Molecular simulation studies of the prion protein: from disease-linked variants to ligand binding

Giulia Rossetti

A thesis submitted for the degree of

PhilosophiæDoctor

2010 October

Supervisor: Prof. Paolo Carloni

Supervisor: Prof. Giuseppe Legname

Supervisor: Dr. Alessandra Magistrato

To my family and to my friends.

They keep my heart safe and my soul strong while my eyes are looking ahead.

Acknowledgements

Acknowledgements

I would like to thank Prof. Paolo Carloni for giving me the opportunity to work in his group. He supported me in my professional growth, teaching me how to manage a research project. Remarkably, he encouraged me to always do a further step, either when I did not see the solution. More important, he motivated me with always new scientific open-mind discussions.

I would like to thank Prof. Giuseppe Legname for his huge contribution to my biology knowledge, giving me the opportunity to join his Prion Lab. I would like to thank Dr. Alessandra Magistrato for helping always me to solve my research doubts with constructive discussions.

I would like to thank Dr. Annalisa Pastore (National Institute for Medical Research, London), Prof. Maria Laura Bolognesi, Prof. Andrea Cavalli and Prof. Marinella Roberti (University of Bologna) for all the fruitful collaborations.

I also thank Dr. Emiliano Ippoliti for his technical support. He is always present when something go wrong and you do not know how to handle it.

I thank Dr. Xevi Biarnes, Dr. Agata Kranjc, Gabriele Giachin, Salvatore Bongarzone and Xiaojing Kong for sharing with me the work on Prion with its difficulties. Special thanks to Gabriele Giachin for helping a theoretical chemistry-background-researcher like me, to find a way in the biology field.

I would like to thank all SBP group in SISSA, my work-family, in particular Paola Lupieri, Valeria Losasso, Roberto Marchese, and its outsider member, Dr. Attilio Vargiu, Dr. Claudio Anselmi, Dr. Arturo Robertazzi, Dr. Fabrizio Marinelli and Dr. Enzo Carnevale. These latter helped me much, especially in my first year of PhD.

Finally, thank to my family and to my friends.

The name in front of this thesis is only one, but it is there thank to all these people.

Contents

List of Figures

LIST OF FIGURES

List of Tables

1

Introduction

Prion diseases or transmissible spongiform encephalopathies (TSEs) are fatal neurodegenerative disorders (198). The crucial event in the development of these diseases is the conformational change of a membrane bound protein, the cellular PrP^C in Figure 3.1, into a disease associated, fibril-forming isoform (199).

Despite their rare incidence, TSEs have captured very large attention from the scientific community due to the unorthodox mechanism by which prion diseases are transmitted. Increasing evidence supports the hypothesis that the infectious agent, responsible for the propagation of TSEs, is the prion or PrP^{Sc}, which propagates itself by imposing its conformation onto its non pathogenic isoform, the host cellular prion protein, PrP^C (147; 199) (see Figure 1.1).

PrP^C belongs to the class of amyloid-forming proteins which are associated with a variety of conformational diseases (52; 76).[1] Actually, inducible proteinopathies such as amyloid AA amyloidosis (239) and Alzheimers (168), turn out to share remarkable similarities to prion diseases (239). Hence, studying the key molecular mechanisms involved in prion diseases may help to understand these other amyloidoses. [2]

[1] These include Alzheimers disease, Parkinsons disease, Huntingtons disease, spinocerebellar ataxias, type II diabetes, amyotrophic lateral sclerosis, as well as diffuse Lewy body dementia and fronto-temporal dementias.

[2] We further notice that the striking parallels of infectious prion disorders to the above-mentioned putatively non- infectious protein misfolding and assembly diseases make it more difficult to delimitate their pathological mechanisms from each other.

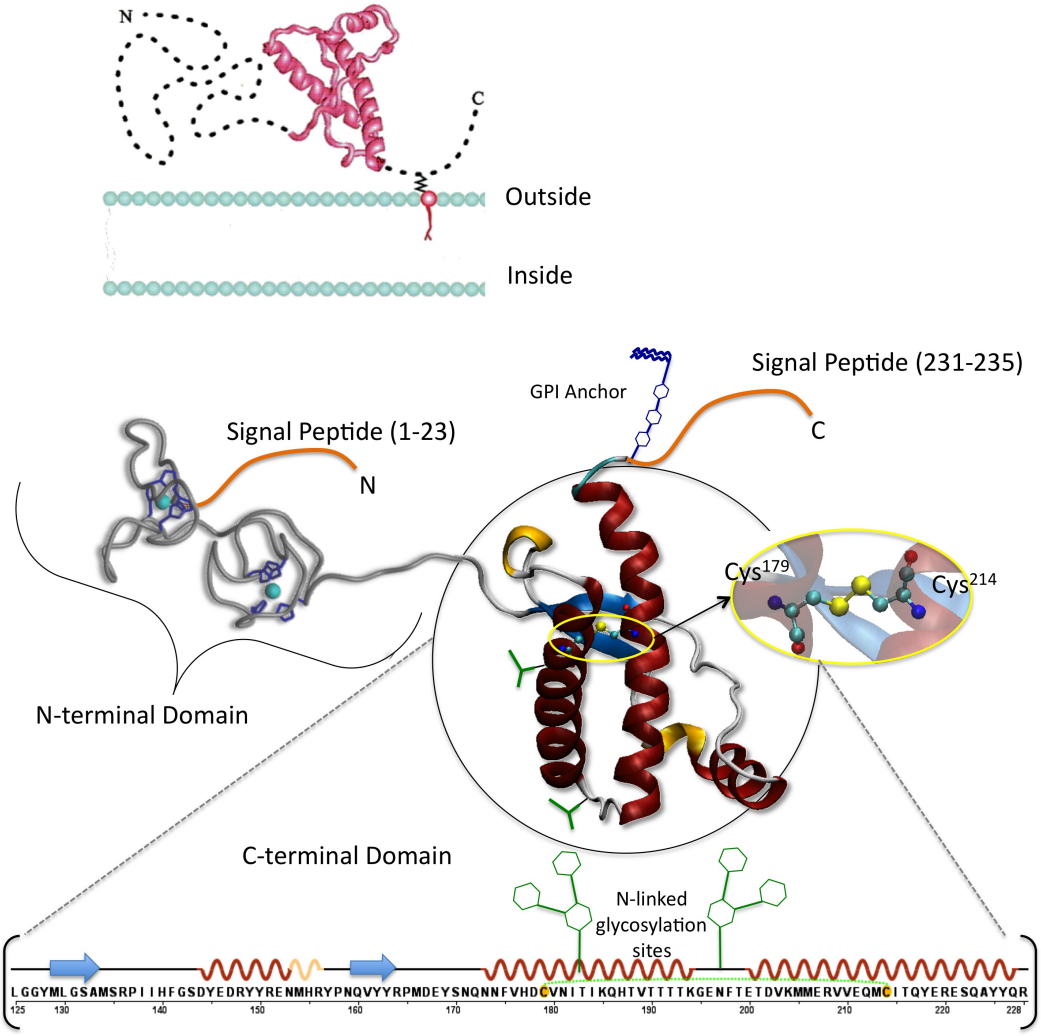


Figure 1.1: The Human Prion Protein - The mature human prion protein features two signal peptides (1-23 and 231-235), an octapeptide repeat region (involved in binding Cu2+ ions (169)), three α-helixes, one disulphide bond (SS) between cysteine residues 179 and 214, and two potential sites for N-linked glycosylation (green forks) at residues 181 and 197. A glycosylphosphatidylinositol anchor (GPI) (yellow box) is attached to the C-terminus of PrP^C.

Little information is available about the infection mechanism by which prion diseases are transmitted. In particular, neither the structural determinants of PrP^{Sc} nor the in vivo conditions which lead to formation of PrP^{Sc}, are known. So far, the tendency of PrP^{C} to undergo induced or spontaneous misfolding has been related to its amino acid sequence (224), to the presence of a highly flexible amino term (196), to the proteins secondary structure elements (2) as well as to the proteins post-translational modifications (206).[1] Most importantly, a relationship between point mutations and TSEs has been firmly established (34; 90; 204; 230). Indeed, in the presence of specific disease-linked mutations, the $PrP^{C} \rightarrow PrP^{Sc}$ conversion appears to occur spontaneously (13; 220; 223). The mutations increase the kinetics of the misfolding process relative to WT (220), possibly mostly because of a destabilization of the native structure and/or increasing the stability of the partially folded intermediate species (13; 150; 220; 223).

Hence, investigating the intrinsic instability of PrP^{C} fold mutants is key for TSEs research.

The $PrP^{C} \rightarrow PrP^{Sc}$ pathogenic conversion process, enhanced by specific mutations, has been related to the global conformational fluctuations of the protein (138; 139). NMR has provide extremely accurate information on structure and conformational properties of proteins involved in neurodegenerative diseases, including prions (38; 54; 55; 73; 92; 113; 115; 153; 203; 204; 236; 247; 252). However, the very large number of disease-linked variants (more than 30), only very few of them have been characterized by NMR so far (115; 252).

Molecular Dynamic (MD) simulations have been successfully used to investigate structure and dynamics of a variety of proteins undergoing fibrillation (65; 68; 79; 104; 125; 156; 174; 190; 192; 233; 250), including the prion protein (94; 118; 143; 194; 208; 253). Hence, such computational approach appears ideal to complement NMR studies in the understanding how disease-linked mutations affect structure and dynamics of the prion protein.

We have therefore decided to carry out MD calculations on all of the disease-linked mutations of the structural domain of the protein (90 % of the total number

[1] Metal ions (53) and cellular protein partners (101; 119) are also thought to be involved in the structural determination of PrP.

of disease-linked mutations), for which the NMR structure is available. We have built a computational protocol able to predict structural facet as well as conformational properties of such variants. The protocol has been validated by NMR experimental data (including those obtained during my thesis by Prof. Legname's Lab (115). Our calculations shed light on common structural traits across the mutants which could play a role for the fibrillation process.

Small ligands have shown to have the opposite effect than disease-linked mutations, as they may decrease in the Gibbs free energy of PrP^C upon binding (Fig. 1.2) (58; 140; 246). In addition, they may suppress conformational rearrangements leading to the $PrP^C \rightarrow PrP^{Sc}$ conversion by cross-linking of distant regions (30; 140; 246). Hence, they might counter react the tendency towards aggregation of the disease-linked mutations as well as the WT. [1] To identify binding poses, molecular docking methodologies are well suited (103). Unfortunately these approaches, whilst fast and useful for enzymes and receptors, are likely to fail here because of the lack of a well-defined binding cavity: only shallow binding sites are present in the structure, as shown in an elegant NMR study by Kuwata and co-workers (140).

Prompted by the relevance of this issue, I tried to establish a computational approach that attempts to include conformational selection (29) and induced fit (132) processes, which play an important role for ligand binding. The work has been done in collaboration with my colleagues Agata Kranjc, Salvatore Bongarzone and Xevi Biarns. We focus on the ligand (2-pyrrolidin-1-yl-N-[4-[4-(2-pyrrolidin-1-yl-acetylamino)-benzyl]-phenyl]-acetamide), for which the free energy of binding to the prion protein has been measured (140). Our protocol includes standard docking methods, Molecular Dynamic (MD) and metadynamics-based (141) free energy simulations in explicit solvent. After an educated guess of the ligand poses using standard docking programs, we have relaxed the ligand/target complex by MD, and sorting out the poses by their binding energy using metadynamics-based (141) free energy calculations. The method follows previous molecular docking approaches designed to account for protein flexibility (11; 36; 47; 227). It then adds a further step by using the metadynamics approach to evaluate poses and potency of the ligand. The accuracy of the protocol

[1] This strategy is applicable regardless of the disease etiology (58).

is established by a comparison not only with thermodynamic data (140) but also with qualitative structural insights obtained by NMR (140).

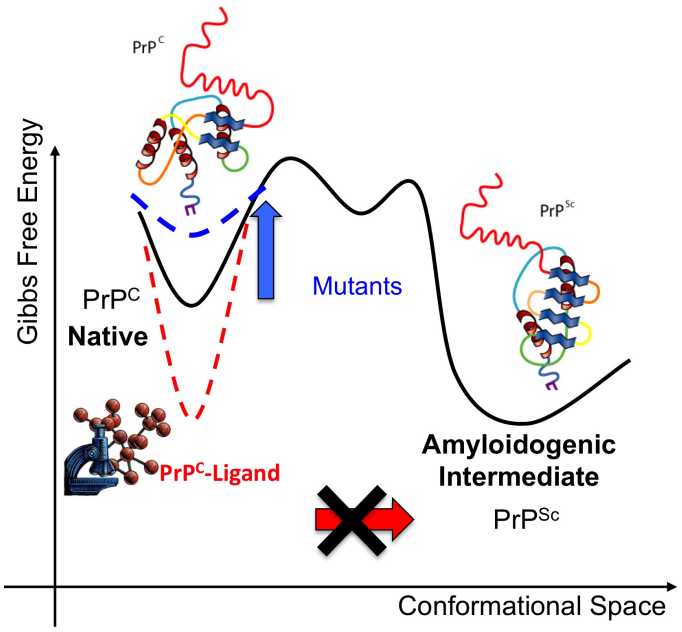


Figure 1.2: PrP^C **to** PrP^{Sc} **conversion** - Qualitative scheme illustrating the Gibbs free energy change in the interconversion, as proposed by Prusiner et al. (199). The interconversion is thought to entail intermediate partially folded conformers (58; 139). The mutation could destabilize PrP^C (hence increasing the population of intermediates), whilst small ligands may stabilize the PrPC conformation (and hence reduce the population of intermediates species) (139; 220).

2

Materials & methods

2.1 Biomolecular Simulations

"Certainly no subject or field is making more progress on so many fronts at the present moment, than biology, and if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms."

Richard P. Feynman from *Six easy pieces* 1963

These words were pronounced about 50 years ago and discussion is still open. How can we 'see' this wiggling and jiggling and understand how it drives biology? Can we believe in modeling?

These are a question often asked by biologists and biochemists. Giving an answer requires not only the understanding of the strengths and limitations of current computational biomolecular modeling and simulation methods, but also their ranges of application. The scepticism encountered among experimentalists of all biomolecular modeling is sometimes misplaced; nevertheless equally misguided is a blind acceptance of modeling results without critical analysis.

Biomolecular modeling is a fertile and growing area, with exciting opportunities and potential applications. It is crucial for the biomolecular modeler to understand the issues of interest to biologists; but more important is to understand the complexity of biological systems and how to tackle them effectively by modeling. Vast amounts of data are being provided by large-scale research efforts in genomics, proteomics, glycomics and structural biology. The challenge for biomolecular modeling is to help in efforts to use these diverse data to develop new drugs, therapies, catalysts and biologically based nanotechnology. (231) Increasingly, computer simulations of biological macromolecules are helping to meet this challenge. In particular, molecular simulations have been more and more being employed to investigate features not directly accessible to experiments. Whereas it is indeed possible to take "still snapshots" of crystal structures and

probe features of the motion of molecules through NMR, no current experimental technique allows access to all the time scales of motion with atomic resolution. Simulations based on fundamental physics offer the potential of filling-in these crucial 'gaps', modeling how proteins and other biomolecules move, fluctuate, interact, react and function. Applications include studies of protein folding and conformational changes, association of proteins with small molecules or other proteins, structure-based drug design, computation of binding free energies for ligands, modeling the dynamics of ion channels and transport across membranes and modeling and analysis of enzyme mechanisms.

Improvements in computer hardware continue to deliver more computational power, which, when combined with theoretical and algorithmic developments, have led to an increasing range and depth of applications of molecular modeling and molecular dynamic in biology. Molecular dynamics (MD) is a specialized computer-based discipline to simulate time evolving atomistic or molecular systems. In such simulations, motion of atoms is determined within the framework of classical mechanics.

A huge part of this thesis benefits of MD techniques and all the results obtained are constantly related with experimental data. Thus, it is crucial to establish relationships between experiments and molecular simulations, and between macroscopic and microscopic world. In the following paragraph I will explain this issues.

2.2 From Microscopic to Macroscopic: Simulations as a bridge between theory and experiments.

An experiment is usually made on a macroscopic sample that contains an extremely large number of atoms or molecules sampling an enormous number of conformations. On the other hand computer simulations generate information at the microscopic level, including atomic positions and velocities. The conversion of this microscopic information to macroscopic observables such as pressure, energy, heat capacities, etc., requires statistical mechanics.

2.2.1 Statistical Mechanics.

Statistical mechanics is fundamental to the study of biological systems by molecular dynamics simulations because it provides the rigorous mathematical expressions that relate macroscopic properties to the distribution and motion of the atoms and the molecules of the N-body system; computer simulations provide instead the means to solve the equation of motion.

Definitions: thermodynamic state and microscopic state.

The *thermodynamic state* of a system is usually defined by a set of parameters (state variables) that are physical observables, like for example, the temperature, T, the pressure, P, and the number of particles, N.

The *mechanical or microscopic state* of a system is defined by the atomic positions, **q**, and momenta, **p**; these can also be considered as coordinates in a multidimensional space called phase space (Γ). A single point in phase space, denoted by G, describes the state of the system.

Thus, the macroscopic or thermodynamic state (macrostate) described by thermodynamic quantities, is statistically interpreted in terms of all the accessible microscopic mechanical states (microstates) where each is characterized by all the N atoms positions,

$$(q_1,q_2,...,q_N) \equiv q^N$$

, and momenta,

$$(p_1,p_2,...,p_N) \equiv p^N$$

, i.e., points in the multidimensional phase space Γ.

An *ensemble* is a *collection* of points in phase space satisfying the conditions of a particular thermodynamic state. Or we can also say that it is a collection of all possible systems, which have different microscopic states but have an identical macroscopic or thermodynamic state.

Common techniques to sample the configurations assumed by a system at equilibrium are Molecular Dynamics (MD), Monte Carlo (MC) sampling and Stochastic/Brownian dynamics. A MD simulation generates a sequence of points in phase space as a function of time; these points belong to the same ensemble, and they correspond to the different conformations of the system and their respective momenta. Thermodynamic quantities are determined in computer simulations as *ensemble averages* over a large number of microscopic configurations assumed by the system under study. This is allowed due to the ergodic hypothesis assumption.

2.2.2 Calculating Averages from a Molecular Dynamics Simulation: the Ergodic hypothesis.

In statistical mechanics, averages corresponding to experimental observables are defined in terms of ensemble averages. Indeed thermodynamic observables can be alternatively modeled by considering at once a collection of identical systems. Each system represents one of all the accessible microstates. If this collection (statistical ensemble) is allowed to evolve in time, its behavior can be characterized by a time dependent distribution function, (qN(t), pN(t)) for the microstates. The instantaneous average value of the observable, O, over the phase space is interpreted as:

$$\frac{\int_{\Gamma} O(q^N(t)p^N(t))\rho(q^N(t)p^N(t))dq^N dp^N}{\int_{\Gamma} \rho(q^N(t)p^N(t))dq^N dp^N} \qquad \text{(eq.1)}$$

If we assume equal probability for all microstates, then the distribution of points in phase space is frozen into one single shape, i.e., the distribution function is time invariant, and the condition:

$$\frac{\rho(q^N(t)p^N(t))dq^N dp^N}{dt} = 0 \qquad \text{(eq.2)}$$

describes the thermodynamic equilibrium and the so-called ensemble average is defined as:

$$\langle O \rangle_{ensemble} = \frac{\int\limits_{\Gamma} O(q^N(t)p^N(t))\rho(q^N(t)p^N(t))dq^N dp^N}{\int\limits_{\Gamma} \rho(q^N(t)p^N(t))dq^N dp^N} \qquad \text{(eq.3)}$$

In a molecular dynamics simulation, the points in the ensemble are calculated sequentially in time, thus, assuming that the equations of motion of the system are solved, each observable can be empirically associated with a function (O), of the instantaneous microstate, $(q^N(t), p^N(t))$, of the system. Nevertheless, the quantity $O(q^N(t), p^N(t))$ is not an observable since measurements are performed in a macroscopic time; thus any microscopic observable is assumed to be a time averaged value:

$$\langle O \rangle_{ensemble} = \lim_{\tau \to \infty} \int_{\tau_0}^{\tau_0 + t} O(q^N(t)p^N(t))dt \qquad \text{(eq.4)}$$

The **ergodic hypothesis** states that in thermodynamic equilibrium, the **time average** and the **ensemble average** are equal.

That is, if one allows the system to evolve indefinitely in time, the system will pass through all possible microstates, and the experimental measurement will coincide with the calculated time and *ensemble averages*.

Such a procedure is well founded only for the so-called *ergodic systems*, which are assumed to fully sample the accessible phase space (hyper)volume during the observation (i.e. simulation) time. It is generally assumed that complex systems, such as the majority of those biologically relevant, are ergodic. Although this is a plausible assumption, it should be pointed out that the ergodic hypothesis is not always true (for some example see e.g. Ref. (84)).

2.2.3 Trajectory Accuracy: Shadow Orbits and the Liapunov instability.

A further drawback affecting (in principle) all kinds of MD simulations is the so-called Liapunov instability (**?**):

Two trajectories differing initially by an infinitesimal amount (Δx) will diverge exponentially in time: $\Delta x(t) = \Delta x(0) exp(\lambda t)$.

Hamiltonian systems, with very few exceptions, are chaotic.

For chaotic systems, like almost all those simulated by MD, the trajectory is extremely sensitive to initial conditions. Any error in the integration of the equation of motion, no matter how small, will always cause the simulated (numerical) trajectory to diverge exponentially from

the true trajectory starting from the same initial conditions. Thus, any imperfect integrator (and all are imperfect) introduces errors that guarantee the trajectory diverges from the true trajectory. How can we know that we are generating the correct results in MD? In some cases, there is actually a shadow orbit that closely follows the integrated trajectory. Or from other point of view, the true trajectory, to which the numerical one overlaps for a certain period of time, is called shadow orbit [1]. The shadow orbit is an exact trajectory for the system, but it starts from a slightly displaced initial point. At the time the Liapunov instability raises up, the numerical trajectory will get far from that specific shadow orbit, but there always will be another one of these to which it is superimposed.

Thus, even if the numerical trajectory diverges from the true one, as long as it conserves the total energy E within a given accuracy ΔE, the sampled configurations will belong to a constant energy hypersurface E^* for which $|E^* - E| < \Delta E$. This means that the numerical trajectory will be, within the error ΔE, representative of some true trajectory in the phase space, although it is not known a priori which one.

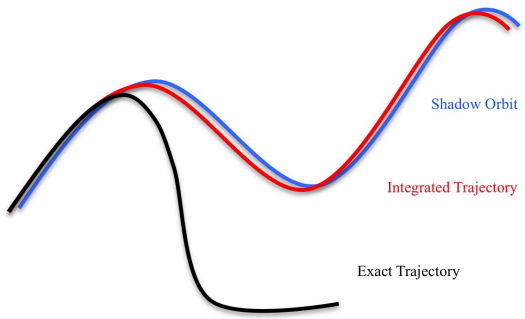


Figure 2.1: Shadow Orbit - Schematic representation of a Shadow Orbit

2.2.4 How Long? How Large?

Molecular dynamics evolves a finite-sized molecular configuration forward in time, in a step-by-step fashion. There are limits on the typical time scales and length scales that can be investigated and the consequences must be considered in analyzing the results.

2.2.4.1 Correlation Time and Correlation Length.

Simulation runs are typically short corresponding to few nanoseconds of real time, and in special cases extending to the microsecond regime. This means that we need to test whether or not a

[1] These are known to exist for hyperbolic systems. Can sometimes be shown to exist, for long times, for more general systems [e.g., see Quinlan and Tremaine, Mon. Not. R. Astron. Soc. 259, 5050 (1992)].

simulation has reached equilibrium before we can trust the averages calculated in it. Moreover, there is a clear need to subject the simulation averages to a statistical analysis, to make a realistic estimate of the errors. How long should we run? This depends on the system and the physical properties of interest. If we have an observable a, the time correlation function is $\langle a(t_0)a(t_0+t)\rangle$; assuming that the system is in equilibrium, this function is independent of the choice of time origin and may be written $\langle a(0)a(t)\rangle$. From eq.4 we can easily define a correlation time

$$\tau_a = \int_0^\tau dt \, \langle a(0)a(t)\rangle \, / \, \langle a^2 \rangle \qquad \text{(eq.5)}$$

for which the measure of a(0) and a(t) became uncorrelated. At the same way, we can define a spatial correlation function $\langle a(0)a(t)\rangle$ relating values computed at different points r apart. Spatial isotropy allows us to write this as a function of the distance between the points, r, rather than the vector r: notably this symmetry is broken in a liquid crystal. Spatial homogeneity, which applies to simple liquids (but not to solids or liquid crystals) allows us to omit any reference to an absolute origin of coordinates. This function decays from a short-range nonzero value to zero over a characteristic distance ξa, the correlation length. It is almost essential for simulation box sizes to be large compared with ξa, and for simulation run lengths τ to be large compared with τa, for all properties of interest a. Only when these two conditions are respected it is guaranteed that reliably-sampled statistical properties are obtained.

2.2.4.2 Accessible time and length scale

An important issue of simulation studies is the accessible time and length scale coverable by microscopic simulations. Figure 2.2 shows a schematic representation for different types of simulations in a length-time-diagram. It is clear that the more detailed a simulation technique operates, the smaller is the accessibility of long times and large length scales. Therefore quantum simulations, where fast motions of electrons are taken into account, are located in the lower left corner of the diagram and typical length and time scales are of order of and ps. Classical molecular dynamics approximates electronic distributions in a rather coarse-grained fashion by putting either fixed partial charges on interaction sites or by adding an approximate model for polarization effects. In both cases, the time scale of the system is not dominated by the motion of electrons, but by the time of intermolecular collision events, rotational motions or intramolecular vibrations, which are orders of magnitude slower than those of electron motions. Consequently, the time step of integration is larger and trajectory lengths are of order ns and accessible lengths of order 10100 A. If tracer particles in a solvent medium are considered, Brownian dynamics can be applied, where the effect of the solvent is hidden in average quantities. Since collision times between tracer particles is very long, larger time steps may be applied. Furthermore, since the solvent is not simulated explicitly, the length scales may be increased considerably. Finally, if one is interested not in a microscopic picture of the simulated system but in macroscopic quantities, the concepts of hydrodynamics may be applied, where the system properties are hidden in effective numbers, e.g. density, viscosity, sound velocity.

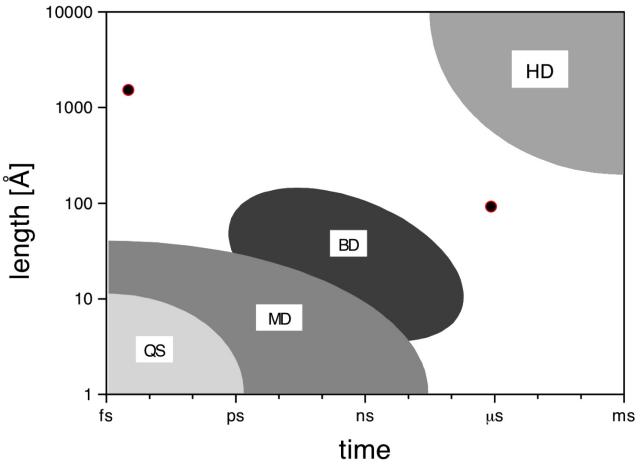


Figure 2.2: Time and Lenght scales - Schematic comparison of time and length scales, accessible to different types of simulation techniques (quantum simulations (QM), molecular dynamics (MD), Brownian dynamics (BD) and hydrodynamics/fluid dynamics (HD)).

2.2.5 Design a Molecular Dynamic Simulation in biomolecular fields.

A key decision in beginning a simulation of a biomolecular system is the choice of an appropriate method for that particular system and for the questions of interest. A modeling method should be capable of delivering a reliable result in a reasonable time(231).

Studies involving multi-nanosecond dynamics simulations are now common. However, expert knowledge is still required, and care needs to be taken to ensure that the application of a biomolecular simulation method to a particular problem is meaningful and useful.

Concerning MD simulations the ingredients are basically three:

1. A model for the interaction between system constituents (atoms, molecules, surfaces etc.) is needed.

2. An integrator is needed, which propagates particle positions and velocities from time t to t + δt. It is a finite difference scheme that moves trajectories discretely in time. The time step t has properly to be chosen to guarantee stability of the integrator, i.e. there should be no drift in the systems energy.

3. A statistical ensemble has to be chosen, where thermodynamic quantities like pressure, temperature or the number of particles are controlled. The natural choice of an ensemble in MD simulations is the microcanonical ensemble (NVE), since the systems Hamiltonian without

external potentials is a conserved quantity. Nevertheless, there are extensions to the Hamiltonian, which also allow simulating different statistical ensembles.

These choices essentially define an MD simulation.

Molecular dynamics simulation used in this thesis are perfomed with a classical description of the system under study (point 1) and the use of Gromacs simulation Package (27; 107) (point 2) in the NPT ensemble (point 3). I will define the interaction model and the ensemble chosen in the following paragraphs.

2.3 Classical Molecular Dynamics Simulations

2.3.1 Introduction

Molecular systems and motion of their constituents, both nuclei and electrons, are known to be accurately described only by laws of quantum mechanics. Implementation of classical laws of mechanics in this scheme involves a series of approximations for the quantum description:

1- The molecular wavefunction (solution) of the molecular Schrdinger equation is separated into nuclear and electronic parts according to **Born-Oppenheimer approximation**, so that motion of nuclei is decoupled from electronic motion due to the fact that nuclei are much heavier. Such decoupling allows their equations of motion to be separated and solved.

2- Nuclei are approximated as classical particles.

3- Third, electronic variables could be either integrated out beforehand and an approximate single potential energy surface (usually representing the electronic ground state) is constructed; or they could be treated within a suitable approximation as active degrees of freedom via the electronic Schrdinger equation, and forces on nuclei are computed by electronic structure calculations that are performed for each generated trajectory.

Accordingly to how the electronic part is treated, MD simulation is branched out into two methodologies:

- Classical Molecular Dynamics, where forces are derived from predefined potential models by analytical gradient applications.
- Ab initio Molecular Dynamics, where forces on nuclei are obtained from the electronic structure calculations.

I will describe only the classical molecular dynamic approach.

2.3.2 Derivation of classical molecular dynamics equations.

A molecular system with the positions of N nuclei, $R = \{R_1, R_2..., R_N\}$, and the n electrons located at $r = \{r_1, r_2..., r_n\}$, is completely described non relativistically by the molecular Schrdinger equation:

$$i\hbar\frac{\partial}{\partial t}\Phi(r, R; t) = H\Phi(r, R; t) \tag{eq.6}$$

Where H is the non-relativistic molecular Hamiltonian, in SI units, of the N nuclei and n electrons. In absence of external fields, it has the form:

$$H(R, r) = -\sum_{i=1}^{N}\frac{\hbar}{2M_I}\nabla^2_{R_I} - \sum_{i=1}^{n}\frac{\hbar}{2M_e}\nabla^2_{r_i} + \sum_{I=1,J>1}^{N,N-1}\frac{Z_i Z_j e^2}{4\pi\varepsilon_0|R_I - R_J|} + \sum_{i=1,j>1}^{n,n-1}\frac{e^2}{4\pi\varepsilon_0|r_i - r_j|} + \sum_{I=1,i=1}^{N,n}\frac{Z_I e^2}{4\pi\varepsilon_0|R_I - r_i|} \tag{eq.7}$$

Where M_I is the mass of nucleus I whose atomic number is Z_I, m_e is the mass of the electron, and e is its charge. The operator ΔR_I and Δr_i act on the coordinates nucleus I and electron i respectively. Thus we can write as:

$$H(R, r) = T_N(R) + T_e(r) + V(r, R) = T_R(R) + H_e(r, R)$$

The electron-nucleus interactions bind electrons to nuclei, and leads to a mathematical inseparable Hamiltonian.

At room temperature the thermal wavelength λ is about 0.1 A, while typical interatomic distances, in liquids and solids, are of the order of 1 A. Thus, a good approximation is to neglect quantum correlations between wave functions of different nuclei, i.e to consider the nuclear wavefunction as an incoherent superimposition of individual nuclear wave packets. In addition, nuclear masses are large enough that such individual wave packets are usually well localized. Formally, we can separate electronic from nuclear degrees of freedom by writing the wavefunction as product of terms depending only on electronic or nuclear positions ("one-determinant" Ansatz):

$$\Phi(r, R; T) \approx \Psi(r; t)\chi(R; t)\exp\left[\frac{1}{\hbar}\int_{t_0}^{t} dt' E_e(t')\right] \tag{eq.8}$$

where the electronic and nuclear wavefunctions are separately normalized:

$$\langle\Psi(r; t) \mid \Psi(r; t)\rangle = 1, \langle\chi(R; t) \mid \chi(R; t)\rangle = 1$$

and the phase factor has the form:

$$\int drdR\Psi^*(r; t)\chi^*(R; t)H_e\Psi(r; t)\chi(R; t)$$

Inserting this last expression in Schrdinger equation, multiply from left by $\langle\Psi(r;t)|$ and $\langle\chi(R;t)|$, integrate over R and r, and apply the energy conservation:

$$\frac{d}{dt}\int \Phi^*(r,R;t)H\Phi(r,R;t) = 0$$

the following system of coupled equations is obtained:

$$i\hbar\frac{\partial\Psi}{\partial t} = -\sum\frac{\hbar}{2m_e}\nabla^2_{r_i}\Psi + \left\{\int drdR\chi^*(R;t)V(r,R)\chi(R;t)\right\}\Psi \qquad \text{(eq.9)}$$

$$i\hbar\frac{\partial\chi}{\partial t} = -\sum\frac{\hbar}{2M_I}\nabla^2_{R_I}\chi + \left\{\int drdR\Psi^*(r;t)H_e\Psi(r;t)\right\}\chi \qquad \text{(eq.10)}$$

which defines the basis of the TDSCF, method introduced as early as 1930 by Dirac.

Each wavefunction above obey Schrdinger equation but with time dependent effective potential obtained by appropriate averages over the other degrees of freedom; both electrons and nuclei move quantummechanically in timedependent effective potentials (or selfconsistently obtained average fields) obtained from appropriate averages (quantum mechanical expectation values $\langle ...\rangle$) over the other class of degrees of freedom. The next step in the derivation of classical molecular dynamics is the task to **approximate the nuclei as classical point particles**. A classical description of nuclei dynamics is achieved by expressing χ in terms of an amplitude factor A and a phase S which are both considered to be real and A > 0 in this polar representation:

$$\chi(R;t) = A(R;t)\exp\left[\frac{iS(R;t)}{\hbar}\right] \qquad \text{(eq.11)}$$

Using the polar representation, in the classical limit $\hbar \to 0$, on the eq.9 and 10, the following set of equations is obtained:

$$\frac{\partial S}{\partial t} + \sum_{I=1}^{N}\frac{(\nabla_I S)^2}{2M_I} + \int dr\Psi^* H_e\Psi = 0 \qquad \text{(eq.12)}$$

$$\frac{\partial A^2}{\partial t} + \sum_{I=1}^{N}\nabla\cdot\left(A^2\frac{\nabla_I S}{M_I}\right) = 0 \qquad \text{(eq.13)}$$

The equation for A is a continuity equation for the density probability $A^2 = |\chi|^2$ of nuclei, which move with classical velocities $\nabla_I S/M_I = p_I/M_I$.

More important for our purpose is the equation eq.12 . Indeed, if we use the connection $P_I \equiv \nabla_I S$, it becomes isomorphic to the Hamilton-Jacobi equation of classical motion for action S and Hamiltonian *H(R,P)=T(P)+V(R)* defined in terms of (generalized) coordinates $\{R_I\}$ and their conjugate momenta $\{P_I\}$. The Newtonian equation of motion $\dot{P}_I = -\nabla_I V(\{R_I\})$

corresponds to eq.12:

$$\begin{aligned} &\tfrac{dP_I}{dt} = -\nabla_I \int dr \Psi^* H_e \Psi \\ &or \\ &M_I \ddot{R}_I(t) = -\nabla_I \int dr \Psi^* H_e \Psi \end{aligned} \qquad \text{(eq.14)}$$

Thus, the nuclei move according to classical mechanics in an effective potential V_e^E due to the electrons. This potential is a function of only the nuclear positions at time t as a result of averaging *He* over the electronic degrees of freedom, i.e. computing its quantum expectation value $\langle\Psi|\, H_e \,|\Psi\rangle$, while keeping the nuclear positions fixed at their instantaneous values RI(t). In other words, nuclei are driven by a mean-field potential due to electrons and containing also a contribution from their kinetic energy. Finally, to get off the nuclear wavefunction also from eq. 9 one replaces the nuclear density $|\chi(\{R_I(t)\};t)|^2$ by a product of delta functions $\prod_I \delta(R_I - R_I(t))$, i.e. incoherent wave packets extremely localized [46]. At the classical limit, the electronic wave equation is:

$$i\hbar\frac{\partial\Psi}{\partial t} = -\sum_i \frac{\hbar^2}{2m_e}\nabla_{r_i}^2\Psi + V(r,R(t))\Psi \qquad \text{(eq.15)}$$

which evolves selfconsistently as the classical nuclei are propagated via eq. 14. Note that now He and thus depend parametrically on the classical nuclear positions $\{R_I(t)\}$ at time t through $V_e^E(\{R_I(t)\})$. These equations represent the so called "Ehrenfest molecular dynamics" scheme. It is clear now that the motion of the nuclei is dictated by the Hamiltonian *He*, which basically contains the quantistic information on the electronic system. Although the TDSCF approach underlying Ehrenfest molecular dynamics clearly is a *meanfield theory*, transitions between electronic states are included in this scheme. Thus, at this stage a further simplification can be invoked by restricting the total electronic wave function Ψ to be the ground state wave function Ψ_0 of He at each instant of time. This should be a good approximation if the energy difference between Ψ_0 and the first excited state Ψ_1 is everywhere large compared to the thermal energy k_{BT}, roughly speaking.

In this limit the nuclei move on a single potential energy surface:

$$V_e^{Ehr} = \int dr \Psi_0^* H_e \Psi_0 \equiv E_0(\{R_I(t)\}) \qquad \text{(eq.16)}$$

which is computed by solving the electronic time-independent Schrdinger equation only for the ground state:

$$H_e\Psi_0 = E_0\Psi_0 \qquad \text{(eq.17)}$$

Now, E_0 is a function of nuclear positions R, and both Ehrenfest and the ground state Born-Oppenheimer potentials are identical.

As a consequence of this observation, it is conceivable to decouple the task of generating the nuclear dynamics from the task of computing the potential energy surface. Assuming the possibility to solve the stationary Schrdinger equation for as many nuclear configurations as possible, the classical molecular dynamics approach is derived by the following three steps

scheme:

1. Solving eq. 17 for many representative nuclear configurations to compute the ground state energy E_0.
2. The generated data points R, $V^{Ehr}(R)$ or some equivalent experimental data points are fitted to a suitable analytical functional form to construct a global potential energy surface.
3. The following Newtonian equation of motion:

$$M_I \ddot{R}_I(t) = -\nabla_I V_e^{Ehr}(\{R_I(t)\}) \quad \text{(eq.18)}$$

is solved by applying analytically the gradient for many different initial conditions to produce the nuclear classical trajectories on this global potential energy surface.

Furthermore the overall internal interaction potential is approximated to V^{appr} which is expanded to pair-wise, three-body, four-body and up to n-body contributions, these contributions are categorized as intermolecular long-range and intramolecular short-range interaction:

$$V^{Ehr} \approx V^{appr}(R) = \sum_{I<J}^{N} V_{IJ}(R_I R_J) + \sum_{I<J<K}^{N} V_{IJK}(R_I R_J R_K) + \sum_{I<J<K<L}^{N} V_{IJKL}(R_I R_J R_K R_L) + ... \quad \text{(eq.19)}$$

Potential expansion is practically truncated at some term to reduce the dimensionality resulting from the increase of the number of active nuclear degrees of freedom. Within the same potential expansion, electronic degrees of freedom do no longer appear explicitly but are effectively included in a functional form of V^{appr} potential.

As a result of this derivation, the essential assumptions underlying classical molecular dynamics become transparent: the electrons follow adiabatically the classical nuclear motion and can be integrated out so that the nuclei evolve on a single BornOppenheimer potential energy surface (typically but not necessarily given by the electronic ground state), which is in general approximated in terms of fewbody interactions.

2.4 Empirical Force Fields

As relevant biological processes usually involve large systems (thousands of atoms or more), and occur in relatively long timescales (from nano to microseconds or more), it is necessary to develop effective parametrized potentials, which are faster to integrate, albeit less accurate, in order to study this kind of systems. The term force field indicates a functional form for this approximation, which relates the configuration of the system ($\{R_i\}, i = 1, ..., N$) to its internal energy U, along with the set of parameters used in that function. In this work, the AMBER(188) force field for description of macromolecules in solution has been used, while the GROMACS(27; 107) package has been used to integrate the equation of motion. The functional

form can be written as:

$$U = \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\varphi - \gamma)] + \\ + \sum_{i<j} \left[4\varepsilon_{ij} \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} + \frac{q_i q_j}{\varepsilon r_{ij}} \right] \quad \text{(eq.20)}$$

Atom bond stretching and angle bending are represented as harmonic terms, while dihedrals or torsional are described by a sinusoidal term. Non-bonded interactions comprise two terms, the first is a Lennard-Jones 6-12 which describes atom-atom repulsion and dispersion interactions, the second is the Coulomb electrostatic term. In eq. eq.20, r and θ are respectively the bond length and valence angle; ϕ is the dihedral or torsion angle and r_{ij} is the distance between atoms i and j. Parameters include the bond force constant and equilibrium distance, K_r and r_{eq}, respectively; the valence angle force constant and equilibrium angle, K_θ, and θ_{eq}, respectively; the dihedral force constant, multiplicity and phase angle, V_n, n, and γ, respectively. The functional form used for out-of-plane distorsions (e.g. in planar groups) is different in different force fields. For instance, in the AMBER force field this term has the same form as that used for proper dihedrals, while in CHARMM an harmonic term is used. Collectively, these parameters represent the internal or intramolecular ones.

Non bonded parameters between atoms i and j include the partial atomic charges, qi, along with the LJ well-depth, ϵ_{ij}, and σ_{ij}, the (finite) distance at which the inter-particle potential is zero. These terms are also referred to as the interaction or external parameters. Typically, ϵ_{ij}, and σ_{ij} are obtained for individual atom types and then combined to yield ϵ_{ij}, and σ_{ij} for the interacting atoms via combining rules. The dielectric constant ϵ is typically set to 1 (corresponding to the permittivity of vacuum) in calculations that incorporate explicit solvent representations.

Van der Waals and electrostatic interactions are calculated between atoms belonging to different molecules or for atoms in the same molecules separated by at least three bonds. 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 in conjunction with periodic boundary conditions, which approximate the exact result to an acceptable error similar to the error in the van der Waals potential.

2.4.1 Long Range Interactions

In simulations of biological systems it is highly convinient to avoid the calculation of all non-bonded pair interactions, as the computational cost would be proportional to the square of the number of atoms. These interactions primarily dictates the dynamics of biomolecules, and cannot be merely truncated beyond a given cutoff when long-ranged. The difference between short and long interactions is the spatial extent of the potential. If the potential drops down

to zero faster than r^d, where r is the separation between two particles and d the dimension of the problem, it is called **short ranged**, otherwise it is **long ranged**. This becomes clear by considering the integral:

$$I = \int \frac{dr^d}{r^n} = \{\infty : n \leq d; finite : n > d$$

i.e a particles potential energy gets contributions from all particles of the universe if n d, otherwise the interaction is bound to a certain region, which is often modeled by a spherical interaction range. Long range interactions essentially require to take all particle pairs into account for a proper treatment of interactions. Coulomb ($\sim \phi_{-1}$) and dipole-dipole ($\sim \phi_{-3}$) should be considered long-range when dealing with three-dimensional systems. This may become a problem, if periodic boundary conditions are imposed to the system, i.e. formally simulating an infinite number of particles (no explicit boundaries imply infinite extent of the system). Therefore one has to devise special techniques to treat this situation.

2.4.2 Ewald Summation Method

The Ewald summation method originates from the physics of crystals, where the problem was to determine the Madelung constant, describing a factor for an effective electrostatic energy in a perfect periodic crystal. Considering the electrostatic energy of a system of particles in a cubic box and imposing periodic boundary conditions, leads to an equivalent problem. At position r_i of particle i, the electrostatic potential, ϕ, can be written down as a lattice sum:

$$\varphi(r_i) = \frac{1}{8\pi\varepsilon_0} \sum_{|n|=0}^{\infty} \sum_{j=1}^{N} \frac{q_j}{|r_{ij} + nL|}$$

thus, for N particles:

$$E = \frac{1}{8\pi\varepsilon_0} \sum_{|n|=0}^{\infty} \left[\sum_{j=1}^{N} \sum_{j=1}^{N} \frac{q_j q_i}{|r_{ij} + nL|} \right]$$

where L is the length of the periodic box , N is the total number of atoms, and n are the direct lattice vectors. i is not equal to j for $|n| = 0$. This equation is conditionally convergent, i.e. the result of the outcome depends on the order of summation. Moreover, the sum extends over infinite number of lattice vectors. Thus, the procedure has to be modified in order to get an absolute convergent sum and to get it fast converging. The original method of Ewald consisted in introducing a convergence factor e^{-ns}, which makes the sum absolute convergent; then transforming it into different fast converging terms and then putting s in the convergence factor to zero. The final result of the calculation can be easier understood from a physical picture. If every charge in the system is screened by a counter charge of opposite sign, which is smeared out, then the potential of this composite charge distribution becomes short ranged (it is similar in electrolytic solutions, where ionic charges are screened by counter charges -

the result is an exponentially decaying function, the Debye potential). In order to compensate for the added charge distribution this has to be subtracted again. The far field of a localized charge distribution is, however, again a Coulomb potential. Therefore this term will be long ranged. The efficiency gain shows up, when the short range interactions are calculated as direct particle-particle contributions in real space, while the long range part of the smeared charge cloud are summed up in reciprocal Fourier space. Choosing as the smeared charge distribution a Gaussian charge cloud of half width $1/\alpha$ the corresponding expression for the energy becomes:

$$\varphi(r_i)\sum_{n}\sum_{j=1}^{N} q_j \frac{erfc(\alpha\,|r_{ij}+nL|}{|r_{ij}+nL|} + \frac{4\pi}{L^3}\sum_{k\neq 0}\sum_{j=1}^{N}\frac{q_j}{|k|^2}e^{-|k|^2/4\alpha}e^{ikr_{ij}} - q_i\frac{2\alpha}{\sqrt{\pi}}$$

The parameter tunes the relative weights of real and reciprocal sums, although the final result is independent of it. An optimal choice for it makes the Ewald sum converge as $N^{3/2}$, which can be further improved to NlnN with the use of Particle-Mesh methods (as the Particle-Mesh Ewald, PME or the Particle-Particle Particle-Mesh, PPPM)(66; 80), making advantage of the Fast Fourier Transform. The last term corresponds to a self-energy contribution which has to be subtracted, as it is considered in the Fourier part. The new equation is an exact equivalent of the first, with the difference that it is an absolute converging expression. Therefore nothing would be gained without further approximation. Since the complimentary error function can be approximated for large arguments by a Gaussian function and the k-space parts decreases like a Gaussian, both terms can be approximated by stopping the sums at a certain lattice vector n and a maximal *kvalue*, *kmax*.

2.4.3 Boundaries

Restriction on the size of a time step is not the only challenge in molecular dynamics methods. Another concerns the finite size effects of the simulated system as its number of particles is far fewer than that in any natural sample, and is most from thousands to maximum few millions. Enclosing the system with a rigid-walled container, most particles would be under the influence of its boundaries through collisions. If we ignore the boundaries most particles would lie at surface whose area tends to be minimized, distorting thus the shape of the system whenever it is a non-spherical. It is of no help to increase particles in a system as the more particles exist, the more particles are at the surface and more undesired effects are encountered. Those peculiarity due to the size limit and the improper treatment of boundaries, makes it unreliable to statistically extract macroscopic bulk properties since the later are calculated in the limit $N \to \infty$, where N is the number of particles. To go over both practical difficulties, periodic boundary conditions are imposed on the relatively small systems in such a way that particles experience forces as if they reside in the bulk.

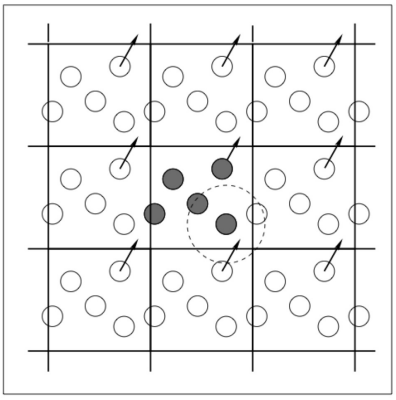


Figure 2.3: Periodic boundary conditions. - Periodic boundary conditions. As a particle moves out of the simulation box, an image particle moves in to replace it. In calculating particle interactions within the cutoff range, both real and image neighbours are included.

2.4.3.1 Periodic boundary conditions (PBC)

When applying periodic boundaries, the fundamental (primitive) simulation cell is replicated infinitely and periodically in all directions. There is no restriction on the shape of the cell other than having the characteristic to completely fill all of space translationally with no overlaps nor voids. It is appropriate to choose a cell shape that reflects the underlying geometry of the system in question. When the interactions are of a short range each side of the replicated primitive cell must be of a length that is at least twice the radius of the spatial cutoff so to keep accuracy. Particles in this case are subjected to the condition such that when a particle leaves the primitive cell, its image from the cell on the opposite side reenters the cell with the same velocity. Herein, boundaries of the cell are no longer rigid but imaginary and their effects are completely absent. When subjecting the system to this condition, the system is not any more invariant (symmetric) under space rotation; henceforth the angular momentum is no longer conserved whereas the linear momentum and mechanical energy are still conserved.

2.4.3.2 Minimum image convention for short range interactions

For short ranged forces, PBC are used in conjunction with the minimum image convention. In this scheme each particle interacts at most with only one image of every other particle n the system. To exclude interactions of a particle with its own images (self-interaction), the assumed cubic simulation cell, as already mentioned, must have a side length of at least as twice as the radius of the cutoff. Interactions terms between pairs further away from each other than the cutoff radius are obviously zero.

2.4.4 Neighbors List

For short-range potentials, not all the n-permutations represent a set of interacting particles since particles at a larger separation than a spatial cutoff radius do not interact. Nevertheless computing the non-bonded contribution to the interatomic forces in an MD simulation involves, in principle, a large number of pairwise calculations: we consider each atom i and loop over all other atoms j to calculate the minimum image separations rij. Let us assume that the interaction potentials are of short range, $\nu(r_{ij}) = 0$ if $r_{ij} > r_{cut}$, the potential cutoff. In this case, the program skips the force calculation, avoiding expensive calculations, and considers the next candidate j. Nonetheless, the time to examine all pair separations is proportional to the number of distinct pairs, 1/2 N (N 1) in an N-atom system, and for every pair one must compute at least r_{ij}^2; this still consumes a lot of time. Some economies result from the use of lists of nearby pairs of atoms. Verlet suggested such a technique for improving the speed of a program. The potential cutoff sphere, of radius rcut, around a particular atom, is surrounded by a skin of rlist radius. 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 rlist. Over the next few MD time steps, only pairs appearing in the list are checked in the force routine. Therefore, in a force routine, not all particles have to tested, whether they are in a range $r < r_{cut}$, but only those particle pairs, stored in the list. Since particles are moving during the simulation, it is necessary to update the list from time to time. List update must be at the correct frequency, a common update is between 10 to 20 time steps. To avoid double counting in the energy summation, only neighbors where (j > 1) are stored. In some cases of three/four body interactions, it is a must to only exclude equal indices, i.e., the list must contain all the pairs (j $\neq$ 1) for evaluation of three-body terms defined by the valence bond angle.

2.4.5 Constrains

It is quite common practice in classical computer simulations not to attempt to represent intramolecular bonds by terms in the potential energy function, because these bonds have very high vibration frequencies (and arguably should be treated in a quantum mechanical way rather than in the classical approximation). Instead, the bonds are treated as being constrained to have fixed length. In classical mechanics, constraints are introduced through the Lagrangian or Hamiltonian formalisms. The general principle of the formalism is the Hamiltons variational principle:

Let L be the Lagrangian of the system; Hamiltons principle states that the physical trajectory taken by that system satisfies:

$$\partial \int_{t_1}^{t_2} L dt = 0$$

for any pair of times t_1, t_2, where variations are taken with respect to q and are fixed at the endpoints. This means that a trajectory followed by a mechanic system in the phase space is the one that minimize the integral $\int L dt$ where L is define as L=K-V. Lagrangian formalism

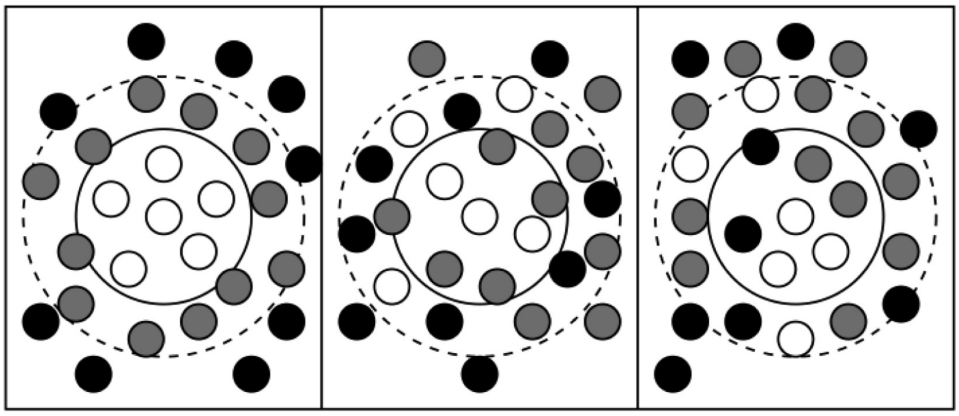


Figure 2.4: The Verlet list - The Verlet list on its construction, later, and too late. The potential cutoff range (solid circle), and the list range (dashed circle), are indicated. The list must be reconstructed before particles originally outside the list range (black) have penetrated the potential cutoff sphere.

allow to treat constrains in a simply and direct way for system in which constrains are only on the position (holonomic) and can be written as $g_l(\{r_i\}) = 0$. Lagrangian equation becomes:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{r}_i}\right) - \frac{\partial L}{\partial r_i} = \sum_{l=1}^{M} \lambda_l \frac{\partial g_l}{\partial r_i}$$

where M are Lagrange multiplier and $_l$ are function of N coordinate and velocities. This means the right equation part could be considered as a generalized force that produces the same effects of imposed constrains. It is easy to derive an exact expression for the multiplier M from the above equations; However, this exact solution is not what we want: in practice, since the equations of motion are only solved approximately, in discrete time steps, the constraints will be increasingly violated as the simulation proceeds. The breakthrough in this area came with the proposal to determine the constraint forces in such a way that the constraints are satisfied exactly at the end of each time step. For the original Verlet algorithm, this scheme is called SHAKE, which calculates the constraint forces λgi necessary to ensure that the end-of-step positions r_i satisfy: $g_l(\{r_i\}) = 0$ An alternative constraint method, LINCS (Linear Constraint Solver) was developed in 1997 by Hess, Bekker, Berendsen and Fraaije(106)LINCS applies Lagrange multipliers to the constraint forces and solves for the multipliers by using a series expansion to approximate the inverse of the Jacobian. This approximation only works for matrices with Eigenvalues smaller than 1, making the LINCS algorithm suitable only for molecules with low connectivity. It is important to realize that a simulation of a system with rigidly constrained bond lengths, is not equivalent to a simulation with, for example, harmonic springs representing the bonds, even in the limit of very strong springs. A subtle, but crucial,

difference lies in the distribution function for the other coordinates. If we obtain the configurational distribution function by integrating over the momenta, the difference arises because in one case a set of momenta is set to zero, and not integrated, while in the other integration is performed, which may lead to an extra term depending on particle coordinates. This is frequently called the metric tensor problem.

2.4.6 MD in NPT Ensemble

The ergodic hypothesis was introduced for a system described by the microcanonical distribution *NVE*. However, the conditions of constant volume V, number of particles N and total energy E do not fit those in which experiments are usually made. Thus, it is necessary to define schemes allowing for the evolution of systems under conditions of constant volume and temperature (NVT), or constant pressure and temperature (NPT), corresponding to typical real-life situations. In this work, the simulations were performed in the NPT ensemble. To simulate the systems in such ensemble, thermostat and barostat algorithms are required to control the temperature and pressure during the MD run.

2.4.7 Nose-Hoover thermostat

A way to sample the NVT ensemble within the framework of MD was introduced about twenty years ago by Nose(180; 181) and reformulated by Hoover. This method modifies Newton equation of motion by adding two non physical variables, thus introducing the following non-Hamiltonian dynamical system where there is added a fictitious degree of freedom, with mass equal to Q. The new extended Hamiltonian has the form:

$$H* = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + \varphi(r_i) + \frac{Q}{2}\zeta^2 + gkT \ln S$$

where $\{r_i\}$,$\{p_i\}$ are coordinates and momenta of the N particles with masses m_i, and S and ζ are coordinates and momenta of fictitious atoms. If ϕ is the interaction potential and g are the degree of freedom, the new equation of motion are:

$$\begin{aligned} \dot{r}_i &= \tfrac{p_i}{m_i} \\ \dot{p}_i &= -\tfrac{d\varphi}{dt} - \zeta p_i \\ \tfrac{d\zeta}{dt} &= \tfrac{\sum \frac{p_i^2}{m_i} - gk_BT}{Q} \end{aligned}$$

These equations sample a microcanonical ensemble in the extended system, however, the energy of the real system is not constant. Nevertheless it can be shown, that the equations of motion sample a canonical ensemble in the real system. The parameter Q controls the strength of the coupling to the thermostat: high values result into a low coupling and viceversa. Although any finite (positive) mass is sufficient to guarantee in principle the generation of a canonical ensemble, if Q is too large, the canonical distribution will only be obtained after very long

simulation times. On the other hand, too small values (tight coupling) may cause high-frequency temperature oscillations.

2.4.8 Berendsen thermostat

A weaker formulation of this approach is the Berendsen thermostat.(25; 26) To maintain the temperature the system is coupled to an external heat bath with fixed temperature T0. The velocities are scaled at each step, such that the rate of change of temperature is proportional to the difference in temperature:

$$\frac{dT(t)}{dt} = \frac{1}{\tau}(T_0 - T(t))$$

where τ is the coupling parameter which determines how tightly the bath and the system are coupled together. This method gives an exponential decay of the system towards the desired temperature. The change in temperature between successive time steps is:

$$\Delta T = \frac{\delta t}{\tau}(T_0 - T(t))$$

Thus, the scaling factor for the velocities is:

$$\lambda^2 = 1 + \frac{\delta}{\tau}\left\{\frac{T_0}{T(t - \delta t/2}\right\}$$

In practice τ is used as an empirical parameter to adjust the strength of the coupling. Its value has to be chosen with care. In the limit $\tau \to \infty$ the Berendsen thermostat is inactive and the run is sampling a microcanonical ensemble. The temperature fluctuations will grow until they reach the appropriate value of a microcanonical ensemble. However, they will never reach the appropriate value for a canonical ensemble. On the other hand, too small values of τ will cause unrealistically low temperature fluctuations. If τ is chosen the same as the time step t, the Berendsen thermostat is nothing else than the simple velocity scaling. Values of $\tau \approx 0.$ 1ps are typically used in MD simulations of condensed-phase systems. The ensemble generated when using the Berendsen thermostat is not a canonical ensemble.

The Andersen method(10) was developed to adjust the pressure in a simulation of interacting particles. In the following description, only systems of pairwise inter- acting particles are treated. The method was later first extended to anisotropic coupling by Parrinello et al.(188) and later also to molecular systems by Nos et al.(180; 181) Andersen proposed to replace the coordinates r_i by scaled coordinates ρ_i defined as:

$$\rho_i = r_i/V^{1/3}$$

Consider the new Lagrangian, in which a new variable Q appears:

$$L\left(\rho^N, \dot{\rho}^N, Q, \dot{Q}\right) = \frac{1}{2}Q^{2/3}\sum_{i=1}^{N} m_i\dot{\rho}_i^2 - \sum_{i<j}^{N} U(Q^{1/3}\rho_{ij}) + \frac{1}{2}M\dot{Q}^2 - p_0 Q$$

If we interpret Q as the volume V, the first two terms on the right are just the Lagrangian of the unscaled system. The third term is a kinetic energy for the motion of Q, and the fourth represent a potential energy associated with Q. Here p_0 and M are constants. A physical interpretation of the additional terms would be: Assume the system is simulated in a container and can be compressed by a piston. Thus, Q, whose value is the volume V, is the coordinate of the piston. p_0 V is the potential derived from an external pressure p_0 acting on the piston and M is the mass of the piston.

2.4.8.1 Parrinello-Rahman barostat

When simulating crystal structures, it is not sufficient only to scale the volume. Parinello and Rahman extended the method proposed by Andersen to let the simulation box also change its shape.(188) Lets start with some notation: The cell can have an arbitrary shape, its volume completely described by three vectors *a,b,c*. The vectors can have different lengths and arbitrary mutual orientations. An alternative description is obtained by arranging the vectors as $\{a, b, c\}$ to form a *3 x 3* matrix h whose columns are the latter vectors. The volume is given by:

$$V = \det h = a \cdot (b \times C)$$

The position r_i of a particle can be written in terms of h and a column of vector s_i, with components ξ_i, η_i and ζ_i as:

$$r_i = hs_i = \xi_i a + \eta_i b + \zeta_i c$$

with $0 \leq \xi_i, \eta_i, \zeta_i \leq 1$. The square of the distance between particle i and j is given by:

$$r_{ij}^2 = s_{ij}^T G s_{ij}$$

where the metric tensor G is $G = h^T h$. Using the latter notation, the Lagrangian can be written as:

$$L = \frac{1}{2}\sum m_i \dot{s}_i^T G \dot{s}_i - \sum\sum U(r_{ij}) + \frac{1}{2}MTr(\dot{h}^T\dot{h}) = pV$$

Deriving the equations of motion is similar to the isotropic case from Andersen.

2.5 Comments on other techniques used.

In the project presented in Chapter 6, part of the calculations was carried out by other members of the biomolecular simulation group of SISSA. These are the metadynamics free-energy calculations which have been performed by Xevi Biarnes and molecular docking which have

been performed by Agata Kranjc and Salvatore Bongarzone. For this reason the theory of free-energy calculations and docking is not presented here but in the computational section and in the Appendix of the corresponding chapter.

3

Prion diseases

3.1 The prion-only hypothesis

Prion diseases, also termed transmissible spongiform encephalopathies (TSEs), are fatal neurodegenerative diseases that affect humans and a wide variety of animals. Human forms of prion diseases include Creutzfeld-Jacob disease (CJD), Gerstamnn-Straussler-Scheinker disease (GSS), fatal familial insomnia (FFI) and kuru (see Appendix A), whereas the animal forms include scrapie (Sc) in sheep and goat, chronic wasting disorder (CWD) in mule, deer and elk, and bovine spongiform encephalopathy (BSE) in cattle (59; 199; 243).

Prion diseases are generally characterized by widespread neurodegeneration and therefore, clinically exhibit symptoms of cognitive and motor dysfunction while, pathologically, TSEs are characterized by neuronal vacuolation, astrocytic gliosis, and amyloid plaques.

Prion diseases may be sporadic (arise spontaneously without any apparent cause), inherited (due to mutations in the gene encoding human prion protein, PRNP) and/or acquired by the transmission of an infectious agent (199): despite of the common morphological and pathophysiological features with other progressive encephalopathies, such as Alzheimers and Parkinsons disorder (6), prion diseases are unique in that they are transmissible. [1]. Hence, the most intriguing aspect of prion diseases is the nature of the infectious pathogen, also referred to as the scrapie agent. This was termed prion (a small **pro**teinaceous **in**fectious particle that is resistant to inactivation by most procedures that modify nucleic acids (197)) by Stanley B. Prusiner to distinguish the infectious pathogen that causes TSEs from viruses and viroids. Naturally, all infectious agents are *proteinaceous* to some degree. What sets prions apart, as proposed by Prusiner, is that the actual infectious principle consists merely of protein and is

[1] Aggregates of amyloid-β (Aβ) peptide associated with Alzheimers disease behave like an infectious agent when injected into the brain of a mouse model of Alzheimers disease, showing a pattern of Aβ deposition that depends on both the host and the agent (168). These findings suggest that what was considered to be a unique feature of prion disease may be a more general property of amyloids

capable of replicating and transmitting infections without the need for informational nucleic acids (147; 197).

The *protein-only hypothesis* [1] states that the agent responsible for the propagation of prion diseases is not a virus but a post-translationally modified form of the cellular prion protein, PrP^C (C=Cellular) [2] denoted as PrP^{Sc} (Sc=Scrapie). Prusiner and co-workers proposed that PrP^{Sc} is the main or only constituent of the prion. Because the sequence of PrP^C is the same as that of PrP^{Sc}, the two species are believed to differ in their conformation (218).

Furthermore, the *protein only hypothesis* states that, when introduced in a normal host, PrP^{Sc} induces the conversion of PrP^C into PrP^{Sc}, and that this conversion is the main event in the propagation of prion diseases. Although still controversial (42), the notion that transmissible spongiform encephalopathies can be propagated by a mechanism involving self-perpetuating changes in protein conformation is supported by a wealth of infection studies (59; 199; 243).

3.2 Toxicity

PrP^C needs to be presented by host neurons for neurodegeneration to occur (4).

Consistently, mice lacking the prion gene are resistant to the disease (37). Indeed, when neurografts propagating PrP^{Sc} were implanted into Prnp knockout mice, no pathological changes were seen in PrP-deficient tissue, even in the immediate vicinity of the grafts (32).

Even if the formation of PrP^{Sc} accompanies neurodegeneration in prion disease, many lines of evidence indicate that *PrP^{Sc} is not intrinsically neurotoxic (4).*

The importance of neuronal expression of PrP^C for prion disease development has been corroborated also by the phenotype of mice with neuron-specific ablation of PrP^C eight weeks

[1] First outlined in general terms by Griffith (96) and enunciated in its updated and detailed form by Prusiner (200)

[2] PrP^C is synthesized in the endoplasmatic reticulum, processed in the Golgi apparatus and then carried, in its mature form, to the cell surface (199). PrP^C is classified as GPI-anchored plasma-membrane glycoprotein (22; 49; 183) and it is found predominantly on the outer surface of neuronal and glial cells. Its physiological function is still unclear; Glycosylphosphatidylinositol (GPI anchor) is a glycolipid that can be attached to the C-terminus of a protein during post-translational modification. It is composed of a phosphatidylinositol group linked through a carbohydrate-containing linker to the C-terminal amino acid of a mature protein. The two fatty acids within the hydrophobic phosphatidylinositol group anchor the protein to the cell membrane. Glypiated (GPI-linked) proteins contain a signal peptide, thus directing them into the endoplasmic reticulum (ER). The C-terminus is composed of hydrophobic amino acids that stay inserted in the ER membrane. The hydrophobic end is then cleaved off and replaced by the GPI-anchor. As the protein processes through the secretory pathway, it is transferred via vesicles to the Golgi apparatus and finally to the extracellular space where it remains attached to the exterior leaflet of the cell membrane. Since the glypiation is the sole means of attachment of such proteins to the membrane, cleavage of the group by phospholipases will result in controlled release of the protein from the membrane

after prion inoculation. Early spongiform changes were reversed, and clinical disease was prevented. This reversal occurred despite the accumulation of extraneuronal PrP^{Sc} (157).

Membrane attachment of PrP^{C} is a prerequisite to initiate neurodegeneration and that the presence of PrP^{Sc} alone does not cause disease (4).

Consistently, transgenic mice expressing only a secreted form of PrP^{C}, lacking its membrane attachment via GPI anchor, have been reported to be refractory to develop clinical signs of prion diseases, although prion inoculation induces PrP^{Sc} formation and aggregation of amyloid plaques (50).

3.3 Structural and Physical Properties

The two PrP isoforms have remarkably different biophysical properties.

PrP^{C} is readily degradable by proteinase K, while PrP^{Sc} is only partially degraded upon proteolytic digestion (162; 183). PrP^{C} is an α-helical, soluble monomeric protein, whereas PrP^{Sc} has a βsheet rich structure and forms large aggregates which are insoluble in non-denaturing detergents and often appear as amyloid fibrils (167; 184; 187).

Human PrP^{C} is a 209 residues protein 3.1, consisting of a largely unordered N-terminal region (residues 23-124) and a folded C-terminal domain (residues 125-231) (247). The protein is highly conserved across mammals (187; 203), indeed its NMR structures expressed by different mammals are similar (54; 55; 92; 153; 247).

The C-terminal globular domain is from residue 125 to 228 in the human numbering. It contains three α-helices comprising the residues 144-154, 173-194 and 200-228 and a very short anti-parallel β-sheets comprising residues 128-131 and 161-164. A single disulfide bond, Cys179-Cys214, connects helices 2 and 3. NMR studies have revealed that the structure of the C-terminal domain is preserved even in the absence of the N-terminal part of the protein (187; 203; 247). The N-terminal region contains characteristic glycine-rich octapeptide repeats (PHGGGWGQ human sequence) which enables the protein to coordinate Cu(II) ions (169; 235). It is unlikely that the N-terminus is randomly coiled also in vivo, since functional studies in transgenic mice imply that the domain comprising amino acids 32-121 carries out important physiological functions (212). The flexible tail of PrP^{C} might acquire a defined structure when PrP^{C} is present within membrane rafts (172).

The transition between human PrP^{C} and PrP^{Sc}, proposed within the *protein only hypothesis*, occurs post translationally without any detectable covalent modifications (218). The molecular basis of the $PrP^{C} \rightarrow PrP^{Sc}$ conversion is currently far from being clear and several models have been proposed (see Appendix B). The earliest event in the conversion involves the formation of a stable intermediate (112; 249), which has not been characterized yet. Nevertheless, the existence of such partially structured monomeric folding intermediate of the prion protein is also indicated by hydrogen-deuterium exchange (139) and high-pressure spectroscopy experiments (176).

PrP^{Sc} is believed to aggregate on a smaller scale and then progress into protofibrils before

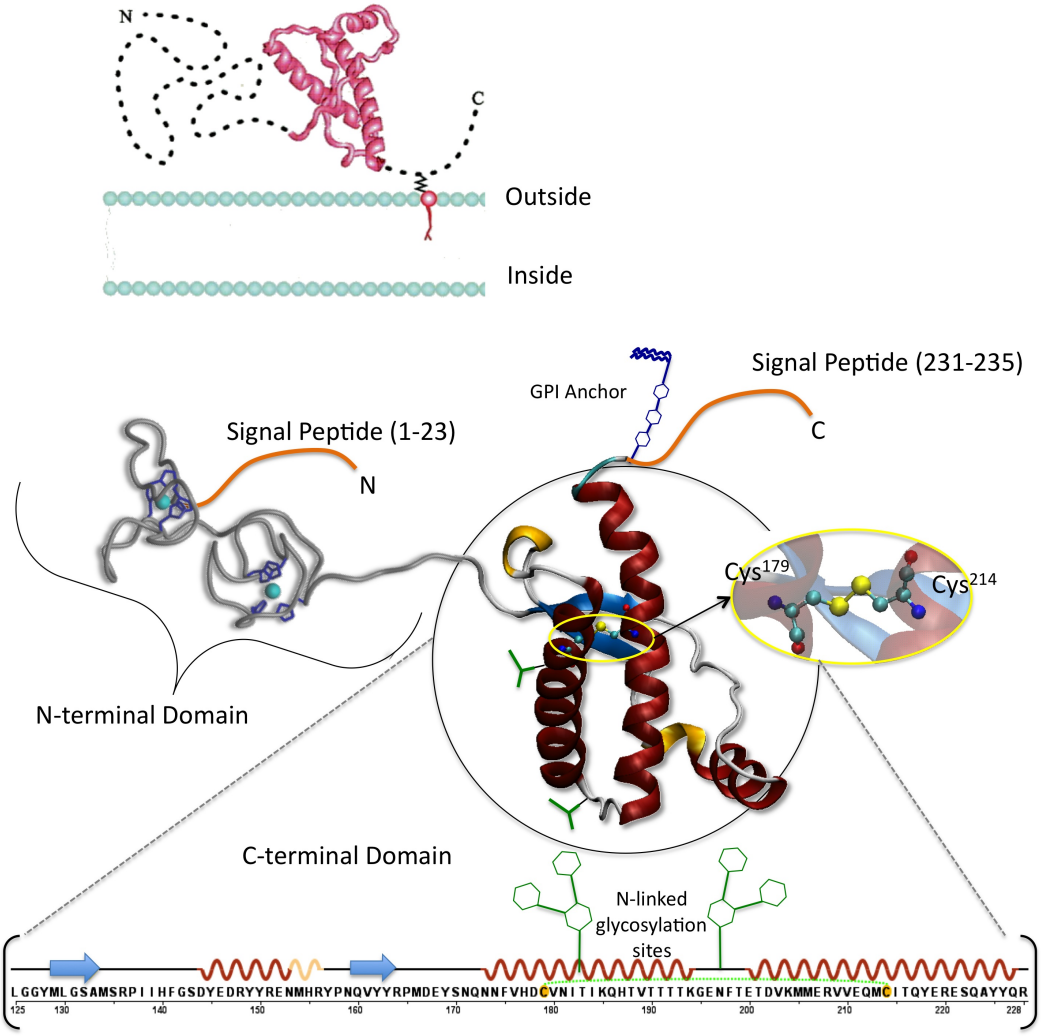


Figure 3.1: The Human Prion Protein - The mature human prion protein. It features two signal peptides (1-23 and 231-235), an octapeptide repeat region (involved in binding Cu2+ ions (169)), three α-helixes, one disulphide bond (SS) between cysteine residues 179 and 214, and two potential sites for N-linked glycosylation (green forks) at residues 181 and 197. A glycosylphosphatidylinositol anchor (GPI) (yellow box) is attached to the C-terminus of PrP.

forming ordered rigid fibrils (5). No atomic-resolution information of any of these structures have been obtained so far (5; 7), due to the insoluble nature of the aggregates (173; 214; 219). Fourier-transform infrared (187) and CD spectroscopy (207) reveal a higher β-sheet content of PrP^{Sc} respect PrP^{C}. Moreover, the discovery of 2D crystals of PrP^{Sc} and of a fibrillogenic miniprion, PrP^{Sc} (with residues 141176 deleted)(244), provided additional structural constraints for modeling PrP^{Sc} (244). According to this, different model describing the mammalian prion structure have been suggested that can be clustered in two groups whose representative structures are: the spiral model (72) and the β-helix model (93).

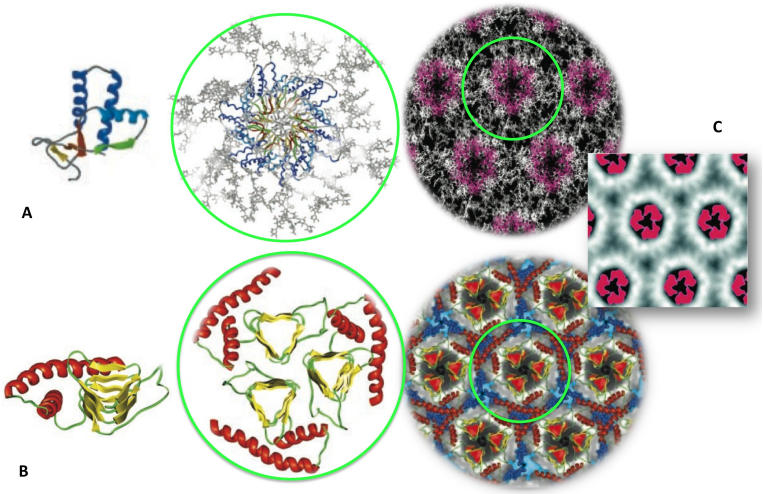


Figure 3.2: Scrapie Models - A) the spiral model (72) and B) the β-helix model (93) on the basis of EM image of the two-dimensional protofibril by Wille et al. (244)(C)

To systematically evaluate the relationship between infectivity, converting activity, and the size of various PrP^{Sc}-containing aggregates, Caughey and co-workers (213) partially disaggregated PrP^{Sc} aggegates. The resulting species were fractionated by size. Intracerebral inoculation of the different fractions into hamsters revealed that nonfibrillar particles, with masses equivalent to 1428 PrP molecules, are the most efficient initiators of TSE disease (213).

Thus, as with other diseases characterized by protein aggregation, such as Alzheimers disease and Parkinsons (16), the formation of large amyloid fibrils might be a protective process that sequesters the more dangerous subfibrillar oligomers of the amyloidogenic peptide or protein into relatively innocuous deposits (43; 126). Fibrils in TSEs, known as prion rods, are composed of PrP 2730, a fragment of PrP^{Sc} resulting from proteinase K digestion of the first $\sim$ 88 residues. Prion rods possess the tinctorial properties of amyloid fibers (162; 198) and resemble amyloid fibrils found in vivo (70).

3.4 Pathogenic Mutations

3.4.1 Disease-linked Mutations accelerate PrP^C to PrP^{Sc} conversion

One of the key arguments in support of the protein-only hypothesis is the evidence linking familial prion diseases (about 15% of human prion diseases (135; 164)) with mutations in the gene coding for human prion protein (59; 199; 243).

Indeed in individuals carrying these mutations the conversion of PrP^C to PrP^{Sc} appears to develop spontaneously (223).

Hence, the primary role of these mutations might be increase the likelihood of misfolding by the thermodynamic destabilization of PrP^C (150; 223) in favor of a folding intermediate that is likely to be a PrP^{Sc} precursor (13; 220). Kinetic studies using both stopped-flow and continuous-flow methods indicate that a number of PrP variants carrying mutations corresponding to inherited prion diseases results in a destabilizing effect of the PrP^C conformer toward a pronounced stabilization and thus increased population of the folding intermediate. For each mutant protein tested, the population of an intermediate was found to be at least 1 order of magnitude higher than that of the fully unfolded state (220). This effect was observed even for mutations with only a minor effect on the global stability of the native state of prion protein (12; 13).

3.4.2 Pathology

Despite the rarity of PRNP mutations, the protracted clinical course of some inherited prion diseases leads to a high population prevalence relative to sporadic and acquired prion diseases (164).

Over 30 different mutations have been described: some are typically associated with particular clinical categories of prion disease; others are associated with a spectrum of clinical phenotypes, often with striking phenotypic variability (164) (238).

Historically, variation in clinical phenotype of inherited prion diseases has been encapsulated by three clinical categories: GSS (Gerstmann, 1936, reviewed by (102)), FFI (155), and familial Creutzfeldt-Jakob disease (fCJD) (Jakob, 1923, described by (34)). The clinical categories of inherited prion diseases may be seen as extremes of phenotype, in reality the syndromes overlap considerably (164).

The existence of phenotypic overlap between individuals with different mutations and even in family members with the same PRNP mutation indicates that accurate classification of inherited human prion diseases should be based upon mutation alone (135).

National Prion Disease Pathology Surveillance Center
Cases Examined[1]
(July 31, 2010)

Year	Total Referrals[2]	Prion Disease	Sporadic	Familial	Iatrogenic	vCJD
1996 & earlier	51	33	28	5	0	0
1997	114	68	59	9	0	0
1998	88	52	44	7	1	0
1999	120	72	64	8	0	0
2000	146	103	89	14	0	0
2001	209	119	109	10	0	0
2002	248	149	125	22	2	0
2003	274	176	137	39	0	0
2004	325	186	164	21	0	1[3]
2005	344	194	157	36	1	0
2006	383	197	166	29	0	2[4]
2007	377	214	187	27	0	0
2008	394	231	204	25	0	0
2009	425	259	216	43	0	0
2010	204	124	85	20	0	0
TOTAL	**3702**[5]	**2177**[6]	**1834**	**315**	**4**	**3**

[1] Listed based on the year of death or, if not available, on year of referral; [2] Cases with suspected prion disease for which brain tissue and/or blood (in familial cases) were submitted; [3] Disease acquired in the United Kingdom; [4] Disease was acquired in the United Kingdom in one case and in Saudi Arabia in the other case; [5] Includes 16 cases in which the diagnosis is pending, and 18 inconclusive cases; [6] Includes 21 (19 from 2010) cases with type determination pending in which the diagnosis of vCJD has been excluded.

Rev 8/1/2010

Figure 3.3: Case of TSEs - Case of TSEs examined Until July 31, 2010 available on the National Prion Disease Pathology Surveillance Centre (http://www.cjdsurveillance.com/) Division of Neuropathology of Case Western Reserve University)

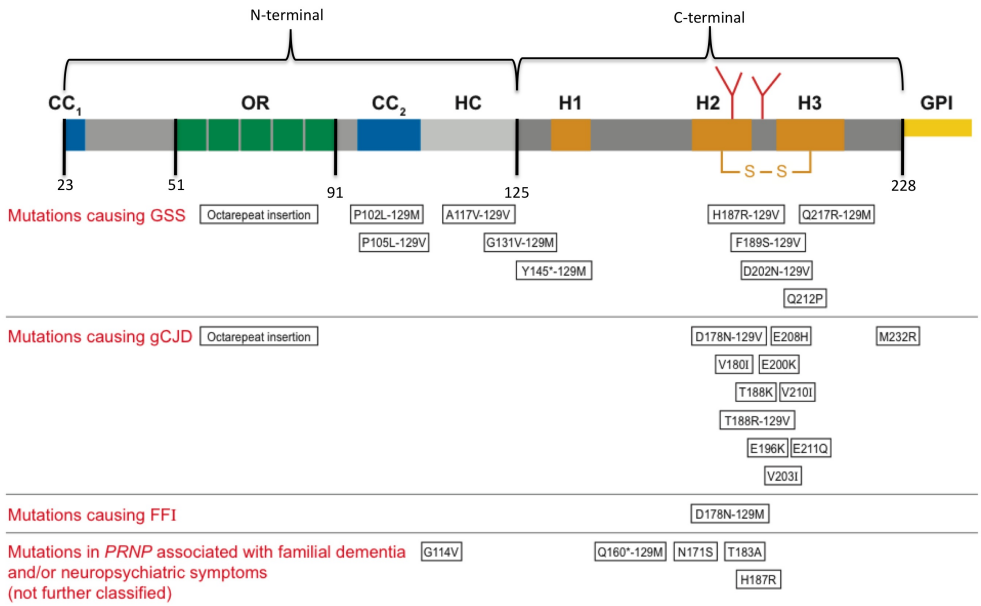


Figure 3.4: Mutations - (revisited from (4)) The human PrP^C protein and its mutants. The mature human PrP^C protein contains 209 amino acid residues. It features two positively charged amino acid clusters denoted CC1 and CC2 (blue boxes), an octapeptide repeat region (OR) (green boxes), a hydrophobic core (HC) (gray box), three α-helixes (H1-H3) (red boxes), one disulphide bond (SS) between cysteine residues 179 and 214, and two potential sites for N-linked glycosylation (red forks) at residues 181 and 197. A glycosylphosphatidylinositol anchor (GPI) (yellow box) is attached to the C-terminus of PrP.

3.4.3 Location of the mutated residues in PrP^C structure

A key question for pharmacological intervention against pathogenic variants is how the identified mutations in PRPN gene affect the conversion of PrP^C to PrP^{Sc} and aggregation. In particular, structural studies on PrP variants containing familial mutations might provide important clues regarding the molecular basis of the disease. On the basis of their occurrence in the sequence, the reported pathogenic mutations can be classified in three different groups (230) [1]:

- Mutations in the flexible N-terminal region
- Mutations in the globular domain [2]
- Mutations in the GPI-signal peptide

Mutations in the flexible N-terminal region

It is likely that the flexible N-terminus is involved in the conversion from PrP^C to PrP^{Sc} (71). In particular, residues from 112 to119 are required for the formation and propagation of PrP^{Sc} (179). The importance of the flexible N-terminus in causing disease is supported also by the existence of a short neurotoxic peptide (residues 106-126), which has a high intrinsic ability to polymerize into amyloid-like fibrils in vitro (34; 83).

There are four residues linked to GSS causing mutations in the flexible N-terminus of human PrP: P102L, P105L/S/T, G114V and A117V. However, there is little evidence for a direct molecular effect of these mutations in the flexible N-terminus on the globular structure of PrP^C or conversion to PrP^{Sc}, also because most of the available studies on these mutants have focused on small fragments of the PrP protein (230). It is possible that β-sheet formation in the N-terminal region is enhanced, but it could also be that the disease-causing ability of these mutants arises due to more complex interactions at the cellular level (230). Jones et al. (120) compared the fibril formation of the 23 144 fragment with the mutations P102L, P105L and A117V to the WT. No significant effects on the amyloidogenicity were found, confirming the results obtained by neurotoxicity studies with small peptides (34; 83).

Mutations in the globular domain

Because of structural information of this domain is available (247), here we are able to discuss the type of interactions the mutated residues form in the protein. On the basis of the type of amino acid replacements involved we can divide mutations in (230):

- Stop Mutations
- Mutations affecting salt-bridge interactions in the globular domain
- Mutations of polar residues in the globular domain

[1] Insertions or deletions in the octarepeat region of the PRPN gene have also been linked to prion disease in humans (24; 91) but they will be not discussed here.

[2] For a number of mutations (S132I, A133V, T193I, E196K), no information is available apart from their occurrence in cases of inherited spongiform encephalopathies.

- Hydrophobic mutations in the globular domain

Stop Mutations. Two stop mutations have been observed in patients that result in a truncated version of PrP: Y145stop and Q160stop, located in α1 and β2 respectively. Further research has mostly been performed on the Y145stop mutant. The largely missing C-terminal globular domain in this mutant is normally required to import the complete PrP into the Endoplasmatic Reticulum (105). Without this happening, a large portion of the truncated PrP is not post-translationally processed (keeping its N-terminal signal peptide) and rapid degradation by the proteasome occurs (77; 248). It was found that a recombinant fragment containing residues 23-144 can convert spontaneously to a fibrillar form, with residues 138-141 being essential for its formation (137).

Mutations affecting salt-bridge interactions in the globular domain. Salt bridges are often involved in stabilization of the tertiary structure of a protein. Six known disease-causing mutations can directly affect salt-bridge interactions in the globular PrP domain: R148H, D178N, E196K, E200K, R208H and E211Q.

R148 is located in helix 1 and can form salt bridges with D144 and E146, thereby his substitution could alter the pattern of salt bridges along helix α1. This network of interactions was shown to confer stability of α1, both experimentally (166; 217) and theoretically (74; 208). Apart from the similarity of R148H PrP^{Sc} aggregates to those observed in a sporadic form of CJD (189), no information on the effect of this mutation is available. The same is for E211Q and E196K.

The D178N mutant, located in helix 2, is probably the most extensively studied PrP mutation and is involved in a large number of disease cases (164). Intriguingly, the mutation causes FFI in combination with M129 and fCJD in combination with V129 (90), although this distinction may not be so clear-cut (163). D178 residues is involved in the salt bridge with R164 (located on β2) and can also be involved in hydrogen bond interactions with the Y128 and Y169 side chains (located at the beginning of strand β1 and the loop between β2 and α2, respectively). Protonation of D178 in MD studies leads to the loss of these interactions, which may facilitate conversion (8). This hydrogen bond is also lost in most of the mutant we studied allowing a more flexibility of the loop and a major exposure of the Y169 toward the surface of the protein.

E200K is the major causative mutation related to fCJD (164), and its phenotype is influenced by the M/V129 polymorphism (201). E200 is located at the beginning of helix 3, thus it was argued that the mutation may affect the stability of helix α3 (86). Either if it does not seem to significantly affect the overall thermodynamic stability of PrP^{C} (223), however this mutations change the surface electrostatic potential of the protein (252) causing abnormalities in interactions of PrP with other proteins in the cell or the cell membrane itself (209).

For R208, the situation is less clear; Zuegg and Gready (1999) (253) argued that a salt bridge between E146 and R208 could stabilize the tertiary structure of PrP^{C} and its loss may facilitate conversion to PrP^{Sc}. MD simulations of WT PrP (residues 125 -228), R208H PrP and WT PrP with neutralized D144 and E146 (to break the putative salt-bridge interaction),

however, showed no significant rearrangements upon breaking the interaction, whereas the introduction of the R208H mutation did (19).

Mutations of polar residues in the globular domain. Three of the polar residue mutations are found in helix 2: T183A, H187R, T188R/K/A.

Mutation of T183 to A disrupts hydrogen bond interaction between the T183 side chain and the backbone nitrogen of Y162, which can help to anchor the native sheet to the core of $\alpha 2$ and $\alpha 3$, preserving the stability of the folds . This mutant indeed affects both in vivo folding and GPI-anchor attachment (**?**) and can cause abnormal glycosylation (95). Accordingly, mouse T183A PrP 121 231 expressed in E. coli aggregates and shows a significantly decreased stability compared with WT (150).

The H187R mutation may disrupt a putative interaction between helix 2 and helix 3 (142). It has been related to both fCJD (98) and GSS disease with abnormal PrP^{Sc} deposits (63) since H187R PrP accumulates in lysosomes instead of moving to the cell surface (98).

Directly adjacent to H187 is T188, which can mutate to R, K or A, causing fCJD. These mutant PrPs show an increased proteinase K resistance and detergent insolubility that can be indicative of a (higher) degree of aggregation (230).

Three other pathogenic mutations of polar residues are located in helix 3: D202N, Q212P, Q217R.

The first is D202, which forms a (conserved) stabilizing capping box with T199 at the top of the helix (204). Combined CD and NMR study of peptides corresponding to helix 3 showed that the D202N mutation completely destabilizes its structure (86). Accordingly, in a human cell model, D202N PrP accumulates and aggregates in the Endoplasmatic Reticulum without reaching a mature conformation (99).

The mutation of Q212P is remarkable as it inserts a proline in the middle of helix 3, which affects its structure (115). The only effect of this mutation currently known is that Q212P PrP accumulates and eventually forms aggresomes in the cytosol upon proteasomal inhibition (170).

The final mutation in helix 3 is Q217R. It lacks a GPI-anchor, which probably causes the observed impaired transport to the cell surface, and it exhibits increased aggregation and proteinase resistance (215). Most of these alterations are temperature-sensitive, indicating that they are due to misfolding of PrP (215). Indeed MD simulation studies have shown that Q212 is involved in a tight hydration site: a water molecule can bind between the backbone carbonyl of S132 (in $\beta 1$), the backbone amide of V161 (in $\beta 2$) and the side chain oxygen of Q217 (69). When Q217 is mutated in R, the large side chain of the arginine replaced the tightly bound water and forced other rearrangements, which may facilitate fibril formation (69).

Hydrophobic mutations in the globular domain. The remaining pathogenic mutations in the globular domain of PrP affect hydrophobic residues. V180I, F198S and V210I have been indicated to cause atypical glycosylation, to interference with GPI-anchor attachment (46**?**) and to affect the folding pathway of PrP (13). Further studies on F198S mutation reveal that it leaves a gap in the hydrophobic core (204) and as expected, the thermodynamic stability of PrP^{C} is significantly affected as in vitro (150) and in silico (251).

Mutations in the GPI-signal peptide

There are three mutations in the GPI-signal peptide associated with fCJD: M232R, M232T and P238S (230).

The mechanism of cytotoxicity of these mutants remains unknown as the signal peptide is cleaved off within 5 min of PrP synthesis and translocation into the Endoplasmatic Reticulum (152). Recently, Gu et al. (2008) (97) showed that M232R and M232T do not interfere with GPI-anchor addition, but do cause PrP being bound to the membrane in an alternative (C-transmembrane) orientation (15).

4

Structural facets of disease-linked human prion protein mutants: a molecular dynamic study

Prion propagation in transmissible spongiform encephalopathies involves the conversion of the cellular prion protein (PrP), PrP^{C}, into the pathogenic conformer PrP^{Sc}. Human familial forms of the disease are linked to specific mutations in the PrP gene, PRNP, and include Gerstmann-Strussler-Scheinker syndrome (GSS), familial Creutzfeldt-Jakob disease (fCJD) and Fatal Familial Insomnia. To gain insights into the molecular basis of these disorders, we performed classical molecular dynamic simulations in aqueous solution on wild type (WT) human PrP (HuPrP), and on three HuPrP variants located in the globular HuPrP domain: two pathological mutations, HuPrP(Q212P) and HuPrP(E200K), linked to GSS and to fCJD respectively, and one protective polymorphism, HuPrP(E219K). A comparison between the predicted structural determinants of WT HuPrP and one of its mutant (HuPrP(E200K)) with the available structures established the accuracy of the methods used. Strikingly, the analyzed disease-linked variants had their major effect on the α_2-α_3 region and the β_2-α_2 loop, regardless of the mutation position. The conformational change of the latter might affect the interactions with cellular partners in the fibrillation process. The protocol proposed here represents a powerful approach for reproducing the structural effects of genetic mutations located in the globular domain of HuPrP.

4.1 Introduction: pathogenic mutations of human prion protein

There are more then twenty mutations of the prion protein gene that are know to be associated with human TSE (230). The purpose of our work is identifying common structural features among these mutants in order to shed light on the molecular basis of prion diseases. Experimental studies suggested that α_2-α_3 region constitute the main seeding element for the

conformational transition and the fibrillogenesis of the PrP. These include: (i) high-pressure NMR hydrogen/deuterion exchange studies (138; 139; 154). (ii) Isotope exchange experiments, which point to relatively large conformational flexibility of the region (247). (iii) Recent NMR studies (2), which proved that the isolated α_2-α_3 region is highly fibrillogenic and forms amyloid fibres morphologically similar to those obtained for the full-length protein. Consistently, in silico studies point to a relatively low helical propensity of the region (73**?**). Interestingly, most point mutations linked to inherited prion diseases (199) are clustered in this region (135; 202) and cause spontaneous conversion (223). The primary role of these mutations might increase the likelihood of misfolding by the thermodynamic destabilization of PrP^C (150; 223) in favor of a folding intermediate (13) that is likely to be a PrP^{Sc} precursor (12; 14). Considering this, structural studies with PrP variants containing familial mutations may provide clues regarding the proposed mechanism and may also shed lights on the role of the α_2-α_3 region that is proposed to be involved in the early stage of the conversion (2; 138; 139).

Here, we report molecular dynamics (MD) simulationbased structural predictions starting from the globular domain of the WT HuPrP deposited structure (38; 247), able to predict structural facets in aqueous solution of pathological mutations located in the globular domain of HuPrP. Interestingly, our calculations uncover common structural features specific for the pathogenic mutants, not present in the WT HuPrP. These features involve the same regions of HuPrP, although the mutations considered are different for position and characteristics. In particular, the conformational flexibility of the β_2-α_2 loop (residues 165-172) and of the α_2-α_3 region (residues 180-205) increase. The latter finding is fully consistent with the proposed role of α_2 and α_3 helices for the conversion (2; 138; 139; 154). For building up our computational protocol we chose to consider only homozygote mutants with high percentage of incidence (more than 1.1%) and carrying the methionine polymorphism [1] at codon-129 (129M). We chose 129M since WT HuPrP with the V129 polymorph displayed an increased lag phase compared with 129M (149). This means that the 129M version formed fibrils readily. In stark contrast, 129V did not form amyloid, even after more incubation time at twice the protein concentration respect the M version (149). Thus the systems chosen are two pathological mutations, HuPrP(E200K) (252) and HuPrP(Q212P) (193) causing familial CJD and GSS, respectively (202), and of the protective mutant HuPrP(E219K) (210; 211), also classified as protective polymorphism.

4.2 Computational Setup

We considered four prion systems in our study: the WT HuPrP (125-228), two pathogenic mutants, HuPrP(E200K) and HuPrP(Q212P), and the protective polymorphism HuPrP(E219K).

[1] Polymorphism in biology occurs when two or more clearly different phenotypes exist in the same population of a species, more in detail, a polymorphis is a allelic variants that change the population with a frequence more than 1%. The human prion protein has a common polymorphism at residue 129, which can be valine or methionine. The polymorphic variation has a profound influence on the ability of the protein to form amyloid fibrils spontaneously (149) (see Appendix A).

For the WT HuPrP we used the coordinates of the deposited NMR structure of HuPrP at pH 7 (PDB code: 1HJN (38; 247), residues 125-228, in Figure 4.1) carrying the methionine polymorphism at codon-129. The disease-causing HuPrP mutants HuPrP(E200K) and HuPrP(Q212P), and the polymorphic HuPrP(E219K) structures were obtained inserting the mutation on the NMR WT HuPrP via the Swiss-Pdb viewer package (100). Also for these mutants we maintained the same methionine polymorphism at codon-129. The four systems were simulated in explicit aqueous solution, inserted into a cubic box (box vector 7.18 nm) of water molecules, ensuring that the solvent shell would extend for at least 1.6 nm around them. Thus, the number of water molecules varied from a minimum of 9251 in the HuPrP(Q212P) system to a maximum of 9345 in the HuPrP(E200K) system. The AMBER99 force field (41; 242) was used for the protein, in combination with the AMBER-adapted Aqvist potential for the counter ions, (1) and the TIP3P force field for the water (122). Neutral pH conditions were realized by setting the protonation states of the ionizable residues according to their pKas. In particular the counter ions used here are the Na+ ions. Simulations were performed in periodic boundary conditions in the NPT ensemble, with temperatures and pressure kept close to the desired value (T = 298 K, P = 1 bar) through the Nos-Hoover (110; 180) and Andersen-Parrinello-Rahman (181; 188) coupling schemes respectively. Long-range electrostatic interactions were treated with the particle mesh Ewald (PME) (66; 80) method, using a grid with a spacing of 0.12 nm, combined with a fourth-order cubic spline interpolation (107) to compute the potential and forces in between grid points. The cutoff radius for the Lennard-Jones interactions, as well as for the real part of PME calculations, was set to 0.9 nm. The LINCS algorithm (106) was used to constrain all bond lengths involving hydrogen atoms and the time-step used was 2 fs. The systems were energy-minimized imposing harmonic position restraints of 1000 kJmol1nm2 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 restraints, the temperature was gradually increased from 0 to 298 K. This was performed in 12 steps, in which the temperature was increased by 25 K in 100 ps of MD. Each system was finally simulated for 50 ns. Four simulations were carried out for each system: the same models were used; however, we assigned statistically independent initial velocities harvested from a Maxwell distribution at the appropriate temperature. This allows generating four sets of statistically independent trajectories sampling the same Boltzmann distribution. All simulations were performed and analyzed with the GROMACS software package (27). g_sas was used for computing the solvent accessible surface area with a probe radius of 0.14 nm. g_mdmat was used to make distance matrix in function of time consisting of the smallest distance between residues pairs. A distance cutoff of 0.5 nm was used as a discrimination criterion for the presence of both salt-bridge (SB) and π-stacking interactions. For the SB interaction, we used the distance from the anionic carboxylate (RCOO-) of either aspartic acid or glutamic acid and the cationic ammonium (RNH_3^+) from lysine or the guanidinium (RNHC(NH2)$^{2+}$) of arginine. For π-stacking interactions, we used the center of mass of atoms belonging to the aromatic ring as the representative point

for the distance criteria [1] The H-bonds were determined based on a cutoff distance of 0.35 nm between Hydrogen and Acceptor and a cutoff angle of 30 degree on Acceptor-Donor-Hydrogen angle.

Clusterization on MD trajectories.

The module g_cluster was used for clusterizing the trajectory by the gromos method (67) and a RMSD cutoff of 0.1 nm; We followed two different approaches to cluster the MD trajectories:

(i) Partial clusterization. We focused here on each single trajectory for each system. The results turn out to be very similar among the trajectories. This strengthens the accuracy of our computational protocol.

(ii) Total clusterization. We focused here on the global set of trajectories for each system. As expected, the results turn out to be very similar to those of (i). The total clusterization is reported in Appendix C.

4.3 Results

Here we perform 50 ns MD simulations for WT HuPrP structure (in our notation MD_WT), HuPrP(E200K) (MD_E200K herafter), HuPrP(Q212P) and HuPrP(E219K). The statistical accuracy of our result was investigated by performing four independent simulations for each system. The resulting findings are very similar for each set of four runs. We first established our setup by comparing MD_WT with the corresponding NMR deposited structure (PDB code 1HJN (38; 247), in Figure 4.1 A and in our notation NMR_WT). Next, we compare MD_E200K (Fig.4.1) with the NMR structure (PDB code: 1FO7 (252), in our notation NMR_E200K). The variant is constructed by an in silico mutation of NMR_WT (see Methods). The calculations allow to establish the predictive power of this procedure (Comparison is also made with the recently solved X-ray structures of WT HuPrPs [2] and its mutants (145), albeit our work focuses on the structure of the protein in water solution). Hence, we used the same in silico mutation setup for predicting the structures of HuPrP(Q212P) and HuPrP(E219K), in Fig. 4.1 B. We focus on the regions for which we observed a larger conformational flexibility in the mutants relative to WT HuPrP. These are the α_2-α_3 region (residues 180-205) and the β_2-α_2 loop (residues 160-175).

4.3.1 WT HuPrP.

The fold of the protein, as obtained by MD calculations, was the same as that of NMR_WT. For the backbone root main square deviation (RMSD) differences between MD_WT HuPrP

[1] The respective orientation of the aromatic rings and thus the type of π-stacking interactions have not been investigated here.

[2] Notice that for this comparison we considered only the monomeric form, as our simulations provided the structural determinants for the monomer in solution

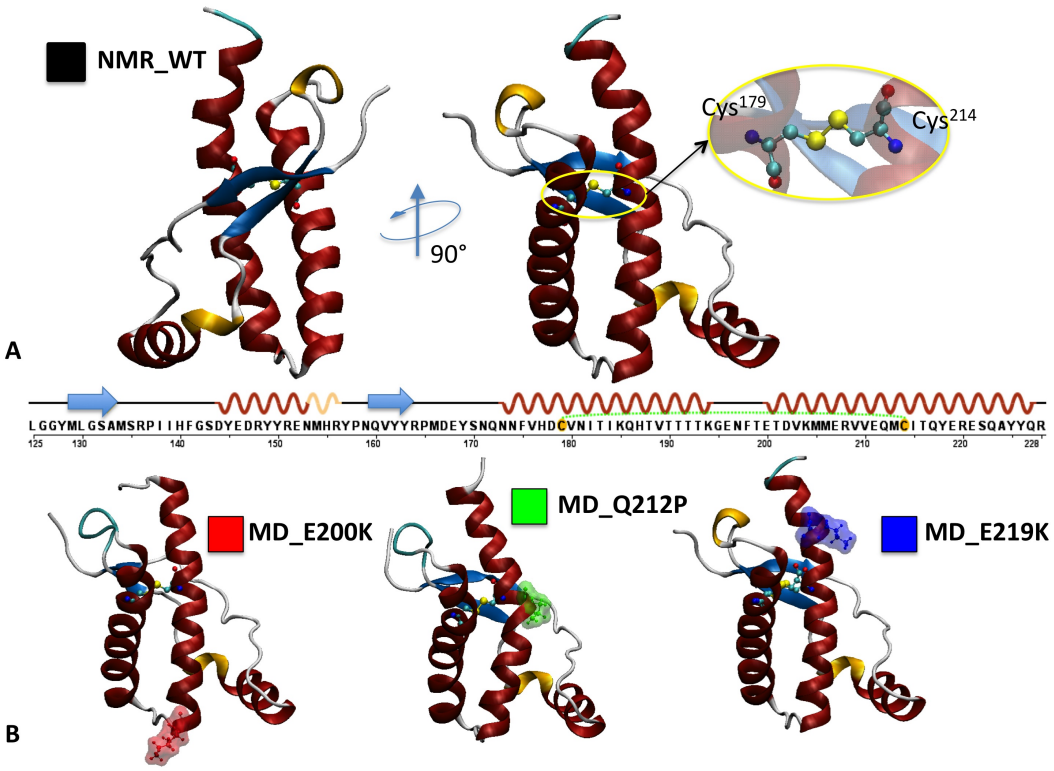


Figure 4.1: Models studied - A) NMR structure of the globular domain of WT HuPrP (PDB code 1HJN(38; 247) and in our notation NMR_WT). The view on the right highlights the only S-S bridge present in the structure. B) The HuPrP mutants MD_E200K, MD_Q212P and MD_E219K, as emerging from the MD simulations. The color-coding of the systems is the same in all the figures (i.e. Black for the MD_WT, red for MD_E200K, green for MD_Q212P and blue for MD_E219K). The snapshots reported here correspond to the most populated cluster. The mutated residue is indicated in ball-and-sticks.

and NMR_WT were smaller than 0.12±0.01 nm for the entire folded domains and less than 0.09±0.01 nm for the non-loop; this was found for all the four simulation performed on the WT (see Tab. 4.2).

	Reference: MD_WT structure (nm)	Reference: NMR corresponding structure (nm)*	
		Folded-domain	Non-loop
MD_WT	0.14±0.05	0.11±0.01	0.09±0.01
MD_WT 1	0.14±0.05	0.11±0.01	0.09±0.01
MD_WT 2	0.12±0.03	0.12±0.03	0.09±0.01
MD_WT 3	0.12±0.03	0.12±0.03	0.09±0.01
MD_E200K	0.20±0.02	0.20±0.02	0.15±0.01
MD_E200K 1	0.13±0.02	0.13±0.02	0.09±0.02
MD_E200K 2	0.13±0.03	0.13±0.02	0.09±0.02
MD_E200K 3	0.14±0.03	0.14±0.02	0.09±0.01
MD_Q212P	0.17±0.02	Not available	
MD_Q212P 1	0.14±0.02	Not available	
MD_Q212P 2	0.15±0.02	Not available	
MD_Q212P 3	0.15±0.02	Not available	
MD_E219K	0.20±0.03	Not available	
MD_E219K 1	0.13±0.02	Not available	
MD_E219K 2	0.12±0.0.2	Not available	
MD_E219K 3	0.12±0.0.2	Not available	

Figure 4.2: RMSD table - Partial clusterization. RMSD values of the system studied respect the MD_WT structure. The structure chosen as the reference is the representative structure of the most populated cluster coming from our cluster analysis.

The most flexible regions are the β_2-α_2 loop and the α_2-α_3 region, as obtained by the plot of the root mean square fluctuation (RMSF) (Fig. 4.3).

The α_2-α_3 region.

Several specific salt bridges (SBs) and hydrophobic long-range interactions in the helical interface have been proposed to play a fundamental role in fold stabilization of α_2-α_3 region (Fig. 4.6 A)(2). Their importance is supported by the observation that some of them are absent in a number of disease cases (Fig. 4.6 B) (2).

The charged residues involved in SBs, five in helix 3 (Glu^{200}, Lys^{204}, Glu^{207}, Arg^{208} and Glu^{211}) and one in helix 2 (Glu^{146}) formed a well-packed electrostatic net. Specifically, four SBs were present (Fig. 4.6 A and 4.4): two of these were formed by Arg^{208} with both $Glu^{207}$and

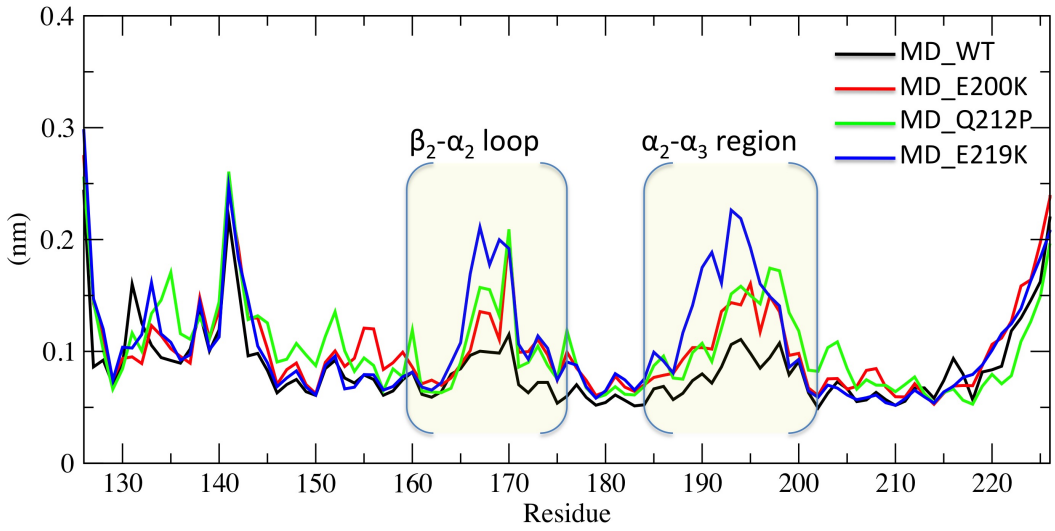


Figure 4.3: Root mean square fluctuations (RMSF) - RMSF of the MD structures as a function of residue number. The regions that showed higher flexibility are highlighted. The first and the last residues have a RMSF value higher than 0.3 nm, since they are in a terminal position. Thus they are omitted here. Notably each point is a mean value calculated on all the simulations for the same system.

Systems		MD_WT 1			MD_WT 2			MD_WT 3			MD_WT 4		
Total population collected (%)		80			81			80			80		
Relative population collected (%)		61.2	21.7	17.1	72.4	13.9	13.7	62.3	20.7	17.0	59.7	23.5	16.8
Cluster		1	2	3	1	2	3	1	2	3	1*	2	3
SB(nm)	*Arg208 -Glu211*	0.39	0.40	0.38	0.40	0.40	0.41	0.41	0.40	0.39	1.02	0.48	0.38
	Glu207-Arg208	0.39	0.44	0.45	0.40	0.39	0.39	0.37	0.43	0.44	0.82	0.45	0.43
	Lys204-Glu207	1.01	0.97	1.05	0.99	0.99	0.96	0.98	0.93	1.04	0.35	0.84	0.99
	Glu200/Lys200-Lys204	0.93	0.37	0.92	0.38	0.92	0.76	0.89	0.47	0.82	0.99	1.28	0.48
	Glu146-Lys204	0.78	0.84	0.35	0.89	0.49	0.76	0.77	0.84	0.38	0.99	0.40	0.99

*Interestingly here a different SB conformation was found in the first cluster: Glu^{146} interacts with Arg^{208} (d_{SB}=0.38 nm) instead of Lys^{204} creating a new SB not observed in all the other simulations of the same system. The second and the third cluster resemble instead the behavior already found.

Figure 4.4: MD_WT SBs - Lys^{204}-Glu^{200}, Lys^{204}-Glu^{146}, Lys^{204}-Glu^{207}, Arg^{208}-Glu^{207} and Arg^{208}-Glu^{211} SBs.The distances are calculated in the first three clusters (able to collect more or less the 80 per cent of the total population) coming from a clustering analysis on the overall trajectory for three independent MD simulations for each system.

Glu^{211} (Block A); the other two were formed by Lys^{204} with Glu^{200} and Glu^{146} (Block B). In Block A, Arg^{208} position remained fixed overall the simulation by an hydrogen bond with Gln^{212}; as a consequence, Arg^{208}-Glu^{207} and Arg^{208}-Glu^{211} SBs remained stable overall the simulation. In Block B, Lys^{204}-Glu^{200} and Lys^{204}-Glu^{146} were instead alternatively formed and broken: Lys^{204} side chain switched indeed from a conformation closed to Glu^{200} to another closed to Glu^{146}. From the cluster analysis of the MD trajectory, we found that, in the most populated cluster Lys^{204} was in a middle position between both. In the second most populated cluster, Lys^{204} was close to Glu^{200} while in the third cluster, Lys^{204} was close to Glu^{146} (Tab. 4.4). A similar result was found in the 20 conformers reported for NMR_WT. In nine clusters Lys^{204} was in a middle position, in seven clusters, Lys^{204} was close to Glu^{200}, and in the other four Lys^{204} was close to Glu^{146}. Hence, our computational setup reproduces this key feature of the NMR structure. In one of the other three MD simulations we observe a different conformation for the SB interactions in one of the cluster: Glu^{146} interacts with Arg^{208} instead of Lys^{204}. This forms a SB not observed in all the other simulations (see 4.4).

We next focus on another key feature of the region. This is the disulphide bond between Cys^{179} and Cys^{214} (Fig. 4.1, right), since disulphide bridges are well known to play key roles in stability, folding and functions of proteins. The structure of a disulfide bond can be described by its χ_{SS} dihedral angle between the Cβ Sγ Sγ Cβ atoms, which is usually close to $\pm$ 90 degrees. Here, χ_{SS} was $\sim$ -90 degrees.

Interaction	π-stacking		(Cys^{179}-Cys^{214}) bridge
Systems	***Tyr*^{169}*-Phe*^{175}** (nm)	***Phe*^{175} *- Tyr*^{218}** (nm)	Cα- Cα(nm)
MD_WT 1	0.45±0.02	0.46±0.03	0.68±0.01
MD_WT 2	0.44±0.02	0.45±0.02	0.68±0.01
MD_WT 3	0.45±0.02	0.46±0.03	0.69±0.01
MD_WT 4	0.45±0.02	0.46±0.02	0.68±0.01

Figure 4.5: MD_WT - π-stacking and (Cys^{179}-Cys^{214}) bridge mean values for each set of simulations.

The β_2-α_2 loop.

It features a characteristic three-centered aromatic cluster based on the π-stacking interactions between Tyr^{169}, Phe^{175} and Tyr^{218} (see 4.6 A and 4.5). A side-chain hydrogen bond (HB) between Tyr^{169} and Asp^{178} keeps the stability of this peculiar configuration. Remarkably, this hydrophobic organization in the β_2-α_2 loop is similar in the NMR structure (38), as well as in the recently reported X-ray structures of WT HuPrPs in its monomeric form (145). We conclude that our computational setup reproduces the key structural facets of this region.

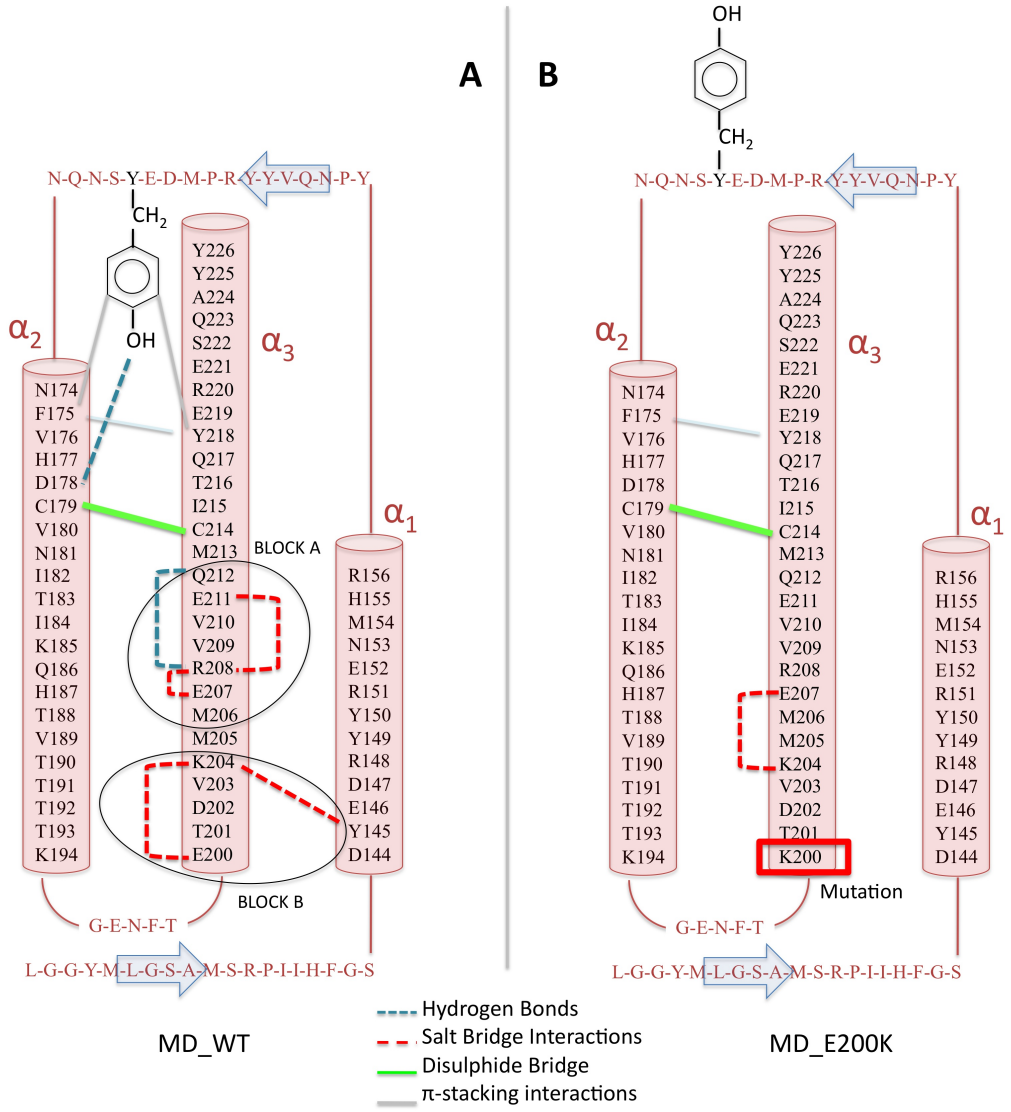


Figure 4.6: Scheme of key interactions in the systems studies. - A scheme of the two different structural behaviors identified for MD_WT and MD_E200K is visualized (we chose the MD_E200K since the structural behaviors of MD_Q212P mutant are similar, see main text). Only the key interactions are reported: the SBs in α_3 and α_2, the position of the Tyr^{169}, the HBs and the π-stackings between Tyr^{169}, Phe^{175} and Tyr^{218}.

4.3.2 HuPrP(E200K).

The 3D structure of the folded domain of this mutant was obtained by replacing Glu^{200} with a Lys in the NMR_WT structure (Fig. 4.1 B). The overall fold turned out to be similar to that of the NMR_E200K (see 4.2), as well as the MD_WT for all the four-independent simulations of this system. However, the RMSF of the residues belonging to the most flexible regions (the β_2-α_2 loop and the α_2-α_3 region) increased (Fig. 4.3).

The α_2-α_3 region.

The mutation affected α_3, where the mutated residue was located. The $Glu^{200} \rightarrow Lys^{200}$ mutation destroyed the SB interactions characterizing that region and as a consequence the two-block organization found in the MD_WT was lost (Fig. 4.6 B): the absence of Lys^{204}-Glu^{200} SB in Block B caused a rearrangement of Lys^{204} side chain, which interacted now with Glu^{207} (see 4.7 and Fig. 4.6 B). The new Lys^{204}- Glu^{207} SB in block B broke the Arg^{208}-Glu^{207} and the Arg^{208}-Glu^{211} SBs in Block A. Indeed, in the most populated cluster, the Arg^{208}-Glu^{207} and the Arg^{208}-Glu^{211} SBs were not observed (see 4.7). We found the same for all the other simulations for this system (see 4.7).

Systems		MD E200K 1			MD E200K 2			MD E200K 3			MD E200K 4		
Total population collected (%)		82			86			85			83		
Relative population collected (%)		73.1	14.2	12.7	59.5	25.5	15.0	87.9	6.9	5.2	56.4	31.6	12.0
Cluster		1	2	3	1	2	3	1	2	3	1	2	3
SB(nm)	*Arg208 -Glu211*	0.97	0.84	0.97	0.94	0.38	0.43	0.92	0.91	0.87	0.93	0.82	1.02
	Glu207-Arg208	1.13	1.10	1.13	0.97	0.45	0.69	0.99	1.09	0.87	1.04	1.02	1.06
	Lys204-Glu207	0.95	0.72	0.65	0.36	0.98	0.94	1.04	0.56	0.99	0.94	0.68	0.38
	Glu200/Lys200-Lys204	1.13	1.22	0.96	0.82	0.70	0.96	0.76	1.20	0.98	0.78	0.94	1.15
	Glu146-Lys204	0.50	1.31	1.01	0.35	1.06	0.66	0.45	0.93	0.55	0.82	0.99	1.33

Figure 4.7: MD_E200K SBs - Lys^{204}-Glu^{200}, Lys^{204}-Glu^{146}, Lys^{204}-Glu^{207}, Arg^{208}-Glu^{207} and Arg^{208}-Glu^{211} SBs.

Due to the new SB net configuration, all the five charged residues of helix 3 were totally exposed to the solvent showing an increased solvent accessible surface (SAS) especially for Glu^{211}, Arg^{208}, Glu^{207} and Glu^{200} (see Fig. 4.8). Remarkably, all of these features were also observed in the NMR structures.

The χ_{SS} in MD_E200K was $\sim$ +90 degrees, as opposed to MD_WT where it was $\sim$ -90 degrees. As a result, the distance between the Cα atoms of these residues was shorter in

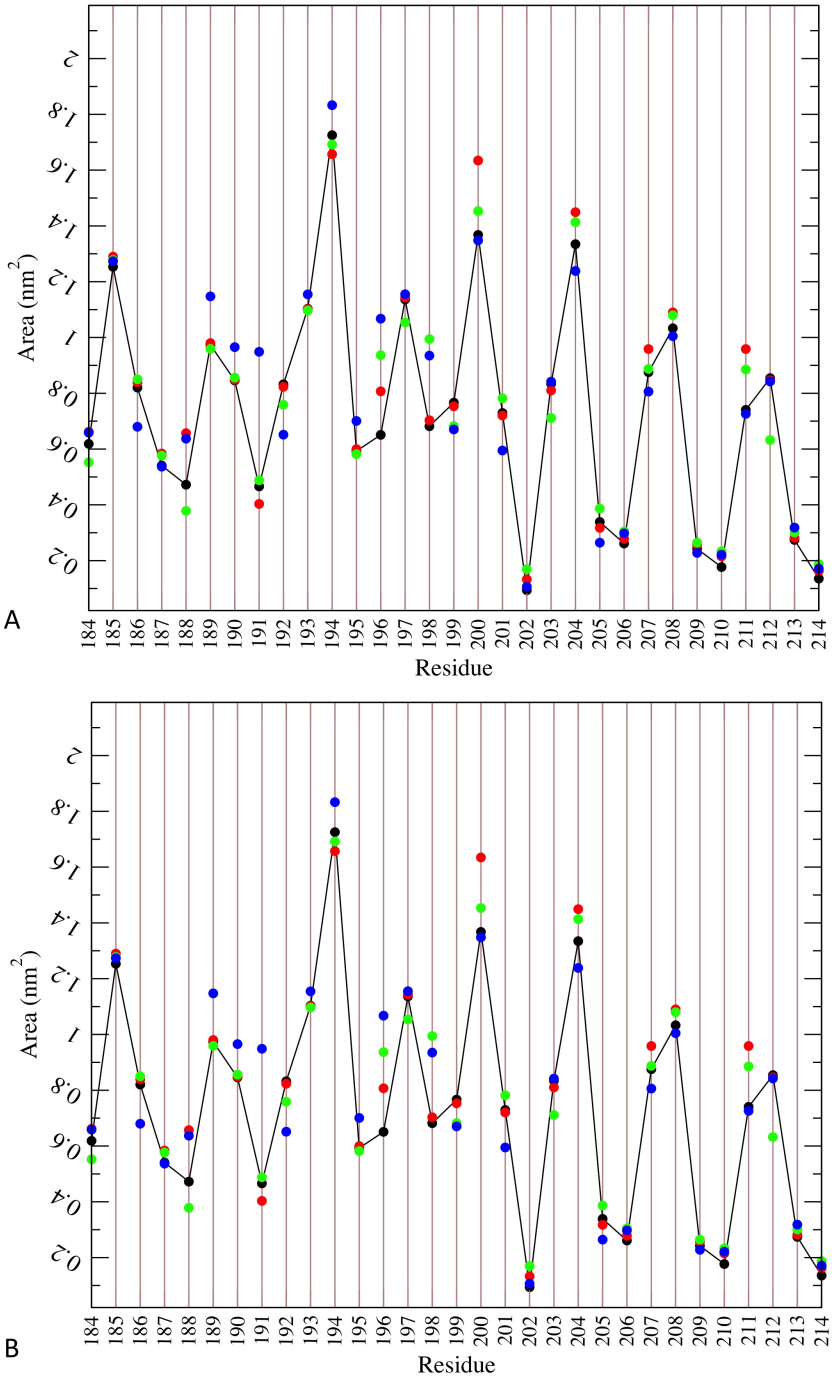


Figure 4.8: Solvent Accessible Surface(SAS) - SAS of the β_2-α_2 loop (A) and the α_2-α_3 region (B). The color code is the same of the figure in the Main Text: MD_WT: black; MD_E200K: red; MD_Q212P: green; MD_E219K: blue. Notably each point is mediated on all the simulation for the same system.

MD_E200K than in the MD_WT (0.65±0.02 nm and 0.68±0.01 nm respectively, see 4.5 and 4.9). Although a difference of 0.03 nm between MD_WT and MD_E200K is small, the same trend was confirmed in the corresponding NMR structures (0.59±0.01 and 0.68±0.03 nm respectively).

The β_2-α_2 loop.

Despite the mutated residue belonging to the helix 3, the β_2-α_2-loop region (residues 160-175) rearranged (Fig. 4.3). Tyr^{169} side chain passed from a gauche+ to a trans conformation and consequently its aromatic ring was solvent exposed (Fig. 4.10 A). This was confirmed by an increase of about 0.5 nm2 in the SAS of this residue compared to the MD_WT (Fig 4.8). Thus, the characteristic three-centered aromatic cluster (Tyr^{169}, Phe^{175} and Tyr^{218}), observed in the MD_WT, was no longer present. Only the π-stacking interaction between Phe^{175} and Tyr^{218} was maintained (see 4.9).

Interaction	π-stacking		(Cys^{179}-Cys^{214}) bridge
Systems	***Tyr*^{169}*-Phe*^{175}** (nm)	***Phe*^{175} *- Tyr*^{218}** (nm)	Cα- Cα(nm)
MD_E200K 1	0.84±0.07	0.50±0.04	0.65±0.02
MD_E200K 2	0.72±0.09	0.47±0.03	0.65 ± 0.02
MD_E200K 3	1.12±0.10	0.46±0.03	0.65±0.02
MD_E200K 4	0.80±0.15	0.47±0.03	0.65±0.02

Figure 4.9: MD_E200K - π-stacking and (Cys^{179}-Cys^{214}) bridge mean values

Finally, Ramachandran angle transition of adjacent residues Ser^{170} and Asn^{171}, not present in the MD_WT, were observed. Remarkably, these calculations reproduced the NMR structure (Fig. 4.10).

We conclude that the NMR structural determinants are reproduced. Therefore, we use our protocol to predict the structure of HuPrP(Q212P) and HuPrP(E219K).

4.3.3 HuPrP(Q212P).

The overall folded domain was similar to that of WT HuPrP (See 4.2, Fig. 4.1 B). Nevertheless, based on the RMSF (Fig. 4.3), we observed also in this case that this mutation caused an increase of the flexibility of the α_2-α_3 and β_2-α_2 regions.

The α_2-α_3 region.

Gln^{212} replacement with Pro caused the breaking of the HB with Arg^{208} present in the MD_WT (Fig. 4.6). The loss of the HB caused an increase of the pitch of the helix (Fig. 4.11) and as a

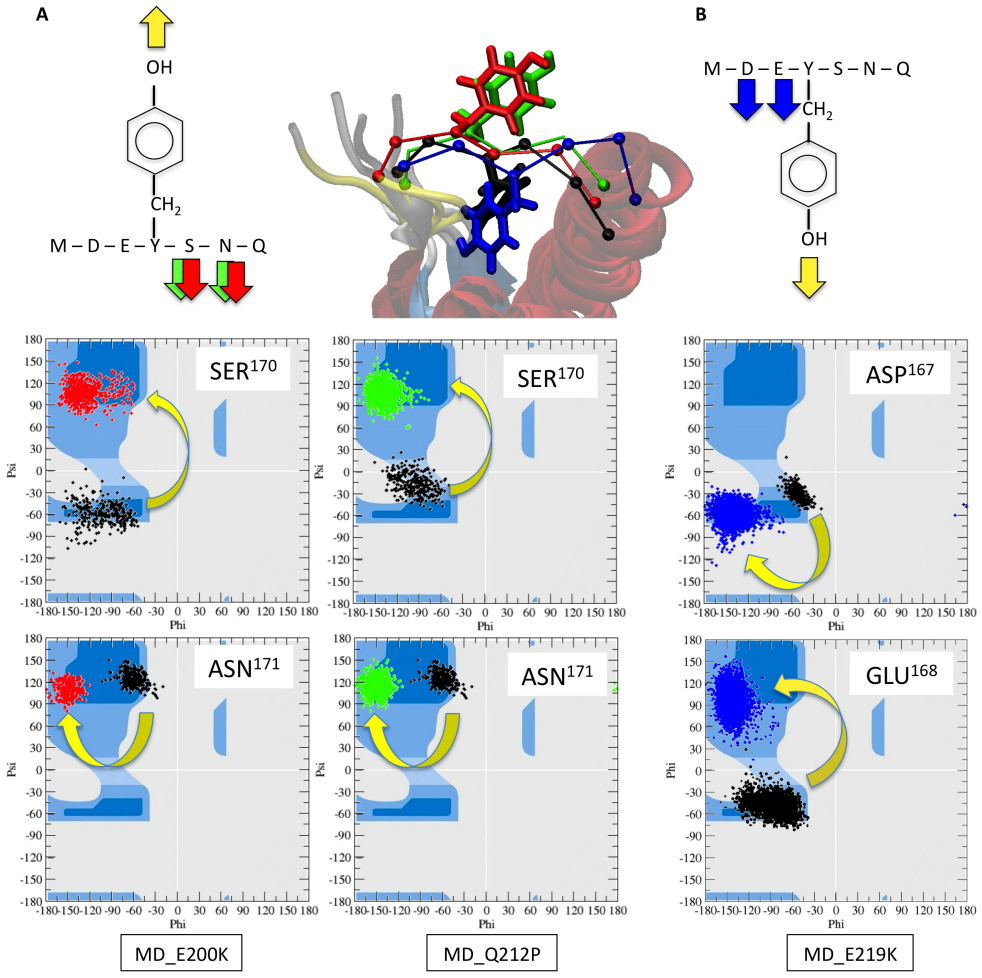


Figure 4.10: Dihedral Angle transitions in the β_2-α_2 loop. - A) Conformation transition of Ser^{170} and Asn^{171} in MD_E200K and MD_Q212P (red and green points respectively) toward the MD_WT (black points). Notably each point is a mean value calculated on all the simulations for the same system. B) Conformation transition of Glu^{168} and Asp^{167} in MD_E219K (blue points) toward the MD_WT (black points). Notably each point is a mean value calculated on all the simulations for the same system. A superimposition of the HuPrP mutant structures on the MD_WT structure in the β_2-α_2 loop is also reported: the position of Tyr^{169} is underlined with a mixed tube/cartoon representation. The snapshots reported correspond to the most populated cluster of one simulation set.

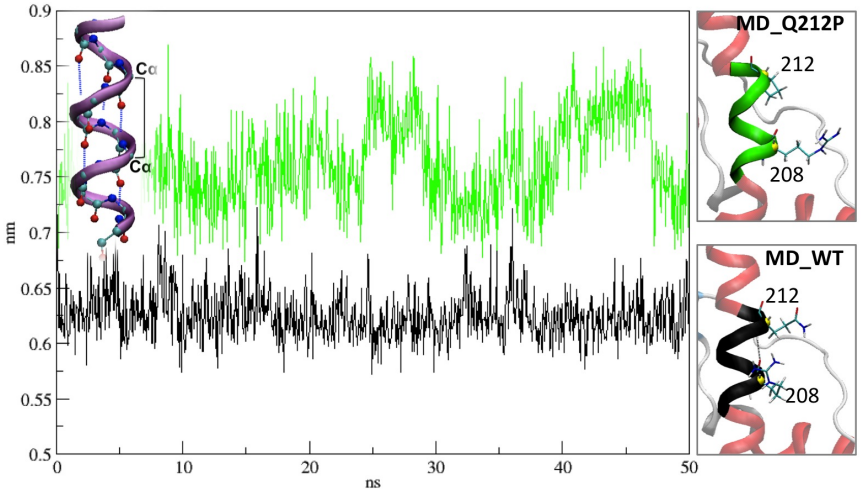


Figure 4.11: α_3 helix pitch of MD_WT toward MD_Q212P. - The Cα(Arg^{208})- Cα(Gln^{212}) distance for MD_WT (black-line) and the Cα(Arg^{208})-Cα(Pro212) distance for MD_Q212P (green-line) are plotted as a function of the simulated time. The mean values are respectively 0.73±0.04 nm and 0.63±0.02 nm. Notably each point is a mean value calculated on all the simulations for the same system.

Systems		MD_Q212P 1			MD_Q212P 2			MD_Q212P 3			MD_Q212P 4		
Total population collected (%)		83			87			83			82		
Relative population collected (%)		55.6	30.2	14.2	69.6	17.4	13.0	76.0	14.2	9.8	72.0	16.4	11.6
Cluster		1	2	3	1	2	3	1	2	3	1	2	3
SB(nm)	*Arg208 -Glu211*	1.09	0.91	0.99	1.04	1.10	0.69	0.38	0.97	0.71	0.79	0.52	0.98
	Glu207-Arg208	0.90	0.92	0.94	0.82	0.81	0.44	0.39	0.74	1.11	0.93	0.40	0.93
	Lys204-Glu207	1.03	0.94	0.52	0.38	0.92	0.84	0.97	0.98	0.40	0.92	0.52	0.82
	Glu200/Lys200-Lys204	0.66	0.64	1.05	1.03	0.76	0.73	0.64	0.38	0.95	1.01	1.05	0.58
	Glu146-Lys204	0.45	0.81	0.79	1.00	0.42	0.72	1.08	1.13	1.14	0.60	1.30	0.89

Figure 4.12: MD_Q212P SBs - Lys^{204}- Glu^{200}, Lys^{204}-Glu^{146}, Lys^{204}-Glu^{207}, Arg^{208}-Glu^{207} and Arg^{208}-Glu^{211} SBs.

result, Arg^{208} rearranged: it weakened its SB with Glu^{211} and Glu^{207} and formed, at times, a SB with Glu^{146} (see 4.12). Thus, also Glu^{146}- Lys^{204} SB weakened.

The cluster analysis reflected that situation: on three main clusters (that collect the 86% of population), the Glu^{146}-Lys^{204} SB was never found (see 4.12). The final SB pattern was strikingly similar to that of HuPrP(E200K), with an increasing of SAS especially for Glu^{211}, Glu^{207} and Lys^{204}. The latter became totally exposed to the solvent.

Finally, we checked the Cα(Cys^{179})- Cα(Cys^{214}) distance (d=0.65±0.02 nm). This was shorter than that in MD_WT (see 4.13), and similar to that in MD_E200K. Also here, Cys^{179} switched toward a $gauche^+$ conformation as found for the HuPrP(E200K) mutant.

The β_2-α_2 loop.

This rearranged very similarly to the loop in MD_E200K: the Tyr^{169} side chain switched from a $gauche^+$ to a trans conformation, becoming exposed to the solvent (Fig. 4.10). In particular, the SAS of this residue increased of about 0.5 nm2 relative to MD_WT (Fig 4.8 B). Also here, the characteristic three-centered aromatic cluster (Tyr^{169}, Phe^{175} and Tyr^{218}) was no longer present (see 4.13). Remarkably Ser^{170} and Asn^{171} rearranged as in MD_E200K mutant (Fig. 4.10 A).

Interaction	π-stacking		(Cys^{179}-Cys^{214}) bridge
Systems	***Tyr*^{169}*-Phe*^{175}** (nm)	***Phe*^{175} *- Tyr*^{218}** (nm)	Cα- Cα(nm)
MD_Q212P 1	0.86±0.05	0.47±0.03	0.65±0.02
MD_Q212P 2	0.93±0.18	0.46±0.02	0.65±0.02
MD_Q212P 3	0.80±0.18	0.46±0.02	0.65±0.01
MD_Q212P 4	0.90±0.22	0.48±0.04	0.66±0.02

Figure 4.13: MD_Q212P - π-stacking and (Cys^{179}-Cys^{214}) bridge mean values

We conclude that the structural features of HuPrP(Q212P) are similar to those of HuPrP(E200K). This is really unexpected, since the mutation is different in chemical properties and it is localized in a different part of the third helix.

4.3.4 HuPrP(E219K).

The overall folded domain remained essentially similar to the WT HuPrP (see 4.2, Fig. 4.1 B). Nevertheless, it had a widespread influence on the flexibility of the same two regions (β_2-α_2 and α_2-α_3) already observed in the pathogenic mutants (Fig. 4.3).

Systems		MD_E219K 1			MD_E219K 2			MD_E219K 3			MD_E219K 4		
Total population collected (%)		76			78.2			80			84		
Relative population collected (\%)		59.2	30.7	10.1	72.0	19.4	8.6	59.8	30.5	9.7	60.6	29.4	10.0
Cluster		1	2	3	1	2	3	1	2*	3	1	2	3
SB(nm)	*Arg208 -Glu211*	0.37	0.39	0.38	0.38	0.38	0.43	0.38	0.99	0.43	0.51	0.50	0.51
	Glu207-Arg208	0.34	0.51	0.36	0.44	0.45	0.45	0.41	0.83	0.42	0.35	0.36	0.36
	Lys204-Glu207	0.89	1.05	0.97	0.9	1.01	0.99	1.08	0.84	0.60	0.97	0.99	0.88
	Glu200/Lys200-Lys204	0.72	0.43	1.00	0.59	0.36	0.82	0.92	0.36	1.30	0.79	0.35	0.94
	Glu146-Lys204	0.75	0.83	0.32	1.23	0.82	0.48	0.38	0.86	0.97	0.70	1.11	0.47

* A different SB conformation was found in the second cluster: *Glu*^{146} interacts with *Arg*^{208} (d_{SB}=0.37 nm) instead of interacting with *Lys*^{204}. Interestingly, this conformation was found also for one of the clusters in the WT system, underlining again the similarities between these two systems.

Figure 4.14: MD_E219K SBs - Lys^{204}- Glu^{200}, Lys^{204}-Glu^{146}, Lys^{204}-Glu^{207}, Arg^{208}-Glu^{207} and Arg^{208}-Glu^{211} SBs

The α_2-α_3 region.

Glu219 is located at the end of α_3 and its mutation did not interfere with the SB pattern of helix 3 (see 4.14): it remained identical to the MD_WT.

The Cα(Cys^{179})- Cα(Cys^{214}) distance was also similar (d=0.67$\pm$0.02 nm) to that of MD_WT and, consequently, longer than that of the pathogenic mutants (see 4.16).

The β_2-α_2 loop.

The three-centered aromatic cluster based on the π-stacking interactions between Tyr^{169}, Phe^{175} and Tyr^{218} was maintained in three of the four simulation performed for this system. More in detail, Tyr^{169} side chain pointed toward the bulk of the protein as in MD_WT (see Fig. 4.10 B and 4.16), however it H-bonded to Arg^{164} instead of Asp^{178}, as in MD_WT (Fig. 4.15).

In one of the other three simulations, a different configuration for Tyr^{169} is observed. The residue is sandwiched between Tyr^{218} and Tyr225, pointing, as in the other cases, towards the bulk of the protein (see 4.16 and side Figure). Glu^{168} and Asp^{167} rearranged to a conformation other than that of the pathogenic mutants (Fig. 4.10 B). Ser^{170} and Asn^{171} instead had the same conformation as in MD_WT. We conclude that the HuPrP(E219K) polymorphism leads to structural determinants close to those of WT HuPrP and opposed to HuPrP(Q212P), and HuPrP(E200K), although it features a mutation close in space to that of HuPrP(Q212P), and

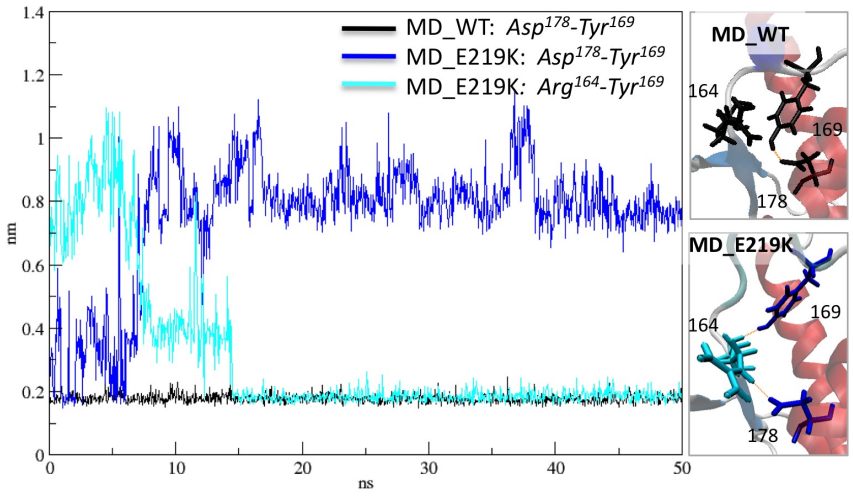


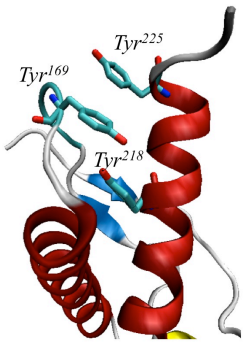
Figure 4.15: Tyr^{169} **HBs.** - Tyr^{169}- Asp^{178} HB for MD_WT, Tyr^{169}- Asp^{178} and Tyr^{169}- Arg^{164} HBs for MD_E219K are represented as a function of the simulated time. Notably this configuration is the same in all the simulation sets. Only an exception is found and it is described in the text

is similar in chemical type to HuPrP(E219K).

4.4 Discussion

It has been postulated that the primary role of inherited point mutations is to increase the likelihood of misfolding by the thermodynamic destabilization of PrP^C (150), (223), triggering in turn an abnormal interaction with other not yet identified co-factors (124; 225), or promoting an aberrant accumulation inside the cell (15). This hypothesis is based on the experimental evidence that the effect of inherited point mutations on the HuPrP structure is the spontaneous formation of PrP^{Sc} in the brain (202). Therefore, structural studies with HuPrP variants containing familial mutations should provide important clues regarding the molecular basis of the disease. One of the crucial questions is the identification of the regions involved in the transition process from -helical to -sheet rich structure. Several experimental studies lead to the proposal that the α_2-α_3 region plays a crucial role during the conversion from PrP^C to PrP^{Sc} (2; 138; 139; 154). We have presented here a MD simulations protocol where starting from the globular domain (residues 125-228) of WT HuPrP NMR structure (38; 247), we were able to predict structural facets in aqueous solution of two pathological mutations, HuPrP(E200K) (252) and HuPrP(Q212P)(193) causing familial CJD and GSS respectively (202), and of the protective polymorphism HuPrP(E219K) (210; 211). All these variants are located in the globular domain of HuPrP. Our computational setup has been established by comparing the structural determinants of MD_WT and MD_E200K with the corresponding NMR structures.

Interaction	π-stacking		(Cys^{179}-Cys^{214}) bridge
Systems	***Tyr*^{169}-*Phe*^{175}** (nm)	***Phe*^{175} - *Tyr*^{218}** (nm)	Cα- Cα(nm)
MD_E219K 1	0.44±0.04	0.43±0.02	0.67±0.01
MD_E219K 2	0.44±0.02	0.46±0.02	0.68±0.01
MD_E219K 3	-----*	0.45±0.02	0.68±0.02
MD_E219K 4	0.45±0.01	0.46±0.02	0.68±0.01



*A different configuration for Tyr169 is here observed. The residue lies between Tyr218 and Tyr225. See Figure on the side.

Figure 4.16: MD_E200K - π-stacking and (Cys^{179}-Cys^{214}) bridge mean values

To improve the statistics of our procedure, four independent, 50 ns long MD simulations have been run for each system (800 ns in total). The results turn out to be very similar across the four runs. The most remarkable result was that the pathogenic mutations affected the same structural regions of the HuPrP, regardless of mutant positions, chemical properties and clinical features. These regions are the α_2-α_3 region and the β_2-α_2 loop. The residues belonging to α_2-α_3 loop region and the β_2-α_2 loop are more flexible than the correspondent ones of WT HuPrP. This was in agreement with the fact that the same regions underwent an essentially complete isotope exchange within less than 30 min, suggesting relatively large conformational flexibility (247). In this respect, it is really remarkable that NMR studies have shown that a completely different pathogenic mutation, H187R, has the same effect on the highly homologous mouse prion protein (17). The larger flexibility of the α_2-α_3 region, observed in the HuPrP(E200K) and HuPrP(Q212P) relative to WT HuPrP might be related to the loss of specific SBs. This is fully consistent with the suggestion, based on in silico studies, that the α-helical fold of α_2-α_3 is assisted by intermolecular interactions other than those of the α-helices themselves (73**?**), as already observed in other α-helical proteins (123; 245). The residues involved in the SBs are clearly very relevant: they are highly conserved and their absence is linked with pathogenic behaviors (2). In spite of their clear importance, so far, a role for protein fibrillation had only been suggested (2). Our calculations corroborate this suggestion. Interestingly, the loss of SBs is associated with a change of the structural parameters of the only disulphide bridge of the

protein, in full agreement with the available experimental data (38; 247; 252). The larger flexibility of the β_2-α_2 loop is due to different features depending on the mutations. In the healthy variant of HuPrP (i.e. MD_WT and MD_E219K), the Tyr^{169} side chain was in gauche+ conformation pointing towards the bulk of the protein. In the pathogenic mutants (i.e. MD_E200K and MD_Q212P) such side chain was in a trans conformation with the consequent exposure of the aromatic ring towards the solvent. Only in these cases the characteristic three-centered aromatic cluster (Tyr^{169}, Phe^{175} and Tyr^{218}), observed in the MD_WT, were no longer present. Interestingly, the crystal structures of HuPrP(D178N) and HuPrP(F198S) mutants effectively showed similar orientation of Tyr^{169} side chain outside the globular part.(145) Tyr^{169} is in the region that might modulate the binding of a hypothetical facilitator of prion conversion involved in the development of TSE (124; 225); thus such conformational change may cause an altered protein-protein interactions. We therefore put forward the hypothesis that the presence of Tyr^{169} outside the loop is a key pathological structural feature. In summary, we suggest that the strongest effect of the HuPrP mutations in the globular domain might always involve the same region of the protein and always follow the same mechanism: on one hand, altering the conformation of the β_2-α_2 loop region and modifying the exposed hydrophobic surface, and on the other, altering the long-range interaction in the 2-3 region. Moreover, this study sheds new light on the molecular basis of inherited prion diseases, introducing novel structural features, which are of great importance for our understanding of these disorders and of the earliest molecular events leading to the conformational transition of PrP^C into a β-sheet-rich fold. Our protocol could be used as a useful high-throughput approach to investigate and predict structural facets of HuPrP disease-causing mutations when increasing the statistics of mutation studied. A statement of warning should be made. Because of the presence of the long unstructured N-term segment, the implications of our findings for the full-length protein in vitro must be taken with great caution. It is possible that the unstructured PrP segment might alter the structure and the stability of the protein in a rather drastic way.

After I had carried out this MD study, Ilk et al. (115) reported the NMR solution structure of the truncated recombinant human PrP from residue 90 to 231 carrying the HuPrP(Q212P) mutation. My predictions are in complete agreement with the NMR data. In particular, Tyr^{169} is completely exposed to solvent and the β_2-α_2 loop is highly flexible in both computation and experiment (data not shown). This further strengthens the validity of my computational protocol.

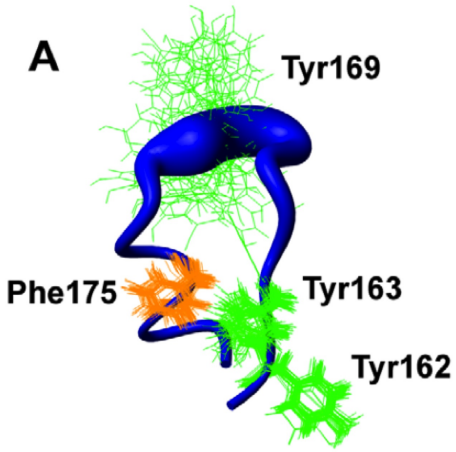


Figure 4.17: HuPrP(90231, M129, Q212P) - pdb id 2KUN; Figure 5 A taken from Ilk et al. (115).

5

Common structural traits across all prion-disease linked mutations.

5.1 Introduction.

In the previous chapter, we have shown that two pathological mutations, HuPrP(Q212P) and HuPrP(E200K) impact on the flexibility of β_2-α_2 loop, on the rearrangement of Tyr^{169} which belongs to that region, and, finally, on the flexibility of the α_2-α_3 region of the globular domain. The latter two are among most flexible regions of the protein (138; 139; 154), (247).), with the largest tendency towards oligomerization in vitro (2; 45). This has lead to the hypothesis that pathogenic conversion process could be related to the global conformational fluctuation of PrP^C (138; 139).

Based on all of these facts, we ask ourselves a key question: Do then all of disease-linked mutations have similar fingerprints as those of HuPrP(Q212P) and HuPrP(E200K)? We have attempted at addressing this issue by carrying out MD simulations on all mutations in the globular domain using the highly efficient computational protocol presented in the previous chapter.

5.2 Incidence and location of mutations of the structure

Let us first review again the disease-linked mutations involving residues of the globular domain of the PrP^C, for which experimental structural information is available. We report the incidence

of all of them (134) and we show their localization in the PrP^C. Some pieces of this information have been already reported in Chapter 2.

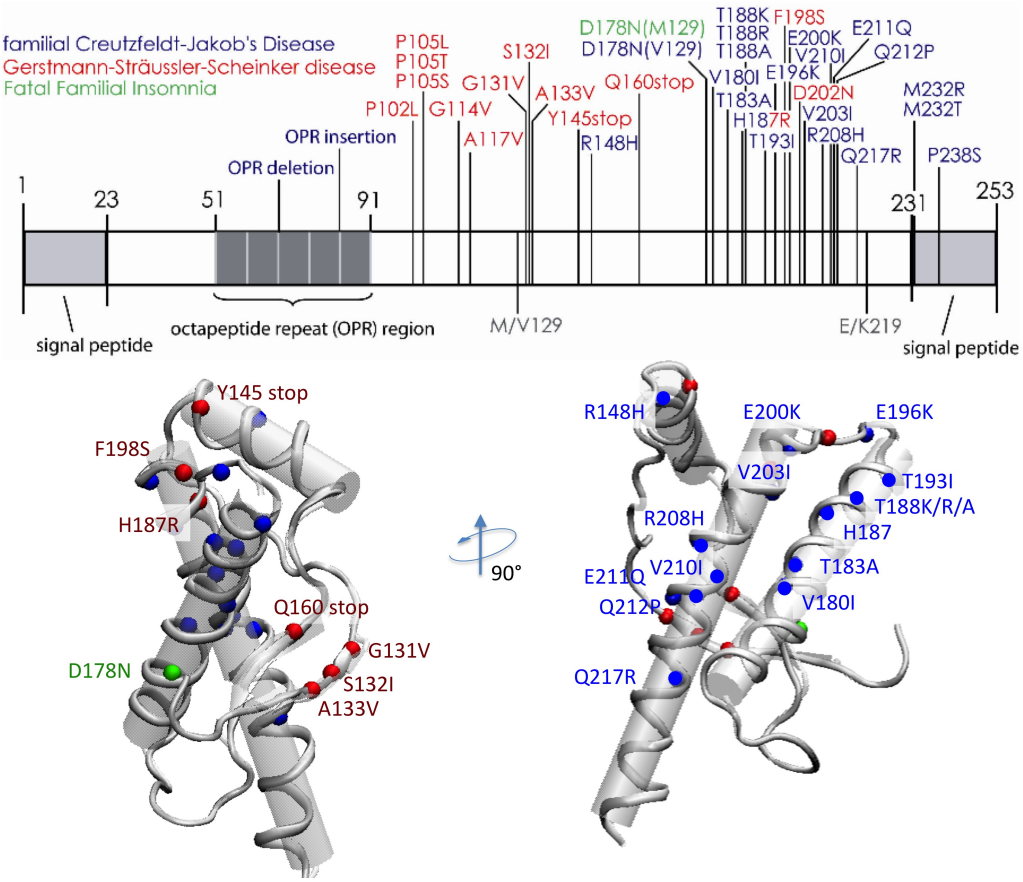


Figure 5.1: Hu PrP Mutants - Location of the mutations in the the globular domain of PrP^C.

More than 30 mutations have been reported (230). The available data are standardized data on TSEs between 1993 and 2002 in the framework of the EUROCJD collaborative surveillance project (134). EUROCJD reports 12 mutations in the globular domain and 2 mutants belongs to N-terminal region. It misses only mutations which very rare and/or are restricted to specific populations (133). The commonest mutation is HuPrP(E200K), followed by HuPrP(V210I) and HuPrP(D178N-129M) (Tab. 5.1). Two mutations involve residues located in 1 (Y145stop and R148H), three in the 1 (G131V, S132I, A133V) and one in the 2 (Q160stop) (Fig. 5.1). All the other mutations in the globular domain are located in the α_2-α_3 region. In particular, the HuPrP(E200K), HuPrP(V210I) and HuPrP(D178N-129M) ones with the highest incidence are the located there.

5.3 Methods.

The computational protocol used is the same used in the previous chapter, except that we decided to perform a single calculation for each mutant instead of four. Indeed, from one hand we have observed that the changes on passing from one simulation to another are really very small. From the other, performing four simulations for all of the mutants would have been highly expensive. Some of the calculations have been performed by my colleague Xiaojing Cong here in SISSA. Calculated properties. The Connelly Solvent Accessible Surface Area (SASA) (see Method section) was calculated by a standard GROMACS utility, which implements the double cube lattice method (78) with a probe radius of 1.4 A. Since the SASA of Tyr^{169} WT PrP^C is 0.40 ± 0.02 (nm^2), we considered the tyrosine Exposed (E) with SASA $\geq$ 0.5 nm^2, Very Exposed (VE) with SASA $\geq$ 0.6 and Outside (O) when SASA$\geq$ 1 nm^2 and Tyr^{169} goes completely outside of the β_2-α_2 loop. The distance between Tyr^{169} and Phe^{175} is defined here as the distance between the centers of mass of atoms belonging to the aromatic rings. That between Tyr^{225} and Met^{166} is defined here as the distance between that between Try^{225} aromatic ring and Met^{166} sulfur. The root mean square fluctuation (RMSF) of atomic positions was calculated with respect to the HuPrP(WT) structure at pH 7 (247).

5.4 Result and Discussion.

Here we investigate whether the key structural facets observed in the diseases linked mutations in the previous chapter (exposure of Tyr^{169}, Root Main Square Fluctuations (RMSF) of β_2-α_2 loop and α_2-α_3 regions) emerge also for all of the other mutants.

5.4.1 Tyr^{169} Exposure.

Tyr^{169} is located in a surface epitope formed by the β_2-α_2 loop and parts of helix α_3 Fig. 5.2.

In the previous chapter, we have shown that (i) in WT and in the protective polymorphism HuPrP(E219K), the Tyr^{169} side chain points toward the bulk of the protein, forming a π-stacking interaction with Phe^{175} (ii) In the two pathogenic mutants E200K and Q212P Tyr^{169} side chain is solvent exposed. The Phe^{175}- Tyr^{169} π-stacking interactions are lost. Also all of the other variants with reported high incidence besides HuPrP(E200K) (HuPrP(V210I), HuPrP(D178N-129M) (134)) feature large rearrangements of Tyr^{169} (Tab. 5.1 and Fig. 5.2-B). In addition, the Phe^{175}- Tyr^{169} π-stacking interactions are lost. This leads to a complete exposure of this residue to the solvent. The same features, albeit less pronounced, are observed for all of the other disease-linked mutations. Tyr^{169} is more exposed to the solvent relative to WT (Tab. 5.1 and Fig. 5.2-C). The Phe^{175}-Tyr^{169} π-interactions are lost for the mutants which experience very large solvent exposure of Tyr^{169} (VE in 5.1). Interestingly, the Tyr^{169}-bearing epitope was previously suggested to be a recognition site in a mouse/human chimera PrP^C for a potential chaperone, so far not identified. This chaperone assists the conversion of PrP^C into PrP^{Sc} (124; 225). Hence, the rearrangement of this residue might affect binding to this

5. COMMON STRUCTURAL TRAITS ACROSS ALL PRION-DISEASE LINKED MUTATIONS.

SYSTEM	**Incidence**	**Available Structure**	∗∗ **Tyr169 Esposure**	π-stacking **Tyr169-Phe175** nm	Aromatic-Sulfur **Tyr225 - Met166** nm
WT	-	NMR (38; 252)	Not Exposed	0.45±0.02	0.49±0.07
G131V-129M	-	0.2	E	0.52±0.02	0.79±0.08
S132I-129M	-	-	VE	0.57±0.07	0.48±0.08
A133V-129M	-	-	E	0.48±0.03	0.78±0.29
R148H-129M	-	-	E	0.48±0.02	0.59±0.17
D178N-129M	14.1	∗ X Ray Dimer (145)	O	1.13±0.02	0.56±0.20
D178N-129V	3.5	∗ X Ray Dimer (145)	O	1.05±0.02	0.56±0.20
V180I-129M	0.7	-	VE	0.61±0.11	0.93±0.26
T183A-129M	-	-	VE	0.74±0.03	0.66±0.14
T183A-129V	-	-	O	1.12±0.11	0.98±0.20
H187R-129V	-	-	O	0.71±0.10	0.68±0.17
T188K-129M	-	-	E	0.52±0.03	0.80±0.23
T188A-129M	-	-	VE	0.73±0.04	0.98±0.27
T188A-129V	-	-	VE	1.12±0.05	0.45±0.05
T188R-129V	-	-	E	0.53±0.03	0.59±0.14
T193I-129M	-	-	E	0.51±0.08	0.60±0.17
E196K-129M	1.1	-	E	0.49±0.03	1.04±0.30
F198S-129V	-	∗ X Ray Dimer(145)	O	0.87±0.05	0.62±0.20
E200K-129M	38.5	NMR(252)	O	0.84±0.07	0.60±0.20
D202N-129M	-	-	VE	0.61±0.12	0.65±0.28
D202N-129V	-	-	E	0.51±0.02	0.53±0.09
V203I-129M	1.1	-	E	0.56±0.06	0.83±0.44
R208H-129M	0.4	-	E	0.48±0.03	0.62±0.07
R208H-129V	-	-	E	0.59±0.08	0.50±0.11
V210I-129V	15.2	-	O	1.12±0.07	0.70±0.20
E211Q-129M	0.9	-	VE	0.73±0.03	0.90±0.30
Q212P-129M	-	NMR(115)	O	0.86±0.05	0.55±0.30
Q217R-129V	-	-	E	0.48±0.02	0.49±0.05
E219K-129M	-	-	Not Exposed	0.44±0.04	0.49±0.08

∗ The structure is in a dimeric conformation (145) ∗∗ Exposed (E) with SASA $\geq$ 0.5 nm^2, Very Exposed (VE) with SASA $\geq$ 0.6 nm^2 and Outside (O) when SASA $\geq$ 1 nm^2 and Tyr169 is outside of the 2-2 loop as in Fig. 5.2

Table 5.1: Table of properties in HuPrP Mutants - Disease-linked mutations for which structural information is available either from NMR and X-ray studies or from the present MD study. The latter is the MD average structure. Several properties are reported: (i) Incidence; (ii) Structure from which data are extracted from; (iii) Tyr^{169} Exposure to the solvent; (iv) Tyr^{169}-Phe^{175} π-stacking distance; (v) Tyr^{225} ring - Met^{166} sulfur distance.

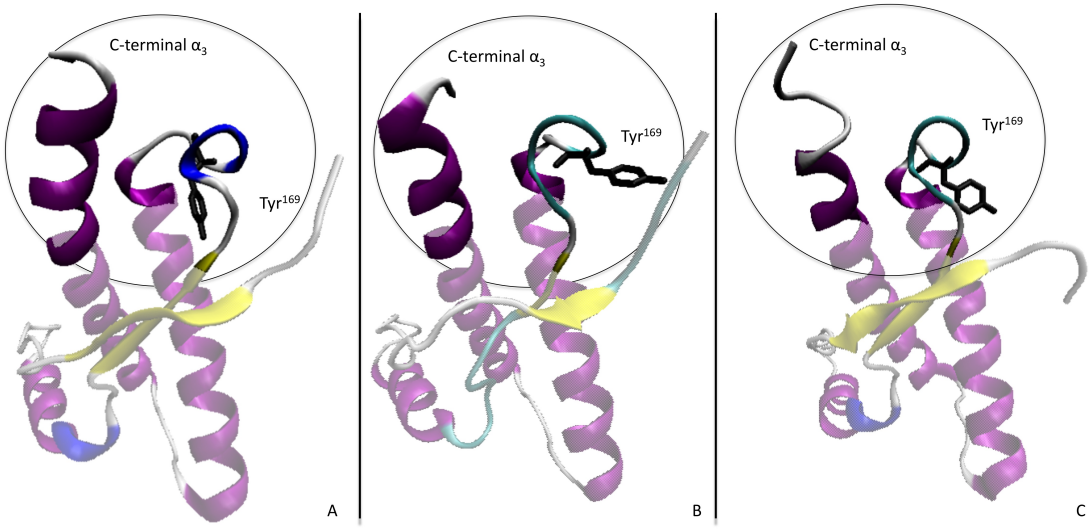


Figure 5.2: Binding Epitope - Epitope formed by the β_2-α_2 loop and parts of helix α_3. A) WT. B) HuPrP(E200K). Similar conformations are seen in HuPrP(D178N-129M/V), HuPrP(T183A), HuPrP(H187R), HuPrP(F198S), HuPrP(E200K), HuPrP(V210I), HuPrP(Q212P). C) HuPrP(Q217R). Similar conformations are seen in all remaining mutants.

chaperone and conversion into PrP^{Sc} in all-disease linked mutations. At the speculative level, we add here that such change of binding might play a role for the experimentally observed increase in fibrillation of such variants.

5.4.2 Flexibility of the β_2-α_2 loop and of the α_2-α_3 region.

The flexibility of these region larger in all the disease-linked mutants than in WT [1].

The increased flexibility of α_2-α_3 loop is in fact (not-surprisingly) accompanied with a change of position of the α_2 and α_3 helices -the key region of the protein for fibrillation (45)) - relative to WT. Indeed, in most of the mutants [2], Tyr^{225}, belonging to helix α_3, does not form contacts with Met^{166}, belonging the β_2-α_2 loop. Consequently, part of helix α_3 is opened and accessible to solvent (Fig. 5.2). This feature is consistent with what found in the NMR structure of HuPrP(Q212P-129V) (115). Because, the α_2-α_3 helices are highly fibrillogenic (2; 45) and their fold is highly stabilized by the protein frame (73; 74), we suggest here that the changes in structure and conformational fluctuations of the α_2-α_3 loop and α_2, α_3 helices might be an important factor for the PrP^C tendency to oligomerize.

[1] Interestingly also in the protective polymorphism HuPrP(E219K) the flexibility of the β_2-α_2 and α_2-α_3 loops is larger respect the WT. Nevertheless, here the rearrangement of these region is completely different from the ones characterizing the pathogenic mutants (see 4.10)

[2] In particular, in all of those variants associated with the largest incidence

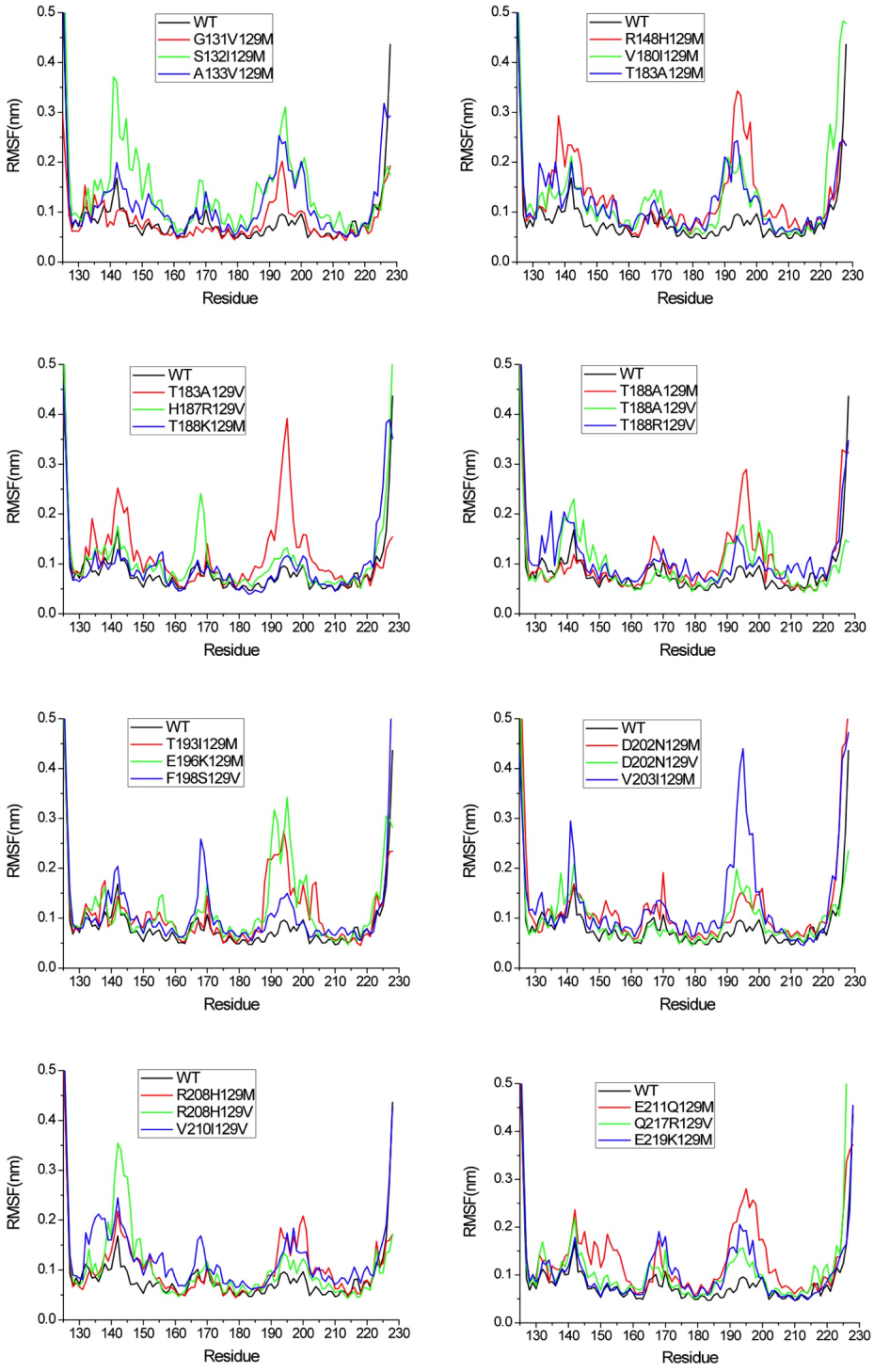


Figure 5.3: RMSF of HuPrP Mutant - RMSF of HuPrP(WT), HuPrP(E219K) (protective polymorphism), and all known disease-mutants calculated for each residue in the globular domain. See Figure 4.3

The increased flexibility of the β_2-α_2 loop might also play a role for PrP^C fibrillation properties. Indeed, the susceptibility towards prion diseases has been correlated with the plasticity of the epitope formed by β_2-α_2 loop and the C-terminal part of the α_3 helix (See (54) and Appendix C).

5.5 Concluding remarks.

Taken together, our calculations carried out on as many as 28 disease-linked mutations, show that mutations are associated with three structural features, which could play an important role for their known fibrillation properties in vivo. We have shown that HuPrP diseases-linked mutations in the globular domain are associated with the same structural fingerprints: from on one hand, they alter the conformation of the β_2-α_2 loop region and modify the exposed hydrophobic surface with the exposure of the Tyr^{169}. Form the other hand they alter the long-range interaction in the α_2-α_3 region. All of these features may play a role for prion/cognate protein recognition (124; 225), and in turn for the fibrillation properties of PrP^C.

6

Docking Ligands on Protein Surfaces: The Case Study of Prion Protein.

Molecular docking of ligands targeting proteins undergoing fibrillization in neurodegenerative diseases is difficult because of the lack of deep binding sites. Here we extend standard docking methods with free energy simulations in explicit solvent to address this issue in the context of the prion protein surface. We focus on a specific ligand (2-pyrrolidin-1-yl-N-[4-[4-(2-pyrrolidin-1-yl-acetylamino)-benzyl]-phenyl]-acetamide), which binds to the structured part of the protein as shown by NMR (Kuwata K, et al. (2007) Proc Natl Acad Sci U S A 104:11921-11926). The calculated free energy of dissociation (7.8 ± 0.9 kcal/mol) is in good agreement with the value derived by the experimental dissociation constant (K_d= 3.9 μM, corresponding to Δ G^0 = -7.5 kcal/mol). Several binding poses are predicted, including the one reported previously. Our prediction is fully consistent with the presence of multiple binding sites, emerging from NMR measurements. Our molecular simulation-based approach emerges therefore as a useful tool to predict poses and affinity of ligands binding to protein surfaces.

6.1 Introduction

Recent developments in molecular docking protocols (MDPs) allow one to predict accurately ligand poses in their target binding sites (11; 129). In several cases, the reason of their success lays in the coupling of traditional scoring function-based approaches with molecular simulation approaches (9) (such as soft harmonic modes (160), molecular dynamics simulations (158; 185) and relaxed complex method (151; 161)). The latter introduce conformational flexibility of the target, accounting for the fact that proteins are in constant motion between different conformational states (40). These may be locally altered when a ligand is bound (40).

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 is common to proteins undergoing fibrillization in neurodegenerative diseases. Here we propose an enhanced molecular docking protocol (EMD in Figure 6.1) that extends MDPs with free energy simulations in explicit solvent to predict the structure and energetics of ligands binding to protein surfaces.

Protein misfolding followed by self-association and subsequent deposition has been observed in the brain tissues of patients affected by different neurodegenerative disorders. Diverse proteins have been shown to follow this process, including amyloid-β (in Alzheimer disease), α-synuclein (in Parkinson disease), huntingtin (in Huntington disease), and prion protein (in prion diseases) (35; 44; 216). The protocol we propose here may be exploited in these cases for the design of ligands that, by stacking onto protein surfaces, may disrupt protein-protein interactions and thus inhibit protein self-assembly.

In this study, we apply our EMD protocol in the context of the cellular form of human prion protein ($HuPrP^C$), responsible for TSE development (see ch. 3), for which neither an early diagnosis nor a cure is currently available (228). Therefore, there is great interest in designing ligands binding to $HuPrP^C$, which may interfere with its conversion and with the interaction with other self-aggregating proteins, such as amyloid-β (144).

Recently, the ligand GN8 (2-pyrrolidin-1-yl-N-[4-[4-(2-pyrrolidin-1-yl-acetylamino)-benzyl]-phenyl]-acetamide,**1** (in Figure 6.2) (140) has been shown to bind Mouse PrP ($MoPrP^C$) in the μM range. $MoPrP^C$ is highly similar to the $HuPrP^C$. The sequence similarity is as high as 98% and the root mean square difference (RMSD) of the backbone between the MD structures of $HuPrP^C$ (PDB code: 1HJM) (38) is the same (0.26 $\pm$ 0.02 nm) as that between the NMR structures of $HuPrP^C$ and $MoPrP^C$ (PDB code: 1AG2) (203) (0.27 nm). Therefore, significant changes of the structure on passing from the mouse to the human protein are not expected. For the 1-$MoPrP^C$ adduct, NMR chemical shifts perturbations of $MoPrP^C$ on protein residues induced by ligand binding have been reported (140). These affect most significantly amino acid residues on one side of the protein surface (Arg156, Asn159@H1-S2 loop, Lys194@H2, Glu196, Thr199@H2-H3 loop and Val210@H3). In addition, Val189 and Thr192, located on the other side of the PrPC surface, are also perturbed (Figure 6.4 A). All these residues are conserved on passing from $MoPrP^C$ to $HuPrP^C$. These perturbations have been ascribed to ligand binding, suggesting that multiple binding sites may be present. An ad hoc model of the 1-$MoPrP^C$ adduct, constructed by docking and energy minimization, exhibited a single binding mode of GN8 connecting Asn159 and Glu196 (140). Subsequent quantum-mechanical studies (116), based on this model, pointed out that these two residues, along with Gln160 and Lys194, are important for the binding. However, such single binding mode was not consistent with the presence of the contacts between the ligand and Val189, Thr192 and Thr199.[1]

The proposed EMD protocol turns out to provide structural prediction consistently with

[1]A similar NMR study to map prion protein binding sites has been made only for another ligand, quinacrine (237). The latter, however, has been later suggested not to bind to PrPC but rather to PrPSc, or other chemical chaperons involved in the prion propagation (127).

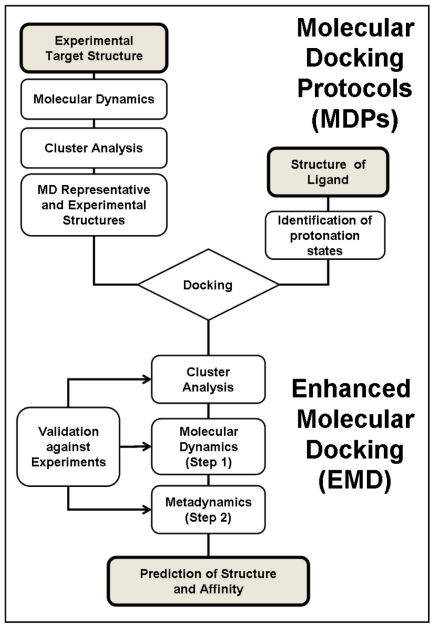


Figure 6.1: EMD - Enhanced Molecular Docking. MDPs are used to guess putative ligand binding regions on target surfaces, based on structural information on the two separated moieties. Structural information on the target may come from experiment and, in some cases, also from molecular simulation. Ligand may be docked onto the entire structure (like in this work) or on a putative binding site. Cluster analysis are used to group MD conformers and/or ligand/target adducts into representative structures. In the EMD approach, MD simulations may be used to relax the structures and to investigate the role of hydration. Enhanced sampling simulation techniques in explicit solvent allow to explore the ligand binding space and to predict free energy of binding. The key ingredient of the protocol proposed in Figure 6.1 is given by the type of free energy approach used. Several powerful methods are available for predicting ligand binding free energies by means of molecular simulation (88; 205). However, predictions have been made so far to targets with binding sites well characterized by X-ray crystallography or NMR experiments. Here we use the metadynamics (141) approach in its bias exchange variant (191). Metadynamic (141) has recently been shown to uncover molecular and energetic aspects protein flexibility (148). This approach provides the free energy as a function of several reaction coordinates (such as geometrical distances, polar contacts, and water-mediated interactions), which characterize the ligand binding to its target and its dissociation from it (33; 82; 87; 232). Although GN8-PrPC interactions energies have been provided by quantum chemical methods (116), no calculation of free energy has been so far reported. As several other techniques (205), it may allow also to simulate the whole molecular recognition process. This in turn may allow characterizing multiple binding sites of the ligand onto the proteins surface, such as those emerging from NMR in the 1-$MoPrP^C$ complex. Comparison against experimental data in this case structural information inferred by NMR chemical shift perturbations as well as with affinity measurements allows to validate the protocol (140)

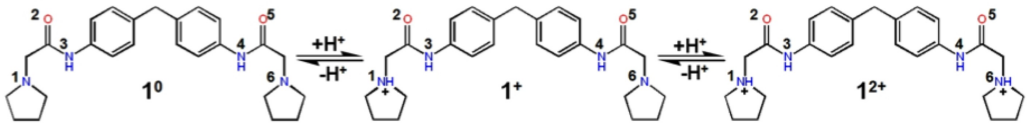


Figure 6.2: GN8 - (2-pyrrolidin-1-yl-N-[4-[4-(2- pyrrolidin-1-yl-acetylamino)-benzyl]-phenyl]-acetamide) molecule.

the NMR data and affinity in agreement with experimental data. The EMD protocol emerges therefore as a useful approach to investigate ligands sticking on protein surfaces.

6.2 Computational Details

6.2.1 Identification of binding sites.

The residues interacting with**1**are located in $HuPrP^C$ C-term, for which the NMR structure is available (residues 125-228, PDB ID: 1HJM) (38). Protonation states were assigned by the web server H++36 assuming pH 7.4. Putative binding sites were identified by (i) molecular simulations (using the GROMACS package (27) and (ii) docking procedure (using the GOLD (121; 234) and the Autodock programs (171)).

(i) Molecular simulations. The protein was inserted into a cubic box of water molecules, ensuring that the solvent shell would extend for at least 0.8 nm around the system. Three sodium counterions were added. The AMBER99 force field (41; 242) was used for the protein. Sodium ions were modeled with the AMBER-adapted Aqvist potential (1). The water molecules were described by the TIP3P model (122). The system was minimized imposing harmonic position restraints of 1000 kJmol-1nm-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 restraints, 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 dynamics was then performed through the Nos-Hoover (110; 180) and Andersen-Parrinello-Rahman (181; 188) coupling schemes. Periodic boundary conditions were applied. The final simulation box equilibrated at around 6.69 x 6.69 x 6.69 nm. Long-range electrostatic interactions were treated with the particle mesh Ewald (PME) (66; 80) method, using a grid with a spacing of 0.12 nm combined with a fourth-order 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 (106) 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 gromos method (67) and as result 20 different conformations were obtained, which were used along with the NMR structure for docking of compound **1**.

(ii) Docking. Titration curves for compound**1**in bulk solution were calculated by the ChemAxon software (48) showing that this molecule is present in two protonation states at pH=7.4: neutral (1^0) and monoprotonated (1^+); while at pH=4.5 it exists mostly in the diprotonated form (1^{2+}) (Figure 6.2 and 6.3).

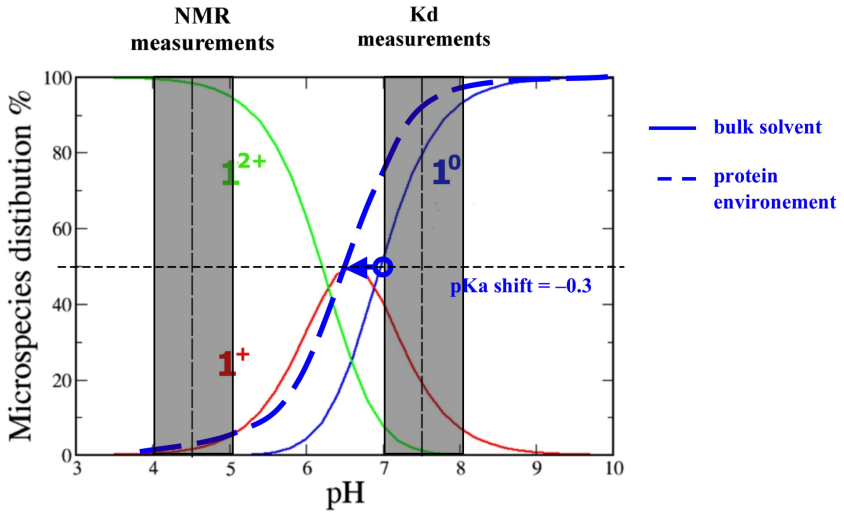


Figure 6.3: Titration curve of GN8 - Titration curve of GN8 as calculated by the ChemAxon software (1^0). At the pH of the Kd measurements (140), GN8 is present in neutral (1^0) and monoprotonated (1^+) states. At the pH of the NMR measurements (140), it exists also in diprotonated (1^{2+}) form. The two pH values are indicated in the figure.

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 (48). Since the acidity of compound**1**is expected to change in the proximity of the protein, electrostatic potential calculations in implicit solvent (18) were addressed for the prion protein in two different conditions: pH=4.5 for NMR measurements conditions and pH=7 for affinity measurements conditions. The electrostatic potential surfaces were calculated using the APBS package (18) and visualized with VMD (114). The calculation parameters were 0.3 grid spacing, 1293 meshes, solvent and protein dielectrics of 78.54 and

40, respectively. All three protomers underwent geometry optimization at the B3LYP/6-31G** level of theory by means of the Gaussian03 software (g03) (85). The optimized structures, 1^0, 1^+ and 1^{2+}, were docked to the NMR structure of $HuPrP^C$ and to its 20 different conformations, as obtained after the cluster analysis of the MD trajectory. The GOLD 3.1 (121; 234) and Autodock 3.0.5 (171) programs were used. In GOLD, the docking area was defined as a sphere of 3.5 nm radius around the His187, so that the whole protein was screened. The ChemScore (CS) (177) and GoldScore (GS) (121) scoring functions were used for ranking. For each protomer and scoring functions, 100 docking runs were performed. In Autodock (171), a Lamarckian genetic search algorithm was used to identify low energy binding sites and orientations of **1** protomers. Binding modes were ranked by a scoring function implemented in the Autodock. A point grid with a spacing 0.0475 nm was used. Point grid was centered to the center of mass of the protein, its dimensions were 12.6 nm x 12.6 nm x 12.6 nm. Gasteiger atom charges were assigned to the protein atoms 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 (31).

6.2.2 Hydration and thermal stability of 1-$HuPrP^C$ adducts.

10-ns MD simulations of the adducts ($HuPrP^C$-1^0, $HuPrP^C$-1^+ and $HuPrP^C$-1^{2+}) allowed for a proper hydration of the system and for identification of collective motions that may be essential for PrPC-ligand interactions. The protomers were bound to the binding region I (Figure 6.4 B). The simulation protocol was the same as for the free protein. For the three ligands, the gaff force field (241; 242) was used. The atomic restrained electrostatic potential (RESP) charges (23; 64) were calculated by using the resp module of AMBER after geometry optimization and electrostatic potential calculations of each protomer at the B3LYP/6-31G** level of theory by means of the g03 software (85).

6.2.3 Dissociation free energy calculations.

The dissociation free energies of 1^0, 1^+ and 1^{2+} were calculated using metadynamics (141) in its bias-exchange variant (191), as a function of collective variables (CVs) which should be relevant for describing the dissociation process. CVs used in this work are: (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 I; (iii) the number of polar contacts between the ligand and the other portion of the protein binding region I; (iv) the number of water bridge contacts between the ligand and the protein binding region I; (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 region I. The choice of these CVs was based on previous ligand-target interaction metadyamics studies (33; 82; 87; 232) as well as by observations based on the former MD simulations (see Appendix E for more details). The calculations do not require in principle the previous knowl-

edge of the protein-ligand adduct structure. However, for computational efficiency we exploit the fact that all the target regions detected from NMR are in the close proximity of region I. Therefore, here we explored only this region. 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 (191). Similar set-up was shown to improve the sampling of the configurational space and the convergence of the results (148; 159; 191; 192; 226). At the end of the different replica simulations, the explored phase space, in terms of the six collective variables used in this study, was clustered using the gromos method (67). 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 (39). Details on this procedure can be found in references](148; 159; 192; 226), and are summarized in SI together with the converged free energy profiles.

Two reference states of the ligand-protein system, bound and unbound, needed to be defined to provide the corresponding dissociation free energy value. The bound state was considered as the lowest free energy cluster. The unbound state was considered to be a cluster showing no contacts with the binding regions 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 dissociation energy of the unbound state was roughly estimated in implicit solvent using an adaptive Poisson Boltzmann solver. The APBS package (18) was used with the same parameters described previously. It was estimated as the difference in solvation energy of the complex minus the solvation energy of each component plus the intermolecular Coulomb interaction. The standard free energy of dissociation was obtained by applying the following relationship: $\Delta\ G^0$=ΔG RT*ln([L]), where ΔG is the total dissociation 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, corresponding to 1 molecule in 6.693 nm3). The standard free energy is related to the dissociation equilibrium constant (K_d) by $\Delta\ G^0$ = RT*ln(K_d).

6.3 Results and Discussion

In this study, we focus on the binding of compound**1**to the surface of $HuPrP^C$ protein. Compound**1**is a symmetric molecule, composed of two pyrrolidine rings connected by acetamides to a diphenylmethane core (Figure 6.2). Two different conditions are considered: at neutral pH, where experimental affinity has been measured (140) and at acidic pH, where NMR chemical shift perturbations (140) have been used to provide information on the amino-acids involved in the binding.

We use the computational protocol summarized in Figure 6.1: (i) Identification of ligand

protonation state at neutral and acidic pH. (ii) Use of MDPs to provide a first guess of putative binding regions (iii). Use of molecular dynamics (MD) simulations to relax the structure in aqueous solution (step 1 in EMD). (iv) Use of metadynamics to predict the energetics of binding of**1**to the protein as well as (v) the binding poses of the compound (step 2 in EMD).

Protonation state of compound 1. This compound can exist in different protonation states, in which none, one or both the tertiary nitrogen atoms of the pyrrolidine rings are protonated (Figure 6.3). At pH=7.4, at which the K_d has been measured, approximate pKa calculations based on ref. 31 (See Methods for details) allow us to suggest that, in water,**1** is present mainly not only in the neutral form (1^0) but also in the monoprotonated one (1^+) (Figure 6.3). In the latter, one of the two pyrrolidine nitrogen atoms is protonated.

At pH=4.5, at which the NMR experiments were performed, the same calculations lead us to the conclusion that, in water, the ligand is mainly diprotonated (1^{2+}), with both pyrrolidine nitrogen atoms protonated. Small amounts of 1^+ are also present. The calculated concentration of 1^0 in bulk water is very low (Figure 6.3). However, one should keep in mind that ligand-protein binding does not occur in pure water and that the influence of the electrostatic field of the protein should be accounted for. Indeed, simple electrostatic potential calculations (see Methods for further details) show an increase of the positive charge density in the region of the protein defined by the NMR contacts (Figure E.2 in Appendix E). This suggests that the protein environment will favor there the accumulation of neutral 1^0. Therefore, binding poses involving the neutral form should be considered even at acidic pH. Based on these results, we performed calculations on all of the three protomers.

Binding regions of $HuPrP^C$ emerging from MDP. The three protomers were docked independently to the $HuPrP^C$ NMR structure and to 20 different conformers obtained from a 20-ns molecular dynamics (MD) simulations of the protein in aqueous solution. The putative binding regions I, II and III were identified (Figure 6.4 B). I is defined by the H2 helix and the loop connecting β-sheet S2 and helix H1. II consists of the H2-H3 helices. III is defined by the H3 helix, the N-term of H2 helix, and the loop between H1 helix and S1 β-sheet. Binding region I is the only site which involves residues changing chemical shifts upon binding of compound**1**and it is closer to all the other residues involved in the binding (140). It was therefore the only one selected for subsequent free energy studies.

$\Delta\delta > 0.9ppm$	$0.9ppm > \delta > 0.5ppm$
Asn159	Arg156
Val189	Glu196
Thr192	Thr199
Lys194	Val210

Table 6.1: NMR chemical shift -NMR chemical shift changes (δ) upon GN8 binding to the MoPrP^C

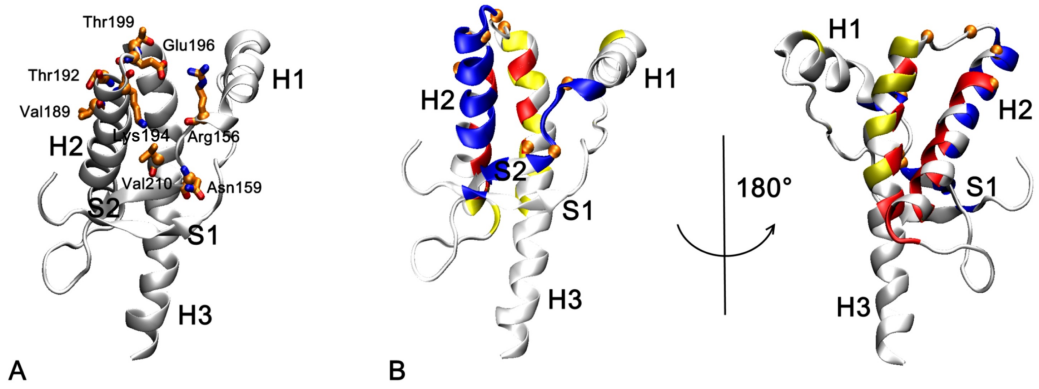


Figure 6.4: Binding Regions - (A) Residues involved in GN8 binding to the prion protein (in licorice), as emerging from chemical shift changes (140). (B) Three different binding regions (I, II and III, shown in blue, red and yellow, respectively), as obtained after MDP procedures (see Figure 6.1). Orange spheres represent compound **1** binding aminoacids defined by NMR chemical shift study. The Figure shows $HuPrP^C$. This is very similar to the $MoPrP^C$ (Sequence Similarity = 98 %), for which experiments have been carried out.

The adducts for each of the three protomers docked at binding region I underwent 10ns of MD calculations in aqueous solution. The ligands maintained completely (1^+ and 1^{2+}) or partially (1^0) the pose identified in the docking (See Appendix E for details). Most importantly, the structural determinants of the three protomers turned out to be consistent with most ligand-protein contacts identified by NMR (Table 6.1). However, the ligand-protein contact with Val189, Thr192, Thr199 and Val210 could not be predicted. The simpler docking approach, combined with energy minimization of protomer 1^0 provided similar results (Figure. 3 in Kuwata et al. (140)).

Free energy calculations were used to explore the ligand binding space in explicit solvent These simulations allowed to identify alternative binding poses for each protomer of the ligand and, for 1^0 and 1^+ , to predict the dissociation free energy. The free energy simulations were performed as a function of 6 collective variables that took into account rearrangements of the ligand and protein, hydrogen bond contacts and water bridges (see Methods). These variables have been already used to characterize ligand-target molecular recognition processes using the metadynamics approach (33; 82; 87; 232).

The $HuPrP^C$-1^0 complex. In the lowest free energy cluster identified by the metadynamics calculations, 1^0 is located in the wide cleft formed by helices H1, H2 and H3 (1^0.B1 in Figure 6.5 A), similarly to the model proposed by Kuwata et al. for $MoPrP^C$ (140). The

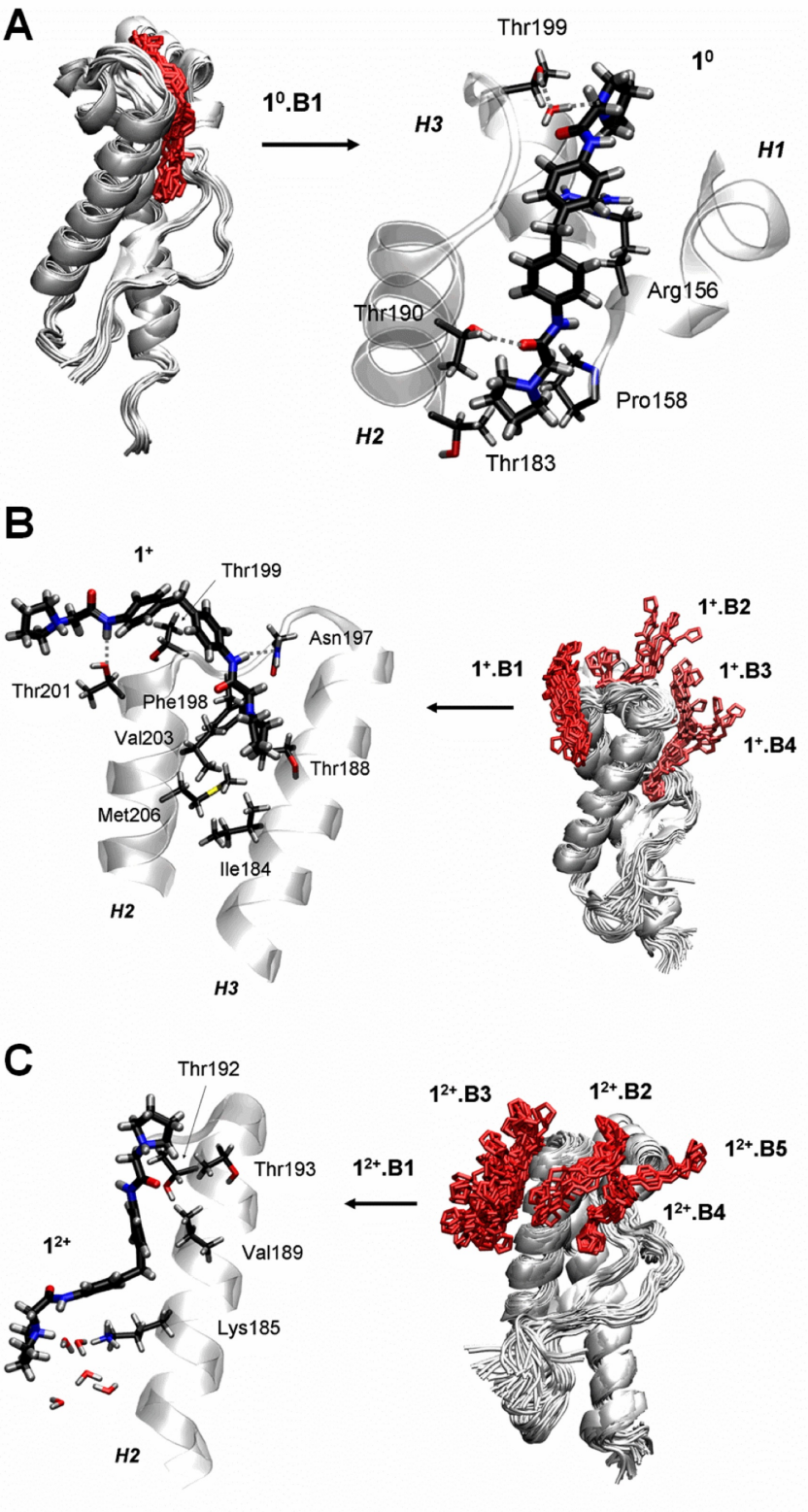


Figure 6.5: Three dimensional structures of HuPrPC in complex with 1 protomers - (A) 1^0, (B) 1^+ and (C) 1^{2+}. These structures correspond to the bound-states free energy minima (1^0.B1, 1^+.B1-B4, 1^{2+}.B1-B5), as calculated with the metadynamics method (see Methods). Close-ups on the ligands and the binding sites are also shown.

contacts 1^0 forms with the $HuPrP^C$ are consistent with the reported chemical shift changes on Glu196, upon GN8 binding (Table 6.1), as well as with a recent quantum chemical study (116). They are also consistent with chemical shift changes on Arg156, Thr199, and Val210 upon GN8 binding (Table 6.1). The phenyl groups of 1^0 form a π -cation interaction with Arg156 and a water-mediated H-bond is present between Thr199 and the pyrrolidine nitrogen (N1; Figure 6.2). The pyrrolidine ring forms hydrophiobic interactions with Val210 (as well as with Pro158, Thr183). The $HuPrP^C$-1^0 complex is further stabilized with a direct H-bond between Thr190 and the carbonyl group of 1^0 (O2; Figure 6.2). The unbound state of $HuPrP^C$-1^0 system corresponds to a conformation in which the ligand has no contact with the protein. The conformation of Lys194 changes upon ligand dissociation (Figure 6.6).

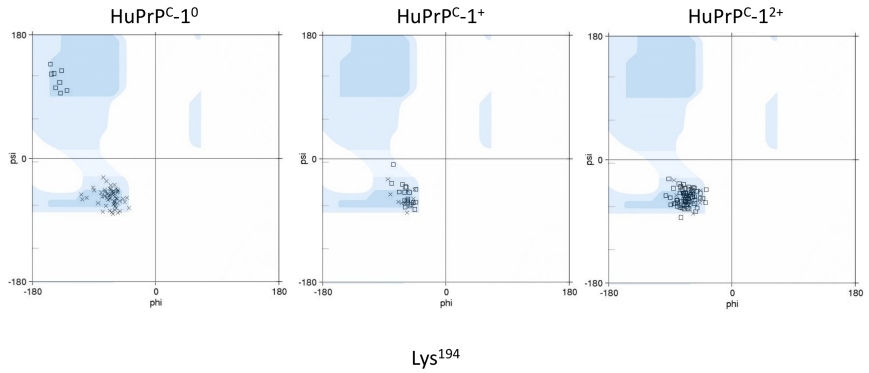


Figure 6.6: Lys194 - Ramachandran plot of residue Lys194 for the three adducts: Crosses indicate the conformation of the residue when the ligand is bound, and empty squares when the ligand is not bound.

This is consistent with the significant chemical shift change reported for this amino acid (140). This contrasts with a recent quantum chemical investigation, which points to the role of Lys194 for the bridging conformation of the GN8/PrPC complex (116). This discrepancy may be due to the fact that here we consider free energies in solution, whilst ref. 20 presents interaction energies in vacuo. Smaller conformational changes were also observed for other residues present in the H2-H3 loop (res. 195-199 Appendix E, Figure

The $HuPrP^C$-1^+ complex. Four different stable conformations of 1^+ were identified on $HuPrP^C$ surface (Figure 6.5 B). In the lowest free energy state, 1^+ lays along the loop connecting helices H2 and H3 (1^+.B1 in Figure 6.5 B). It forms a remarkable hydrophobic interaction with Thr199, consistently with the chemical shift changes reported for this residue (140). The amidic nitrogen atoms of 1^+ (N3 and N4; Figure 6.2) 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 (Table 6.1). Additionally, the neutral pyrrolidine ring of

1^+ is kept by the hydrophobic cleft formed by Ile184, Thr188, Phe198, Val203 and Met206 further stabilizing the complex. No water-mediated interactions were observed between 1^+ and $HuPrP^C$. The free energy difference between the bound state of $HuPrP^C$-1^+ (1^+.B1 in Figure 6.5 B) and the corresponding unbound state, with the corrections described above, turns out to be 8.6 kcal/mol. This is similar to that predicted for 1^0 and in good agreement with experimental data.

The $HuPrP^C$-1^{2+} complex. Five different stable conformations of 1^{2+} were identified on $HuPrP^C$ surface (Figure 6.5 C). In the most stable conformation, 1^{2+} binds yet in another position of $HuPrP^C$, laying along helix H2 (1^{2+}.B1 in Figure 6.5 C). Half part of 1^{2+} is in close contact with $HuPrP^C$ surface in the cleft formed by Val189, Thr192 and Thr193. Indeed, these positions were reported to interact directly with the ligand according to NMR experiments (Table 6.1). Two layers of water molecules are present between the protein surface and the rest of the molecule, presumably due to the presence of Lys185. In the other accessible conformations, 1^{2+} covers different regions of the protein surface (See Appendix E). The interaction with Lys194 is conserved in the majority of them. This result is consistent with the chemical-shift changes of this residue upon ligand binding (140). The dissociation free energy of 1^{2+} was not calculated as, according to our calculations based on pKa estimations, this protomer is not present at the conditions in which the K_d was measured (140).

7

Conclusion

TSEs arise with the post-translational conversion of the ubiquitous cellular form of the prion protein, PrP^C, into its pathogenic form, PrP^{Sc}, without any detectable covalent modifications. There is evidence that the $PrP^C \rightarrow PrP^{Sc}$ conversion initiates an autocatalytic reaction, which leads to the accumulation of amyloid in the central nervous system and, through still unknown mechanisms, to neurodegeneration (147; 238; 240).

In the presence of pathogenic mutations, the $PrP^C \rightarrow PrP^{Sc}$ conversion occurs spontaneously (13; 220; 223). Therefore these mutations might increase the likelihood of misfolding by thermodynamic destabilization of PrP^C (13; 220) (Fig. 1.2) [1]

In this thesis, we have investigated PrP^C disease-linked variants and the adduct with a ligand by molecular simulation methods. Our goal has been to characterize the influence of disease-linked mutations on the structural determinants as well as to build up a computational protocol able to predict poses and energetics of ligands binding to PrP^C.

First, we have focused on the destabilizing effect of all pathogenic mutations on PrP^C globular domain in the early stage of the conversion.

MD calculations, based on the NMR structure(247), reproduce the structural determinants of HuPrP(E200K) (252) mutant and HuPrP(Q212P), (115), for which NMR structural information are available. This establishes the reliability of our computational protocol. Our calculations are then carried out on all of the pathogenic mutants of PrP^C globular domain (overall 28 variants, Fig. 5.1). All the variants turn out to features common traits, independently from position and chemical nature of the mutations. These features are: (i) an increase of the flexibility of the β_2-α_2 loop, of the α_2-α_3 region and a larger solvent exposure of Tyr169 (which belong to the β_2-α_2 loop) relative to the wild type prion protein. Strikingly, the flexibility of β_2-α_2 and α_2-α_3 is related to pathogenicity of TSEs (2; 45; 54), and it might also affect the interactions of PrP^C with cellular partners during fibrillation (124; 225). Indeed, the β_2-α_2 loop and parts of helix α_3 form an epitope that might serve as recognition site for

[1] Remarkably, a class of antiprion compound has been discovered which may act stabilizing PrP^C structure (30).

a potential chaperone assisting PrP^C →PrP^{Sc} conversion (124). In this respect, also the rearrangement of Tyr169 exposure might play a crucial role for fibrillation. Its exposure toward the solvent modifies the hydrophobic surface of the protein and alters π-staking interactions in the recognition site. Remarkably, in all of the pathogenic mutants with high incidence on the population (namely HuPrP(E200K) (HuPrP(V210I), HuPrP(D178N-129M) (134), Tyr169 features the largest degree of solvent exposure. In addition, part of helix α_3 is opened and accessible to solvent (Fig. 5.2).

We thus identify some hot spots in PrP^C structure, rather sensible to mutations that could give crucial hints in understanding the early stage of $PrP^C \rightarrow PrP^{Sc}$ conversion. The knowledge of PrP^C regions most susceptible to the effect of pathogenic mutations might provide important hints for a molecular-based drug design approaches (228).

Next, I have contributed to the development of a docking procedure able to characterize binding of small ligands on shallow binding sites, where multiple binding pose are possible, such as in the case of prion protein. This is challenging for standard molecular docking procedures, which might encounter difficulties in ligands binding to (multiple) shallow binding sites. We further notice that in docking procedures which take into account protein flexibility, (11; 36; 47; 227), probably very important here, might provide a significant number of false-positive solutions (11).

Here, we have established a computational approach that combines standard docking methods with molecular dynamics and metadynamics-based free energy simulations (141) (Fig. 6.1). We have focused on a specific ligand (2-pyrrolidin-1-yl-N-[4-[4-(2-pyrrolidin-1-yl-acetylamino)-benzyl]-phenyl]-acetamide), which may stabilize PrP^C, binding to the structured part of the protein (as shown by NMR (141)).

Our procedure is able to account for protein flexibility and it provides detailed information about the binding process, binding affinity and ligand-target interactions. Applying our procedure on PrP^C, we found a multiple-pose binding pattern for the examined compound that could not be obtained by applying only standard protocols. Three binding poses are identified, in complete agreement with NMR data (140). Most importantly, the predicted dissociation free energy turned out to be in very good agreement with experimental data (140).

Our protocol could be now used to predict the potency of ligands interacting on protein surfaces or with target proteins without a unique binding site. These include most proteins involved in neurodegeneration possessing a structured moiety (20; 51; 81; 104; 175; 221; 229).

List Of Publications

Papers related to the thesis:

G. Rossetti, G. Giachin, G. Legname, P. Carloni, Structural facets of disease-linked human prion protein mutants: A molecular dynamic study Proteins, (Jul 23, 2010).

A. Kranjc, S. Bongarzone, **G. Rossetti**, X. Biarnes, A. Cavalli, M. L. Bolognesi, M. Roberti, G. Legname, P. Carloni, Docking Ligands on Protein Surfaces: The Case Study of Prion Protein J. Chem. Theory Comput. 5, 2565 (Sep 8, 2009).

Other papers:

G. Rossetti, A. Magistrato, A. Pastore, P. Carloni, Hydrogen Bonding Cooperativity in polyQ -Sheets from First Principle Calculations Journal of Chemical Theory and Computation 6, 1777 (Jun 8, 2010).

G. Rossetti, A. Magistrato, A. Pastore, F. Persichetti, P. Carloni, Structural properties of polyglutamine aggregates investigated via molecular dynamics simulations J Phys Chem B 112, 16843 (Dec 25, 2008).

D. Flock, **G. Rossetti**, I. Daidone, A. Amadei, A. Di Nola, Aggregation of small peptides studied by molecular dynamics simulations Proteins 65, 914 (Dec 1, 2006).

Appendix A

Human Prion Diseases

A.1 Human Prion Diseases

Prion diseases are also referred as transmissible spongiform encephalopathies (TSE). They occur in humans and animals, primarily affecting the central nervous system. They can be sporadic (spontaneous), familial (genetic/inherited) or acquired (transmitted by infection). The hallmark of these diseases is the presence of microscopic vacuolization of the brain tissue, called spongiform degeneration (meaning that the tissue deteriorates, developing a spongy texture), and an abnormal protein, called scrapie prion protein (PrP^{Sc}), prion or abnormal prion protein.

Humans are susceptible to several prion diseases that include:

- CJD: Creutzfeld-Jacob Disease.

- GSS: Gerstmann-Straussler-Scheinker syndrome

- FFI: Fatal familial Insomnia.

- Kuru

These original classifications were based on a clinical evaluation of a patients family history symptoms and are still widely used, however more recent and accurate molecular diagnosis of the disease is gradually taking the place of this classification. The incidence of sporadic CJD is about 1 per million per year. GSS occurs at about 2% of the rate of CJD. It is estimated that 1 in 10,000 people are infected with CJD at the time of death. These data are likely to be underestimates since prion diseases may be misdiagnosed as other neurological disorders. The diseases are characterized by loss of motor control, dementia, paralysis wasting and eventually death, typically following pneumonia.

A.1.1 Polymorphisms

All prion disease, inherited forms, iatrogenic or sporadic, are influenced by an aminoacis polymorphism [1] resulting in a Methionine (M) to Valine (V) substitution at PrP codon 129 (184).

This polymorphism is crucial to the etiology and neuropathology of prion disease; it is found that homozygosity at codon 129 predisposes individuals to CJD (186) and to date, only individuals homozygous for methionine (Met/Met) have succumbed to variant CJD (60; 108), and iatrogenic CJD occurs predominately in homozygotes with an excess of valine homozygotes in cases related to exposure to contaminated human pituitary hormones (62).

The importance of sequence homology is further emphasized by the observation that in some families with inherited prion diseases, the age of onset is significantly later in individuals heterozygous to this polymorphism (61; 75; 195). Thus, heterozygosity with respect to this polymorphism appears to confer substantial resistance to prion disease.

Elderly survivors of the kuru epidemic (an acquired prion disease largely restricted to the Fore linguistic group of the Papua New Guinea Highlands, which was transmitted during endocannibalistic feasts) who had multiple exposures at mortuary feasts, are, in marked contrast to younger unexposed Fore, predominantly PRNP 129 heterozygotes. Kuru imposed strong balancing selection on the Fore, essentially eliminating PRNP 129 homozygotes.

Worldwide PRNP haplotype diversity and coding allele frequencies have suggested that strong balancing selection at this locus occurred during the evolution of modern humans (165).

A second polymorphism resulting in glutamate (E) to lysine (K) substitution at codon 219 has been identified in the Japanese population, and it has also been reported to have an effect on the susceptibility of individuals to prion disease (128).

A.1.2 Inherited Prion Diseases

Familial TSEs are associated with an autosomal dominant [2] PRNP gene alteration. Depending on the particular PRNP mutation involved, there is variability in the clinical and pathological findings, the age of onset, and the duration. Familial TSEs account for 1020% of all TSE cases in humans and include fCJD, GSS, and FFI.

The clinical categories may be seen as phenotypic extremes of a continuum (5), because in reality the syndromes overlap considerably (165).

fCJD have been associated with point mutations affecting the region between the second

[1] Polymorphism is a discontinuous genetic variation that results in different forms or types of individuals among the members of a single species. Genetic polymorphism promotes diversity within a population. It often persists over many generations because no single form has an overall advantage or disadvantage over the others regarding natural selection. A common example is the different allelic forms that give rise to different blood types in humans.

[2] Autosomal dominant is one of several ways that a trait or disorder can be passed down through families. If a disease is autosomal dominant, it means you only need to get the abnormal gene from one parent in order for you to inherit the disease. One of the parents may often have the disease.

and the third helix of the C terminus region, insertions in the octarepeat region, in the N terminus, and even a premature termination codon at position 145. The inheritance is, in all cases, autosomal dominant, often with very high penetrance (133).

GSS syndrome is a rare, inherited autosomal dominant disease that is associated with mutations in the PRNP gene. GSS syndrome is characterized by chronic progressive ataxia, terminal dementia, a long clinical duration (210 yr), and multicentric amyloid plaques that can be visualized by antibodies directed against the prion protein.

FFI is characterized by sleep disturbances as well as vegetative and focal neurological signs as a result of thalamic lesions. The clinical phenotype depends on the D178N point mutation of the PRNP gene coupled with a methionine at codon 129.

Appendix B

Biophysical Mechanisms of Prion Toxicity

B.1 Models describing the replications of prions

According to the *protein only hypothesis*, there are at least two models to explain the conformational conversion of PrP^C into PrP^{Sc} and prion propagation:

- The heterodimer refolding mechanism, also known as template-assistance model or monomer-directed conversion (198);

- The nucleated polymerization mechanism (117).

These models differ with respect to the role of ordered prion protein oligomers in the conversion reaction.

The heterodimer refolding mechanism model postulates that PrP^C is thermodynamically less stable than PrP^{Sc} but spontaneous conversion to PrP^{Sc} is kinetically limited. A critical step in the conversion would be the formation of a heterodimer between PrP^C and PrP^{Sc} monomers. PrP^{Sc} in this complex would act as a template, inducing a conformational transition of PrP^C. Here, oligomerization is a consequence, not a cause, of conversion.

A different view is presented by the nucleated polymerization mechanism model, according to which the infectious species is not the PrP^{Sc} monomer but a PrP^{Sc} aggregate. This model postulates that monomeric PrP^C and PrP^{Sc} exist in an equilibrium far displaced toward PrP^C. Stabilization of PrP^{Sc} occurs only upon formation of an oligomer large enough to act as a stable nucleus. Monomeric PrP^C would subsequently add to the nucleus, adopting the structure of PrP^{Sc}. The rate-limiting step in the nucleated polymerization model is not conversion but nucleation. This step, responsible for the lag phase in the spontaneous conversion reaction, can be bypassed by addition of preformed PrP^{Sc} aggregates.

While both these mechanisms are theoretically plausible, there is little evidence for the

existence of a stable PrP^{Sc} monomer. Moreover, prion infectivity is associated with PrP^{Sc}-containing aggregates, not with PrP monomers (213).

B.2 Cell-Free Conversion Studies.

An important development in TSE research was the finding that PrP^{C} can be converted in vitro into a PrP^{Sc}-like conformation (a protease-resistant state, PrP-res) from infected animals with normal PrP^{C} (42; 111; 131).

These cell-free experiments, pioneered by Caughey and co-workers, demonstrated that PrP conversion consists of two kinetically distinct steps: (i) binding of PrP^{C} to PrP-res oligomer, followed by (ii) conversion of bound PrP^{C} to the PrP-res conformation.

In this in vitro conversion, formation of new PrP-res molecules invariably required the presence of oligomeric seeds, and the newly created PrP-res always remained tightly associated with the original PrP-res, indicating that aggregation is an inseparable aspect of PrP conversion (42; 111; 131).

This arguing against the existence of a stable -sheet-rich monomer of PrP

B.3 Recombinant Prion Protein.

Since biophysical studies with brain-derived PrP pose major experimental challenges, many laboratories have resorted to bacterially expressed recombinant prion protein (rPrP), which can be readily purified in large quantities. Although the recombinant protein lacks both glycolsylation and the GPI anchor, its secondary and tertiary structures appear to be identical to those of brain-derived PrP^{C} (113).

Thus, rPrP provides a useful tool for studying the physiochemical properties and conformational transitions of the prion protein (220).

Early studies revealed that the transition of the human rPrP (fragment 90-231) to an oligomeric -sheet structure is especially effective at mildly acidic pH and in the presence of low concentrations of chemical denaturants (222). Moreover, the α-helix to β-sheet transition occurs concomitantly with oligomerization of the protein, according to cell-free conversion studies. (222) However, the α-helix to β-sheet conversion reaction described in these early studies was non-autocatalytic, failing to mimic self-propagation of infectious prions.

Auto-catalytic (seeded) conversion of rPrP was first generated by disulfide oxidation and reduction (146), leading to a model of prion propagation based on the domain-swapped structure as observed in a crystallographic study with human rPrP (130). There is, however, no evidence that disulfide reshuffling occurs during PrP^{C} to PrP^{Sc} conversion in vivo. More recently, it was shown that rPrP can be converted to classical amyloid fibrils without the reduction of the native disulfide bond (21) and in the absence of any denaturing agents or detergents (56).

B.3.1 Synthetic Prions

Intracerebral injection of recombinant PrP fibrils into transgenic mice over expressing N-terminally truncated PrP^C was reported to result in a transmissible neurological disorder, though characterized by a very long incubation time as compared to those of classical TSE diseases (147). This observation has led to the conclusion that these aggregates represented synthetic prions, associated with an unusually slow strain of the disease. While these findings are undoubtedly significant, there are a number of unresolved questions pertaining to this study (57).

Appendix C

Additional Details on Chapter 5

	Reference: MD_WT structure (nm)	Reference: NMR corresponding structure (nm)*	
		Folded-domain	Non-loop
MD_WT_tot	0.13±0.04	0.11±0.02	0.09±0.01
MD_E200K_tot	0.15±0.03	0.15±0.02	0.09±0.02
MD_Q212P_tot	0.16±0.03	Not available	
MD_E219K_tot	0.14±0.03	Not available	

Figure C.1: RMSD - RMSD values of the system studied respect the WT structure. The structure chosen as the reference is the representative structure of the most populated cluster coming from our cluster analysis.

Interaction	π-stacking		(Cys^{179}-Cys^{214}) bridge
Systems	Tyr^{169}-Phe^{175} (nm)	Phe^{175} - Tyr^{218} (nm)	Cα- Cα (nm)
MD_WT_tot	0.45±0.02	0.46±0.03	0.68±0.01
MD_E200K_tot	0.87±0.10	0.48±0.03	0.65 ± 0.02
MD_Q212P_tot	0.87±0.16	0.47±0.03	0.65 ± 0.02
MD_E219K_tot	0.44±0.02	0.45±0.02	0.68±0.01

Figure C.2: π -stacking and (Cys^{179}-Cys^{214} - π -stacking and (Cys^{179}-Cys^{214}) bridge mean values for each set of simulations.

System	$R=\Delta E_{tot}/\Delta E_{kin}$	Average E_{tot} (KJ mol^{-1})	Drift (KJ mol^{-1})
MD_WT	0.048	-315260	-0.00217825
MD_WT 1	0.048	-315220	-0.00218826
MD_WT 2	0.043	-351255	0.000134186
MD_WT 3	0.043	-351255	-0.00013589
MD_E200K	0.054	-314972	-0.00514264
MD_E200K 1	0.043	-317077	-0.0001443
MD_E200K 2	0.045	-317105	-0.0007398
MD_E200K 3	0.042	-317101	-0.0036171
MD_Q212P	0.052	-314763	-0.00498458
MD_Q212P 1	0.043	-314832	-0.0026909
MD_Q212P 2	0.051	-314835	-0.0028808
MD_Q212P 3	0.042	-314802	-0.00237203
MD_E219K	0.057	-314331	-0.00572916
MD_E219K 1	0.052	-314399	-0.00253388
MD_E219K 2	0.041	-314384	-0.00215787
MD_E219K 3	0.0542	-314400	-0.00090530

Figure C.3: Convergence - Convergence

Appendix D

TSEs susceptibility

TSEs susceptibility has been correlated both with the aminoacidic composition of the β_2-α_2 loop and with long-range interactions between residues in the β_2-α_2 loop and residues at the C-terminal of the 3 helix (54). Indeed, within the highly conserved PrP scaffold [1], the β_2-α_2 loop residues (165-173, in human numbering) show local structure variations among species.

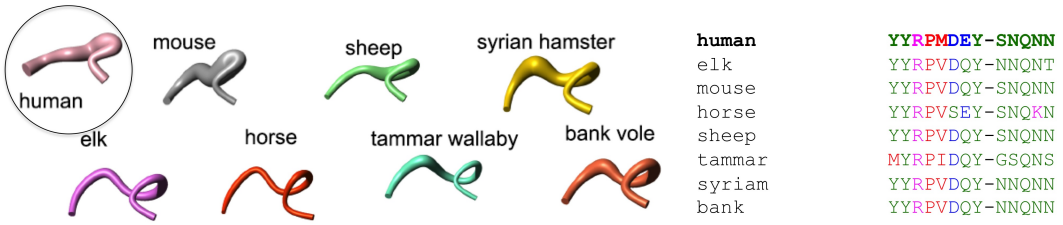


Figure D.1: β_2-α_2 loop in different species - Polypeptide segment 162174 in the energy-minimized NMR structures (92) and multiple sequence alignment of 162-174 residues of PrPC in different species (182)

The NMR-derived flexibility of the loop of PrPC in mouse, sheep and syriam hamster, for which prion disease cases have been reported is larger than that in PrPC of elk (92), bank vole (55), tammar wallaby (54) and, very recently, horse (Perez et al. available on line). Interestingly, in contrast to elk and bank voles, which have a well-defined β_2-α_2 loop (55; 92), and which have independently been shown to be highly susceptible to TSE infection, no TSE has been reported so far in marsupials nor in horses (3; 178). Christen et al. (54) show that the high structural

[1] Sequence Identity of the examined Species across the globular domain is very high 97 % or more

definition of the β_2-α_2 loop in tammar wallaby and horse PrP is entirely due to long-range interactions between loop residues and C-terminal part of 3 (54; 55), whereas in the PrPs of elk and bank vole, the high structural definition of the loop is related to apparently strictly local effects from amino acid substitutions within the loop (55; 92). Thus, the results of the NMR studies on the examined species (3; 55; 92; 178)(Perez et al. available on line), imply that plasticity of the epitope formed by the β_2-α_2 loop and helix 3 appears to determine the susceptibility of a given species to development of a TSE, rather than the isolated β_2-α_2 loop conformation.

Appendix E

Additional details for the PrP docking protocol

E.1 Details on 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 (121; 234) and the Autodock programs (171) were used to predict ligand-protein interactions.

E.1.1 Ligand-protein docking

GOLD (Genetic Optimization for Ligand Docking) (121; 234) 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 dened scoring function. In this work GOLD 3.0.1 was used. AutoDock (171) 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.

E.1.1.1 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 predened systematic manner (103). 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 (103). 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 (89; 109).

Genetic Algorithms Genetic algorithms (GA) (89; 109) 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 favourable 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. 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. 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.

Island-based genetic algorithm In GOLD a so-called island-based genetic algorithm is employed. It means that rather than manipulate only one large population of chromosomes (as described in previous Section), 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.

Lamarckian genetic algorithm The Autodock uses a genetic algorithm 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. 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 dened as free energies, here they are only used to predict ligand binding poses.

E.1.1.2 GOLD scoring functions

The GOLD program has implemented two scoring functions, ChemScore (CS) (177) and GoldScore (GS) (121). 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). 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. 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).

E.1.1.3 Autodock scoring function

In AutoDock the implemented scoring function is defined as an empirical binding free energy function.The are three terms in vacuo, namely Lennard-Jones dispersion, directional hydrogen bond and screened Coulomb electrostatic potential. Then there is a term which is a measure of the unfavorable entropy of ligand binding due to the restriction of conformational degrees of freedom. Finally, 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 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 to give the desolvation energy.

E.2 Details of MD of GN8 protomers in complex with HuPrP^C.

The GN8-HuPrP^C adducts that emerged from docking calculations underwent 10ns of molecular dynamics (MD) simulations in aqueous solution. The ligands maintained completely (1^+ and 1^{2+}) or partially (1^0) the poses identified in the docking. After 3ns of MD simulations, 1^0 partially slided out of the binding pocket. In its final conformation, the GN8 amide oxygen (O2, Figure 1) forms H-bond interactions with Lys194. The pyrrolidine ring forms hydrophobic interactions with methylene groups of Glu196 and Arg156. The two aromatic rings form also hydrophobic interactions with the methylene groups of Lys194. Water molecules can be involved in protein-ligand recognition by forming mediated H-bonds between the protein and the ligand. In this regard, water-mediated H-bond contacts were identified between HuPrP^C and protomers during the MD simulations. The amidic nitrogen of 1^0 protomer forms water mediated H-bond with carbonyl oxygen of Arg156, with hydroxyl group of Thr190 or imidazole nitrogen of His187. Instead, 1^+ and 1^{2+} kept their initial docked structure during the MD. The amide nitrogen atom of both protomers (N3; Figure 1) H-bonds to the backbone oxygen of Glu196. The carbonyl oxygen (O5, Figure 1) of 1^+ H-bonds to the backbone amino group of Asn159. In addition, phenyl rings of 1^+ form hydrophobic interactions with the methylene groups of Arg156 and T-stacking interactions with the His187. During the MD the protonated nitrogen atom (N1, Figure 1) of 1^+ forms direct or water mediated H-bond with the carboxylate side chains of Asp122. The protonated N6 atom of 1^{2+} forms water mediated H-bond with the side chain oxygen of Asn159. Electrostatic interaction is present among carboxylate of Glu196 and N1 atom of 1^{2+}. Hydrophobic interactions between the diphenylmetane fragment of 1^{2+} and methylene groups of Lys194 and Arg156 stabilize further the binding of 1^{2+}.

E.3 Details of bias-exchange metadynamics simulations performed here.

In metadynamics (141) 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 our previous work (28; 82; 232) (Tab. E.1). Here we have combined metadynamics with its bias exchange variant (191). Six independent metadynamics simulations were run in parallel. Each replica was biased by different one-dimensional time-dependent potential, which was built as a function of each of the collective variable defined above. Exchanges among replicas were attempted every 10 ps using a metropolis acceptance criterion (191). Similar set-ups have been shown to improve the sampling of the configurational space and the convergence of the results (148; 191; 192; 226). 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 to 22.5 ns and the other between times 22.5 ns to 30 ns. Only the portion of the two average profiles that agreed with each other within 1.5 k_{BT} were used for further analysis. (see Figure E.1 A and B). An analysis of the converged bias potentials together with the trajectories allows quantifying the relative free energy between all the different states detected during the bias exchange metadynamics simulations. This methodology is based on the weighted histogram analysis method (WHAM) introduced by Kumar et al. (136). Details on this procedure can be found in references (148; 192; 226). Briefly, the method is based on 1) clustering the phase space 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 the WHAM and using the information from the converged free energy profiles. Between 15ns and 30ns, the trajectories of the different replicas were evolving under the action of converged bias potentials (Figure E.1 A and B). Thus, only this part of the simulation data was used for extracting statistical information.

collective variable	width of the Gaussian	height of the Gaussian
distance	0.02 0.04 nm	0.05 0.14 kcalmol-1
polar contacts A	0.1 0.2	0.05 0.14 kcalmol-1
polar contacts B	0.2 0.3	0.05 0.14 kcalmol-1
water bridges	0.8 1	0.05 0.14 kcalmol-1
RMSD system	0.01 0.02 nm	0.05 0.14 kcalmol-1
RMSD binding site	0.01 0.02 nm	0.05 0.14 kcalmol-1

Table E.1: Parameters of the biasing potential in the metadynamics calculations.

E. ADDITIONAL DETAILS FOR THE PRP DOCKING PROTOCOL

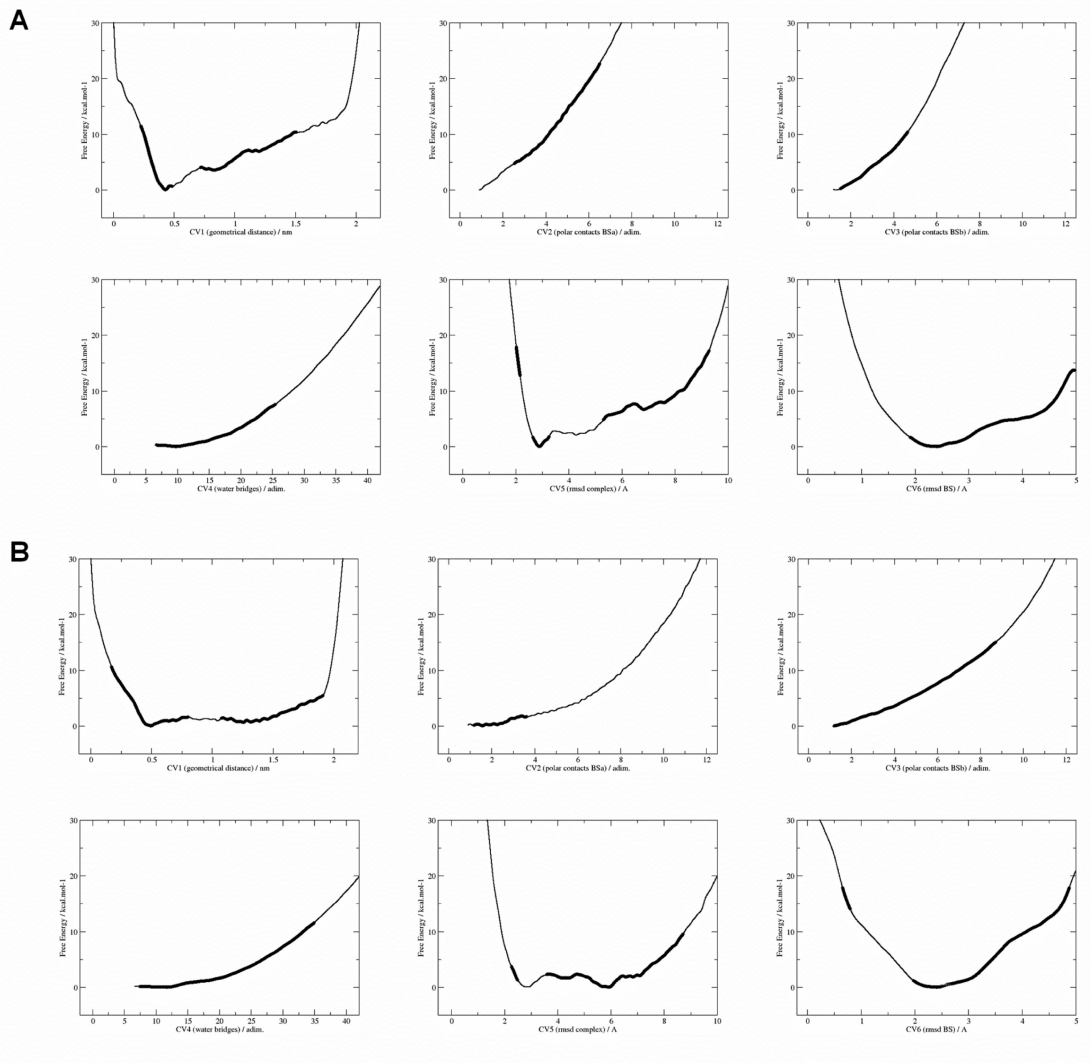


Figure E.1: Free energy profiles - Free energy profiles at the end of the simulation of (A) 1^0 and (B) 1^+ protonation states as a function of each collective variable. The portion of each profile that is converged at the end of the simulation is highlighted in a bold line (For details see Methods).

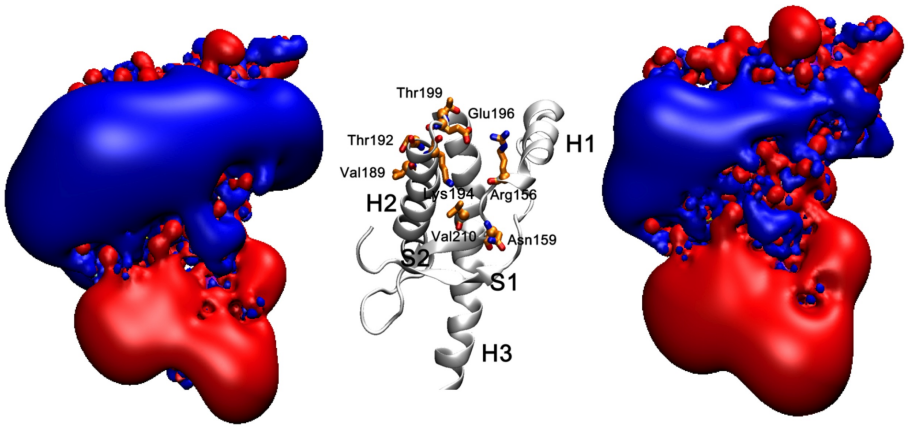


Figure E.2: Electrostatic potential - Electrostatic potential generated by HuPrP^C at different experimental conditions. Left: NMR measurements conditions (pH=4); Right: Affinity measurements conditions (pH=7). Red volume represents the regions at -1.2 eV; Blue volume represents the regions at +1.2 eV. Center: Residues involved in GN8 binding to the prion protein (in licorice), as emerging from chemical shift changes.

E. ADDITIONAL DETAILS FOR THE PRP DOCKING PROTOCOL

Bibliography

[1] J Aaqvist. Ion-water interaction potentials derived from free energy perturbation simulations. *The Journal of Physical Chemistry*, 94(21):8021–8024, 1990. doi: 10.1021/j100384a009. 43, 74

[2] M Adrover, K Pauwels, S Pringent, C De Chiara, Z Xu, C Chapuis, A Pastore, and H Rezaei. Prion fibrillization is mediated by a native structural element which comprises the helices h2 and h3. *Journal of Biological Chemistry*, pages 1–20, 2010. 3, 42, 46, 58, 59, 63, 67, 83

[3] U Agrimi, R Nonno, G Dell'Omo, MA Di Bari, M Conte, B Chiappini, E Esposito, G Di Guardo, O Windl, G Vaccari, and H-P Lipp. Prion protein amino acid determinants of differential susceptibility and molecular feature of prion strains in mice and voles. *PLoS Pathog*, 4(7):e1000113, 2008. 97, 98

[4] A Aguzzi, F Baumann, and J Bremer. The prion's elusive reason for being. *Annu Rev Neurosci*, 31:439–77, 2008. 30, 31, 36

[5] A Aguzzi and AM Calella. Prions: protein aggregation and infectious diseases. *Physiological Reviews*, 89(4):1105–52, 2009. 33, 88

[6] A Aguzzi and C Haass. Games played by rogue proteins in prion disorders and alzheimer's disease. *Science*, 302(5646):814–818, 2003. 29

[7] A Aguzzi and T O'Connor. Protein aggregation diseases: pathogenicity and therapeutic perspectives. *Nat Rev Drug Discov*, 9(3):237–248, 2010. 33

[8] DO Alonso, SJ DeArmond, FE Cohen, and V Daggett. Mapping the early steps in the ph-induced conformational conversion of the prion protein. *P Natl Acad Sci Usa*, 98(6):2985–2989, 2001. 38

[9] H Alonso, AA Bliznyuk, and JE Gready. Combining docking and molecular dynamic simulations in drug design. *Medicinal Research Reviews*, 26(5):531–68, 2006. 71

[10] H Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384–2393, 1980. 26

[11] N Andrusier, E Mashiach, R Nussinov, and HJ Wolfson. Principles of flexible protein-protein docking. *Proteins*, 73(2):271–89, 2008. 4, 71, 84

[12] AC Apetri, K Maki, H Roder, and K Surewicz. Early intermediate in human prion protein folding as evidenced by ultrarapid mixing experiments. *J Am Chem Soc*, 128(35):11673–11678, 2006. 34, 42

[13] AC Apetri and K Surewicz. The effect of disease-associated mutations on the folding pathway of human prion protein. *Journal of Biological Chemistry*, 279(17):18008–18014, 2004. 3, 34, 39, 42, 83

[14] AC Apetri and WK Surewicz. Kinetic intermediate in the folding of human prion protein. *J Biol Chem*, 277(47):44589–92, 2002. Apetri, Adrian C Surewicz, Witold K NS38604/NS/NINDS NIH HHS/United States Research Support, U.S. Gov't, P.H.S. United States The Journal of biological chemistry J Biol Chem. 2002 Nov 22;277(47):44589-92. Epub 2002 Sep 27. 42

[15] A Ashok and RS Hegde. Selective processing and metabolism of disease-causing mutant prion proteins. *PLoS Pathog*, 5(6):e1000479, 2009. 40, 58

[16] S Auer, F Meersman, CM Dobson, and M Vendruscolo. A generic mechanism of emergence of amyloid protofilaments from disordered oligomeric aggregates. *PLoS Comp Biol*, 4(11):e1000222, 2008. 33

[17] S-H Bae, G Legname, A Serban, SB Prusiner, P Wright, and H Dyson. Prion proteins with pathogenic and protective mutations show similar structure and dynamics. *Biochemistry-Us*, 48(34):8120–8128, 2009. doi: 10.1021/bi900923b. 59

[18] NA Baker, D Sept, S Joseph, MJ Holst, and JA McCammon. Electrostatics of nanosystems: application to microtubules and the ribosome. *P Natl Acad Sci Usa*, 98(18):10037–41, 2001. 75, 77

[19] K Bamdad and H Naderi-Manesh. Contribution of a putative salt bridge and backbone dynamics in the structural instability of human prion protein upon r208h mutation. *Biochemical and Biophysical Research Communications*, 364(4):719–24, 2007. 39

[20] M Bartolini and V Andrisano. Strategies for the inhibition of protein aggregation in human diseases. *Chembiochem*, 11(8):1018–35, 2010. 84

[21] IV Baskakov. Autocatalytic conversion of recombinant prion proteins displays a species barrier. *J Biol Chem*, 279(9):7671–7, 2004. 92

[22] K Basler, B Oesch, M Scott, D Westaway, M Wlchli, DF Groth, MP McKinley, SB Prusiner, and C Weissmann. Scrapie and cellular prp isoforms are encoded by the same chromosomal gene. *Cell*, 46(3):417–28, 1986. 30

[23] C Bayly, P Cieplak, W Cornell, and P Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model. *The Journal of Physical Chemistry*, 97(40):10269–10280, 1993. doi: 10.1021/j100142a004. 76

BIBLIOGRAPHY

[24] JA Beck, S Mead, TA Campbell, A Dickinson, DP Wientjens, EA Croes, CM Van Duijn, and J Collinge. Two-octapeptide repeat deletion of prion protein associated with rapidly progressive dementia. *Neurology*, 57(2):354–6, 2001. 37

[25] H Berendsen, J Grigera, and T Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91(24):6269–6271, 1987. 26

[26] H Berendsen, J Postma, W Vangunsteren, A Di Nola, and J Haak. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690, 1984. 26

[27] HJC Berendsen, D van der Spoel, and R van Drunen. Gromacs: a message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 1995. 14, 18, 43, 74

[28] X Biarns, A Ardvol, 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(35):10686–93, 2007. 102

[29] DD Boehr, R Nussinov, and PE Wright. The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol*, 5(11):789–96, 2009. 4

[30] ML Bolognesi, HN AiTran, M Staderini, A Monaco, A Lpez-Cobeas, S Bongarzone, X Biarns, P Lpez-Alvarado, N Cabezas, M Caramelli, P Carloni, JC Mendez, and G Legname. Discovery of a class of diketopiperazines as antiprion compounds. *ChemMedChem*, 5(8):1324–1334, 2010. 4, 83

[31] G Bottegoni, W Rocchia, M Recanatini, and A Cavalli. Aclap, autonomous hierarchical agglomerative cluster analysis based protocol to partition conformational datasets. *Bioinformatics*, 22:58 – 65, 2006. 76

[32] S Brandner, S Isenmann, A Raeber, M Fischer, A Sailer, Y Kobayashi, S Marino, C Weissmann, and A Aguzzi. Normal host prion protein necessary for scrapie-induced neurotoxicity. *Nature*, 379(6563):339–43, 1996. 30

[33] 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(25):9147–55, 2005. 73, 76, 79

[34] P Brown, L Cervenkov, JW Boellaard, D Stavrou, LG Goldfarb, and DC Gajdusek. Identification of a prnp gene mutation in jakob's original creutzfeldt-jakob disease family. *Lancet*, 344(8915):130–1, 1994. 3, 34, 37

[35] M Bucciantini, E Giannoni, F Chiti, F Baroni, L Formigli, J Zurdo, N Taddei, G Ramponi, CM Dobson, and M Stefani. Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature*, 416(6880):507–11, 2002. 72

[36] MG Bursavich and DH Rich. Designing non-peptide peptidomimetics in the 21st century: inhibitors targeting conformational ensembles. *Journal of Medicinal Chemistry*, 45(3):541–58, 2002. 4, 84

[37] H Beler, A Aguzzi, A Sailer, RA Greiner, P Autenried, M Aguet, and C Weissmann. Mice devoid of prp are resistant to scrapie. *Cell*, 73(7):1339–47, 1993. 30

[38] L Calzolai and R Zahn. Influence of ph on nmr structure and stability of the human prion protein globular domain. *J Biol Chem*, 278(37):35592–6, 2003. 3, 42, 43, 44, 45, 49, 58, 60, 66, 72, 74

[39] V Campana, L Zentilin, I Mirabile, A Kranjc, P Casanova, M Giacca, SB Prusiner, G Legname, and C Zurzolo. Development of antibody fragments for immunotherapy of prion diseases. *The Biochemical journal*, 418(3):507–15, 2009. 77

[40] HA Carlson. Protein flexibility and drug design: how to hit a moving target. *Current Opinion in Chemical Biology*, 6:447 – 452, 2002. 71

[41] DA Case, TE Cheatham, T Darden, H Gohlke, R Luo, KM Merz, A Onufriev, C Simmerling, B Wang, and RJ Woods. The amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26(16):1668–88, 2005. 43, 74

[42] B Caughey. Interactions between prion protein isoforms: the kiss of death? *Trends Biochem Sci*, 26(4):235–42, 2001. 30, 92

[43] B Caughey and PT Lansbury. Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu Rev Neurosci*, 26:267–98, 2003. 33

[44] A Cavalli, ML Bolognesi, A Minarini, M Rosini, V Tumiatti, M Recanatini, and C Melchiorre. Multi-target-directed ligands to combat neurodegenerative diseases. *J Med Chem*, 51:347 – 372, 2008. 72

[45] N Chakroun, S Prigent, CA Dreiss, S Noinville, C Chapuis, F Fraternali, and H Rezaei. The oligomerization properties of prion protein are restricted to the h2h3 domain. *FASEB J*, 24(9):3222–31, 2010. 63, 67, 83

[46] S Chasseigneaux, S Hak, I Laffont-Proust, O De Marco, M Lenne, J-P Brandel, J-J Hauw, J-L Laplanche, and K Peoc'h. V180i mutation of the prion protein gene associated with atypical prpsc glycosylation. *Neurosci Lett*, 408(3):165–9, 2006. 39

[47] S Chaudhury and JJ Gray. Conformer selection and induced fit in flexible backbone protein-protein docking using computational and nmr ensembles. *Journal of Molecular Biology*, 381(4):1068–87, 2008. 4, 84

[48] ChemAxon. Calculator plugins were used for structure property prediction and calculation., 2008. 75

[49] B Chesebro, R Race, K Wehrly, J Nishio, M Bloom, D Lechner, S Bergstrom, K Robbins, L Mayer, and JM Keith. Identification of scrapie prion protein-specific mrna in scrapie-infected and uninfected brain. *Nature*, 315(6017):331–3, 1985. 30

[50] B Chesebro, M Trifilo, RE Race, K Meade-White, C Teng, R LaCasse, L Raymond, C Favara, G Baron, S Priola, B Caughey, E Masliah, and M Oldstone. Anchorless prion protein results in infectious amyloid disease without clinical scrapie. *Science*, 308(5727):1435–1439, 2005. 31

[51] F Chiti and CM Dobson. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem*, 75:333–66, 2006. 84

[52] Z Chiti, OM Knutsen, S Betmouni, and JRT Greene. An integrated, temporal study of the behavioural, electrophysiological and neuropathological consequences of murine prion disease. *Neurobiol Dis*, 22(2):363–73, 2006. 1

[53] CJ Choi, A Kanthasamy, V Anantharam, and AG Kanthasamy. Interaction of metals with prion protein: possible role of divalent cations in the pathogenesis of prion diseases. *Neurotoxicology*, 27(5):777–87, 2006. 3

[54] B Christen, S Hornemann, FF Damberger, and K Wuthrich. Prion protein nmr structure from tammar wallaby (macropus eugenii) shows that the beta2-alpha2 loop is modulated by long-range sequence effects. *J Mol Biol*, 389(5):833–45, 2009. Christen, Barbara Hornemann, Simone Damberger, Fred F Wuthrich, Kurt Research Support, Non-U.S. Gov't England Journal of molecular biology J Mol Biol. 2009 Jun 26;389(5):833-45. Epub 2009 Apr 23. 3, 31, 69, 83, 97, 98

[55] B Christen, DR Prez, S Hornemann, and K Wthrich. Nmr structure of the bank vole prion protein at 20 degrees c contains a structured loop of residues 165-171. *Journal of Molecular Biology*, 383(2):306–12, 2008. 3, 31, 97, 98

[56] NJ Cobb, AC Apetri, and K Surewicz. Prion protein amyloid formation under native-like conditions involves refolding of the c-terminal alpha-helical domain. *Journal of Biological Chemistry*, 283(50):34704–34711, 2008. 92

[57] NJ Cobb and K Surewicz. Prion diseases and their biochemical mechanisms. *Biochemistry*, 48(12):2574–85, 2009. 93

[58] FE Cohen, KM Pan, Z Huang, M Baldwin, RJ Fletterick, and SB Prusiner. Structural clues to prion replication. *Science*, 264(5158):530–1, 1994. 4, 5

[59] J Collinge. Prion diseases of humans and animals: their causes and molecular basis. *Annu Rev Neurosci*, 24:519–50, 2001. 29, 30, 34

[60] J Collinge, J Beck, T Campbell, K Estibeiro, and RG Will. Prion protein gene analysis in new variant cases of creutzfeldt-jakob disease. *Lancet*, 348(9019):56, 1996. 88

[61] J Collinge, J Brown, J Hardy, M Mullan, MN Rossor, H Baker, TJ Crow, R Lofthouse, M Poulter, and R Ridley. Inherited prion disease with 144 base pair gene insertion. 2. clinical and pathological features. *Brain*, 115 (Pt 3):687–710, 1992. 88

[62] J Collinge, MS Palmer, and AJ Dryden. Genetic predisposition to iatrogenic creutzfeldt-jakob disease. *Lancet*, 337(8755):1441–2, 1991. 88

[63] M Colucci, FJ Moleres, Z-L Xie, A Ray-Chaudhury, S Gutti, CM Butefisch, L Cervenakova, W Wang, LG Goldfarb, Q Kong, B Ghetti, SG Chen, and P Gambetti. Gerstmann-strussler-scheinker: a new phenotype with 'curly' prp deposits. *J Neuropath Exp Neur*, 65(7):642–51, 2006. 39

[64] W Cornell, P Cieplak, C Bayly, I Gould, KM Merz, D Ferguson, D Spellmeyer, T Fox, JW Caldwell, and PA Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc*, 117(19):5179–5197, 1995. 76

[65] V Daggett. Protein folding-simulation. *Chem Rev*, 106(5):1898–916, 2006. 3

[66] T Darden, D York, and L Pedersen. Particle mesh ewald: An n-log(n) method for ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993. 21, 43, 74

[67] X Daura, K Gademann, B Jaun, D Seebach, WFv Gunsteren, and AE Mark. Peptide folding: When simulation meets experiment. *Angew. Chem. Int. Ed.*, 38(1-2):236–240, 1999. 44, 75, 77

[68] GMS De Mori, G Colombo, and C Micheletti. Study of the villin headpiece folding dynamics by combining coarse-grained monte carlo evolution and all-atom molecular dynamics. *Proteins*, 58(2):459–71, 2005. 3

[69] A De Simone, GG Dodson, CS Verma, A Zagari, and F Fraternali. Prion and water: tight and dynamical hydration sites have a key role in structural stability. *P Natl Acad Sci Usa*, 102(21):7535–40, 2005. 39

[70] SJ DeArmond, MP McKinley, RA Barry, MB Braunfeld, JR McColloch, and SB Prusiner. Identification of prion amyloid filaments in scrapie-infected brain. *Cell*, 41(1):221–35, 1985. 33

[71] ML DeMarco and V Daggett. Molecular mechanism for low ph triggered misfolding of the human prion protein. *Biochemistry-Us*, 46(11):3045–54, 2007. 37

[72] ML DeMarco, J Silveira, B Caughey, and V Daggett. Structural properties of prion protein protofibrils and fibrils: an experimental assessment of atomic models. *Biochemistry*, 45(51):15573–82, 2006. 33

[73] R Dima and D Thirumalai. Exploring the propensities of helices in prpc to form beta sheet using nmr structures and sequence alignments. *Biophysical Journal*, 83(3):1268–1280, 2002. 3, 42, 59, 67

[74] RI Dima and D Thirumalai. Probing the instabilities in the dynamics of helical fragments from mouse prpc. *P Natl Acad Sci Usa*, 101(43):15335–40, 2004. 38, 67

[75] SR Dlouhy, K Hsiao, MR Farlow, T Foroud, PM Conneally, P Johnson, SB Prusiner, ME Hodes, and B Ghetti. Linkage of the indiana kindred of gerstmann-strussler-scheinker disease to the prion protein gene. *Nat Genet*, 1(1):64–7, 1992. 88

BIBLIOGRAPHY

[76] CM Dobson. Protein misfolding, evolution and disease. *Trends Biochem Sci*, 24(9):329–32, 1999. 1

[77] B Drisaldi, RS Stewart, C Adles, LR Stewart, E Quaglio, E Biasini, L Fioriti, R Chiesa, and DA Harris. Mutant prp is delayed in its exit from the endoplasmic reticulum, but neither wild-type nor mutant prp undergoes retrotranslocation prior to proteasomal degradation. *J Biol Chem*, 278(24):21732–43, 2003. 38

[78] F Eisenhaber, P Lijnzaad, P Argos, C Sander, and M Scharf. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *Journal of Computational Chemistry*, 16(3):273–284, 1995. 65

[79] AH Elcock. Molecular simulations of cotranslational protein folding: fragment stabilities, folding cooperativity, and trapping in the ribosome. *PLoS Comput Biol*, 2(7):e98, 2006. 3

[80] U Essmann, L Perera, M Berkowitz, T Darden, H Lee, and L Pedersen. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103(19):8577–8593, 1995. 21, 43, 74

[81] A Esteras-Chopo, G Morra, E Moroni, L Serrano, M Lopez de la Paz, and G Colombo. A molecular dynamics study of the interaction of d-peptide amyloid inhibitors with their target sequence reveals a potential inhibitory pharmacophore conformation. *Journal of Molecular Biology*, 383(1):266–80, 2008. 84

[82] G Fiorin, A Pastore, P Carloni, and M Parrinello. Using metadynamics to understand the mechanism of calmodulin/target recognition at atomic detail. *Biophysical Journal*, 91(8):2768–77, 2006. 73, 76, 79, 102

[83] G Forloni, N Angeretti, R Chiesa, E Monzani, M Salmona, O Bugiani, and F Tagliavini. Neurotoxicity of a prion protein fragment. *Nature*, 362(6420):543–6, 1993. 37

[84] D Frenkel and B Smit. Understanding molecular simulation: from algorithms to applications. page 638, 2002. 10

[85] M Frisch, H Schlegel, G Scuseria, M Robb, J Cheeseman, J Montgomery, T Vreven, K Kudin, J Burant, J Millam, S Iyengar, J Tomasi, V Barone, B Mennucci, M Cossi, G Scalmani, N Rega, G 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 Knox, H Hratchian, J Cross, C Adamo, J Jaramillo, R Gomperts, R Stratmann, O Yazyev, A Austin, R Cammi, C Pomelli, J Ochterski, P Ayala, K Morokuma, G Voth, P Salvador, JJ Dannenberg, V Zakrzewski, S Dapprich, A Daniels, M Strain, O Farkas, D Malick, A Rabuck, K Raghavachari, J Foresman, J Ortiz, Q Cui, A Baboul, S Clifford, J Cioslowski, B Stefanov, G Liu, A Liashenko, P Piskorz, I Komaromi, R Martin, D Fox, T Keith, M Al-Laham, C Peng, A Nanayakkara, M Challacombe, P Gilli, B Johnson, W Chen, M Wong, C Gonzalez, and JA Pople. *Gaussian 03*. 2003. 76

[86] M Gallo, D Paludi, DO Cicero, K Chiovitti, E Millo, A Salis, G Damonte, A Corsaro, S Thellung, G Schettini, S Melino, T Florio, M Paci, and A Aceto. Identification of a conserved n-capping box important for the structural autonomy of the prion alpha 3-helix: the disease associated d202n mutation destabilizes the helical conformation. *Int J Immunopathol Pharmacol*, 18(1):95–112, 2005. 38, 39

[87] FL Gervasio, A Laio, and M Parrinello. Flexible docking in solution using metadynamics. *J Am Chem Soc*, 127(8):2600–7, 2005. 73, 76, 79

[88] MK Gilson and H-X Zhou. Calculation of protein-ligand binding affinities. *Annual review of biophysics and biomolecular structure*, 36:21–42, 2007. 73

[89] D Goldberg. *Genetic algorithms in search optimization and machine learning.* Addison- Wesley, Reading, MA, 1989. 100

[90] LG Goldfarb, P Brown, M Haltia, F Cathala, WR McCombie, J Kovanen, L Cervenkov, L Goldin, A Nieto, and MS Godec. Creutzfeldt-jakob disease cosegregates with the codon 178asn prnp mutation in families of european origin. *Ann Neurol*, 31(3):274–81, 1992. 3, 38

[91] LG Goldfarb, P Brown, WR McCombie, D Goldgaber, GD Swergold, PR Wills, L Cervenakova, H Baron, CJ Gibbs, and DC Gajdusek. Transmissible familial creutzfeldt-jakob disease associated with five, seven, and eight extra octapeptide coding repeats in the prnp gene. *P Natl Acad Sci Usa*, 88(23):10926–30, 1991. 37

[92] AD Gossert, S Bonjour, DA Lysek, F Fiorito, and K Wthrich. Prion protein nmr structures of elk and of mouse/elk hybrids. *P Natl Acad Sci Usa*, 102(3):646–50, 2005. 3, 31, 97, 98

[93] C Govaerts, H Wille, S Prusiner, and F Cohen. Evidence for assembly of prions with left-handed beta 3-helices into trimers. *Proc Natl Acad Sci U S A*, 101(22):8342–8347, 2004. 33

[94] M Grabenauer, C Wu, P Soto, J-E Shea, and MT Bowers. Oligomers of the prion protein fragment 106-126 are likely assembled from beta-hairpins in solution, and methionine oxidation inhibits assembly without altering the peptide's monomeric conformation. *J Am Chem Soc*, 132(2):532–9, 2010. 3

[95] E Grasbon-Frodl, H Lorenz, U Mann, RM Nitsch, O Windl, and HA Kretzschmar. Loss of glycosylation associated with the t183a mutation in human prion disease. *Acta Neuropathol*, 108(6):476–84, 2004. 39

[96] JS Griffith. Self-replication and scrapie. *Nature*, 215(5105):1043–4, 1967. 30

[97] Y Gu, A Singh, S Bose, and N Singh. Pathogenic mutations in the glycosylphosphatidylinositol signal peptide of prp modulate its topology in neuroblastoma cells. *Mol Cell Neurosci*, 37(4):647–56, 2008. 40

[98] Y Gu and N Singh. Doxycycline and protein folding agents rescue the abnormal phenotype of familial cjd h187r in a cell model. *Brain Res Mol Brain Res*, 123(1-2):37–44, 2004. 39

[99] Y Gu, S Verghese, S Bose, M Mohan, and N Singh. Mutant prion protein d202n associated with familial prion disease is retained in the endoplasmic reticulum and forms 'curly' intracellular aggregates. *J Mol Neurosci*, 32(1):90–6, 2007. 39

[100] N Guex and MC Peitsch. Swiss-model and the swiss-pdbviewer: an environment for comparative protein modeling. *Electrophoresis*, 18(15):2714–23, 1997. 43

[101] NS Hachiya, M Imagawa, and K Kaneko. The possible role of protein x, a putative auxiliary factor in pathological prion replication, in regulating a physiological endoproteolytic cleavage of cellular prion protein. *Med Hypotheses*, 68(3):670–3, 2007. 3

[102] JA Hainfellner, S Brantner-Inthaler, L Cervenkov, P Brown, T Kitamoto, J Tateishi, H Diringer, PP Liberski, H Regele, and M Feucht. The original gerstmann-strussler-scheinker family of austria: divergent clinicopathological phenotypes but constant prp genotype. *Brain Pathol*, 5(3):201–11, 1995. 34

[103] I Halperin, B Ma, H Wolfson, and R Nussinov. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*, 47(4):409–43, 2002. 4, 99

[104] FE Herrera, A Chesi, KE Paleologou, A Schmid, A Munoz, M Vendruscolo, S Gustincich, HA Lashuel, and P Carloni. Inhibition of alpha-synuclein fibrillization by dopamine is mediated by interactions with five c-terminal residues and with e83 in the nac region. *PLoS ONE*, 3(10):e3394, 2008. 3, 84

[105] J Heske, U Heller, KF Winklhofer, and J Tatzelt. The c-terminal globular domain of the prion protein is necessary and sufficient for import into the endoplasmic reticulum. *J Biol Chem*, 279(7):5435–43, 2004. 38

[106] B Hess, H Bekker, H Berendsen, and J Fraaije. Lincs: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997. 24, 43, 74

[107] B Hess, C Kutzner, D van der Spoel, and E Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4(3):435–447, 2008. doi: 10.1021/ct700301q. 14, 18, 43

[108] AF Hill, RJ Butterworth, S Joiner, G Jackson, MN Rossor, DJ Thomas, A Frosh, N Tolley, JE Bell, M Spencer, A King, S Al-Sarraj, JW Ironside, PL Lantos, and J Collinge. Investigation of variant creutzfeldt-jakob disease and other human prion diseases with tonsil biopsy samples. *Lancet*, 353(9148):183–9, 1999. 88

[109] J Holland. *Adaptation in natural and artificial systems.* University of Michigan Press, Ann Arbor, MI, 1975. 100

[110] W Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A*, 31(3):1695–1697, 1985. 43, 74

[111] M Horiuchi and B Caughey. Prion protein interconversions and the transmissible spongiform encephalopathies. *Structure*, 7(10):R231–40, 1999. 92

[112] S Hornemann and R Glockshuber. A scrapie-like unfolding intermediate of the prion protein domain prp(121-231) induced by acidic ph. *Proc Natl Acad Sci U S A*, 95(11):6010–4, 1998. Hornemann, S Glockshuber, R Research Support, Non-U.S. Gov't United states Proceedings of the National Academy of Sciences of the United States of America Proc Natl Acad Sci U S A. 1998 May 26;95(11):6010-4. 31

[113] S Hornemann, C Schorn, and K Wthrich. Nmr structure of the bovine prion protein isolated from healthy calf brains. *EMBO Rep*, 5(12):1159–64, 2004. 3, 92

[114] W Humphrey, A Dalke, and K Schulten. Vmd: visual molecular dynamics. *J Mol Graph*, 14(1):33–8, 27–8, 1996. 75

[115] G Ilc, G Giachin, M Jaremko, Jaremko, F Benetti, J Plavec, I Zhukov, and G Legname. Nmr structure of the human prion protein with the pathological q212p mutation reveals unique structural features. *PLoS ONE*, 5(7):e11715, 2010. 3, 4, 39, 60, 61, 66, 67, 83

[116] T Ishikawa, T Ishikura, and K Kuwata. Theoretical study of the prion protein based on the fragment molecular orbital method. *J Comput Chem*, 30(16):2594–601, 2009. 72, 73, 81

[117] JT Jarrett and PT Lansbury. Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in alzheimer's disease and scrapie? *Cell*, 73(6):1055–8, 1993. 91

[118] H-F Ji, H-Y Zhang, and L Shen. The role of electrostatic interaction in triggering the unraveling of stable helix 1 in normal prion protein. a molecular dynamics simulation investigation. *J Biomol Struct Dyn*, 22(5):563–70, 2005. 3

[119] T Jin, Y Gu, G Zanusso, M Sy, A Kumar, M Cohen, P Gambetti, and N Singh. The chaperone protein bip binds to a mutant prion protein and mediates its degradation by the proteasome. *J Biol Chem*, 275(49):38699–704, 2000. 3

[120] EM Jones, K Surewicz, and WK Surewicz. Role of n-terminal familial mutations in prion protein fibrillization and prion amyloid propagation in vitro. *J Biol Chem*, 281(12):8190–6, 2006. 37

[121] G Jones, P Willett, RC Glen, AR Leach, and R Taylor. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, 267(3):727–48, 1997. 74, 76, 99, 101

[122] W Jorgensen, J Chandrasekhar, J Madura, R Impey, and M Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983. 43, 74

[123] S Kamtekar and MH Hecht. Protein motifs. 7. the four-helix bundle: what determines a fold? *FASEB J*, 9(11):1013–22, 1995. 59

BIBLIOGRAPHY

[124] K Kaneko, L Zulianello, M Scott, CM Cooper, AC Wallace, TL James, FE Cohen, and SB Prusiner. Evidence for protein x binding to a discontinuous epitope on the cellular prion protein during scrapie prion propagation. *Proc Natl Acad Sci U S A*, 94(19):10069–74, 1997. 0027-8424 (Print) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. 58, 60, 65, 69, 83, 84

[125] M Karplus and JA McCammon. Molecular dynamics simulations of biomolecules. *Nat Struct Biol*, 9(9):646–52, 2002. 3

[126] R Kayed, E Head, JL Thompson, TM McIntire, SC Milton, CW Cotman, and CG Glabe. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science*, 300(5618):486–9, 2003. 33

[127] L Kirby, CR Birkett, H Rudyk, IH Gilbert, and J Hope. In vitro cell-free conversion of bacterial recombinant prp to prpres as a model for conversion. *J Gen Virol*, 84(Pt 4):1013–20, 2003. 72

[128] T Kitamoto, T Muramoto, S Mohri, K Doh-Ura, and J Tateishi. Abnormal isoform of prion protein accumulates in follicular dendritic cells in mice with creutzfeldt-jakob disease. *Journal of virology*, 65(11):6292–5, 1991. 88

[129] DB Kitchen, H Decornez, JR Furr, and J Bajorath. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature reviews Drug discovery*, 3(11):935–49, 2004. 71

[130] KJ Knaus, M Morillas, W Swietnicki, M Malone, K Surewicz, and VC Yee. Crystal structure of the human prion protein reveals a mechanism for oligomerization. *Nat Struct Biol*, 8(9):770–4, 2001. 92

[131] DA Kocisko, JH Come, SA Priola, B Chesebro, GJ Raymond, PT Lansbury, and B Caughey. Cell-free formation of protease-resistant prion protein. *Nature*, 370(6489):471–4, 1994. 92

[132] DE Koshland. Application of a theory of enzyme specificity to protein synthesis. *P Natl Acad Sci Usa*, 44(2):98–104, 1958. 4

[133] GG Kovacs, MW Head, T Bunn, L Laszlo, RG Will, and JW Ironside. Clinicopathological phenotype of codon 129 valine homozygote sporadic creutzfeldtx2013;jakob disease. *Neuropathology Applied Neurobiology*, 26(5):463–472, 2000. 64, 89

[134] GG Kovacs, M Puopolo, A Ladogana, M Pocchiari, H Budka, C van Duijn, SJ Collins, A Boyd, A Giulivi, M Coulthart, N Delasnerie-Laupretre, JP Brandel, I Zerr, HA Kretzschmar, J de Pedro-Cuesta, M Calero-Lara, M Glatzel, A Aguzzi, M Bishop, R Knight, G Belay, R Will, E Mitrova, and EUROCJD. Genetic prion disease: the eurocjd experience. *Hum Genet*, 118(2):166–74, 2005. 64, 65, 84

[135] GG Kovacs, G Trabattoni, JA Hainfellner, JW Ironside, RSG Knight, and H Budka. Mutations of the prion protein gene phenotypic spectrum. *J Neurol*, 249(11):1567–82, 2002. 34, 42

[136] S Kumar, JM Rosenberg, D Bouzida, RH Swendsen, and PA Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992. 103

[137] B Kundu, NR Maiti, EM Jones, KA Surewicz, DL Vanik, and WK Surewicz. Nucleation-dependent conformational conversion of the y145stop variant of human prion protein: structural clues for prion propagation. *P Natl Acad Sci Usa*, 100(21):12069–74, 2003. 38

[138] K Kuwata, YO Kamatari, K Akasaka, and TL James. Slow conformational dynamics in the hamster prion protein. *Biochemistry*, 43(15):4439–46, 2004. Kuwata, Kazuo Kamatari, Yuji O Akasaka, Kazuyuki James, Thomas L Comparative Study Research Support, Non-U.S. Gov't United States Biochemistry Biochemistry. 2004 Apr 20;43(15):4439-46. 3, 42, 58, 63

[139] K Kuwata, H Li, H Yamada, G Legname, SB Prusiner, K Akasaka, and TL James. Locally disordered conformer of the hamster prion protein: a crucial intermediate to prpsc? *Biochemistry*, 41(41):12277–83, 2002. Kuwata, Kazuo Li, Hua Yamada, Hiroaki Legname, Giuseppe Prusiner, Stanley B Akasaka, Kazuyuki James, Thomas L Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. United States Biochemistry Biochemistry. 2002 Oct 15;41(41):12277-83. 3, 5, 31, 42, 58, 63

[140] K Kuwata, N Nishida, T Matsumoto, YO Kamatari, J Hosokawa-Muto, K Kodama, HK Nakamura, K Kimura, M Kawasaki, Y Takakura, S Shirabe, J Takata, Y Kataoka, and S Katamine. Hot spots in prion protein for pathogenic conversion. *P Natl Acad Sci Usa*, 104(29):11921–6, 2007. 4, 5, 72, 73, 75, 77, 78, 79, 81, 82, 84

[141] A Laio and M Parrinello. Escaping free-energy minima. *P Natl Acad Sci Usa*, 99(20):12562–6, 2002. 4, 73, 76, 84, 102

[142] E Langella, R Improta, O Crescenzi, and V Barone. Assessing the acidbase and conformational properties of histidine residues in human prion protein (125228) by means of pka calculations and molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics*, 64(1):167–177, 2006. 39

[143] E Langella, VR Improta, and V Barone. Checking the ph-induced conformational transition of prion protein by molecular dynamics simulations: effect of protonation of histidine residues. *Biophys J*, 87:3623 – 3632, 2004. 3

[144] J Laurn, DA Gimbel, HB Nygaard, JW Gilbert, and SM Strittmatter. Cellular prion protein mediates impairment of synaptic plasticity by amyloid- oligomers. *Nature*, 457(7233):1128–1132, 2009. 72

[145] S Lee, L Antony, R Hartmann, KJ Knaus, K Surewicz, and VC Yee. Conformational diversity in prion protein variants influences intermolecular beta;-sheet formation. *The EMBO Journal*, pages 1–12, 2009. 44, 49, 60, 66

[146] S Lee and D Eisenberg. Seeded conversion of recombinant prion protein to a disulfide-bonded oligomer by a reduction-oxidation process. *Nat Struct Biol*, 10(9):725–30, 2003. 92

[147] G Legname, IV Baskakov, HO Nguyen, D Riesner, FE Cohen, SJ DeArmond, and SB Prusiner. Synthetic mammalian prions. *Science*, 305(5684):673–6, 2004. 1, 30, 83, 93

[148] V Leone, G Lattanzi, C Molteni, and P Carloni. Mechanism of action of cyclophilin a explored by metadynamics simulations. *PLoS Comp Biol*, 5(3):e1000309, 2009. 73, 77, 102, 103

[149] PA Lewis, MH Tattum, S Jones, D Bhelt, M Batchelor, AR Clarke, J Collinge, and GS Jackson. Codon 129 polymorphism of the human prion protein influences the kinetics of amyloid formation. *J Gen Virol*, 87(Pt 8):2443–9, 2006. 42

[150] S Liemann and R Glockshuber. Influence of amino acid substitutions related to inherited human prion diseases on the thermodynamic stability of the cellular prion protein. *Biochemistry-Us*, 38(11):3258–3267, 1999. doi: 10.1021/bi982714g. 3, 34, 39, 42, 58

[151] J-H Lin, AL Perryman, JR Schames, and JA McCammon. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J Am Chem Soc*, 124(20):5632–3, 2002. 71

[152] R Linden, VR Martins, MAM Prado, M Cammarota, I Izquierdo, and RR Brentani. Physiology of the prion protein. *Physiol Rev*, 88:673 – 728, 2008. 40

[153] F Lopez Garcia, R Zahn, R Riek, and K Wuthrich. Nmr structure of the bovine prion protein. *Proc Natl Acad Sci U S A*, 97(15):8334–9, 2000. Lopez Garcia, F Zahn, R Riek, R Wuthrich, K Research Support, Non-U.S. Gov't United states Proceedings of the National Academy of Sciences of the United States of America Proc Natl Acad Sci U S A. 2000 Jul 18;97(15):8334-9. 3, 31

[154] X Lu, PL Wintrode, and WK Surewicz. Beta-sheet core of human prion protein amyloid fibrils as determined by hydrogen/deuterium exchange. *Proc Natl Acad Sci U S A*, 104:1510 – 1515, 2007. 42, 58, 63

[155] E Lugaresi, R Medori, P Montagna, A Baruzzi, P Cortelli, A Lugaresi, P Tinuper, M Zucconi, and P Gambetti. Fatal familial insomnia and dysautonomia with selective degeneration of thalamic nuclei. *N Engl J Med*, 315(16):997–1003, 1986. 34

[156] B Ma and R Nussinov. Polymorphic c-terminal beta-sheet interactions determine the formation of fibril or addl-like globulomer for the alzheimer abeta42 dodecamer. *The Journal of biological chemistry*, 2010. 3

[157] G Mallucci, A Dickinson, J Linehan, P-C Klhn, S Brandner, and J Collinge. Depleting neuronal prp in prion infection prevents disease and reverses spongiosis. *Science*, 302(5646):871–4, 2003. 31

[158] M Mangoni, D Roccatano, and A Di Nola. Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation. *Proteins*, 35(2):153–62, 1999. 71

[159] F Marinelli, F Pietrucci, A Laio, and S Piana. A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations. *PLoS Comp Biol*, 5(8):e1000452, 2009. 77

[160] A May and M Zacharias. Accounting for global protein deformability during protein-protein and protein-ligand docking. *Biochim Biophys Acta*, 1754(1-2):225–31, 2005. 71

[161] JA McCammon. Target flexibility in molecular recognition. *Biochim Biophys Acta*, 1754(1-2):221–4, 2005. 71

[162] MP McKinley, RK Meyer, L Kenaga, F Rahbar, R Cotter, A Serban, and SB Prusiner. Scrapie prion rod formation in vitro requires both detergent extraction and limited proteolysis. *J Virol*, 65(3):1340–51, 1991. 31, 33

[163] CA McLean, E Storey, RJ Gardner, AE Tannenberg, L Cervenkov, and P Brown. The d178n (cis-129m) "fatal familial insomnia" mutation associated with diverse clinicopathologic phenotypes in an australian kindred. *Neurology*, 49(2):552–8, 1997. 38

[164] S Mead. Prion disease genetics. *Eur J Hum Genet*, 14(3):273–81, 2006. 34, 38

[165] S Mead, MPH Stumpf, J Whitfield, JA Beck, M Poulter, T Campbell, JB Uphill, D Goldstein, M Alpers, EMC Fisher, and J Collinge. Balancing selection at the prion protein gene consistent with prehistoric kurulike epidemics. *Science*, 300(5619):640–3, 2003. 88

[166] S Megy, G Bertho, SA Kozin, P Debey, GHB Hoa, and J-P Girault. Possible role of region 152-156 in the structural duality of a peptide fragment from sheep prion protein. *Protein Sci*, 13(12):3151–60, 2004. 38

[167] RK Meyer, MP McKinley, KA Bowman, MB Braunfeld, RA Barry, and SB Prusiner. Separation and properties of cellular and scrapie prion proteins. *P Natl Acad Sci Usa*, 83(8):2310–4, 1986. 31

[168] M Meyer-Luehmann, J Coomaraswamy, T Bolmont, S Kaeser, C Schaefer, E Kilger, A Neuenschwander, D Abramowski, P Frey, AL Jaton, J-M Vigouret, P Paganetti, DM Walsh, PM Mathews, J Ghiso, M Staufenbiel, LC Walker, and M Jucker. Exogenous induction of cerebral beta-amyloidogenesis is governed by agent and host. *Science*, 313(5794):1781–1784, 2006. 1, 29

[169] GL Millhauser. Copper and the prion protein: methods, structures, function, and disease. *Annu Rev Phys Chem*, 58:299 – 320, 2007. 2, 31, 32

[170] RS Mishra, S Bose, Y Gu, R Li, and N Singh. Aggresome formation by mutant prion proteins: the unfolding role of proteasomes in familial prion disorders. *J Alzheimers Dis*, 5(1):15–23, 2003. 39

BIBLIOGRAPHY

[171] G Morris, D Goodsell, R Halliday, R Huey, W Hart, R Belew, and A Olson. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem*, 19(14):1639–1662, 1998. 74, 76, 99

[172] N Naslavsky, R Stein, A Yanai, G Friedlander, and A Taraboulos. Characterization of detergent-insoluble complexes containing the cellular prion protein and its scrapie isoform. *J Biol Chem*, 272(10):6324–31, 1997. 31

[173] A Nazabal, S Hornemann, A Aguzzi, and R Zenobi. Hydrogen/deuterium exchange mass spectrometry identifies two highly protected regions in recombinant full-length prion protein amyloid fibrils. *J Mass Spectrom*, 44(6):965–977, 2009. 33

[174] HD Nguyen and CK Hall. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *P Natl Acad Sci Usa*, 101(46):16180–5, 2004. 3

[175] G Nicastro, L Masino, V Esposito, RP Menon, A De Simone, F Fraternali, and A Pastore. Josephin domain of ataxin-3 contains two distinct ubiquitin-binding sites. *Biopolymers*, 91(12):1203–14, 2009. 84

[176] EM Nicholson, H Mo, SB Prusiner, FE Cohen, and S Marqusee. Differences between the prion protein and its homolog doppel: a partially structured state with implications for scrapie formation. *Journal of Molecular Biology*, 316(3):807–15, 2002. 31

[177] JWM Nissink, CW Murray, MJ Hartshorn, ML Verdonk, JC Cole, and R Taylor. A new test set for validating predictions of protein-ligand interaction. *Proteins*, 49(4):457–71, 2002. 76, 101

[178] R Nonno, MA Di Bari, F Cardone, G Vaccari, P Fazzi, G Dell'Omo, C Cartoni, L Ingrosso, A Boyle, R Galeno, M Sbriccoli, H-P Lipp, M Bruce, M Pocchiari, and U Agrimi. Efficient transmission and characterization of creutzfeldt-jakob disease strains in bank voles. *PLoS Pathog*, 2(2):e12, 2006. 97, 98

[179] EM Norstrom and JA Mastrianni. The agaaaaga palindrome in prp is required to generate a productive prpsc-prpc complex that leads to prion propagation. *J Biol Chem*, 280(29):27236–43, 2005. 37

[180] S Nose. A molecular dynamics method for simulations in the canonical ensemble. *Mol Phys*, 52(2):255 – 268, 1984. 25, 26, 43, 74

[181] S Nose and M Klein. Constant pressure molecular dynamics for molecular systems. *Mol Phys*, 50(5):1055 – 1076, 1983. 25, 26, 43, 74

[182] C Notredame, DG Higgins, and J Heringa. T-coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, 302(1):205–17, 2000. 97

[183] B Oesch, D Westaway, M Wlchli, M McKinley, S Kent, R Aebersold, R Barry, P Tempst, D Teplow, L Hood, S Prusiner, and C Weissmann. A cellular gene encodes scrapie prp 27-30 protein. *Cell*, 40(4):735–746, 1985. 30, 31

[184] F Owen, M Poulter, J Collinge, and TJ Crow. Codon 129 changes in the prion protein gene in caucasians. *Am J Hum Genet*, 46(6):1215–6, 1990. 31, 88

[185] Y Pak and S 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. 71

[186] MS Palmer, AJ Dryden, JT Hughes, and J Collinge. Homozygous prion protein genotype predisposes to sporadic creutzfeldt-jakob disease. *Nature*, 352(6333):340–2, 1991. 88

[187] KM Pan, M Baldwin, J Nguyen, M Gasset, A Serban, D Groth, I Mehlhorn, Z Huang, RJ Fletterick, and FE Cohen. Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *P Natl Acad Sci Usa*, 90(23):10962–6, 1993. 31, 33

[188] M Parrinello and A Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys*, 52(12):7182–7190, 1981. 18, 26, 27, 43, 74

[189] M Pastore, SS Chin, KL Bell, Z Dong, Q Yang, L Yang, J Yuan, SG Chen, P Gambetti, and W-Q Zou. Creutzfeldt-jakob disease (cjd) with a mutation at codon 148 of prion protein gene: relationship with sporadic cjd. *Am J Pathol*, 167(6):1729–38, 2005. 38

[190] S Piana, P Carloni, and M Parrinello. Role of conformational fluctuations in the enzymatic reaction of hiv-1 protease. *Journal of Molecular Biology*, 319(2):567–83, 2002. 3

[191] S Piana and A Laio. A bias-exchange approach to protein folding. *The journal of physical chemistry B*, 111(17):4553–9, 2007. 73, 76, 77, 102

[192] S Piana, A Laio, F Marinelli, M Van Troys, D Bourry, C Ampe, and JC Martins. Predicting the effect of a point mutation on a protein fold: the villin and advillin headpieces and their pro62ala mutants. *Journal of Molecular Biology*, 375(2):460–70, 2008. 3, 77, 102, 103

[193] P Piccardo, S Kish, L Ang, K Young, O Bugiani, F Tagliavini, S Dlouhy, and B Ghetti. Prion protein isoforms in the new variant of gerstmann-straussler-scheinker disease q212p. *J Neuropath Exp Neur*, 57(5):518–518, 1998. 42, 58

[194] A Pietropaolo, L Muccioli, C Zannoni, and E Rizzarelli. Conformational preferences of the full chicken prion protein in solution and its differences with respect to mammals. *Chemphyschem*, 10(9-10):1500–1510, 2009. 3

[195] M Poulter, HF Baker, CD Frith, M Leach, R Lofthouse, RM Ridley, T Shah, F Owen, J Collinge, and J Brown. Inherited prion disease with 144 base pair gene insertion. 1. genealogical and molecular studies. *Brain*, 115 (Pt 3):675–85, 1992. 88

[196] SA Priola, K Meade-White, VA Lawson, and B Chesebro. Flexible n-terminal region of prion protein influences conformation of protease-resistant prion protein isoforms associated with cross-species scrapie infection in vivo and in vitro. *J Biol Chem*, 279(14):13689–95, 2004. 3

[197] SB Prusiner. Novel proteinaceous infectious particles cause scrapie. *Science*, 216:136 – 144, 1982. 29, 30

[198] SB Prusiner. Molecular biology of prion diseases. *Science*, 252(5012):1515–22, 1991. 1, 33, 91

[199] SB Prusiner. Prions. *P Natl Acad Sci Usa*, 95(23):13363–13383, 1998. 1, 5, 29, 30, 34, 42

[200] SB Prusiner, D Groth, A Serban, R Koehler, D Foster, M Torchia, D Burton, SL Yang, and SJ DeArmond. Ablation of the prion protein (prp) gene in mice prevents scrapie and facilitates production of anti-prp antibodies. *P Natl Acad Sci Usa*, 90(22):10608–12, 1993. 30

[201] G Puoti, G Rossi, G Giaccone, T Awan, PM Lievens, CA Defanti, F Tagliavini, and O Bugiani. Polymorphism at codon 129 of prnp affects the phenotypic expression of creutzfeldt-jakob disease linked to e200k mutation. *Ann Neurol*, 48(2):269–70, 2000. 38

[202] Kong Q, Surewicz WK, Peterson RB, Zou W, Chen SG, and e al. *Inherited Prion Diseases.* Prion Biology and Diseases. New york: Cold spring harbor monograph edition, 2004. 42, 58

[203] R Riek, S Hornemann, G Wider, M Billeter, R Glockshuber, and K Wuthrich. Nmr structure of the mouse prion protein domain prp(121-321). *Nature*, 382(6587):180–2, 1996. 3, 31, 72

[204] R Riek, G Wider, M Billeter, S Hornemann, R Glockshuber, and K Wuthrich. Prion protein nmr structure and familial human spongiform encephalopathies. *P Natl Acad Sci Usa*, 95(20):11667–72, 1998. 3, 39

[205] T Rodinger and R Poms. Enhancing the accuracy, the efficiency and the scope of free energy simulations. *Current opinion in structural biology*, 15(2):164–70, 2005. 73

[206] PM Rudd, AH Merry, MR Wormald, and RA Dwek. Glycosylation and prion protein. *Current opinion in structural biology*, 12(5):578–86, 2002. 3

[207] J Safar, PP Roller, DC Gajdusek, and CJ Gibbs. Thermal stability and conformational transitions of scrapie amyloid (prion) protein correlate with infectivity. *Protein Sci*, 2(12):2206–16, 1993. 33

[208] S Santini and P Derreumaux. Helix h1 of the prion protein is rather stable against environmental perturbations: molecular dynamics of mutation and deletion variants of prp(90-231). *Cell Mol Life Sci*, 61(7-8):951–60, 2004. 3, 38

[209] E Schiff, V Campana, S Tivodar, S Lebreton, K Gousset, and C Zurzolo. Coexpression of wild-type and mutant prion proteins alters their cellular localization and partitioning into detergent-resistant membranes. *Traffic*, 9(7):1101–15, 2008. 38

[210] S Shibuya, J Higuchi, R-W Shin, J Tateishi, and T Kitamoto. Protective prion protein polymorphisms against sporadic creutzfeldt-jakob disease. *The Lancet*, 351(9100):419–419, 1998. 42, 58

[211] S Shibuya, J Higuchi, RW Shin, J Tateishi, and T Kitamoto. Codon 219 lys allele of prnp is not found in sporadic creutzfeldt-jakob disease. *Ann Neurol*, 43(6):826–8, 1998. 42, 58

[212] D Shmerling, I Hegyi, M Fischer, T Blttler, S Brandner, J Gtz, T Rlicke, E Flechsig, A Cozzio, C von Mering, C Hangartner, A Aguzzi, and C Weissmann. Expression of amino-terminally truncated prp in the mouse leading to ataxia and specific cerebellar lesions. *Cell*, 93(2):203–14, 1998. 31

[213] JR Silveira, GJ Raymond, AG Hughson, RE Race, VL Sim, SF Hayes, and B Caughey. The most infectious prion protein particles. *Nature*, 437(7056):257–61, 2005. 33, 92

[214] V Sim and B Caughey. Ultrastructures and strain comparison of under-glycosylated scrapie prion fibrils. *Neurobiology of Aging*, 30(12):2031–2042, 2009. 33

[215] N Singh, G Zanusso, SG Chen, H Fujioka, S Richardson, P Gambetti, and RB Petersen. Prion protein aggregation reverted by low temperature in transfected cells carrying a prion protein gene mutation. *J Biol Chem*, 272(45):28461–70, 1997. 39

[216] C Soto. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat Rev Neurosci*, 4(1):49–60, 2003. 72

[217] J Speare. The role of helix 1 aspartates and salt bridges in the stability and conversion of prion protein. *Journal of Biological Chemistry*, 278(14):12522–12529, 2003. 38

[218] N Stahl, M Baldwin, D Teplow, L Hood, B Gibson, A Burlingame, and S Prusiner. Structural studies of the scrapie prion protein using mass spectrometry and amino acid sequencing. *Biochemistry-Us*, 32(8):1991–2002, 1993. doi: 10.1021/bi00059a016. 30, 31

[219] J Stoehr, N Weinmann, H Wille, T Kaimann, L Nagel-Steger, E Birkmann, G Panza, SB Prusiner, M Eigen, and D Riesner. Mechanisms of prion protein assembly into amyloid. *P Natl Acad Sci Usa*, 105(7):2409–2414, 2008. 33

[220] K Surewicz, EM Jones, and AC Apetri. The emerging principles of mammalian prion propagation and transmissibility barriers: Insight from studies in vitro. *Accounts Chem Res*, 39(9):654–662, 2006. 3, 5, 34, 83, 92

[221] A Surguchov. Molecular and cellular biology of synucleins. *Int Rev Cell Mol Biol*, 270:225–317, 2008. 84

[222] W Swietnicki, M Morillas, SG Chen, P Gambetti, and K Surewicz. Aggregation and fibrillization of the recombinant human prion protein huprp90-231. *Biochemistry-Us*, 39(2):424–31, 2000. 92

[223] W Swietnicki, RB Petersen, P Gambetti, and WK Surewicz. Familial mutations and the thermodynamic stability of the recombinant human prion protein. *J Biol Chem*, 273(47):31048–52, 1998. 3, 34, 38, 42, 58, 83

BIBLIOGRAPHY

[224] A Tahiri-Alaoui, AC Gill, P Disterer, and W James. Methionine 129 variant of human prion protein oligomerizes more rapidly than the valine 129 variant: implications for disease susceptibility to creutzfeldt-jakob disease. *J Biol Chem*, 279(30):31390–7, 2004. 3

[225] GC Telling, M Scott, J Mastrianni, R Gabizon, M Torchia, FE Cohen, SJ DeArmond, and SB Prusiner. Prion propagation in mice expressing human and chimeric prp transgenes implicates the interaction of cellular prp with another protein. *Cell*, 83(1):79–90, 1995. 0092-8674 (Print) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. 58, 60, 65, 69, 83

[226] N Todorova, F Marinelli, S Piana, and I Yarovsky. Exploring the folding free energy landscape of insulin using bias exchange metadynamics. *The journal of physical chemistry B*, 113(11):3556–64, 2009. 77, 102, 103

[227] M Totrov and R Abagyan. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Current opinion in structural biology*, 18(2):178–84, 2008. 4, 84

[228] CR Trevitt and J Collinge. A systematic review of prion therapeutics in experimental models. *Brain*, 129(Pt 9):2241–65, 2006. 72, 84

[229] VN Uversky, CJ Oldfield, and AK Dunker. Intrinsically disordered proteins in human diseases: introducing the d2 concept. *Annual review of biophysics*, 37:215–46, 2008. 84

[230] MW van der Kamp and V Daggett. The consequences of pathogenic mutations to the human prion protein. *Protein Eng Des Sel*, 22(8):461–468, 2009. 3, 37, 39, 40, 41, 64

[231] MW van der Kamp, KE Shaw, CJ Woods, and AJ Mulholland. Biomolecular simulation and modelling: status, progress and prospects. *J R Soc Interface*, 5:S173–S190, 2008. 7, 13

[232] AV Vargiu, P Ruggerone, A Magistrato, and P Carloni. Dissociation of minor groove binders from dna: insights from metadynamics simulations. *Nucleic Acids Res*, 36(18):5910–21, 2008. 73, 76, 79, 102

[233] M Vendruscolo. Determination of conformationally heterogeneous states of proteins. *Current opinion in structural biology*, 17(1):15–20, 2007. 3

[234] ML Verdonk, JC Cole, MJ Hartshorn, CW Murray, and RD Taylor. Improved protein-ligand docking using gold. *Proteins*, 52(4):609–23, 2003. 74, 76, 99

[235] JH Viles, FE Cohen, SB Prusiner, DB Goodin, PE Wright, and HJ Dyson. Copper binding to the prion protein: structural implications of four identical cooperative binding sites. *P Natl Acad Sci Usa*, 96(5):2042–7, 1999. 31

[236] JH Viles, D Donne, G Kroon, SB Prusiner, FE Cohen, HJ Dyson, and PE Wright. Local structural plasticity of the prion protein. analysis of nmr relaxation dynamics. *Biochemistry-Us*, 40(9):2743–53, 2001. 3

[237] M Vogtherr, S Grimme, B Elshorst, DM Jacobs, K Fiebig, C Griesinger, and R Zahn. Antimalarial drug quinacrine binds to c-terminal helix of cellular prion protein. *Journal of Medicinal Chemistry*, 46(17):3563–4, 2003. 72

[238] JDF Wadsworth and J Collinge. Molecular pathology of human prion disease. *Acta Neuropathol*, 2010. 34, 83

[239] LC Walker, H Levine, MP Mattson, and M Jucker. Inducible proteopathies. *Trends Neurosci*, 29(8):438–43, 2006. 1

[240] F Wang, X Wang, CG Yuan, and J Ma. Generating a prion with bacterially expressed recombinant prion protein. *Science*, 327(5969):1132–5, 2010. Wang, Fei Wang, Xinhe Yuan, Chong-Gang Ma, Jiyan R01NS060729/NS/NINDS NIH HHS/United States Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't United States Science (New York, N.Y.) Science. 2010 Feb 26;327(5969):1132-5. Epub 2010 Jan 28. 83

[241] J Wang, P Cieplak, and PA Kollman. How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, 21(12):1049–1074, 2000. 76

[242] J Wang, RM Wolf, JW Caldwell, PA Kollman, and DA Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–74, 2004. 43, 74, 76

[243] C Weissmann. Molecular biology of transmissible spongiform encephalopathies. *FEBS Letters*, 389(1):3–11, 1996. 29, 30, 34

[244] H Wille, MD Michelitsch, V Guenebaut, S Supattapone, A Serban, FE Cohen, DA Agard, and SB Prusiner. Structural studies of the scrapie prion protein by electron crystallography. *P Natl Acad Sci Usa*, 99(6):3563–8, 2002. 33

[245] W Xian, PJ Connolly, M Oslin, AC Hausrath, and JJ Osterhout. Fundamental processes of protein folding: Measuring the energetic balance between helix formation and hydrophobic interactions. *Protein Sci*, 15(9):2062–2070, 2006. 59

[246] N Yamamoto and K Kuwata. Regulating the conformation of prion protein through ligand binding. *The journal of physical chemistry B*, 113(39):12853–12856, 2009. 4

[247] R Zahn, A Liu, T Lhrs, R Riek, C von Schroetter, F Lpez Garca, M Billeter, L Calzolai, G Wider, and K Wthrich. Nmr solution structure of the human prion protein. *P Natl Acad Sci Usa*, 97(1):145–50, 2000. 3, 31, 37, 42, 43, 44, 45, 58, 59, 60, 63, 65, 83

[248] G Zanusso, RB Petersen, T Jin, Y Jing, R Kanoush, S Ferrari, P Gambetti, and N Singh. Proteasomal degradation and n-terminal protease resistance of the codon 145 mutant prion protein. *J Biol Chem*, 274(33):23396–404, 1999. 38

[249] H Zhang, J Stockel, I Mehlhorn, D Groth, MA Baldwin, SB Prusiner, TL James, and FE Cohen. Physical studies of conformational plasticity in a recombinant prion protein. *Biochemistry*, 36(12):3543–53, 1997. Zhang, H Stockel, J Mehlhorn, I Groth, D Baldwin, M A Prusiner, S B James, T L Cohen, F E AG02132/AG/NIA NIH HHS/United States AG08967/AG/NIA NIH HHS/United States NS14069/NS/NINDS NIH HHS/United States etc. Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. United states Biochemistry Biochemistry. 1997 Mar 25;36(12):3543-53. 31

[250] J Zhang, W Li, J Wang, M Qin, L Wu, Z Yan, W Xu, G Zuo, and W Wang. Protein folding simulations: from coarse-grained model to all-atom model. *IUBMB Life*, 61(6):627–43, 2009. 3

[251] Y Zhang, L Dai, M Iwamoto, and Z Ouyang. Molecular dynamics study on the conformational transition of prion induced by the point mutation: F198s. *Thin Solid Films*, 499(1-2):224–228, 2006. 39

[252] Y Zhang, W Swietnicki, MG Zagorski, WK Surewicz, and FD Snnichsen. Solution structure of the e200k variant of human prion protein. implications for the mechanism of pathogenesis in familial prion diseases. *J Biol Chem*, 275(43):33650–4, 2000. 3, 38, 42, 44, 58, 60, 66, 83

[253] J Zuegg and JE Gready. Molecular dynamics simulations of human prion protein: importance of correct treatment of electrostatic interactions. *Biochemistry*, 38(42):13862–76, 1999. 3, 38