and migrate to the central nervous system by neuroinvasion. These observations make inefficient anti-prion treatments based exclusively on active immunotherapy. Since it has been shown that small proteins can be rapidly delivered and spread in the brain after i.c. injection or after nasal administration [211, 212], we have evaluated, in cell culture, an alternative approach to treat TSEs based on passive immunotherapy. To this aim, we have produced a smaller version of the D18Fab, the D18scFv in which the variable regions of the light and heavy chains of Fab were fused together. The sequence corresponding to the scFv (Figure 6.3A) was cloned in a bacterial expression system (Figure 6.3B). In order to directly test for interaction and anti-prion activity, D18scFv was evaluated for its binding to PrP. We tested the ability of D18scFv to bind recombinant PrP and we found that D18scFv specifically interact with PrP and competes with the Fab for its binding.

6.3.1 Molecular modeling of D18scFv-PrP^{C} complex

Due to the interest on how D18scFv interacts with the PrP^{C}, and consequently prevents formation of PrP^{Sc}, we have used bioinformatics approaches to construct a structural model of D18scFv-PrP^{C} complex. Firstly, the 3D structure of D18scFv was modeled based on anti-prion protein P scFv fragment (PDB code: 2HH0) [206]. The anti-prion protein P scFv fragment was crystallized in a complex with a small peptide. Due to the presence of a small peptide, the scFv fragment represents an "open" conformation, where complementarity determining regions 3 (CDR3) are further apart than in non-complexed antibody structures. Therefore, this structure was taken as a template for constructing the D18scFv model, since it allows an easy approach of the PrP^{C} to the cavity of the antibody in the docking procedure.

Like other single chain variable fragments also D18scFv is composed by heavy (VH) and light (VL) hypervariable domains. CDRs of the heavy chain (H29-H33, H53-H57, H98-H105) and light chain (L141-L146, L164-L166, L205-L210) form the paratope that interacts with the PrP residues Ser132-Arg156 [27] (Figure 6.4A).

A D18scFv-PrP^{C} complex was built by docking of PrP^{C} (PDB code: 1HJM) to the model of D18scFv (Figure 6.4A). According to the docking results $Arg_{PrP}151$ represent the antigenic determinant, a key residue for interactions with the D18scFv (Figure 6.4B). An antigenic determinant is a site on the antigen that the immune system responds to by making antibody and is frequently represented by one residue on the antigen. $Arg_{PrP}151$ anchors the PrP to the cavity of the antibody forming H-bonds with $Tyr_{VH}32$, $Asn_{VH}33$, $Asp_{VH}35$ and $Tyr_{VL}210$, along with van der Waals interactions with $Trp_{VL}205$ (Figure 6.4B). In addition the Cβ atom of $Arg_{PrP}151$ forms hydrophobic interactions with the CE1 and CZ atoms of $Tyr_{VH}32$.

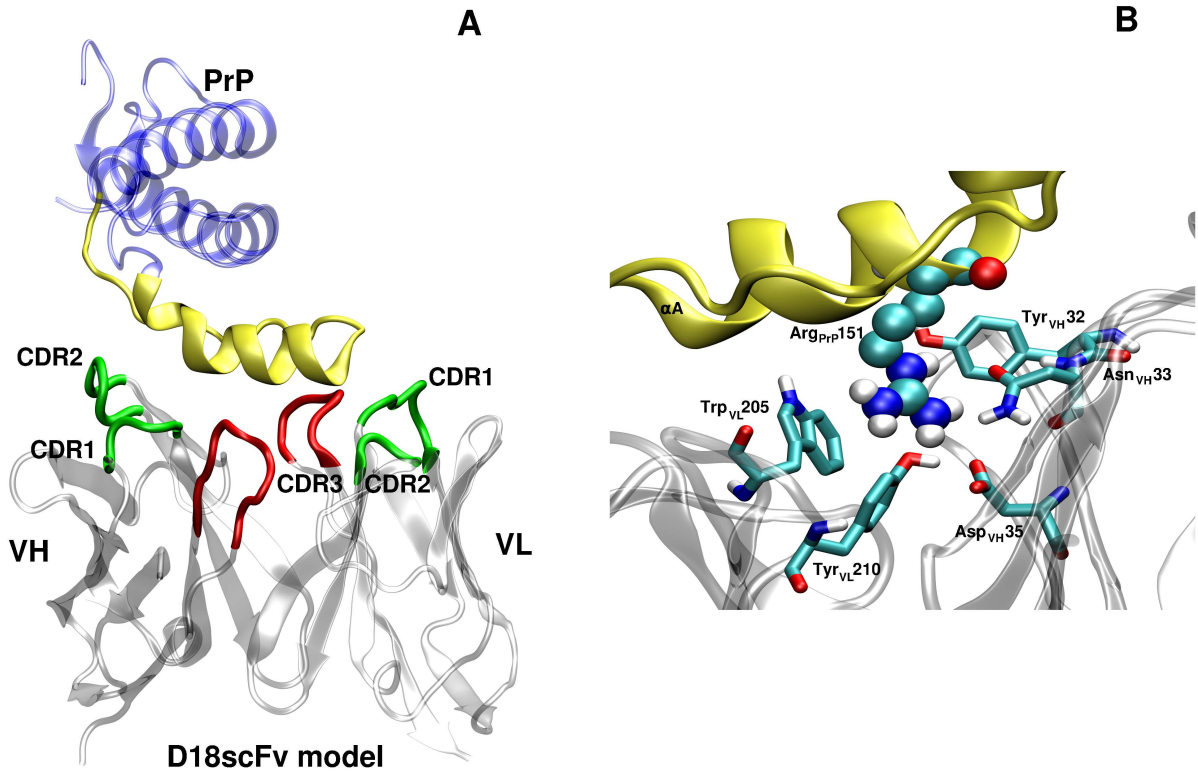

Figure 6.4: Model of interactions between D18scFv and PrP. **A:** Complex of the D18scFv (silver) and PrP (blue). Epitope residues of PrP (yellow) interact with the six CDR loops (red - CDR3, green - CDR1,2) of the antibody fragment. **B:** Arg151 anchors the PrP to the antibody's cavity through the H-bond and van der Waals interactions with Tyr32, Asn33, Asp35, Trp205 and Tyr210.

$Arg_{PrP}151$ seems important for recognition. But how D18scFv stabilizes the PrP^C? A theoretical study was done [213], where it was shown that PrP^C βA-αA loop (residues 135 to 140 in the PrP epitope) is directly involved in protofibril formation (Figure 6.5A). Under certain conditions (e.g. low pH) α helical formation of PrP^C starts to convert to β strand rich PrP^{Sc} structure. A loop connecting β A strand and α A helix of PrP^C changes from random coil to β strand. The newly formed β strand interacts with βA strand from other PrP^{Sc} molecule leading to the fibrillation. When PrP^C is bound to the D18scFv, the βA-αA loop is stabilized by H-bond interactions (Figure 6.5B) and cannot convert to β strand. Due to this stabilization the formation of fibrils is prevented. The interactions stabilizing βA-αA loop are formed by $Arg_{PrP}136$ and $His_{PrP}140$ that are part of the long loop connecting βA strand and α helix of PrP. The hydrogen atom from NH_2 group of $Arg_{PrP}136$ interacts with the side chain oxygen atom of $Thr_{VH}55$ and the nitrogen atom Nε2 of $His_{PrP}140$ accepts the H-bond from the amidic oxygen of $Asn_{VL}208$ or from the carbonyl group of $Ser_{VL}207$ (Figure 6.5B).

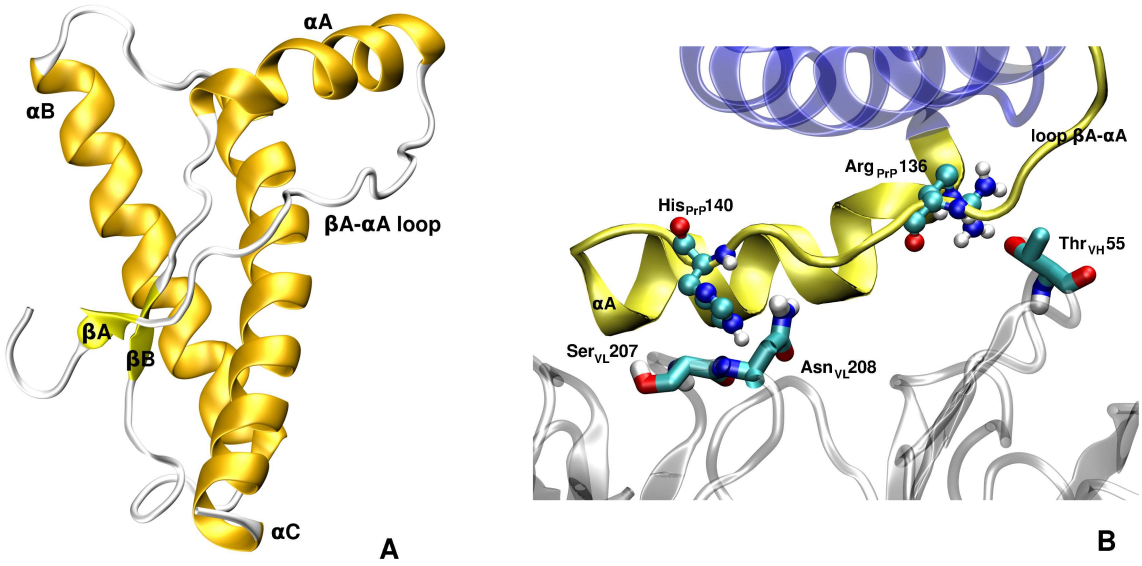

Figure 6.5: **A:** It was argued that loop connecting βA strand and αA helix can convert to β strand, interacts with βA strand of another PrP^{Sc} molecule and therefore enable prion aggregation. **B:** Conversion of βA-αA loop is prevented when PrP^{C} is bound to D18scFv due to the H-bond interactions between the $Arg_{PrP}136$ and Thr55 among with the interactions of $His_{PrP}140$ with the Asn208 or Ser207.

A series of other H-bonds and hydrophobic interactions contribute further to the scFv-PrP complex stability.

Additional H-bond interactions with the antibody are mostly formed by residues present in the α helix of the PrP (Figure 6.6): $Asp_{PrP}144$-$Lys_{VL}167$, $Tyr_{PrP}145$-$Tyr_{VH}100$, $Asp_{PrP}147$-$Trp_{VL}205$ & $Tyr_{VL}146$, $Arg_{PrP}148$-$Tyr_{VH}32$, $Glu_{PrP}152$-$Tyr_{VH}52$ and $His_{PrP}155$-$Tyr_{VH}52$. In the cavity of the antibody there are only few hydrophobic residues, namely $Trp_{VH}47$, $Val_{VH}57$ and $Trp_{VL}205$. Among them just $Val_{VH}57$ forms hydrophobic interactions with the Cβ atom of $Ser_{PrP}135$. In addition $Tyr_{VH}50$ and $Tyr_{VH}52$ form hydrophobic interactions with $Ile_{PrP}138$ and $Met_{PrP}154$, respectively. All other hydrophobic interactions involve the methyl/methylene groups of polar or charged residues.

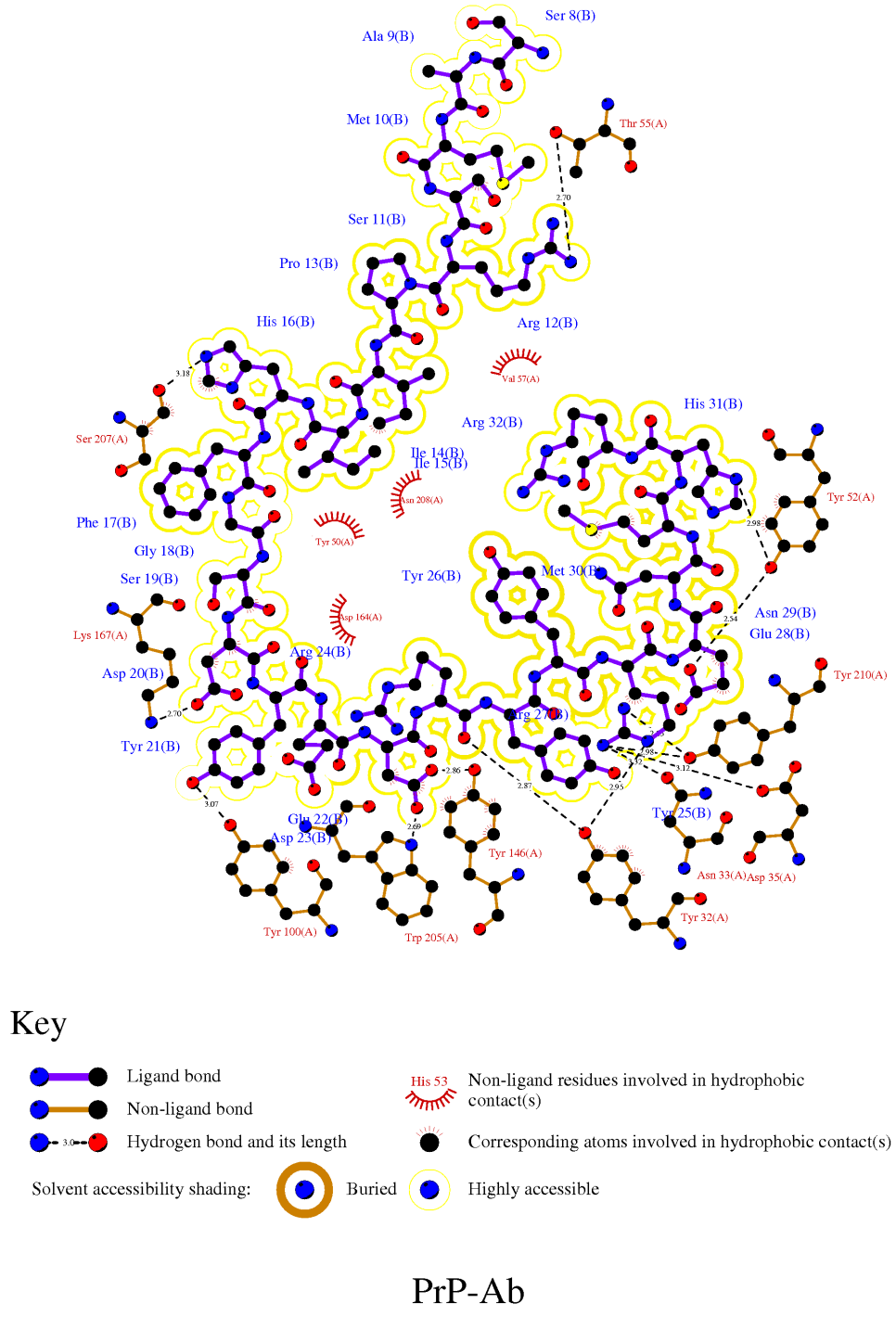

Figure 6.6: Hydrophobic and H-bond interactions of the PrP epitope (Ser8 - Arg32; B) with D18 scFv (A), as depicted by LIGPLOT [13]. Please note that the numbering of PrP residues shown above differs from the original one used in the text: Ser8 corresponds to Ser132 and Arg32 to Arg156 in the original numbering (conversion: ligplot res+125=PrP res).

6.3.2 In vitro studies of scFv

The next step in experiments was to test D18scFv for its effect on scrapie replication in scrapie infected cell lines (Figure 6.7).

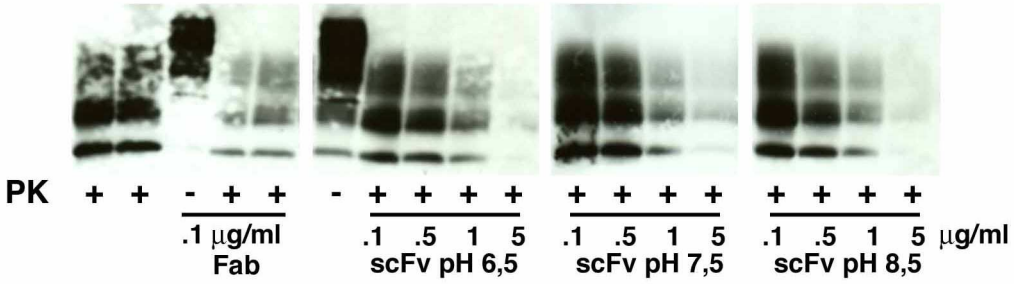

Figure 5

Figure 6.7: Effect of D18scFv in curing scrapie infected GT1 cells. ScGT1 cells were incubated for 1 week with different concentrations (0.1; 0.5; 1 and 5 μg/ml) of scFv refolded at different pH (pH 6.5; 7.5 and 8.5). As controls, R2 (10μg/ml) and D18Fab (1 and 5 μg/ml) were used. PrP^{Sc} present in cell lysates was digested with Proteinase K (PK +) as described in Experimental and detected by western blot by using SAF61 antibody. Note that scFv reduces PrP^{Sc} level of 20%±4, 50%±7, 80%±3 and 95%±1 by using respectively 0.1, 0.5, 1 and 5 μg/ml D18scFv. Blots of PK-resistant assays were quantified by ImageJ (n=3) and Student T-test was used to calculate significance of data ($p<0.05$). No difference was found among the scFvs refolded at the different pH.

scFv was found to reduce PrP^{Sc} levels similarly to what was found for the Fab in ScN2a cells [28]. Interestingly, the same concentration of Fab and scFv is required to achieve the same extent of reduction in PrP^{Sc} level. However, since the scFv is approximately half molecular weight compared with the Fab, D18scFv is 2-fold less effective than Fab with respect to the reduction of the PrP^{Sc} level in ScGT1 cells. After 1 week treatment, PrP^{Sc} was reduced to undetectable levels but was not completely eliminated from the cells because it reappeared after 1 additional week of growth in the absence of the scFv. However, after 2 or 3 weeks of scFv treatment PrP^{Sc} remained undetectable for up to 3 weeks of culture without adding any scFv, as it has been shown for the original Fab [28]. These data indicate that the small monovalent antibody fragment, scFv, might be a very useful tool in therapy because it is able to interfere with prion metabolism (either reducing conversion or promoting degradation) in infected cells. The big improvement of our strategy compared to the previous gene transferring-based approaches [202] is that our molecule might have effects both peripherically and into the brain, both of them sites of prion replication. In vivo the putative anti-prion effects of the D18scFv expressed as recombinant protein might be enhanced by combining passive and active immunotherapeutic approaches.

Recently, different gene transfer approaches in anti-prion prophylaxis of infected animals have been successfully used by using either lentivirus carrying a soluble prion antagonist [202] or AAV transducing either anti-LPR/LR [32] or anti-prion protein specific scFvs [202]. Therefore, in order to have a continuous supply of our recombinant scFv to be eventually used in immunotherapy in animals and to test the effectiveness of these two viral systems, we cloned the gene encoding for D18scFv (Figure 6.3A) either in a defective lentiviral vector (Figure 6.8A, left panel) or in an AAV vector (Figure 6.8A, right panel). Lentivirus carrying the D18scFv was titered and tested for its effectiveness in reducing PrP^{Sc} content in infected ScGT1 and ScN2a cell cultures. One week posttransduction D18scFv was expressed (Figure 6.8B) and the levels of PrP^{Sc} were reduced of 80%±7 and 90%±3 respectively in ScGT1 and ScN2a cells by using the highest viral titer tested (10x106 TU) (Figure 6.8C). Interestingly, the effect on PrP^{Sc} content after lentivirus transduction was more pronounced in ScN2a cells, where even a lower titer of $10x10^4$ TU was effective (Figure 6.8C). Alternatively, we also produced a recombinant (rec) AAV2 expressing D18scFv. Titers of $10x10^{10}$ up to $10x10^{12}$ TU were used to transduce both ScGT1 and ScN2a cells and transgene expression (Figure 6.8B). One week post-transduction only the highest titer tested ($10x10^{12}$ TU) was able to reduce PrP^{Sc} level of 20%±5 in ScN2a cells (Figure 6.8C).

We have shown that, at least in cell culture models, lentivirus transduces D18scFv in a functional form able to prevent or interfere with PrP^{Sc} conversion and seem to represent a better delivery system compared to AAV. However, AAV-scFv remains a valuable potential tool for the in vivo treatment of prion disease. How these viruses act in reducing PrP^{Sc} levels has to be still elucidated. One possibility could be that anti-PrP scFvs block conversion reaction directly. Alternatively, PrP^{Sc} reduction could be an indirect effect of down-regulation of PrP^{C} expression. Nevertheless, the eventual regulation of PrP^{C} expression is specifically due to the scFv, because viruses expressing other constructs (GFP-PrP) do not reduce PrP^{Sc} content in scGT1 cells.

6.4 Conclusions

In conclusion, we validated here the use of D18scFv as potential anti-prion immunotherapeutic both as purified protein and after viral transduction in cultured neuronal cells. The availability of our lentiviral and AAV2 vectors expressing the same therapeutic molecule will now allow us to directly compare the efficacy of the two viral vectors in sick animals *in vivo*. With bioinformatics approach we modeled D18scFv-PrP^{C} complex and defined $Arg_{PrP}151$ as the antigenic determinant important for the recognition between the two proteins. We proposed that H-bond interactions of $Arg_{PrP}136$ and $His_{PrP}140$ with the D18scFv stabilize PrP^{C} and therefore prevent fibrillation process. All these data give new suggestions for mutational studies that will elucidate the exact mechanism of action for the D18scFv.

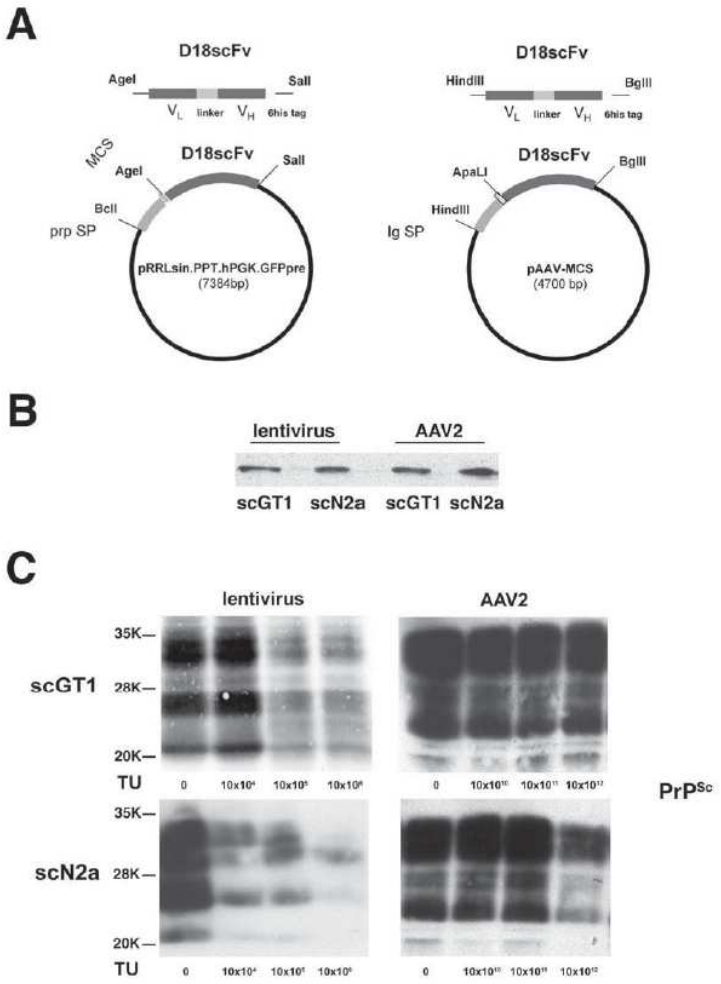

Figure 6.8: Production and cellular test of lentivirus and AAV2 containing the D18scFv. **A:** Cloning of the D18scFv in pTTLsin.PPT.hPGK.GFPpre and in pAAVMCS. For lentiviral vector production (left panel), the gene encoding for D18scFv (Figure 6.3A) was cloned in defective lentiviral vector (pTTLsin.PPT.hPGK.GFPpre) between AgeI and SalI restriction enzyme sites. The signal peptide of the PrP sequence was cloned upstream between BclI and AgeI. For AAV vector production (right panel), the gene encoding for D18scFv was cloned in pAAV-MCS between HindIII and BglII restriction enzyme sites. For secretion, the signal peptide of the immunoglobulin (Ig) sequence was added between HindIII and ApaLI. **B:** Expression of D18scFv after transduction of ScGT1 and ScN2a cells. $10x10^4$ TU of lentivirus and $10x10^{10}$ TU of rAAV2 were used to transduce both ScGT1 and ScN2a cells. One week post-infection, D18scFv expression was tested by western blot by using an anti- 6his tag antibody. **C:** Effect of the lentivirus and rAAV2 carrying the D18scFv in curing scrapie infected neuronal cell lines. Titers of $10x10^4$-$10x10^6$ TU of the lentivirus (left panels) and $10x10^{10}$-$10x10^{12}$ TU of rAAV2 (right panels) carrying the D18scFv were used to infect both ScGT1 and ScN2a cells. One week post-infection, PrP^{Sc} present in cell lysates was digested with proteinase K (PK +) as described in Experimental and detected by western blot by using SAF61 antibody. Blots of PK-resistant assays were quantified by ImageJ (n=3) and Student T-test was used to calculate significance of data ($p<0.05$). By using increasing titer of lentivirus, levels of PrP^{Sc} were reduced of 2%±7, 75%±3 and 80%±7 in ScGT1 and 76%±2, 80%±7 and 90%±3 in ScN2a cells. By using increasing titer of AAV2, levels of PrP^{Sc} were reduced of 2%±3, 5%±3 and 5%±4 in ScGT1 and 4%±2, 4%±7 and 20%±5 in ScN2a cells. Note that levels of PrP^{Sc} were significantly reduced in both cell systems by using the highest lentiviral titer tested ($10x10^6$ TU).

Chapter 7

A Computational Protocol for Docking Ligands to the PrP

Abstract

Molecular docking of ligands targeting proteins undergoing fibrillation in neurodegenerative diseases is difficult because of the lack of deep binding sites in these proteins. Here we combine standard docking methods with MD simulations and bias-exchange metadynamics based-free energy calculations in explicit solvent to address this issue in the context of the prion protein. We focus on the GN8 ligand (2-pyrrolidin-1-yl-N-[4-[4-(2- pyrrolidin-1-yl-acetylamino)-benzyl]-phenyl]-acetamide) , which has been shown to bind to the structured part of the protein, increasing its stability. Our calculations enable to identify alternative GN8 binding poses that satisfy all of the contacts emerging from NMR, in contrast to a previous model. Most importantly, our results provided an estimation of the experimental free energy of binding which is in very good agreement with the experimental one. We conclude that our approach is a useful tool to predict poses and affinity of ligands binding to proteins with fibrillation properties.

This work was carried out in collaboration with S. Bongarzone, G. Rossetti, X. Biarnes (Statistical and Biological Physics Sector, SISSA/ISAS).

7.1 Introduction

Recent developments in molecular docking protocols (MDPs) allow to predict accurately ligand poses in their target binding sites [214, 215]. In many cases, the reason of their success lies in the coupling of traditional scoring function-based approaches with molecular simulations [216, 217] which introduces conformational flexibility of the target. This is very useful because proteins are in constant motion between different conformational states [218, 219] that may be further altered when a ligand is bound [219]. Different methods were proposed to address these issues, among them: soft harmonic modes [220], molecular dynamics simulations [221, 222, 223, 216] and relaxed complex method [224].

In spite of these successes, there are still many important cases for which MDPs are challenged. These include, the prediction of the poses of transition-metal and/or alkylating drugs, of ligands causing large structural changes, and of ligands *not* binding to specific pockets. The latter case

is crucial in most targets undergoing fibrillation in neurodegenerative diseases. Here we use the metadynamics along with MDP (Figure 7.1) to predict ligand binding poses onto protein surfaces for which a specific binding pocket is not defined. Metadynamics [33] has recently been shown to uncover molecular and energetics aspects of the drug binding mechanisms when the binding sites are well characterized by X-ray crystallography or NMR experiments [225, 226, 227, 228].

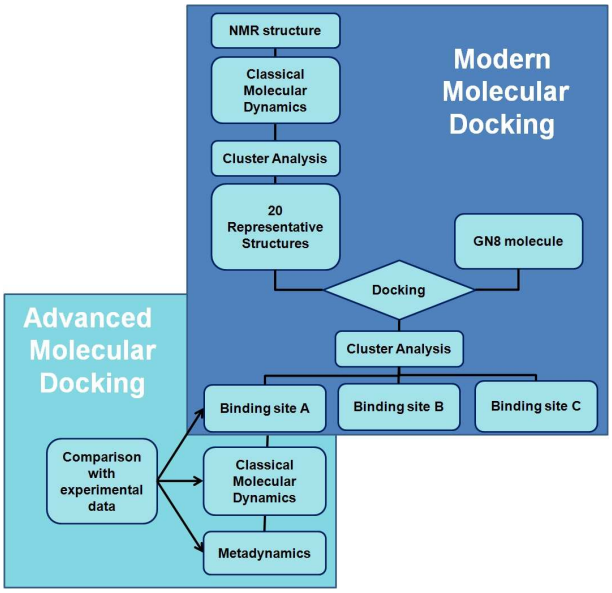

Figure 7.1: Our hybrid computational protocol combines docking, MD simulations and metadynamics technique (Advanced Molecular Docking).

We focus on a small organic molecule able to bind to the cellular form of human prion protein ($hPrP^C$), which can interfere with its conversion to the pathogenic form ($hPrP^{Sc}$, scrapie prion protein) [26]. This in turn is involved in the epidemics of bovine spongiform encephalopathy (BSE) and, possibly, in the new variant Creutzfeldt-Jakob disease (nvCJD) [24, 191, 229, 230, 231], for which there is neither early diagnosis nor a cure [232, 191, 229, 230].

Recently, chemical shift perturbations and binding affinity of a PrP-binding ligand, GN8 (2-pyrrolidin-1-yl-N-[4-[4-(2- pyrrolidin-1-yl-acetylamino)-benzyl]-phenyl]-acetamide, Figure 7.2) [35], have been measured for the mouse prion protein ($mPrP^C$), which is highly similar to the $hPrP^C$ (Seq. Similarity=98%).

The reported chemical shifts perturbations of $mPrP^C$ induced by GN8 binding affect most significantly amino acid residues on one side of the protein surface (Asn159@H1-S2 loop, Val189, Thr192, Lys194@Helix H2 and Glu196@ H2-H3 loop). In addition, Val189 and Thr192, located on the other side of the PrP surface, are also perturbed [35] (Figure 7.4A and Table 7.1). This suggests that multiple binding sites may be present. In fact, an *ad hoc* model of the $mPrP^C$-GN8 adduct, constructed by docking and energy minimization exhibited a single binding mode of GN8 connecting Asn159 and Glu196 [35]: such single binding mode could not explain the contacts with Val189 and Thr192.

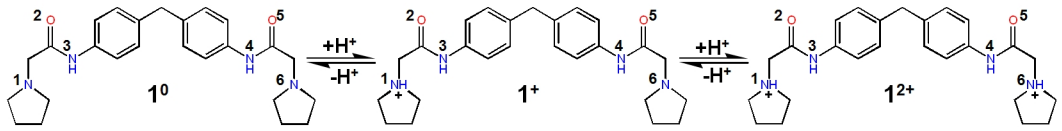

Figure 7.2: GN8 (2-pyrrolidin-1-yl-N-[4-[4-(2- pyrrolidin-1-yl-acetylamino)-benzyl]-phenyl]-acetamide) molecule.

Here we have applied our computational protocol to identify ligand binding modes of the GN8 molecule on the surface of the hPrP^C. Our calculations provide multiple binding modes of the ligands, consistent with all of the NMR contacts. Most importantly, it provides energetics in agreement with experimental data. The computational approach emerges therefore as a useful approach to investigate ligands sticking on protein surfaces.

σ>0.9 ppm	0.5ppm<σ<0.9 ppm
Asn159	Arg156
Val189	Glu196
Thr192	Thr199
Lys194	

Table 7.1: NMR chemical shift changes (σ) upon GN8 binding to the mPrPC .

7.2 Computational details

Identification of binding sites. The residues interacting with GN8 are located in hPrP^C C-term, for which the NMR structure is available (residues 125-228, PDB ID: 1HJM) [233]. Protonation states were assigned by the web server H++ [234, 235] assuming pH 7.4.

Putative binding sites were identified by **(i) molecular simulations** (using the GROMACS package [155] and **(ii) docking procedure** (using the GOLD [56, 57, 58, 236] and the Autodock programs [59]).

(i) Molecular simulations. The protein was inserted into a cubic box of water molecules, ensuring that the solvent shell would extend for at least 8 Å around the system. Three sodium counterions were added. The AMBER99 force field [149, 237] was used for the protein. Sodium ions were modeled with the AMBER-adapted Aqvist potential [161]. The water molecules were described by the TIP3P model [66]. The system was minimized imposing harmonic position constraints of 1000 kJ/(mol*nm^2) on solute atoms, allowing the equilibration of the solvent without distorting the solute structure. After an energy minimization of the solvent and the solute without harmonic constraints, the temperature was gradually increased from 0 to 298 K. This was performed by increasing the temperature from 0 to 298 K in 12 steps in which the temperature was increased by 25 K in 100 ps of MD. Constant temperature-pressure (T=298 K, P =1 bar) 20-ns MD dynamics was then performed through the Nose´-Hoover [81, 80] and Andersen-Parrinello-Rahman [238] coupling schemes. Periodic boundary conditions were applied. Long-range electrostatic interactions were treated with the particle mesh Ewald (PME) [76, 78] method, using a grid with a spacing of 0.12 nm combined with a fourth-order

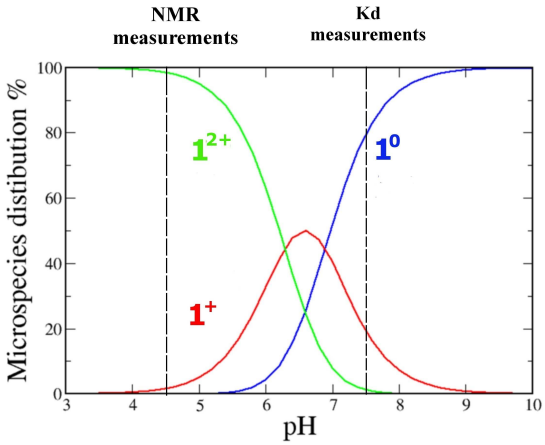

Figure 7.3: Titration curve of GN8. At the pH of the Kd measurements , it is present in neutral ($\mathbf{1}^0$) and monoprotonated ($\mathbf{1}^+$) states. At the pH of the NMR measurements , it exists in diprotonated ($\mathbf{1}^{2+}$) form.

B-spline interpolation to compute the potential and forces in between grid points. The cutoff radius for the Lenard-Jones interactions as well as for the real part of PME calculations was set to 0.9 nm. The pair list was updated every 2 steps, and the LINCS algorithm [151] was used to constrain all bond lengths involving hydrogen atoms allowing us to use a time step of 2 fs.

The MD trajectory of prion protein alone was clustered with the program gromos [239] and as result 20 different conformations were obtained, which were used along with the NMR structure for the docking of GN8 compound.

(ii) Docking. Titration curves, as calculated by the ChemAxon software [240], suggested that GN8 molecule is present in two protonation states (neutral, $\mathbf{1}^0$ and monoprotonated, $\mathbf{1}^+$) in aqueous solution at pH=7.4, while at pH=4.5 exists also in diprotonated form ($\mathbf{1}^{2+}$) (Figure 7.2 and 7.3). This method has been used because it appears to be rather reliable: in a calculation of pKa of 1000 molecules, less than 0.5 % calculations turned out to differ by more than 0.5 pH unit from the experimental value [240]. All three protomers underwent geometry optimization at the B3LYP/6-31G** level of theory by means of the Gaussian03 software (g03) [241]. The optimized structures, $\mathbf{1}^0$, $\mathbf{1}^+$ and $\mathbf{1}^{2+}$, were docked to the NMR structure of hPrP^C and to its 20 different conformations.

The GOLD 3.1 [57] and Autodock 3.0.5 [59] programs were used. In GOLD, the docking area was defined as a sphere of 35 Å radius around the His187, so that the whole protein was screened. The ChemScore (CS) [242, 243, 62] and GoldScore (GS) [244, 245] scoring functions were used for ranking. For each protomer and scoring functions, 100 docking runs were performed. In Autodock, a Lamarckian genetic search algorithm was used to identify low energy binding sites and orientations of GN8 protomers. Binding modes were ranked by a scoring function implemented in the Autodock. A point grid with a spacing 0.475 Å was used. Point grid was centered to the center of mass of the protein, its dimensions were 126 Å x 126 Å x 126 Å. Gasteiger atom charges were assigned to the proteins using AutoDockTools. Water molecules were excluded from the protein before docking. 100 randomly seeded runs were performed. The binding poses were identified by AClAP 1.0 clustering procedure [246, 247].

Hydratation and thermal stability of GN8- hPrPC adducts. 10-ns MD simulations of the adducts (hPrP^C-$\mathbf{1}^0$, hPrP^C-$\mathbf{1}^+$ and hPrP^C-$\mathbf{1}^{2+}$) allowed for a proper hydration of the system and for identification of collective motions that may be essential for PrP-ligand interactions. The protomers were bound to the binding site **I**. The simulation protocol was the same as for the free protein. For the three ligands, the gaff force field [248] was used. The atomic restrained electrostatic potential (RESP) charges [249, 250] were calculated by using the resp module of AMBER after geometry optimization and wave function relaxation.

Binding free energy calculations. The binding free energies of $\mathbf{1}^0$, $\mathbf{1}^+$ and $\mathbf{1}^{2+}$ were calculated using metadynamics [33] in its bias-exchange variant [34], as a function of collective variables (CVs) which should be relevant for describing the dissociation process. Those chosen here included: (i) the distance between the center of mass of the ligand and the center of mass of the protein binding region; (ii) the number of polar contacts between the ligand and one portion of the protein binding region; (iii) the number of polar contacts between the ligand and the other portion of the protein binding region; (iv) the number of water bridge contacts between the ligand and the protein binding region; (v) the RMSD difference of the system with respect to an equilibrated MD structure taken from the previous Section; (vi) the RMSD fluctuation of the residues defining the protein binding site. This set of CVs does not require the previous knowledge of the protein-ligand adduct structure, but is limited to the exploration of a region around the previously defined binding site **I**. The choice of such CVs was dictated by previous ligand-target interaction metadynamics studies [225, 227, 226, 228] as well as by observations based on the former MD simulations. For further details on this theoretical approach see Appendix 7.4.

Six independent metadynamics simulations were run in parallel. Each replica was biased by different one-dimensional time-dependent potentials, which were built as a function of each of the collective variables defined above. Exchanges among replicas were attempted every 10 ps by a Metropolis acceptance criterion [34]. Similar set-up has shown to improve the sampling of the configurational space and the convergence of the results [34, 251]. At the end of the different replica simulations, the explored phase space, in terms of the six collective variables used here, was clustered using the gromos method [239]. The clustering radius for each collective variable was set to 0.1 nm, 0.2, 0.4, 2.5, 0.05 nm, 0.02 nm, respectively. The free energy corresponding to each cluster was then reconstructed from the populations of clusters observed during the simulations. The free energy value was corrected by the corresponding bias potentials acting on that cluster as in a usual weighted histogram analysis [252].

Two reference states of the ligand-protein system, bound and unbound, needed to be defined to provide the corresponding binding free energy value. The bound state was here considered as the lowest free energy cluster. The unbound state was considered to be a cluster showing no contacts with the binding site **I** (lowest values of CVs ii and iii) and at the same time with a higher RMSD with respect to the initial docked structure (highest value of CV v). Given the size of the simulation box the ligand is not fully detached from the protein in its unbound state. Therefore, the residual binding energy in the unbound state was roughly estimated in implicit solvent using an adaptive Poisson Boltzmann solver [253]. It was estimated as the difference in solvation energy of the complex minus the solvation energy of each component plus the intermolecular Coulombic interaction. The standard free energy of binding was obtained by applying the following relationship: $\Delta G^0 = \Delta G + RT\log([L])$, where ΔG is the total binding free energy as a result of our simulation, R is the molar constant, and $[L]$ is the concentration of the ligand in our simulation box (i.e. 5.5 mM). The standard free energy is related to the dissociation equilibrium constant (Kd) by $\Delta G^0 = -RT\log(Kd)$.

7.3 Results

In this study, we set up a computational protocol to dock small molecules targeting cavity-less proteins such as hPrP^C. As prototype we used known active compound GN8 (2-pyrrolidin-1-yl-N-[4-[4-(2-pyrrolidin-1-yl-acetylamino)-benzyl]-phenyl]-acetamide), for which there are NMR chemical shifts and energetic information available [35].

Putative GN8 binding sites of hPrP^C. GN8 is a symmetric molecule, composed by two phenyl and pyrrolidin rings connected by acetamides. pKa calculations allowed us to suggest that at pH=7.4, at which the Kd has been measured, the GN8 is present in neutral ($\mathbf{1}^0$) and monoprotonated ($\mathbf{1}^+$) form; while at pH=4.5, at which the NMR experiments have been performed, exists in the diprotonated ($\mathbf{1}^{2+}$) form (Figure 7.2 and 7.3). Therefore, the three protomers were docked onto the hPrP^C NMR structure and its 20 conformers. Notice that the sequence identity between hPrP^C used in calculations and mPrP^C used in experiments is very high (98%). In particular, all of the residues involved in GN8 binding are fully conserved. In addition, the RMSD of the C_α between the 20 hPrP^C conformational structures is not very dissimilar from that between mPrP^C and hPrP^C NMR structures (2.4 ±0.1 Å nm and 2.7 Å, respectively). Therefore, we do not expect large changes of the results upon performing simulations either in the mouse or human protein.

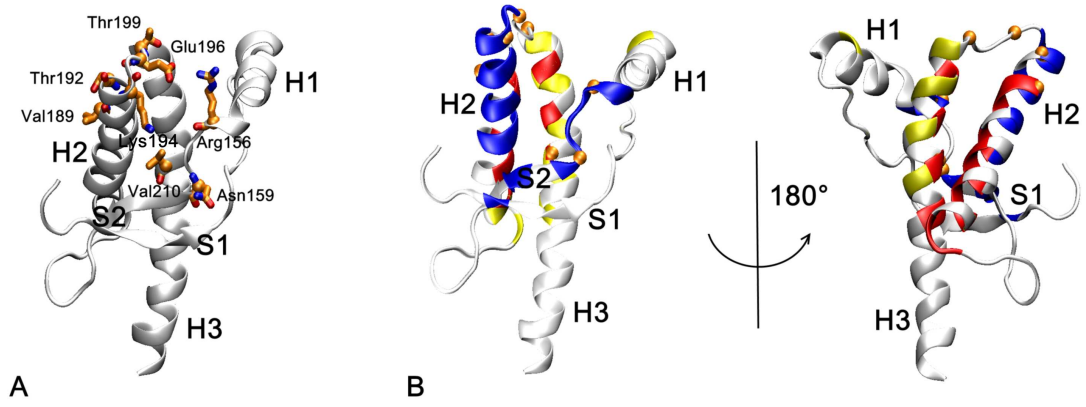

Figure 7.4: **A:** The NMR-defined binding site. Residues that exhibit significant chemical shift as reported in ref. [35] are shown in licorice. **B:** The three different binding sites **I** (shown in blue), **II** (depicted in red) and **III** (highlighted in yellow) obtained after docking of GN8 to diverse conformational structures of hPrP^C. Orange balls represent amino acids defined by NMR chemical shift study to interact with GN8 compound.

As a result of docking procedure, binding sites **I**, **II** and **III** were determined (Figure 7.4B). **I** is defined by H2 helix and the loop connecting β-sheet S2 and helix H1. **II** is defined by H2 and H3 helices. **III**, is represented by H3 helix, H2 helix N-term, and the loop between H1 helix and S1 β-sheet. **I** is the only site which involves residues changing chemical shifts upon GN8 binding and it was the only one selected for further calculations.

Adducts with the three protomers bound to this location underwent therefore 10-ns MD calculations in aqueous solution. After $\sim$ 3 ns of MD simulations, $\mathbf{1}^0$ partially slided out of the binding pocket **I**. In its final conformation, the GN8 amidic oxygen (O2, Figure 7.2) formed H-bond interactions with Lys194. The pyrrolidine ring formed hydrophobic interactions with methylene groups of

Glu196 and Arg156. The two aromatic rings formed also hydrophobic interactions with the methylene groups of Lys194. Water molecules can be involved in protein-ligand recognition by forming mediating hydrogen bonds between the protein and the ligand. In this regard, water-mediated H-bond contacts were identified between hPrP^C and protomers during the MD simulations. The amidic nitrogen of 10 protomer formed water mediated H-bond with carbonyl oxygen of Arg156, with hydroxyl group of Thr190 or imidazole nitrogen of His187.

Instead, $\mathbf{1}^+$ and $\mathbf{1}^{2+}$ remained stable during the MD simulations relative to initial docked structure. The amidic nitrogen atom of both protomers (N3; Figure 7.2) H-bonded to backbone oxygen of Glu196. The carbonyl oxygen (O5, Figure 7.2) of $\mathbf{1}^+$ H-bonded with the backbone amino group of Asn159. In addition, phenyl rings of $\mathbf{1}^+$ formed hydrophobic interactions with the methylene groups of Arg156 and T-stacking interactions with the His187. During the MD the protonated nitrogen atom (N1, Figure 7.2) of $\mathbf{1}^+$ formed direct or water mediated H-bond with the carboxylate side chains of Asp122. The protonated N6 atom of $\mathbf{1}^{2+}$ formed water mediated H-bond with the side chain oxygen of Asn159. Electrostatic interaction was present among carboxylate of Glu196 and N1 atom of $\mathbf{1}^{2+}$. Hydrophobic interactions between the diphenylmetane fragment of $\mathbf{1}^{2+}$ and methylene groups of Lys194 and Arg156 stabilized further the binding of $\mathbf{1}^{2+}$.

The structural determinants emerging from our MDP protocol turned out to be consistent with all of the ligand-protein contacts identified by NMR (Table 7.1) except for Val189, Thr192 and Thr199. A docking approach, combined with energy minimization provided similar results (Figure 3 in ref. [35]). Based on these results, we proceed using a more sophisticated method which attempts at identifying ligand poses in agreement with experiments.

This is **bias-exchange metadynamics** (BE-META) simulations, which allows to reconstruct the binding free energy profile associated to the simulated process of dissociation. The free energy was here calculated as a function of a large number of collective variables necessary for describing the complex binding process (see methods).

hPrP^C-$\mathbf{1}^0$ complex. $\mathbf{1}^0$ global minimum is located in the wide cleft formed by helices H1, H2 and H3, similarly to I and the minimum identified by Kuwata et al. for mPrP^C . The contacts $\mathbf{1}^0$ formed with the hPrP^C are consistent with the reported chemical shift changes on Arg156, Thr199, Glu196 upon GN8 binding . The phenyl groups of $\mathbf{1}^0$ form a cation-pi interaction with Arg156 and a water-mediated hydrogen bond is present between Thr199 and the pyrrolidine nitrogen (N1; Figure 7.2). Other water mediated interactions observed for $\mathbf{1}^0$ in previous MD simulations are not observed here. In contrast to MD, Thr190 forms a direct hydrogen bond with the carbonyl group of $\mathbf{1}^0$ compound (O2; Figure 7.2). This is an indication that MD simulations were hindered in a pre-stationary state in which the release of intermediate water molecules is necessary to reach the global minimum. This now arises when combining MD simulations with metadynamics. The hPrP^C-$\mathbf{1}^0$ complex is further stabilized by hydrophobic interactions between the pyrrolidine ring and Thr183 and Pro158.

The unbound state of hPrP^C-$\mathbf{1}^0$ system ($\mathbf{1}^0$.U1 in Figure 7.6A) corresponds to a conformation in which the ligand is not forming any contact with the protein. Lys194 showed a remarkable conformational change of the backbone upon ligand binding (Figure 7.5). This is consistent with the significant chemical shift change reported for this amino acid . Smaller conformational changes were observed for other residues present in the H2-H3 loop. These rearrangements were not observed with previously used MDP techniques, because they are induced during the ligand binding process simulated here.

This unbound state is 5.5 kcal·mol^{-1} $\pm$ 0.9 kcal·mol^{-1} higher in energy with respect to the global minimum. The ligand is not completely detached from the protein (although already 5 layers of water molecules between the two moieties are present). The remaining free energy for complete dissociation

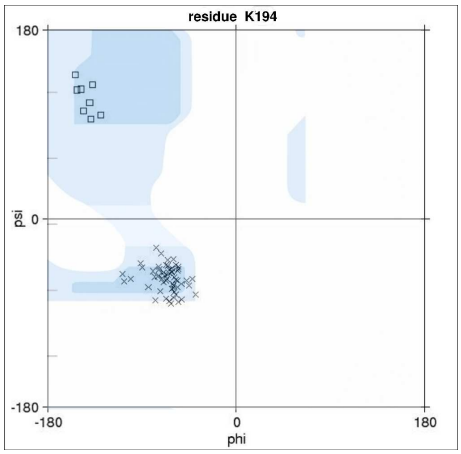

Figure 7.5: Ramachandran plot of residue Lys194 for the unbound state (empty squares), and bound states (crosses) of the hPrP^C-$\mathbf{1}^0$ adduct.

was estimated to be -0.7 kcal·mol^{-1} by means of the Poisson-Boltzmann equation (see methods). Thus, we estimate the dissociation free energy to be 4.7 kcal·mol^{-1} in our simulation conditions. Considering also the concentration of the species in the simulation box, the standard free energy of dissociation is 7.8 kcal·mol^{-1}. This is in very good agreement with the experimental value of 7.5 kcal·mol^{-1} (corresponding to Kd=3.9 μM) reported by Kuwata et al. [35].

hPrP^C-$\mathbf{1}^+$ complex. In the global minimum, $\mathbf{1}^+$ lies along the loop that connects helices H2 and H3 ($\mathbf{1}^+$.B1 in Figure 7.6A). It forms a remarked hydrophobic interaction with Thr199, consistently with the chemical shift changes reported for this group . The amidic nitrogen atoms of $\mathbf{1}^+$ (N3 and N4; Figure 7.2) are H-bonded to Thr201 and Asn197 respectively. This induces a subtle conformational change of the Glu196 and Asn197 backbone upon ligand binding which may be the reason for the chemical shift displacement reported experimentally for Glu196. Additionally, the neutral pyrrolidine ring of $\mathbf{1}^+$ is kept by the hydrophobic cleft formed by Val203, Met206, Ile184, Thr188 and Phe198, further stabilizing the complex. No water-mediated interactions were observed between $\mathbf{1}^+$ and hPrP^C.

The free energy difference between the bound state of $\mathbf{1}^+$ ($\mathbf{1}^+$.B1 in Figure 7.6A) and the corresponding unbound state is 5.3 kcal·mol^{-1}. The remaining free energy for complete dissociation was estimated to be 0.2 kcal·mol^{-1} (see methods). Thus, we estimate the dissociation free energy to be 5.5 kcal·mol$^{-1}\pm$ 0.9 kcal·mol^{-1}. Correcting by the concentration of the species in the simulation box, the standard free energy of dissociation turns out to be 8.6 kcal·mol^{-1}, that is a similar value than that predicted value for $\mathbf{1}^0$ and in good agreement with experiments.

hPrP^C-$\mathbf{1}^{2+}$ complex. The most stable conformation of $\mathbf{1}^{2+}$ binds yet in another position of hPrP^C, laying along helix H2 ($\mathbf{1}^{2+}$.B1 in Figure 7.6A). Half part of $\mathbf{1}^{2+}$ is in strict contact with hPrP^C surface in the cleft formed by Thr192, Thr193 and Val189 by means of hydrophobic interactions. Indeed, these positions were reported to interact directly with the ligand according to NMR experiments . However, the second part of $\mathbf{1}^{2+}$ is quite distanced from binding surface due to the presence of Lys185. Two water molecule layers in between the PrP/ligand interface are responsible for the fluctuation of the ligand in the binding pocket. A value of free energy here was not calculated, as this protomer is present in the conditions of the NMR experiment , but not in the conditions in which the affinity has been measured.

In summary our protocol enabled us to identify the alternative binding poses of the GN8 compound to hPrP^C and to provide an accurate estimation of the experimental dissociation constant. The multiple binding sites pattern that we observed from our simulations (Figure 7.6B) allows for justifying the NMR-contacts observed in distant positions of hPrP^C which are located in opposite sites of the protein surface (Arg156, Thr199, Glu196, Thr192, Val189) . This pattern could not be observed applying only standard MD, and was not observed in the study reported by Kuwata et al. [35] where a single binding mode was proposed.

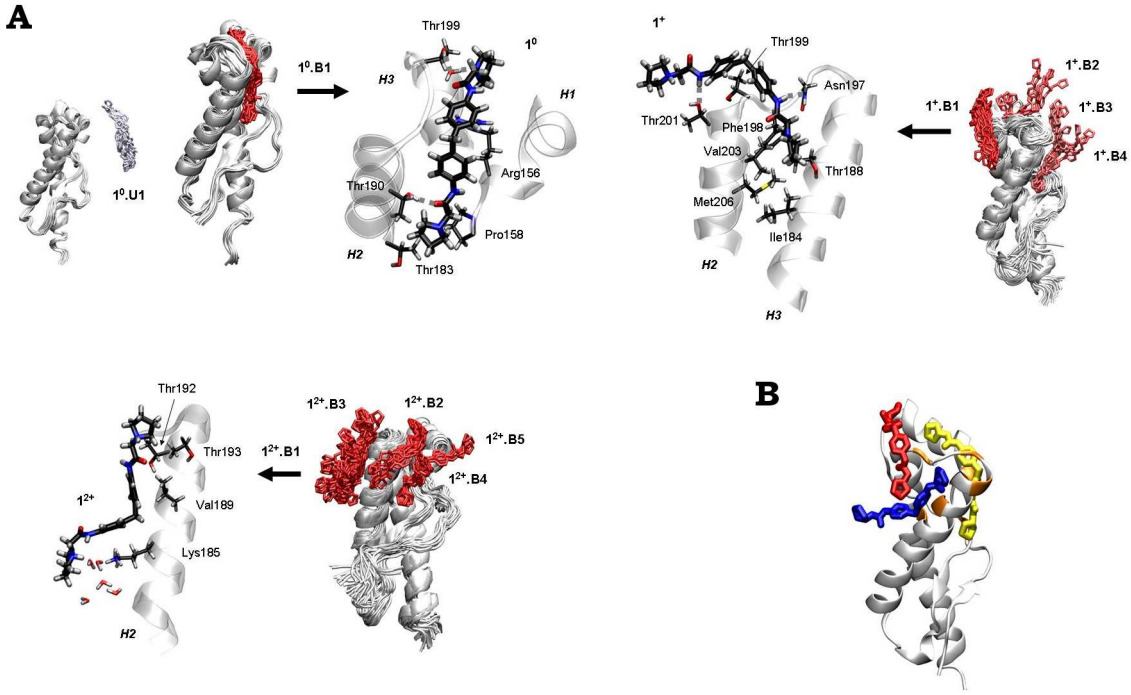

Figure 7.6: **A:** Representative three dimensional structures of the hPrP^C complexes with the $\mathbf{1}^0$, $\mathbf{1}^+$ and $\mathbf{1}^{2+}$ protomers. Ligand structures are colored according to their relative free energy values (B-bound state; U-unbound state). **B:** Comparison of the GN8 binding modes for each of the three protonation states on hPrP^C surface. Yellow: neutral state ($\mathbf{1}^0$); Red: protonated state ($\mathbf{1}^+$); Blue: diprotonated state ($\mathbf{1}^{2+}$). The amino acids showing a change in NMR chemical shift upon ligand binding [35] are highlighted in orange.

7.4 Conclusions

Predicting the binding mode of small ligands to the protein surface is a challenging task from the computational point of view. We have presented here a hybrid protocol that overcomes these limitations. Our protocol combines molecular docking with rare events simulation techniques (molecular dynamics coupled to bias-exchange metadynamics). The integration of docking algorithms with molecular dynamics based simulations has already shown to be convenient in computer assisted drug design [221, 224]. We showed here that this can be extended with free energy calculations in order to get an accurate prediction of the binding modes and the thermodynamic stability of the complexes. Similar approaches have already been followed [225, 228, 227].

This protocol has been validated with the binding of a small compound (GN8) [35] to the cellular

form of prion protein ($hPrP^C$), the main agent involved in prion diseases. GN8 compound can adopt different protonation states at physiological and acid pH, at which experiments have been carried out. We identified different binding patterns for each of them. Neutral state binds in a specific region of $hPrP^C$ surface along the shallow cleft formed by helices H1, H2 and H3, while the protonated state bind in multiple sites of $hPrP^C$ surface. The characteristic binding of this molecule is in agreement with the distant contacts with the prion protein reported experimentally by NMR studies [35]. The calculated free energy agrees with the experimental value.

Given the lack of deep binding pockets along $hPrP^C$ protein structure, it is reasonable that a small molecule will not bind specifically to a single site. NMR data shows that this is indeed the case [35]. Consistently, we have observed a multi-binding mode for the GN8 ligand, not emerging from standard MDP protocols. Our simulations predict an unspecific binding of the different protonation states of GN8 compound on multiple sites of $hPrP^C$ surface. The different binding modes allow justifying the experimentally predicted contacts of GN8 compound with distant regions of the protein [35]. These alternative binding modes could be predicted only when simulating the whole binding process of GN8 compound from solution to the protein surface for each protonation state independently. In conclusion, molecular simulations approaches appear as a valuable tool to predict poses and energetic of ligands binding to protein surfaces.

Concluding remarks

We have addressed issues related to the structural predictions and docking of systems relevant for molecular pharmacology, especially for proteins of great importance in neurobiology.

Ca^{2+}-gated potassium (hBK_{Ca}) and chloride (bestrophins) channels play a key role for the signaling in neuronal, muscular and epithelial tissues [9, 10, 11, 12, 118, 119, 117]. Failure of their physiological function leads to degenerative diseases [14, 15, 7]. Understanding the structural determinants is crucial to address the problems related to the function. To gain structural information of these channels, we have here applied molecular modeling techniques.

The structural determinants of two different hBK_{Ca} Ca^{2+}-binding domains were predicted. These are the RCK1 domain (Regulator of Conductance of K^+) and the Ca^{2+} bowl. A structural model of the RCK1 domain was constructed based on two templates: the RCK domains of *E. coli* [97] and MthK K^+ [91] channels. Our model points to a key role of the salt bridge among K513 and D546. This is consistent with experiments of site directed mutagenesis, which showed the importance of the predicted salt bridge in stabilizing the domain. Next, the Ca^{2+} bowl including five Ca^{2+} ions was constructed and its stability was studied by MD simulations. Two Ca^{2+} ions remained stably bound. One binding site was formed by residues D959, D961, D963 and E967. Mutations to Ala of Asp residues forming this binding site were shown experimentally to reduce Ca^{2+} binding and/or sensitivity from 50% to 90% with respect to the wild type channel [21, 22], consistently with our model. The second ion was bound to E946, D950, D957 and D962. All these residues except Asp950 when mutated to Ala reduced Ca^{2+} binding and/or sensitivity from 30% to 90% with respect to the wild type BK_{Ca} [21, 22]. Again, these results are consistent with our predictions.

Bestrophin is suggested to be a Ca^{2+}-gated Cl^- channel [6]. It contains a Ca^{2+}-binding motif very similar to that of hBK_{Ca} Ca^{2+} bowl [20, 21]. Structural models of the Asp-rich domain and its mutants were constructed including different numbers of Ca^{2+} ions. By performing extensive molecular simulations, two well formed Ca^{2+}-binding sites were identified. Residues involved in Ca^{2+}-binding, D301, D304, E306 and D312, were proposed for experimental mutational studies. Electrophysiological experiments indeed confirmed that mutations D301A and D304A decreased the Ca^{2+}-activated current amplitude, as predicted by our modeling. On the other hand, mutations E306A and D312A did not cause reduction in the current consistently with the simulations: in fact, neutral residues such as Asn, Gln and Ser replace the aforesaid residues in the Ca^{2+} coordination shell. These results support the importance of the Asp-rich domain in Ca^{2+}-dependent activation of the bestrophins and shed light on its structural features.

The second part of the thesis focuses on docking of mini-antibodies and ligands to a protein involved in neurodegeneration, the prion protein. Many key proteins involved in neurodegenerative diseases (such as α-synuclein in Parkinson disease, β-amyloid in Alzheimer disease or prion protein in prion disease) exhibit the ability to aggregate, what consequently causes the damage of neurons [254].

Therefore, due to these similar features, finding a treatement for one of these diseases may help to address also the others.

We studied the interactions between the small antibody fragment D18scFv and the prion protein in collaboration with experimental groups. Our experiments have shown that the small antibody fragment D18scFv binds to the prion protein and prevents its conversion to the pathological form. To get insight into the mechanism by which D18scFv prevents the conversion and propagation of prion proteins, a structural model of D18scFv was constructed and the prion protein was docked onto it. Based on our results, we suggested that Arg151 is the antigenic determinant of prion protein, i.e. it is the key residue for the recognition between the two proteins. In addition, we proposed that interactions of $\mathrm{Arg}_{PrP}136$ and $\mathrm{His}_{PrP}140$ with D18scFv stabilize PrP^{C} and therefore prevent the fibrillation process.

An alternative to immunotherapy is the design of small ligands binding to the prion protein. Small compounds may stabilize the cellular prion protein conformation or modify the binding of molecular chaperon, which enhances prion protein conformational conversion [26, 190, 191]. Unfortunately, prion proteins lack the cavities (unlike enzymes and receptors) which allow for efficient docking and high throughput screening, first steps in design of new ligands. Here we have addressed this issue by proposing a new computational protocol to address the problem of ligands binding to a protein surface. To this aim, standard docking methods were combined with MD simulations and free energy calculations. The protocol was tested on the known anti-prion compound (GN8), and compared with experimental data [35]. It turns out that the combination of the three methods of the protocol is necessary to satisfy contacts experimentally detected by NMR spectroscopy [35]. In addition, this approach allows to predict the binding free energy, which for GN8 [35] is in very good agreement with the experimental value.

List of publications

Papers related to the thesis:

1. **A. Kranjc**, F. W. Grillo, J. Rievaj, A. Boccaccio, F. Pietrucci, P. Carloni, A. Menini, C. Anselmi; Regulation of bestrophins by Ca^{2+}: a theoretical and experimental study. *PLoS ONE*, 2009, in press.

2. V. Campana, L. Zentilin, I. Mirabile, **A. Kranjc**, P. Casanova, M. Giacca, S.B. Prusiner, G. Legname and C. Zurzolo; Antibody fragments for immunotherapy of prion diseases. *Biochem. J.* 418 (2009) 507-515.

3. **A. Kranjc**, C. Anselmi, P. Carloni and F. E. Blaney; Structural Models of Human Big Conductance Calcium- and Voltage-gated Potassium Channels. *Comp. Phys. Comm.* 177 (2007) 21-26.

4. **A. Kranjc**, S. Bongarzone, G. Rossetti, X. Biarnes, P. Carloni; Docking Ligands on cavity-less proteins. The case study of prion protein. *to be submitted* within February, 2009.

Other papers:

1. K. Galesa, U. Bren, **A. Kranjc** and J. Mavri; Carcinogenicity of Acrylamide: A Computational Study. *J. Agr. Food Chem.*, 56 (2008) 8720-8727.

2. **A. Kranjc**, J. Mavri; Guanine Alkylation by Ethylene Oxide: Calculation of Chemical Reactivity. *J. Phys. Chem. A,* 110 (2006) 5740-5744.

Acknowledgments

I would like to thank my supervisor Prof. Paolo Carloni for giving me the opportunity to be a part of his group and for leading me through my PhD study. He taught me not only how to handle the research, but he also helped me to do a step further in my personal growth.

I would like to thank Dr. Claudio Anselmi for all fruitful collaborations and for helping me to solve many technical problems and research doubts.

I would like to thank Alejandro Giorgetti for teaching me the first steps in homology modeling. I am grateful to my first external collaborators Iain M. Mclay, Frank E. Blaney, Anna-Maria M. Capelli and Aldo G. Feriani from GSK (Stevenage, UK and Verona Italy) for scientific discussions, for funding my first PhD project and for their kindness. A special thank goes to Prof. Anna Menini, Juraj Rievaj, Federico W. Grillo and Anna Boccaccio who carried out experiments that supported our theoretical work, for many hours of discussions and for all their friendship. Thank you to all collaborators in the project of prion proteins, especially to Prof. Maria Laura Bolognesi, Prof. Andrea Cavalli and Prof. Marinella Roberti (University of Bologna) for giving me the opportunity to learn docking at their department, for all kindness and fruitful discussions. Thank you to Prof. Giuseppe Legname (SISSA/ISAS) and to the group of prion boys and girls, Salvatore Bongarzone, Xevi Biarnes and Giulia Rossetti with whom we shared difficulties in solving problems related to prion protein. I would like to thank Dr. Janez Mavri, who was the first person encouraging me to start living my dreams of doing a research and for giving me the opportunity to collaborate with him also in the time of my PhD.

Thank you to the SBP group. We were like a family where everyone was always open for helping the others. We were devoted to the science but knew also how to have fun.

Thank you to my husband Fabio, to my parents and to my big family for believing in me, for all their love and support that helped me to finish my PhD.

Appendix

Additional computational details for the PrP docking protocol

In metadynamics a history dependent potential that guides the evolution of the system is constructed as a sum of Gaussian functions centered on the value of each collective variable along the simulation. Here, a new Gaussian term was added to the historic dependent potential every 1 ps. The metadynamics parameters of this bias potential were set up as in our previous work [228, 227, 255] (Table 7.2).

Here we have combined metadynamics with its bias exchange variant [34]. Six independent metadynamics simulations were run in parallel. Each replica was biased by different one-dimensional time-dependent potentials, which were built as a function of each of the collective variables defined above. Exchanges among replicas were attempted every 10 ps using a Metropolis acceptance criterion [34]. Similar set-ups have shown to improve the sampling of the configurational space and the convergence of the results [34, 251]. The different simulation replicas of the system were run for 30 ns each (total sampling time of 180 ns). Along these simulations the ligand gets reversibly bound and unbound to different parts of the protein surface. As in standard metadynamics, the replicas were allowed to run until the different bias potentials were converged. To ensure the convergence of the bias potentials, two average profiles were computed: one between times 15 ns and 22.5 ns and the other between times 22.5 ns and 30 ns. Only portions of the two average profiles that agreed with each other within 1.5 k_BT were used for further analysis. (Figure 7.7). An analysis of the converged bias potentials allows quantifying the relative free energy among all different bound states detected during the simulation.

Laio and collaborators have recently introduced a methodology to reconstruct, from a bias exchange metadynamics simulation, the multidimensional free energy landscape from the corresponding lower dimensional projections [*F. Mareinelli, F. Pietrucci, A. Laio and S. Piana, submitted to PLOS Comp Biol, 2008*]. This methodology is based on the weighted histogram analysis method (WHAM) introduced by Kumar et al. [252]. Briefly, the method is based on 1) clustering the phase space

collective variable	width of the Gaussian	height of the gaussian
distance	0.02 - 0.04 nm	0.05 - 0.14 kcal·mol^{-1}
polar contacts A	0.1 - 0.2	0.05 - 0.14 kcal·mol^{-1}
polar contacts B	0.2 - 0.3	0.05 - 0.14 kcal·mol^{-1}
water bridges	0.8 - 1.0	0.05 - 0.14 kcal·mol^{-1}
RMSD system	0.01 - 0.02 nm	0.05 - 0.14 kcal·mol^{-1}
RMSD binding site	0.01 - 0.02 nm	0.05 - 0.14 kcal·mol^{-1}

Table 7.2: Parameters of the biasing potential in the metadynamics calculations.

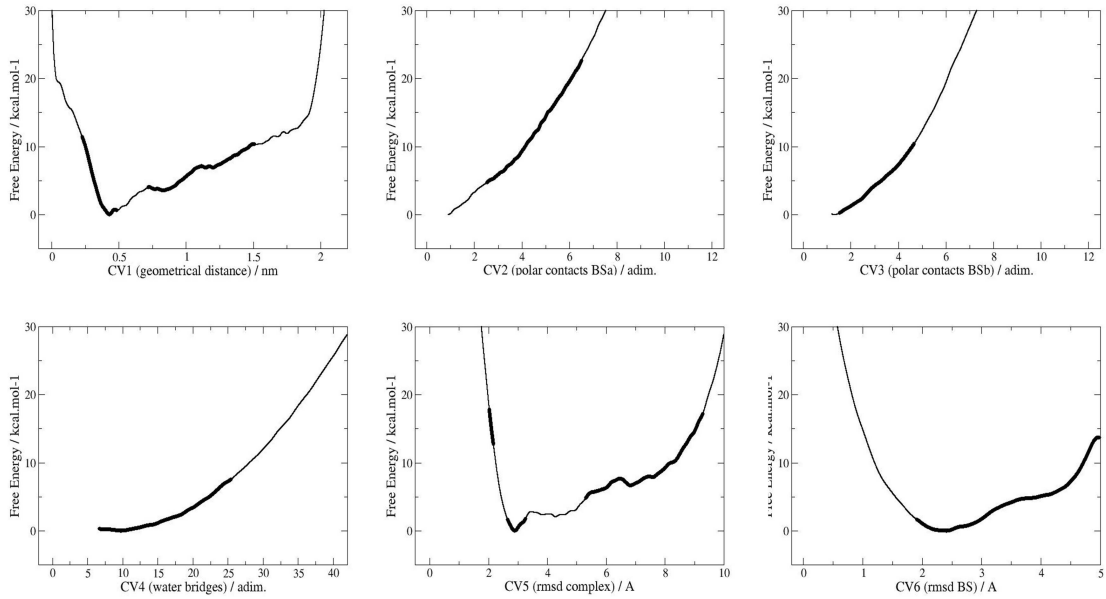

Figure 7.7: Free energy profiles at the end of the simulation of $\mathbf{1}^0$ protomer as a function of each collective variable.

explored during the simulations as a function of the collective variables, 2) evaluating the population of each cluster along each biased trajectory, 3) assigning the unbiased free energy to each cluster by means of WHAM, using the information from the converged free energy profiles. Between 15 ns and 30 ns, the trajectories of the different replicas were evolving under the action of converged bias potentials (Figure 7.7). Thus, only this part of the simulation data was used for extracting statistical information.

Bibliography

[1] F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer Jr, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 112:535–542, 1977.

[2] Berman, H. M. and Westbrook, J. and Feng, Z. and Gilliland, G. and Bhat, T. N. and Weissig, H. and Shindyalov, I. N. and Bourne, P. E. The protein data bank. *Nucleic Acids Res*, 28:235–242, 2000.

[3] A. Bairoch and R. Apweiler. The swiss-prot protein sequence database and its supplement trembl in 2000. *Nucleic Acids Res.*, 28:45–48, 2000.

[4] A. Sali and T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *J.Mol.Biol.*, 234:779 – 815, 1993.

[5] A. Marty. Ca-dependent K channels with large unitary conductance in chromaffin cell membranes. *Nature*, 291:497–500, 1981.

[6] H. Sun, T. Tsunenari, K.W. Yau, and J. Nathans. The vitelliform macular dystrophy protein defines a new family of chloride channels. *Proc.Natl.Acad.Sci.U.S.A*, 99:4008 – 4013, 2002.

[7] Wickenden AD. K+ channels as therapeutic drug targets. *Pharmacol. Ther.*, 94:157–182, 2002.

[8] V Calderone. Large-conductance, Ca2+-activated K+ channels: Function, pharmacology and drugs. *Curr. Med. Chem.*, 9:1385–1395, 2002.

[9] O. H. Petersen and Y. Maruyama. Calcium-activated potassium channels and their role in secretion. *Nature*, 307:693–696, 1984.

[10] R. Robitaille and M. Charlton. Presynaptic calcium signals and transmitter release are modulated by calcium-activated potassium channels. *J Neurosci*, 12:297–305, 1992.

[11] G. Perez and L. Toro. Differential modulation of large-conductance KCa channels by PKA in pregnant and nonpregnant myometrium. *Am. J. Physiol. Cell Physiol.*, 266:C1459–1463, 1994.

[12] H. Kume, K. Mikawa, K. Takagi, and M. I. Kotlikoff. Role of G proteins and KCa channels in the muscarinic and beta-adrenergic regulation of airway smooth muscle. *Am. J. Physiol. Lung Cell Mol. Physiol.*, 268:L221–229, 1995.

[13] B. Hille. *Ionic channels of excitable membranes. IInd Ed.* Sinauer Associates Inc: Sunderland, 1992.

[14] A. Marquardt, H. Stohr, L.A. Passmore, F. Kramer, A. Rivera, and B.H. Weber. Mutations in a novel gene, vmd2, encoding a protein of unknown properties cause juvenile-onset vitelliform macular dystrophy (best's disease). *Hum.Mol.Genet.*, 7:1517 – 1525, 1998.

[15] K. Petrukhin, M.J. Koisti, B. Bakall, W. Li, G. Xie, T. Marknell, O. Sandgren, K. Forsman, G. Holmgren, S. Andreasson, M. Vujic, A.A. Bergen, V. McGarty-Dugan, D. Figueroa, C.P. Austin, M.L. Metzker, C.T. Caskey, and C. Wadelius. Identification of the gene responsible for best macular dystrophy. *Nat.Genet.*, 19:241 – 247, 1998.

[16] H.C. Hartzell, Z. Qu, K. Yu, Q. Xiao, and L.T. Chien. Molecular physiology of bestrophins: multifunctional membrane proteins linked to best disease and other retinopathies. *Physiol Rev.*, 88:639 – 672, 2008.

[17] D. Marchant, K. Yu, K. Bigot, O. Roche, A. Germain, D. Bonneau, V. Drouin-Garraud, D.F. Schorderet, F. Munier, D. Schmidt, Neindre P. Le, C. Marsac, M. Menasche, J.L. Dufier, R. Fischmeister, C. Hartzell, and M. Abitbol. New vmd2 gene mutations identified in patients affected by best vitelliform macular dystrophy. *J.Med.Genet.*, 44:e70 – 0, 2007.

[18] K. Yu, Z. Qu, Y. Cui, and H.C. Hartzell. Chloride channel activity of bestrophin mutants associated with mild or late-onset macular degeneration. *Invest Ophthalmol.Vis.Sci.*, 48:4694 – 4705, 2007.

[19] K. Yu, Q. Xiao, G. Cui, A. Lee, and H.C. Hartzell. The best disease-linked cl- channel hbest1 regulates cav1 (l-type) ca2+ channels via src-homology-binding domains. *J.Neurosci.*, 28:5660 – 5670, 2008.

[20] M. Schreiber and L. Salkoff. A novel calcium-sensing domain in the bk channel. *Biophys.J.*, 73:1355 – 1363, 1997.

[21] L. Bao, C. Kaldany, E.C. Holmstrand, and D.H. Cox. Mapping the bkca channel's "ca2+ bowl": side-chains essential for ca2+ sensing. *J.Gen.Physiol.*, 123:475 – 489, 2004.

[22] J.Z. Sheng, A. Weljie, L. Sy, S. Ling, H.J. Vogel, and A.P. Braun. Homology modeling identifies c-terminal residues that contribute to the ca2+ sensitivity of a bkca channel. *Biophys.J.*, 89:3079 – 3092, 2005.

[23] T. Tsunenari, J. Nathans, and K.W. Yau. Ca2+-activated cl- current from human bestrophin-4 in excised membrane patches. *J.Gen.Physiol.*, 127:749 – 754, 2006.

[24] S.B. Prusiner. Novel proteinaceous infectious particles cause scrapie. *Science*, 216:136 – 144, 1982.

[25] F. E. Cohen and K. M. Pan and Z. Huang and M. Baldwin and R. J. Fletterick and S. B. Prusiner. Structural clues to prion replication. *Science*, 264:530–531, 1994.

[26] C. R. Trevitt and J. Collinge. A systematic review of prion therapeutics in experimental models. *Brain*, 129:2241 – 2265, 2006.

[27] Williamson, R. A., Peretz, D., Pinilla, C., Ball, H., Bastidas, R. B., Rozenshteyn, R., Houghten, R. A., Prusiner, S. B. and Burton, D. R. Mapping the prion protein using recombinant antibodies. *J Virol*, 72:9413–9418, 1998.

[28] Peretz, D., Williamson, R. A., Kaneko, K., Vergara, J., Leclerc, E., Schmitt-Ulms, G., Mehlhorn, I. R., Legname, G., Wormald, M. R., Rudd, P. M., Dwek, R. A., Burton, D. R. and Prusiner, S. B. Antibodies inhibit prion propagation and clear cell cultures of prion infectivity. *Nature*, 412:739–743, 2001.

[29] Cardinale, A., Filesi, I., Vetrugno, V., Pocchiari, M., Sy, M. S. and Biocca, S. Trapping prion protein in the endoplasmic reticulum impairs prpc maturation and prevents prpsc accumulation. *J Biol Chem*, 280:685–694, 2005.

[30] Vetrugno, V., Cardinale, A., Filesi, I., Mattei, S., Sy, M. S., Pocchiari, M. and Biocca, S. Kdel-tagged anti-prion intrabodies impair prp lysosomal degradation and inhibit scrapie infectivity. *Biochem Biophys Res Commun*, 338:1791–1797, 2005.

[31] Zuber, C., Knackmuss, S., Rey, C., Reusch, U., Rottgen, P., Frohlich, T., Arnold, G. J., Pace, C., Mitteregger, G., Kretzschmar, H. A., Little, M. and Weiss, S. Single chain fv antibodies directed against the 37 kda/67 kda laminin receptor as therapeutic tools in prion diseases. *Mol Immunol*, 45:144–151, 2008.

[32] Zuber, C., Mitteregger, G., Schuhmann, N., Rey, C., Knackmuss, S., Rupprecht, W., Reusch, U., Pace, C., Little, M., Kretzschmar, H. A., Hallek, M., Buning, H. and Weiss, S. Delivery of single-chain antibodies (scfvs) directed against the 37/67 kda laminin receptor into mice via recombinant adeno-associated viral vectors for prion disease gene therapy. *J Gen Virol*, 89:2055–2061, 2008.

[33] A. Laio and M. Parrinello. Escaping free-energy minima. *Proc.Natl.Acad.Sci.U.S.A*, 99:12562 – 12566, 2002.

[34] Piana S and Laio A. A bias-exchange approach to protein folding. *J. Phys. Chem. B*, 111:4553–4559, 2007.

[35] Kazuo Kuwata, Noriyuki Nishida, Tomoharu Matsumoto, Yuji O. Kamatari, Junji Hosokawa-Muto, Kota Kodama, Hironori K. Nakamura, Kiminori Kimura, Makoto Kawasaki, Yuka Takakura, Susumu Shirabe, Jiro Takata, Yasufumi Kataoka, and Shigeru Katamine. Hot spots in prion protein for pathogenic conversion. *Proc Natl Acad Sci U S A*, 104:11921 – 11926, 2007.

[36] M. A. Marti-Renom, A. C. Stuart, A. Fiser, R. Sanchez, F. Melo, and A. Sali. Comparative protein structure modeling of genes and genomes. *Annu. Rev.Biophys. Biomol. Struct.*, 29:291–325, 2000.

[37] C. J. Epstein, R. F. Goldberger, and C. B. Anfinsen. The genetic control of tertiary protein structure. model systems. *Cold Spring Harb. Symp. Quant. Biol.*, 28:439–49, 1963.

[38] C. Chothia and A. M. Lesk. The relation between the divergence of sequence and structure in proteins. *EMBO J.*, 5:823–826, 1986.

[39] C. Sander and R. Schneider. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins: Struct., Funct., Genet.*, 9:56–68, 1991.

[40] SF Altschul, W Gish, W Miller, EW Myers, and DJ Lipman. Basic local alignment search tool. *J. Mol. Biol.*, 215(3):403–410, 1990.

[41] L. Holm and C. Sander. Protein folds and families: sequence and structure alignments. *Nucleic Acid Res.*, 27:244–47, 1999.

[42] Pearson W.R. Rapid and sensitive sequence comparison with fastp and fasta. *Methods in Enzymology*, 183:63–98, 1990.

[43] Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Res.*, 25:3389–3402, 1997.

[44] M. O. Dayhoff and R. M. Schwartz and B. C. Orcutt. A model of evolutionary change in proteins. *In "Atlas of Protein Sequence and Structure"*, 5(3):345–352, 1978.

[45] S. Henikoff and J. G. Henikoff. Amino-acid substitution matrices from protein blocks. *Proc. Nat. Acad. Sci. U.S.A.*, 89:10915–10919, 1992.

[46] D. G. Higgins and P. M. Sharp. Clustal: a package for performing multiple sequence alignment on a microcomputer. *Gene*, 73:237–244, 1988.

[47] J. D. Thompson, D. G. Higgins, and T. J. Gibson. Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 22:4673–4680, 1994.

[48] A Sali and TL Blundell. Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.*, 234(3):779–815, 1993.

[49] N. Guex and M. C. Peitsch. Swiss-model and the swiss-pdbviewer: An environment for comparative protein modeling. *Electrophoresis*, 18:2714–2723, 1997.

[50] M. J. Bower, F. E. Cohen, R. L. Jr. Dunbrack. Prediction of protein side-chain rotamers from a backbone dependent rotamer library: a new homology modeling tool. *J. Mol. Biol.*, 267:1268–82, 1997.

[51] P. A. Bates, L. A. Kelley, R. M. MacCallum, and M. J. Sternberg. Enhancement of protein modeling by human intervention in applying the automatic programs 3d-jigsaw and 3d-pssm. *Proteins*, 45 (Suppl. 5):39–46, 2001.

[52] T. Schwede and J. Kopp and N. Guex and M. C. Peitsch. Swiss-model: An automated protein homolgy-modeling server. *Nucleic Acids Res.*, 31:3381–85, 2004.

[53] A. D. MacKerell Jr.; D. Bashford; M. Bellott; R. L. Dunbrack Jr.; J. D. Evanseck; M. J. Field; S. Fischer; J. Gao; H. Guo; S. Ha; D. Joseph-McCarthy; L. Kuchnir; K. Kuczera; F. T. K. Lau; C. Mattos; S. Michnick; T. Ngo; D. T. Nguyen; B. Prodhom; W. E. Reiher III; B. Roux; M. Schlenkrich; J. C. Smith; R. Stote; J. Straub; M. Watanabe; J. Wiorkiewicz-Kuczera; D. Yin; M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102:3586–3616, 1998.

[54] A. L. Morris and M. W. MacArthur and E. G. Hutchinson and J. M. Thornton. Stereochemical quality of protein structure coordinates. *Proteins*, 12:345–364, 1992.

[55] R. A. Laskowski and M. W. MacArthur and D. S. Moss and J. M. Thornton. Procheck: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, 26:283–291, 1993.

[56] G. Jones, P. Willett, and R. C. Glen. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol*, 245:43–53, 1995.

[57] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, 267:727–748, 1997.

[58] J. W. M. Nissink, C. Murray, M. Hartshorn, M. L. Verdonk, J. C. Cole, and R. Taylor. A new test set for validating predictions of protein-ligand interactions. *Proteins: Struct., Funct., Genet.*, 49:457–471, 2002.

[59] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Heart, R. K. Belew, A. J. Olson. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J. Comp. Chem.*, 19:1639–1662, 1998.

[60] S. J. de Vries, A. D. J. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar, and A. M. J. J. Bonvin. Haddock versus haddock: new feauters and performance of haddock2.0 on the capri targets. *Proteins*, 69:726–733, 2007.

[61] C. Dominguez and R. Boelens and A. M. J. J. Bonvin . Haddock: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.*, 125:1731–1737, 2003.

[62] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor. Improved protein-ligand docking using gold. *Proteins: Struct., Funct., Genet.*, 52:609–623, 2003.

[63] I. Halperin and B. Ma and H. Wolfson and R. Nussinov. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*, 47:409–443, 2002.

[64] D. E. Goldberg. *Genetic algorithms in search optimization and machine learning.* Addison-Wesley, Reading, MA, 1989.

[65] J. H. Holland. *Adaptation in natural and artificial systems.* University of Michigan Press, Ann Arbor, MI, 1975.

[66] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein. Comparison of simple potential functions for simulating liquid water. *J.Chem.Phys.*, 79:926 – 935, 1983.

[67] W. L. Jorgensen and J. Tirado-Rives. The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem Soc.*, 110:1657–1666, 1988.

[68] D. Frenkel and B. Smit. *Understanding Molecular Simulation.* Academic Press, 2001.

[69] H. J. C. Berendsen, D. Van der Spoel, and R. Vandrunen. GROMACS - a message-passing parallel molecular-dynamics implementation. *Comput. Phys. Commun.*, 91(1-3):43–56, 1995.

[70] B. Hess, H. Bekker, H. J. C. Berendsen, and G. E. M. J. Fraaije. Lincs: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18:1463–1472, 1997.

[71] L. Verlet. Computer experiments on classical fluids .I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159:98–&, 1967.

[72] R. W. Hockney, S. P. Goel, and J. W. Eastwood. Quiet high-resolution computer models of a plasma. *J. Comput. Phys.*, 14:148–158, 1974.

[73] J. W. Ponder and D. A. Case. Force fields for protein simulations. *Adv. Protein Chem.*, 66:27–85, 2003.

[74] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.

[75] Bayly C.I., Cieplak P., Cornell W.D., and Kollman P.A. Well-behaved electrostatic potential based method using charge restraints for determining atom-centered charges: The model. *J.Phys.Chem.*, 97:10269, 1993.

[76] T. A. Darden and D. York. Particle mesh ewald - an n.log(n) method for ewald sums in large systems. *J. Chem. Phys.*, 98:10089, 1993.

[77] P.P. Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik*, 369:253–287, 1921.

[78] U. Essman, L. Perera, M. L. Berkowitz, T. A. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103:8577, 1995.

[79] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.

[80] W.G. Hoover. Canonical dynamics - equilibrium phase-space distributions. *J Pathol*, 31:1695 – 1697, 1985.

[81] Nose S. molecular dynamics method for simulation in the canonical ensemble. *Mol.Physics*, 52:255–268, 1984.

[82] S Ghatta, D Nimmagadda, XP Xu, and ST O'Rourke. Large-conductance, calcium-activated potassium channels: Structural and functional implications. *Pharmacol. Ther.*, 110:103–116, 2006.

[83] XM Xia, XH Zeng, and CJ Lingle. Multiple regulatory sites in large-conductance calcium-activated potassium channels. *Nature*, 418:880–884, 2002.

[84] BS Pallotta. N-Bbromoacetamide removes a calcium-dependent component of channel opening from calcium-activated potassium channels in rat skeletal-muscle. *J. Gen. Physiol.*, 86(5):601–611, 1985.

[85] M Garcia-Calvo, HG Knaus, OB McManus, KM Giangiacomo, GJ Kaczorowski, and ML Garcia. Purification and reconstituition of the high-conductance, Calcium-Activated Potassium Channel from tracheal smooth-muscle. *J. Biol. Chem.*, 269(1):676–682, 1994.

[86] L Diaz, P Meera, J Amigo, E Stefani, O Alvarez, L Toro, and R Latorre. Role of the S4 segment in a voltage-dependent calcium-sensitive potassium (hSlo) channel. *J. Biol. Chem.*, 273(49):32430–32436, 1998.

[87] TI Brelidze, XW Niu, and KL Magleby. A ring of eight conserved negatively charged amino acids doubles the conductance of BK channels and prevents inward rectification. *Proc. Natl. Acad. Sci. USA*, 100:9017–9022, 2003.

[88] P Meera, M Wallner, M Song, and L Toro. Large conductance voltage- and calcium-dependent K+ channel, a distinct member of voltage-dependent ion channels with seven N-terminal transmembrane segments (SO-S6), an extracellular N terminus, and an intracellular (S9-S10) C terminus. *Proc. Natl. Acad. Sci. USA*, 94(25):14066–14071, 1997.

[89] JY Shi, G Krishnamoorthy, YW Yang, L Hu, N Chaturvedi, D Harilal, J Qin, and JM Cui. Mechanism of magnesium activation of calcium-activated potassium channels. *Nature*, 418(6900):876–880, 2002.

[90] XH Zeng, XM Xia, and CJ Lingle. Divalent cation sensitivity of BK channel activation supports the existence of three distinct binding sites. *J. Gen. Physiol.*, 125:273–286, 2005.

[91] YX Jiang, A Lee, JY Chen, M Cadene, BT Chait, and R MacKinnon. Crystal structure and mechanism of a calcium-gated potassium channel. *Nature*, 417(6888):515–522, 2002.

[92] YX Jiang, A Lee, JY Chen, M Cadene, BT Chait, and R MacKinnon. The open pore conformation of potassium channels. *Nature*, 417(6888):523–526, 2002.

[93] M Schreiber and L Salkoff. A novel calcium-sensing domain in the BK channel. *Biophys. J.*, 73(3):1355–1363, 1997.

[94] P Orio, P Rojas, G Ferreira, and R Latorre. New disguises for an old channel: MaxiK channel beta-subunits. *News Physiol. Sci.*, 17:156–161, 2002.

[95] L Bao and DH Cox. Gating and ionic currents reveal how the BKCa channel's Ca2+ sensitivity is enhanced by its beta 1 subunit. *J. Gen. Physiol.*, 126(4):393–412, 2005.

[96] X Qian, CM Nimigean, XW Niu, BL Moss, and KL Magleby. Slo1 tail domains, but not the Ca2+ bowl, are required for the beta 1 subunit to increase the apparent Ca2+ sensitivity of BK channels. *J. Gen. Physiol.*, 120(6):829–843, 2002.

[97] YX Jiang, A Pico, M Cadene, BT Chait, and R MacKinnon. Structure of the RCK domain from the E. coli K+ channel and demonstration of its presence in the human BK channel. *Neuron*, 29(3):593–601, 2001.

[98] M Kvansakul, JC Adams, and E Hohenester. Structure of a thrombospondin C-terminal fragment reveals a novel calcium core in the type 3 repeats. *EMBO J.*, 23(6):1223–1233, 2004.

[99] SB Long, EB Campbell, and R MacKinnon. Crystal structure of a mammalian voltage-dependent Shaker family K+ channel. *Science*, 309(5736):897–903, 2005.

[100] N Guex and MC Peitsch. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*, 18(15):2714–2723, 1997.

[101] D van der Spoel, E Lindhal, B Hess, AR van Buuren, E Apol, PJ Meulenhoff, DP Tieleman, ALTM Sijbers, KA Feenstra, R van Drunen, and HJC Berendsen. Gromacs User Manual version 3.2. , ():, 2004.

[102] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. Di Nola, and J.R. Haak. Molecular dynamics with coupling to an external bath. *J.Chem.Phys.*, 81:3684 – 3690, 1984.

[103] DA Doyle, JM Cabral, RA Pfuetzner, AL Kuo, JM Gulbis, SL Cohen, BT Chait, and R MacKinnon. The structure of the potassium channel: Molecular basis of K+ conduction and selectivity. *Science*, 280(5360):69–77, 1998.

[104] T Harris, AR Graber, and M Covarrubias. Allosteric modulation of a neuronal K+ channel by 1-alkanols is linked to a key residue in the activation gate. *Am. J. Physiol.Cell Physiol.*, 285(4):C788–C796, 2003.

[105] D Kerschensteiner, F Monje, and M Stocker. Structural determinants of the regulation of the voltage-gated potassium channel Kv2.1 by the modulatory alpha-subunit Kv9.3. *J. Biol Chem.*, 278(20):18154–18161, 2003.

[106] AJ Labro, AL Raes, I Bellens, N Ottschytsch, and DJ Snyders. Gating of Shaker-type channels requires the flexibility of S6 caused by prolines. *J. Biol. Chem.*, 278(50):50724–50731, 2003.

[107] N Ottschytsch, A Raes, D Van Hoorick, and DJ Snyders. Obligatory heterotetramerization of three previously uncharacterized Kv channel alpha-subunits identified in the human genome. *Proc. Natl. Acad. Sci. U. S. A.*, 99(12):7986–7991, 2002.

[108] N Ottschytsch, AL Raes, JP Timmermans, and DJ Snyders. Domain analysis of Kv6.3, an electrically silent channel. *J.Physiol.-London*, 568(3):737–747, 2005.

[109] JK Jaiswal. Calcium - how and why? *J. Biosci.*, 26(3):357–363, 2001.

[110] L Bao, C Kaldany, EC Holmstrand, and DH Cox. Mapping the BKCa channel's "Ca2+ bowl": Side-chains essential for Ca2+ sensing. *J. Gen. Physiol.*, 123(5):475–489, 2004.

[111] JZ Sheng, A Weljie, L Sy, SZ Ling, HJ Vogel, and AP Braun. Homology modeling identifies C-terminal residues that contribute to the Ca2+ sensitivity of a BKCa channel. *Biophys. J.*, 89(5):3079–3092, 2005.

[112] Z. Qu, R. Fischmeister, and C. Hartzell. Mouse bestrophin-2 is a bona fide cl- channel: identification of a residue important in anion binding and conduction. *J.Gen.Physiol.*, 123:327 – 340, 2004.

[113] Z. Qu, Y. Cui, and C. Hartzell. A short motif in the c-terminus of mouse bestrophin 4 inhibits its activation as a cl channel. *FEBS Lett.*, 580:2141 – 2146, 2006.

[114] T. Tsunenari, H. Sun, J. Williams, H. Cahill, P. Smallwood, K.W. Yau, and J. Nathans. Structure-function analysis of the bestrophin family of anion channels. *J.Biol.Chem.*, 278:41114 – 41125, 2003.

[115] K. Kunzelmann, V.M. Milenkovic, M. Spitzner, R. Barro Soria, and R. Schreiber. Calcium-dependent chloride conductance in epithelia: is there a contribution by bestrophin? *Pflugers Arch.*, 454:879 – 889, 2007.

[116] J.B. Stanton, A.F. Goldberg, G. Hoppe, L.Y. Marmorstein, and A.D. Marmorstein. Hydrodynamic properties of porcine bestrophin-1 in triton x-100. *Biochim.Biophys.Acta*, 1758:241 – 247, 2006.

[117] S. Pifferi, G. Pascarella, A. Boccaccio, A. Mazzatenta, S. Gustincich, A. Menini, and S. Zucchelli. Bestrophin-2 is a candidate calcium-activated chloride channel involved in olfactory transduction. *Proc.Natl.Acad.Sci.U.S.A*, 103:12929 – 12934, 2006.

[118] V. Duta, A.J. Szkotak, D. Nahirney, and M. Duszyk. The role of bestrophin in airway epithelial ion transport. *FEBS Lett.*, 577:551 – 554, 2004.

[119] R. Barro-Soria, M. Spitzner, R. Schreiber, and K. Kunzelmann. Bestrophin 1 enables ca2+ activated cl- conductance in epithelia. *J.Biol.Chem.*, 577:0 – 0, 2006.

[120] R. Barro-Soria, R. Schreiber, and K. Kunzelmann. Bestrophin 1 and 2 are components of the ca2+ activated cl- conductance in mouse airways. *Biochim.Biophys.Acta*, 1783:1993 – 2000, 2008.

[121] A. Boccaccio and A. Menini. Temporal development of cyclic nucleotide-gated and ca2+-activated cl- currents in isolated mouse olfactory sensory neurons. *J.Neurophysiol.*, 98:153 – 160, 2007.

[122] S.J. Kleene. The electrochemical basis of odor transduction in vertebrate olfactory cilia. *Chem.Senses*, 98:0 – 0, 2008.

[123] A. Caputo, E. Caci, L. Ferrera, N. Pedemonte, C. Barsanti, E. Sondo, U. Pfeffer, R. Ravazzolo, O. Zegarra-Moran, and L. J. V. Galietta. TMEM16A, a membrane protein associated with calcium-dependent chloride channel activity. *Science*, 322:590–594, 2008.

[124] B. C. Schroeder, T. Cheng, Y. N. Jan, and L. Y. Jan. Expression cloning of TMEM16A as a calcium-activated chloride channel subunit. *Cell*, 134:1019–1029, 2008.

[125] Y. D. Yang, H. Cho, J. Y. Koo, M. H. Tak, Y. Cho, W-S. Shim, S. P. Park, J. Lee, B. Lee, B-M. Kim, R. Raouf, Y. K. Shin, and U. Oh. TMEM16A confers receptor-activated calcium-dependent chloride conductance. *Nature*, 455:1210–U36, 2008.

[126] C. Hartzell, Z. Qu, I. Putzier, L. Artinian, L.T. Chien, and Y. Cui. Looking chloride channels straight in the eye: bestrophins, lipofuscinosis, and retinal degeneration. *Physiology.(Bethesda.)*, 20:292 – 302, 2005.

[127] K. White, A. Marquardt, and B.H. Weber. Vmd2 mutations in vitelliform macular dystrophy (best disease) and other maculopathies. *Hum.Mutat.*, 15:301 – 308, 2000.

[128] F. Horling. The vmd2 database. *Hum.Mutat.*, 15:0 – 0, 2006.

[129] Z.Q. Qu, K. Yu, Y.Y. Cui, C. Ying, and C. Hartzell. Activation of bestrophin cl- channels is regulated by c-terminal domains. *J.Biol.Chem.*, 282:17460 – 17467, 2007.

[130] V.M. Milenkovic, A. Rivera, F. Horling, and B.H. Weber. Insertion and topology of normal and mutant bestrophin-1 in the endoplasmic reticulum membrane. *J.Biol.Chem.*, 282:1313 – 1321, 2007.

[131] L. Salkoff, A. Butler, G. Ferreira, C. Santi, and A. Wei. High-conductance potassium channels of the slo family. *Nat.Rev.Neurosci.*, 7:921 – 931, 2006.

[132] A. Butler, S. Tsunoda, D.P. McCobb, A. Wei, and L. Salkoff. mslo, a complex mouse gene encoding "maxi" calcium-activated potassium channels. *Science*, 261:221 – 224, 1993.

[133] D.P. McCobb, N.L. Fowler, T. Featherstone, C.J. Lingle, M. Saito, J.E. Krause, and L. Salkoff. A human calcium-activated potassium channel gene expressed in vascular smooth muscle. *Am.J.Physiol*, 269:H767 – H777, 1995.

[134] C.B. Carlson, J. Lawler, and D.F. Mosher. Structures of thrombospondins. *Cell Mol.Life Sci.*, 65:672 – 686, 2008.

[135] A. Kranjc, C. Anselmi, P. Carloni, and F.E. Blaney. Structural models of human big conductance calcium- and voltage-gated potassium channels. *Comp.Phys.Comm.*, 177:21 – 26, 2007.

[136] M. Kvansakul, J.C. Adams, and E. Hohenester. Structure of a thrombospondin c-terminal fragment reveals a novel calcium core in the type 3 repeats. *EMBO J.*, 23:1223 – 1233, 2004.

[137] T.M. Misenheimer and D.F. Mosher. Calcium ion binding to thrombospondin 1. *J.Biol.Chem.*, 270:1729 – 1733, 1995.

[138] H. Chen, M. Deere, J.T. Hecht, and J. Lawler. Cartilage oligomeric matrix protein is a calcium-binding protein, and a mutation in its type 3 repeats causes conformational changes. *J.Biol.Chem.*, 275:26538 – 26544, 2000.

[139] C.B. Carlson, D.A. Bernstein, D.S. Annis, T.M. Misenheimer, B.L. Hannah, D.F. Mosher, and J.L. Keck. Structure of the calcium-rich signature domain of human thrombospondin-2. *Nat.Struct.Mol.Biol.*, 12:910 – 914, 2005.

[140] R. Piskorowski and R.W. Aldrich. Calcium activation of bkca potassium channels lacking the calcium bowl and rck domains. *Nature*, 420:499 – 502, 2002.

[141] X. Qian, C.M. Nimigean, X. Niu, B.L. Moss, and K.L. Magleby. Slo1 tail domains, but not the ca2+ bowl, are required for the beta 1 subunit to increase the apparent ca2+ sensitivity of bk channels. *J.Gen.Physiol*, 120:829 – 843, 2002.

[142] H. Stohr, A. Marquardt, I. Nanda, M. Schmid, and B.H. Weber. Three novel human vmd2-like genes are members of the evolutionary highly conserved rfp-tm family. *Eur.J.Hum.Genet.*, 10:281 – 284, 2002.

[143] D.S. Gerhard, L. Wagner, E.A. Feingold, C.M. Shenmen, L.H. Grouse, G. Schuler, S.L. Klein, S. Old, R. Rasooly, P. Good, M. Guyer, A.M. Peck, J.G. Derge, D. Lipman, F.S. Collins, W. Jang, S. Sherry, M. Feolo, L. Misquitta, E. Lee, K. Rotmistrovsky, S.F. Greenhut, C.F. Schaefer, K. Buetow, T.I. Bonner, D. Haussler, J. Kent, M. Kiekhaus, T. Furey, M. Brent, C. Prange, K. Schreiber, N. Shapiro, N.K. Bhat, R.F. Hopkins, F. Hsie, T. Driscoll, M.B. Soares, T.L. Casavant, T.E. Scheetz, M.J. Brown-stein, T.B. Usdin, S. Toshiyuki, P. Carninci, Y. Piao, D.B. Dudekula, M.S. Ko, K. Kawakami, Y. Suzuki, S. Sugano, C.E. Gruber, M.R. Smith, B. Simmons, T. Moore, R. Waterman, S.L. Johnson, Y. Ruan, C.L. Wei, S. Mathavan, P.H. Gunaratne, J. Wu, A.M. Garcia, S.W. Hulyk, E. Fuh, Y. Yuan, A. Sneed, C. Kowis, A. Hodgson, D.M. Muzny, J. McPherson, R.A. Gibbs, J. Fahey, E. Helton, M. Ketteman, A. Madan, S. Rodrigues, A. Sanchez, M. Whiting, A. Madari, A.C. Young, K.D. Wetherby, S.J. Granite, P.N. Kwong, C.P. Brinkley, R.L. Pearson, G.G. Bouffard, R.W. Blakesly, E.D. Green, M.C. Dickson, A.C. Rodriguez, J. Grimwood, J. Schmutz, R.M. Myers, Y.S. Butterfield, M. Griffith, O.L. Griffith, M.I. Krzywinski, N. Liao, R. Morin, D. Palmquist, A.S. Petrescu, U. Skalska, D.E. Smailus, J.M. Stott, A. Schnerch, J.E. Schein, S.J. Jones, R.A. Holt, A. Baross,

M.A. Marra, S. Clifton, K.A. Makowski, S. Bosak, and J. Malek. The status, quality, and expansion of the nih full-length cdna project: the mammalian gene collection (mgc). *Genome Res.*, 14:2121 – 2127, 2004.

[144] F. Kramer, H. Stohr, and B.H. Weber. Cloning and characterization of the murine vmd2 rfp-tm gene family. *Cytogenet.Genome Res.*, 105:107 – 114, 2004.

[145] E.E. Slawson, C.D. Shaffer, C.D. Malone, W. Leung, E. Kellmann, R.B. Shevchek, C.A. Craig, S.M. Bloom, J. Bogenpohl, J. Dee, E.T. Morimoto, J. Myoung, A.S. Nett, F. Ozsolak, M.E. Tittiger, A. Zeug, M.L. Pardue, J. Buhler, E.R. Mardis, and S.C. Elgin. Comparison of dot chromosome sequences from d. melanogaster and d. virilis reveals an enrichment of dna transposon sequences in heterochromatic domains. *Genome Biol.*, 7:R15 – 0, 2006.

[146] M.D. Adams, S.E. Celniker, R.A. Holt, C.A. Evans, J.D. Gocayne, P.G. Amanatides, S.E. Scherer, P.W. Li, R.A. Hoskins, R.F. Galle, R.A. George, S.E. Lewis, S. Richards, M. Ashburner, S.N. Henderson, G.G. Sutton, J.R. Wortman, M.D. Yandell, Q. Zhang, L.X. Chen, R.C. Brandon, Y.H. Rogers, R.G. Blazej, M. Champe, B.D. Pfeiffer, K.H. Wan, C. Doyle, E.G. Baxter, G. Helt, C.R. Nelson, G.L. Gabor, J.F. Abril, A. Agbayani, H.J. An, C. Andrews-Pfannkoch, D. Baldwin, R.M. Ballew, A. Basu, J. Baxendale, L. Bayraktaroglu, E.M. Beasley, K.Y. Beeson, P.V. Benos, B.P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M.R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K.C. Burtis, D.A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J.M. Cherry, S. Cawley, C. Dahlke, L.B. Davenport, P. Davies, B. de Pablos, A. Delcher, Z. Deng, A.D. Mays, I. Dew, S.M. Dietz, K. Dodson, L.E. Doup, M. Downes, S. Dugan-Rocha, B.C. Dunkov, P. Dunn, K.J. Durbin, C.C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A.E. Gabrielian, N.S. Garg, W.M. Gelbart, K. Glasser, A. Glodek, F. Gong, J.H. Gorrell, Z. Gu, P. Guan, M. Harris, N.L. Harris, D. Harvey, T.J. Heiman, J.R. Hernandez, J. Houck, D. Hostin, K.A. Houston, T.J. Howland, M.H. Wei, C. Ibegwam, M. Jalali, F. Kalush, G.H. Karpen, Z. Ke, J.A. Kennison, K.A. Ketchum, B.E. Kimmel, C.D. Kodira, C. Kraft, S. Kravitz, D. Kulp, Z. Lai, P. Lasko, Y. Lei, A.A. Levitsky, J. Li, Z. Li, Y. Liang, X. Lin, X. Liu, B. Mattei, T.C. McIntosh, M.P. McLeod, D. McPherson, G. Merkulov, N.V. Milshina, C. Mobarry, J. Morris, A. Moshrefi, S.M. Mount, M. Moy, B. Murphy, L. Murphy, D.M. Muzny, D.L. Nelson, D.R. Nelson, K.A. Nelson, K. Nixon, D.R. Nusskern, J.M. Pacleb, M. Palazzolo, G.S. Pittman, S. Pan, J. Pollard, V. Puri, M.G. Reese, K. Reinert, K. Remington, R.D. Saunders, F. Scheeler, H. Shen, B.C. Shue, I. Siden-Kiamos, M. Simpson, M.P. Skupski, T. Smith, E. Spier, A.C. Spradling, M. Stapleton, R. Strong, E. Sun, R. Svirskas, C. Tector, R. Turner, E. Venter, A.H. Wang, X. Wang, Z.Y. Wang, D.A. Wassarman, G.M. Weinstock, J. Weissenbach, S.M. Williams, WoodageT, K.C. Worley, D. Wu, S. Yang, Q.A. Yao, J. Ye, R.F. Yeh, J.S. Zaveri, M. Zhan, G. Zhang, Q. Zhao, L. Zheng, X.H. Zheng, F.N. Zhong, W. Zhong, X. Zhou, S. Zhu, X. Zhu, H.O. Smith, R.A. Gibbs, E.W. Myers, G.M. Rubin, and J.C. Venter. The genome sequence of drosophila melanogaster. *Science*, 287:2185 – 2195, 2000.

[147] S. Henikoff and J.G. Henikoff. Amino acid substitution matrices from protein blocks. *Proc.Natl.Acad.Sci.U.S.A*, 89:10915 – 10919, 1992.

[148] M.A. Marti-Renom, A.C. Stuart, A. Fiser, R. Sanchez, F. Melo, and A. Sali. Comparative protein structure modeling of genes and genomes. *Annu.Rev.Biophys.Biomol.Struct.*, 29:291 – 325, 2000.

[149] J. Wang, P. Cieplak, and P.A. Kollman. How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? *J.Comput.Chem.*, 21:1049 – 1074, 2000.

[150] J.W. Ponder and D.A. Case. Force fields for protein simulations. *Adv.Protein Chem.*, 66:27 – 85, 2003.

[151] B. Hess, H. Bekker, H.J.C. Berendsen, and J.G.E.M. Fraaije. Lincs: A linear constraint solver for molecular simulations. *J.Comp.Chem.*, 18:1463 – 1472, 1997.

[152] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, and L.G. Pedersen. A smooth particle mesh ewald method. *J.Chem.Phys.*, 103:8577 – 8593, 1995.

[153] G. Schaftenaar and J.H. Noordik. Molden: a pre- and post-processing program for molecular and electronic structures. *J.Comput.-Aided Mol.Design*, 14:123 – 134, 2000.

[154] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, III Cheatham, T.E., S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman. Amber, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp.Phys.Commun.*, 91:1 – 41, 1995.

[155] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen. Gromacs:a message- passing parallel molecular dynamics implementation. *Comp.Phys.Comm.*, 91:43 – 56, 1995.

[156] E. Lindhal, B. Hess, and D. van der Spoel. Gromacs 3.0: a package for molecular simulation and trajectory analysis. *J.Mol.Mod.*, 7:306 – 317, 2001.

[157] W. Humphrey, A. Dalke, and K. Schulten. Vmd: visual molecular dynamics. *J.Mol.Graph.*, 14:33 – 38, 1996.

[158] V. Costa and P. Carloni. Calcium binding to the transmembrane domain of the sarcoplasmic reticulum ca2+-atpase: Insights from molecular modeling. *Proteins*, 50:104 – 113, 2003.

[159] S. Marchand and B. Roux. Molecular dynamics study of calbindin d9k in the apo and singly and doubly calcium-loaded states. *Proteins*, 33:265 – 284, 1998.

[160] M. Dal Peraro, S. Raugei, P. Carloni, and M.L. Klein. Solute-solvent charge transfer in aqueous solution. *Chemphyschem.*, 6:1715 – 1718, 2005.

[161] J. Aqvist. Ion water interaction potentials derived from free-energy perturbation simulations. *J.Phys.Chem.*, 94:8021 – 8024, 1990.

[162] G.M. Bradbrook, T. Gleichmann, S.J. Harrop, J. Habash, J. Raftery, J. Kalb, J. Yariv, I.H. Hillier, and J.R. Helliwell. X-ray and molecular dynamics studies of concanavalin-a glucoside and mannoside complexes - relating structure to thermodynamics of binding. *J.Chem.Soc.-Faraday Trans.*, 94:1603 – 1611, 1998.

[163] J.W. Ponder and D.A. Case. Force fields for protein simulations. *Adv.Protein Chem.*, 66:27 – 85, 2003.

[164] J. Wang, P. Cieplak, and P.A. Kollman. How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? *J.Comput.Chem.*, 21:1049 – 1074, 2000.

[165] M. Kvansakul, J.C. Adams, and E. Hohenester. Structure of a thrombospondin c-terminal fragment reveals a novel calcium core in the type 3 repeats. *EMBO J.*, 23:1223 – 1233, 2004.

[166] T.M. Misenheimer and D.F. Mosher. Calcium ion binding to thrombospondin 1. *J.Biol.Chem.*, 270:1729 – 1733, 1995.

[167] H. Chen, M. Deere, J.T. Hecht, and J. Lawler. Cartilage oligomeric matrix protein is a calcium-binding protein, and a mutation in its type 3 repeats causes conformational changes. *J.Biol.Chem.*, 275:26538 – 26544, 2000.

[168] K. Spiegel, A. Magistrato, P. Maurer, P. Ruggerone, U. Rothlisberger, P. Carloni, J. Reedijk, and M.L. Klein. Parameterization of azole-bridged dinuclear platinum anticancer drugs via a qm/mm force matching procedure. *J.Comput.Chem.*, 29:38 – 49, 2008.

[169] F. Horling. The vmd2 database. *J.Comput.Chem.*, 29:0 – 0, 2006.

[170] H.C. Hartzell, Z. Qu, K. Yu, Q. Xiao, and L.T. Chien. Molecular physiology of bestrophins: multifunctional membrane proteins linked to best disease and other retinopathies. *Physiol Rev.*, 88:639 – 672, 2008.

[171] J.W. Torrance, M.W. Macarthur, and J.M. Thornton. Evolution of binding sites for zinc and calcium ions playing structural roles. *Proteins*, 71:813 – 830, 2007.

[172] R.E. Bulo, D. Donadio, A. Laio, F. Molnar, J. Rieger, and M. Parrinello. "site binding" of ca2+ ions to polyacrylates in water: A molecular dynamics study of coiling and aggregation. *J.Mol.Graph.*, 40:3437 – 3442, 2007.

[173] Laio A and Parrinello M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.*, 99:12562–12566, 2002.

[174] Z. Qu, R.W. Wei, W. Mann, and H.C. Hartzell. Two bestrophins cloned from xenopus laevis oocytes express ca2+-activated cl- currents. *J.Biol.Chem.*, 278:49563 – 49572, 2003.

[175] N. Garcia-Sanz, P. Valente, A. Gomis, A. Fernandez-Carvajal, G. Fernandez-Ballester, F. Viana, C. Belmonte, and A. Ferrer-Montiel. A role of the transient receptor potential domain of vanilloid receptor i in channel gating. *J.Neurosci.*, 27:11641 – 11650, 2007.

[176] Q. Xiao, A. Prussia, K. Yu, Y y. Cui, and C. H. Hartzell. Regulation of Bestrophin Cl Channels by Calcium: Role of the C Terminus. *J. Gen. Physiol.*, 132:681–692, 2008.

[177] Y. Cao, R. A. Musah, D. B. Goodin, and D. E. Mcree. Probing the strength and character of an asp-his-x hydrogen bond by introducing buried charges. *to be published*, 0:0, 0.

[178] Will, R. G., Ironside, J. W., Zeidler, M., Cousens, S. N., Estibeiro, K., Alperovitch, A., Poser, S., Pocchiari, M., Hofman, A. and Smith, P. G. A new variant of creutzfeldt-jakob disease in the uk. *Lancet*, 347:921–925, 1996.

[179] Prusiner, S. B. Prions. *Proc Natl Acad Sci U S A*, 95:13363–13383, 1998.

[180] Calzolai, L. and Zahn, R. Influence of ph on nmr structure and stability of the human prion protein globular domain. *J Biol Chem*, 278:35592–35596, 2003.

[181] PC Pauly and DA Harris. Copper stimulates endocytosis of the prion protein. *J. Biol. Chem.*, 273(50):33107–33110, 1998.

[182] NT Watt and NM Hooper. The prion protein and neuronal zinc homeostasis. *Trends Biochem. Sci.*, 28(8):406–410, 2003.

[183] DR Brown. Copper and prion disease. *Brain Res. Bulletin*, 55(2):165–173, 2001.

[184] LB Chiarini, ARO Freitas, SM Zanata, RR Brentani, VR Martins, and R Linden. Cellular prion protein transduces neuroprotective signals. *EMBO J.*, 21(13):3317–3326, 2002.

[185] S Mouillet-Richard, M Ermonval, C Chebassier, JL Laplanche, S Lehmann, JM Launay, and O Kellermann. Signal transduction through prion protein. *Science*, 289(5486):1925–1928, 2000.

[186] J Collinge, MA Whittington, KCL Sidle, CJ Smith, MS Palmer, AR Clarke, and JGR Jefferys. Prion protein is necessary for normal synaptic function. *Nature*, 370(6487):295–297, 1994.

[187] GR Mallucci, S Ratte, EA Asante, J Linehan, I Gowland, JGR Jefferys, and J Collinge. Post-natal knockout of prion protein alters hippocampal CA1 properties, but does not result in neurodegeneration. *EMBO J.*, 21(3):202–210, 2002.

[188] E Graner, AF Mercadante, SM Zanata, OV Forlenza, ALB Cabral, SS Veiga, MA Juliano, R Roesler, R Walz, A Minetti, I Izquierdo, VR Martins, and RR Brentani. Cellular prion protein binds laminin and mediates neuritogenesis. *Mol. Brain Res.*, 76(1):85–92, 2000.

[189] L Solforosi, JR Criado, DB McGavern, S Wirz, M Sanchez-Alavez, S Sugama, LA DeGiorgio, BT Volpe, E Wiseman, G Abalos, E Masliah, D Gilden, MB Oldstone, B Conti, and RA Williamson. Cross-linking cellular prion protein triggers neuronal apoptosis in vivo. *Science*, 303(5663):1514–1516, 2004.

[190] Glenn C. Telling, Michael Scott, James Mastrianni, Ruth Gabizon, Marilyn Torchia, Fred E. Cohen, Stephen J. DeArmond, and Stanley B. Prusiner. Prion propagation in mice expressing human and chimeric prp transgenes implicates the interaction of cellular prp with another protein. *Cell*, 83:79–90, 1995.

[191] K. Kaneko, L. Zulianello, M. Scott, C. M. Cooper, A. C. Wallace, T. L. James, F. E. Cohen, and S. B. Prusiner. Evidence for protein X binding to a discontinuous epitope on the cellular prion protein during scrapie prion propagation. *Proc. Nat. Acad. Sci. U. S. A.*, 94:10069–10074, 1997.

[192] Horiuchi, M. and Caughey, B. Specific binding of normal prion protein to the scrapie form via a localized domain initiates its conversion to the protease-resistant state. *Embo J.*, 18:3193–3203, 1999.

[193] C. R. Trevitt and J. Collinge. A systematic review of prion therapeutics in experimental models. *Brain*, 129:2241–2265, 2006.

[194] Weissmann, C. and Aguzzi, A. Approaches to therapy of prion diseases. *Annu Rev Med*, 56:321–344, 2005.

[195] Ludewigs, H., Zuber, C., Vana, K., Nikles, D., Zerr, I. and Weiss, S. Therapeutic approaches for prion disorders. *Expert Rev Anti Infect Ther*, 5:613–630, 2007.

[196] Heppner, F. L. and Aguzzi, A. Recent developments in prion immunotherapy. *Curr Opin Immunol*, 16:594–598, 2004.

[197] Enari, M., Flechsig, E. and Weissmann, C. Scrapie prion protein accumulation by scrapie-infected neuroblastoma cells abrogated by exposure to a prion protein antibody. *Proc Natl Acad Sci U S A*, 98:9295–9299, 2001.

[198] Sigurdsson, E. M., Sy, M. S., Li, R., Scholtzova, H., Kascsak, R. J., Kascsak, R., Carp, R., Meeker, H. C., Frangione, B. and Wisniewski, T. Anti-prion antibodies for prophylaxis following prion exposure in mice. *Neurosci Lett*, 336:185–187, 2003.

[199] White, A. R., Enever, P., Tayebi, M., Mushens, R., Linehan, J., Brandner, S., Anstee, D., Collinge, J. and Hawke, S. Monoclonal antibodies inhibit prion replication and delay the development of prion disease. *Nature*, 422:80–83, 2003.

[200] Polymenidou, M., Heppner, F. L., Pellicioli, E. C., Urich, E., Miele, G., Braun, N., Wopfner, F., Schatzl, H. M., Becher, B. and Aguzzi, A. Humoral immune response to native eukaryotic prion protein correlates with anti-prion protection. *Proc Natl Acad Sci U S A*, 101 Suppl 2:14670–14676, 2004.

[201] Donofrio, G., Heppner, F. L., Polymenidou, M., Musahl, C. and Aguzzi, A. Paracrine inhibition of prion propagation by anti-prp single-chain fv miniantibodies. *J Virol*, 79:8330–8338, 2005.

[202] Wuertzer, C. A., Sullivan, M. A., Qiu, X. and Federoff, H. J. Cns delivery of vectored prion-specific single-chain antibodies delays disease onset. *Mol Ther*, 16:481–486, 2008.

[203] Genoud, N., Ott, D., Braun, N., Prinz, M., Schwarz, P., Suter, U., Trono, D. and Aguzzi, A. Antiprion prophylaxis by gene transfer of a soluble prion antagonist. *Am J Pathol*, 172:1287–1296, 2008.

[204] M. K. Boehm, A. L. Corper, T. Wan, M. K. Sohi, B. J. Sutton, J. D. Thornton, P. A. Keep, K. A. Chester, R. H. Begent, and S. J. Perkins. Crystal structure of the anti-(carcinoembryonic antigen) single-chain fv antibody mfe-23 and a model for antigen binding based on intermolecular contacts. *Biochem. J.*, 346:519–528, 2000.

[205] Y-R. Kim, J-S. Kim, S-H. Lee, W-R. Lee, J-N. Sohn, Y-C. Chung, H-K. Shim, S-C. Lee, M-H. Kwon, and Y-S. Kim. Heavy and Light Chain Variable Single Domains of an Anti-DNA Binding Antibody Hydrolyze Both Double- and Single-stranded DNAs without Sequence Specificity. *J. Biol. Chem.*, 281:15287–15295, 2006.

[206] Luginbuhl, B., Kanyo, Z., Jones, R. M., Fletterick, R. J., Prusiner, S. B., Cohen, F. E., Williamson, R. A., Burton, D. R. and Pluckthun, A. Directed evolution of an anti-prion protein scfv fragment to an affinity of 1 pm and its structural interpretation. *J Mol Biol*, 234:75–97, 2006.

[207] de Vries, S. J., van Dijk, A. D., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T. and Bonvin, A. M. Haddock versus haddock: new features and performance of haddock2.0 on the capri targets. *Proteins*, 69:726–733, 2007.

[208] Dominguez, C., Boelens, R. and Bonvin, A. M. Haddock: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc*, 125:1731–1737, 2003.

[209] S. J. Hubbard and J. M. Thornton. *NACCESS*. Department of Biochemistry and Molecular Biology, University College, London, 1993.

[210] V. Campana and L. Zentilin and I. Mirabile and A. Kranjc and P. Casanova and M. Giacca and S. Prusiner and G. Legname and C. Zurzolo. Development of antibody fragments for immunotherapy of prion diseases. *Biochem. J., Epub ahead of print*, 2008.

[211] Ross, T. M., Martinez, P. M., Renner, J. C., Thorne, R. G., Hanson, L. R. and Frey, W. H. Intranasal administration of interferon beta bypasses the blood-brain barrier to target the central nervous system and cervical lymph nodes: a non-invasive treatment strategy for multiple sclerosis. *J Neuroimmunol*, 151:66–77, 2004.

[212] Thorne, R. G., Hanson, L. R., Ross, T. M., Tung, D. and Frey, W. H. Delivery of interferon-beta to the monkey nervous system following intranasal administration. *Neuroscience*, 152:785–797, 2008.

[213] M. L. DeMarco and V. Daggett. From conversion to aggregation: Protofibril formation of the prion protein. *Proc. Natl. Acad. Sci. U. S. A.*, 101:2293–2298, 2004.

[214] Douglas B. Kitchen, Helene Decornez, John R. Furr, and Jurgen Bajorath. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*, 3:935 – 949, 2004.

[215] Sergio Filipe Sousa, Pedro Alexandrino Fernandes, and Maria Joao Ramos. Protein-ligand docking: current status and future challenges. *Proteins*, 65:15 – 26, 2006.

[216] J. H. Lin, A. L. Perryman, J. R. Schames, and J. A. McCammon. Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *J. Am. Chem. Soc.*, 124:5632–5633, 2002.

[217] JH Lin, AL Perryman, JR Schames, and JA McCammon. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers*, 68:47–62, 2003.

[218] Simon J. Teague. Implications of protein flexibility for drug discovery. *Nat Rev Drug Discov*, 2:527 – 541, 2003.

[219] Heather A. Carlson. Protein flexibility and drug design: how to hit a moving target. *Current Opinion in Chemical Biology*, 6:447 – 452, 2002.

[220] Andreas May and Martin Zacharias. Accounting for global protein deformability during protein-protein and protein-ligand docking. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics*, 1754:225 – 231, 2005.

[221] R. Mangoni, D. Roccatano, and A. Di Nola. Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation. *Nat Rev Neurosci*, 35:153 – 162, 1999.

[222] Youngshang Pak and Shaomeng Wang. Application of a molecular dynamics simulation method with a generalized effective potential to the flexible molecular docking problems. *Dev Cell*, 104:354 – 359, 2000.

[223] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, and J.A. McCammon. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*, 98:10037 – 10041, 2001.

[224] J.Andrew McCammon. Target flexibility in molecular recognition. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics*, 1754:221 – 224, 2005.

[225] F.L. Gervasio, A. Laio, and M. Parrinello. Flexible docking in solution using metadynamics. *J Biol Chem*, 127:2600 – 2607, 2005.

[226] D Branduardi, FL Gervasio, A Cavalli, M Recanatini, and M Parrinello. The role of the peripheral anionic site and cation-pi interactions in the ligand penetration of the human AChE gorge. *J. Am. Chem. Soc.*, 127:9147–9155, 2005.

[227] A.V. Vargiu, P. Ruggerone, A. Magistrato, and P. Carloni. Dissociation of minor groove binders from dna: insights from metadynamics simulations. *Nucl.Acids Res.*, 36:5910 – 5921, 2008.

[228] G. Fiorin, A. Pastore, P. Carloni, and M. Parrinello. Using metadynamics to understand the mechanism of calmodulin/target recognition at atomic detail. *Drug Discov Today*, 91:2768 – 2777, 2006.

[229] J.R. Silveira, B. Caughey, and G.S. Baron. Prion protein and the molecular features of transmissible spongiform encephalopathy agents. *Curr Top Microbiol Immunol*, 284:1 – 50, 2004.

[230] G.C. Telling, M. Scott, J. Mastrianni, R. Gabizon, M. Torchia, F.E. Cohen, S.J. DeArmond, and S.B. Prusiner. Prion propagation in mice expressing human and chimeric prp transgenes implicates the interaction of cellular prp with another protein. *Cell*, 83:79 – 90, 1995.

[231] David A. Hilton. Pathogenesis and prevalence of variant creutzfeldt-jakob disease. *J Pathol*, 208:134 – 141, 2006.

[232] Byron Caughey and Gerald S. Baron. Prions and their partners in crime. *Nature*, 443:803 – 810, 2006.

[233] Luigi Calzolai and Ralph Zahn. Influence of ph on nmr structure and stability of the human prion protein globular domain. *J Biol Chem*, 278:35592 – 35596, 2003.

[234] J.C. Gordon, J.B. Myers, T. Folta, V. Shoja, L.S. Heath, and A. Onufriev. H++: a server for estimating pk(a)s and adding missing hydrogens to macromolecules. *Nucl.Acids Res.*, 33:W368 – W371, 2005.

[235] R. Anandakrishnan and A. Onufriev. Analysis of basic clustering algorithms for numerical estimation of statistical averages in biomolecules. *Proc Natl Acad Sci U S A*, 15:165 – 184, 2008.

[236] M.L. Verdonk, J.C. Cole, M.J. Hartshorn, C.W. Murray, and R.D. Taylor. Improved protein-ligand docking using gold. *Nucl.Acids Res.*, 52:609 – 623, 2003.

[237] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R.J. Woods. The amber biomolecular simulation programs. *J Comput Aided Mol Des*, 26:1668 – 1688, 2005.

[238] M. Parrinello and A. Rahman. Polymorphic transitions in single-crystals - a new molecular-dynamics method. *Dev Cell*, 52:7182 – 7190, 1981.

[239] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, and a. E. Mark. Peptide folding: When simulation meets experiment. *Angew. Chem.-Int. Edit.*, 38:236, 1999.

[240] ChemAxon. Calculator plugins were used for structure property prediction and calculation. *Marvin 5.0.0*, 2008, http://www.chemaxon.com.

[241] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, T. Vreven Jr., K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, A. D. Daniels, O. Farkas, A. D. Rabuck, K. Raghavachari, and J. V. Ortiz. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.

[242] Eldridge M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor. Improved protein-ligand docking using gold. *J. Comput-Aided Mol. Des.*, 11:425–445, 1997.

[243] C.A. Baxter, C.W. Murray, D.E. Clark, D.R. Westhead, and M.D. Eldridge. Flexible docking using tabu search and an empirical estimate of binding affinity. *J Virol*, 33:367 – 382, 1998.

[244] G. Jones, P. Willett, and R.C. Glen. Molecular recognition of receptor-sites using a genetic algorithm with a description of desolvation. *J Neurochem*, 245:43 – 53, 1995.

[245] G. Jones, P. Willett, R.C. Glen, A.R. Leach, and R. Taylor. Development and validation of a genetic algorithm for flexible docking. *J Neurochem*, 267:727 – 748, 1997.

[246] Giovanni Bottegoni, Walter Rocchia, Maurizio Recanatini, and Andrea Cavalli. Aclap, autonomous hierarchical agglomerative cluster analysis based protocol to partition conformational datasets. *Bioinformatics*, 22:58 – 65, 2006.

[247] Giovanni Bottegoni, Andrea Cavalli, and Maurizio Recanatini. A comparative study on the application of hierarchical-agglomerative clustering approaches to organize outputs of reiterated docking runs. *J Chem Inf Model*, 46:852 – 862, 2006.

[248] J.M. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, and D.A. Case. Development and testing of a general amber force field. *J Med Chem*, 26:114 – 114, 2005.

[249] C.I. Bayly, P. Cieplak, W.D. Cornell, and P.A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the resp model. *J Virol*, 97:10269 – 10280, 1993.

[250] W.D. Cornell, P. Cieplak, C.I. Bayly, and P.A. Kollman. Application of resp charges to calculate conformational energies, hydrogen-bond energies, and free-energies of solvation. *Eur J Med Chem*, 115:9620 – 9631, 1993.

[251] S. Piana, A. Laio, F. Marinelli, M. Van Troys, D. Bourry, C. Ampe, and J. C. Martins. Predicting the effect of a point mutation on a protein fold: The villin and advillin headpieces and their Pro62Ala mutants. *J. Mol. Biol.*, 375(2):460–470, 2008.

[252] S. Kumar, D. Bouzida, R.H. Swendsen, P.A. Kollman, and J.M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules .1. the method. *Proc Natl Acad Sci U S A*, 13:1011 – 1021, 1992.

[253] Baker N.A., Sept D., Joseph S., Holst M.J., and McCammon J.A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA.*, 98:10037–10041, 2001.

[254] F. Neff and X. Wei and C. Noelker and M. Bacher and Y. Du and R. Dodel. Immunotherapy and naturally occurring autoantibodies in neurodegenerative disorders. *Autoimmun Rev.*, 7:501–507, 2008.

[255] X. Biarnes, A. Ardevol, A. Planas, C. Rovira, A. Laio, and M. Parrinello. The conformational free energy landscape of beta-d-glucopyranose. implications for substrate preactivation in beta-glucoside hydrolases. *J. Am. Chem. Soc.*, 129:10686–10693, 2007.