Molecular Simulation Studies on the Prion Protein Variants: Insights into the Intriguing Effects of Mutations

Xiaojing Cong

A thesis submitted for the degree of
Doctor of Philosophy

Ph.D. programme: Functional and Structural Genomics

Supervisors: Prof. Giuseppe Legname, Ph.D.
Prof. Paolo Carloni, Ph.D.
Giulia Rossetti, Ph.D.

天行健，君子以自强不息。地势坤，君子以厚德载物。

—— 《周易》

Acknowledgements

During the four years as a PhD student in SISSA, I had the great opportunity to be in the collaboration between the Computational Biophysics Lab of Prof. Paolo Carloni in German Research School of Simulation Sciences (Jülich, Germany) and the Prion Lab of Prof. Giuseppe Legname in SISSA. I am grateful to my supervisors, Prof. Carloni and Prof. Legname, for giving me this opportunity to work with two groups of excellent researchers.

I would like to thank Prof. Carloni for his constant guidance, encouragement and continuous support through the years of my study in this international collaboration. He gave me the possibility to stay in his lab where I have constantly benefited from the face-to-face discussions with him and the lab members, from their direct help with both scientific and technical problems, from the substantial computational facilities in the lab and from the numerous colloquiums given by researchers from all over the world. Prof. Carloni led me to multiple prospective research subjects and to work with experts in different fields, from which I could rapidly expand my knowledge and skills for a future career. He taught me to assess and carry out a potential project always by examining carefully and thoroughly the experimental data together with the capacity and accuracy of the computational approach. The years working with Prof. Carloni are also valuable experiences that significantly improved my efficiency in work, developed my communication skills and strengthened my personality.

I would like to address my sincere gratitude to Prof. Legname, who welcomed me to join the Prion Lab and gave me vital advice for the thesis work to be successful. Prof. Legname introduced me to the world of prion biology that binds my computational work with fascinating aspects of the cellular life, which has been the source of my motivation over the years. The discussions with Prof. Legname have always been inspiring and agreeable. The encouragement, reassurance and respect I received from Prof. Legname have been important in establishing my confidence and persistency in carrying out the research work.

I would like to express my true thanks to Dr. Giulia Rossetti who has worked with me step-by-step in realizing the many detailed, non-trivial tasks in every project. The fruits of this thesis would not have been possible without the uncountable days, evenings and even weekends that she spent on working with me. I would like to give my appreciation to Dr. Gabriele Giachin for the many essential discussions and suggestions on the biological and structural aspects of every research subject presented here. Dr. Rossetti, Dr. Gianchin and Erica Sarnataro have provided me with imperative help and suggestions in writing up this thesis as well as the articles for publication. My thanks extend to Salvatore Bongarzone and Nicola Casiraghi for the very pleasant experiences of working together, side-by-side, on our research subjects.

The members of Prof. Carloni's lab in Jülich were always ready to help when I knocked on the door or contacted online with questions or problems. Dr. Emiliano Ippoliti, especially, has given me so much help in solving technical problems, in applications for computing recourses and in proof reading my thesis. Dr. Choung Nguyen, Dr. Vo Cam Quy and Domenica Dibenedetto shared with me many happy moments during the years together as colleagues and friends.

The time I spent in SISSA and in Trieste, full of wonderful experiences and memories, is tightly bound with my dear friends and colleagues: Xiaochuan Ge, Zhaleh Ghaemi, Fahimeh Baftizadeh, Shima Seyed-Allaei, Nhu Trang Do, Rolando Hong, Olga Puccioni, Paola Mengotti, Duvan Franco, Daniele Granata, Luca Laneselli, Gianpaolo Gobbo, Giuseppe Facchetti, Juan Manuel Carmona Loaiza, as well as many other members of SISSA SBP group. Their support and friendship composed a big part of my first two years of life abroad, which turned the difficulties and loneliness into laughs, tears and unique memories. Although the two last years I was staying mostly in Germany, every time I went back to SISSA I always received very cheerful welcome from them.

I have a very special thanks to my boyfriend and soul mate. He has always been there for me, standing by my side over the years. His love, patience and understanding without a single doubt are the secure base for me to carry on with my studies. He experienced all the ups and downs of my research and assured me over and over again that to every problem there is a solution and for whatever happens I am not alone. The special thanks also extend to his family, his father, mother and brother, who have been seeing me as a family member, taking care of me and supporting me without holding back.

Finally, the most heartfelt thanks go to my mother. She has sacrificed her life for me and provided unconditional love. Her upbeat, optimistic attitude and encouragement boost my confidence and strength along the way. This thesis is dedicated to my mother, albeit any amount of gratitude shown to her is woefully inadequate.

List of Abbreviations (in alphabetical order)

Aβ:	Amyloid beta
ADI:	angular dispersion index
BGS:	biased Gaussian steps
Bo:	bovine
BSE:	bovine spongiform encephalopathies
CAA:	cerebral amyloid angiopathy
^{Ctm}PrP:	the transmembrane form of prion protein with the C-terminus in the endoplasmic reticulum lumen
^{cy}PrP:	cytosolic PrP
EM:	electron microscopy
ER:	endoplasmic reticulum
fCJD:	familial Creutzfeldt-Jakob Disease
FFI:	fatal familial insomnia
FFT:	fast Fourier transform
GAG:	sulphated glycosaminoglycans
GD:	the globular domain
GPI:	glycosylphosphatidyl inositol
GSS:	Gerstmann-Sträussler-Scheinker Syndrome
HB:	hydrogen bond
Hu:	human
HX-MS:	hydrogen-deuterium exchange mass spectroscopy
iCJD:	iatrogenic Creutzfeldt-Jakob Disease
LJ:	Lennard-Jones
LRP1:	the lipoprotein receptor-related protein 1
MC:	Monte Carlo
MD:	molecular dynamics
MM:	molecular mechanics
Mo:	mouse
Mo/HuPrP chimera:	chimeric mouse/human hybrid prion protein

CMPrP (or MoPrP chimera):	chimeric mouse prion protein containing artificial mutations
NCAM:	the neural cell adhesion molecule
N-term:	the N-terminal domain
^{Ntm}PrP:	the transmembrane form of prion protein with the N-terminus in the endoplasmic reticulum lumen
OR:	octapeptide repeat
Ov:	ovine
PBC:	periodic boundary conditions
PM:	pathogenic mutation
PME:	particle mesh Ewald
PP:	protective polymorphism
PROFASI:	PROtein Folding and Aggregation Simulator
PrP^{C}:	the cellular prion protein
PrP^{Sc}:	prions or the scrapie form of the prion protein
REM:	replica-exchange method
REMC:	replica-exchange Monte Carlo
R_g:	radius of gyration
RMSD:	Root Main Square Deviation
RMSF:	Root Main Square Fluctuation
SASA:	solvent-accessible surface area
SB:	salt bridge
sCJD:	sporadic Creutzfeldt-Jakob Disease
ScN2a:	mouse neuroblastoma
SD:	standard deviation
^{Sec}PrP:	secretory prion protein
sFI:	sporadic familial insomnia
SHa:	Syrian hamster
SS:	secondary structure
STE:	the “stop transfer effector” region of prion protein
STI1:	the stress-inducible protein 1
Tg:	transgenic

TM1:	the transmembrane region of prion protein
TSE:	transmissible spongiform encephalopathies
vCJD:	variant Creutzfeldt-Jakob Disease
vdW:	van der Waals
WT:	wild type

Table of Contents

Chapter 1 Introduction

Prion diseases, or transmissible spongiform encephalopathies (TSE), are a group of rare fatal neurodegenerative maladies that affect humans and animals[1; 2]. The fundamental breakthrough in TSE research was the discovery of the "prion" —proteinaceous infectious particle— and the verification of the "protein-only" hypothesis[3], which states that prions could self-propagate by converting the cellular prion protein (PrP^{C}) into the scrapie form, PrP^{Sc} (or prions), and lead to neurodegeneration without using any nucleic acids. The concept of prions may unify neurodegenerative diseases under a common pathogenic mechanism[4]. Indeed, growing evidence shows that TSE may share similar pathogenesis with common neurodegenerative syndromes such as Alzheimer's disease and Parkinson's disease[5; 6; a], for which there are currently no cure. Similar to Alzheimer's and Parkinson's diseases, TSE share two remarkable characteristics with other neurodegenerative diseases[4]: i) more than 80% of the cases are sporadic; ii) the familial (inherited) types of these maladies are late-onset, although many of the disease-linked mutations are expressed in embryogenesis. Today, PrP is one of the most studied models for protein misfolding mechanism and TSE serve as an excellent model for studying many other neurodegenerative diseases. Understanding the molecular mechanism of the PrP misfolding process may profoundly influence the development of diagnostics and effective therapies for neurodegenerative diseases in general.

Investigating human (Hu) PrP TSE-linked mutations (more than 50 currently identified mutations, linked to ~15% of the cases) may be very instrumental in this respect, as it can provide hints on the molecular basis of the PrP^{C}→PrP^{Sc} conversion. These mutations cause spontaneous TSE, which are likely due to modifications in the native structure of PrP^{C}. They are located all over the structure. Polymorphisms (i.e. non-pathogenic, naturally occurring mutations) in the PrP gene have been found to influence the etiology and neuropathology of the disease in both humans[8] and sheep[9]. In transgenic (Tg) mice, artificial mutations can determine the susceptibility to the infection of different prion strains[10; 11; 12]. Intriguingly, mouse (Mo) PrP[b] containing artificial mutations (denoted MoPrP chimera, hereinafter) have very different effects *in vitro*[13]: some MoPrP chimera were found to resist PrP^{Sc} infection, whereas some others did not; some of the resistant MoPrP chimeras even exhibited a protective effect (known as the dominant-negative effect) over the co-expressed endogenous wild-type (WT) MoPrP^{C}. Most mutations are located in the folded globular

---

[a] The WHO estimated ca. 7.7 million new cases of Alzheimer's disease in the year 2010 and increasing numbers in the next decades due to global aging[7]. The European Parkinson's Disease Association (EPDA) estimated ca. 6.3 million people living with Parkinson's disease in the year 2011 (www.epda.eu.com).

[b] MoPrP shares over 90% sequence identity with HuPrP.

domain (GD) while fewer are located in the intrinsically disordered N-terminal domain (N-term) (Fig. 1.1). The N-term of PrP^C has been suggested to serve multiple functions *in vivo*, which likely relies on the structural flexibility of this domain[14]. Therefore, characterizing the structural features of the N-term is central for investigating not only the mutations in this domain, but also the physiological role of the N-term.

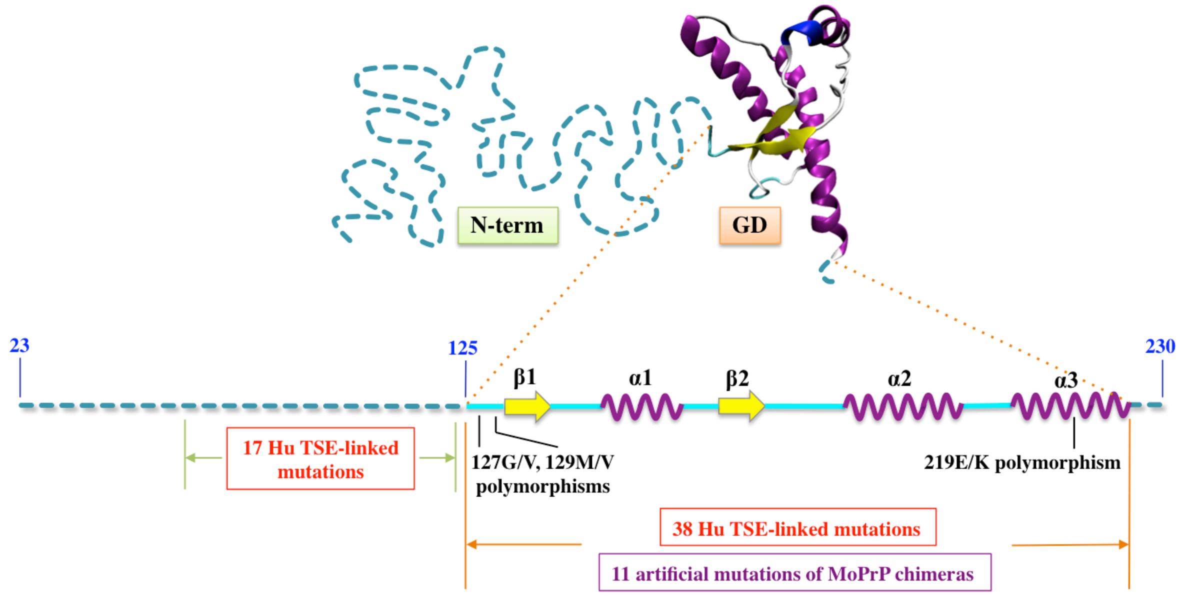


Figure 1.1: Schematic representation of PrP^C structure and the distribution of the mutations and polymorphisms. The intrinsically disordered N-term (residues 23-124, in HuPrP numbering) is depicted in dashed lines. The folded GD (residues 125-230) is shown in cartoon representation, which contains two antiparallel β-strands (in yellow) and three α-helices (in purple). The two signal peptides (residues 1-22 and 231-253) that are cleaved off during the PrP^C biosynthesis are not shown.

The Prion Biology Lab of Prof. Legname has solved the NMR structure of $HuPrP^C$ GD variants in solution[15]. The Lab showed, along with other NMR and X-ray studies[16], that Hu TSE-linked mutations could induce structural features in the protein distinct from those in the WT $HuPrP^C$. These experimental methods are very powerful yet relatively slow. Hence, determining the structures of more than 30 variants represents a challenge. Recently, my colleague Giulia Rossetti (from Prof. Carloni's Lab), along with other researches[16], has shown that molecular dynamics (MD) simulations could predict PrP^C variant structures with high accuracy whilst requiring very short time[17]. Building on this work, in this thesis we first extended the computational protocol to studying the impact of *all* the known Hu TSE-linked mutations[c] in HuPrP^{C} GD. We next applied the same approach to study the GD structure of MoPrP chimeras which contain one or two residues from Hu or sheep PrP sequence. By studying these PrP variants, we aim to identify the structural determinants of the mutants that may play a role in the $\mathrm{PrP}^{C} \rightarrow \mathrm{PrP}^{Sc}$ conversion. Our calculations discovered that these mutants exhibit different structural features from those of the WT PrP GD mainly in two common regions that are likely the "hot spots" in the protein misfolding process. These features can be classified into different types that are correlated to the types of mutants (i.e. pathogenic, resistant or dominant-negative), thus hinting to the molecular mechanisms of PrP^{Sc} formation and propagation.

These results, although very interesting, are limited to the PrP GD. Studying the structural determinants in the intrinsically disordered N-term is highly challenging and indeed structural information is available only for short fragments in this domain, which contain 10 to 20 residues[14]. Here we have predicted the structure of the entire PrP N-term and the impact of the Hu TSE-linked mutations in this domain using a novel Monte Carlo-based simulation approach, PROFASI[18]. PROFASI has already shown to provide structural predictions in a disordered protein such as α-synuclein[19]. Our results are consistent with available experimental data and therefore firmly allow us to provide the first overview on the structural determinants of all Hu TSE-linked mutations in PrP.

---

[c] We did not study nonsense PMs that code for truncated HuPrP. PMs that were identified after we had carried out the study were not included. These are I138M, D167N, N173K, V176G, D202G, V203G, V209M, E211N and Q212H.

Chapter 2 The prion biology

2.1 Prion diseases

Hu TSE are classified according to various phenotypes: i) sporadic (spontaneous) TSE, including sporadic Creutzfeldt-Jakob Disease (sCJD), sporadic familial insomnia (sFI); ii) familial (inherited) TSE, including familial Creutzfeldt-Jakob Disease (fCJD), fatal familial insomnia (FFI), Gerstmann-Sträussler-Scheinker Syndrome (GSS), prion protein cerebral amyloid angiopathy (PrP-CAA); iii) transmitted (infected) TSE, including iatrogenic Creutzfeldt-Jakob Disease (iCJD), variant Creutzfeldt-Jakob Disease (vCJD) and Kuru[20]. There are still other phenotypes of human TSE which are yet unclassified. In recent history, human TSE have occurred in epidemic form due to endocannibalism in Papua New Guinea where tribes suffered from Kuru, which is the first TSE shown to be transmissible[21]. Major public health concerns were raised by the verification that vCJD was caused by the same prion strain as bovine spongiform encephalopathies (BSE)[22; 23] which infected more than 2 million UK cattle in the epizootic of the 1990s.

2.1.1 The mysterious "toxic PrP species"

Although the conversion of PrP^{C} to PrP^{Sc} is clearly central to pathogenesis, the neurotoxicity does not correlate with PrP^{Sc} levels in the brain[24; 25; 26]. Moreover, the distribution of PrP^{Sc} deposits does not necessarily reflect the clinical signs, and PrP^{Sc} is not directly toxic to neurons, which do not express PrP^{C}[27; 28; 29; 30]. These lines of evidence argue that the infectious PrP^{Sc} is not directly neurotoxic, but rather its propagation produces toxic PrP species[31]. The fact that GPI-anchored PrP^{C} is required for neurodegeneration[29; 32; 33] suggests that PrP^{C} acts as a receptor for PrP^{Sc} and mediates toxic signaling[34]. This is further supported by a recent study[35] describing two distinct phases of PrP^{Sc} propagation during the incubation period in mice infected with PrP^{Sc}: phase 1 where PrP^{Sc} propagates exponentially; and phase 2 where PrP^{Sc} remains at a plateau level until the onset of clinical disease. Interestingly, the propagation rate in phase 1 is not proportional to the expression level of PrP^{C}, whereas the length of phase 2 is inversely proportional to PrP^{C} expression level. Moreover, the PrP^{Sc} titer attained in the brain during phases 1 and 2 is independent of PrP^{C} expression level. All these studies clearly demonstrate that PrP^{Sc} infectivity and propagation are uncoupled from the neurotoxicity in prion diseases. More and more evidence suggests that the neurotoxic species consist of small oligomeric forms of PrP which are off-pathway for PrP^{Sc} propagation, while the toxic signaling is mediated by PrP^{C} and likely its interaction with cellular partners[36].

2.2 PrP^{C} structure and function

2.2.1 PrP biosynthesis and trafficking

PrP is highly conserved across mammalian species[37]. In fact, most *in vivo* and *in vitro* studies are made with mouse (Mo) and Syrian hamster (SHa) PrP, which have over 90% sequence identity with human (Hu) PrP (Fig. 2.1A). In this thesis the term "PrP" generally refers to mammalian PrP unless otherwise stated (and same for PrP^{C}, PrP^{Sc}, etc.). The Hu prion gene *PRNP* encodes a 253-residue precursor HuPrP including a 22-residue signal peptide at the N-terminus and a 23-residue signal peptide at the C-terminus (reviewed in ref. [38; 39; 40]). The precursor HuPrP is first imported into the endoplasmic reticulum (ER) where the signal peptides are cleaved and the protein is glycosylated, modified by a C-terminal glycosylphosphatidyl inositol (GPI) anchor and properly folded. The protein subsequently transits in the Golgi where additional modifications are made to its glycans and GPI anchor before being exported to the cell surface[38]. The mature form of HuPrP^{C} is a 208-residue N-glycosylated protein attached to the cell surface by the GPI anchor (Fig. 2.1B)[41].

During co-translational translocation at the ER, four topological forms of PrP can be generated (Fig. 2.1C)[38; 42]. Most PrP nascent chains pass completely through the translocon to generate secretory PrP (^{Sec}PrP) and follow the traditional exocytic pathway towards the cell-surface[38]. Two transmembrane forms, however, span the ER membrane lipid bilayer by a hydrophobic transmembrane region (TM1, residue ~113-135[d]). One form (denoted ^{Ntm}PrP) integrates into the lipid bilayer with the N-terminus in the ER lumen and the C-terminus retained in the cytosol, whereas the other form (^{Ctm}PrP) integrates in the opposite orientation[43]. A fourth form fails to engage the translocon and is released in the cytosol to generate cytosolic PrP (^{cy}PrP)[44; 45 46]. An increased production of the atypical forms, ^{Ntm}PrP, ^{Ctm}PrP and ^{cy}PrP, can be detrimental *in vivo*[43; 47; 48]. ^{Ctm}PrP is found in abnormally high level in the brain tissue from GSS patients carrying the pathogenic mutation A117V[43]. Expressing the equivalent mutation in Tg mice leads to increased ^{Ctm}PrP (5–20% of total PrP, compared with ~1% in wild-type mice) and neurodegeneration[47]. Expression of artificial MpPrP mutants that generate ^{cy}PrP can cause neurodegenerative disease in Tg mice[48]. The abnormal translocation of PrP has been proposed as one of the potential toxic mechanisms *in vitro* and *in vivo*[49; 50; 51]. The translocation process is suggested to be regulated by the TM1 region and a preceding hydrophilic "stop transfer effector" region (STE, residue 103-111)[52; 53]. TM1 and STE act in concert to control the recognition, orientation, and integration at the translocon during PrP biosynthesis[38; 42; 52], which interacts not only with the ER membrane but also trans-acting

---

[d] The residue numbers refer to HuPrP in Uniprot database (www.uniprot.org, entry ID P04156) and hereafter, unless otherwise stated.

factors in the ER membrane[52] and in the cytosol[54]. The molecular mechanism of abnormal PrP topology and the related neurotoxic pathway are still elusive.

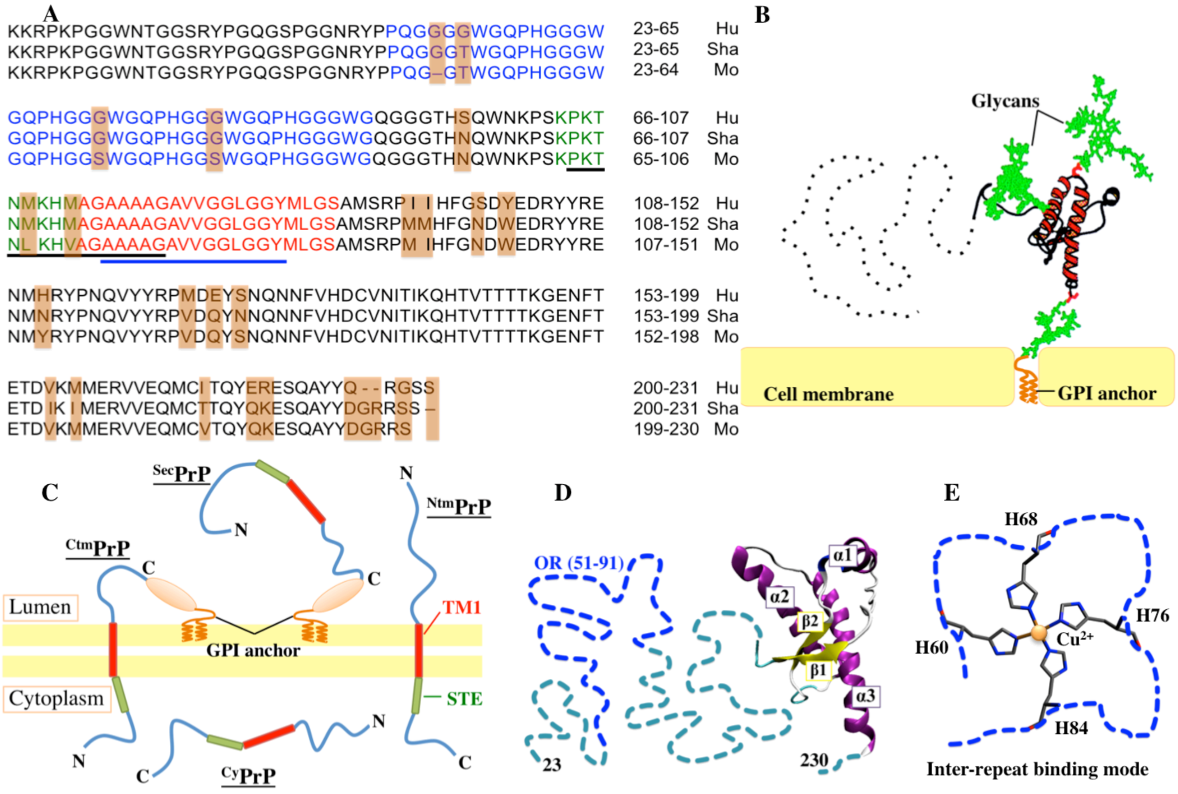


Figure 2.1: A) Sequence alignment of Hu, Mo and SHa PrP excluding their signal peptides. The three proteins differ in 25 residues (shadowed in orange). Residues in the OR, STE and TM1 are colored in blue, green and red, respectively. Binding sites in MoPrP^{C} identified for vitronectin and STI1 are underlined in black and blue, respectively. **B)** PrP^{C} is a N-glycosylated protein attached to the cell surface by the GPI anchor. **C)** Four topology forms of PrP: ^{Sec}PrP, ^{Ntm}PrP, ^{Ctm}PrP and ^{cy}PrP. **D)** Molecular structure of HuPrP^{C}. The folded GD structure resolved by solution-NMR[55] is shown in cartoon, while the disordered N-term and 3 C-terminal residues are sketched in dashed lines. The two signal peptides are now shown. **E)** Putative "inter-repeat" Cu^{2+} binding mode in the HuPrP^{C} OR region.

2.2.2 Mammalian PrP^{C} structure and functional domains

Solution-NMR studies have resolved the molecular structure of HuPrP^{C}[55; 56], MoPrP^{C}[57; 58], SHaPrP^{C}[59], and PrP^{C} of many other mammalian species such as sheep (Ov)[60], bovine (Bo)[61; 62], etc. These structures have closely similar folds in the C-terminal globular domain (GD, residues 125-230), which contains a bundle of three α-helices and two short antiparallel β-strands (Fig. 2.1D). A disulfide bridge between α2 and α3 helices stabilizes the GD. The NMR studies also found that the N-terminal domain of PrP^{C} (N-term, residue 23-124) is largely unstructured and flexible in all these species.

PrP^{C} is highly abundant in the developing and mature nervous system where it is expressed by neuronal and glial cells[20]. After intensive studies over the years, the physiological function of PrP^{C} is still surrounded by uncertainties. As thoroughly reviewed by Aguzzi[20], *in vivo* and *in vitro* experiments have suggested a range of possible neuroprotective functions of PrP^{C}, such as antiapoptotic or antioxidative effects, neurotransmitter release in synapses, mediation of neurite outgrowth and survival, maintenance of the white matter, as well as functions in cell adhesion. Correspondingly, PrP^{C} is hypothesized to perform its functions through interactions with its cellular partners. Experiments using Hu, Mo, SHa or Bo PrP^{C} and different methods such as yeast-two hybrid, coimmunoprecipitations and chemical cross-linking have proposed a large number of interaction partners which include proteins, metal ions, membrane, lipids, nucleic acids and small molecules[20]. Many of these partners are suggested to interact with the intrinsic disordered N-term[14], which is therefore believed to be a broad-spectrum molecular sensor which endows PrP^{C} with multiple functions[14].

Chemicals which have been established to interact with PrP^{C} N-term *in vivo* include sulphated glycosaminoglycans (GAG)[63] and Cu^{2+} ions[e; 34; 42]. Sulphated GAG is detected in PrP^{Sc} plaques in Hu and SHa brains infected by GSS or CJD[63]. Solution-NMR demonstrates that herapin, a sulphated GAG, binds to SHaPrP^{C} N-term at residues K23, K24, K27, K101, and histidine residues in the so-called octapeptide-repeat region (OR region, residues 51-90, Fig. 1B)[70]. Cu^{2+} ions, whose concentration in the synaptic cleft may be as high as 100-250 μM[71], bind to Hu/Mo/SHa PrP^{C} *in vivo*[72; 73; 74; 75]. Remarkable decreases of Cu^{2+} are observed in the *post mortem* brains of patients affected by TSE[72] or in brains of scrapie-infected mice[73; 76]. Copper-poor diets were found to reduce the survival time of scrapie-infected mice significantly, whereas enhanced administration of copper significantly delayed prion disease onset[76]. Alterations of the Cu^{2+}-PrP^{C} binding process has been suggested to play a role in PrP^{C} aggregation[77; 78], antioxidant activity regulation[73; 79; 80] and synaptic function[81; 82], affecting PrP^{Sc} formation, amplification and the disease onset[75; 77; 78; 83; 84; 85; 86; 87]. On the other hand, it may affect cellular copper uptake and trafficking[82; 88]. As many as six equivalent Cu^{2+}

---

[e] Other metal ions such as Zn^{2+} have also been associated with Hu/Mo/SHa PrP^{C}[64], but their relevance *in vivo* has not been firmly established. Among these, Zn^{2+} shows the next highest affinity for HuPrP^{C}[65]. Zn^{2+} ions bind exclusively in the OR region in the inter-repeat mode with a much lower affinity (~200 μM) than Cu^{2+}[66]. The weak Zn^{2+} binding could still be relevant *in vivo* because of the high concentration of zinc in the brain which reaches up to 300 μM at peak levels in the synaptic cleft of glutaminergic neurons[67]. Cu^{2+} and Zn^{2+} distribution in Mo brain is dependent on MoPrP^{C} expression levels[68]. At low Cu^{2+} availability, Zn^{2+} is able to modify the Cu^{2+} distribution among the possible Cu^{2+} binding modes in SHaPrP^{C}[66]. On the other hand, Zn^{2+} is shown *in vitro* to drive the N-term of MoPrP^{C} to interact with GD[69].

ions can bind in PrP^C N-term at physiological pH[89], in the presence or absence of the GD[90; f]. Four of these binding sites are in the OR region, each containing an HGGGW motif centered at the histidine residue (either H60, H68, H76 or H84)[95]. At low Cu^{2+} availability, however, these four histidines may cooperate to bind one Cu^{2+} ion in a so-called "inter-repeat" binding mode (Fig. 2.1E) with K_d in nanomolar range (several-magnitude lower than that of a single repeat)[89]. The other two binding sites are located at residues H96 and H111 with the affinity comparable to the inter-repeat binding mode[89]. The metal coordination at H96 and H111 is highly debated, as it appears to depend on the length and sequence of the PrP fragments used and on the buffer and pH[89; 90; 94; 96; 97; 98; 99; 100; 101]. A recent study from our group found that these two binding sites could also cooperate to bind one Cu^{2+} ion[102]. Interestingly, this study also shows that a pathogenic mutation Q212P in HuPrP^C GD could influence the Cu^{2+} coordination at the H96 and H111 binding sites[102], implicating a role of abnormal Cu^{2+} binding in the pathology of familial Hu TSE.

MoPrP^C N-term, which has 93% sequence identical to HuPrP^C N-term (Fig. 2.1A) binds to a variety of proteins[14]. Its protein partners are assumed to bind also to HuPrP^C N-term[14]. These proteins include (i) vitronectin[103], a secreted extracellular glycoprotein, possibly supporting axonal growth[104]. Vitronectin interacts with MoPrP^C residues 104–118 (in Mo numbering, Fig. 2.1) with an estimated K_d of 12 nM[104]. (ii) The stress-inducible protein 1 (STI1)[105], a secreted heat-shock protein, triggering neuroprotection[105]. STI1 binds at MoPrP^C residues 114–127 (Fig. 2.1A) with an estimated K_d of 140 nM[105]. (iii) The Amyloid beta (Aβ) multimers that are related to Alzheimer's diseases[106; 107; 108]. Aβ multimers have been suggested to bind residues 95-110 of MoPrP^C N-term[106; 107; 108]. (iv) The lipoprotein receptor-related protein 1 (LRP1), a transmembrane protein, which is associated with a clathrin-dependent internalization mechanism[109]. LRP1 binds to MoPrP^C N-term with nanomolar affinity[109]. (v) The neural cell adhesion molecule (NCAM)[110]. The binding may play a role in the cell adhesion process. NCAM binds to both N-term and GD of MoPrP^C[32]. The MoPrP^C binding sites for LRP1 and NCAM have not been precisely determined[14].

Despite the enormous investigations that have provided knowledge in PrP^C function and functional domains, the lack of structural information on the N-term has long hampered progress towards a full comprehension.

---

[f] Few studies suggest that Cu^{2+} can bind also to PrP^C GD[91; 92], for which the binding affinity is estimated to be ~8 μM using an N-terminal truncated murine MoPrP(121-231). However, the full-length MoPrP(23-231) is almost saturated at 5 μM Cu^{2+} concentration under the same experimental conditions. The Cu^{2+} binding at GD (above 5 equivalent Cu^{2+}) is considered to be hardly physiologically relevant [93; 94], thus it is not discussed here.

2.3 PrP^{Sc} strains and infectivity

2.3.1 PrP^{Sc} strains

The PrP^{C}→PrP^{Sc} conversion is the result of a post-translational process whereby most α-helical motifs in PrP^{C} are replaced by β-structures[27; 111]. To date, the molecular structure of PrP^{Sc} is unknown[112]. Several structural models have been proposed (Fig. 2.2) based on MD simulations[113] or low-resolution biophysical techniques such as X-ray fiber diffraction[114], electron microscopy (EM)[115] and Hydrogen-deuterium exchange mass spectroscopy (HX-MS)[116]. In fact, PrP^{Sc} exist in different conformations related to multiple PrP^{Sc} strains that are distinct in many aspects[117; 118]. *In vivo*, different strains are distinguished by differential incubation periods and distinct patterns of pathological phenotypes[119; 120; 121]. *In vitro*, they demonstrate distinct biochemical features, as distinguished by proteinase K digestion, glycosylation pattern, thermal or chemical denaturation, conformation-dependent immunoassay, infrared spectroscopy and metal binding[122; 123; 124; 125]. Indeed, these features are strain-specific because they are maintained when a strain passes across several mammalian species including humans, such as in the case of BSE strain[120]. Similarly, when Tg mice expressing a chimeric Mo/Hu hybrid PrP (denoted Mo/HuPrP chimera) are inoculated with different HuPrP^{Sc} strains, the biological features of each specific strain are preserved[126]. These findings along with other lines of experimental data strongly suggest that PrP^{Sc} strains derive from the capacity of PrP^{Sc} to adopt different conformational and aggregation states[117; 118].

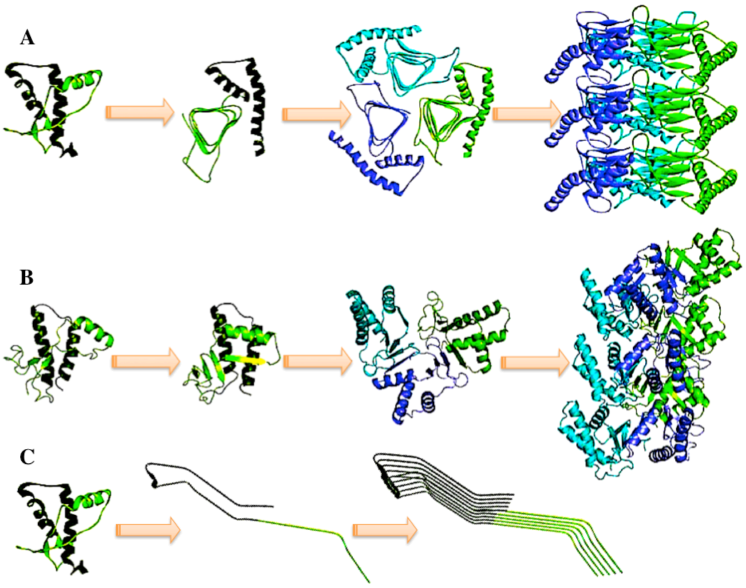


Figure 2.2: The conformational conversion from PrP^{C} to three proposed models of PrP^{Sc}: **A)** the β-helical model[115], **B)** the β-spiral model[113] and **C)** the parallel in-register extended β-sheet model[116]. This figure is adapted from ref. [112].

2.3.2 The "species barrier" and the "dominant-negative" effect

Although PrP is highly conserved in mammalian species, transmission of TSE across species is typically far less efficient than within species[127]. This is known as the "species barrier", which was presumed to arise from differences in the primary sequence across species. This issue has received wide attention since the discovery that vCJD in humans was caused by the BSE strain in cattle[22; 23]. This strain is able to transmit efficiently to a range of species while maintaining its biological features[128]. Another striking example of the prion transmission across species came from several studies using Tg mice expressing chimeric Mo/HuPrP[10; 11; 12; 129]. Expressing HuPrP^{C} in Tg mice did not abbreviate the incubation time upon infection with HuPrP^{Sc}, indicating that these Tg mice are not more susceptible to HuPrP^{Sc} than the WT mice[129]. By contrast, in the same study Tg mice expressing a Mo/HuPrP chimera were found highly susceptible to HuPrP^{Sc}. Tg mice expressing MoPrP chimeras containing mutations from other species have been reported to exhibit both transmission barriers against certain prion strains[10] and spontaneous conversion into prions[11; 12]. Investigation in Mo neuroblastoma (ScN2a) cells showed that upon MoPrP^{Sc} infection the "species barrier" could be raised by single-residue substitutions in MoPrP with Hu or Ov residues [13]. More intriguingly, some of these MoPrP chimeras also acted as the so-called "dominant-negatives"[g] to protect the endogenous WT MoPrP^{C} from being converted to MoPrP^{Sc}.

These facts argue that the transmission barrier across species is not correlated to the differences in the primary sequence. In the late 1990s, Collinge already commented that "species-strain barrier" or "transmission barrier" might be more appropriate than "species barrier" for describing the effects[134]. Accumulating experimental observations suggest that the "species barrier" is a conformational barrier related to distinct PrP^{Sc} conformations[31; 117]. Therefore, PrP^{Sc} strains and transmission barrier are different observations of the same origin[31]. Then how can one or few residue substitutions such as in the case of MoPrP chimeras determine the susceptibility of the host to a same PrP^{Sc} strain?

It has been proposed that the loop between the $\beta2$ strand and the $\alpha2$ helix in PrP^{C} (denoted $\beta2$-$\alpha2$ loop, residue 165-175, Fig. 2.1 and 1-2C) mediates the transmission barrier. This was first supported by solution-NMR studies which found that this loop could be highly flexible in some mammalian species including Hu[55; 56] and Mo[57; 58] or well structured in some other species such as

---

[g] The dominant-negative effects of residue substitutions on PrP^{C} have been reported and deeply investigated in humans and in sheep. In humans, the naturally occurring polymorphism E219K was found to protect against sCJD when expressed in heterozygosis[130; 131]. In Ov, three polymorphisms are acutely linked to the occurrence of Ov scrapie: A136V, R154H and Q171R/H. These generate five commonly observed alleles: ARQ, ARR, AHQ, ARH and VRQ. ARR and AHQ are associated with resistance to Ov scrapie, while ARQ, ARH and VRQ are associated with susceptibility[132; 133].

SHa[135], elk[136], bank vole[137], wallaby[138], rabbit[139] and horse[140]. Moreover, the C-terminal part of α3 helix (residue 215-228) might modulate the β2-α2 loop conformation by long-range interactions[138]. Single-residue substitutions within these two regions of MoPrP^C could cause the otherwise flexible β2-α2 loop to adopt rigid structures[137; 138; 141]. Remarkably, Tg mice denoted tg1020 which overexpress MoPrP containing two of these substitutions, S170N and N174T, developed spontaneous and progressive prion-like disease[11]. This *de novo* generated disease in tg1020 mice could be transmitted to Tg mice overexpressing WT MoPrP and then to WT mice in a second passage. By contrast, tg1020 mice exhibited a very prolonged incubation time when infected with WT MoPrP^{Sc} strain, RML[11], arguing that the artificial mutations generated transmission barrier for the RML strain. Later studies where the tg1020 mice were challenged with PrP^{Sc} from five different species (Mo, SHa, Bo, Ov and deer) demonstrated altered transmission barriers compared to Tg mice overexpressing WT MoPrP[10; 142]. These *in vivo* studies further support the hypothesis that the β2-α2 loop is involved in the conformational selection for PrP^{Sc} strains.

The finding that expressing HuPrP does not render Tg mice more susceptible to HuPrP^{Sc} [129] suggests that other factors might be involved in the PrP^C→PrP^{Sc} conversion. To explain this phenomenon, the concept of "protein X", an unknown factor that mediates the PrP^C→PrP^{Sc} conversion, was proposed. Although many interaction factors have been suggested for PrP^{Sc}, the existence of protein X has never been proved[20]. Nevertheless, more and more evidence indicates that PrP^{Sc} does not replicate by directly imposing its structure onto PrP^C[36]. More future researches are required to reveal the complicated picture of PrP^{Sc} propagation.

2.4 Mutations and polymorphism in the human *PRNP* gene

The familial types of Hu TSE (including fCJD, GSS, FFI and PrP-CAA) account for about 15% of the cases, which are caused by pathogenic mutations (PMs) in the *PRNP* gene[15; 143] (Fig. 2.3).

Most of the currently identified PMs are missense mutations coding for a single-residue substitution. As many as 33 missense PMs are located in HuPrP^C GD. *In vitro,* they may accelerate the protein misfolding process[144; 145; 146]. Regarding the structural role of these PMs, NMR[147; 148; 149] and MD[16; 17; 150; 151; 152] studies on HuPrP^C or MoPrP^C have revealed that PMs affect secondary structure elements of the GD and render the GD more flexible than that in the WT. This increases the propensity of the protein for misfolding, consistent with experimental evidence[145; 146; 153]. Seven missense PMs are instead located in the N-term of HuPrP^C: Q42P is likely associated with fCJD[154]; G114V is reported with both fCJD- and GSS-like phenotypes[155; 156], whereas P102L[h], P105L/S/T,

---

[h] The correspondent mutation of P102L in MoPrP^C alters the incubation time in transgenic mice upon infection by PrP^{Sc} from Hu, SHa, Ov or murine sources in a strain-dependent manner[157; 158].

and A117V are associated with GSS[159]. They cause the disease without affecting the kinetics of the $HuPrP^C \rightarrow HuPrP^{Sc}$ conversion *in vitro*[145; 146; 160]. This hints at the presence of processes *in vivo* that lead these mutations *indirectly* to the formation of $HuPrP^{Sc}$, the hallmark of Hu TSE. However, structural information on the effects of these mutations is currently lacking. GSS patients carrying A117V have an abnormally high level of ^{Ctm}PrP in the brain[43]. P102L and G114V have also been shown to increase the formation of ^{Ctm}PrP *in vitro*[161; 162]. As mentioned in section 1.2, $HuPrP^C$ N-term on the one hand contains the TM1 and STE regions, which regulate the protein topology; on the other hand it interacts with a broad range of cellular partners such as Cu^{2+}, sulphated GAG and likely several protein partners. Therefore, the PMs in $HuPrP^C$ N-term might affect the protein topology and/or alter the interaction with cellular partners, thus affecting the protein function. Another 3 missense PMs located in the C-terminal signal peptide are not discussed here, since the signal peptide is cleaved off during $HuPrP^C$ biosynthesis.

Five PMs are nonsense mutations, Y145stop, Q160stop, Y163stop, Y225stop and Y226stop, which code to prematurely terminate the protein sequence at the mutation position[163; 164]. Although these PMs are rare, they all cause peculiar vascular amyloid fibril deposits generally defined cerebral amyloid angiopathy (CAA) that also occurs in other neurodegenerative disorders including Alzheimer's disease[164].

In the OR region, from one to nine OR insertions have been reported. Most of the OR insertions lead to fCJD, except for an 8-OR insertion whose phenotype is unclassified. A 2-OR deletion in this region might cause fCJD[165], albeit OR deletions are not yet firmly associated to Hu TSE.

A methionine/valine polymorphism at codon 129 (129M/V) appears to influence Hu TSE phenotype when coupled with certain PMs, although the polymorphism itself does not promote the disease[159]. Another polymorphism, E219K, present in ~6% of the Japanese population appears to be protective against sCJD[131]. A polymorphism at codon 127 (G127V) reported only in the natives of Papua New Guinea is protective against Kuru[166]. Other mutations have been reported in *PRNP* that are not clearly polymorphism or pathogenic[15; 143], thus they are not discussed here.

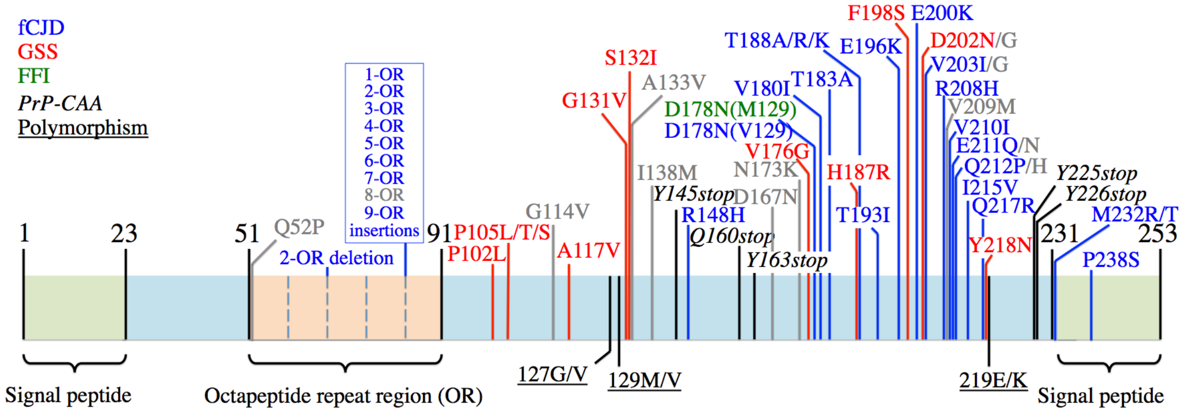


Figure 2.3: Polymorphism and pathogenic mutations in Hu *PRNP* gene. Mutations associated with fCJD, GSS, FFI and PrP-CAA are colored in blue, red, green, and *black* (italic), respectively. Polymorphisms are underlined. Mutations with unclassified disease phenotypes are colored in gray.

Chapter 3 Materials and methods

3.1 Molecular simulation of proteins

Proteins are central to the health and function of cellular life. These complex macromolecules serve a fascinating variety of functions, which lie in their diverse molecular structures. Researchers have sought to uncover the structure–function relationships of proteins at atomic level using techniques such as high-resolution X-ray crystallography and NMR spectroscopy. Nevertheless, proteins are characterized by their dynamic features rather than static structures[167]. A large part of the properties of a protein or a biomolecular system is beyond the access of experimental measurements that in general correspond to averages of a quantity over space and time. Moreover, a rapidly increasing number of proteins are being identified as (at least partially) "intrinsically disordered", which lack stable tertiary structures *in vitro*[168]. These intrinsically disordered proteins perform essential functions by virtue of their flexibility[168]. However, they represent a major challenge to structural biology. The science and art of molecular simulations has revolutionized the understanding in protein dynamics and function. Through model building and computation, molecular simulations have brought to reality the dream to 'watch' proteins in action at atomic resolution[169]. On the one hand, computer simulations can determine not only averages but also distributions and often time series of any definable quantity. Thus, they represent an indispensable tool to interpret experimental data by providing insights into the experimental observations. On the other hand, computer simulations are widely used to predict protein structures and properties that are difficult or expensive to realize under environmental conditions.

Protein simulations encompass *ab initio* and semi-empirical quantum mechanics, empirical molecular mechanics (MM) and many other established procedures and methods. Here, only all-atom MM simulations will be described, which are the methods employed in this thesis. The term MM refers to the use of classical potential energy functions (e.g., harmonic oscillator or Coulombic potentials) to model molecular systems[170]. The potential energy of a system in MM is calculated using force fields. In all-atom MM methods, the system typically consists of spherical atoms each assigned with a radius (usually the van der Waals radius), polarizability, and a constant net charge. The bonded interactions are represented by springs connecting the atoms, and the non-bonded atoms might be treated as inelastic hard spheres or may interact according to a Lennard-Jones potential. Each force field is parameterized to be internally consistent, typically through agreement with some experimental values and theoretically calculated results. MM approaches are widely applied in molecular dynamics simulations, Monte Carlo simulations, and ligand-docking simulations.

The development of protein simulation has been continuously driven by the steady growth of computing power as well as the advancement in force fields, algorithms and experimental techniques[171]. Protein modeling and simulation has expanded in both quality and scope, albeit caveats and inherent approximations are clearly recognized in the field[169]. Today, protein simulations are capable of handling more and more complex biomolecular systems and achieving profound comprehension of intriguing questions in biological systems. These questions range from resolving the equilibrium structure of a small peptide to the kinetics of protein folding and the functioning of a supramolecular aggregate; from the energetics of hydrogen-bond formation in proteins to the binding affinities of ligands/drugs to their target. Combined with constantly evolving experimental techniques, computer simulations are indispensable tools to pursue many fundamental questions concerning the biological motions and functions of complex systems such as enzyme-substrate complexes, ion channels, transmembrane receptors and ribosomes[169].

3.2 Classical molecular dynamics simulation

3.2.1 Introduction

In statistical mechanics, experimental observables are described in terms of ensemble[i] averages taken over a large number of microscopic states of the system considered simultaneously. Molecular dynamics (MD) simulations generate a sequence of microscopic states of the system in phase space as a function of time. These states correspond to different configurations belonging to the same ensemble and satisfy the conditions of a particular thermodynamic state. MD simulations, in the most widely used versions, determine the trajectories of atoms and molecules by numerically solving the Newton's equations of motion for a system of N interacting particles. Internal forces between the particles and potential energy of the system are defined by force fields. The results of MD simulations can be used to quantify macroscopic thermodynamic properties of the system assuming the ergodic hypothesis is valid for it: if the system is allowed to evolve in time indefinitely, it will eventually pass through all possible states and the time averages of the system are equal to the statistical ensemble averages. An ensemble average is given by

$$\langle A \rangle_{ensemble} = \iint dp^N \, dr^N A\left(p^N, r^N\right) \rho\left(p^N, r^N\right), \tag{3.1}$$

where $A\left(p^N, r^N\right)$ is the observable of interest, and is expressed as a function of the momenta, $p^N = \left(p_1, p_2, \ldots p_N\right)$, and the positions, $r^N = \left(r_1, r_2, \ldots r_N\right)$, of the N particles. Consider a canonical ensemble that describes the statistical distribution in a system that can exchange energy via very weak contact with a temperature bath and eventually come to equilibrium. It is generally the most

---

[i] An ensemble is a collection of different microscopic states of a system with the probability distribution that corresponds to the macroscopic properties of the system.

useful ensemble in practice since we most often deal with systems in thermal equilibrium (constant T) with their surroundings. The probability density of the canonical ensemble is given by

$$\rho\left(p^N,r^N\right)=\frac{1}{Z}\exp\left[-H\left(p^N,r^N\right)/k_BT\right], \tag{3.2}$$

where H is the Hamiltonian, T is the temperature, k_B is Boltzmann's constant and Z is the partition function

$$Z=\iint dp^N\,dr^N\exp\left[-H\left(p^N,r^N\right)/k_BT\right], \tag{3.3}$$

where the integral is generally very difficult to calculate because all possible microstates of the system must be considered. Alternatively in MD simulations, the microstates in the ensemble are calculated sequentially in time, which gives the time average of the observable A:

$$\langle A\rangle_{time}=\lim_{\tau\to+\infty}\frac{1}{\tau}\int_0^{\tau}A\left(p^N(t),r^N(t)\right)dt\approx\frac{1}{M}\sum_{t=1}^{M}A\left(p^N,r^N\right), \tag{3.4}$$

where τ is the simulation time, M is the number of frames in the simulation and $A\left(p^N,r^N\right)$ is the instantaneous value of A. The right-end side of Eq. (3.4) accurately approximates the ensemble average only if MD simulations sample a sufficient amount of phase space (and this has to be done using a feasible amount of computer resources), i.e. the ergodicity is satisfied.

MD simulations have the peerless advantage (with respect to other sampling methods like the Monte Carlo method) that they can thoroughly describe protein dynamics[167]. Provided a high-resolution structure as a starting point, the precise position of each atom at any instant of time can be followed, along with the corresponding energies. It can describe the motions of individual particles as a function of time and disentangle the correlations between motions, which are obscured in ensemble-average experiments. MD simulations provide fine details of a system at a precision and on a time scale that are otherwise inaccessible[167]. They serve to extend our understanding of the modeled systems and inspire new experiments. Since the first simulation of a protein in 1976[172], MD simulations, along with a range of complementary computational approaches, have become valuable tools for solving the puzzle of protein structures and functions[170]. They are widely applied to various studies such as molecular design, structure determination and refinement, protein folding, investigation in molecular stability, conformational and allosteric properties, molecular recognition and the properties of complexes[169]. The last few years have seen challenging applications of MD simulations such as those involving complex systems like membrane-embedded receptors in dimer bound with ligand[173], and those that extensively probe into protein folding process at microsecond scale[174].

3.2.2 Force fields

An MD simulation consists of the numerical, step-by-step solution of the Newton's equations of motion

$$m_i \ddot{r}_i = F_i, \tag{3.5}$$

where F_i is the force exerted on particle i, m_i and r_i are the mass and coordinate of atom i, respectively. The force acting on atom i can be expressed as the derivative of the potential energy U of the system with respect to its coordinate r_i:

$$F_i = -\frac{d}{dr_i} U. \tag{3.6}$$

Therefore, the Newton's equations can be rewritten as

$$-\frac{dU}{dr_i} = m_i \frac{d^2 r_i}{dt_i^2}. \tag{3.7}$$

The core of any force field is the functional form of U with respect to the coordinates, r^N. The combination of the potential energy function with the parameters used in that function constitute a *force field*. Biomolecular systems are commonly well described by empirical force fields[175] using an additive potential energy function formed by a collection of simple functions that represent the intramolecular and intermolecular interactions:

$$\begin{aligned}
U\left(r^N\right) &= U_{\text{bonded}}\left(r^N\right) + U_{\text{non-bonded}}\left(r^N\right), \\
U_{\text{bonded}}\left(r^N\right) &= \sum_{\text{bonds}} K_b\left(b-b_0\right)^2 + \sum_{\text{angles}} K_\theta\left(\theta-\theta_0\right)^2 + \sum_{\text{dihedral}} K_\chi\left(1+\cos\left(n\chi-\delta\right)\right) \\
&\quad + \sum_{\text{impropers}} K_{\text{imp}}\left(\phi-\phi_0\right)^2, \\
U_{\text{non-bonded}}\left(r^N\right) &= \sum_{\text{nonbond}} \left[\varepsilon_{ij}\left(\frac{R_{\min_{ij}}}{r_{ij}}\right)^6 - \left(\frac{R_{\min_{ij}}}{r_{ij}}\right)^{12}\right] + \frac{q_i q_j}{\varepsilon r_{ij}}.
\end{aligned} \tag{3.8}$$

In the bonded terms, bonds, angles and out-of-plane distortions (improper dihedral angles) are treated harmonically, and dihedral rotations are described by a sinusoidal term. The non-bonded interactions include a Lennard-Jones (LJ) 6-12 potential, that is typically used to describe the atom-atom van der Waals (vdW) interactions, and a Coulomb potential, which describes electrostatic interactions. In Eq. (3.8), b is the bond length, θ is the valence angle, χ is the dihedral angle, ϕ is the improper angle, and r_{ij} is the distance between atoms i and j. The parameters in Eq. (3.8) represent the actual force field. Parameters in the bonded terms include: the bond force constant and equilibrium distance, K_b and b_0, respectively; the valence angle force constant and equilibrium angle, K_θ, and θ_0, respectively; the dihedral force constant, multiplicity and phase angle, K_X, n, and δ, respectively; and the improper force constant and equilibrium improper angle, K_{imp} and ϕ_0, respectively. Parameters in the non-bonded interactions between atoms i and j include: the LJ well depth, ε_{ij}, and the minimum interaction radius, $R_{\min_{ij}}$, which are used to treat the vdW repulsion and dispersion interactions; and the partial atomic charges, q_i, q_j, and the dielectric constant, ε, which are used in the Coulomb term. Typically, ε_i and $R_{\min_i}$ are obtained for individual atom types and then combined to yield ε_{ij} and $R_{\min_{ij}}$. The terms in Eq. (3.8) are common to the majority of biomolecular force fields, such as the widely used all-atom protein force fields AMBER[176], CHARMM[177] and OPLS[178]. In fact, this simple form for the potential energy function is good enough to allow for studies on systems of 100,000 or more atoms with the current computational capabilities.

The potential energy function along with the parameters allow for the energies and forces of the selected molecules to be evaluated. The approaches employed to calculate the parameters of the force fields ultimately determine their applicability and quality. For a variety of force fields the parameter optimization is based on fitting and comparisons with quantum mechanical results for model compounds representative of the biomolecules of interest. For example, the alanine dipeptide is the quintessential model for optimization of the protein backbone parameters, often being supplemented with data from the glycine and proline dipeptides (Fig. 2.1), as these represent the amino acids in which the covalent structure of the backbone differs from the remaining 18 natural amino acids.

Additional adjustment or alternate terms beyond Eq. (3.8) are often introduced to yield higher accuracy, such as higher order terms to treat the bond and angle terms and/or cross terms between bonds and angles or angles and dihedrals[175].

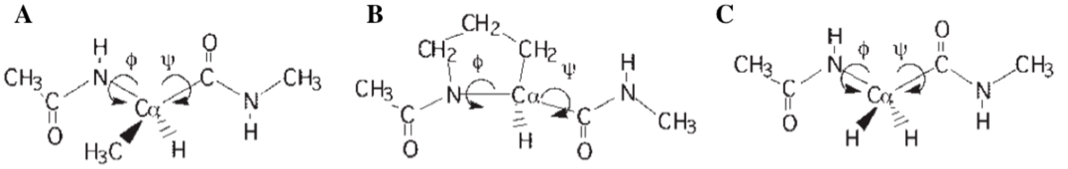


Figure 2.1: A) Alanine, **B)** proline, and **C)** glycine dipeptides often used for parameter development of protein backbone. Shown in the figures are the ϕ, ψ dihedral angles that define the Ramachandran plot[179].

3.2.3 The integration algorithms

There is no analytical solution to the Newton's equations of motion, which must be solved numerically

$$F_i = m_i a_i,$$
$$a_i = \frac{dv_i}{dt}, \tag{3.9}$$
$$v_i = \frac{dr_i}{dt}.$$

Among the many possible methods to perform step-by-step numerical integration of these equations, the favorable ones should take into account the following characteristics: i) there may be short and long timescales, and the algorithm must cope with both; ii) calculating the forces typically involves a sum over pairs of atoms, which is computationally intensive and should be performed as infrequently as possible; and iii) the trajectory of coordinates should stay on the constant-energy hypersurface over long timescales (much longer than typical correlation times of the dynamical properties of interest), in order to sample the correct ensemble. Consistent with these features, the most common simulation algorithms are of low order, i.e. without involving high-order derivatives of positions, which are faster and allow larger time steps without jeopardizing energy conservation. In particular, the old Verlet algorithm[180], which requires only one evaluation of the forces at each time step, fulfills all these requirements and due to its simplicity is one of the most employed nowadays. The Verlet algorithm exists in various, essentially equivalent versions. The leapfrog[181] version of the Verlet algorithm is described below, which is the default algorithm in the Gromacs 4 simulation package[182] used in this thesis. Leapfrog uses positions r_i at time t and velocities v at time t-½Δt, and updates r and v using the forces $F(t)$ determined by r_i at time t:

$$v\left(t+\frac{1}{2}\Delta t\right)=v\left(t-\frac{1}{2}\Delta t\right)+\frac{\Delta t}{m}F\left(t\right), \tag{3.10}$$

$$r_i\left(t+\frac{1}{2}\Delta t\right)=r_i\left(t\right)+\Delta t v\left(t+\frac{1}{2}\Delta t\right). \tag{3.11}$$

It produces trajectories that are identical to the original Verlet algorithm[180]:

$$r_i\left(t+\Delta t\right)=2r_i\left(t\right)-r_i\left(t-\Delta t\right)+\frac{1}{m}F\left(t\right)\Delta t^2+O\left(\Delta t^4\right). \tag{3.12}$$

Therefore, the leapfrog algorithm is (locally) a third order method and it is time-reversible. The advantage of leapfrog with respect to the original Verlet algorithm is that the velocities are calculated explicitly. The disadvantage is that velocities are not calculated at the same time as positions, therefore it cannot calculate kinetic energy contribution to the total energy at the same time as the positions are defined. The equations of motion are modified for temperature coupling and pressure coupling, and extended to include the conservation of constraints, all of which will be described in later sections.

3.2.4 Periodic boundary conditions and neighbor list

Biomolecular MD simulations are typically performed in small-sized systems, which means that atoms on the outer surface can have large artificial effects on the measured properties. Periodic boundary conditions (PBC) are usually employed to eliminate the surface effects, unless the behaviors near the surface are of particular interest. In PBC, a so-called "primary box" is generated which contains the molecules of interest and the bulk material. The primary box is replicated infinitely in all directions to give a periodic array. Thus the primary box is surrounded by images of itself with the same size, shape and content (Fig. 2.2A). The shape of the boxes is space-filling and all the boxes are separated by open boundaries so that particles can freely enter or leave any box. In the course of the simulation, if atoms leave the box, their images simultaneously enter the box through the opposite side. PBC enables a simulation to be performed using a relatively small number of atoms in such a way that the atoms experience forces as if they were in a bulk solution. Assuming the interactions are of short-range (treatment of long-range interactions will be discussed in a later section), the "minimum-image convention" is applied. In fact, with short-range interactions a cutoff distance, r_{cut}, can be introduced to save computational resources, so that interactions beyond the cutoff range ($r_{ij} > r_{\mathrm{cut}}$) are not calculated. If the cutoff is not longer than half of the shortest box vector, then in order to calculate the interaction of an individual particle i with all the replicas of another particle j it is sufficient to consider only the closest image among the particles j to i.

In principle, computing the non-bonded terms in Eq. (3.7) involves the evaluation of a large number of pairwise interactions between each atom i and all interacting atoms j. Although the use of cutoff avoids expensive calculations, it is still time consuming to examine and calculate all distinct pairs, which are $\frac{1}{2}N(N-1)$ in a system of N atoms. The "neighbor list" technique was introduced by Verlet to improve the computing speed. The cutoff range r_{cut} gives a sphere of radius r_{cut} around an atom i within which we can find all the atoms j that can interact with i. A larger range r_{list} is defined to identify a larger sphere around atom i for determining a list of neighbor atoms (Fig. 2.2B) as possible interacting atoms. At the first step in a simulation, for each atom i a list of all the neighbors j is constructed (for which $r_{ij} < r_{\mathrm{list}}$). Over the next few MD time steps, only pairs between i and the atoms appearing in the corresponding neighbor list are examined for the force computation. The neighbor lists have to be reconstructed rather frequently, before an unlisted atom can enter the interaction range r_{cut} of the corresponding atom. Therefore, the choice for the best value of r_{list} has to take into account both the value of the updating frequency of the neighbor list and the computing time needed to evaluate the interactions of all the pairs defined by r_{list}.

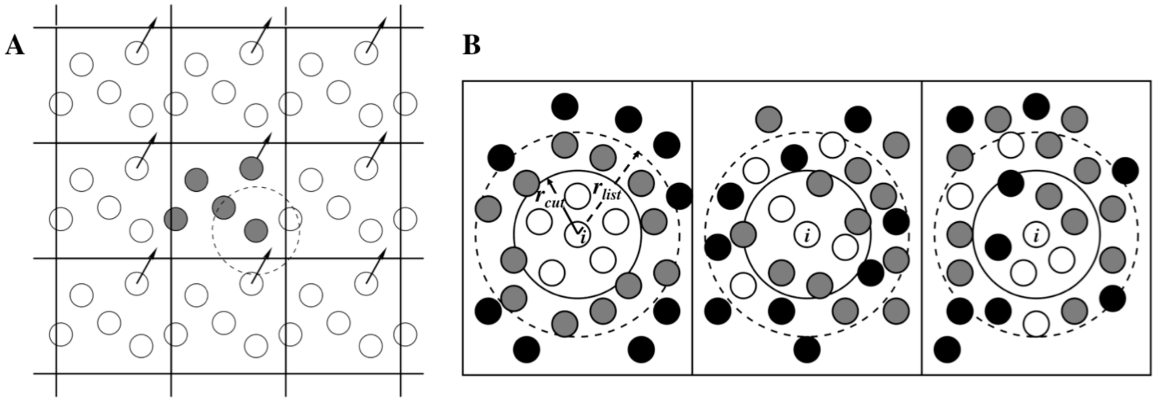


Figure 2.2: **A)** 2D illustration of periodic boundary conditions applied on a cubic primary box. When an atom moves out of the simulation box, an image particle moves in from the opposite side to replace it. **B)** 2D illustration of the neighbor list on its (re)construction (left) and a few MD time steps later (middle). Interacting atoms (white circles) and atoms in the neighbor list (gray filled circles) of atom *i* are indicated. The neighbor list must be reconstructed before any unlisted atoms (black filled circles) have entered the range of r_{cut}, as shown on the right.

3.2.5 Constrains

In biomolecular simulations, it is common practice to constrain intramolecular bonds with very high vibration frequencies in order to improve the computational efficiency. For instance, bonds involving hydrogen atoms require a time step of 0.5 fs in order to conserve energy. Instead of treating these bonds by terms in the potential function, they can be constrained to have a fixed length. This allows for larger time steps up to 2 fs and significantly reduces the computing costs of simulations of given timescales.

3.2.5.1 The SHAKE algorithm

The SHAKE algorithm[183] was the first one developed to satisfy bond geometry in MD simulations. Using a set of coordinates (r_i) and a set of new unconstrained coordinates (r_i') generated at the next time step, SHAKE changes r_i' to a set of coordinates r_i'' that fulfills a list of distance constraints. For the original Verlet algorithm, SHAKE determines the constraint forces in such a way that the constraints are satisfied at the end of each time step.

Let us take a fixed bond length *b* between atoms 1 and 2 as an example. A constraint equation and the time derivative of the constraint can be written as

$$\sigma(r_1, r_2) = (r_1 - r_2)^2 - b^2 = 0, \tag{3.13}$$

$$\dot{\sigma}(r_1, r_2) = 2(v_1 = v_2)(r_1 - r_2) = 0. \tag{3.14}$$

Assuming the equations of motion must fulfill *n* holonomic constraints:

$$\sigma_k\left(r_1 \ldots r_2\right)=0; \quad k=1 \ldots n, \tag{3.15}$$

the constraints are added as zero terms to the potential $U(r)$, multiplied by Lagrange multiplier $\lambda_k\left(t\right)$, and consequently the forces are now given by

$$F\left(t\right)=-\frac{\partial}{\partial r_i}\left(U+\sum_{k=1}^{n}\lambda_k\sigma_k\right), \tag{3.16}$$

where λ_k must be solved to fulfill the constraint equations Eq. (3.15). The second part of the right hand side determines the constraint forces G_i

$$G_i=-\sum_{k=1}^{n}\lambda_k\frac{\partial\sigma_k}{\partial r_i}. \tag{3.17}$$

In the Verlet or leapfrog algorithm, the evolution of the unconstrained coordinates r_i' and the constrained coordinates r_i'' (on which are applied the constraint forces G_i) are respectively:

$$r_i'\left(t+\Delta t\right)=2r_i'\left(t\right)-r_i'\left(t-\Delta t\right)+\frac{F'\left(t\right)}{m}\Delta t^2, \tag{3.18}$$

$$\begin{aligned}r_i''\left(t+\Delta t\right)&=2r_i''\left(t\right)-r_i''\left(t-\Delta t\right)+\frac{F\left(t\right)}{m}\Delta t^2\\&=2r_i''\left(t\right)-r_i''\left(t-\Delta t\right)+\frac{F'\left(t\right)+G\left(t\right)}{m}\Delta t^2.\end{aligned} \tag{3.19}$$

SHAKE solves the Lagrange multipliers (and hence the displacements $G_i(t)\,\Delta\, t^2/m$) by solving a set of coupled constraint equations iteratively to satisfy each constraint equation in turn until convergence to a given tolerance.

However, SHAKE has the drawback that no solutions may be found when displacements are large[184]. This is because the coupled bonds are handled one by one and correcting one bond may tilt a coupled bond up to a point that the method does not converge.

3.2.5.2 The LINCS algorithm

The MD simulations in this thesis employ the LINCS (Linear Constraint Solver) algorithm[184] which is the default constraint method in Gromacs 4. LINCS is faster and more stable than SHAKE, although it can only be used with bond constraints and isolated angle constraints, such as the proton angle in OH. Similar to SHAKE, LINCS resets new bonds to their correct lengths after an unconstrained update. The method is non-iterative, as it always uses two steps (Fig. 2.3). The first step sets the projection of the new bonds onto the directions of the old bonds. The second step corrects the bond lengths to the prescribed lengths.

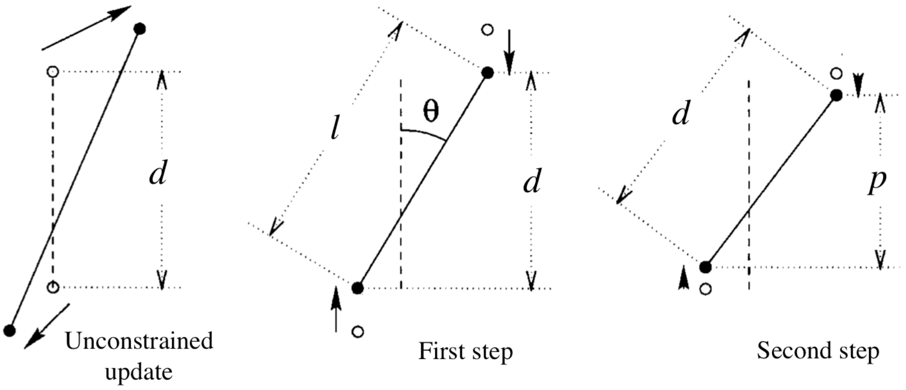


Figure 2.3: Schematic illustration of the two-step reset of constrained bonds by LINCS after an unconstrained update for one time step. The old bonds are shown in dashed line and the new bonds in solid lines. $l = d \cos \theta$ and $p = \sqrt{2d^2 - l^2}$.

The equations of motion Eq. (3.5) can be rewritten as

$$\frac{d^2\mathbf{r}}{dt^2} = \mathbf{M}^{-1}\mathbf{F}, \tag{3.20}$$

where $\mathbf{r}$ is the $3N$ coordination vector, $\mathbf{F}$ is the $3N$ force vector and $\mathbf{M}$ is a $3N \times 3N$ diagonal matrix containing the masses of the atoms. Eq. (3.16) then becomes

$$\mathbf{M}\frac{d^2\mathbf{r}}{dt^2} = -\frac{\partial}{\partial \mathbf{r}}\left(U + \boldsymbol{\lambda\sigma}\right), \tag{3.21}$$

Introducing a new notation for the gradient matrix of the constraint equations $B_{hk} = -\partial\sigma_h / \partial r_k$, then $\mathbf{B}$ is an $n \times 3N$ matrix containing the directions of the constraints. Eq. (3.21) can now be simplified as

$$-\mathbf{M}\frac{d^2\mathbf{r}}{dt^2} + \mathbf{B}^T\boldsymbol{\lambda} + \mathbf{F} = 0. \tag{3.22}$$

Because the constraint equations equal zero, the first and second time derivatives of the constraints are zero as well:

$$\frac{d\boldsymbol{\sigma}}{dt} = \mathbf{B}\frac{d\mathbf{r}}{dt} = 0, \tag{3.23}$$

$$\frac{d^2\boldsymbol{\sigma}}{dt^2} = \mathbf{B}\frac{d^2\mathbf{r}}{dt^2} + \frac{d\mathbf{B}}{dt}\frac{d\mathbf{r}}{dt} = 0. \tag{3.24}$$

To solve for $\boldsymbol{\lambda}$, left-multiply Eq. (3.22) by $\mathbf{BM}^{-1}$ and use Eq. (3.24) to get

$$\frac{d\mathbf{B}}{dt}\frac{d\mathbf{r}}{dt} + \mathbf{BM}^{-1}\mathbf{B}^T\boldsymbol{\lambda} + \mathbf{BM}^{-1}\mathbf{F} = 0. \tag{3.25}$$

Thus the constraint forces are

$$\mathbf{B}^T\boldsymbol{\lambda} = -\mathbf{B}^T\left(\mathbf{B}\mathbf{M}^{-1}\mathbf{B}^T\right)^{-1}\mathbf{B}\mathbf{M}^{-1}\mathbf{F} - \mathbf{B}^T\left(\mathbf{B}\mathbf{M}^{-1}\mathbf{B}^T\right)^{-1}\frac{d\mathbf{B}}{dt}\frac{d\mathbf{r}}{dt}, \qquad (3.26)$$

which can be substituted into Eq. (3.22) to give the constrained equations of motion. Defining an abbreviation $\mathbf{T} = \mathbf{M}^{-1}\mathbf{B}^T\left(\mathbf{B}\mathbf{M}^{-1}\mathbf{B}^T\right)^{-1}$, the equations of motion become

$$\frac{d^2\mathbf{r}}{dt^2} = \left(\mathbf{I} - \mathbf{T}\mathbf{B}\right)\mathbf{M}^{-1}\mathbf{F} - \mathbf{T}\frac{d\mathbf{B}}{dt}\frac{d\mathbf{r}}{dt}. \qquad (3.27)$$

I-TB is a projection matrix that sets the constrained coordinates to zero, $\mathbf{BM}^{-1}\mathbf{F}$ is an n vector of second derivatives of the bond lengths in the direction of the bonds, and **T** is a $3N \times n$ matrix that transforms motions in the constrained coordinates into motions in Cartesian coordinates without changing the equations of motion of the unconstrained coordinates. The last term in Eq. (3.27) represents centripetal forces caused by rotating the bonds.

3.2.6 Long-range interactions

Non-bonded interactions are extremely important for understanding the properties of biomolecules. They can be divided into short-range and long-range interactions. In a 3D system, a force is defined to be short ranged if it decreases with distance quicker than r^3. As mentioned earlier, short-range interactions are commonly handled by imposing a cutoff, r_{cut}, to the potential, beyond which the potential is set to 0. Long-range interactions, however, present a more complicated problem because a cutoff could introduce a considerable energetic discontinuity into the system, which leads to unstable simulations. The determination of these interactions (which have to be treated as infinite-ranged) represents the most computationally expensive part of empirical force field calculations: the simulation cell and all its periodic images must be considered. The neighboring-list technique would diminish the effect of cutting off the potential over a range of several tens or hundreds of nanometers, while current biomolecular simulations typically have box lengths of the order of few nanometers. Thus, to account for long-range interactions, one must consider the effect of the periodic images.

With respect to the long-range LJ term, the region beyond the cutoff distance is typically treated as being homogenous (i.e., an “average” over all LJ atom types in the system)[185; 186]. This is realized by introducing corrections to account for the truncated LJ contribution[185; 186]. It is often assumed that the LJ contribution at longer distances is insignificant due to the r^6 distance dependence. However, it should be emphasized that the summation over all the truncated contributions becomes significant, especially in systems dominated by aliphatic groups, as in lipids. For the electrostatic interactions, the Ewald summation method introduced by Ewald in 1921[187] is of particular importance in this field. It takes advantage of crystal symmetry combined with reciprocal space to treat the long-range electrostatic interactions of systems satisfying periodic boundary conditions. A modified, numerically more efficient version, the particle mesh Ewald (PME) method[188], has been implemented in a number of widely used simulation packages. Both the Ewald summation method and the PME method are described in detail below.

3.2.6.1 Standard Ewald summation

Consider a neutral system in a lattice made up of *N* particles with charges $q_1, q_2, \ldots, q_N$ at positions $r_1, r_2, \ldots, r_N$. The real space lattice vectors are $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$, which need not be orthogonal. All the particles are assumed to be spherically homogeneous (point-like in biomolecular systems). The problem is to calculate the electrostatic potential experienced by particle *i* in the presence of all the other particles in the system. Under PBC, The electrostatic energy of the system is defined by

$$E = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} q_i q_j \sum_{|\mathbf{n}|=0}^{\infty}{}' \left|\mathbf{r}_{ij} + \mathbf{n}\right|^{-1} = \frac{1}{2}\sum_{i=1}^{N} q_i \varphi\left(\mathbf{r}_i\right), \tag{3.28}$$

where the sum over $\mathbf{n}$ is the a sum over lattice vectors $\mathbf{n} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$, and $\psi\left(\mathbf{r}_i\right)$ is the electrostatic potential. The prime in Eq. (3.28) indicates that the term for which $\left|\mathbf{r}_{ij} + \mathbf{n}\right| = 0$ is neglected. Hence, the potential $\psi\left(\mathbf{r}_i\right)$ does not include the infinite "self-energy" of the point charges, and is not a solution of Poisson's equation.

The Ewald summation introduces a localized "screening" charge density in spherical Gaussian distribution[j], which is centered at each point charge of equal magnitude and opposite sign. The screening charge density $\rho_S\left(\mathbf{r}\right)$ is a sum of the local densities.

$$\rho_S\left(\mathbf{r}\right) = -\sum_{j=1}^{N} q_j \left(\frac{\beta}{\sqrt{\pi}}\right)^3 \exp\left(-\beta^2 \left|\mathbf{r} - \mathbf{r}_j\right|^2\right), \tag{3.29}$$

where β is a positive parameter that determines the width of the Gaussians. The potential $\psi\left(\mathbf{r}_i\right)$ in Eq. (3.28) can be written

$$\psi\left(\mathbf{r}_i\right) = \left(\psi\left(\mathbf{r}_i\right) + \psi_S\left(\mathbf{r}_i\right)\right) - \psi_S\left(\mathbf{r}_i\right) = \psi_{\text{real}}\left(\mathbf{r}_i\right) + \psi_{\text{recip}}\left(\mathbf{r}_i\right), \tag{3.30}$$

where $\psi\left(\mathbf{r}_i\right)$ is the potential contributed by a Gaussian distribution of charges centered at each particle site, with the same total charge as those of the real particles; $\psi_S\left(\mathbf{r}_i\right)$ is the potential of the screening charges with an additional Gaussian distribution of opposite sign at each point charge. Since the charge distribution on a reference charge point is not considered to contribute to the

---

[j] In some methods non-Gaussian distributions are used but Gaussian distribution is typically chosen for Ewald sum.

potential $\psi(\mathbf{r}_i)$, by taking the sum of $\psi(\mathbf{r}_i)$and $\psi_S(\mathbf{r}_i)$ the Gaussian distributions counteract, so that the rapid convergence of the sum can be achieved by choosing a suitable width of the Gaussian distributions without affecting the total energy. This yields the real-space sum, $\psi_{\text{recip}}(\mathbf{r}_i)$, which is a sum of short-range potentials. The reciprocal sum, $\psi_{\text{recip}}(\mathbf{r}_i)$, is the long-range potential associated with the periodic charge density $-\rho_S(\mathbf{r})$ which converges smoothly in the real space and it had long been a challenge for the MD community.

The reciprocal potential $\psi_{\text{recip}}(\mathbf{r}_i)$ in Eq. (3.30) can be calculated by solvating Poisson equation in Fourier space. The Poisson equation reads

$$\nabla^2\psi_{\text{recip}}(\mathbf{r}_i) = -4\pi\rho_S(\mathbf{r}), \qquad (3.31)$$

Expanding $\psi_{\text{recip}}(\mathbf{r}_i)$ and $\rho_S(\mathbf{r})$ in Fourier series

$$\psi_{\text{recip}}(\mathbf{r}_i) = \sum_{\mathbf{m}} \hat{\varphi}_{\text{recip}}(\mathbf{m})\exp(i\mathbf{m}\mathbf{r}_i), \qquad (3.30)$$

$$\rho_S(\mathbf{r}) = \sum_{\mathbf{m}} \hat{\rho}_S(\mathbf{m})\exp(i\mathbf{m}\mathbf{r}_j), \qquad (3.31)$$

where $\mathbf{m}$ is the reciprocal lattice vector $\mathbf{m} = m_1\mathbf{a}_1^* + m_2\mathbf{a}_2^* + m_3\mathbf{a}_3^*$ with m_1, m_2, m_3 integers not all zero. Substituting Eq. (3.30), (3.31) into Eq. (3.31), we obtain

$$\hat{\psi}_{\text{recip}}(\mathbf{m}) = \frac{4\pi}{\mathbf{m}^2}\hat{\rho}_S(\mathbf{m}). \qquad (3.32)$$

With the choice of ρ_S in Eq. (3.29), the reciprocal potential $\psi_{\text{recip}}(\mathbf{r})$ can be written[189]

$$\begin{aligned}\psi_{\text{recip}} &= \sum_{j=1}^{N} q_j \frac{1}{\pi V}\sum_{|\mathbf{m}|\neq 0}^{\infty}\frac{1}{m^2}\exp\left(-\frac{\pi^2\mathbf{m}^2}{\beta^2}\right)\exp(2\pi i\mathbf{m}\cdot\mathbf{r}_{ij}) \\ &= \frac{1}{\pi V}\sum_{|\mathbf{m}|\neq 0}^{\infty}\frac{1}{m^2}\exp\left(-\frac{\pi^2\mathbf{m}^2}{\beta^2}\right)\exp(2\pi i\mathbf{m}\cdot\mathbf{r}),\end{aligned} \qquad (3.34)$$

where V is the volume of the lattice.

The potential $\psi_{\text{real}}(\mathbf{r}_i)$ is produced by the fraction of point charges that are not screened and it is evaluated at the reference point

$$\psi_{\text{real}}(\mathbf{r}_i) = {\sum_{j=1}^{N}}' q_j \sum_{|\mathbf{n}|=0}^{\infty}\frac{\operatorname{erfc}(\beta|\mathbf{r}_{ij}+\mathbf{n}|)}{|\mathbf{r}_{ij}+\mathbf{n}|} - \frac{\beta}{\sqrt{\pi}}q_j, \qquad (3.35)$$

where erfc(x) is the complementary error function.

$$\operatorname{erfc}(x)=\frac{2}{\sqrt{\pi}}\int_x^{\infty}\exp\left(-t^2\right)dt\,. \tag{3.36}$$

Finally, the analytic expression for the electrostatic energy is

$$\begin{aligned}
E &= E_{\text{real}}+E_{\text{recip}}+J\left(\mathbf{D},P,\varepsilon'\right)\\
&=\frac{1}{2}\sum_{i=1}^{N}q_i\left[\psi_{\text{real}}\left(\mathbf{r}_i\right)+\psi_{\text{recip}}\left(\mathbf{r}_i\right)\right]+J\left(\mathbf{D},P,\varepsilon'\right)\\
&=\frac{1}{2}\sum_{i=1}^{N}q_i\left[\begin{array}{l}\left(\displaystyle\sum_{j=1}^{N}{}'q_j\sum_{|\mathbf{n}|=0}^{\infty}\frac{\operatorname{erfc}\left(\beta\left|\mathbf{r}_{ij}+\mathbf{n}\right|\right)}{\left|\mathbf{r}_{ij}+\mathbf{n}\right|}-\frac{\beta}{\sqrt{\pi}}q_j\right)\\
+\left(\displaystyle\sum_{j=1}^{N}q_j\frac{1}{\pi V}\sum_{|\mathbf{m}|\neq 0}^{\infty}\frac{1}{m^2}\exp\left(-\frac{\pi^2\mathbf{m}^2}{\beta^2}\right)\exp\left(2\pi i\mathbf{m}\cdot\mathbf{r}_{ij}\right)\right)\end{array}\right]+J\left(\mathbf{D},P,\varepsilon'\right).
\end{aligned} \tag{3.37}$$

The last term $J\left(\mathbf{D},P,\varepsilon'\right)$ in Eq. (3.37) corrects for the non-uniform field associated with a macroscopic (but finite) crystal in a dielectric continuum[190]. It depends on the dipole moment **D** of the unit cell, the shape P of the macroscopic boundary conditions, and the external dielectric constant ε'.

E in Eq. (3.35) can be recast in the Fourier space to be more efficient convergent, which is known as the Fast Fourier Poisson method[190]. This method replaces the interaction of each point charge with $\psi_{\text{recip}}\left(\mathbf{r}\right)$ by the interaction of the introduced charge density having the same net charge at the same location. This is accomplished by splitting $\psi_{\text{recip}}\left(\mathbf{r}\right)$ of Eq. (3.37) into two integrals,

$$\begin{aligned}
E_{\text{recip}} &= \frac{1}{2}\sum_{i=1}^{N}q_i\psi_{\text{recip}}\left(\mathbf{r}_i\right)=\frac{1}{2}\int\rho\left(\mathbf{r}'\right)\psi_{\text{recip}}\left(\mathbf{r}'\right)d^3\mathbf{r}'\\
&=\frac{1}{2}\int\left[\rho\left(\mathbf{r}'\right)+\rho_S\left(\mathbf{r}'\right)\right]\psi_{\text{recip}}\left(\mathbf{r}'\right)d^3\mathbf{r}'-\frac{1}{2}\int\rho_S\left(\mathbf{r}'\right)\psi_{\text{recip}}\left(\mathbf{r}'\right)d^3\mathbf{r}',
\end{aligned} \tag{3.38}$$

where $\rho\left(\mathbf{r}\right)$ is the point charge density

$$\rho\left(\mathbf{r}\right)=\sum_{i=1}^{N}q_i\delta\left(\mathbf{r}-\mathbf{r}_i\right). \tag{3.39}$$

The first integral in Eq. (3.38) is canceled out by the real-space sum leaving only the second integral of order N, which can be numerically estimated in a much more efficient way.

3.2.6.2 The PME method

Inspired by the *particle-mesh* method of Hackney and Eastwood[191], Darden et al.[188] investigated the use of the fast Fourier transform (FFT) algorithm to speed up the calculation of the reciprocal sum in the Ewald summation. In computing the direct sum, the Ewald parameter α is chosen to reduce the real-space sum from $O(N^2)$ to $O(N)$ problem by employing a cutoff approach to numerically estimate it, The reciprocal sum is approximated by using 3D-FFT with convolutions on a grid where particle charges are interpolated on.

The reciprocal term in Eq. (3.37) can be simplified to be more efficiently convergent. Expanding $\exp\left(2\pi i\mathbf{m}\mathbf{r}_{ij}\right)$ and simplifying, one obtains $\cos 2\pi i\mathbf{m}\mathbf{r}_{ij}$. Thus,

$$E_{\text{recip}} = \frac{1}{\pi V}\sum_{\mathbf{m}\neq 0}\frac{1}{\mathbf{m}^2}\exp\left(-\frac{\pi^2\mathbf{m}^2}{\beta^2}\right)\sum_{i,j}^{N} q_i q_j \cos\left(2\pi i\mathbf{m}\mathbf{r}_{ij}\right). \tag{3.40}$$

Then by using trigonometry identities, E_{recip} is converted as follows

$$\begin{aligned}
E_{\text{recip}} &= \frac{1}{\pi V}\sum_{\mathbf{m}\neq 0}\frac{1}{\mathbf{m}^2}\exp\left(-\frac{\pi^2\mathbf{m}^2}{\beta^2}\right)\sum_{i,j}^{N} q_i q_j \cos\left(2\pi i\mathbf{m}\mathbf{r}_{ij}\right) \\
&= \frac{1}{\pi V}\sum_{\mathbf{m}\neq 0}\frac{1}{\mathbf{m}^2}\exp\left(-\frac{\pi^2\mathbf{m}^2}{\beta^2}\right)\sum_{i,j}^{N} q_i q_j \begin{bmatrix}\cos\left(2\pi i\mathbf{m}\mathbf{r}_i\right)\cos\left(2\pi i\mathbf{m}\mathbf{r}_j\right) \\ +\sin\left(2\pi i\mathbf{m}\mathbf{r}_i\right)\sin\left(2\pi i\mathbf{m}\mathbf{r}_j\right)\end{bmatrix} \\
&= \frac{1}{\pi V}\sum_{\mathbf{m}\neq 0}\frac{1}{\mathbf{m}^2}\exp\left(-\frac{\pi^2\mathbf{m}^2}{\beta^2}\right)\left\{\left[\sum_{i=1}^{N} q_i \cos\left(2\pi i\mathbf{m}\mathbf{r}_i\right)\right]^2 + \left[\sum_{i=1}^{N} q_j \sin\left(2\pi i\mathbf{m}\mathbf{r}_i\right)\right]^2\right\}.
\end{aligned} \tag{3.41}$$

Here, introducing the structure factor[192] known in protein crystallography

$$S\left(\mathbf{m}\right) = \sum_{k=1}^{N} q_k \exp\left(2\pi i\mathbf{m}\mathbf{r}_k\right), \tag{3.42}$$

so that

$$\begin{aligned}
E_{\text{recip}} &= \frac{1}{\pi V}\sum_{\mathbf{m}\neq 0}\frac{1}{\mathbf{m}^2}\exp\left(-\frac{\pi^2\mathbf{m}^2}{\beta^2}\right)S\left(\mathbf{m}\right)S\left(-\mathbf{m}\right) \\
&= \frac{1}{\pi V}\sum_{\mathbf{m}\neq 0}\frac{1}{\mathbf{m}^2}\exp\left(-\frac{\pi^2\mathbf{m}^2}{\beta^2}\right)\left|S\left(\mathbf{m}\right)^2\right|.
\end{aligned} \tag{3.43}$$

Therefore, the sum over i and j of order N^2 in E_{recip} are transformed into two sums of order N in Eq. (3.41) and finally into a single sum of order N in Eq. (3.43).

In the PME approach , the structure factor $S(\mathbf{m})$ in Eq. (3.43) is approximated by

$$\begin{aligned}
S\left(\mathbf{m}\right) \simeq \tilde{S}\left(\mathbf{m}\right) &= \sum_{k_1,k_2,k_3}\mathbf{Q}\left(k_1,k_2,k_3\right)\exp\left[2\pi i\left(\frac{m_1k_1}{K_1}+\frac{m_2k_2}{K_2}+\frac{m_3k_3}{K_3}\right)\right] \\
&= \boldsymbol{F}\left(\mathbf{Q}\right)\left(m_1,m_2,m_3\right),
\end{aligned} \tag{3.44}$$

where $\mathbf{Q}$ is the charge matrix containing charge values obtained by interpolating the point charges to a uniform grid of dimensions $K_1 \times K_2 \times K_3$ that fills the simulation box. $\boldsymbol{F}(\mathbf{Q})$ is the 3D-FFT of $\mathbf{Q}$. Combining Eq. (3.43) with Eq. (3.44), the reciprocal energy can be also approximated by

$$\tilde{E}_{\text{recip}} = \frac{1}{\pi V} \sum_{\mathbf{m} \neq 0} \frac{1}{\mathbf{m}^2} \exp\left(-\frac{\pi^2 \mathbf{m}^2}{\beta^2}\right) \boldsymbol{F}(\mathbf{Q})(\mathbf{m}) \boldsymbol{F}(\mathbf{Q})(-\mathbf{m}). \tag{3.45}$$

Therefore, this sum is computed with $O(N\ log(N))$ operations. Eq. (3.45) can be rewritten as a convolution

$$\tilde{E}_{\text{recip}} = \frac{1}{2} \sum_{m_1=0}^{K_1-1} \sum_{m_2=0}^{K_2-1} \sum_{m_3=0}^{K_3-1} \mathbf{Q}(m_1, m_2, m_3)(\psi_{\text{recip}} * \mathbf{Q})(m_1, m_2, m_3), \tag{3.46}$$

where ψ_{recip} is the reciprocal pair potential and "*" indicates a convolution. The **Q** matrix is first computed over a 3D uniform grid, and then it is transformed using inverse 3D-FFT to obtain the structure factors. The reciprocal energy is then calculated using Eq. (3.46) with the aid of FFT to compute the convolution $\psi_{\text{recip}} * \mathbf{Q}$.

3.2.7 MD in NPT Ensemble

The *NPT* ensemble is an isothermal-isobaric ensemble where the number of particles (*N*), the pressure (*P*) and the temperature (*T*) are conserved. It corresponds most closely to laboratory conditions with a system open to ambient temperature and pressure. Therefore, MD simulations are commonly performed in NPT ensemble, which is realized by applying a thermostat and a barostat algorithm to the equations of motion. Below I describe the most widely employed thermostat and barostat approaches including those used in this thesis.

3.2.7.1 Nosé thermostat and Nosé-Hoover thermostat

Various methods have been developed to control the temperature, which allow for sampling the canonical (or "*NVT*") ensemble. In canonical ensemble, the number of particles, the volume and the temperature are fixed to the prescribed values. Among these methods, the Nosé thermostat[193] has received a great amount of attention because it is deterministic, reversible and not restricted to systems with a large number of particles. A modified version of this method, the Nosé-Hoover thermostat[194], has become one of the most widely used methods to produce a simulation that samples the canonical ensemble.

An extended system

Given a system of *N* particles with coordinates r^N in a fixed volume *V*, and potential energy $U(r^N)$, the basic idea in the Nosé thermostat[193] is to introduce an additional degree of freedom *s*, which represents an external system for the physical system and acts as an heat reservoir. The thermal interaction between the heat reservoir and the physical system is an exchange of kinetic energy between them, which is expressed via the scaling of the real velocities of the particles

$$v_i = s\dot{r}_i, \qquad (3.47)$$

and this results in the temperature fluctuating around a target value T_{eq}. A potential energy $gk_BT_{eq}\ln s$ is associated with s, where $g = f+1$ (f being the number of degrees of freedom in the system) and k_B is the Boltzmann's constant. This choice of potential energy is essential for producing the canonical ensemble, as it will be shown later.

The Lagrangian and Hamiltonian of the extended system of particles and s can be written as

$$L_e\left(r,s,\dot{r},\dot{s}\right) = \sum_i \frac{m_i}{2} s^2 \dot{r}_i^2 - U\left(r\right) + \frac{Q}{2}\dot{s}^2 - gk_BT_{eq}\ln s, \qquad (3.48)$$

$$\begin{aligned} H_e\left(p_i,r,p_s,s\right) &= \sum_i \frac{p_i^2}{2m_i s^2} + U\left(r\right) + \frac{p_s^2}{2Q} + gk_BT_{eq}\ln s \\ &= H_r\left(p_i',r'\right) + \frac{p_s^2}{2Q} + gk_BT_{eq}\ln s, \end{aligned} \qquad (3.49)$$

where H_r is the Hamiltonian of the real system

$$H_r\left(p_i',r'\right) = \sum_i \frac{p_i'^2}{2m_i} + U\left(r'\right), \qquad (3.50)$$

and $p_i' = p_i / s$, $r' = r$ are the canonical variables associated to the real particles. The momenta are given by

$$p_i = \frac{\partial L_e}{\partial \dot{r}_i} = m_i s^2 \dot{r}_i, \qquad (3.51)$$

$$p_s = \frac{\partial L_e}{\partial \dot{s}} = Q\dot{s}. \qquad (3.52)$$

In the extended system the total Hamiltonian H_e is conserved.

The Nosé method produces the canonical ensemble. In fact, if H_e is the only conserved quantity, the microcanonical ensemble for the extended system is described by the partition function

$$Z = \frac{C_0}{N!}\int dp_s \int ds \int dp \int dr\, \delta\left(H_e - E\right), \qquad (3.53)$$

where C_0 is a normalization constant, E is the (constant) extended-system energy, $\delta(x)$ is the Dirac delta function, $dp = dp_1 dp_2 ... dp_N$, $dr = dr_1 dr_2 ... dr_N$ Transforming p_i and r_i into the real variables, the volume element becomes $dpdr = s^f dp'dr'$. Therefore,

$$Z = \frac{C_0}{N!}\int dp_s \int ds \int dp' \int dr' s^f \delta\left[H_r\left(p_i',r'\right) + \frac{p_s^2}{2Q} + gk_BT_{eq}\ln s - E\right]. \qquad (3.54)$$

Using the equivalence relation $\delta\left[g\left(s\right)\right] = \delta\left(s - s_0\right) / g'\left(s\right)$ where s_0 is the root of $g(s)$=0,

$$H_r(p_i',r') + \frac{p_s^2}{2Q} + gk_BT_{eq}\ln s_0 - E = 0, \tag{3.55}$$

$$s_0 = \exp\left[-\frac{H_r\left(p_i',r'\right) + p_s^2/2Q - E}{gk_BT_{eq}}\right], \tag{3.56}$$

thus,

$$\begin{aligned}
Z &= \frac{C_0}{N!gk_BT_{eq}}\int dp_s\int dp'\int dr'\int ds\ s^{f+1}\delta\left\{s - \exp\left[-\frac{H_r\left(p_i',r'\right) + p_s^2/2Q - E}{gk_BT_{eq}}\right]\right\} \\
&= \frac{C_0}{N!gk_BT_{eq}}\int dp_s\int dp'\int dr' \exp\left[-\frac{f+1}{g}\frac{H_r\left(p_i',r'\right) + p_s^2/2Q - E}{k_BT_{eq}}\right] \\
&= \frac{C_0}{N!gk_BT_{eq}}\exp\left(\frac{f+1}{g}\frac{E}{k_BT_{eq}}\right)\int dp_s \exp\left(-\frac{f+1}{g}\frac{p_s^2}{2Qk_BT_{eq}}\right)\int dp'\int dr' \exp\left[-\frac{f+1}{g}H_r\left(p_i',r'\right)\right] \\
&= \frac{C_0}{N!}\exp\left(\frac{f+1}{g}\frac{E}{k_BT_{eq}}\right)\left[\frac{2\pi Q}{\left(f+1\right)gk_BT_{eq}}\right]^{\frac{1}{2}}\int dp'\int dr' \exp\left[-\frac{f+1}{g}\frac{H_r\left(p_i',r'\right)}{k_BT_{eq}}\right],
\end{aligned} \tag{3.57}$$

When g=f+1,

$$Z = \frac{C_0}{N!}\exp\left(\frac{E}{k_BT_{eq}}\right)\frac{1}{g}\left[\frac{2\pi Q}{k_BT_{eq}}\right]^{\frac{1}{2}}\int dp'\int dr' \exp\left[-\frac{H_r\left(p_i',r'\right)}{k_BT_{eq}}\right] = CZ_c, \tag{3.58}$$

where Z_c is the partition function for the canonical ensemble

$$Z_c = \frac{1}{N!}\int dp'\int dr' \exp\left[-H_r\left(p_i',r'\right)/k_BT_{eq}\right]. \tag{3.59}$$

This demonstrates that by choosing the potential energy for s as $\left(f+1\right)k_BT_{eq}\ln s$, the partition function of the extended system is equivalent to that of the canonical ensemble except for a constant factor C. Under the ergodic hypothesis, the time average of time-independent quantities, which are functions of p_i/s and r_i in the extended system, are exactly the same as that in the canonical ensemble

$$\lim_{t_0\to\infty}\frac{1}{t_0}\int_0^{t_0} A(p/s,r)dt = \left\langle A(p/s,r)\right\rangle = \left\langle A(p',r')\right\rangle_c \tag{3.60}$$

where $\left\langle \ldots \right\rangle_c$ denote the average in the canonical ensemble. During simulations, canonical trajectories in the real molecular systems are generated by sampling the microcanonical ensemble for the extended system.

It is worth pointing out that the above result is derived from the assumption that the extended system has no conserved qualities other than H_e. In fact, the total momentum $\sum_i p_i$ and angular momentum $\sum_i r_i \times p_i$ are also conserved, which gives rise to slight deviations from the canonical ensemble.

The equations of motions for the extended system are

$$\frac{dr_i}{dt} = \frac{\partial H_e}{\partial p_i} = \frac{p_i}{m_i s^2}, \tag{3.61}$$

$$\frac{dp_i}{dt} = -\frac{\partial H_e}{\partial r_i} = -\frac{\partial \phi}{\partial r_i}, \tag{3.62}$$

$$\frac{ds}{dt} = \frac{\partial H_e}{\partial p_s} = \frac{p_s}{Q}, \tag{3.63}$$

$$\frac{dp_s}{dt} = -\frac{\partial H_e}{\partial s} = \frac{1}{s}\left[\sum_i \frac{p_i^2}{m_i s^2} - gkT_{eq}\right]. \tag{3.64}$$

Combining (3.64) and (3.63), we get a second order differential equation for s:

$$\frac{d^2 s}{dt^2} = \frac{2}{Qs}\left[\sum_i \frac{p_i^2}{2m_i s^2} - \frac{gk_B T_{eq}}{2}\right]. \tag{3.65}$$

Eq. (3.65) indicates that the change of s with time makes the kinetic energy oscillate about the value $gk_B T_{eq}/2$. When the kinetic energy is larger than $gk_B T_{eq}/2$, the quantity in the bracket is positive, s experiences a positive acceleration and tends to be larger. This tends to make the kinetic energy smaller. On the contrary, when the kinetic energy is smaller than $gk_B T_{eq}/2$, s tends to get smaller and make the kinetic energy become larger. Under equilibrium state, the time average of $d^2 s/dt^2$ vanishes, hence the time average of the kinetic energy coincides with the externally set temperature T_{eq}. The parameter Q controls the time scale of the temperature fluctuation. A small Q gives rapid response to the temperature change while a large Q leads to slow temperature fluctuations.

Real time sampling and Nosé-Hoover thermostat

In the previous section, Eq. (3.57) is achieved by sampling data points at integer multiples of virtual time unit Δt. Nosé called this sampling as “virtual time sampling” in which the real time interval is unequal. The relation of real time and virtual time is $s = \frac{dt}{dt'}$, so s can be thought of as a time scaling factor. Real time sampling can be carried out by using equal intervals in real time with $t_1 = \int_0^{t_0} dt/s$ and it gives the following result

$$\begin{aligned}\lim_{t_1\to\infty}\frac{1}{t_1}\int_0^{t_1}A\left(p/s,r\right)dt' &= \lim_{t_1\to\infty}\frac{t_0}{t_1}\frac{1}{t_0}\int_0^{t_0}A\left(p/s,r\right)\frac{dt}{s}\\ &=\left[\lim_{t_0\to\infty}\frac{1}{t_0}\int_0^{t_0}A\left(p/s,r\right)\frac{dt}{s}\right]\Bigg/\left(\lim_{t_0\to\infty}\frac{1}{t_0}\int_0^{t_0}\frac{dt}{s}\right)\\ &=\left\langle\frac{A\left(p/s,r\right)}{s}\right\rangle\Bigg/\left\langle\frac{1}{s}\right\rangle.\end{aligned}\tag{3.66}$$

To produce canonical ensemble, this average should be identical to $\left\langle A\left(p',r'\right)\right\rangle_c$, so one expects

$$\lim_{t_1\to\infty}\frac{1}{t_1}\int_0^{t_1}A\left(p/s,r\right)\frac{dt}{s}=\frac{C'}{N!}\int dp'\int dr'\int ds'\delta\left[H_r\left(p',r',s'\right)-E\right]\frac{A\left(p',r'\right)}{s'},\tag{3.67}$$

which can be achieved by choosing $g = f$.

Nosé[193] transformed the equations of motion Eq.(3.61)~(3.64) into equations for the real variables $r_i'= r$, $p_i' = p_i / s$, $t' = \int\frac{dt}{s}$, $s' = s$ and $p_s' = p_s / s$

$$\frac{dr_i'}{dt'}=\frac{p_i'}{m_i},\tag{3.68}$$

$$\frac{dp_i'}{dt'}=-\frac{\partial U}{\partial r_i'}-s'p_s'p_i'/Q,\tag{3.69}$$

$$\frac{ds'}{dt'}=\frac{p_s's'^2}{Q},\tag{3.70}$$

$$\frac{dp_s'}{dt'}=\frac{1}{s'}\left[\sum_i\frac{p_i'^2}{m_i}-gk_BT_{eq}\right]-\frac{s'p_s^2}{Q}.\tag{3.71}$$

The equations become complex but can be solved in real time by setting $g = f$. Hoover[194] simplified the above equations by introducing a new variable $\zeta\equiv p_s'/Q$. The equations of motion are then rewritten in coordinate evolution equations as

$$\frac{dr_i'}{dt'}=\frac{p_i'}{m_i},\tag{3.72}$$

$$\frac{dp_i'}{dt'}=-\frac{\partial\phi}{\partial r_i'}-\zeta p_i',\tag{3.73}$$

$$\frac{d\zeta}{dt'}=\frac{2}{Q}\left(\sum_i\frac{p_i'}{2m_i}-\frac{gk_BT_{eq}}{2}\right).\tag{3.74}$$

It is evident that ς controls the kinetic energy oscillating around the value $gk_BT_{eq}/2$. These equations describe the Nosé-Hoover thermostat, which is more suitable for production runs and it has been widely used for constant-temperature simulations.

3.2.7.2 The Andersen barostat and the Parrinello-Rahman barostat

Among the many classes of pressure control methods, the Berendsen barostat, the Andersen barostat and the constraint methods have been the most used over past decades. Here I discuss the Andersen barostat[195] and an extended version of it, the Parrinello-Rahman barostat[196] which was used in the work of this thesis.

The Andersen barostat

The Andersen barostat[195] considers a system of N particles in a cubic PBC box of volume V. V is set to be a dynamical variable rather than a fixed quantity in order to describe the volume fluctuations at constant pressure. The coordinates $r_1, r_2 ... r_N$ of the particles were then replaced by scaled coordinates, $\rho_1, \rho_2 ... \rho_N$,

$$\rho_i = r_i \Big/ V^{\frac{1}{3}}, \quad i = 1,2,...,N \quad , \tag{3.75}$$

where each component of ρ_i is a dimensionless number between 0 and 1. The Lagrangian of this scaled system is

$$L_A\left(\rho_i, V, \dot{\rho}_i, \dot{V}\right) = \sum_i \frac{m_i}{2} V^{\frac{2}{3}} \dot{\rho}_i^2 - U\left(V^{\frac{1}{3}}\rho\right) + \frac{M}{2}\dot{V}^2 - \alpha V. \tag{3.76}$$

The last two terms represent the kinetic energy term and the potential energy term associated with V, respectively; M and α are constants. The Lagrangian can be interpreted as follows: suppose the particles are in a container whose volume can be compressed by a piston, the variable V represents the coordinate of the piston, thus, M is the mass of the piston and α is the external pressure acting on the piston. The Hamiltonian of the scaled system is

$$H_A\left(\rho_i, V, p_i, p_V\right) = \sum_i \frac{p_i^2}{2m_i V^{2/3}} + U\left(V^{\frac{1}{3}}\rho_i\right) + \frac{p_V^2}{2M} + \alpha V, \tag{3.77}$$

where p_i is the momentum conjugate to ρ_i , and p_V is the momentum conjugate to V:

$$p_i = \frac{\partial L_A}{\partial \dot{\rho}_i} = m_i V^{\frac{2}{3}} \dot{\rho}_i, \quad p_V = \frac{\partial L_A}{\partial \dot{V}} = M\dot{V} \tag{3.78}$$

The equations of motion are therefore:

$$\dot{\rho}_i = \frac{\partial H_A}{\partial p_i} = \frac{p_i}{m_i V^{2/3}}, \tag{3.79}$$

$$\dot{p}_i = -\frac{\partial H_A}{\partial \rho_i} = -V^{\frac{1}{3}}\frac{\partial U}{\partial \rho_i}, \tag{3.80}$$

$$\dot{V} = \frac{\partial H_A}{\partial p_V} = \frac{p_V}{M}, \tag{3.81}$$

$$\dot{p}_V = -\frac{\partial H_A}{\partial V} = \sum_i \frac{p_i^2}{3 m_i V^{5/3}} - \rho V^{-\frac{2}{3}} U'\left(V^{\frac{1}{3}}\rho\right) - \alpha, \tag{3.82}$$

and combining the last two equations we get:

$$\ddot{V} = \frac{1}{M}\left[\left(\frac{2}{3}\sum_i \frac{p_i^2}{2m_i} - \frac{1}{3}\sum_i \frac{\partial U_i}{\partial \rho_i}\rho_i\right)\Big/ V + \alpha\right] = \frac{\alpha - P}{M}. \tag{3.83}$$

P is the instantaneous pressure defined by

$$P = \frac{2}{3V}\left(E_k - W\right) = \frac{2}{3V}\left(\sum_i \frac{p_i^2}{2m_i} - \frac{1}{2}\sum_i \frac{\partial U_i}{\partial \rho_i}\rho_i\right), \tag{3.84}$$

where *W* is the internal virial.

It can be proved that under the ergodic hypothesis, the time average of any function $X\left(\rho, V, p, p_V\right)$ of the scaled system is equal to the averages of *X* over an isoenthalpic-isobaric (*NPH*) ensemble in which the pressure is α, except for negligible errors[195]. In fact, the partition function of *NE* ensemble for the scaled system is

$$Z_A = \frac{C_0}{N!}\int d\rho \int dV \int dp \int dp_V \delta\left(H_A - E\right), \tag{3.85}$$

where $d\rho = d\rho_1 d\rho_2 ... d\rho_N$, $dp = dp_1 dp_2 ... dp_N$, and therefore the time average of $X\left(\rho, V, p, p_V\right)$ is

$$\begin{aligned} \left\langle X \right\rangle_A &= \frac{1}{N! Z_A}\int d\rho \int dV \int dp \int dp_V \delta\left(H_A - E\right) X\left(\rho, V, p, p_V\right) \\ &= \frac{\int d\rho \int dV \int dp \int dp_V \delta\left[\sum_i \frac{p_i^2}{2 m_i V^{2/3}} + U\left(V^{\frac{1}{3}}\rho_i\right) + \frac{p_V^2}{2M} + \alpha V - E\right] X\left(\rho, V, p, p_V\right)}{\int d\rho \int dV \int dp \int dp_V \delta\left[\sum_i \frac{p_i^2}{2 m_i V^{2/3}} + U\left(V^{\frac{1}{3}}\rho_i\right) + \frac{p_V^2}{2M} + \alpha V - E\right]}. \end{aligned} \tag{3.86}$$

Transforming the scaled variables into real ones $\rho_i = r_i / V^{1/3}$, $p_i = p_i' V^{1/3}$,

$$\left\langle X\right\rangle_A=\left\langle X\right\rangle=\frac{\int dr\int dV\int dp'\int dp_V\delta\left[\sum_i\frac{p_i'^2}{2m_i}+U\left(r_i\right)+\frac{p_V^2}{2M}+\alpha V-E\right]X\left(r,V,p'\right)}{\int dr\int dV\int dp'\int dp_V\delta\left[\sum_i\frac{p_i'^2}{2m_i}+U\left(r_i\right)+\frac{p_V^2}{2M}+\alpha V-E\right]}$$

$$=\frac{\int dr\int dV\int dp'\int dp_V\delta\left[H_r\left(r,p',V\right)+\alpha V-\left(E-\frac{p_V^2}{2M}\right)\right]X\left(r,V,p'\right)}{\int dr\int dV\int dp'\int dp_V\delta\left[H_r\left(r,p',V\right)+\alpha V-\left(E-\frac{p_V^2}{2M}\right)\right]}.\tag{3.87}$$

Consider the isoenthalpic-isobaric ensemble partition function

$$Z_{NPH}=\frac{C_0}{N!}\int dr\int dV\int dp'\delta\left[H_r\left(r,p',V\right)+PV-H\right],\tag{3.88}$$

then Eq. (3.87) can be written as

$$\left\langle X\right\rangle=\frac{\int dp_V\int dr\int dV\int dp'\delta\left[H_r\left(r,p',V\right)+\alpha V-\left(E-\frac{p_V^2}{2M}\right)\right]X_{NPH}\left(N,\alpha,E-\frac{p_V^2}{2M}\right)}{\int dp_V\int dr\int dV\int dp'\delta\left[H_r\left(r,p',V\right)+\alpha V-\left(E-\frac{p_V^2}{2M}\right)\right]}\tag{3.89}$$

which is closely related to the isoenthalpic-isobaric ensemble average of X except for the p_v integration. To eliminate this difference, Andersen expanded X_{NPH}

$$X_{NPH}\left(N,\alpha,E-\frac{p_V^2}{2M}\right)=X_{NPH}\left(N,\alpha,E\right)-\frac{p_V^2}{2M}\frac{\partial X_{NPH}\left(N,\alpha,E\right)}{\partial H}+O\left(N^{-2}\right),\tag{3.90}$$

and substituted it into Eq. (3.87) to get

$$\begin{aligned}\left\langle X\right\rangle&=X_{NPH}\left(N,\alpha,E\right)\\&\quad-\frac{\partial X_{NPH}\left(N,\alpha,E\right)}{\partial H}\frac{\int dp_V\int dr\int dV\int dp'\delta\left[H_r\left(r,p',V\right)+\alpha V-\left(E-\frac{p_V^2}{2M}\right)\right]\frac{p_V^2}{2M}}{\int dp_V\int dr\int dV\int dp'\delta\left[H_r\left(r,p',V\right)+\alpha V-\left(E-\frac{p_V^2}{2M}\right)\right]}+O\left(N^{-2}\right)\\&=X_{NPH}\left(N,\alpha,E\right)-\left\langle\frac{p_V^2}{2M}\right\rangle\frac{\partial X_{\mathrm{NPH}}\left(N,\alpha,E\right)}{\partial H}+O\left(N^{-2}\right)\\&=X_{NPH}\left(N,\alpha,E-\left\langle\frac{p_V^2}{2M}\right\rangle\right)+O\left(N^{-2}\right),\end{aligned}\tag{3.91}$$

where $O(N^{-2})$ is the error term proportional to $\partial^2X_{NPH}/\partial H^2$ for an intensive property X (for an extensive X, it should be of order N^{-1}).

The above result indicates that the Andersen barostat trajectory gives time averages equal to those of the *NPH* ensemble with $P = \alpha$ and $H = E - \left\langle p_V^2 / 2M \right\rangle$. As Andersen pointed out, the value of $\left\langle p_V^2 / 2M \right\rangle$ is proportional to the temperature *T* fixed by the chosen *E* of the scaled system ($\left\langle p_V^2 / 2M \right\rangle = kT/2$) and it is therefore independent of *M*. From Eq. (1.10) we can see that the mass of the piston, *M*, determines a response time of the pressure fluctuations caused by the box size oscillations, thus, an appropriate value of *M* must be chosen to ensure the response time is short enough for the pressure to be approximately constant. Andersen recommends that *M* be chosen to make the response time approximately equal to that for a sound wave to cross the simulation box.

The Parrinello-Rahman barostat

The Andersen barostat does not allow the simulation box to change its shape. To overcome this problem, Parrinello and Rahman[196] extended the Andersen method to enable a variable simulation box shape. In the Parrinello-Rahman barostat, a time-dependent metric tensor is introduced which allows the volume and the shape of the simulation box to vary with time. Basically, the box vectors are set to follow an equation of motion, and the equations of motion of the particles are also changed as in the Andersen barostat.

The simulation box can have an arbitrary shape described by three vectors, **a**, **b** and **c**. The vectors can have different lengths and arbitrary mutual orientations. Define a 3 × 3 matrix, **h**, whose columns are these three vectors. The box volume is given by

$$V = \det \mathbf{h} = \mathbf{a} \cdot \left(\mathbf{b} \times \mathbf{c}\right), \tag{3.92}$$

The position $\mathbf{r}_i$ of a particle can be written in terms of **h** and a column vector $\mathbf{s}_i$, with components ξ_i, η_i and ζ_i as

$$\mathbf{r}_i = \mathbf{h}\mathbf{s}_i = \xi_i \mathbf{a} + \eta_i \mathbf{b} + \zeta_i \mathbf{c}, \tag{3.93}$$

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

$$\mathbf{r}_{ij}^2 = \mathbf{s}_{ij}^T \mathbf{G} \mathbf{s}_{ij}, \tag{3.94}$$

where the metric tensor **G** is $\mathbf{G} = \mathbf{h}^T \mathbf{h}$. Using the latter notation, the Lagrangian of the scaled system can be written as

$$L = \frac{1}{2} \sum m_i \dot{\mathbf{s}}_i^T \mathbf{G} \dot{\mathbf{s}}_i - \sum \sum U\left(\mathbf{r}_{ij}\right) + \frac{1}{2} M T\mathbf{r}\left(\dot{\mathbf{h}}^T \dot{\mathbf{h}}\right) - pV. \tag{3.95}$$

Derivation of the equations of motion for the Parrinello-Rahman barostat system is similar to the case of the Andersen barostat.

3.2.8 Solvation effects: implicit and explicit models

Most proteins exist within an aqueous environment[k]. When simulating such proteins, it is common to assume that the protein is fully solvated in pure or ion-containing water. An explicit representation of the surrounding solvent can provide an accurate treatment of solute-solvent interactions. Although the discrete nature of water plays an important role in protein thermodynamics and functions, it typically increases the system size by about one order of magnitude compared to the size of the solute alone. In addition, a long simulation timescale is needed to average instantaneous solute-solvent interactions in order to obtain meaningful results concerning the solute structure, dynamics or energetics. Implicit solvent models represent the solvent as a continuous medium instead of individual explicit solvent molecules. Although implicit models largely improve the computational efficiency, they are associated with many limitations such as the lacking in the treatment of ionization effects and of specific interactions between biomolecules and solvent. Therefore, caution must be taken in choosing the appropriate solvent models in biomolecular simulations.

3.2.8.1 Implicit solvent models

Implicit solvent models reduce explicit solute-solvent interactions to their mean field characteristics, which are expressed as a function of the solute configuration alone. In many cases of protein simulations, the properties of interest are not largely affected by the detailed atomistic behavior of solvent molecules. In such cases, averaging over unimportant atomistic details of solvent greatly decreases the number of degrees of freedom in the system and generally yields significant computational efficiency. They allow for better direct estimations of free energies of solvation than explicit solvent models[197]. Therefore, implicit solvent methods are often applied to estimate free energy of solute-solvent interactions processes such as folding or conformational transitions of biomolecules, protein-protein or protein-ligand association, or transport of drugs across biological membranes.

There are two basic types of implicit solvent methods: models based on solvent-accessible surface areas (SASA), and the more recent continuum dielectric models based on the Poisson Boltzmann equation[198]. Various modifications and combinations of the different methods exist[198] and many successful applications have been reported[199]. However, implicit solvent models alone are not adequate to mimic all aspects of the solvent environment. A number of studies have shown that when the properties of the water molecules are substantially different from those of the bulk

---

[k] Other proteins such as transmembrane ones are not discussed here.

solvent, implicit solvent models could produce large artifacts compared to explicit solvent models. These include water near a membrane bilayer[200], ion dynamics in gramicidin channels[201], as well as the role of solvent in proton binding free energies for proteins[202].

3.2.8.2 Explicit solvent models

In explicit solvent models, water is represented by all-atom force fields. Explicit solvent models contribute specific interactions in addition to the dielectric screening effects, which are often important for mediating protein structure or function[203]. For instance, i) water molecules often mediate protein-protein or protein-ligand interactions; ii) explicit solvents play an important role in simulations for the accurate consideration of electrostatic effects and for the valid decomposition of free energies[204]; iii) explicit solvents are often required to study the detailed role of solvent in mediating interactions during protein folding[205].

For cases when explicit solvents are necessary, there is a wide range of explicit water models available[170]. The most popular models being used in biomolecular simulations include TIP3P[206], TIP4P[206],TIP5P[207], SPC[208], and SPC/E[208]. Each of these models is optimized to one or more physical properties of water, such as the radial distribution function, diffusivity, density etc. But none of these models can simultaneously reproduce all the properties of water. Commonly, the parameters in the water models are adjusted so that the enthalpy of vaporization and the density of water are reproduced in simulations[209]. All of the above models have a dipole moment of about 2.3 D instead of the experimental gas-phase value of 1.85 D[210]. The temperature dependence of the density of water is not described well by any of these models perhaps except the TIP5P model[211]. These most popular models treat the molecules as completely rigid, thus thet are all consistent with the SHAKE and LINCS constraints discussed above.

3.3 Classical Monte Carlo simulations

Monte Carlo (MC) simulations represent a significant alternative to MD simulations for investigating structural and thermodynamic properties of a system[212]. The term MC generally refers to simulations that use a stochastic approach to generate a set of representative configurations under given thermodynamic conditions such as temperature and volume. Classical MC simulations perform importance sampling of systems at equilibrium. The configurations are drawn from a probability distribution, often the classical Boltzmann distribution, to obtain thermodynamic properties or minimum-energy structures.

In its general form, a MC simulation proceeds as follows: starting from an initial configuration of particles in a system, a MC move is attempted that changes the configuration. This move is accepted or rejected based on an acceptance criterion which guarantees that the configurations are sampled in the desired ensemble distribution, and with the correct weight. After the acceptance or rejection of a move, the value of a property of interest is calculated. The results of many such steps are collated. Once sufficient sampling has been achieved, an accurate average value of this property can be obtained. With the application of statistical mechanics, it is possible to calculate the equilibrium thermodynamic properties of the system. Therefore, the ergodic hypothesis is one important necessary condition for MC simulations.

One attractive aspect of classical MC simulations is that only the potential energy is normally used in stepping through configurations. Unlike molecular dynamics simulations, MC simulations are free from solving Newton's equations of motion and calculation of forces (although some biased MC approaches do utilize force data). This allows for much more efficient calculations, which can lead to great speedups of up to 10 orders or more in the sampling of equilibrium properties[170]. Specific MD moves can also be combined in a simulation, allowing the modeler great flexibility in the approach to a specific problem[212]. In addition, Monte Carlo methods are generally easily parallelizable, with some techniques ideal for large CPU clusters.

However, there are a number of issues that hamper the use of MC simulations with large biomolecules. One issue is that it is difficult to define efficient moves for macromolecules —moves that change the configuration in a sufficiently large magnitude but also avoid generating energetically infeasible configurations. Conventional MC methods are inefficient for exploring the configurational space of large biomolecules when compared to MD simulations[213]. However, some work has eased this issue for proteins[214; 215; 216]. Another main issue is MC simulations of proteins in explicit solvent. It is difficult to conduct large-scale moves. Any move that significantly alters the internal coordinates of the protein without also moving the solvent molecules will likely result in a large overlap of atoms and thus, the rejection of the move. MC simulations using implicit solvent do not suffer from these drawbacks. Therefore, MC simulations of proteins are commonly conducted with coarse-grained models[217]. In addition, MC methods give no dynamical information about the time evolution of structural events. Last but not least, there is a lack of general, good, freely available programs for MC simulations of proteins because the choice of MC moves and acceptance criteria vary for one specific problem to another. It is worth noticing that a MC module has been added to CHARMM[218]. Another MC simulation program package, PROFASI (PROtein Folding and Aggregation Simulator)[18] has been made freely accessible for studying protein folding and aggregation. PROFASI has been employed in the preparation of this thesis and details of PROFASI will be described in a latter section.

3.4 Classical Monte Carlo simulations

Monte Carlo (MC) simulations represent a significant alternative to MD simulations for investigating structural and thermodynamic properties of a system[212]. The term MC generally refers to simulations that use a stochastic approach to generate a set of representative configurations under given thermodynamic conditions such as temperature and volume. Classical MC simulations perform importance sampling of systems at equilibrium. The configurations are drawn from a probability distribution, often the classical Boltzmann distribution, to obtain thermodynamic properties or minimum-energy structures.

In its general form, a MC simulation proceeds as follows: starting from an initial configuration of particles in a system, a MC move is attempted that changes the configuration. This move is accepted or rejected based on an acceptance criterion that guarantees the configurations are sampled in the desired ensemble distribution with the correct weight. After the acceptance or rejection of a move, the value of a property of interest is calculated. The results of many such steps are collated. Once sufficient sampling has been achieved, an accurate average value of this property can be obtained.

One attractive aspect of classical MC simulations is that only the potential energy is normally used in stepping through configurations. Unlike molecular dynamics simulations, MC simulations do not solve Newton's equations of motion and do not need to calculate forces (although some biased MC approaches do utilize force data). This allows for much more efficient calculations, which can lead to great speedups of orders of magnitude in the sampling of equilibrium properties [170]. Specific MD moves can also be combined in a simulation, allowing the modeler great flexibility in the approach to a specific problem[212]. In addition, Monte Carlo methods are generally easily parallelizable, with some techniques ideal for large CPU clusters.

However, there are a number of issues that hamper the use of MC simulations with large biomolecules. One issue is that it is difficult to define efficient moves for macromolecules— moves that change the configuration in a sufficiently large magnitude but also avoid generating energetically infeasible configurations. Conventional MC methods are inefficient for exploring the configurational space of large biomolecules when compared to MD simulations[213]. However, some work has eased this issue for proteins[214; 215; 216]. Another main issue is performing MC simulations of proteins in explicit solvent. It is difficult in general to conduct large-scale moves. Any move that significantly alters the internal coordinates of the protein without also moving the solvent molecules will likely result in a large overlap of atoms and thus, the rejection of the move. This implies that in these condition the sampling will be inefficient. MC simulations using implicit solvent do not suffer from these drawbacks. In addition, MC methods give no dynamical information about the time evolution of structural events. Last but not least, there is a lack of general, good, freely available programs for MC simulations of proteins because the choice of MC moves and acceptance criteria vary for one specific problem to another. It is worth noticing that a MC module has been added to CHARMM[218]. Another MC simulation program package, PROFASI (PROtein Folding and Aggregation Simulator)[18] has been made freely accessible for studying protein folding and aggregation. PROFASI has been employed in the work of this thesis and details of PROFASI will be described here in a latter section.

3.4.1 The Metropolis method

The Metropolis algorithm introduced by Metropolis *et al.*[219] is a Markov chain technique which is by far the most common MC simulation scheme[217]. In this scheme, a problem is described in terms of a thermodynamic system at potential energy, E, and temperature, T. With a constant T, the initial configuration is perturbed and the change in energy dE is computed. If dE is negative, the new configuration is accepted. If dE is positive, it is accepted with a probability given by a Boltzmann factor. This algorithm can be summarized in the following steps:

Step 1. Pick the configuration C_n (the initial configuration $n=0$ can be any configuration of the system).

Step 2. Generate a trial configuration C_t (usually a configuration similar to C_n obtained by performing some change or move on it) and compute the probability ratio $R= P(C_t)/P(C_n)$.

Step 3. Pick a random number p with value between 0 and 1. Make $C_{n+1}= C_t$, if $p \leq R$. Otherwise, make $C_{n+1}= C_n$.

Note that the probability of accepting a trial configuration C_t is

$$P_{C_n,C_T} = \begin{cases} R = \dfrac{P\left(C_t\right)}{P\left(C_n\right)}, & \text{if} \quad P\left(C_t\right) < P\left(C_n\right), \\ 1, & \text{otherwise.} \end{cases} \tag{3.96}$$

Step 2 and step 3 are iterated to obtain a set of N configurations, $C_1, C_2, \ldots, C_N$, where N is a sufficiently large number. It can be proved (below) that such an algorithm indeed produces an ensemble of configurations that satisfies the Boltzmann distribution for the current temperature T

$$\lim_{N \to \infty} \frac{N_C}{N} = P\left(C\right), \tag{3.97}$$

where N_C is the number of time the configurations C of the system has been visited in the MC run, and $P(C)$ is the Boltzmann distribution, $P\left(C\right) = \left[e^{-E(C)/(k_B T)}\right] \Big/ Z\left(T\right)$.

To prove that the algorithm produces an ensemble of configurations that satisfies Eq. (3.97), it needs to be shown that any initial distribution N_C/N evolves towards the distribution $P(C)$, and once such a distribution is reached it remains at equilibrium. According to steps 2 and 3, for any pair of states C and C', the number of configurations generated in state C' by applying the algorithm to the $N(C)$ configurations in state C is $N(C)P_{C,C'}$, where $P_{C,C'}$ is the probability of accepting the trial configuration C' when $C_n = C$. In addition, the number of configurations generated in state C' by applying the algorithm to the $N(C')$ configurations in state C' is $(1- P_{C,C'})N(C')$. Therefore, the total number $N_{\text{tot}}(C')$ of configurations in state C' after the application of the algorithm to configuration C and C' is

$$N_{\text{tot}}(C') = N(C') + \Delta N(C'), \tag{3.98}$$

where $\Delta N(C') = N(C)P_{C,C'} - N(C')P_{C',C}$ is the net change in the number of configurations in state C', relative to $N(C')$. According to Eqs. (3.96) and (3.98),

$$\Delta N(C') = N(C) - N(C')\frac{P(C)}{P(C')}, \tag{3.99}$$

when $P(C') > P(C)$ and

$$\Delta N(C') = N(C)\frac{P(C')}{P(C)} - N(C'), \tag{3.100}$$

when $P(C') < P(C)$. Therefore, according to Eqs. (3.99) and (3.100), $\Delta N(C') > 0$ when $N(C')/N < P(C')$ and $\Delta N(C') < 0$ when $N(C')/N > P(C')$. The algorithm evolves any arbitrary distribution towards the equilibrium distribution where $N_C/N = P(C)$. In addition, Eqs. (3.99) and (3.100) indicate that $\Delta N(C') = 0$ when both $N(C)/N = P(C)$ and $N(C')/N = P(C')$ are satisfied, i.e. the algorithm does not alter the relative population of the states when the ensemble distribution is equal to the equilibrium distribution.

The most important aspect of the method is that the algorithm is able generate an ensemble of configurations with the probability distribution simply by computing the probability ratios $P(C')/P(C)$.

Under ergodic hypothesis, the Boltzmann average for any observable property can be obtained as a simple arithmetic average of the property values from individual accepted configurations:

$$\langle A \rangle \approx \bar{A} = \frac{1}{N}\sum_{C} N_C A(C), \tag{3.101}$$

where $A(C)$ is the value of the observable A for state C and $\bar{A}$ is the Monte Carlo estimator of $\langle A \rangle$ associated with the finite number of configurations N.

3.4.2 PROFASI: PROtein Folding and Aggregation Simulator

3.4.2.1 The force field in PROFASI

PROFASI uses an effective all-atom potential[220] to simulate protein folding and aggregation. The model contains all atoms of the protein chains, including hydrogen atoms, but no explicit water molecules. It assumes fixed bond lengths, bond angles, and peptide torsion angles (180°), so that each amino acid has the Ramachandran dihedral angles, ϕ and ψ, as well as a number of side-chain torsion angles as its degrees of freedom. The interaction potential is composed of four terms

$$E = E_{\text{loc}} + E_{\text{ev}} + E_{\text{hb}} + E_{\text{sc}}. \tag{3.102}$$

The first term E_{loc} is a local potential accounting for interactions between atoms separated by a few covalent bonds, such as the electrostatic interaction between adjacent peptide units along the chain. The other three terms are non-local: E_{ev} represents excluded-volume effects, E_{hb} is a hydrogen-bond potential, and E_{sc} represents residue-specific, pairwise, additive approximations for hydrophobic attraction between nonpolar side chains and electrostatic interactions among charged side-chains. Energy parameters in Eq. (3.102) are given in entropy unit, eu (1 eu = 1 kcal/(mol·K)). The factor for conversion from eu to kcal/mol are determined by calibration against the experimental melting temperature for a peptide[220]. The exact forms of each term in Eq. (3.102) are given below.

The local potential $E_{\text{loc}} = E_{\text{loc}}^{(1)} + E_{\text{loc}}^{(2)} + E_{\text{loc}}^{(3)}$ can be divided into two backbone terms, $E_{\text{loc}}^{(1)}$ and $E_{\text{loc}}^{(2)}$, and one side-chain term, $E_{\text{loc}}^{(3)}$. Defining a so-called "peptide unit" to be composed of the backbone C′O group of one residue and the backbone NH group of the next residue, then $E_{\text{loc}}^{(1)}$ represents interactions between partial charges of neighboring peptide units along the peptide chain:

$$E_{\text{loc}}^{(1)} = \kappa_{\text{loc}}^{(1)} \sum_{\text{n.n.}} \sum_{i} \sum_{j} \frac{q_i q_j}{r_{ij} / \text{Å}}, \tag{3.103}$$

where the outer sum runs over all pairs of nearest-neighbor peptide units, and each of the two inner sums runs over atoms in one peptide unit (except for proline for which the sum runs over only C′ and O). The partial charges are given as $q_i = \pm 0.20$ for H and N atoms and $q_i = \pm 0.42$ for C′ and O atoms, respectively[221]. The constant $\kappa_{\text{loc}}^{(1)}$ has the value of 6 eu, corresponding to a dielectric constant of $\varepsilon \approx 41$. Two peptide units that are not nearest neighbors along the chain interact through hydrogen bonding (see below) rather than through $E_{\text{loc}}^{(1)}$.

The term $E_{\text{loc}}^{(2)}$ reads

$$E_{\text{loc}}^{(2)} = \kappa_{\text{loc}}^{(2)} \sum_{\text{non-Gly}} \left[f\left(u_I\right) + f\left(v_I\right) \right] + \kappa_{\text{loc,G}}^{(2)} \sum_{\text{GLy}} \left(\cos\psi_I + 2\cos 2\psi_I \right). \tag{3.104}$$

where $\kappa_{\text{loc}}^{(2)} = 1.2$ eu, $\kappa_{\text{loc,G}}^{(2)} = -0.15$ eu, I is a residue index, and

$$\begin{aligned} u_I &= \min\left[d\left(\text{H}_I, \text{N}_{I+1}\right), d\left(\text{N}_I, \text{H}_{I+1}\right) \right] - d\left(\text{H}_I, \text{H}_{I+1}\right), \\ v_I &= \min\left[d\left(\text{O}_I, \text{C}'_{I+1}\right), d\left(\text{C}'_I, \text{O}_{I+1}\right) \right] - d\left(\text{O}_I, \text{O}_{I+1}\right), \\ f\left(x\right) &= \max\left(0, \tanh 3x\right). \end{aligned} \tag{3.105}$$

The function $f(u_I)$ is positive if the H_IH_{I+1} distance, $d(H_I,H_{I+1})$, is smaller than both the H_IN_{I+1} and N_IH_{I+1} distances, and zero otherwise. This term thus provides an energy penalty when H_I and H_{I+1} are exposed to each other (it is omitted if residue I or I+1 is a proline). Similarly, $f(v_I)$ is positive when O_I and O_{I+1} are exposed to each other. The first sum in Eq. (3.104) provides an additional OO and HH repulsion for neighboring peptide units, unless the residue flanked by the two peptide units is a glycine. This repulsion makes doubling of hydrogen bonds less likely. Glycine has markedly different backbone energetics compared to other residues. Therefore, the second sum in Eq. (3.104) provides an energy penalty for glycine backbone dihedral ψ values around ±120°, which are sterically allowed but relatively rare in known protein structures[222].

$E_{\mathrm{loc}}^{(3)}$ is an explicit torsion angle potential for side-chain angles, χ_i:

$$E_{\mathrm{loc}}^{(3)} = \sum_i \kappa_{\mathrm{loc},i}^{(3)} \cos n_i \chi_i, \tag{3.106}$$

where $\kappa_{\mathrm{loc},i}^{(3)}$ and n_i are constants with different values that define four classes of side-chain angles in amino acids (Table 2.1).

Table 2.1: Classification of side-chain angles, χ_i. The parameters of the torsion angle potential $E_{\mathrm{loc}}^{(3)}$ are ($\kappa_{\mathrm{loc},i}^{(3)}$, n_i)= (0.6 eu, 3) for class I, ($\kappa_{\mathrm{loc},i}^{(3)}$, n_i)= (0.3 eu, 3) for class II, ($\kappa_{\mathrm{loc},i}^{(3)}$, n_i)= (0.4 eu, 3) for class III, and ($\kappa_{\mathrm{loc},i}^{(3)}$, n_i)= (-0.4 eu, 3) for class IV.

Residue	χ_1	χ_2	χ_3	χ_4
Ser, Cys, Thr, Val	I			
Ile, Leu	I	I		
Asp, Asn	I	IV		
His, Phe, Tyr, Trp	I	III		
Met	I	I	II	
Glu, Gln	I	I	IV	
Lys	I	I	I	I
Arg	I	I	I	III

The second term in Eq. (3.102), E_{ev}, has the form

$$E_{\mathrm{ev}} = \kappa_{\mathrm{ev}} \sum_{i<j} \left[\frac{\lambda_{ij}\left(\sigma_i + \sigma_j\right)}{r_{ij}} \right]^{12}, \tag{3.107}$$

where $\kappa_{\mathrm{ev}} = 0.10$ eu, and the summation is over pairs of atoms (i and j). The values of the radii, σ_i, are given as 1.77, 1.75, 1.53, 1.42, and 1.00 Å for S, C, N, O, and H atoms, respectively. The σ_{i} values agree reasonably well with the statistical analysis of standard radii and volumes in proteins[223]. The parameter λ_{ij} is a “reduction factor” with the value 0.75 for all pairs except for those connected by three covalent bonds, for which $\lambda_{ij} = 1$. A cutoff of $4.3\lambda_{ij}$ Å is employed in the evaluation of the pairs to be considered in the sum of E_{ev} to speedup the calculations.

The third term in Eq. (3.102), E_{hb}, describes hydrogen bonds between NH and CO groups. Two neighboring peptide units, which interact through the local potential E_{loc}, are excluded in the hydrogen-bond term. E_{hb} has the following form

$$E_{hb} = \varepsilon_{\mathrm{hb}}^{(1)} \sum_{\text{bb-bb}} u\left(r_{ij}\right) v\left(\alpha_{ij}, \beta_{ij}\right) + \varepsilon_{\mathrm{hb}}^{(2)} \sum_{\text{sc-bb}} u\left(r_{ij}\right) v\left(\alpha_{ij}, \beta_{ij}\right), \tag{3.108}$$

where $\varepsilon_{\mathrm{hb}}^{(1)}$ = 3.0 eu and $\varepsilon_{\mathrm{hb}}^{(2)}$ = 2.3 eu set the strengths of backbone-backbone and side-chain-backbone bonds, respectively, r_{ij} is the HO distance, α_{ij} is the NHO angle, and β_{ij} is the HOC angle. The functions $u(r)$ and $v(\alpha, \beta)$ are given by

$$u\left(r\right) = 5\left(\frac{\sigma_{\mathrm{hb}}}{r}\right)^{12} - 6\left(\frac{\sigma_{\mathrm{hb}}}{r}\right)^{10},$$

$$v\left(\alpha, \beta\right) = \begin{cases} \left(\cos\alpha\cos\beta\right)^{1/2} & \text{if } \alpha, \beta > 90° \\ 0 & \text{otherwise} \end{cases}, \tag{3.109}$$

where σ_{hb} = 2.0 Å. A cutoff of 4.5 Å is used for $u(r)$, i.e. $u(r) = 0$ if $r > 4.5$ Å.

The last term in Eq. (3.102), E_{sc}, is formed by two terms, $E_{\mathrm{sc}} = E_{\mathrm{hp}} + E_{\mathrm{ch}}$. E_{ch} represents interactions among side-chain charges. E_{hp} accounts for the effects of all the other relevant interactions, especially effective hydrophobic attraction. E_{hp} and E_{ch} have a similar form

$$E_{\mathrm{hp}} = -\sum_{I<J} M_{IJ}^{(\mathrm{hp})} C_{IJ}^{(\mathrm{hp})}, \quad E_{\mathrm{ch}} = -\sum_{I<J} M_{IJ}^{(\mathrm{ch})} C_{IJ}^{(\mathrm{ch})}, \tag{3.110}$$

where the sums run over residue pairs IJ, $C_{IJ}^{(\mathrm{hp})}$ and $C_{IJ}^{(\mathrm{ch})}$ are contact measures with values between 0 and 1, and $M_{IJ}^{(\mathrm{hp})}$ and $M_{IJ}^{(\mathrm{ch})}$ are energy parameters. Details of the parameters $C_{IJ}^{(\mathrm{hp})}$, $C_{IJ}^{(\mathrm{ch})}$, $M_{IJ}^{(\mathrm{hp})}$ and $M_{IJ}^{(\mathrm{ch})}$ are given in ref. [220].

PBC is applied in PROFASI. Since all the interactions in PROFASI are short-ranged, it is sufficient to use a periodic box length slightly larger than the maximum end-to-end distance of the peptide/protein being simulated[18].

3.4.2.2 The MC algorithm in PROFASI

The simulation algorithms used in PROFASI are based on MC methods rather than MD. The conformational updates currently implemented in PROFASI include three different conformational updates: i) single-variable updates of side-chain and backbone angles, respectively, using simple Metropolis criteria [219] ii) backbone updates using biased Gaussian steps (BGS)[224], and iii) rigid-body translations and rotations of whole chains, which are useful updates for multi-chain systems.

The single-variable update, when applied to a backbone angle, tends to lead to a highly nonlocal deformation of the chain, thus, is likely to be rejected if the chain is compact. However, this update can be a very powerful method for simulating extended chains[225; 226]. The BGS are semi-local updates that are less drastic than the single-variable updates. The BGS simultaneously update n ($n \leq 8$) adjacent torsion angles ($\tau_\mathrm{i}, \ldots, \tau_n$) along the backbone in a manner that keeps the ends of the segment approximately fixed. Given a backbone conformation $\bar{\tau} = \left(\tau_1, \ldots, \tau_n\right)$, a tentative new conformation $\bar{\tau}'$ is generated with a bias toward local deformations. Specifically, using a conformation-dependent $n \times n$ matrix $\mathbf{G}$ such that $\left(\bar{\tau}' - \bar{\tau}\right)^T \mathbf{G} \left(\bar{\tau}' - \bar{\tau}\right) \approx 0$ for changes corresponding to local deformations, $\bar{\tau}'$ is drawn from the Gaussian distribution

$$W\left(\bar{\tau} \to \bar{\tau}'\right) = \frac{\left(\det \mathbf{A}\right)^{1/2}}{\pi^3} \exp\left[-\left(\bar{\tau}' - \bar{\tau}\right)^T \mathbf{A} \left(\bar{\tau}' - \bar{\tau}\right)\right],$$
$$\mathbf{A} = \frac{a}{2}\left(\mathbf{1} + b\mathbf{G}\right), \tag{3.111}$$

where $\mathbf{1}$ denotes the $n \times n$ unit matrix, and a and b are tunable parameters controlling the acceptance rate and the degree of bias toward local deformations, respectively. Typical values for these parameters are $a = 300(\mathrm{rad})^{-2}$ and $b = (\mathrm{rad}/\text{Å})^{-2}$. The probability for accepting the new conformation $\bar{\tau}'$ is

$$P\left(\bar{\tau} \to \bar{\tau}'\right) = \min\left(1, \frac{W\left(\bar{\tau}' \to \bar{\tau}\right)}{W\left(\bar{\tau} \to \bar{\tau}'\right)} \exp\left(-\frac{E' - E}{k_B T}\right)\right), \tag{3.112}$$

where E and E' are the energy of $\bar{\tau}$ and $\bar{\tau}'$, respectively.

These updates are employed together in PROFASI for simulations in the canonical ensemble. In addition, improved sampling techniques such as simulated tempering[227] and replica-exchange method[228] are also implemented in PROFASI. The latter method is discussed in the next section.

3.5 Replica-exchange method for improved sampling

3.5.1 The sampling problem

One of the major challenges in biomolecular simulations is the sampling[229], because of the large number of degrees of freedom and the rugged energy landscape of biomolecules. Conformational transitions of proteins systems often occur orders-of-magnitude slower than the timescales accessible to standard simulation techniques[230]. Thus, methods that simply rely on the native sampling ability of the thermal motion of biomolecular systems at room temperature are often insufficient because of the far too short simulation time. Even with sufficiently accurate force fields it is still very difficult to simulate the folding processes or conformational transitions for proteins[230; 231]. Simulations may be trapped in local minima, thus they cannot achieve ergodicity.

Simulations of the same system starting from different initial conditions may give different results[232].

To improve sampling, a variety of advanced techniques have been developed, which have been discussed in several reviews[229; 233; 234; 235]. Most of the techniques could be grouped into three categories[233]: i) those that modify the potential energy; ii) those that reduce the degrees of freedom; and iii) variants of the replica-exchange (also known as parallel tempering) method[228] (REM). Here, I will focus on the original REM, which was used to carry out the work presented in Chapter 5 of this thesis.

3.5.2 Replica-exchange method

The dynamic behavior of a system is intrinsically dictated by its free energy, and crossing energy barriers is often the first step leading to the conformational changes of proteins. Elevated temperature can significantly enhance the ability of a system to cross energy barriers. It is thus the most commonly used strategy to improve sampling in simulations. In this type of method, the entropic component of the free energy is exploited to enhance barrier crossing and sampling. Based on this idea, REM[228] employs a set of replicated simulations at a range of temperatures. At regular intervals, the non-interacting simulation replicas are allowed to switch temperature with each other, thereby making it possible for low temperature replicas to gradually migrate up to higher temperatures and back again. The temperature range is chosen such that the low temperature values correspond to native conditions and the high values enable the system with sufficient energy to rapidly pass energy barriers. Originally developed for MC simulations[228], REM has been adapted also for MD algorithms[236]. Below the basic technique of REM in MC simulations is described.

Consider two replicas of the Metropolis MC simulation of a system, performed at temperature T_1 and T_2, respectively. At a certain MC step, an update is made by exchanging the temperatures of the two replicas. The exchange is a stochastic process based on a Metropolis criterion. It occurs with a probability $P(T_1 \leftrightarrow T_2)$ which depends on the reference temperatures of two replicas (T_1, T_2) and their instantaneous potential energies (U_1, U_2)[237]:

$$P\left(T_1 \leftrightarrow T_2\right) = \min\left(1, e^{\Delta}\right),$$
$$\Delta = \left(\frac{1}{k_B T_1} - \frac{1}{k_B T_2}\right)\left(U_1 - U_2\right). \tag{3.113}$$

If the difference between the two temperatures, T_1 and T_2, is small enough, the energy histograms of the two simulation replicas should overlap. In other words, the system configurations sampled at temperature T_1 are likely to appear during the simulation at T_2. This gives the acceptance probability of exchanging temperatures between the two replicas. In practice, this update can be generalized to more than two systems. It is assumed that, in order to be efficient, each replica of the simulations should spend equal amounts of time at each temperature and thereby facilitate the flux of configurations from low to high temperatures and back[238]. Therefore, the temperatures have to be chosen in a way that gives the same acceptance probability between all adjacent pairs of replicas over the entire temperature range. The exact choice of temperatures is crucial for optimal performance of the REM. The exchange frequency is determined by both the probability and the number of exchange trials. A simple geometric distribution of temperatures does not yield equal acceptance probabilities[239]. To determine an optimum set of temperatures is not straightforward. Recently, a "temperature predictor" has been developed to find the set of temperatures in a given range that give a desired exchange probability[240].

By a careful choice of temperatures and number of replicas, REM allows tunneling between metastable states of the system and improving convergence to a global optimum. This approach is successful in exploring energy landscapes of peptides and small proteins[229]. However, simulations of large systems or with explicit solvent are computationally expensive, and pose a challenge for REM to obtain converged results.

Chapter 4 Common Structural Traits across Pathogenic Mutants of the Human Prion Protein and Their Implications for Familial Prion Diseases

Familial Hu TSE are associated with over 50 pathogenic mutations (PMs) in the gene coding for HuPrP (the *PRNP* gene). Most of the currently identified PMs are missense mutations (coding for a single-residue substitution) located in the folded globular domain of HuPrP^{C}. We performed 50 ns of molecular dynamics (MD) simulations in aqueous solution for HuPrP^{C} globular domain containing each of these PMs in order to investigate their influence on the protein structure. Overall, 1.6 μs of MD data for these mutants are presented here. The calculations are based on the AMBER99 force field, which has been shown to reproduce very accurately the structural features of the wild-type (WT) HuPrP^{C} and a few variants for which experimental structural information is available. The mutants exhibit structural traits different from those of WT HuPrP and of the protective polymorphism HuPrP(E219K-129M). These traits include the loss of salt bridges in the α2–α3 region and the loss of π-stacking interactions in the β2–α2 loop and between the loop and the C-terminal of α3 helix. In addition, in the majority of the mutants, the α3 helix is more flexible and residue Y169 in the β2–α2 loop is more solvent exposed than those in WT HuPrP and HuPrP(E219K-129M). The presence of similar traits in this large spectrum of PMs hints to a role of these fingerprints in their known disease-causing properties. Overall, the regions most affected by the PMs in terms of structure and/or flexibility are likely to be those involved in the conversion to PrP^{Sc} and in the interactions with cellular partners. These regions thus emerge as optimal targets for antibody- and ligand-binding studies.

4.1 Introduction

As introduced in Chapter 2, HuPrP^{C} is mainly in α-helix conformation, while HuPrP^{Sc} is rich in β-structures[241; 242; 243; 244; 245; 246; 247; 248]. Solution-NMR studies[249; 250] have established that HuPrP^{C} features a largely disordered N-terminal tail (residues 23-124), a folded C-terminal globular domain (residues 125-230, denoted GD hereinafter), and two signal peptides (residues 1-22 and 231-253) (see Fig. 2.1D in Chapter 2). The GD structure has been resolved at pH 7[250], which contains three α-helices (α1, α2 and α3) and two short anti-parallel β-strands (β1 and β2) (Fig. 2.1D). A disulfide bond between residues C179 and C214 links the middle of the α2 and α3 helices. The α2-α3 region is believed to be crucial for HuPrP^{C}→HuPrP^{Sc} conversion[147; 249; 250; 251; 252; 253; 254; 255; 256; 257; 258; 259; 260; 261; 262; 263; 264; 265; 266; 267; 268; 269; 270; 271; 272; 273]. Familial types of Hu TSE are associated with over 50 PMs in the *PRNP* gene with broad chemical properties and diverse worldwide incidence[154; 159; 274] (see Fig.

2.3 in Chapter 2). As many as 33 PMs are missense mutations located in the GD (Fig. 2.3) and they can lead to spontaneous formation of $\mathrm{HuPrP}^{\mathrm{Sc}}$ in the brain[275]. These PMs may accelerate the kinetics of the misfolding process *in vitro* compared to WT HuPrP[276] (HuPrP(WT), hereinafter). This may be caused, in part, by destabilization of the native HuPrP structure and/or by stabilization of the partially unfolded intermediate HuPrP species[153; 267; 276; 277; 278] that are likely to be $\mathrm{HuPrP}^{\mathrm{Sc}}$ precursors[267; 276].

NMR[147; 259; 260] and X-ray studies[279] have provided the underlying structural effects of several such PMs. MD simulations have given insights into the structural determinants and the plasticity of another 10 mutants containing PMs[261; 262; 263; 264; 265; 266; 267; 269; 270; 271; 272; 273; 280; 281; 282; 283; 284; 285] with high incidences[286], as well as those of the protective polymorphism HuPrP(E219K-129M) [263, 1] (HuPrP(PP), hereinafter). These studies have shown that the secondary and tertiary structural elements are similar across all these variants. In the MD simulations, Root Main Square Fluctuation (RMSF)[263; 265; 266; 269; 272; 281], Root Main Square Deviation (RMSD)[261; 287] and secondary-structure content[270; 283] calculations lead to the suggestion that the α2-α3 region of the investigated mutants are more flexible than that of HuPrP(WT). This hints to the possible presence of structural fingerprints across the PMs that are relevant for the conversion to $\mathrm{HuPrP}^{\mathrm{Sc}}$.

Here we investigate the effect of PMs in $\mathrm{HuPrP}^{\mathrm{C}}$ GD by using a MD protocol based on the AMBER99 force field[288; 289]. This protocol has been validated against structural data of a few HuPrP mutants in a previous study of our lab[263]. In particular, thisstudy[263] has found that the β_2-α_2 loop and the α_2-α_3 region in two PMs, HuPrP(Q212P-129M) and HuPrP(E200K-129M), differ significantly from those in HuPrP(WT) and HuPrP(PP), in full agreement with NMR structural information[147; 260]. In this work we extend this MD protocol to study 30 HuPrP mutants[m] containing PMs located in GD (Fig. 4.1). Comparison is made with results of HuPrP(WT) and HuPrP(PP) obtained previously by applying the same protocol[263] as well as with available NMR/X-ray structures of HuPrP variants[147; 260; 279; 290]. Overall, 1.6 μs MD simulations are performed here. The calculations point to the presence of several structural traits in all or most of the PMs. These traits hint to the molecular basis of the early stage of the $\mathrm{HuPrP}^{\mathrm{C}} \rightarrow \mathrm{PrP}^{\mathrm{Sc}}$ conversion.

---

[1] Hereinafter, each variant is denoted by the mutation (the first five characters inside the parentheses) followed by the M/V polymorphisms at position 129 (last four characters inside the parentheses) associated with the mutation.

[m] These include all the known PMs in the GD at the time of this study along with their associated polymorphism at position 129. PMs that are identified after we have carried out this study are not included here.

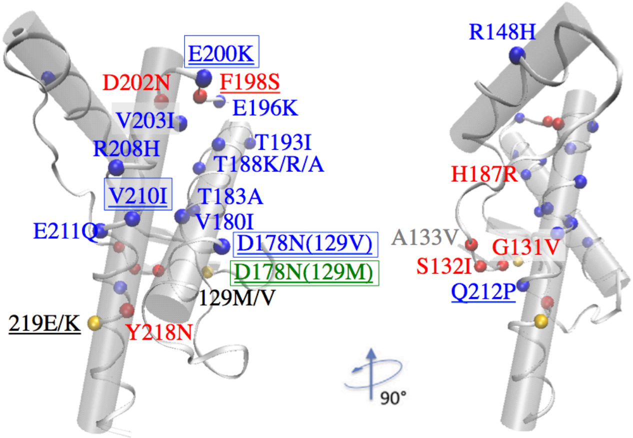


Figure 4.1: The location of the PMs and polymorphisms in HuPrP GD studied in this work. The PMs are colored according to the associated Hu TSE phenotypes: blue, familial Creutzfeldt-Jakob's disease (fCJD); red, Gerstmann-Sträussler-Scheinker disease (GSS); green, fatal familial insomnia (FFI); gray, unclassified phenotype. The polymorphisms are colored in black. Experimental structural information is available for the underlined variants (E200K-129M[260], V210I-129M[148], Q212P-129M[147] and E219K-129M[290] from NMR, F198S-129V[279] and D178N-129M/V[279] from X-ray crystallography). PMs with the highest worldwide incidence are boxed[286].

4.2 Computational details

4.2.1 The simulation systems

The structural models of the mutants were based on HuPrP(WT) GD NMR structure (residues 125-228, PDB entry 1HJN) at pH 7[250]. 1HJN is the only solution structure of HuPrP(WT) resolved at pH 7, which corresponds to the desired pH of our MD simulations. The best representative conformer of 1HJN is not indicated, whilst the minimized average structure, 1HJM[250], may not represent well the structure ensemble[291]. Therefore, a cluster analysis[292] was performed on the 20 conformers in 1HJN to identify a representative as the initial structure for the MD simulations. Using the procedure validated previously[263], the PMs and associated polymorphism were introduced into HuPrP(WT) using the Swiss-Pdb Viewer (also known as DeepView) 4.0[293] to obtain the initial models of 30 mutants.

Four histidines (H140, H155, H177, and H187) are present in the GD of HuPrP(WT). At pH 7, the histidine side chains may be neutral or, to a smaller extent, positively charged. To assign their protonation state at pH 7, we noticed the following points. i) H140, H155, and H177 are solvent-exposed, so that one or two of their side-chain nitrogen atoms could be protonated. ii) H187 forms a single intramolecular hydrogen bond (HB) which involves H187 N_ε atom and R156 backbone carbonyl. Hence, it is very likely that N_ε is protonated and the N_ε-H group acts as the HB donor. On the other hand, N_δ is solvent-exposed, thus it could be protonated or deprotonated. iii) Test calculations performed here (see Appendix 1.1 for details) along with an earlier study[294] point to an instability of the structure featuring the charged, di-protonated states. Instead, by protonating all the histidine residues at N_ε atom only, the simulations could reproduce accurately the structural determinants of the NMR structure of HuPrP(WT) resolved at pH 7 (discussed in Appendix 1.1 and ref. [263]). Inspection of all of the 30 mutant structures showed that point i) and point ii) hold true for them as well. Hence, we chose to protonate the four histidine residues at N_ε in all of the mutants considered here. With such a choice of the protonation state, the calculations turned out to reproduce fairly well the structural determinants of the available NMR structures of HuPrP variants including HuPrP(E200K-129M)[260], HuPrP(V210I-129M)[148], HuPrP(Q212P-129M)[147] and HuPrP(PP)[290] (discussed in section 4.3 and ref. [263]).

The aspartic acid, glutamic acid, lysine and arginine residues were considered in their ionized form. The N- and C-termini of the protein were considered charged, same as in the HuPrP(WT) NMR structure[250]. Following the protocol of ref. [263], we solvated each model in a box of explicit water, ensuring that the solvent shell extends for at least 16 Å around the protein. Thus, each system contained 10,140 to 10,160 water molecules and 2 to 4 Na^+ counter ions. The total number of atoms ranged from 31,600 to 33,400. Periodic boundary conditions were applied.

4.2.2 MD simulations

Same as in ref. [263], the AMBER99 force field[288; 289] was used for the protein, while the Åqvist parameters[295] and the TIP3P model[296] were used for the counter ions and the water molecules, respectively. Long-range electrostatic interactions were treated with the particle mesh Ewald method[297]. A Fourier spacing of 1.2 Å combined with a fourth-order cubic spline interpolation was used. A 12 Å cutoff was used for van der Waals interactions as well as the real-space part of the electrostatic interactions[298]. All bond lengths were constrained with the LINCS algorithm and the time step was set to 2 fs. NPT ensemble (T=298K, P=1bar) MD simulations were carried out by coupling the systems with the Nosé-Hoover thermostat[299] and the Parrinello-Rahman Barostat[300; 301].

The systems underwent 1000 steps of steepest-descent energy minimization with 5000 $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$ harmonic position restraints on the protein, followed by 2500 steps of steepest-descent and 2500 steps of conjugate-gradient minimization without restraints. For each system, the starting temperature of MD simulation was 2K. The system was then gradually heated in 12 steps of 100-ps simulation (from 2K to 25K, 50K, 75K, 100K, 125K, 150K, 175K, 200K, 225K, 250K, 275K and 298K). For each step, the velocity was generated consistent with a Maxwell-Boltzmann distribution at the corresponding temperature. After that, 200-ps equilibration simulations were performed. Finally, 50-ns production MD simulations were carried out. This exact protocol had already been validated in the previous study on HuPrP variants[263], where four sets of 50-ns independent MD runs were carried out for HuPrP(WT), HuPrP(E200K-129M), HuPrP(Q212P-129M) and HuPrP(E219K-129M). The results obtained turned out to be very similar in the four sets. Moreover, they reproduced very accurately all of the available experimental information. Thus, here we performed one single 50-ns MD run for each mutant, since it had been shown to provide reliable insights into the structural determinants of PMs in ref. [263]. All the simulations were carried out using GROMACS 4.0.7[302] package.

4.2.3 Data analyses

Solvent-accessible surface area. The solvent-accessible surface area was calculated using the GROMACS g_sas module with a probe radius of 1.4 Å. The van der Waals radii used were 1.5 Å for C, 1.2 Å for F, 0.4 Å for H, 1.1 Å for N, 1.05 Å for O, and 1.6 Å for S.

Salt bridge and π-stacking interactions. A distance cutoff of 5.0 Å was used as a geometry criterion for the presence of both salt bridge (SB) and π-stacking interactions between the interacting groups. For SB, the distance, d_{SB}, was calculated between the center of mass of anionic carboxylate ($RCOO^-$) of either aspartic acid or glutamic acid and that of the cationic ammonium (RNH_3^+) of lysine or the guanidinium [$RNHC(NH_2)_2^+$] of arginine. Because these SBs were dynamically formed and broken during the simulations, an average d_{SB} value over the entire trajectory did not describe well the presence/absence of the SBs. Therefore, we performed a cluster analysis on each trajectory for each SB using the GROMOS method[303] implemented in the GROMACS g_cluster module with an RMSD cutoff of 1.0 Å. The main clusters were selected, which cover ~80% of the overall trajectory. The SBs in the representative conformer of each cluster were analyzed according to the geometry criterion. For evaluating π-stacking interactions, the distance, d_π, was calculated between the centers of mass of non-hydrogen atoms belonging to the aromatic rings.

HB interactions. HBs were examined using the geometry criterion including a cutoff of 3.5 Å for the distance between the HB donor and acceptor and a cutoff of 30° for the acceptor–donor–hydrogen angle.

Protein flexibility. We used directional statistics in one-dimension space[304] to calculate the spread of N values of the Ramachandran angles, $(\phi_1,\ldots\ \phi_j,\ldots\phi_N)$ and $(\psi_1,\ldots\psi_j,\ldots\psi_N)$, during the simulations. In this method[305], each angle value is represented as a vector in a complex space: $e^{i\omega_j}$, where ω_j is the angle formed with the real axis. $\omega = \phi$ or ψ. The vector is completely defined by its direction, which depends on ω_j, because the modulus is one. The spread of the set of ω_j values is defined in terms of an angular dispersion index (ADI_ω)[304]. This can be expressed by

$$\mathrm{ADI}_\omega = \frac{1-\dfrac{1}{N}\left|\sum_{j=1}^{N} e^{i2\omega_j}\right|}{2\left(\dfrac{1}{N}\left|\sum_{j=1}^{N} e^{i\omega_j}\right|\right)^2}. \tag{4.1}$$

Within this definition, the spread is minimum when $\mathrm{ADI}_{\omega}= 0$: all N unit vectors are aligned on the same direction and the vector from N data points has the length $\sum_{j=1}^{N} e^{i\omega_j} = N$. The spread is maximum and $\sum_{j=1}^{N} e^{i\omega_j} = 0$, if ADI_{ω} tends to infinity. The ADI of the residue is then defined as the geometrical mean of the angular dispersion of the ϕ and ψ dihedral angles:

$$\mathrm{ADI}=\sqrt{\mathrm{ADI}_{\phi} * \mathrm{ADI}_{\psi}}. \tag{4.2}$$

It is a measure of the backbone local fluctuations: the larger the ADI, the larger the local flexibility and/or the local rearrangements. It has the following advantages over RMSF (see Appendix 1.2 for more details): i) it requires no alignment of the different configurations along the trajectory to a reference structure; ii) it does not suffer for edge effects, i.e. it does not increase systematically in the terminus regions; iii) the value for each residue is strictly dependent on the residue itself.

ADI may be used to characterize the flexibility of a secondary-structure element by averaging its values over the residues belonging to this element. We defined $\mathrm{ADI\_}\alpha_i$ (i = 1, 2, 3) as the mean value of the ADI calculated for all the residues belonging to each of the three helices in HuPrP. The error assciated to $\mathrm{ADI\_}\alpha_i$ is the standard error of the mean $\sigma/\sqrt{n}$, where n is the number of residues belonging to $\alpha 1$, $\alpha 2$ or $\alpha 3$ and σ is the standard deviation of the corresponding ADI values.

The ADI calculations have been implemented into a tcl script, which can be utilized with VMD[306].

4.3 Results

Our MD simulations show that the secondary-structure elements present in HuPrP(WT) are largely preserved in all the 30 PMs in the 50-ns timescale. The RMSD of C_{α} atoms relative to HuPrP(WT) range from 2.1 to 3.2 Å in the mutants (Table 4.1).

Table 4.1: RMSD[i] of C_α atoms during the simulations of HuPrP variants with respect to the NMR structure of HuPrP(WT).

HuPrP variant	RMSD (Å)	HuPrP variant	RMSD (Å)
WT	2.4±0.3	E196K-129M	2.8±0.4
G131V-129M	2.3±0.1	F198S-129V	2.3±0.3
S132I-129M	2.3±0.4	E200K-129M	2.0±0.2
A133V-129M	2.2±0.3	E200K-129V	2.3±0.4
R148H-129M	3.0±0.3	D202N-129M	2.7±0.4
D178N-129M	2.4±0.3	D202N-129V	2.7±0.4
D178N-129V	2.3±0.4	V203I-129M	2.8±0.4
V180I-129M	2.5±0.3	R208H-129M	2.3±0.3
T183A-129M	3.1±0.2	R208H-129V	2.9±0.2
T183A-129V	3.1±0.4	V210I-129M	2.7±0.2
H187R-129V	2.4±0.3	V210I-129V	2.8±0.2
T188K-129M	3.1±0.2	E211Q-129M	2.4±0.3
T188A-129M	2.7±0.3	Q212P-129M	2.0±0.2
T188A-129V	2.4±0.3	Q217R-129V	2.5±0.6
T188R-129V	2.6±0.3	Y218N-129V	3.9±0.4
T193I-129M	3.2±0.3	PP	2.0±0.2

[i] Standard deviations (SDs) are indicated by "±".

The 30 PMs lose completely or partly a SB network which embraces the α2-α3 region and the α1 helix (see Appendix 1.3 for detailed data). This network is stably present in the simulations of HuPrP(WT) and in HuPrP(PP) (Fig. 4.2 and Table 4.2), consistent with the NMR structures[250; 290]. It involves five residues in α3 (E200, K204, E207, R208, and E211), two residues in α1 (E146 and R156) and one residue in the α2–α3 loop (E196) (Fig. 4.2 and Table 4.2). The disruption of the SB network has been observed also in the available NMR structures of the PMs, namely HuPrP(E200K-129M)[260], HuPrP(V210I-129M)[148] and HuPrP(Q212P-129M)[147]. In particular, the R156-E196 and the E146-K204 SBs play a role in preserving the tertiary structure of the GD[252; 265; 266; 269]. The former connects the α2-α3 region to the C-terminal part of α1 (C-α1, hereinafter), and the latter connects α3 to α1, respectively. These SBs are lost in the majority of the 30 PMs (Fig. 4.2 and Table 4.2). The disruption of the SB network is accompanied by the formation, in most but not all the PMs, of a SB between K204 and E207 in the α3 helix, whereas this SB is absent in HuPrP(WT) or HuPrP(PP) (Fig. 4.2 and Table 4.2).

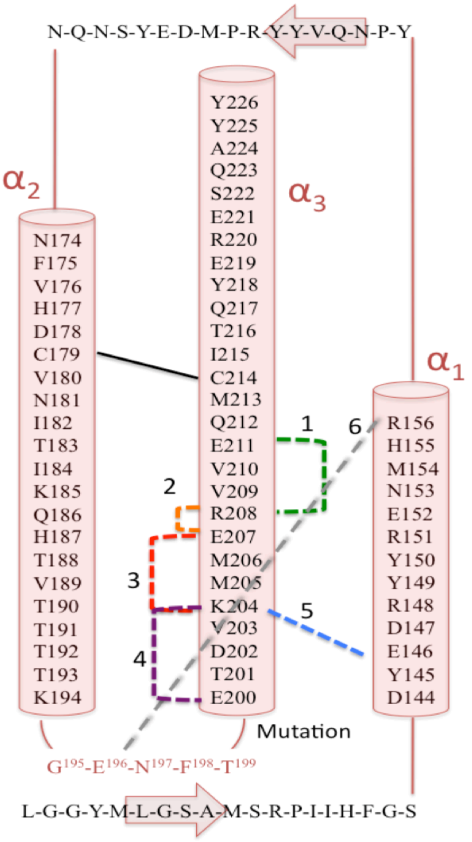


Figure 4.2: Schematic representation of the SB network embracing the α2-α3 region and the α1 helix. Each SB is indicated with dotted line and a corresponding number: **1**, E211-R208 SB; **2**, E207-R208 SB; **3**, E207-K204; **4**, E200-K204 SB; **5**, E146-K204 SB; **6**, E196-R156 SB. HuPrP(WT) and HuPrP(PP) feature the SBs **1**, **2**, **4**, **5** and **6**. Most of these SBs are lost in the PMs, whilst a new SB, **3**, is formed. The SBs are identified using a cluster analysis (see 3.2.2 for the method). Helices and β-strands are indicated with cylinders and arrows respectively. The C179-C214 disulfide bond is shown with a solid line.

Table 4.2: The presence of SBs in the SB network[i] embracing the α2-α3 region and the α1 helix in the main clusters of each HuPrP variant.

HuPrP variant	α3	α1 with α3	α1 with the α2-α3 loop	HuPrP variant	α3	α1 with α3	α1 with the α2-α3 loop
WT	1, 2, 4	5	6	E196K-129M	3	–	–
G131V-129M	3	–	6	F198S-129V	–	–	6
S132I-129M	1	–	6	E200K-129M	3	–	6
A133V-129M	1, 2	5	6	E200K-129V	1	5	6
R148H-129M	3	–	–	D202N-129M	–	–	–
D178N-129M	–	–	–	D202N-129V	3	–	6
D178N-129V	–	–	–	V203I-129M	–	–	–
V180I-129M	–	–	–	R208H-129M	–	–	–
T183A-129M	–	–	–	R208H-129V	–	–	6
T183A-129V	4	–	–	V210-129M	1, 2	–	–
H187R-129V	–	–	–	V210I-129V	1, 2	–	–
T188K-129M	–	–	–	E211Q-129M	–	–	-
T188A-129M	–	5	–	Q212P-129M	1, 2	–	–
T188A-129V	–	5	6	Q217R-129V	–	–	–
T188R-129V	3	–	6	Y218N-129V	–	–	6
T193I-129M	–	5	–	PP	1, 2, 4	5	6

[ii] The numbers represent different SBs, same as those used in Fig. 4.2 (1, E211-R208; 2, E207-R208; 3, E207-K204; 4, E200-K204; 5, E146-K204; 6, E196-R156).

In the β2–α2 loop, hydrophobic intramolecular interactions between the loop and C-α3 (Fig. 4.3A) are undermined in the PMs compared to those in HuPrP(WT) and HuPrP(PP) (Table 4.3). These interactions include the Y169-F175-Y218 (already reported in ref. [263] for HuPrP(E200K-129M) and HuPrP(Q212P-129M)) and the M166-Y225 π-stacking interactions (Fig. 4.3B). The latter appear stable in the NMR structures of HuPrP(WT)[250] and HuPrP(PP)[290] (Table 4.5). These results reproduce the features observed in the NMR and X-ray structures of HuPrP(E200K-129M)[260], HuPrP(V210I-129M)[148], HuPrP(Q212P-129M)[147], HuPrP(F198S-129V)[279], and HuPrP(D178N-129M/V)[279] (Table 4.4).

Table 4.3: π-stacking interactions in the β2-α2 loop and C-α3 region described as average d_{π} values (defined in section 4.2.3).

HuPrP variant	Y169-F175[i] d_{π} (Å)	F175-Y218[i] d_{π} (Å)	M166-Y225[ii] d_{π} (Å)
WT	4.5±0.2	4.6±0.03	4.9±0.7
G131V-129M	5.2±0.2	6.0±0.03	7.9±0.5
S132I-129M	5.7±0.7	5.2±0.03	4.7±0.8
A133V-129M	4.8±0.3	5.3±0.02	7.7±2.9
R148H-129M	4.8±0.2	5.4±0.02	6.1±2.0
D178N-129M	7.2±0.6	5.3±0.03	5.6±2.0
D178N-129V	6.6±1.0	5.2±0.03	5.2±1.5
V180I-129M	6.8±0.8	5.2±0.03	10.3±2.3
T183A-129M	7.4±0.3	5.5±0.03	10.8±1.3
T183A-129V	4.8±0.3	5.3±0.02	6.6±1.4
H187R-129V	10.7±1.6	5.1±0.02	6.8±1.7
T188K-129M	5.1±0.3	5.4±0.03	9.6±2.3
T188A-129M	7.3±0.4	5.4±0.03	9.8±2.7
T188A-129V	6.5±0.3	5.0±0.03	4.5±0.5
T188R-129V	5.4±0.3	5.4±0.03	9.6±3.2
T193I-129M	5.4±1.0	5.3±0.03	5.6±1.6
E196K-129M	4.9±0.3	5.3±0.03	10.4±3.0
F198S-129V	9.6±2.8	5.3±0.03	6.2±2.0
E200K-129M	8.4±0.7	5.0±0.04	6.0±2.0
E200K-129V	4.9±0.3	5.1±0.03	6.0±1.3
D202N-129M	5.4±0.9	5.3±0.03	6.3±3.0
D202N-129V	5.1±0.2	5.3±0.03	5.3±0.9
V203I-129M	5.6±0.6	5.3±0.03	8.3±4.4
R208H-129M	4.8±0.3	5.3±0.02	6.2±0.7
R208H-129V	6.0±0.9	5.2±0.03	6.7±2.9
V210I-129M	10.9±1.0	5.1±0.02	5.9±1.0
V210I-129V	10.5±1.0	5.1±0.02	7.0±2.0
E211Q-129M	7.3±0.3	5.4±0.03	10.3±2.5
Q212P-129M	8.6±0.5	4.7±0.03	5.5±3.0
Q217R-129V	5.5±0.9	5.1±0.03	4.9±0.5
Y218N-129V	6.4±0.10	–	6.3±1.5
PP	4.4±0.4	4.3±0.02	4.9±0.8

[i] Centroid distance between the aromatic rings (only carbon atoms considered).

[ii] Distance between M166 sulfur atom and Y225 aromatic centroid.

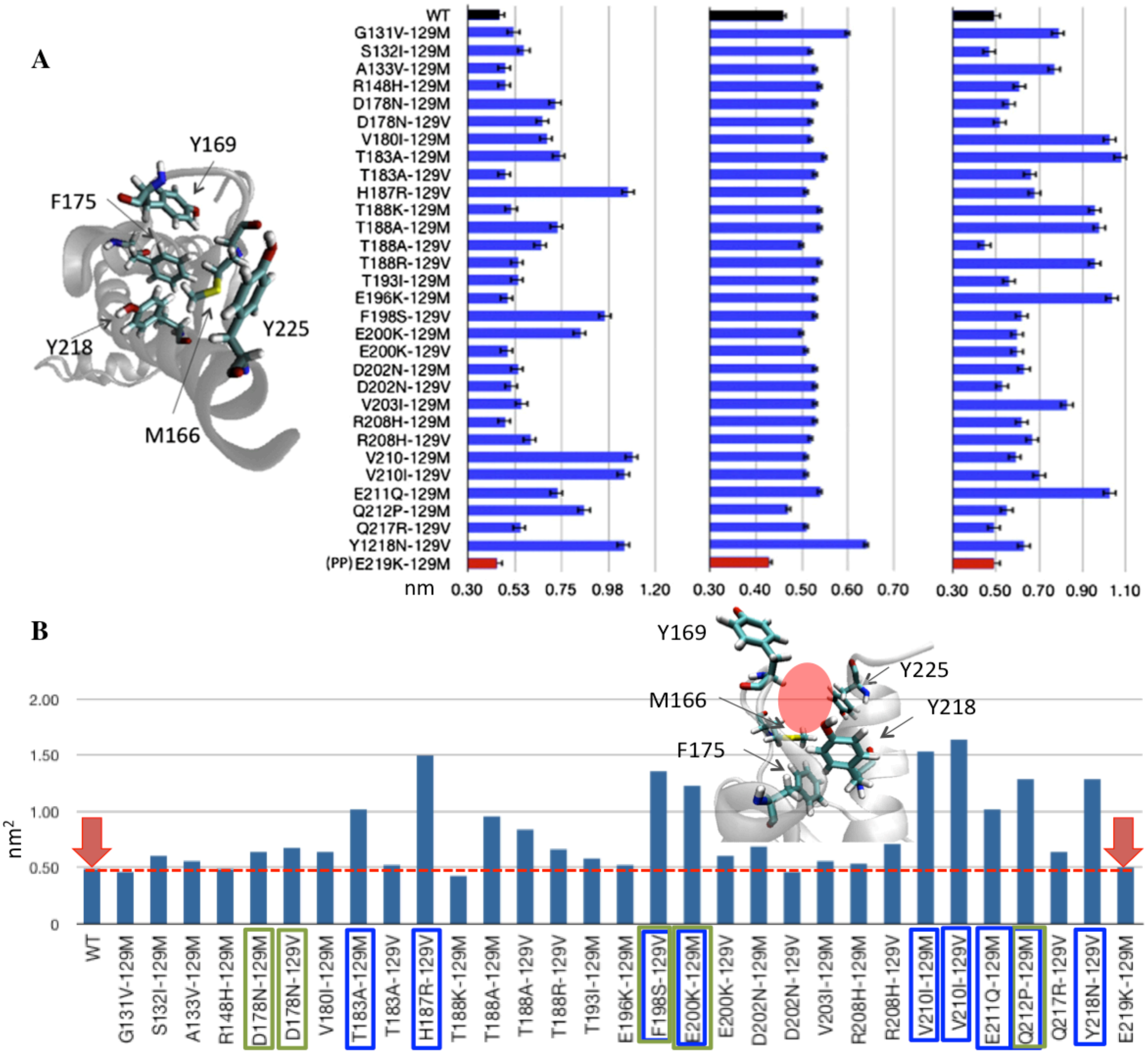


Figure 4.3: Selected properties emerging from the MD simulations. **A)** Y169-F175, F175-Y218, and M166-Y225 π-stacking interactions. The values in HuPrP(WT) and HuPrP(PP) are highlighted with black and red bars, respectively. The interactions are described here in terms of distances between the centers of mass of the aromatic rings. The cartoon represents the interactions in HuPrP(WT). **B)** Y169 solvent exposure. The values in HuPrP(WT) and HuPrP(PP) are highlighted with black and red bars, respectively. The red dotted line underlines the value for HuPrP(WT). The PMs in which Y169 is totally outside the loop and completely solvent-exposed are boxed in blue and green for the calculated and experimental structures[147; 260; 279], respectively. These PMs feature a large cavity formed by the β2-α2 loop and C-α3, which is indicated with a red shadow in the cartoon representation (upper part).

Table 4.4: Calculated and experimental π-stacking interactions in the β2-α2 loop and C-α3 region.

HuPrP variant	Y169-F175 (Å)		F175-Y218 (Å)		M166-Y225 (Å)	
	Experimental[i]	Calculated	Experimental	Calculated	Experimental	Calculated
WT[250]	5.3	4.5±0.2	5.3	4.6±0.3	4.9	4.9±0.7
D178N-129M[279, ii]	14.1	7.2±0.6	5.3	5.3±0.3	5.6	5.6±2.0
D178N-129V[279]	10.7	6.6±1.0	5.2	5.2±0.3	—[iii]	5.2±1.5
F198S-129V[279, ii]	14.2	9.6±2.8	5.6	5.3±0.3	5.8	6.2±2.0
E200K-129M[260]	14.2	8.4±0.7	5.5	5.0±0.4	6.0	6.0±2.0
Q212P-129M [32]	13.2	8.6±0.5	7.7	4.7±0.3	15.0	5.5±3.0
V210I-129M[148]	14.6	10.9±1.0	8.5	5.1±0.2	17.1	5.9±1.0
PP[290]	9.6	4.4±0.4	4.8	4.3±0.2	5.9	4.9±0.8

[i] Measured as the d_π values (defined in section 4.2.3) in the X-ray structure or the best representative conformer in the NMR structure.

[ii] Crystallized as a dimer[279].

[iii] Missing side chain.

The Y169 residue in the β2-α2 loop turns out to be more solvent-exposed in most PMs than in HuPrP(WT) and HuPrP(PP). This is evaluated by calculating the solvent-accessible surface area of Y169 along the trajectory of each HuPrP variant (Table 4.5). Remarkably, Y169 has been found completely outside the loop in HuPrP(T183A-129M), HuPrP(E200K-129M), HuPrP(H187R-129V), HuPrP(F198S-129V), HuPrP(V210I-129M/V), HuPrP (E211Q-129M), HuPrP(Q212P-129M), and HuPrP (Y218N-129V). As already noted in ref. [263], the predicted conformational change reproduces what is found in the available NMR structures of HuPrP(Q212P-129M) and HuPrP(E200K-129M)[147; 260]. Here, it is shown to also reproduce this feature in the X-ray structures of HuPrP(F198S-129V)[279] and the NMR structure of HuPrP(V210I-129M)[148]. It is worth noting that in the recently published NMR structure of HuPrP(PP)[290], the β2–α2 loop appears flexible and in some of the conformers Y169 is solvent-exposed. However, as the structure was resolved at pH 5.5, it may differ from that at pH 7. Indeed, acidic pH has been shown to decrease the stability and affect the structure of HuPrP[250].

The flexibility of helices α2 and α3 has been suggested to increase on passing from HuPrP(WT) to some PMs[261; 263; 265; 266; 269; 270; 272; 281; 283; 287]. The flexibility was evaluated mainly by calculating RMSF[263; 265; 266; 269; 272; 281]. Here we use ADI_α_2 and ADI_α_3 for characterizing the flexibility (as described in section 4.2.2). In most of the PMs, ADI_α_3 turns out to be larger than that of HuPrP(PP) and of HuPrP(WT) (Fig. 4.4A)[n]. This is caused, at least in part, by the loss of the SBs involving α3 on passing from HuPrP(WT) and HuPrP(PP) to the PMs. RMSD calculations (data not shown) and simple visual inspection of the investigated systems suggest that no significant

---

[n] The only exceptions are HuPrP(E200K-129V) and HuPrP(F198S-129V), whose ADI_α_3 are smaller than that of HuPrP(WT).

rearrangements of the helix take place. Hence, α3 appear to be more flexible in most of the PMs than in HuPrP(PP) (and HuPrP(WT)).

Table 4.5: Solvent-accessible surface area (SASA) of residue Y169 calculated over the MD trajectory of each HuPrP variant.

HuPrP variant	Y169 SASA (Å^2)[i]		HuPrP variant	Y169 SASA (Å^2)[a]	
WT	NE	4.9±1.3	E196K-129M	E	5.2±1.1
G131V-129M	E	4.6±0.8	F198S-129V	O	13.5±3.3
S132I-129M	VE	6.0±1.1	E200K-129M	O	12.3±1.1
A133V-129M	E	5.5±1.1	E200K-129V	VE	6.0±1.3
R148H-129M	NE	4.9±0.9	D202N-129M	VE	6.8±1.1
D178N-129M	VE	6.4±1.1	D202N-129V	NE	4.6±0.8
D178N-129V	VE	6.7±1.2	V203I-129M	E	5.6±1.3
V180I-129M	VE	6.4±1.5	R208H-129M	E	5.3±0.9
T183A-129M	O	10.2±0.9	R208H-129V	VE	7.1±1.3
T183A-129V	E	5.2±0.8	V210I-129M	O	15.3±2.8
H187R-129V	O	15.0±2.5	V210I-129V	O	16.4±2.4
T188K-129M	NE	4.0±0.9	E211Q-129M	O	10.2±1.0
T188A-129M	VE	9.6±1.5	Q212P-129M	O	12.9±1.6
T188A-129V	VE	8.4±1.0	Q217R-129V	VE	6.4±1.6
T188R-129V	VE	6.6±1.3	Y218N-129V	O	12.9±1.3
T193I-129M	E	5.8±1.0	PP	NE	5.0±1.2

[i] Not Exposed (NE): SASA≤ 5.0 Å^2, Exposed (E): SASA > 5.0 Å^2, Very Exposed (VE): SASA ≥ 5.0 Å^2, Outside (O): SASA ≥ 10.0 Å^2. SDs are indicated by "±".

The ADI_α2 values of HuPrP(WT) are smaller than those of the PMs, but those of HuPrP(PP) are not (Fig. 4.4B). A large rearrangement of C-α2 (T193-K194) is likely to significantly contribute to the relatively large value of ADI_α2 in HuPrP(PP). This rearrangement is accompanied by the formation of a HB between H155 and E196, as well as by an increase of intramolecular contacts in the protein (Fig. 4.4C). This HB is only partially formed in HuPrP(F198S-129V) and HuPrP(T193I-129M) and is not present in HuPrP(WT) and in the other 28 PMs (Table 4.6). At the speculative level, this might increase the stability of the protein in this variant. Interestingly, in the NMR structure of HuPrP(PP) at pH 5.5 (which was resolved after this study had been carried out), the conformation at C-α2 turn out very similar to that in our simulations. C-α2 is in shorter distance to the α1 helix than those in HuPrP(WT), which enhances the intramolecular contacts in this region, similar to that in our simulations.

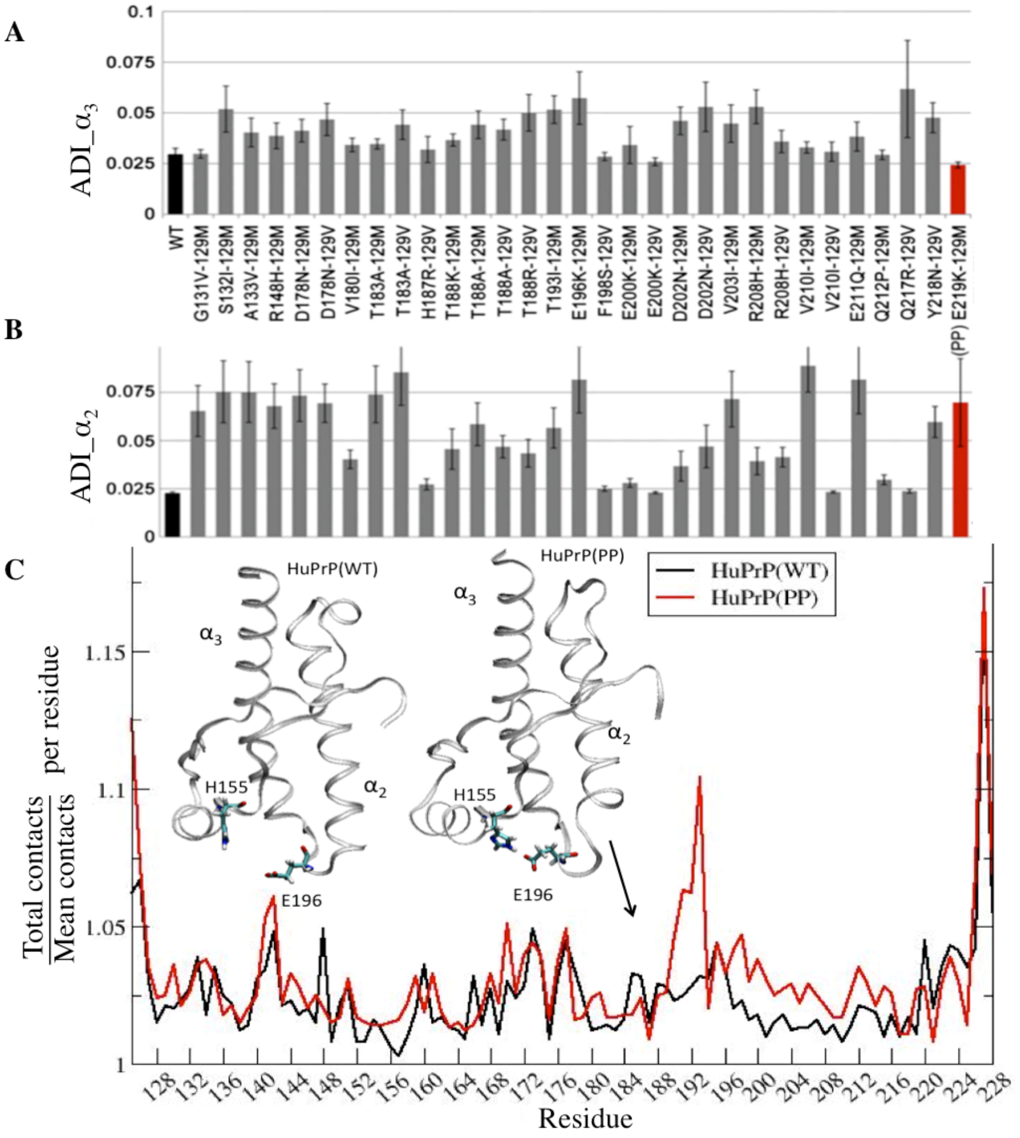


Figure 4.4: Flexibility analysis of α2 and α3 helices. **A)** ADI_α_3 and **B)** ADI_α_2 for all the HuPrP variants; HuPrP(WT) and HuPrP (PP) are highlighted with black and red bars, respectively; error bars indicate the standard error of the mean (see 3.2.2 for the definition). **C)** The observed conformational rearrangement in HuPrP(PP) at C-α2. Selected conformations of HuPrP(WT) and HuPrP(PP) are shown in cartoon representation. Residues H155 and E196 (in licorice) form a HB between in HuPrP(PP). The intramolecular contacts counted over the whole trajectories are plotted per residue for HuPrP(WT) (black line) and HuPrP(PP) (red line). HuPrP(PP) shows an increase of contacts in the region of the α2-α3 loop and the N-terminal of α3 relative to HuPrP(WT).

Table 4.6: Hydrogen bond (HB) between residues H155 and E196 during the MD simulations of each HuPrP variant, calculated as number of HB per frame.

HuPrP variant	H155-E196 HB[i]	HuPrP variant	H155-E196 HB	HuPrP variant	H155-E196 HB
WT	0.00	T188K-129M	0.00	V203I-129M	0.00
G131V-129M	0.00	T188A-129M	0.00	R208H-129M	0.00
S132I-129M	0.00	T188A-129V	0.00	R208H-129V	0.00
A133V-129M	0.01	T188R-129V	0.00	V210I-129M	0.30
R148H-129M	0.14	T193I-129M	0.84	V210I-129V	0.00
D178N-129M	0.00	E196K-129M	–	E211Q-129M	0.00
D178N-129V	0.00	F198S-129V	0.75	Q212P-129M	0.20
V180I-129M	0.00	E200K-129M	0.00	Q217R-129V	0.12
T183A-129M	0.00	E200K-129V	0.14	Y218N-129V	0.00
T183A-129V	0.26	D202N-129M	0.00	**PP**	**1.00**
H187R-129V	0.28	D202N-129V	0.00		

[i] Calculated as average number of HB per frame.

4.4 Discussion

In this study, we have carried out 1.6 μs MD simulations in explicit solvent for 30 HuPrP mutants containing known PMs in the GD. Along with the previous study on HuPrP(WT), HuPrP(PP), HuPrP(E200K-129M) and HuPrP(Q212P-129M), the simulations have been compared with experimental data[263] and other MD studies reported earlier[261; 262; 264; 265; 266; 267; 268; 269; 270; 271; 272; 281; 307]. The results point to common structural traits in the mutants independent of the position and the chemical nature of the PMs. These common traits include: i) the disruption of a specific SB network present in HuPrP(WT) and HuPrP(PP) in all PMs, with consequent increase in flexibility of α_3 in most PMs; ii) the loss or the weakening of hydrophobic interactions present in HuPrP(WT) and HuPrP(PP) in all PMs; iii) an increased solvent-exposure of Y169 in the β2-α2 loop relative to HuPrP(WT) and HuPrP(PP) in most PMs. These findings may have biological relevance as discussed below.

i) The partial or complete loss of the SB network in the α2–α3 region in all PMs may lead to a destabilization of the PrP fold in those regions. Indeed, according to *in silico* studies, the PrP fold of α2 and α3 is assisted by intramolecular interactions with the rest of the protein[308; 309; 310] rather than those of the α-helices themselves (as it has been already observed in other proteins with a significant content of α-helices[310; 311; 312]). The α_2-α_3 region is *per se* highly fibrillogenic[249; 255; 256; 257; 258; 310; 313], hence the disruption of intramolecular interactions (such as SBs) may play a key role in the $PrP^C \rightarrow PrP^{Sc}$ conversion. Most importantly, the residues involved in the SB interactions (Fig. 4.2, Table 4.2) are highly conserved[258] and their absence is linked with pathogenic mutants[258]. Possibly due to the loss of SBs, α3 is more flexible in most PMs.

ii) The weakening of hydrophobic interactions between the β2-α2 loop and C-α3 in the PMs is consistent with *in vitro* and *in vivo* studies[10; 11; 137; 138; 141; 142], which indicate that the long-range interactions involving these regions play a role in prion disease susceptibility. As already mentioned in Chapter 2, the β2-α2 loop could be highly flexible (same as in HuPrP[55; 56]) or well structured in some other mammalian species[135 136; 137; 138; 139; 140]. Linking the β2-α2 loop to C-α3 can partially inhibit oligomerization in ovine PrP[314]. Single-residue mutations in the β2-α2 loop and C-α3 can cause the otherwise flexible β2-α2 loop to adopt rigid structures[137; 138; 141]. Such mutations, when expressed in transgenic mice, could lead to spontaneous and progressive prion-like disease[11] or alter the susceptibility to different $\mathrm{PrP}^{\mathrm{Sc}}$ strains[10; 142].

(iii) The larger solvent-exposure of Y169 in the majority of PMs than in HuPrP(WT) and HuPrP(PP) is consistent with the fact that this residue, along with the rest of the β2-α2 loop and C-α3, forms a disease-linked epitope for a monoclonal antibody[315], a hypothetical facilitator of prion conversion involved in the development of TSE[316]. Therefore, the PMs are expected to interact differently with this antibody relative to HuPrP(WT). At the speculative level, we expect that this in turn may affect the $\mathrm{HuPrP}^{\mathrm{C}} \rightarrow \mathrm{HuPrP}^{\mathrm{Sc}}$ conversion. We further noticed that the rearrangement of Y169 creates a cavity (Fig. 4.3B), which might accommodate planar compounds that are able to inhibit prion fibrillization[317; 318; 319; 320].

4.5 Conclusion

Our calculations based on AMBER99 force field[288; 289] identify key structural features (few of which have already been reported for some PMs[147; 260; 261; 263; 270]) that emerge as common traits among most or all of the PMs concerning HuPrP GD.

From a physiological point of view, the altered conformation observed in the PMs might cause a different affinity for cellular membranes and, consequently, an aberrant localization of HuPrP in cellular compartments, favoring the formation of altered endoplasmic reticulum topologies[43; 321]. Independent evidence derived from cell-line assays on some of the PMs studied here shows that these mutations affect folding and maturation of $\mathrm{PrP}^{\mathrm{C}}$ in the secretory pathway of neuronal cells[322; 323].

The structure-function relationships suggested in this work may contribute to understanding the molecular basis of familial Hu TSE. They may also serve to provide a molecular explanation for $\mathrm{PrP}^{\mathrm{Sc}}$ formation in the more common sporadic forms of TSE.

Finally, the highlighted "hot spot" regions identified here correspond to the optimal targets proposed in several antibody- and ligand-binding studies[324; 325; 326; 327] as the key regions for $\mathrm{PrP}^{\mathrm{C}} \rightarrow \mathrm{PrP}^{\mathrm{Sc}}$ conversion or for interactions with cofactors.

We close this section by pointing out that our MD calculations are subject to limitations, as in any modeling study. In particular, the timescale of the simulations is several orders of magnitude shorter than that of the NMR experiments with which comparison has been made. However, in this specific case, the AMBER99-based MD protocol used here had been shown to reproduce strikingly well the available NMR structures[263]. Importantly, the predictions turned out to be rather independent of the initial conditions of the MD simulations[263]. Thus, for this specific system, one may be confident about the predictive power of the simulations as far as the NMR structural determinants are concerned.

Chapter 5 Dominant-negative Effects in Prion Diseases: Insights from MD Simulations on Mouse Prion Protein Chimeras

In Chapter 4, it has been shown that pathogenic mutations in HuPrP affect the protein structure in a common manner, which hints to the underlying mechanism of familial Hu TSE. In transgenic mice, mutations in mouse (Mo) PrP can determine the susceptibility of the mice to the infection of different PrP^{Sc} strains. Some of these mutations also show a dominant-negative effect, thus halting the replication process through which endogenous wild-type MoPrP is converted into MoPrP^{Sc}. Using all-atom MD simulations in explicit solvent, here we study and compare the structures of HuPrP, MoPrP, 10 Hu/Mo PrP chimeras and 1 Mo/sheep PrP chimera. Overall, ~2 μs of MD simulations are carried out. The simulations reveal distinct structural features in these PrP variants, which are related to the different responses to PrP^{Sc} infection. Our findings suggest that the interactions between α1 helix and the N-terminal of α3 helix in PrP^{C} are critical in the PrP^{C}→PrP^{Sc} conversion, whereas the β2-α2 loop conformation plays a role in the dominant-negative effect.

5.1 Introduction

The link between mutations in PrP and TSE has been further verified using transgenic (Tg) mice expressing mutations in the MoPrP gene[328; 329; 330]. When mutations associated with familial Hu TSE were expressed in Tg mice, these mice could spontaneously develop TSE-like neurodegenerative diseases[328; 329; 330]. Tg mice expressing both MoPrP and HuPrP did not exhibit abbreviated incubation times when infected with HuPrP^{Sc}. However, Tg mice expressing chimeric PrP in which a 9-residue segment was replaced with HuPrP sequence were found highly susceptible to HuPrP^{Sc} and exhibited abbreviated incubation times[129]. Recently, Tg mice expressing MoPrP chimeras (CMPrP) containing residues from other species have been reported to exhibit either transmission barriers against certain PrP^{Sc} strains[10] or spontaneous generation of PrP^{Sc}[11; 12]. Interestingly, a variety of CMPrP investigated in scrapie-infected Mo neuroblastoma (ScN2a) cells showed diverse susceptibility or resistance to MoPrP^{Sc} infection[13]. Some of these CMPrP also acted as "dominant-negatives" to inhibit the endogenous wild-type (WT) MoPrP from being converted into PrP^{Sc}. The dominant-negative effects of residue substitutions in PrP have been reported and deeply investigated in humans and in sheep. Naturally occurring polymorphisms in PrP have been known to influence the etiology and neuropathology of TSE both in humans[8] and in sheep[9]. In humans, the polymorphism E219K was found to be protective against sporadic CJD when expressed in heterozygosis[130; 131]. In sheep, three polymorphisms are acutely linked to the occurrence of sheep scrapie: A136V, R154H and Q171R/H. These generate five commonly observed alleles:

ARQ, ARR, AHQ, ARH and VRQ. ARR and AHQ are associated with resistance to scrapie, whereas ARQ, ARH and VRQ are linked with susceptibility to scrapie[132; 133].

Taken together, these experimental data suggest that subtle changes in PrP^C structure can alter PrP^{Sc} propagation, possibly by (de)stabilizing the physiological folding of PrP^C and/or affecting its interactions with some yet unknown factors. Here, we apply our well-established MD protocol (used in Chapter 4) to study the atomistic structures of 11 CMPrP globular domain (GD) (Table 5.1), which are then compared with the GD structures of WT MoPrP and HuPrP. Among these 11 CMPrP, 10 CMPrP are Hu/Mo chimeras that contain 1 or 2 mutations corresponding to the Hu codons. Another CMPrP, CMPrP(Q172R)[o], corresponds to a natural ovine dominant-negative polymorphism[331; 332]. All these 11 CMPrP have been previously investigated in ScNa2 cells for their response to $MoPrP^{Sc}$ infection[13]. We include CMPrP(Q172R) in this study because the mutation is located in the PrP^C $\beta 2$-$\alpha 2$ loop which composes an epitope crucial for prion susceptibility[10; 12]. Although CMPrP(Q172R) may not have the same significance as the Hu/Mo CMPrP, studying it provides additional information and confirms our findings with Hu/Mo CMPrP. The origin of these mutations as well as their *in vivo* and *in vitro* effects are listed in Table 5.1. This study may explain the transmission barrier between Hu and Mo PrP^{Sc} strains, as well as the dominant-negative effects of certain PrP polymorphisms.

---

[o] Hereinafter, each CMPrP is denoted (in parenthesis) with the mutation(s) in MoPrP numbering according to the equivalent residue(s) in HuPrP. This is to keep a coherent numbering throughout the thesis.

Table 5.1: Experimental observations of the CMPrP studied by MD simulations.

Mutation(s) in CMPrP	Experimental system[i]	Resistance to MoPrP^{Sc} replication[ii]	Dominant-negative effect[iii]	Origin of the mutation(s)
Q168E	ScN2a cells[13]	#	-	Hu codon
N171S	ScN2a cells[13]	-	-	Hu mutation linked to prion-like disorders[333]
Q172R	ScN2a cells[13]	+	+	Sheep protective polymorphism[331; 332]
V215I	ScN2a cells[13]	#	+	Hu codon
V215I_Q219E	ScN2a cells[13]	+	-	Hu codons
Q217R	ScN2a cells[13]	+	+	Hu GSS-linked mutation[156]
Q217R_Q219E	ScN2a cells[13]	+	-	R217: Hu GSS-linked mutation[156] E219: Hu codon
Q219K	ScN2a cells[13], Tg mice[334]	+	+	Hu protective polymorphism[335]
Q219E	ScN2a cells[13]	+	-	Hu codon
K220R	ScN2a cells[13]	-	-	Hu codon
K220R_Q219E	ScN2a cells[13]	+	-	Hu codons

[i] All the experiments were performed *in vitro* except for Q219K where the same findings have been reported also in Tg mice.

[ii] "+" indicates the CMPrP is NOT converted into PrP^{Sc}. "#" indicates the CMPrP is less efficiently converted than the control (WT MoPrP). "-" indicates the CMPrP shows no resistance to MoPrP^{Sc}.

[iii] "+" indicates the CMPrP inhibits the co-expressed MoPrP from being converted to PrP^{Sc} (the dominant-negative effect). "-"indicates the CMPrP shows no dominant-negative effect.

5.2 Computational details

5.2.1 The simulation Systems

The structural models of MoPrP and the CMPrP GD were based on the NMR structure of MoPrP GD (PDB entry 1XYX) resolved at pH 4.5[58], the only solution structure available so far for MoPrP. Note that 1XYX is the refined structure of an earlier PDB entry 1AG2[336] resolved at the same pH. The best representative conformer (model 1) of the 1XYX ensemble[58] was used. Following the same procedure used in Chapter 4 (section 4.2.1), the mutations in CMPrP were introduced into WT MoPrP using Swiss-PdbViewer (DeepView 4.0)[293]. The structural model of HuPrP was the one used in Chapter 4 (section 4.2.1).

The aspartic acid, glutamic acid, lysine and arginine residues were considered in their ionized form at pH 7. Three histidine residues, H140, H177, and H187, are present in the GD of MoPrP and CMPrP. They have the same local environment as the corresponding histidines in HuPrP (already described in Chapter 4, section 4.2.1). Same as in Chapter 4, test simulations were carried out to determine the protonation states of H140, H177, and H187 in these simulation systems. The test calculations suggest that H140, H177, and H187 adopt the same N_ε-protonation neutral state as those in HuPrP (see Appendix 2.1 for details). Other protonation states introduced instability to the protein structure (Appendix 2.1).

5.2.2 MD Simulations

The simulations were performed using the same protocol described in Chapter 4 (section 4.2.2). Each system contained 9,540 to 10,720 water molecules and 0 to 2 Na^+ counter ions. The total number of atoms ranged from 29,700 to 33,240. However, the latest AMBER99SBidln (instead of AMBER99) force field[337] was used to treat the protein. A newer GROMACS version 4.5[338] was used to perform the simulations. Because our initial structural models are based on the PDB 1XYX resolved at pH 4.5, these models may differ considerably from the structures at pH 7. Therefore, we performed three independent MD runs on each system, starting from different random seeds for the initial velocity generation.

5.2.3 Data analyses

All the analyses were carried out with GROMACS 4.5 modules and VMD[339] except for the secondary-structure analysis which was performed using DSSP software[340].

5.3 Results

Three independent MD runs are carried out on HuPrP, MoPrP and each of the 11 CMPrP (Table 5.1). During all the MD runs the systems appear to reach equilibrium within 10 ns, as indicated by the RMSD of Cα atoms as a function of simulation time (Fig. A2.1 in Appendix 2). Thus, the last 50 ns of equilibrated trajectory are taken from each MD run for analysis, which yield 1.95 μs of dynamics data.

All CMPrP maintain the 3D fold in GD similar to MoPrP and HuPrP during the simulations. Large fluctuations are observed in the β2-α2 loop (residues 167-171) and the C-terminal of α3 helix (C-α3, residues 219-231) (Fig. 5.1). Indeed, these two fragments were poorly defined in MoPrP NMR structure, indicating structural disorder or increased mobility[341]. Excluding the flexible terminal residues (125-127 and 220-231), the GD of CMPrP during our simulations show RMSD of C_α atoms no more than 2.7±0.4 Å from the MoPrP NMR structure[341] or 2.8±0.7 Å from the HuPrP NMR structure[55] (Table 5.2). Consistently, the radius of gyrations (R_g) of C_α atoms averaged along the x, y and z axes ranges from 13.3±0.1 Å to 13.6±0.2 Å for residues 128-219 of each CMPrP (Table 5.2). All CMPrP except CMPrP(K220R) show smaller RMSF of Cα atoms than those of MoPrP. These fluctuations are similar to those of HuPrP (Fig. 5.1). CMPrP(K220R) exhibits large fluctuations in residues 133-157 which contain the β1-α1 loop and the α1 helix, as well as in residues 190-199 which correspond to the N-terminal of α3 (N-α3) (Fig. 5.1). These fluctuations are due to a loss of intramolecular interactions between α1 and N-α3, which are reported in detail below.

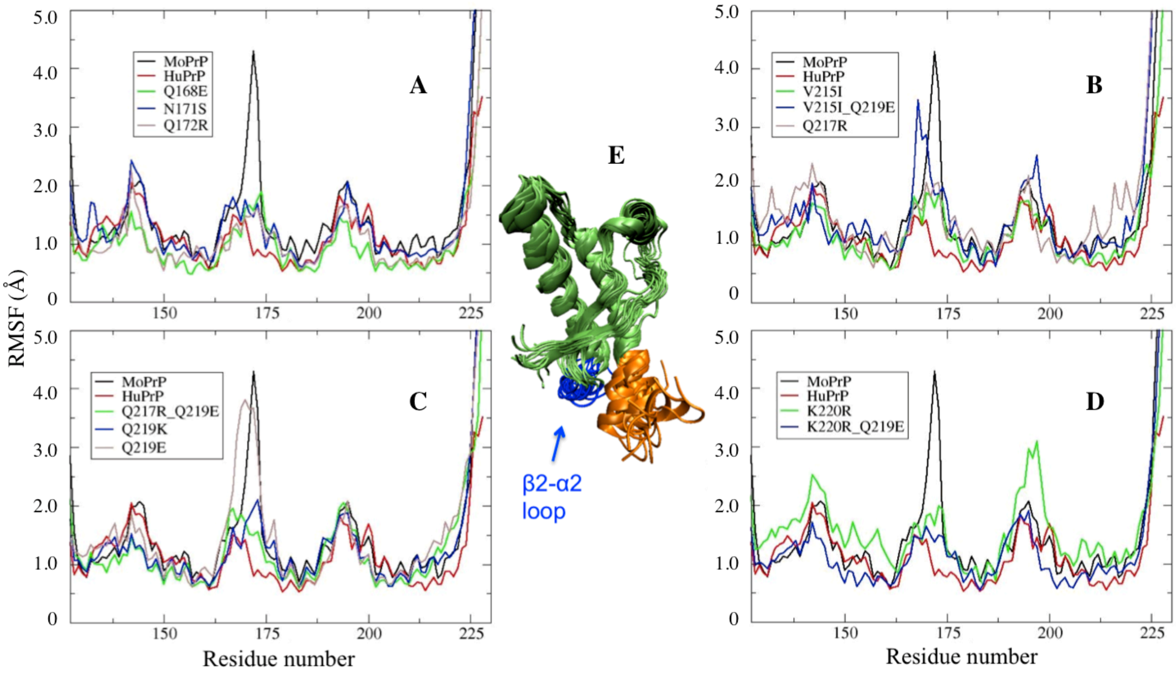


Figure 5.1: A-D) RMSF of C_α atoms in GD averaged over three independent 50-ns MD trajectories of each CMPrP, in comparison with those of MoPrP and HuPrP. The plot of each CMPrP is labeled with the mutation(s) in the CMPrP. **E)** The final conformation during the MD of MoPrP, of HuPrP and of all the CMPrP are superimposed on the HuPrP NMR structure[56]. The β2-α2 loop (residues 167-171) and C-α3 (residues 219-231) that experience the largest fluctuations are colored in blue and orange, respectively.

Table 5.2: RMSD of C_α atoms and R_g of residues 128-219 of MoPrP, HuPrP and all the CMPrP during MD. The RMSD are calculated relative to the NMR structures of MoPrP (PDB code: 1XYX) and HuPrP (PDB code: 1HJN), respectively.

PrP variant	MD run	RMSD (Å)		R_g (Å)	PrP variant	MD run	RMSD (Å)		R_g (Å)
		Ref. 1XYX	Ref. 1HJN				Ref. 1XYX	Ref. 1HJN	
MoPrP	1	1.8±0.4	2.4±0.2	13.6±0.2	CMPrP (V215I_Q219E)	1	1.7±0.1	2.0 ±0.2	13.4±0.1
	2	1.8±0.2	2.1±0.2	13.3±0.1		2	1.9±0.2	2.1±0.2	13.4±0.1
	3	1.8±0.2	1.9±0.1	13.3±0.1		3	1.6±0.2	2.2±0.2	13.4±0.1
HuPrP	1	1.9±0.2	1.7±0.2	13.5±0.2	CMPrP (Q217R)	1	1.7±0.2	2.2±0.2	13.3±0.1
	2	1.7±0.2	2.2±0.2	13.2±0.1		2	1.3±0.1	2.2±0.2	13.3±0.1
	3	1.8±0.2	2.0±0.2	13.3±0.1		3	2.4±0.2	2.1±0.2	13.6±0.1
CMPrP (Q168E)	1	1.8±0.2	2.0 ±0.2	13.5±0.1	CMPrP (Q217R_Q219E)	1	1.8±0.6	2.0±0.2	13.3±0.1
	2	2.4±0.2	2.1±0.2	13.7±0.1		2	1.7±0.2	2.1±0.6	13.5±0.1
	3	1.7±0.2	2.0±0.2	13.3±0.1		3	1.7±0.2	1.8±0.2	13.5±0.1
CMPrP (N171S)	1	1.8±0.2	2.0±0.2	13.5±0.1	CMPrP (Q219K)	1	1.7±0.3	2.0 ±0.2	13.6±0.2
	2	2.1±0.2	2.3±0.2	13.5±0.1		2	1.4±0.2	2.3±0.2	13.3±0.1
	3	2.2±0.2	2.1±0.2	13.3±0.1		3	1.6±0.2	2.1±0.2	13.4±0.1
CMPrP (Q172R)	1	1.4±0.2	2.2±0.2	13.4±0.1	CMPrP (Q219E)	1	1.9±0.2	2.2±0.2	13.6±0.2
	2	1.5±0.2	1.9±0.2	13.4±0.1		2	1.6±0.2	2.0±0.2	13.4±0.1
	3	1.6±0.2	2.6±0.2	13.2±0.1		3	1.8±0.2	2.1±0.3	13.4±0.2
CMPrP (V215I)	1	1.6±0.2	2.0±0.2	13.4±0.1	CMPrP (K220R)	1	2.6±0.4	2.8±0.6	13.5±0.2
	2	1.7±0.2	2.0 ±0.2	13.5±0.1		2	1.8±0.2	2.4±0.2	13.3±0.1
	3	1.7±0.2	2.1±0.2	13.3±0.1		3	1.9±0.3	1.9±0.2	13.5±0.1
					CMPrP (K220R_Q219E)	1	1.4±0.2	2.1±0.2	13.3±0.1
						2	1.5±0.2	2.1±0.2	13.3±0.1
						3	1.6±0.2	2.3±0.2	13.3±0.1

We analyzed the interactions between α1 and N-α3 in all simulated systems. Strikingly, 8 CMPrP (Q168E, Q172R, V215I_Q219E, Q217R, Q217R_Q219E, Q219K, Q219E, K220R_Q219E) exhibit stable HB interactions between residues Y149 and D202, Y157 and D202 (Table 5.3, Fig. 5.2A). These HBs appear less stable in MoPrP and HuPrP (Table 5.3). A salt bridge (SB) between R156 and E196 is observed in these 8 CMPrP as well as MoPrP and HuPrP (Table 3, Fig. 5.2B). These HB and SB interactions keep the α1 helix in close contact with N-α3, conserving the hydrophobic interactions among residues F141, Y149, Y157, M205 and M206 at this interface. The same interactions are observed also in CMPrP(V215I), albeit less pronounced than in the above 8 CMPrP. By contrast, CMPrP(N171S) and CMPrP(K220R) lack these HB and SB interactions during most of the simulation time (Table 5.3, Fig. 2C and D). Without them, the hydrophobic interactions are undermined and the α1 helix shifts apart from N-α3 (Fig. 2E). Correspondingly, the average Cα-distance between residue D144 (at C-α1) and residue D202 (at N-α3) increases significantly in CMPrP(N171S) and CMPrP(K220R) (17.7±1.0 Å and 17.5±1.5 Å, respectively). In sharp contrast, this distance is below 16.0 Å in 7 CMPrP (Q172R, V215I_Q219E, Q217R, Q217R_Q219E, Q219K, Q219E, K220R_Q219E). In CMPrP(Q168E), CMPrP(V215I), MoPrP and HuPrP this distance is between 16.0 and 17.0 Å (Table 5.3).

Next we examined the β2-α2 loop where all CMPrP and HuPrP exhibited smaller RMSF than MoPrP. We analyzed the secondary-structure elements of this loop during each simulation. Residues 169-171 showed an exchange between coil and bend conformations in all the systems. Residues 166-168, however, exhibited distinct exchange patterns in different PrP variants. In MoPrP, residues 166-168 feature an exchange pattern mainly between 3_{10}-helix (68.9%) and turn (29.6%) conformations, whereas in HuPrP an exchange pattern among bend (48.2%), turn (35.5%) and coil (16.2%) is observed. Out of the 11 CMPrP studied here, seven (N171S, Q172R, V215I, Q217R, Q219K, K220R and K220R_Q219E) show the '3_{10}-helix/turn'-exchange pattern at residues 166-168, similar to that of MoPrP (Table 5.4). In such a Mo-like pattern, 3_{10}-helix covers over 45% of the simulation time while coil occurs in less than 3%. The 3_{10}-helix conformation is stabilized by two HBs: one is formed between the backbone of Q168 N atom and P165 O atom, the other is between the backbone of Y169 N atom and V166 O atom. The other 4 CMPrP (Q168E, V215I_Q219E, Q217R_Q219E and Q219E) manifested a more Hu-like pattern: the 'bend/turn'-exchange dominates the loop conformation, whereas 3_{10}-helix covers less than 20% and coil accounts for over 10% (Table 5.4). A HB formed between D167 backbone N atom and S170 sidechain O atom stabilizes the bend conformation. Alternatively, a HB between the E168 backbone N atom and M166 backbone O atom contributes to the turn conformation. Our analyses did not detect any SB or hydrophobic interactions that particularly contribute to these specific loop conformations.

Table 5.3: Selected intramolecular interactions between α1 and N-α3 of MoPrP, of HuPrP and of all the CMPrP during MD. Related experimental findings are also reported.

PrP variant	MD run / NMR structure	HB[i] Y149-D202	HB[i] Y157-D202	SB[i] R156-E196	D144-D202 Cα distance (Å)[ii]	Experimental findings[13; 334]	
						Resistance to MoPrP^{Sc} replication[iii]	Dominant-negative effect[iv]
MoPrP	1XYX	16/20	7/20	1/20	14.8±0.8	-	-
	1	0.3	0.3	0.0	16.7±1.1		
	2	0.8	0.6	0.3			
	3	0.4	0.8	0.7			
HuPrP	1HJN	5/20	9/20	2/20	15.0±1.0	N/A	N/A
	1	0.9	0.8	0.7	16.5±1.0		
	2	0.5	0.3	0.9			
	3	1.0	0.3	0.5			
CMPrP (Q168E)	1	1.0	0.9	0.6	17.0±1.5	#	-
	2	1.1	1.0	0.7			
	3	0.9	1.0	0.9			
CMPrP (N171S)	1	0.0	0.0	0.0	17.7±1.0	-	-
	2	0.0	0.0	0.1			
	3	0.1	0.0	0.0			
CMPrP (Q172R)	1	1.0	1.0	0.7	15.8 ±0.4	+	+
	2	1.0	1.0	1.0			
	3	1.0	0.9	0.2			
CMPrP (V215I)	1	1.1	1.0	0.5	16.5±1.0	#	+
	2	0.7	0.5	0.0			
	3	0.5	0.4	0.2			
CMPrP (V215I_Q219E)	1	1.0	1.0	0.8	15.7±0.4	+	-
	2	1.1	0.9	0.5			
	3	1.0	0.9	0.6			
CMPrP (Q217R)	1	1.0	1.0	0.7	15.6±0.4	+	+
	2	1.0	1.0	0.9			
	3	1.1	1.0	0.4			
CMPrP (Q217R_Q219E)	1	1.1	0.9	0.4	15.8±0.4	+	-
	2	1.1	1.0	0.5			
	3	1.0	0.8	0.3			
CMPrP (Q219K)	1	1.1	1.0	0.5	15.9±0.4	+	+
	2	1.1	1.0	0.5			
	3	1.1	1.0	0.5			
CMPrP (Q219E)	1	1.0	0.9	0.9	15.9±0.4	+	-
	2	1.1	1.0	0.5			
	3	1.0	1.0	0.9			
CMPrP (K220R)	1	0.0	0.0	0.0	17.5±1.5	-	-
	2	0.8	0.3	0.0			
	3	0.1	0.1	0.4			
CMPrP (K220R_Q219E)	1	1.0	1.0	0.7	15.9±0.4	+	N/A
	2	1.1	0.9	0.7			
	3	1.0	0.9	0.3			

[i] Calculated as the average number per frame in the MD run, or the number of conformers out of the total number of conformers in the NMR structure. The value of HB Y149-D202 can be slightly >1, because in a small number of frames D202 has the chance to form two HBs with Y149.

[ii] Averaged over 3 runs.

[iii, iv] The symbols have the same indications as those in Table 5.1.

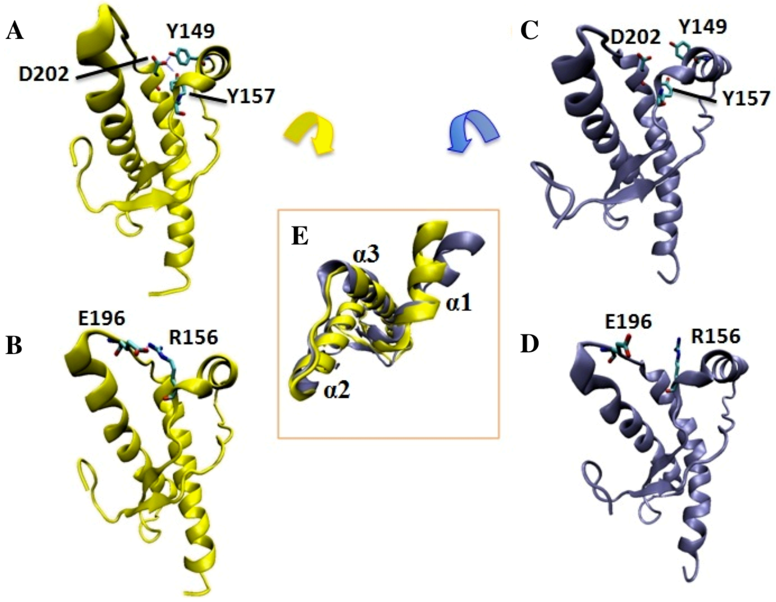


Figure 5.2: Selected HB and SB interactions between α1 helix and N-α3 are formed during 98% and 67%, respectively, of the simulation time of CMPrP(Q172R) which is resistant to MoPrP^{Sc} replication *in vitro*[13] (**A** and **B**). These interactions are absent in the non-resistant CMPrP(N171S) (**C** and **D**). **E)** The α1 helix shifts apart from N-α3 in CMPrP(N171S) (blue), relative to that in CMPrP(Q172R) (yellow). These features are common across the other resistant and non-resistant CMPrP.

Table 5.4: Percentage content of secondary-structure elements in residues 166-168 of MoPrP, of HuPrP and of all the CMPrP during the MD (average over three runs of each system).

PrP variant	3_{10}-helix (%)	Turn (%)	Bend (%)	Coil (%)
MoPrP	68.9	29.6	4.2	2.9
HuPrP	0	35.5	48.2	16.2
CMPrP(Q168E)	14.4	54.9	16.8	13.2
CMPrP(N171S)	49.0	37.5	10.4	1.4
CMPrP(Q172R)	62.2	30.0	6.3	1.5
CMPrP(V215I)	45.6	42.9	8.9	2.1
CMPrP(V215I_Q219E)	8.9	21.1	42.3	27.7
CMPrP(Q217R)	57.0	27.2	10.8	5.1
CMPrP(Q217R_Q219E)	14.1	31.2	29.4	25.3
CMPrP(Q219K)	62.8	24.3	5.7	1.5
CMPrP(Q219E)	20.0	38.5	18.6	22.2
CMPrP(K220R)	45.1	42.5	11.2	1.1
CMPrP(K220R_Q219E)	67.2	25.4	5.9	1.6

5.4 Discussion

Our simulations highlight stronger interactions between α1 helix and N-α3 in 7 CMPrP (Q172R, V215I_Q219E, Q217R, Q217R_Q219E, Q219K, Q219E, and K220R_Q219E) than those in MoPrP and HuPrP. Indeed, the average distance between C-α1 and N-α3 appears shorter in the 7 CMPrP than that in MoPrP and HuPrP. Notably, all these 7 CMPrP were reported to highly resist the MoPrP^{Sc} replication in ScN2a cells[13] (Table 5.1). Although less pronounced, these features are also observed in the MD of CMPrP(Q168E) and CMPrP(V215I). The latter two CMPrP were reported with only mild resistance in ScN2a cells[13]. By contrast, CMPrP(N171S) and CMPrP(K220R) lack these interactions during the MD. They exhibit longer distances between C-α1 and N-α3 than those in MoPrP and HuPrP. Correspondingly, these 2 CMPrP are both susceptible to MoPrP^{Sc} infection. These data indicate that the resistance to MoPrP^{Sc} is related to stronger intramolecular interactions between α1 helix and N-α3 in CMPrP. Our previous study in Chapter 4[150] suggests that a loss of SB interactions between α1 and N-α3 plays a role in the spontaneous conversion of Hu pathogenic mutants to PrP^{Sc}. Recently. a solution-NMR study on Hu protective polymorphism E219K has discovered stronger interactions between N-α3 and α1 as well as a few residues proceeding α1[342]. Interestingly, a number of MD studies suggest that the region from C-α2 to N-α3 is more flexible in HuPrP pathogenic mutants than in WT HuPrP[16; 150; 151; 343; 344; 345; 346; 347; 348]. Many other studies have pointed out that the α2-α3 region is *per se* highly fibrillogenic[255; 256; 349; 350; 351; 352]. Hence, the α2-α3 region is likely to be stabilized by interactions with α1 and its flanking residues. This hypothesis is in line with several investigations on the hydrophobic core of PrP amyloid fibrils. Hydrogen exchange, site-directed spin labeling, electron paramagnetic resonance spectroscopy and solid-state NMR studies have suggested that the core of PrP amyloid fibrils is composed of the region spanning the α2 helix, the α2-α3 loop and a major part of α3 helix of PrP^{C}[353; 354; 355]. These studies indicate a dissociation of the α2-α3 region from the rest of PrP during the conversion. Therefore, enhanced interactions between α1 and N-α3 might inhibit the dissociation and the conversion, such as seen in our simulations of the MoPrP^{Sc}-resistant CMPrP. However, some other studies suggest that amyloid fibrils can be formed by the structured regions preceding the α1 helix, likely via an extended β-structure involving residues 102-124. This hypothesis is supported by the extensive studies[356; 357; 358] on the pathogenic HuPrP mutation Y145stop, which contains only residues 23-144. These studies indicate that in the absence of the C-terminal residues 145-230, the fibril core is composed mostly of residues 102–139. While the misfolding mechanism of PrP remains obscure, a growing body of evidence suggests that distinct aggregate structures can be formed through multiple pathways by different PrP^{Sc} strains[359; 360; 361; 362; 363]. This may explain the discrepancy between our findings and some others derived from different PrP variants or PrP^{Sc} strains.

The β2-α2 loop of PrP conformation has been deeply investigated in various mammalian PrPs[136; 137; 364]. As already introduced in Chapter 2, it has been suggested that the loop plasticity is related to the transmissible barrier across PrP^{Sc} strains[58; 60; 137; 138; 140; 365; 366; 367; 368]. Solution-NMR studies on MoPrP and HuPrP structures found that at room temperature (293-298K) a line broadening of NMR signals was reported for the β2-α2 loop in all the Mo and Hu PrP constructs due to local conformational exchanges[55; 56; 58; 369]. A recent solution-NMR study discovered that at 310K, the β2-α2 loop of MoPrP was characterized by 3_{10}-helix, while Y169A/G mutations resulted in β-turn conformation[370]. This study suggests that at room temperature the β2-α2 loop of MoPrP is disordered, likely due to conformational exchange between 3_{10}-helix and other conformations. These findings imply that the loop conformation may play an essential role by binding to an unknown factor required for prion propagation. Consistently, previous studies on pathogenic HuPrP mutants (including ours in Chapter 4)[17; 150; 152] indicate that the conformation of the β2-α2 loop is related to the familial Hu TSE. In this study, we identify a '3_{10}-helix/turn' conformational exchange pattern of the β2-α2 loop in MoPrP and a 'bend/turn/coil' pattern in HuPrP. The intramolecular interactions that contribute to the 3_{10}-helix turn out to be the same HBs as those observed in the MoPrP NMR structure at 310K, which involve residues P165, V166, Q168 and Y169[370]. Another 2 HBs are identified to control the bend and turn conformations, which involve residues M166, D167, E168 and S170. These specific interactions in the β2-α2 loop may be the determinants of the loop conformation. Seven CMPrP (N171S, Q172R, V215I, Q217R, Q219K, K220R and K220R_Q219E) exhibit the Mo-like pattern during the MD. Among these, two of them (CMPrP(N171S) and CMPrP(K220R)) have been shown to be susceptible to MoPrP^{Sc} in ScN2a cells, whereas the other four (CMPrP(Q172R), CMPrP(V215I), CMPrP(Q217R), and CMPrP(Q219K)) showed the dominant-negative effect[13]. CMPrP(K220R_Q219E) was resistant to MoPrP^{Sc}[13] but no data have been reported about the dominant-negative effect. The other four CMPrP studied here (CMPrP(Q168E), CMPrP(V215I_Q219E), CMPrP(Q217R_Q219E) and CMPrP(Q219E)) show a more Hu-like pattern. They turn out to be the ones resistant to MoPrP^{Sc} but not to have dominant-negative effect[13]. Our results indicate that the Mo-like conformational exchange pattern in the β2-α2 loop is necessary for the dominant-negative effect if the CMPrP is also resistant to MoPrP^{Sc}.

Based on these findings, we propose the following mechanism for the dominant-negative effect: the Mo-like conformation at the β2-α2 loop can bind to an unknown factor that facilitates PrP^{Sc} propagation. When the resistant CMPrP have the Mo-like β2-α2 loop conformation, they can bind to this factor but cannot be converted to PrP^{Sc}. They stay bound with the factor and keep its binding site inaccessible for the WT MoPrP. Therefore, WT MoPrP is not converted to PrP^{Sc}. In the case of non-resistant CMPrP, such as CMPrP(N171S) and CMPrP(K220R), they can bind to the factor but are easily converted to PrP^{Sc} and released from it. The binding site is then free again to bind new CMPrP or WT MoPrP. Therefore, WT MoPrP is also converted to PrP^{Sc} and no dominant-negative effect is observed.

Evidence exists that PrP^{Sc} propagation requires an accessory cofactor[371; 372; 373]. Recently, Deleault et al. demonstrated that a small-molecule cofactor can regulate the conformation, strains and infectivity of PrP^{Sc} *in vivo* [374]. However, some other studies argue against the intervention of any ancillary components[375] and suggest that the dominant-negative effect is only due to the heterozygosis of *PRNP* (in humans) or *Prnp* (in other species) genes[376; 377; 378]. This alternative model, denoted as the "stone fence model", suggests that human individuals heterozygous for E/K at codon 219 are protected by sporadic CJD because the PrPs deriving from two allelic variants interfere with each other in the conversion process due to their incompatible structures. The *in vivo* and *in vitro* evidence supporting this model is reviewed in[379]. However, an ultimate structural explanation of the dominant-negative effect is still missing in prion biology.

Considering the previous findings on CMPrP[13; 129], which have inspired this work, we propose the above possible mechanism of the dominant-negative effect. Nevertheless, further investigations are needed to verify this hypothesis. In particular, the alternative hypothesis that the dominant-negative effects are due to *PRNP* or *Prnp* heterozygosis may suggest novel rationally designed MD experiments aimed to investigate the interaction between the two different allelic variants to mimic the heterozygosis condition.

Chapter 6 Role of prion disease-linked mutations in the intrinsically disordered N-terminal domain of the prion protein

Among the currently identified PMs in the human *PRNP* gene, seven missense mutations are located in the naturally unfolded N-term of HuPrP. Intriguingly, and in sharp contrast to many PMs in the GD, these seven PMs do not affect the *in vitro* misfolding kinetics of the protein. Here we hypothesize that the neurotoxicity of these PMs arises from changes in structural determinants of the N-term, which affect the protein binding with its cellular partners and/or the co-translational translocation during the biosynthesis of the protein. We test this idea by predicting the conformational ensemble of WT and mutated MoPrP N-term, whose sequence is almost identical to that of the human one and for which the largest number of *in vivo* data is available. The conformational properties of the WT are consistent with those inferred experimentally. Importantly, the PMs turn out to affect in a subtle manner the intra-molecular contacts in the N-term at the putative binding sites for Cu^{2+} ions, sulphated glycosaminoglycans and other known prion protein's cellular partners *in vivo*. The PMs also alter the local structural features of the transmembrane domain and adjacent stop transfer effector, which act together to regulate the protein topology. These results corroborate the hypothesis that PMs affect the prion protein binding to its functional interactors and/or the translocation.

6.1 Introduction

Numerous studies have shown that PMs in the HuPrP GD can alter the GD's secondary-structure (SS) elements and structure flexibility[16; 147; 148; 150; 152] and accelerate the misfolding process *in vitro*[156]. Our previous work in Chapter 4 suggests that PMs and chimeric mutations in HuPrP or MoPrP GD affect common regions in the GD structure that are crucial for the protein's stability and interactions with cellular partners.

Seven missense PMs have been identified in the N-term of HuPrP (N-term_HuPrP hereinafter, Fig. 6.1A): Q52P has been recently found and it is likely associated with familial Creutzfeldt–Jakob disease (fCJD)[154], G114V has been reported with both fCJD[155] and Gerstmann-Sträussler-Scheinker syndrome (GSS)[156], whereas P102L, P105L/S/T, and A117V are associated with GSS[156]. The correspondent mutation of P102L in Mo PrP^{C} alters the incubation time in transgenic mice upon infection by PrP^{Sc} from human, hamster, sheep or murine sources in a strain-dependent manner[157; 158]. Intriguingly, these PMs may cause the disease without affecting the kinetics of the $HuPrP^{C}$ to $HuPrP^{Sc}$ conversion *in vitro*[156]. This hints at the presence of processes *in vivo* that lead these PMs *indirectly* to pathogenic forms of HuPrP, the hallmark of all Hu TSE, although structural information on the effects of these PMs is currently lacking.

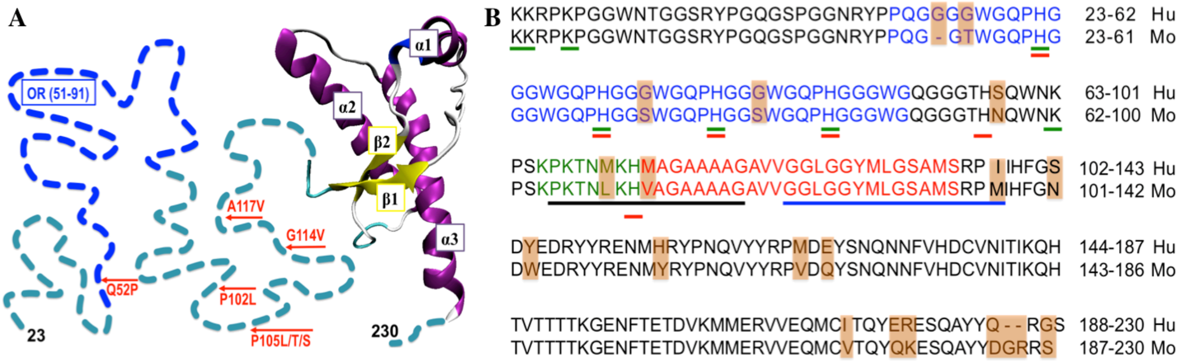


Figure 6.1: A) Molecular structure of HuPrP^C. The GD NMR structure[56] is shown in cartoon. The disordered N-term and three C-terminal residues, not present in the NMR structure, are sketched in dashed lines. Red arrows indicate the PMs studied in this work. **B)** Alignment of HuPrP^C and MoPrP^C sequences, obtained using BLAST[380]. The distinct residues in each of the two proteins are shadowed in orange. Those belonging to OR, STE, and TM1 are colored in blue, green and red, respectively. Residues at putative binding-sites for Cu^{2+}, sulphated GAG, vitronectin and STI1 (Table 6.1) are underlined in red, green, black and blue, respectively.

On the one hand, N-term_HuPrP, along with the equivalent domain in MoPrP (N-term_MoPrP, featuring 93% sequence identity with N-term_HuPrP, see Fig. 6.1B), contains a major part of the so-called transmembrane domain (termed TM1, comprising roughly residues 112–135) and the preceding "stop transfer effector" (STE, a hydrophilic region containing roughly residues 104–111)[43] (Fig. 6.1B). STE and TM1 act in concert to control the co-translational translocation at the endoplasmic reticulum (ER) during the biosynthesis of HuPrP^C and of Mo and Syrian hamster (SHa) PrP^C[38; 42]. Four topological forms of the protein can be generated during the translocation[38] (Fig. 2.1 in Chapter 2): ^{Sec}PrP, ^{Ntm}PrP, ^{Ctm}PrP and ^{cy}PrP (see Chapter 2 section 2.2.1 for details). Modified structural features of TM1 and STE can increase the proportion of the atypical forms ^{Ntm}PrP, ^{Ctm}PrP and ^{cy}PrP, which can be detrimental *in vivo*[38; 43; 47]. It is worth noting that 5 out of the 7 PMs in this study are located in the STE and TM1 regions ('STE/TM1', hereinafter).

On the other hand, N-term_HuPrP is a broad-spectrum molecular sensor[14]. Indeed, along with N-term_MoPrP, it interacts with chemicals including: i) sulphated glycosaminoglycans (GAG)[381], which bind to the four histidines in the so-called octapeptide-repeat region (OR, residues 51-91, see Fig. 6.1A and B), as well as several lysines at the N-terminus[70] (Table 6.1); ii) Cu^{2+} ions[382], which may bind at six histidine residues[383], including those located in the OR (Table 6.1). In addition, the Mo protein has been shown to interact with vitronectin[103], the stress-inducible protein 1 (STI1)[105] and amyloid-β (Aβ) multimers[106; 107; 108], which may bind N-term_MoPrP at residues 104–118, 114–127and 96-111, respectively, as well as with the lipoprotein receptor-related protein 1 (LRP1)[109] and the neural cell adhesion molecule (NCAM)[32] (Table 6.1).

Table 6.1: Binding partners of Hu/Mo/SHa PrP^{C} N-term so far identified by *in vivo* studies. The affinity and the binding site identified *in vitro* for these partners are also reported.

Binding partner	Proposed role(s) *in vivo*	PrP^{C} interacting residues (*in vitro*)	Affinity (*in vitro*)	References
Sulphated GAG	In the formation of Hu/Mo/SHa PrP^{Sc}[381; 384]	In Hu/Mo/SHa PrP^{C}: K23, K24, K27, H61, H69, H77, H85, K101[70]	N/A	Ref. [70; 381; 384]
Cu^{2+} ions	Antioxidant activity regulation, copper homeostasis, calcium signaling, in the progression of the disease[382]	In Hu/Mo/SHa PrP^{C} OR: H61, H69, H77, H85	Inter-repeat mode: nanomolar range; Single-repeat mode: low-micromolar range	Reviewed in ref. [382; 383]
		In Hu/Mo/SHa PrP^{C} outside OR[a]: H96, H111	Nanomolar range	
Vitronectin	Supporting axonal growth[103]	In MoPrP^{C}: residues 105–119[103]	12 nM[103]	Ref. [103]
STI1	Triggering neuroprotection[105]	In MoPrP^{C}: residues 115–128[105]	140 nM[105]	Ref. [105]
Aβ multimers	Suppression of long-term potentiation[106; 107; 108]	In MoPrP^{C}: residues 96–111[106; 107; 108]	N/A	Ref. [106; 107; 108]
LRP1	In a clathrin-dependent internalization mechanism[109]	N/A	Nanomolar range[109]	Ref. [109]
NCAM	In the cell adhesion process[110]	N/A	N/A	Ref. [32; 110]

[a] A few studies using N-terminally truncated murine PrP^{C} GD have also detected a weak binding site in the GD with affinity ~10 μM [91; 92].

On these premises, here we make the plausible hypothesis that the PMs in the N-term of the HuPrP cause prion diseases by (i) modifying the conformational properties of the STE/TM1 region and its interactions with trans-acting factors in the ER membrane and in the cytosol[385], and/or (ii) by affecting the interactions of N-term_HuPrP with chemicals and/or with protein partners *in vivo*. We tested this hypothesis by biocomputing methods. We focused on MoPrP^{C} because of its high sequence identity (particularly in the N-term) to HuPrP^{C}, and because experiments of protein binding have been so far reported mainly for this species. Our calculations[p] predict the conformational ensemble of WT N-term_MoPrP and that of N-term_MoPrP containing the 7 Hu PMs using replica-exchange (RE) Monte Carlo (MC) simulations[227] based on an implicit-solvent

---

[p] We do not consider: i) the PMs present in MoPrP but not in HuPrP and artificial mutations[386; 387] because their effect in humans is currently unknown; ii) the PMs in the signal peptides which are cleaved off during the biosynthesis of the protein[388]; iii) the interactions with metal ions other than Cu^{2+} (e.g. Zn^{2+}) because their relevance *in vivo* has not been firmly established[382].

all-atom potential[18]. The latter has been shown to be effective in investigating the folding of structurally diverse sets of proteins, as well as in studying the conformational properties of naturally unfolded proteins[19; 220; 389].

We found that N-term_MoPrP is composed of several regions characterized by different SS, in agreement with biophysical data[56; 390; 391; 392; 393]. The PMs do not change dramatically the conformational ensemble of N-term_MoPrP. Rather, they affect in a subtle manner the intra-molecular contacts in regions responsible for interactions with cellular partners, as well as the local structural features of STE/TM1. These in turn may alter the protein function[14] and biosynthesis[38].

6.2 Computational details

6.2.1 Bioinformatics

A structural model of the full-length MoPrP^C was first constructed by homology modeling. HHpred web server[394] was used to select homologous templates (Table 6.2) from structured proteins in RCSB Protein Data Bank (www.pdb.org). These templates were provided to MODELLER 9.10[395] for multi-template homology modeling. Firstly, 100 preliminary models were constructed. The best model was then selected which had the best DOPE score (calculated in Modeller 9.10) and high stereochemical quality (analyzed using the PROCHECK program[396]). The full-length model was truncated at residue 128 to obtain the N-term_MoPrP. The Swiss-PdbViewer (DeepView) 4.0[293] was used to introduce the 7 mutations (Q52P, P102L, P105L/T/S, G114V and A117V) into the N-term_MoPrP for generating initial models of the 7 PMs. These 8 models were used as educated guesses for the subsequent molecular simulations and by no means represented a structural prediction.

The bioinformatics tool PONDR-FIT[397] was used to predict the intrinsic disorder disposition of N-term_MoPrP. SS prediction tools including PSIPRED[398], Jpred3[399], Porter[400], RaptorX[401] and CABS-fold[402] were also used to predict the SS in this domain.

Table 6.2: Structure templates for homology modeling of the full-length MoPrP.

Protein name (residues)	Structure type	PDB entry (chain)	Fragment used as template	Target residues in MoPrP^{C}	Sequence identity to target (%)	References
MoPrP (121-232)	NMR	1XYX (A)	121-128	120-127	100	Ref. [341]
HuPrP octapeptide (61-68)	NMR	1OEH (A)	61-68	52-59, 60-67, 68-75, 76-83, 84-91	88	Ref. [56]
HuPrP pathogenic mutant Q212P (90-231)	NMR	2KUN (A)	90-121	89-120	88	Ref. [403]
HuPrP pathogenic mutant V210I (90-231)	NMR	2LEJ (A)	85-121	84-120	92	Ref. [148]
The Thermus thermophilus protein TTHA0988	X-ray	3OEP (A)	163-169	27-32	83	Ref. [404]
Bovine angiogenin	X-ray	1AGI (A)	85-91	33-39	57	Ref. [405]
Hu Ribonuclease 1, R31L, R32L mutant	X-ray	2E0J (A)	87-93	33-39	86	Ref. [406]
The malonyl coenzyme A-acyl carrier protein transacylase	X-ray	2QC3 (A)	6-11	38-43	83	Ref. [407]
Phycocyanin	X-ray	2VML (B)	69-75	44-50	86	N/A

6.2.2 Molecular simulations

REMC simulations[227] using the PROFASI (PROtein Folding and Aggregation Simulator) code[18] were carried out for WT N-term_MoPrP and each of the 7 PMs. PROFASI uses an all-atom model for the protein including hydrogen atoms, and implicit water solvent. The model assumes fixed bond lengths, bond angles, and peptide torsion angles (180°), so that each amino acid has the Ramachandran dihedral angles ϕ and ψ as well as a number of side-chain torsion angles as its degrees of freedom. The interaction potential is composed of four terms: $E = E_{loc} + E_{ev} + E_{hb} + E_{sc}$. The term E_{loc} is a local potential which accounts for interactions between atoms separated by a few covalent bonds, such as the electrostatic interaction between adjacent peptide units along the chain. The other three terms are nonlocal in sequence. The excluded volume term E_{ev} is a $1/r^{12}$ repulsion between pairs of atoms. E_{hb} represents two kinds of hydrogen bonds: backbone-backbone bonds, and bonds between charged side chains and the backbone. The last term E_{sc} represents simple pairwise additive approximations for hydrophobic attraction between nonpolar side chains and electrostatic interactions among charged side-chains. The interaction potential assumes the presence of a solvent environment for the proteins although solvent molecules are not explicitly represented in the simulations. To explore the conformation space, we used a Markov Chain MC procedure. At any step, a conformation update involving a change in one or more backbone and side-chain torsion angles is proposed, and is either accepted or rejected using a Metropolis[219] criterion. A semi-local backbone update, called Biased Gaussian Steps[224] producing smooth local deformations of the chains was used at the lower temperatures to improve the sampling of compact structures.

In each case studied here, the system was enclosed in a periodic cubic box of size (300 Å)^3. We used a replica exchange (also known as parallel tempering) procedure[408] to enhance the sampling in our simulations. Simultaneous MC simulations were performed at 8 temperatures: 293.0 K, 299.7 K, 306.5 K, 313.5 K, 320.6 K, 327.9 K, 335.4 K and 343.0 K. Periodically, conformations of replicas at consecutive temperatures were swapped with a probability $P_{swap}^{ij} = \exp\left(\left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right)\left(E_i - E_j\right)\right)$, where T_k is the temperature and E_k is the energy of replica k, and k_B is the Boltzmann constant. The exchange procedure improves sampling at the lower temperatures by stochastically seeding them with independent states obtained at high temperatures, and the above exchange probability ensures equillibrium is maintained at each temperature even for the replica exchange step.

In order to assess the reliability of the statistics obtained in these simulations, we examined the random walk of the replicas in temperature space (see Fig. A3.1 in Appendix 3). Since, at the higher temperatures the conformation of the system decorrelates fast by jumping over energy barriers, the statistics collected at the lower temperatures consists of uncorrelated samples, if the replica has visited a high temperature in between. The histogram of energy at different temperatures (see Fig. A3.2 in Appendix 3) demonstrates a massive overlap between each consecutive pair of temperatures. This shows that there were no sharp transitions in temperature in our system. The length of the simulations was chosen to allow our replicas to reach both ends of the temperature range repeatedly, so that our histograms received data from uncorrelated sample conformations from different parts of the energy landscape. We calculated the cumulative averages[409] of the the radius of gyration (R_g) and of the SS contents over the sampled trajectory. The cumulative average was defined by $\sum_{i=1}^{N} X^i/N$, where N was the total number of frames sampled, i was the frame number and X^i was the property of interest (the R_g or the SS contents) calculated for each frame. The cumulative averages appeared to converge rather than systematically drifting (see Fig. A3.3 in Appendix 3). This combined with the fact that the simulations continued to explore uncorrelated states by moving back and forth across temperatures gives us reasonable confidence that our simulations have captured the statistical properties of these systems to the accuracy of the force field. In order to further reduce the effect of auto-correlation in our statistics, the conformations were sampled every 10^4 cycles (3.7 10^6 MC steps). We performed 1.8×10^8 cycles (or 6.7×10^{10} MC steps) of simulations for each system and the first 10% of each run was discarded for thermalization. Production trajectories obtained at 293K were extracted, which yielded 2000 frames per protein for analysis.

6.2.3 Data analyses

The following properties were calculated with the production trajectory obtained for each N-term_MoPrP variant. i) The average R_g was calculated using the g_gyrate module in Gromacs 4.5[338]. ii) The SS in each frame, as well as the average SS content in the trajectory, was calculated with the DSSP software[340]. iii) The intra-molecular contact map for each frame, which was a distance matrix consisting of the smallest distance within 6 Å between residue pairs, was calculated using the g_mdmat module in Gromacs 4.5. The total number of contacts per residue was also calculated over the entire production trajectory. Every time two residues approached each other within 6-Å distance, this was counted as one contact for each residue. iv) The backbone flexibility during the trajectory was calculated using the *angular dispersion index* (*ADI*) analysis[305] described in Chapter 4, section 4.2.3.

A clustering analysis was carried out on the contact map data in order to characterize the intra-molecular contacts in each trajectory. The contact map of each frame was converted into an array of length $M \times M$ by concatenating matrix rows, where M was the number of residues in the protein. Distances between frames were computed using the Euclidean distance (Eq. (6.1)) as metric and stored in a proximity matrix. This distance indicated the similarity between two frames regarding their intra-molecular contacts: the shorter the distance, the more similar the two frames. Finally, a hierarchical clustering[410] was performed by using an average linkage method[411]. In this method, the distance between two clusters is the average distance between pairs of observations, one in each cluster. Main clusters were obtained using a distance cutoff on the hierarchical tree so that the 10 most populated clusters cover about 70% of the entire data set. An average contact map in each cluster was computed to represent the intra-molecular contacts in the cluster members.

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \ldots (p_n - q_n)^2}. \tag{6.1}$$

6.3 Results

A total of 5.4×10^{11} MC steps of simulations were performed on bioinformatics-based structural models of N-term_MoPrP and the 7 PMs (6.7×10^{10} MC steps for each protein). The number of MC steps performed is expected to be sufficient for the naturally unfolded proteins studied here. The simulations obtained at 293 K are chosen for analyses in this work because most of the *in vitro* experiments with which we compare our results have been carried out at room temperature[56; 390; 391; 392; 393]. We also calculated the SS contents (in %) for the wild-type N-term_MoPrP^C simulated at 306.5 K and 313.5 K in this work. These are the nearest to the physiological temperature. As it might be expected, the α-helix content decreases and the coil structure increases with respect to the results of 293 K (see Table A3.1 in Appendix 3). However, the changes are in the range of the standard deviation along the simulations. Hereafter we describe the structural determinants of the resulting proteins' conformational ensembles at 293 K.

N-term_MoPrP WT. The equilibrated trajectory of N-term_MoPrP provides an ensemble of conformations containing transient SS elements (Fig. 6.2A) with R_g of 18.5±2.9 Å. On average, it contains 19±8% α-helix, 8±5% β-sheet, 7±3% β-bridge, 27±5% β-turn, 12±4% bend, 4±3% 3_{10}-helix and 1±1% π-helix. However, different regions appear to adopt different SS. According to the occurrence of diverse SS, N-term_MoPrP can be roughly divided into 6 sub-regions (Table 6.3). Notably at residues 99-117 and 118-125 we observe relatively high occurrence of α-helix (64±9% and 33±8%, respectively), which is in line with the prediction by PONDR-FIT that this region has relatively low intrinsic disorder disposition (Fig. A3.4 in Appendix 3). Consistently, commonly used bioinformatic tools for SS prediction (such as PSIPRED, Jpred, Porter, Raporx and CABS-fold) also indicate that residues ~104-120 have the propensity to form α-helix (Fig. A3.4).

Our results are consistent with the finding by solution NMR that the OR peptide comprising residues 61-68 in N-term_HuPrP acquires a stable loop/β-turn conformation at pH ~6[393]. In our simulations indeed we find the loop/β-turn conformations in the OR region resembling the NMR structure. These conformations show backbone RMSD of less than 2.5 Å from the NMR structure, which composes 8%, 11%, 3%, 9% and 13% of the five sequential OR peptide in the OR region, respectively. We also find antiparallel β-sheets/β-bridges formed between adjacent OR peptides.

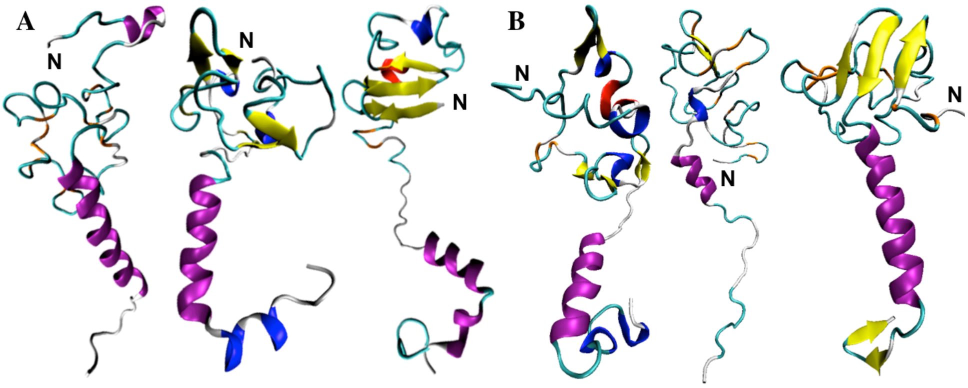


Figure 6.2: Selected conformations of **A)** WT N-term_MoPrP and **B)** one PM (N-term_MoPrP_Q52P) from our simulations. These contain transient α-helix (in violet), β-sheet (yellow), β-bridge (orange), β-turn (cyan), 3_{10}-helix (blue) and π-helix (red) elements.

Table 6.3: Content of transient SS elements (in %[i]) in each region of N-term_MoPrP averaged over the trajectory. According to the occurrence of different SS, we identify 6 sub-regions: i) residues 23-30, mainly in coil/β-turn/bend conformation; ii) residues 31-50, mainly in β-turn/coil/bend/β-bridge conformations; iii) residues 51-91 (OR) mainly in β-turn/coil/bend/β-sheet conformations; iv) residues 90-99 mainly in coil/β-turn/bend/β-sheet conformations; v) residues 100-118 which contain the highest percentage of α-helix among all 6 sub-regions and vi) residues 119-126 which display a comparable percentage of α-helix and β-turn.

Residues	α-helix	β-sheet	β-bridge	β-turn	Bend	3_{10}-helix	π-helix	Coil
23-128	19±8	8±5	7±3	27±5	12±4	4±3	1±1	25±5
23-30	3±2	3±3	6±2	22±5	10±4	2±2	0±1	55±9
31-50	7±4	8±5	12±4	33±6	13±4	5±3	0±1	22±5
51-91 (OR)	6±2	12±5	9±4	36±7	13±5	6±3	1±1	18±4
90-99	5±2	12±5	5±3	23±5	16±4	4±2	0±2	35±7
100-118	64±9	1±2	0±1	7±3	8±3	2±2	0±1	16±6
119-126	33±8	3±6	2±3	33±5	9±3	5±3	1±1	15±4

[i] Standard deviations (SD) are indicated by "±".

In residues 106–126 (known as the amyloidogenic region), we observe transient helical conformation spanning all over this region. However, the first part (residues 106–118) shows mainly α-helix motif, while residues 119–126 display a comparable percentage of α-helix and β-turn. These results are consistent with the reported CD, NMR and FTIR studies on $HuPrP^C$ fragments[390; 391; 392]. Together, these studies indicate that the equivalent region in $HuPrP^C$ may adopt α-helix, β-sheet or coil structures under different experimental conditions[390; 391; 392]. We notice that this region roughly corresponds to the STE/TM1 in N-term_MoPrP, which comprises residues 104–128.

N-term_MoPrP PMs. The R_g of the 7 PMs studied here is comparable to that of the WT (Table 6.4). This indicates that the mutations do not alter the compactness of the N-term_MoPrP. The average SS contents in the PMs are also mostly similar to those of the WT (Table 6.4, Fig. 6.2B).

Differences between the WT and PMs are mainly in the β-sheet and α-helix contents in residues 100–128 (Table 6.4). Among these, N-term_MoPrP_P102L has 11% more β-sheet content and 11% less α-helix content in this region than those in the WT. This result is consistent with the *in vitro* data of $MoPrP^C$ containing the P102L mutation[412]. N-term_MoPrP_P105L/T/S mutations lead to an increase in α-helix content by 9%, 11% and 8%, respectively. The PMs feature different backbone's conformational flexibility (calculated here with the T-PAD analysis[305], Fig. A3.2) from that of N-term_MoPrP WT. In particular, the flexibility of STE/TM1 increases in N-term_MoPrP_Q52P, N-term_MoPrP_P102L and N-term_MoPrP_A117V. It decreases in N-term_MoPrP_G114V and N-term_MoPrP_P105L/T/S with respect to the WT (Fig. 6.3).

Table 6.4: R_g and content of transient SS elements (in %[i]) of the entire systems studied (top) and of residues 100-128 (bottom).

Variant	R_g (Å)	α-helix	β-sheet	β-bridge	Bend	β-turn	3_{10}-helix	π-helix	Coil
	N-term								
N-term_MoPrP	18.5±2.9	19±8	8±5	7±3	12±4	27±5	4±3	1±1	24±5
N-term_MoPrP _Q52P	17.1±2.5	17±8	14±7	7±3	12±4	29±5	3±3	1±2	22±6
N-term_MoPrP _P102L	19.6±3.0	16±10	12±10	7±3	11±4	29±5	4±3	0±1	24±6
N-term_MoPrP _P105L	19.3±2.9	23±8	9±7	7±3	11±4	29±5	4±3	0±1	22±6
N-term_MoPrP _P105S	20.1±2.9	21±7	8±5	7±5	11±3	27±5	5±4	0±1	23±5
N-term_MoPrP _P105T	19.7±3.0	22±8	8±7	7±3	11±4	28±5	4±4	0±1	24±6
N-term_MoPrP _G114V	18.7±2.9	20±7	9±5	7±7	12±3	28±4	4±3	0±1	25±5
N-term_MoPrP _A117V	18.8±2.9	17±8	10±7	7±3	11±4	29±5	4±3	0±1	25±5
	Residues 100-128								
N-term_MoPrP	12.3±1.4	52±10	2±2	1±1	8±5	15±4	3±3	1±1	20±4
N-term_MoPrP _Q52P	12.2±1.4	48±10	4±4	2±2	9±4	15±5	3±3	0±1	19±5
N-term_MoPrP _P102L	11.9±1.2	41±11	13±4	1±2	6±4	17±5	3±3	0±1	19±4
N-term_MoPrP _P105L	12.2±1.4	61±12	3±2	1±1	4±4	14±5	2±3	0±1	14±5
N-term_MoPrP _P105S	12.4±1.2	63±13	2±2	1±2	4±4	14±4	2±2	0±1	14±4
N-term_MoPrP _P105T	12.3±1.3	60±11	3±2	1±1	14±6	14±4	2±3	0±1	15±4
N-term_MoPrP _G114V	11.9±1.2	53±10	2±3	1±2	7±4	14±6	2±2	1±1	19±5
N-term_MoPrP _A117V	11.8±1.4	46±10	4±3	1±1	8±5	16±4	3±3	0±1	21±5

[i] SDs are indicated by "±".

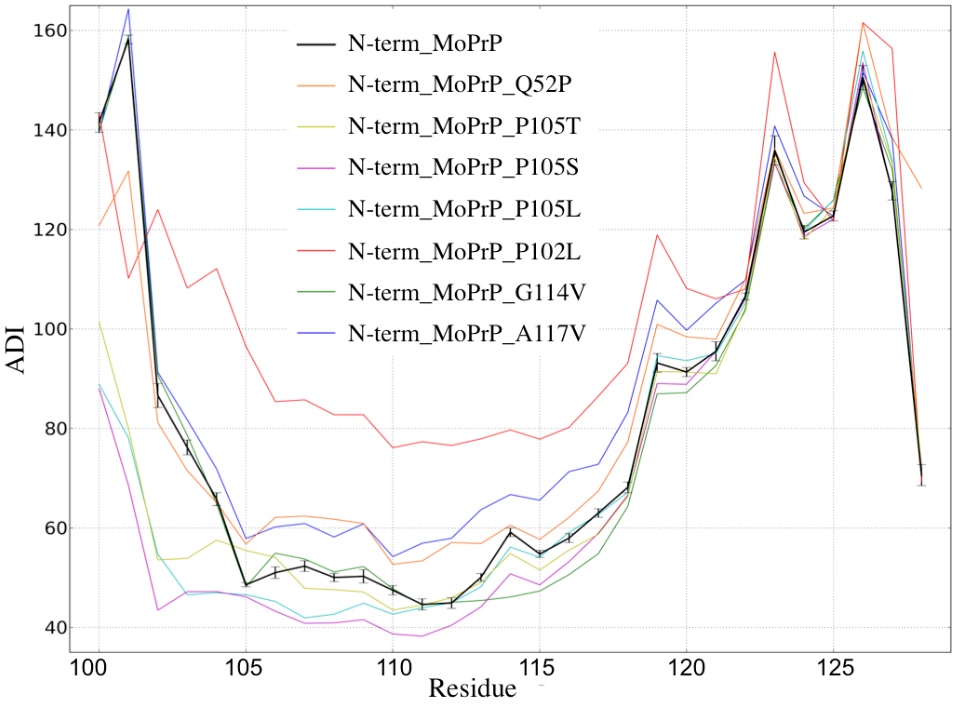


Figure 6.3: Flexibility of backbone units of residues 100-128. The higher the ADI value[305], the more flexible the backbone. The SDs in the 10 simulation runs of the WT N-term_MoPrP are indicated with vertical bars, whereas those of the PMs are shown in Appendix 3 Fig. A3.6.

6.4 Discussion

Our calculations suggest that N-term_MoPrP WT features a large ensemble of conformations containing transient SS, rather than completely disordered ones. The result is consistent with experimental data[56; 390; 391; 392; 393].

While many PMs in the GD are known to modify significantly the folded structure and to increase its flexibility[16; 147; 150; 156], our calculations suggest that those in the N-term do not impact significantly the global structural properties of the N-term. This is consistent with *in vitro* experiments, which have shown that PMs in N-term_HuPrP do not affect the thermostability or misfolding kinetics of the protein molecule[156]. However, the PMs at the N-term do modify local features from those of the WT *at the binding sites for known cellular partners*. This might affect the interactions with these partners and interfere with the related physiological functions.

Specifically, (i) the intra-molecular contacts of residues binding Cu^{2+} and sulphated GAG (Table 6.1) decrease on passing from the WT to the PMs, as shown by the total number of contacts per residue during the simulations of each system (Fig. 6.4). A cluster analysis on the intra-molecular contact maps of each system (described in “Materials and Methods”) result in ten main clusters that characterize the intra-molecular contacts during the simulations. These contact maps of the most populated cluster for each system (Fig. 6.5) demonstrate that the major differences between N-term_MoPrP WT and the PMs are in the residues binding Cu^{2+} and sulphated GAG (i.e. the OR region and the H111 Cu^{2+}-binding site (Table 6.1)). In the less populated clusters of contact maps (Appendix 3.1), minor differences between N-term_MoPrP WT and the PMs are observed also in other regions, which vary from one PM to another. These minor differences are not statistically significant, thus they are not discussed here. Alteration of the Cu^{2+}-PrP^C binding process has been suggested to play a role for Hu/Mo PrP^C aggregation and for the progression of the disease[382]. Interestingly, a recent study in our lab shows that the PM Q212P in the $HuPrP^C$ GD can influence the Cu^{2+} binding mode at H96 and H111[413], implicating a role of abnormal Cu^{2+} binding in the pathology of PMs in $HuPrP^C$. Sulphated GAG binding to PrP^C is likely involved in the formation of PrP^{Sc} *in vivo*, although its role remains to be established[381]. In addition, ii) the PMs affect the SS, the flexibility and increase the hydrophobicity of STE/TM1 (Table A3.2 in Appendix 3). The latter contains the putative binding sites for *in vivo* binding partner proteins such as vitronectin[103] and STI1[105]. This might affect the biological function of these interactions, which involves the signaling for axonal growth[103] and that for neuroprotection[105], respectively.

The PM Q52P in the OR region affects the flexibility of STE/TM1 (Fig. 6.3), while the other 6 PMs in STE/TM1 also alter the intra-molecular contacts in the OR region (Fig. 6.4 and 6.5). This suggests that the PMs may affect the transient interactions between the OR region and STE/TM1. However, these observed alternations on passing from the WT to the PMs in (i) the intra-molecular contacts and (ii) the backbone flexibility turn out not to be quantitatively correlated. The number of contacts takes into account the backbone and side-chain atoms and is related to the 3D fold of the protein conformations sampled in the simulations. The PAD index, instead, measures the fluctuations of the backbone dihedral angles of the residues during the simulations. Therefore, the local backbone flexibility might not be directly related to the protein's 3D fold.

Growing evidence point to the role of PrP^C in mediating the memory impairment by Aβ multimers *in vivo*[106; 107]. On the one hand, Aβ oligomers have been found to bind PrP^C N-term, which is required for the suppression of long-term potentiation by Aβ oligomers *in vivo*[106; 108]. On the other hand, *in vitro* experiments suggest a Cu^{2+}-dependent mechanism that links the neurotoxicity of Aβ oligomers to PrP^C[414]. The fact that $MoPrP^C$ residues 96-111 in the N-term are essential for the binding of both Aβ oligomers[106] and Cu^{2+} suggests the significance of this region in the yet undefined role of PrP^C in Alzheimer's disease. Our findings that PMs affect the structural features of this region are in line with the hypothesis that a Cu^{2+}-dependent mechanism may unify the pathogenesis of neurodegenerative diseases such as TSE, Alzheimer's disease and Parkinson's disease[415].

The altered local features in STE/TM1 might also impact on the protein's interactions with trans-acting factors in the cytosol and in the ER membrane[385]. This result is consistent with the *in vitro* data that PMs P102L, P105L and A117V increase the interactions between MoPrP$^{\mathrm{C}}$ STE/TM1 and a membrane mimetic at pH 7[416]. These changes may affect the relative proportions of the $^{\mathrm{Sec}}$PrP, $^{\mathrm{Ntm}}$PrP, $^{\mathrm{Ctm}}$PrP and $^{\mathrm{Cy}}$PrP topological variants in human and mice[38]. This in turn could have repercussions on the disease, as elevated $^{\mathrm{Ctm}}$PrP level has been related to both familial (in GSS patients) and infectious (in transgenic mice) prion diseases[43; 47].

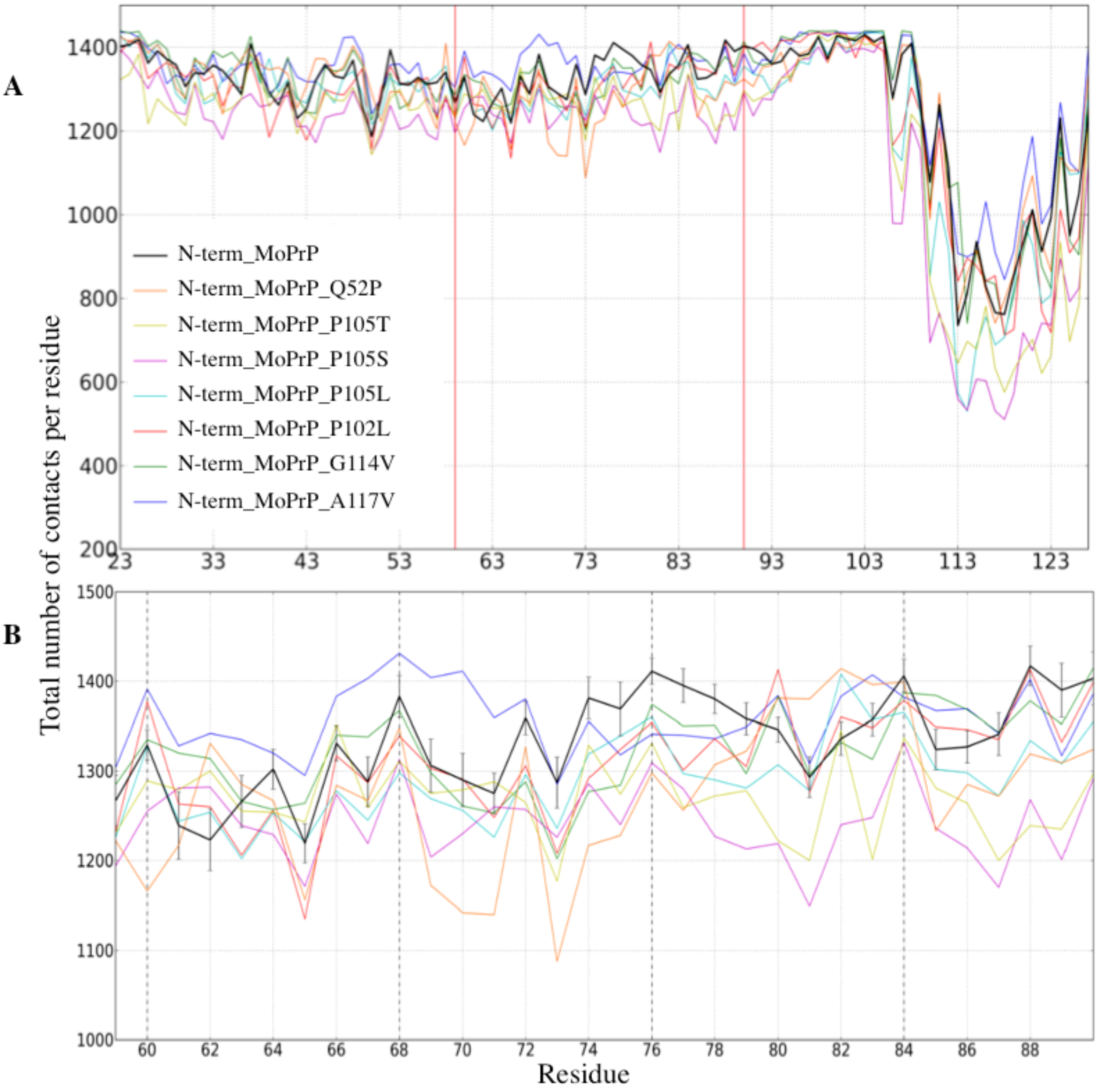


Figure 6.4: Total number of intra-molecular contacts per residue in the WT and mutated N-term_MoPrP. **A**) Total number of contacts between residues in the entire N-term over the trajectory. The SDs of the 10 simulation runs of each system are shown in Appendix 3 Fig. A3.7. **B**) Total number of contacts in the OR over the trajectory where the four histidine residues are indicated with vertical dashed lines. The SDs in the 10 simulation runs of the WT N-term_MoPrP are indicated with vertical bars, whereas those of the PMs are in the similar range (Fig. A3.7).

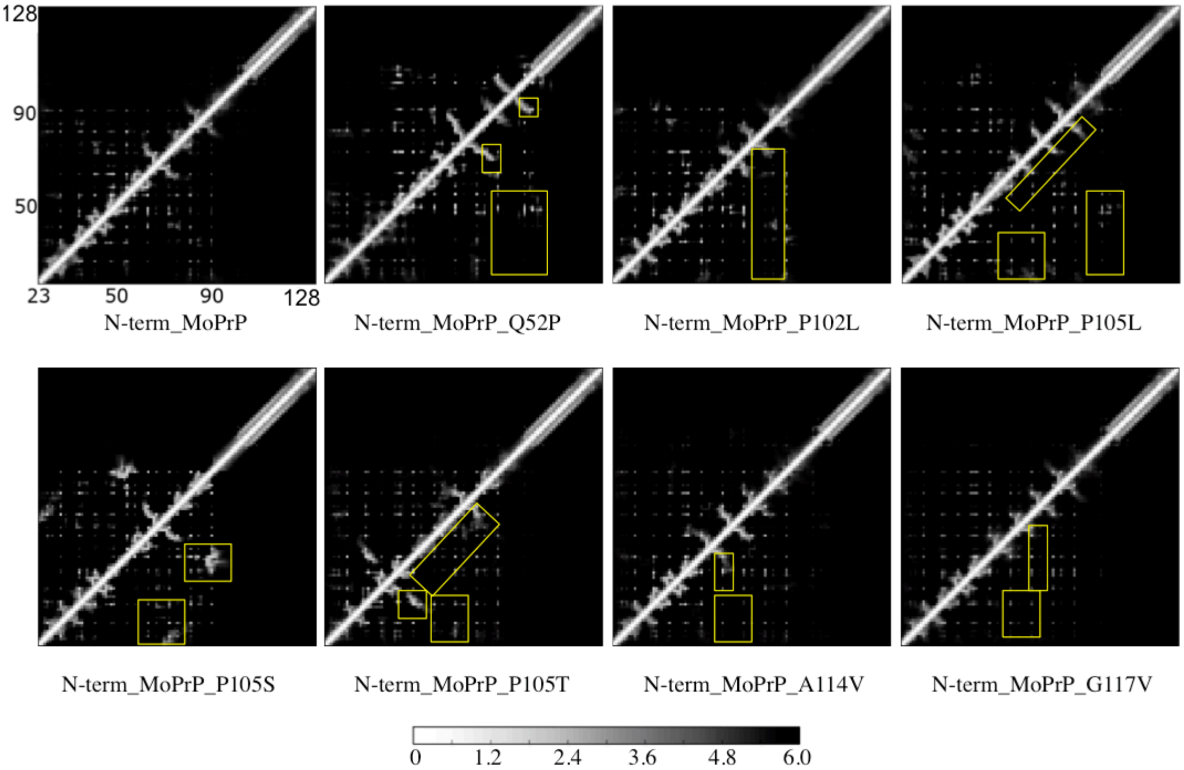


Figure 6.5: Contact maps illustrating the intra-molecular interactions in WT and mutated N-term_MoPrP. Contacts within 6 Å are plotted in gray scale according to the distances as shown at the bottom. Residue numbers are labeled for the N-term_MoPrP and are the same for all the PMs. For each protein, only the most populated cluster is shown (see Appendix 3.1 for the rest of the clusters and the detailed data of each cluster). Major differences between the WT and the PMs are highlighted in the lower half of the (diagonally symmetric) maps with yellow boxes.

Implications for PrP^{C}'s molecular recognition. Experimental structural information is available for several fragments of PrP^{C} N-term in complex with antibodies. These include: residues 119-122 from HuPrP^{C} in β-strand (Abskharon et al., unpublished data), residues 104–113 from SHaPrP (equivalent to HuPrP 104–113) in bend conformation[417], residues 107–115 from bovine (Bo) PrP[q] (equivalent to HuPrP 96–104) in bend conformation[418] and a MoPrP 2-OR peptide (residues 68–83) in an extended conformation[419]. In addition, herapin binds to N-term from SHaPrP and stabilizes a repeated loop/β-turn conformation encompassing 4 OR[70]. Interestingly, the former three are present in our calculated ensemble. This indicates a possible "conformational selection" mechanism[420] upon binding in these cases. The extended conformation of the MoPrP 2-OR peptide

---

[q] SHaPrP and BoPrP both share high (>85%) sequence identity with HuPrP and MoPrP (Table A3.3 in Appendix 3). The N-term and OR region are highly conserved in these PrP species (Table A3.3).

is embedded in a groove of its cognate antibody (named POM2 Fab)[419] which prevents the formation of the loop/β-turn seen by NMR[393] and by our simulation. Similarly, the 4-OR conformation in N-term from SHaPrP bound with herapin is absent in our simulation, although, as discussed in the Results section, the conformation of each single OR is similar to that of the NMR structure[393] and it is present in our simulation. Hence, we suggest that the binding of the POM2 Fab antibody and herapin to the OR involves an "induced fit" mechanism[420], since the resulting OR conformations are not formed before binding.

Our test simulations on full-length WT MoPrP^C (discussed below), the above-mentioned β-strand structure in residues 119-122 (denoted β_0, hereinafter) extends the two-strand β-sheet in the GD into a three-strand antiparallel β-sheet. Such extension has been reported previously in the crystal structure of HuPrP^C bound with an antibody (Abskharon et al., unpublished data) and in the all-atom MD simulations on truncated SHaPrP^C at acidic pH[348] or upon the G131V or M129V mutations in truncated SHaPrP^C[345]. Our simulations on full-length WT MoPrP^C and N-term_MoPrP show that β_0 can be formed with or without the GD (Table 6.3 and Table A3.4 in Appendix 3), suggesting the propensity of residues 119-126 to form transient β-strand. The occurrence of β_0 along with other β-sheet structures within residues 100-128 (Table 6.4) is higher in the P102L mutant than that in the other mutants and the WT, which appears to be accompanied by a decrease of α-helix in this mutant, consistent with the *in vitro* data of full-length MoPrP^C_P10L[412]. This was not observed in the all-atom MD simulations of a truncated SHaPrP^C (90-231) containing this PM[346], in which, however, the authors pointed out that the discrepancy might be explained by the relatively short timescale (4-8 ns) of their MD simulations.

Limitations. As any modeling study, this work has limitations. Firstly, we did not include the GD in the calculations. Test calculations on full-length WT MoPrP show that its N-term conformational ensemble exhibits very similar features to those of the N-term_MoPrP (Table A3.4 in Appendix 3). On the one hand, this suggests that the absence of GD does not notably influence the conformational ensemble of the N-term. On the other, simulating the full-length MoPrP (containing 208 residues) for all of the systems investigated here might face issues of convergence and limitations of the accuracy of the force field used. The latter is calibrated for systems much smaller than the full-length MoPrP[18; 421; 422]. Secondly, without explicit water molecules, some potentially important questions cannot be addressed, such as the effect of ionic strength. Here, however, we stress that the full agreement between available experimental data and our findings does validate the computational approach, which has been successfully applied to a variety of intrinsically disordered proteins[19; 389].

6.5 Conclusion

The seven PMs do not dramatically impact the conformational ensemble. However, they do alter the local conformational features in STE/TM1, which are in control of PrP topology during its biosynthesis[38; 43]. This region also contains the estimated binding sites for cellular partners such as vitronectin and STI1[103; 105]. Furthermore, the PMs specifically affect the intra-molecular contacts in the OR at the putative binding sites for Cu^{2+}[383] and sulphated GAG[70], as well as in the Cu^{2+}-binding site around residue H111[90] (Table 6.1). Hence, unlike most of the PMs in the GD, which affect HuPrP thermostability and misfolding kinetics[156], the PMs at the N-term induce local structural modifications in key regions for HuPrP biosynthesis[38; 43], for metal binding[383] and for interactions with other cellular partners[14].

Chapter 7 Conclusions

Pathogenic mutations in the PrP gene (*PRNP* in humans or *Prnp* in other mammalian species) are associated with spontaneous generation of TSE[328; 329; 330]. Polymorphisms (i.e. non-pathogenic, naturally occurring mutations) in the PrP gene are found to influence the etiology and neuropathology of the disease both in humans[8] and in sheep[9]. In transgenic mice, artificial mutations could determine the susceptibility to the infection of different prion strains[10; 11; 12]. *In vitro*, chimeric MoPrP variants (denoted MoPrP chimeras) which contain artificial mutations react very differently to PrP^{Sc} infection[13]: some were resistant to the PrP^{C}→PrP^{Sc} conversion and others were not; some of the resistant mutants even showed dominant-negative effect to protect the endogenous WT MoPrP^{C} from the conversion.

The intriguing roles of these various mutations likely lie in their impact on the molecular structure of PrP^{C}, which affects the PrP^{C}→PrP^{Sc} conversion and the disease development. Understanding the underlining molecular mechanism can revolutionize the research on many other neurodegenerative diseases that are major concerns for public health. Structural biology approaches such as NMR spectroscopy, although very powerful in studying the structures of PrP variants, are relatively slow, and this hampers the progress in probing the roles of different mutations. Moreover, experimental methods currently face immense difficulties in studying the intrinsically disordered N-term of PrP. In this thesis, we addressed these challenging issues by using extensive molecular simulation approaches. We investigated the structural features of more than 37 PrP^{C} variants, and characterized the impact of pathogenic mutations, polymorphisms and artificial mutations on the structural determinants in PrP^{C}.

First, we investigated the pathogenic missense mutations (point mutations coding for single-amino acid substitutions) in the HuPrP^{C} GD, aiming to identify their impact on the folded structure of HuPrP^{C}. MD simulations were carried out on 30 variants of HuPrP^{C} GD containing pathogenic missense mutations (Fig. 4.1 in Chapter 4), using our well-established MD protocol[17] that had been validated against available NMR structures of HuPrP^{C} WT and mutants. Remarkably, despite the diverse location and chemical nature of the mutations, all the mutants turned out to feature common structural traits in the GD, distinct from those of the WT and of the protective polymorphism. These common traits (relative to the WT) include: i) increased flexibility of the α2-α3 region due to a loss of specific intramolecular interactions in this region and between this region and the α1 helix; ii) weakened intramolecular contacts in the β2-α2 loop and between the β2-α2 loop and the α3 helix; iii) larger solvent exposure of residue Y169 in the β2-α2 loop, especially in the mutants with high worldwide incidence. The findings suggest that these regions are the hot spots in HuPrP^{C} structure vulnerable to the destabilizing effect of mutations. Altered structural features in these spots may be involved in the early stage of the $PrP^C \rightarrow PrP^{Sc}$ conversion. Consistently, a number of other studies suggest that the $\alpha 2$-$\alpha 3$ region is metastable under acidic pH[256] or high pressure[255], and it has high tendency to form oligomers[352], amyloids[351] or fibrils[349; 350]. The $\beta 2$-$\alpha 2$ loop and its long-range interactions with the C-terminal of $\alpha 3$ helix are likely responsible for the transmission barrier across PrP^{Sc} strains, as indicated by *in vitro* and *in vivo* studies[10; 11; 137; 138; 141; 142]. Therefore, the $\beta 2$-$\alpha 2$ loop and the C-terminal of $\alpha 3$ helix may form an epitope for molecular recognition by PrP^{Sc} and/or other interaction partners.

Encouraged by the above findings, we have then applied the same approach to study the GD structure of a group of 11 MoPrP chimeras. Our findings are as follows. i) The specific interactions between the N-terminal of $\alpha 3$ helix and the $\alpha 1$ helix are stronger in the resistant chimeras than those in the non-resistant chimeras and in the WT $MoPrP^C$. This suggests that strong intramolecular interactions in this region are crucial for maintaining the folded GD structure and preventing the conformational conversion upon PrP^{Sc} infection. This is consistent with the finding in our previous study that human pathogenic mutations remarkably deteriorate the intramolecular interactions in this region and increase the flexibility of the $\alpha 2$-$\alpha 3$. ii) The chimeras show exchanges between different secondary structures in the $\beta 2$-$\alpha 2$ loop, which can be categorized into two distinct patterns, a 3_{10}-helix/turn pattern like in $MoPrP^C$ and a bend/turn pattern like in $HuPrP^C$. All the non-resistant chimeras adopt the Mo-like pattern, same as the dominant-negative chimeras. The other resistant (non-dominant-negative) chimeras show the Hu-like pattern. Thus, the dominant-negative effect occurs when the resistant chimeras adopt the Mo-like pattern in the $\beta 2$-$\alpha 2$ loop conformation. These indicate that the $PrP^C \rightarrow PrP^{Sc}$ conversion process requires specific conformation in the PrP^C $\beta 2$-$\alpha 2$ loop that is compatible with the PrP^{Sc} strain or with some unknown factor. This in turn allows the dominant-negative chimeras to interfere with the conversion of the endogenous $MoPrP^C$ into PrP^{Sc}. This agrees with our previous study, which suggests that the $\beta 2$-$\alpha 2$ loop serves as (part of) a recognition site in the disease development.

Lastly, we investigated the impact of all the 7 Hu TSE-linked missense mutations on the structure of the disordered N-term of MoPrP. Using replica-exchange MC simulations and an effective all-atom potential for proteins, we were able to obtain the large conformational ensemble of the N-term of WT MoPrP and that of the 7 mutants. The N-term turns out to feature an ensemble of transient secondary structures, instead of completely disordered conformations. The mutations in the N-term do not alter dramatically the conformational ensemble. Unlike the missense mutations in the GD that have global impact on the GD structure, these in the N-term cause subtle, local structural changes. The changes are mainly in the regions that control PrP topology during its biosynthesis and in the residues involved in binding with putative cellular partners. Therefore, these mutations may not affect the thermostability of PrP^C, which is consistent with *in vitro* experiments[156]. Rather, they may cause the diseases by leading to detrimental amount of atypical topological species of the protein and/or malfunction of the protein and its partners.

In conclusion, we have shown that biocomputing is instrumental to provide hints to possible pathogenic mechanisms of all reported mutations in prion diseases.

Appendices

Appendix 1 Supplementary materials for Chapter 4

Appendix 1.1 Test simulations on the histidine protonation state

We performed 10-ns test MD simulations based on AMBER99 force field[423; 424] force field for HuPrP(WT) in which either H140, or H155, or H177 or H187 side chain was in the charged state (denoted HuPrP(WT_140HIP), HuPrP(WT_155HIP), HuPrP(WT_177HIP) and HuPrP(WT_187HIP), respectively). We used the same computational protocol as described in Chapter 4, section 4.2.2. Comparison was made with the first 10 ns of MD of HuPrP(WT), in which the four histidines are in the neutral state and they are protonated at the N_ε atom.

The RMSD of HuPrP(WT) turned out to be slightly smaller than that of HuPrP(WT_140HIP), HuPrP(WT_155HIP), HuPrP(WT_177HIP) and HuPrP(WT_187HIP), for most of the dynamics (Fig. A1.1.1). A comparison between initial and final MD structures suggests that the presence of any positively charged histidine residues affects the conformation of nearby segments as well as that of other regions of the proteins (Fig. A1.1.2). Consistently, other MD simulations of HuPrP(WT) using the GROMOS96 53 A6 force field[425] have shown that the fluctuations of HuPrP(WT) along the dynamics are smaller when all the histidines are at their neutral state than when they are positively charged[294].

We conclude that the presence of the positively charged histidines causes instability in the structure of HuPrP GD.

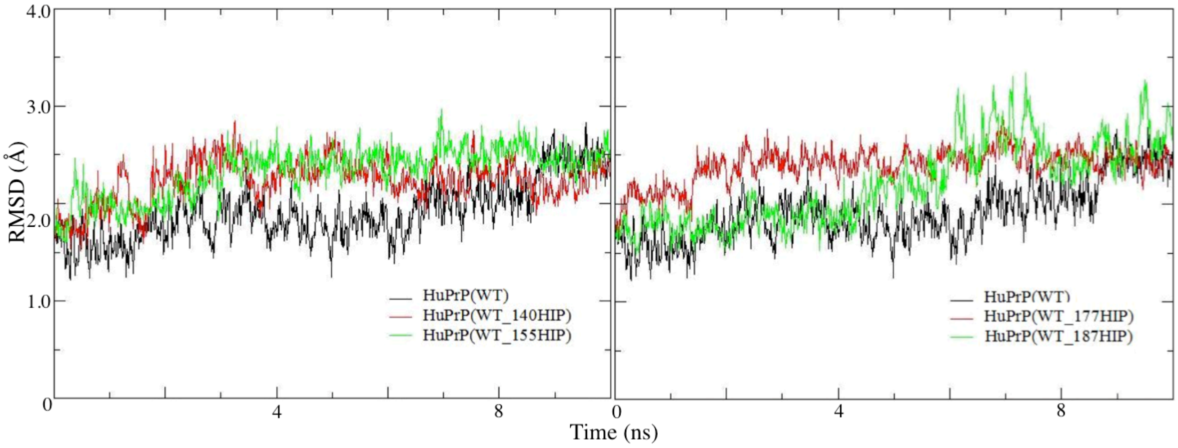


Figure A1.1.1: RMSD of C_α atoms of HuPrP(WT), HuPrP(WT_140HIP), HuPrP(WT_155HIP), HuPrP(WT_177HIP) and HuPrP(WT_187HIP) with respect to the initial conformations, plotted as a function of simulation time.

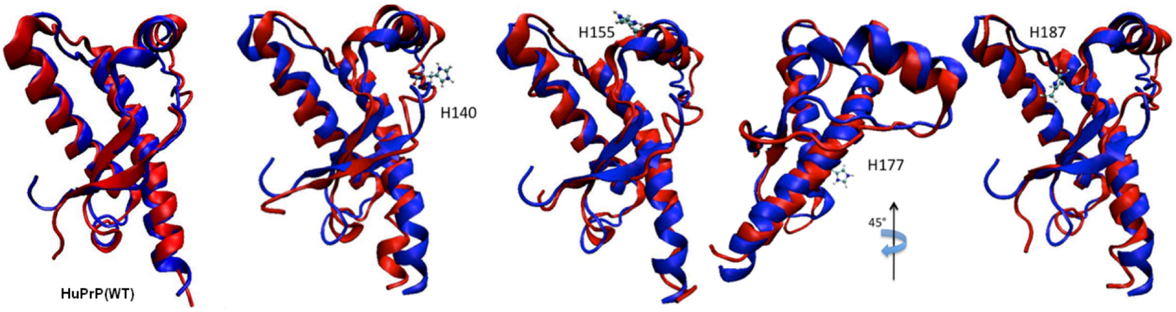


Figure A1.1.2: Superimposition between initial (blue) and final (red) MD structures of HuPrP(WT), HuPrP(WT_140HIP), HuPrP(WT_155HIP), HuPrP(WT_177HIP) and HuPrP(WT_187HIP). The proteins are represented in cartoons. Charged histidine residues are shown in ball-and-sticks.

Appendix 1.2 ADI versus RMSF

In Chapter 4, section 4.2.3, we state that ADI has three advantages over RMSF. In the following paragraph we discuss each of these points.

i) *ADI needs no fitting of the different configurations along the trajectory to a reference structure*. Two different conformations of the same protein, sampled with MD, may be compared: a) by using atomic coordinates (for example by calculating their RMSF), b) by using internal coordinates, specifically the dihedral angles of each residue.

By using atomic coordinates it is necessary to eliminate (or minimize) the effects of rigid body movements of one conformation with respect to the other (roto-translations), in order to highlight only the effects due to protein rearrangements. This is usually accomplished by fitting one configuration over the other. By using dihedrals angles, instead, roto-translations effects are intrinsically removed. Indeed the dihedral angles ϕ and ψ (involving respectively the (C, N, C_α, C) and (N, C_α, C, N) atoms) depend on the position of the last atom with respect to the other three in each set of four atoms. Thus they allow to define univocally the residues' internal coordinates.

ii) *ADI does not suffer for edge effects*. The RMSF calculation is preceded by a fitting of the different protein configurations. However this is not a straightforward operation, especially when the flexibility of the protein is relevant. The standard fitting procedures tend indeed to minimize the overlap of most of the protein atoms, leading to a systematic increase of the deviations of the atoms in the terminus regions, usually more flexible. So the RMSF exhibits a steady increase for residues belonging to these regions. By using the dihedral angles to calculate the residue-by-residue flexibility this issue can be overcome: the fitting procedure is avoided and the movements of each residue are treated independently from those of the neighboring residues. This can be appreciated by analyzing Figure A1.2.1, where the residue-by-residue RMSF (in Å) and ADI calculated from MD simulation of HuPrP(PP) are reported.

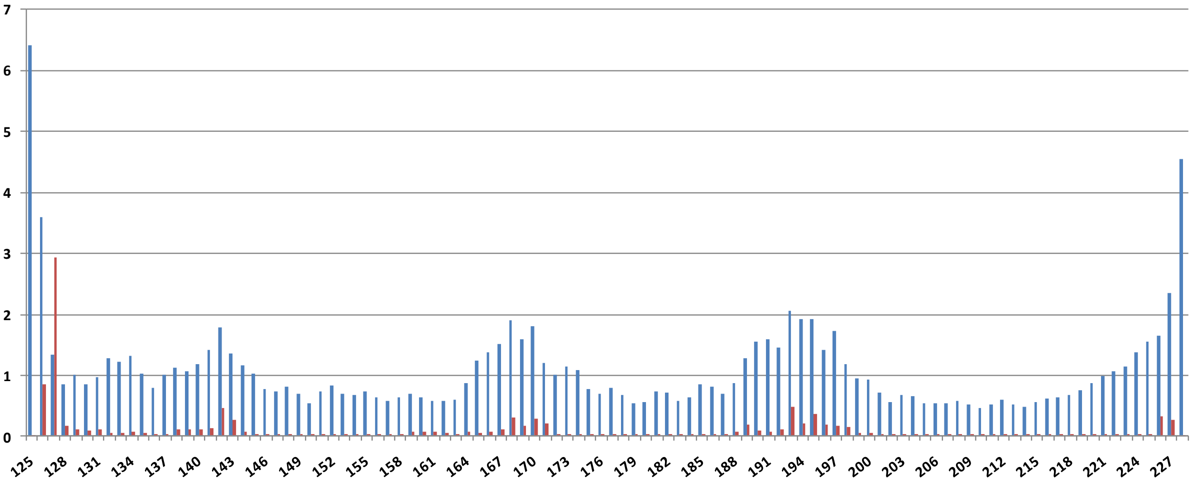


Figure A1.2.1: Values of RMSF (blue bars, in Å) and ADI (red bars) for HuPrP(PP).

iii) *The ADI value for each residue is strictly dependent on the single residue*. Figure A1.2.2 plots the ψ values of two residues as a function of simulation time for HuPrP(Q212P-129M).

A) The residue experiences a transition between two Ramachandran regions.

B) The residue remains in the same Ramachandran region along the whole trajectory with slight fluctuations.

The corresponding values of ADI and RMSF of C_α atoms are reported in Table A1.2.1. The values of ADI allow distinguishing between these very different cases: S170 exhibits the largest value of ADI, indeed the two Ramachandran regions are equally populated, indicating a conformational transition. E200 has a lower ADI. It indeed samples only one Ramachandran region and remains in the same conformation for the whole simulation. The RMSF values do follow the opposite trend (Table A1.2.1).

Table A1.2.1: Measure of flexibility values of ADI and RMSF for selected residues of HuPrP(Q212P-129M).

Residue	ADI	RMSF (Å)
S170	0.63	1.20
E200	0.03	1.39

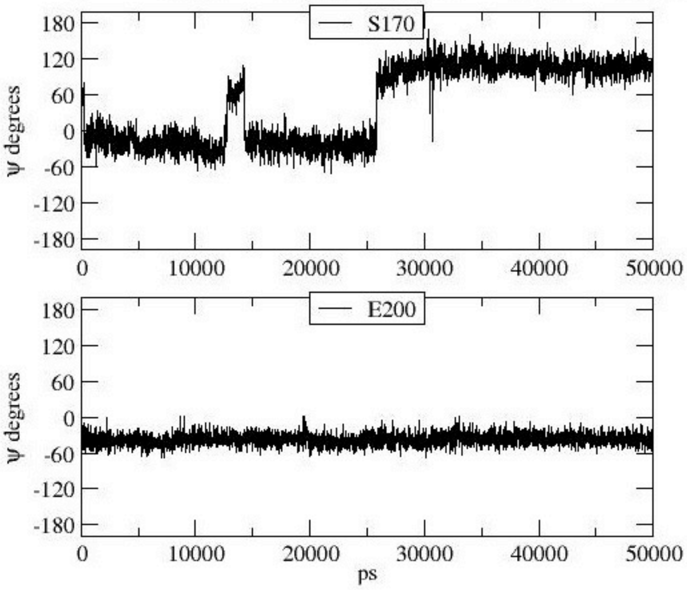


Figure A1.2.2: Values of dihedral angle ψ evolved in residue S170 (top) and E200 (bottom) during the MD trajectory of HuPrP(Q212P-129M).

The reason for such discrepancy can be explained by Fig. A1.2.3, where the values of RMSF and ADI for all the residues of the considered mutant are compared. Arrowheads point to the above-mentioned residues S170 and E200: they both fall in the more flexible loop regions (the β2-α2 and the α2-α3 loops respectively). As a consequence, their RMSF values, being influenced by those of the neighboring residues, are larger, while their ADI values are specific of these residues and very different from those of the neighboring residues.

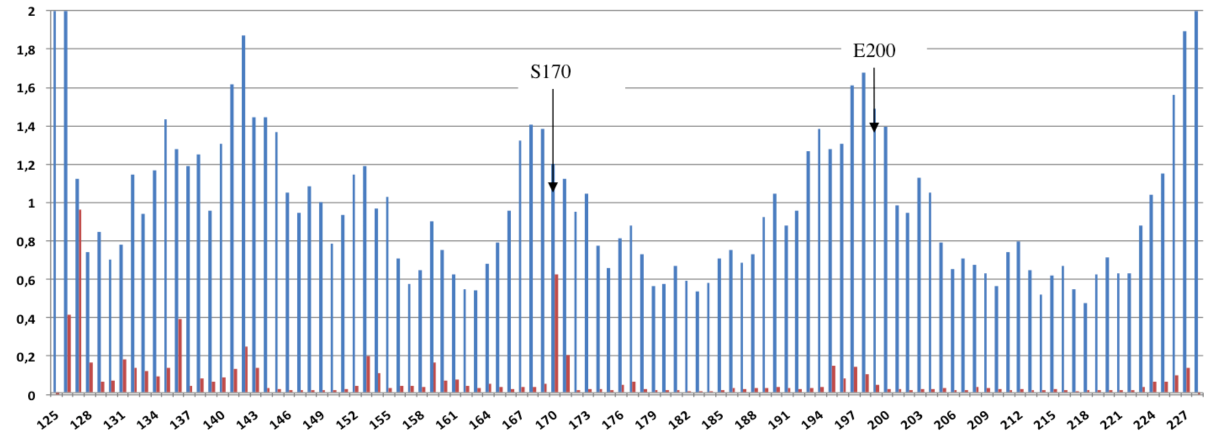


Figure A1.2.3. Values of RMSF (blue bars, in Å) and ADI (red bars) for HuPrP(Q212P-129M).

Appendix 1.3 The SB network in the HuPrP variants

Table A1.3.1: The d_{SB} (nm) values (defined in Chapter 4, section 4.2.3) calculated for the main clusters extracted from the MD trajectory of each HuPrP variant, which covered ~80% of the entire conformer population in the trajectory. The cluster configurations of E207-K204 and E200-K204 in HuPrP(WT) and HuPrP(PP) correspond to SBs alternatively formed and broken: K204 side chain switched indeed from a conformation closed to E200 to another closed to E146. Consistently, K204 is in a middle position between both in the most populated cluster, close to E200 in the second and to E146 in the third.

Table A1.3.1-1

HuPrP variant	WT			G131V-129M				S132I-129M				A133V-129M		
Total population collected (%)	80			74				74				83		
Relative population collected (%)	61.2	21.7	17.1	72.7	12.4	9.7	5.1	32.4	25.4	22.4	19.7	57.1	28.1	14.8
Cluster	1	2	3	1	2	3	4	1	2	3	4	1	2	3
E211-R208	3.9	4.0	3.8	7.0	3.8	4.0	3.9	5.0	7.2	4.7	10.7	4.0	8.1	3.9
E207-R208	3.9	4.4	4.5	8.8	4.2	4.6	4.8	11.3	8.6	9.9	9.4	4.4	9.6	4.4
E207-K204	10.1	9.7	10.5	3.5	5.7	3.4	9.4	3.5	9.7	6.6	4.5	9.9	3.3	9.5
E200-K204	9.3	3.7	9.2	11.1	11.6	10.9	5.1	12.7	11.0	4.0	10.3	5.7	8.8	3.7
E146-K204	7.8	8.4	3.5	13.2	10.9	11.7	12.5	13.1	5.0	14.0	11.8	4.6	13.2	11.4
E196-R156	3.9	4.0	4.0	4.0	4.1	6.6	4.4	5.1	3.9	13.9	3.9	4.5	4.8	4.1

Table A1.3.1-2

HuPrP variant	R148H-129M			D178N-129M			D178N-129V			V180I-129M			T183A-129M		
Total population collected (%)	80			78			76			89			89		
Relative population collected (%)	55.9	30.4	13.7	74.1	14.9	11.0	53.9	25.3	20.8	72.9	15.4	11.6	58.3	27.6	14.1
Cluster	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
E211-R208	9.1	4.0	3.9	8.3	8.9	8.7	10.3	9.5	9.5	7.6	4.4	4.0	8.6	8.0	7.8
E207-R208	10.0	4.2	4.3	10.3	10.7	10.0	9.5	9.3	8.3	10.1	8.7	4.2	12.4	10.3	10.3
E207-K204	3.2	9.0	3.5	8.5	3.7	3.2	7.9	3.2	3.3	10.9	10.7	6.6	9.5	11.0	8.2
E200-K204	9.1	6.3	10.3	7.6	10.4	9.6	9.5	17.4	9.2	8.5	4.6	6.9	7.0	4.9	5.0
E146-K204	13.6	13.8	12.0	9.5	12.6	16.3	9.1	11.8	10.6	6.3	7.3	12.6	11.9	9.3	10.7
E196-R156	7.1	6.9	7.4	4.0	4.0	4.0	8.4	13.9	10.5	3.9	4.1	3.8	6.3	4.0	4.0

Table A1.3.1-3

HuPrP variant	T183A-129V				H187R-129V				T188K-129M			T188A-129M		
Total population collected (%)	77				81				94			97		
Relative population collected (%)	59.2	16.9	15.1	8.8	62.7	16.0	15.2	6.1	57.7	37.0	5.3	83.2	11.0	5.8
Cluster	1	2	3	4	1	2	3	4	1	2	3	1	2	3
E211-R208	8.9	6.9	4.5	9.4	9.5	9.3	4.8	10.0	9.7	10.0	9.2	8.8	3.9	3.8
E207-R208	10.0	9.6	9.9	9.3	8.9	9.8	8.7	7.8	8.4	9.9	9.1	11.6	4.3	6.4
E207-K204	8.7	8.6	3.5	8.3	8.0	9.7	5.9	3.6	8.0	8.3	9.9	14.1	10.7	9.5
E200-K204	3.6	3.9	10.3	10.2	5.7	5.2	10.5	8.1	7.0	3.7	4.4	14.5	9.6	10.1
E146-K204	14.4	11.6	11.9	8.8	6.4	6.3	9.3	9.0	14.0	13.4	15.3	3.3	7.4	3.7
E196-R156	6.5	4.3	3.8	4.1	4.1	5.8	4.2	8.3	8.9	4.2	8.8	11.2	9.3	10.1

Table A1.3.1-4

HuPrP variant	T188A-129V			T188R-129V			T193I-129M			E196K-129M			F198S-129V		
Total population collected (%)	79			83			84			79			80		
Relative population collected (%)	66.8	23.3	9.9	54.5	23.0	22.4	63.5	24.3	12.2	40.4	36.0	23.4	80.1	11.2	8.7
Cluster	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
E211-R208	9.4	10.2	4.9	10.8	10.5	10.7	7.8	4.6	4.8	8.9	9.2	4.0	8.5	8.7	3.7
E207-R208	8.7	8.9	11.5	9.9	10.8	9.1	12.3	11.1	10.4	7.0	9.1	8.0	9.5	8.7	4.4
E207-K204	9.5	3.5	14.8	3.9	6.0	3.8	15.6	14.3	9.3	3.6	3.8	8.2	8.8	9.9	7.5
E200-K204	10.1	9.0	13.0	11.5	8.6	9.5	12.0	14.5	12.5	12.4	10.8	8.6	8.2	5.6	9.1
E146-K204	3.7	12.6	3.4	12.0	11.2	13.3	3.7	6.8	13.7	11.7	11.6	9.2	5.9	7.2	7.1
E196-R156	3.8	3.9	4.4	3.9	3.6	3.6	6.5	7.3	5.6	-	-	-	4.5	3.7	4.2

Table A1.3.1-5

HuPrP variant	E200K-129M			E200K-129V				D202N-129M			D202N-129V			V203I-129M		
Total population collected (%)	86			77.2				76			86			79		
Relative population collected (%)	59.5	25.5	15.0	60.1	23.1	9.6	7.3	68.9	23.0	8.2	55.7	36.8	7.5	58.9	37.5	13.6
Cluster	1	2	3	1	2	3	4	1	2	3	1	2	3	1	2	3
E211-R208	9.4	3.8	4.3	4.6	4.1	4.5	4.9	10.3	7.0	8.4	8.3	10.0	7.0	9.5	8.9	10.1
E207-R208	9.7	4.5	6.9	9.5	8.7	8.1	9.4	9.8	6.2	9.5	10.3	10.1	9.0	9.4	8.1	9.1
E207-K204	3.6	9.8	9.4	10.2	8.7	9.4	10.3	8.2	3.5	3.5	3.7	8.3	9.3	3.6	9.6	9.1
E200-K204	8.2	7.0	9.6	-	-	-	-	14.8	13.7	11.6	9.0	3.6	4.7	8.8	7.0	6.4
E146-K204	3.5	10.6	6.6	3.5	7.7	5.4	5.6	7.3	14.0	12.9	13.3	14.1	11.3	11.8	8.0	8.6
E196-R156	3.6	3.7	7.5	4.1	4.4	4.0	4.0	7.2	6.7	4.4	3.7	4.0	10.2	9.4	4.7	12.1

Table A1.3.1-6

HuPrP variant	R208H-129M			R208H-129V			V210I-129M			V210I-129V			E211Q-129M		
Total population collected (%)	81			89			79.8			78			79		
Relative population collected (%)	45.3	29.6	25.1	69.4	18.5	12.1	52.6	24.6	22.8	60.3	29.3	10.4	47.8	20.8	18.7
Cluster	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
E211-R208	-	-	-	-	-	-	4.8	10.2	8.2	3.8	3.9	7.8	-	-	-
E207-R208	-	-	-	-	-	-	4.4	8.1	7.2	4.4	4.3	9.8	9.4	4.9	9.0
E207-K204	6.6	11.0	8.5	8.2	8.4	7.2	3.9	3.5	3.4	8.6	8.8	3.7	3.4	11.2	3.3
E200-K204	9.9	10.4	11.3	6.9	6.8	5.5	10.1	10.7	9.4	7.6	3.5	9.7	7.4	10.1	7.0
E146-K204	10.6	6.8	10.5	11.5	11.8	11.8	12.3	9.5	12.3	8.6	11.3	10.4	13.5	4.0	10.7
E196-R156	7.3	4.2	6.7	3.7	3.9	3.7	7.2	4.0	5.5	7.9	7.4	4.9	9.1	10.1	15.3

Table A1.3.1-7

HuPrP variant	Q212P-129M			Q217R-129V				Y218N-129V			PP		
Total population collected (%)	83			81				81			84		
Relative population collected (%)	76.0	14.2	9.8	72.2	13.3	7.9	6.6	61.2	25.8	13.0	60.6	29.4	10.0
Cluster	1	2	3	1	2	3	4	1	2	3	1	2	3
E211-R208	3.8	9.7	7.1	8.5	4.6	3.8	9.8	9.6	9.7	10.5	5.1	5.0	5.1
E207-R208	3.9	7.4	11.1	10.0	9.2	4.0	8.2	10.9	9.6	9.3	3.5	3.6	3.6
E207-K204	9.7	9.8	4.0	8.6	5.4	11.5	8.1	6.9	3.9	7.5	9.7	9.9	8.8
E200-K204	6.4	3.8	9.5	8.9	10.8	4.4	8.8	9.5	12.1	10.1	5.9	3.5	9.4
E146-K204	10.8	11.3	11.4	5.2	10.0	9.1	4.7	10.9	10.9	8.9	4.6	11.1	4.7
E196-R156	5.3	5.7	11.2	5.3	4.4	5.7	4.0	3.7	5.9	3.9	4.5	5.4	5.6

Appendix 2 Supplementary materials for Chapter 5

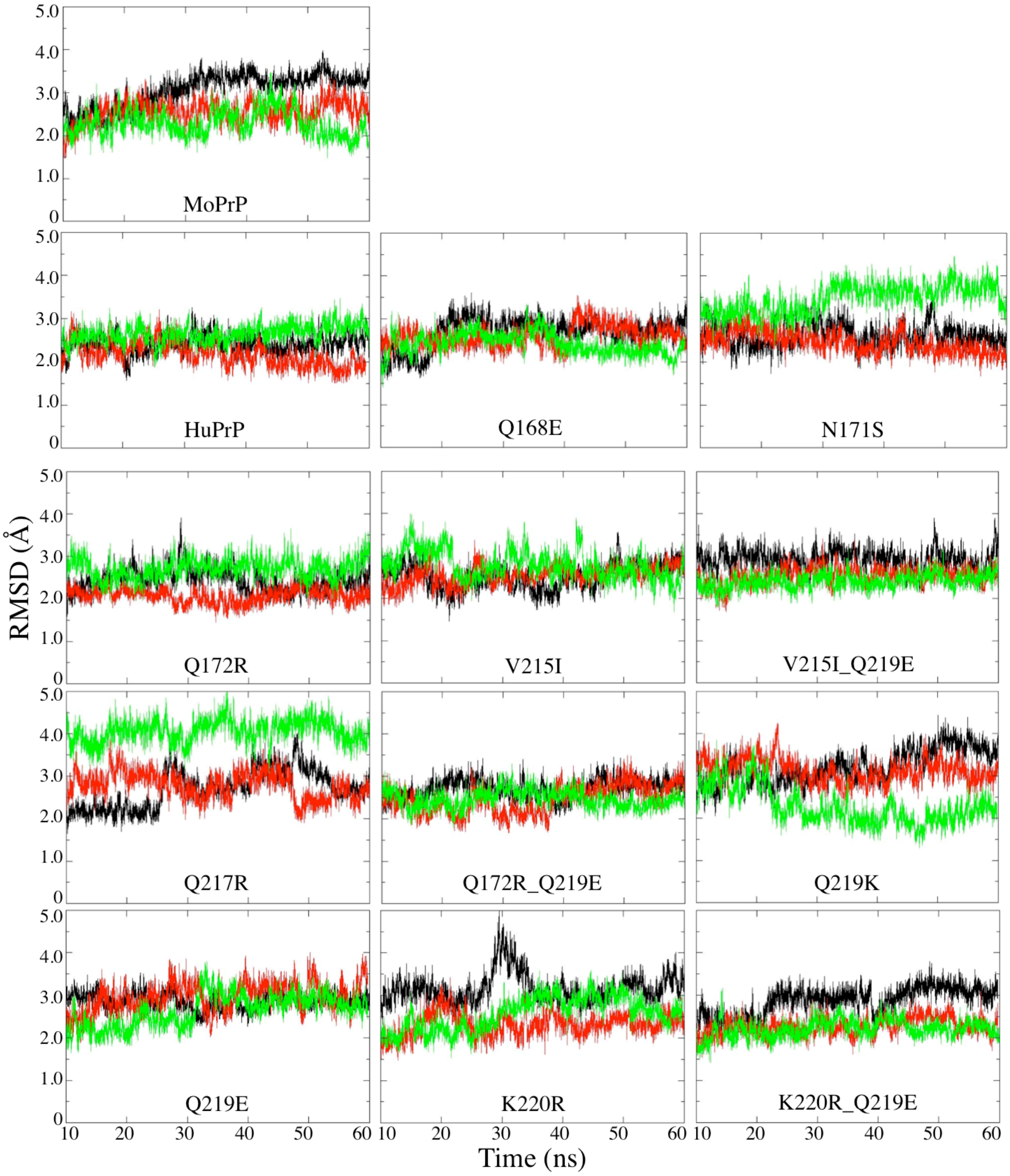


Figure A2.1: RMSD of the C_{α} atoms of MoPrP, HuPrP and all the CMPrP with respect to the initial models, plotted as a function of simulation time. Three independent MD runs in each system are shown in black, red and green, respectively.

Appendix 2.1 Test simulations on the histidine protonation state

We carried out 20-ns MD simulations on 7 forms of the WT MoPrP in explicit solvent. These forms differed from each other in the protonation state of the 3 histidines (H140, H177 and H187) present in MoPrP GD (Table A2.1.1). The simulation method was the same as that described in Chapter 5, section 5.2.2. All the systems appeared to reach equilibrium within 10 ns of MD, as indicated by the RMSD of C_α atoms as a function of simulation time (data not shown). The RMSF of Cα atoms in each system (Fig. A2.1.1) were then calculated from the last 10 ns of the simulations.

MoPrP_HIE, in which all the 3 histidines were in neutral state and protonated at the N_ε atoms (Table A2.1.1) turned out to have the smallest RMSF (Fig. A2.1.1a). The other systems showed much larger RMSF (Fig. A2.1.1b-e) than MoPrP_HIE. We concluded that MoPrP is likely to exist in the HIE form in solution at pH 7. This is the same protonation state as that of HuPrP at pH 7. This result was not unexpected since these histidines have the same local environment as those in HuPrP.

Table A2.1.1: Protonation state[i] of the 3 histidines in the 7 forms of MoPrP considered here in Appendix 2.1.

Form	HIE	140HID	177HID	187HID	140HIP	177HIP	187HIP
H140	N_ε	N_δ	N_ε	N_ε	N_ε, N_δ	N_ε	N_ε
H177	N_ε	N_ε	N_δ	N_ε	N_ε	N_ε, N_δ	N_ε
H187	N_ε	N_ε	N_ε	N_δ	N_ε	N_ε	N_ε, N_δ

[i] For each form, the protonated nitrogen atom(s) is (are) indicated.

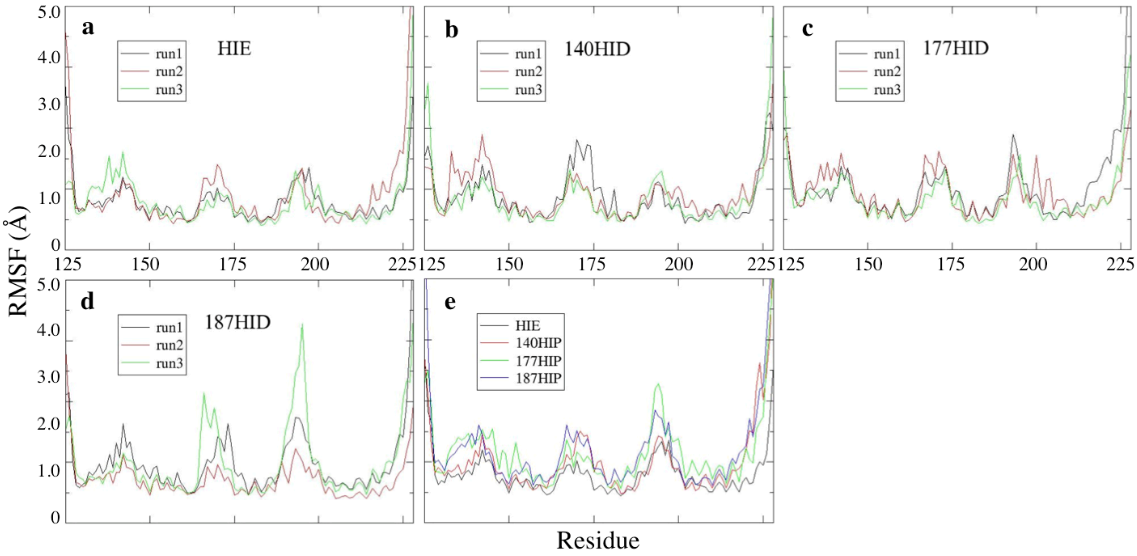


Figure A2.1.1: RMSF of C_α atoms of the 7 forms of MoPrP in different histidine protonation states during 10 ns of equilibrated test simulations. **a-d)** Three independent runs were carried out on each neutral state. On average, the HIE state gave the smallest RMSF. **e)** Three charged states showed apparent larger RMSF compared to the HIE state, therefore only one run was performed on these charged states.

Appendix 3 Supplementary materials for Chapter 6

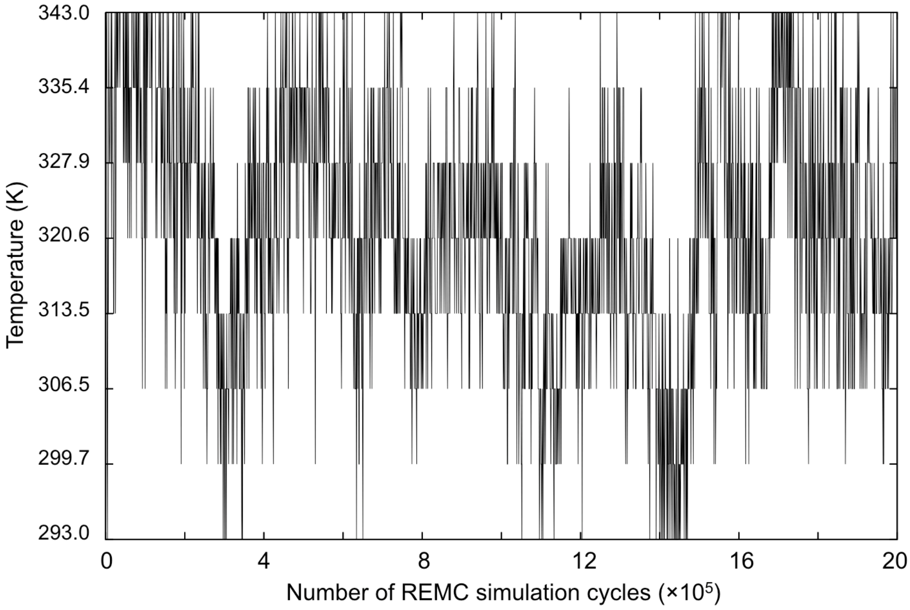


Figure A3.1: Combined histogram of energy at different temperatures during a typical run of our simulations. The observed energy distribution is unimodal at each temperature and there is a substantial overlap in the energy histogram at any pair of consecutive temperatures, leading to a high acceptance probability for the replica exchange move. In the systems examined in this study, there were no observed bottle-necks complicating the flow of replicas across the temperatures.

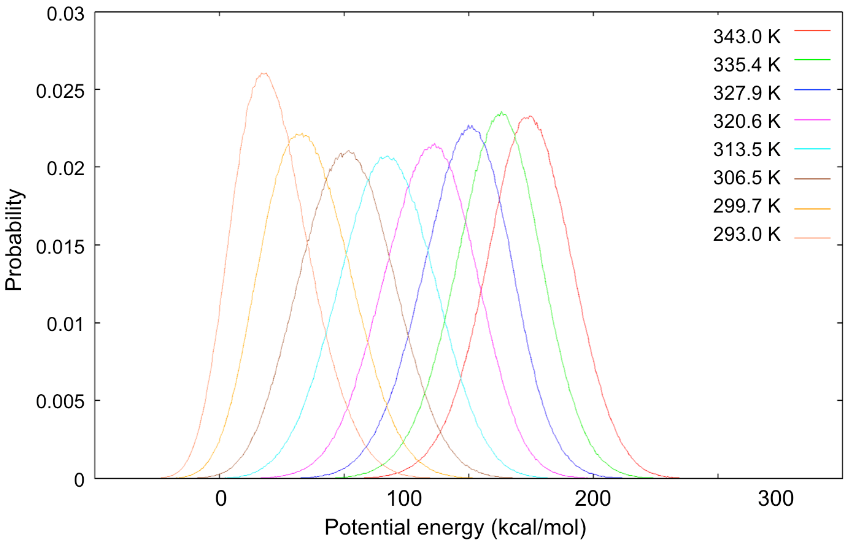


Figure A3.2: The random walk of a representative replica (in a typical run) in the temperature space of our simulations. Our simulation runs are long enough to permit several round trips of each replica from one extreme of temperature to the other extreme and back.

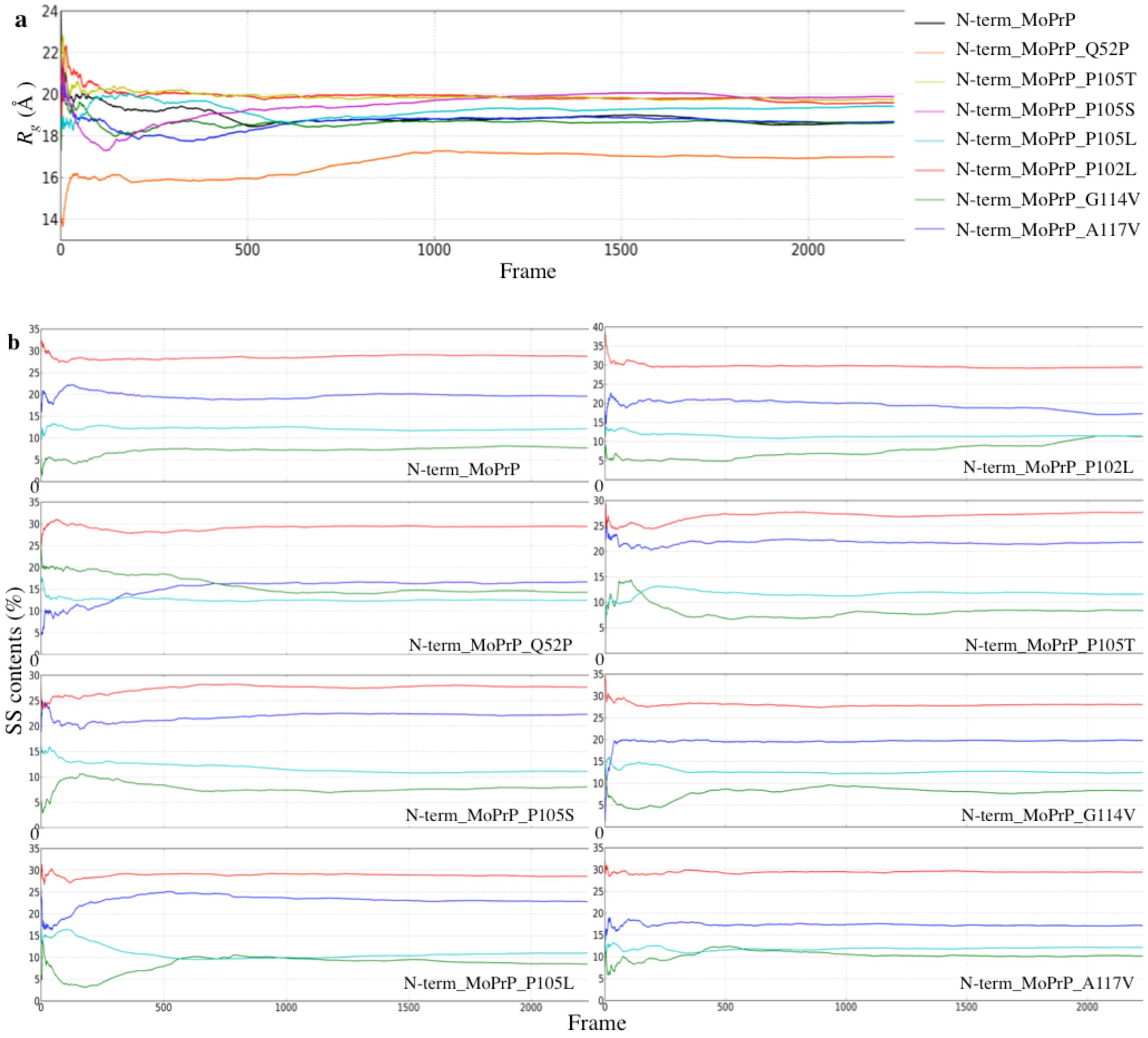


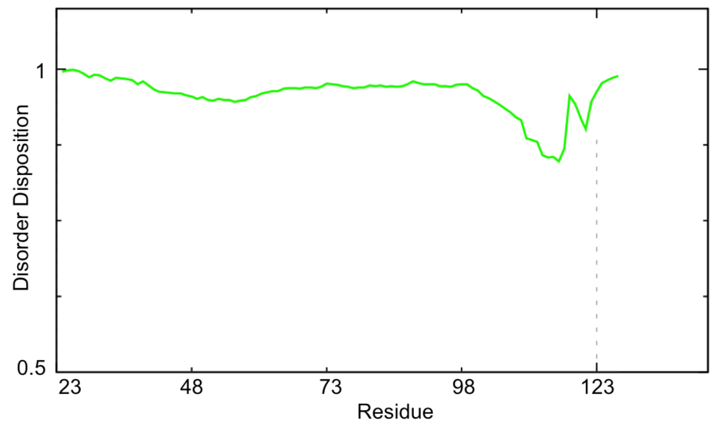
Figure A3.3: Cumulative averages of **a)** R_g and **b)** selected SS contents as a function of the MC frames of N-term_MoPrP and the PMs at 293 K. The color code in **b)** is: α-helix (blue), β-sheet (green), β-turn (red) and bend (cyan).

Table A3.1: Content of transient secondary structure (SS) elements (in %[i]) in N-term_MoPrP averaged over the REMC simulation trajectory at 293 K, 306.5 K and 313.5 K, respectively.

REMC	α-helix	β-sheet	β-bridge	β-turn	Bend	3_{10}-helix	π-helix	Coil
293 K	19±8	8±5	7±3	27±5	12±4	4±3	1±1	25±5
306.5 K	14±8	6±6	6±3	28±4	15±3	5±4	0±2	31±6
313.5 K	11±8	5±6	6±5	28±6	16±5	5±4	0±1	34±6

[i] Standard deviations (SD) are indicated by "±".

a



b

```
  25        35        45        55        65        75  Residue number
  |         |         |         |         |         |
KKRPKPGGWNTGGSRYPGQGSPGGNRYPPQGGTWGQPHGGGWGQPHGGSWGQP   Sequence
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC   PSIPRED
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC   Jpred3
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC   Porter
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC   RaptorX
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC   CASB-fold

         85        95       105       115       125     Residue number
         |         |         |         |         |
HGGSWGQPHGGGWGQGGGTHNQWNKPSKPKTNLKHVAGAAAAGAVVGGLGGY    Sequence
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHCCCCCC    PSIPRED
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHCCCCCC    Jpred3
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHCCCCCCCCCC    Porter
CCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHCCCC    RaptorX
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHCCCCCCC    CASB-fold
```

Figure A3.4: a) The intrinsic disorder disposition of MoPrP^{C} N-terminal residues (23-128, here in Mo numbering) predicted by PONDR-FIT. **b)** Predictions of SS in N-term_MoPrP using five different bioinformatics tools, PSIPRED[398], Jpred3[399], Porter[400], RaptorX[401] and CABS-fold[402]. The letter code for predicted SS: C, random coil; H, α-helix.

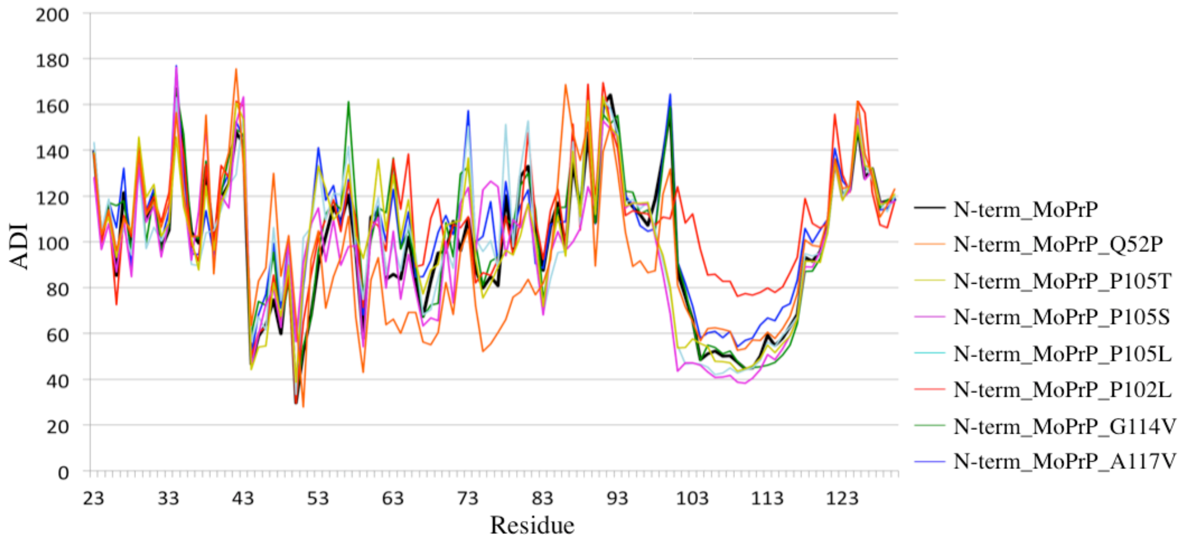


Figure A3.5: Backbone flexibility per residue in WT and mutated N-term_MoPrP calculated using the ADI analysis[305] over the frames sampled from the production trajectories. The higher the ADI value, the more flexible the backbone.

Figure A3.6: Flexibility of backbone units of residues 100-128 measured by the ADI analysis [305]. The SDs in the 10 simulation runs of each system are indicated with vertical bars.

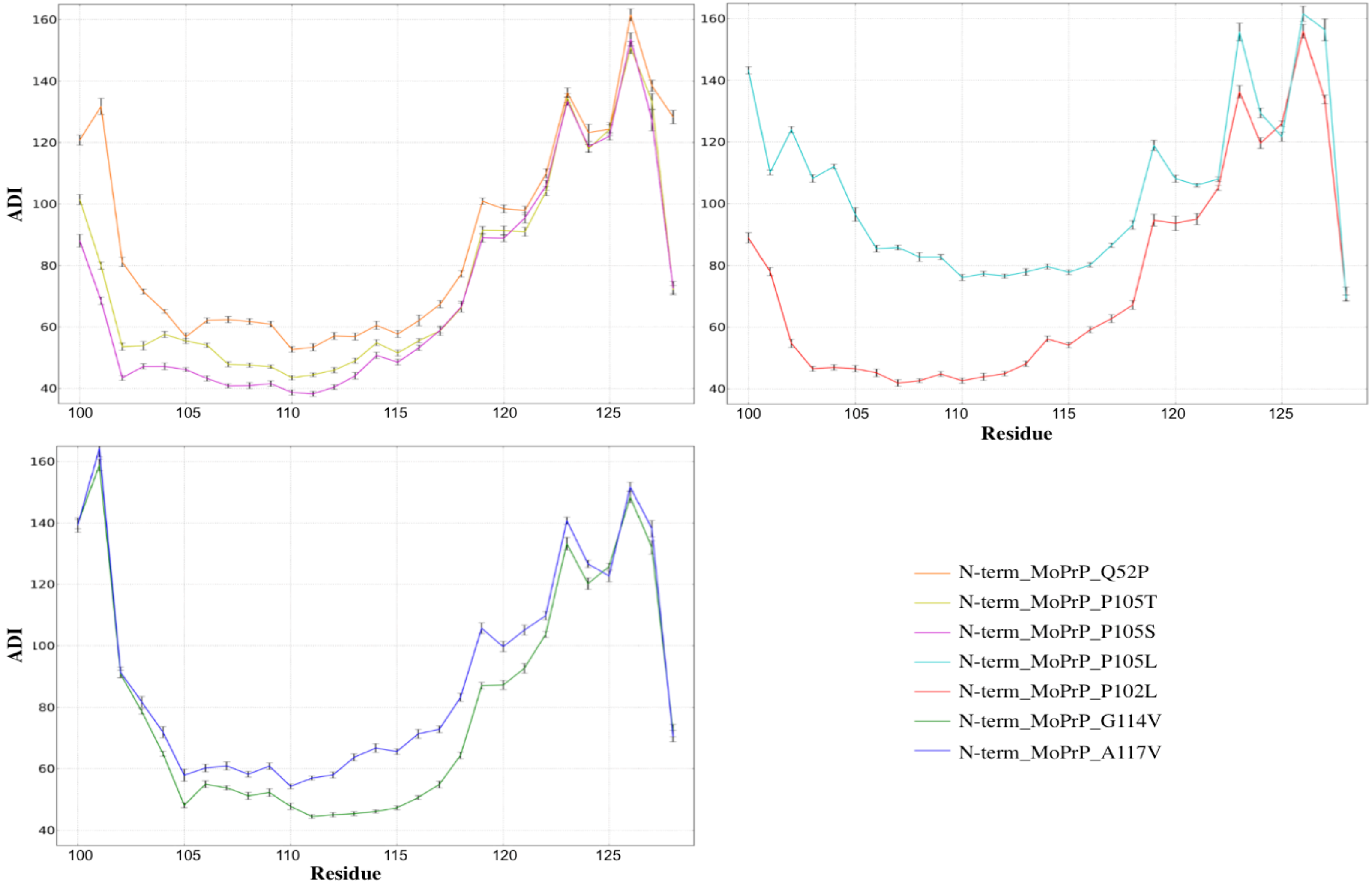


Figure A3.7: Total number of intra-molecular contacts per residue in the WT and mutated N-term_MoPrP. The SDs in the 10 simulation runs of each system are indicated with vertical bars.

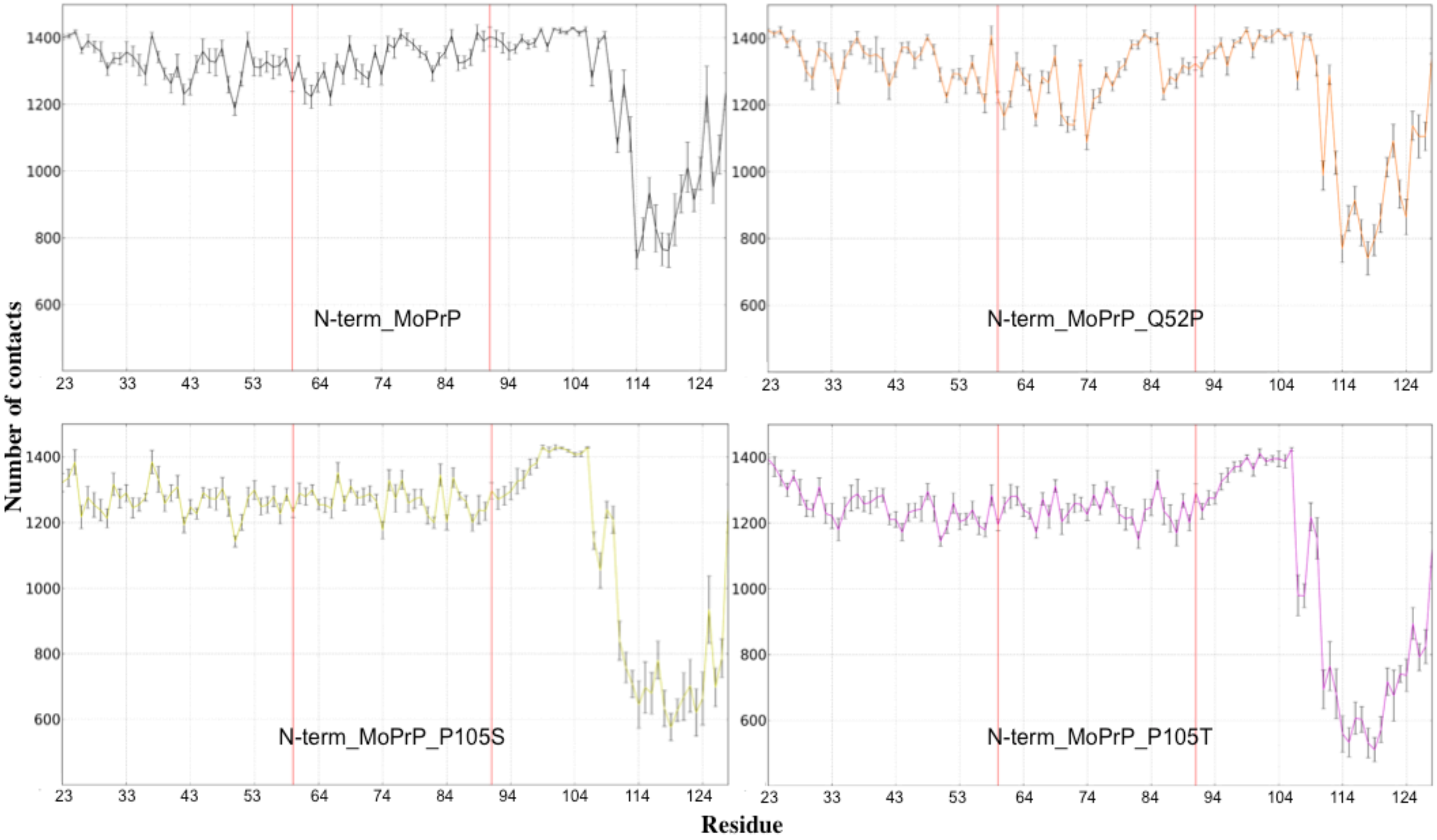


Figure A3.7-1

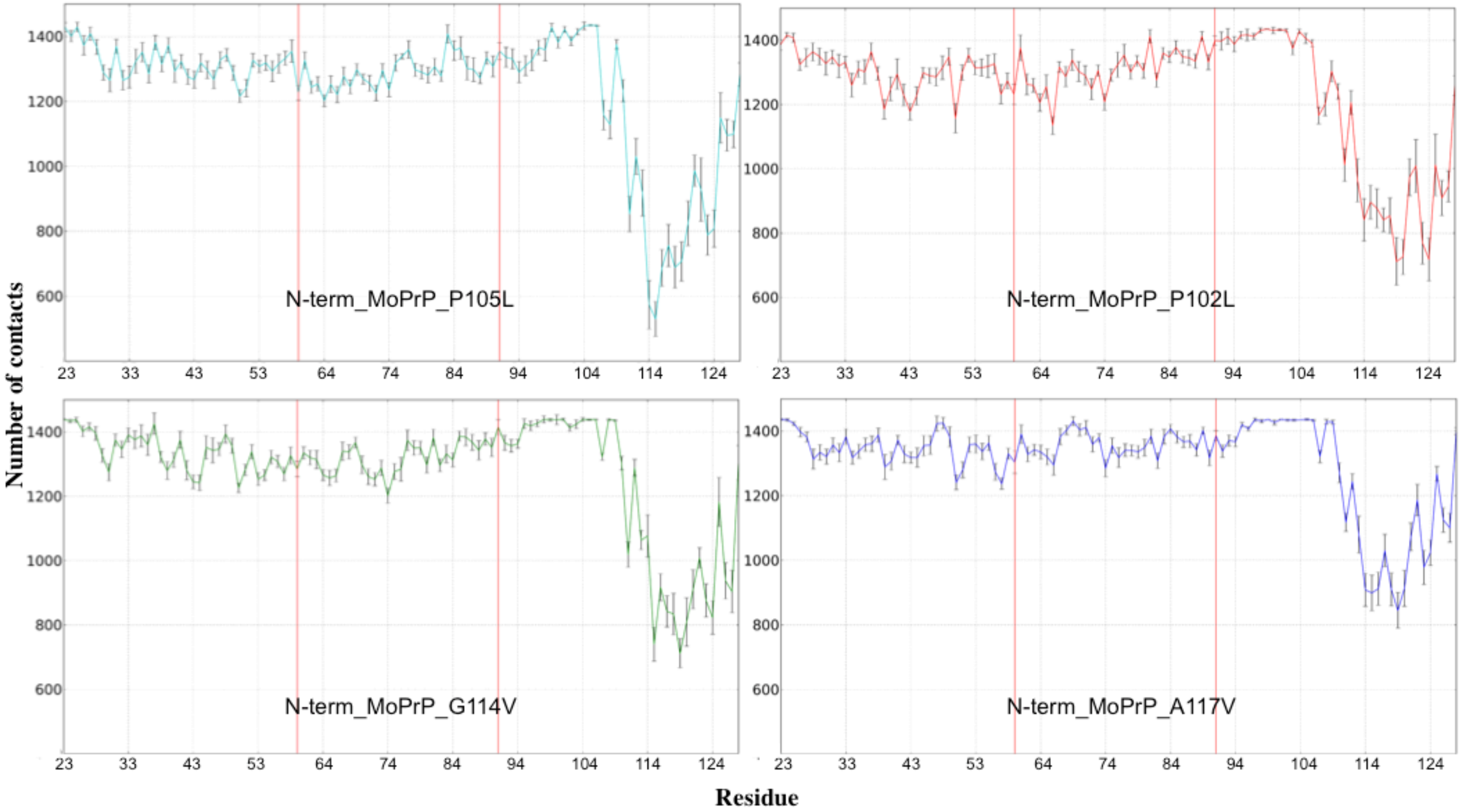
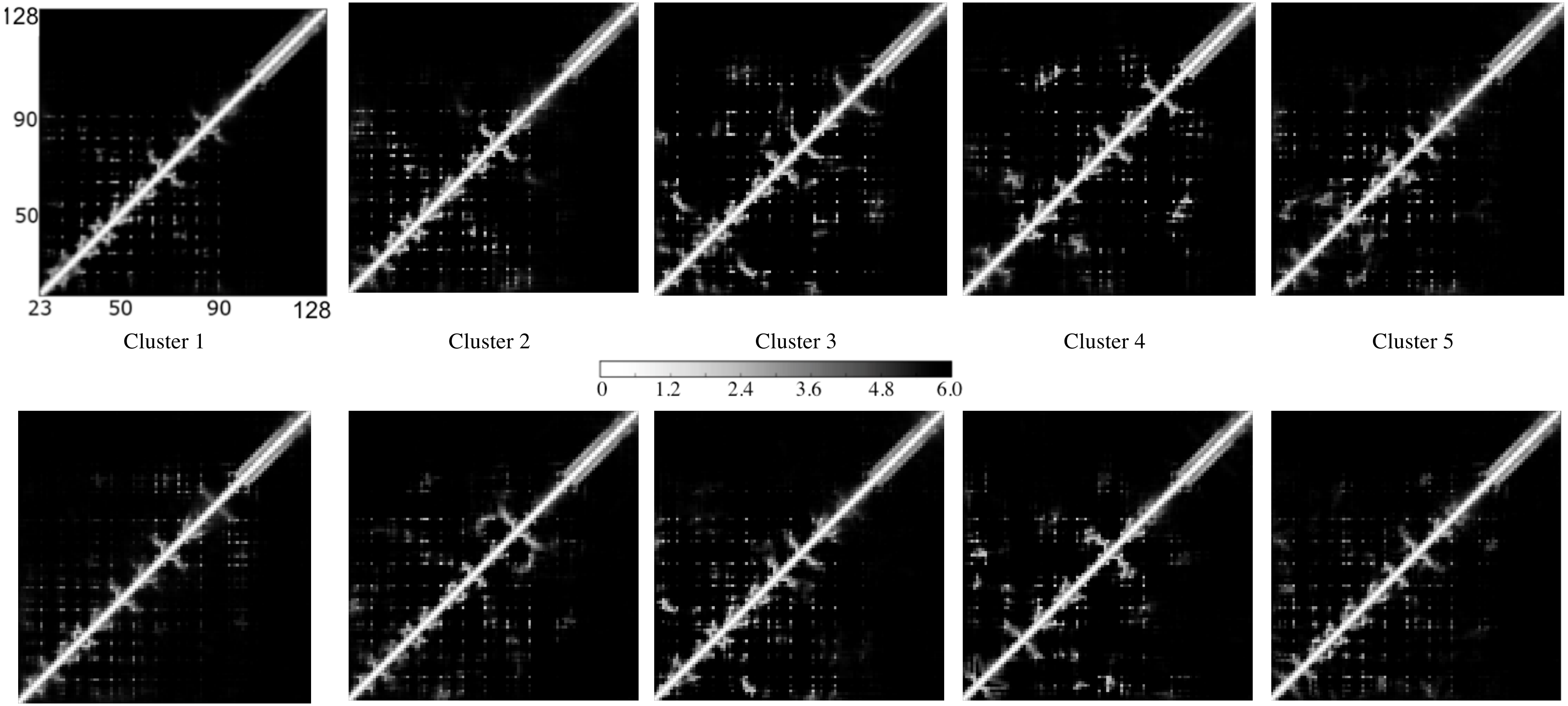

N-term_MoPrP_P105L
N-term_MoPrP_P102L
N-term_MoPrP_G114V
N-term_MoPrP_A117V
Number of contacts
Residue
1400
1200
1000
800
600
23
33
43
53
64
74
84
94
104
114
124

Figure A3.7-2

Appendix 3.1 Intra-molecular contacts in the WT and mutated N-term_MoPrP

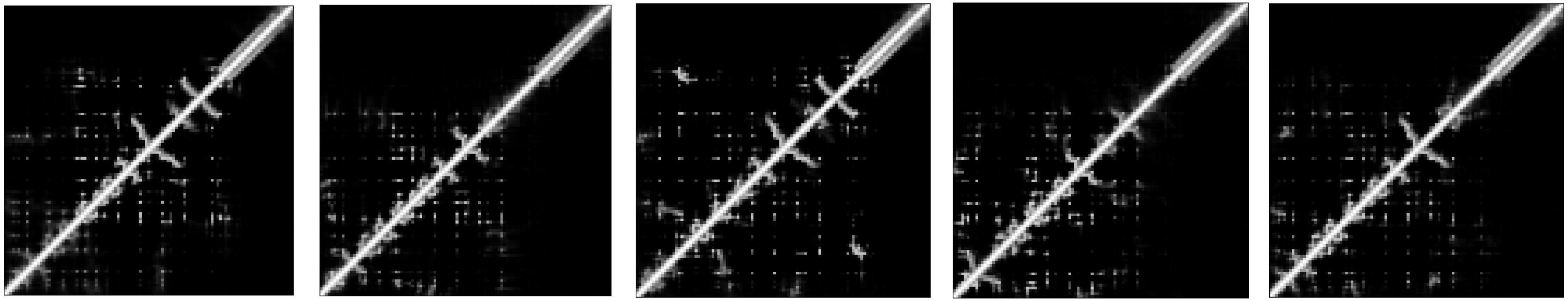
Figure A3.1.1: Contact maps illustrating the intra-molecular interactions in the N-term_MoPrP and PMs. For each protein, the 10 most populated clusters are shown. Contacts within 6 Å are plotted in white. Residue numbers are labeled in the contact map of N-term_MoPrP cluster 1 and are the same for all the other contact maps.



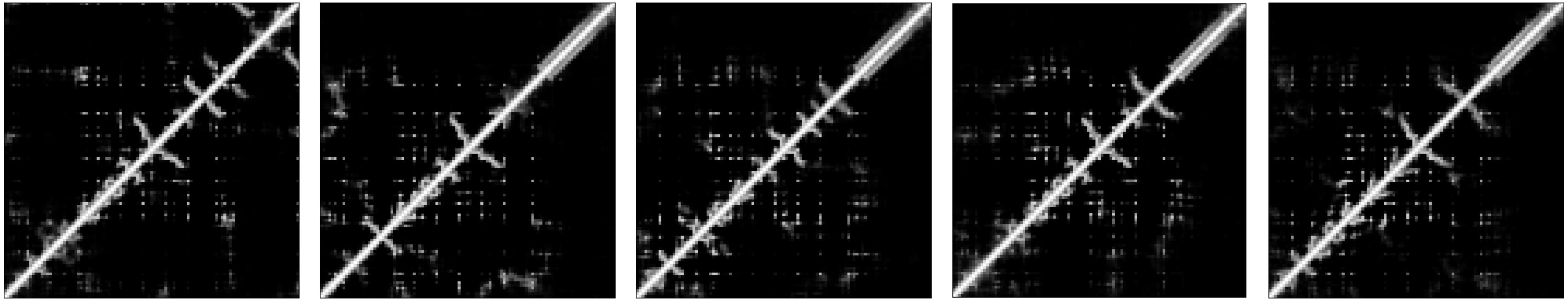
Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5

Cluster 6 Cluster 7 Cluster 8 Cluster 9 Cluster 10

Figure A3.1.1-1: Contact maps of N-term_MoPrP.

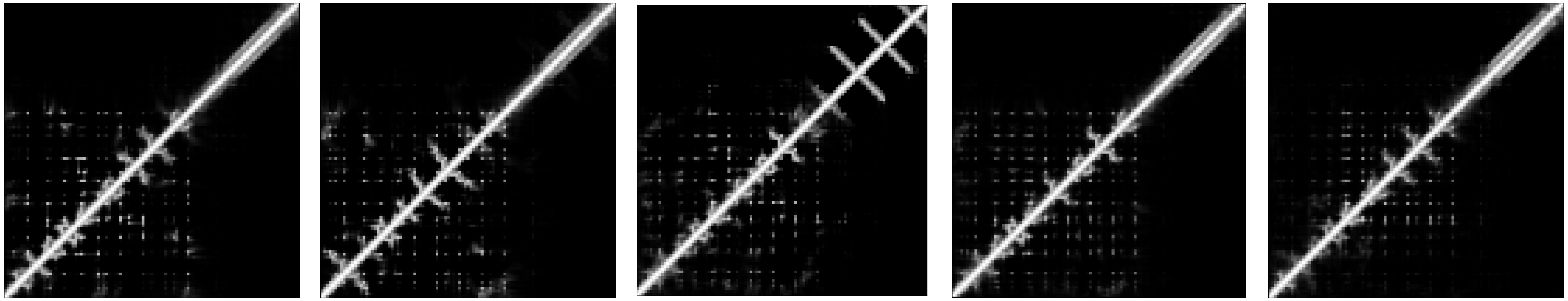


Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5

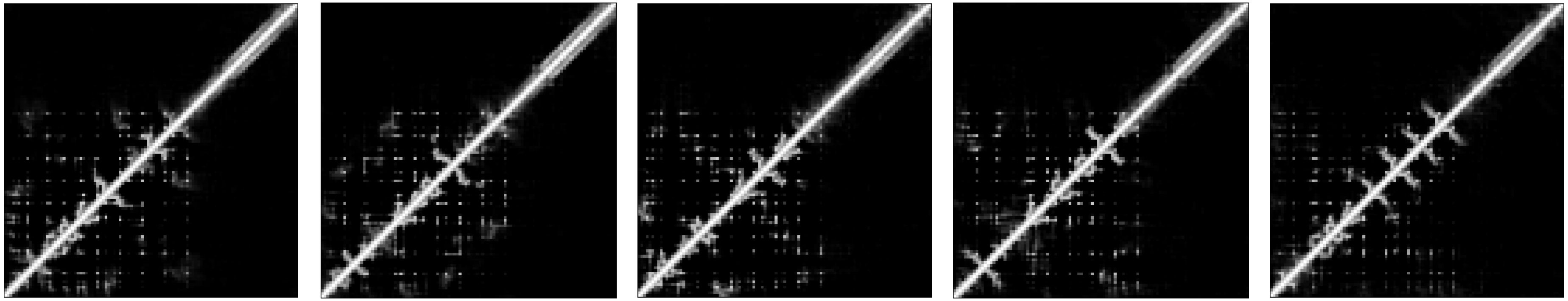


Cluster 6 Cluster 7 Cluster 8 Cluster 9 Cluster 10

Figure A3.1.1-2: Contact maps of N-term_MoPrP_Q52P.

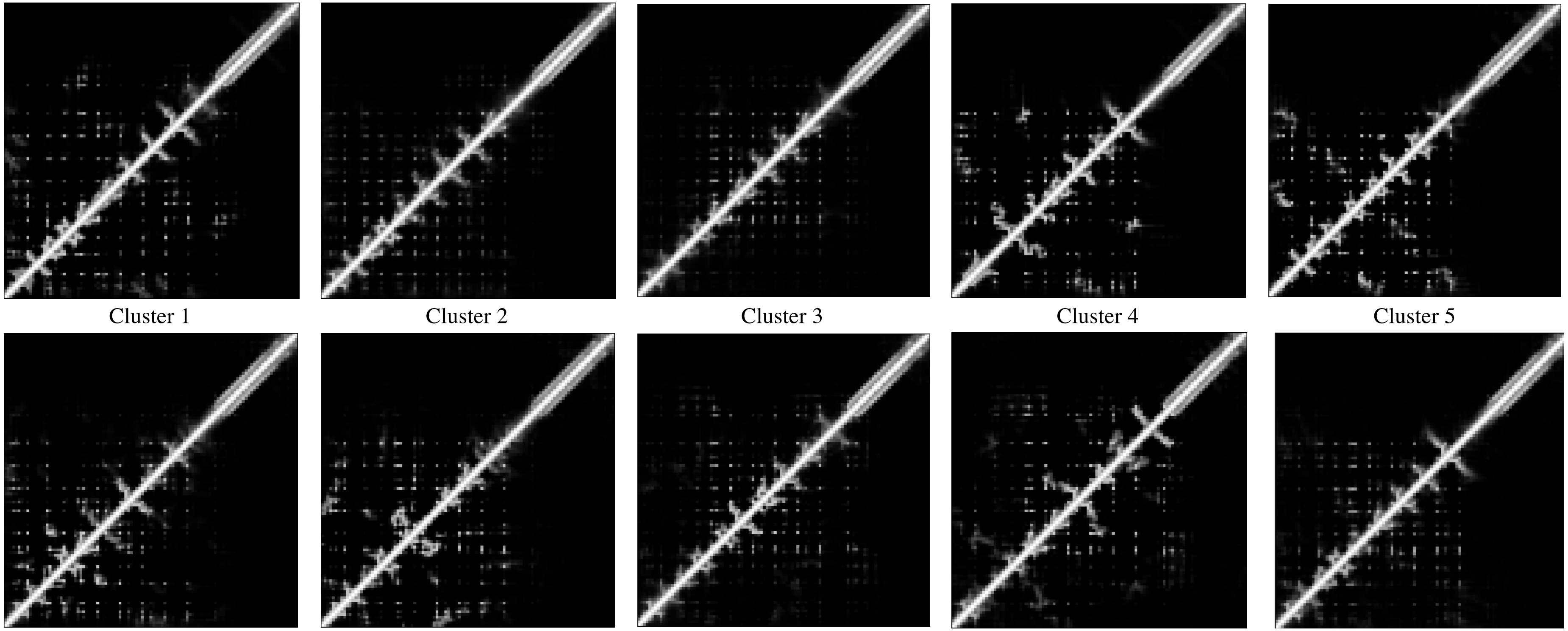


Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5



Cluster 6 Cluster 7 Cluster 8 Cluster 9 Cluster 10

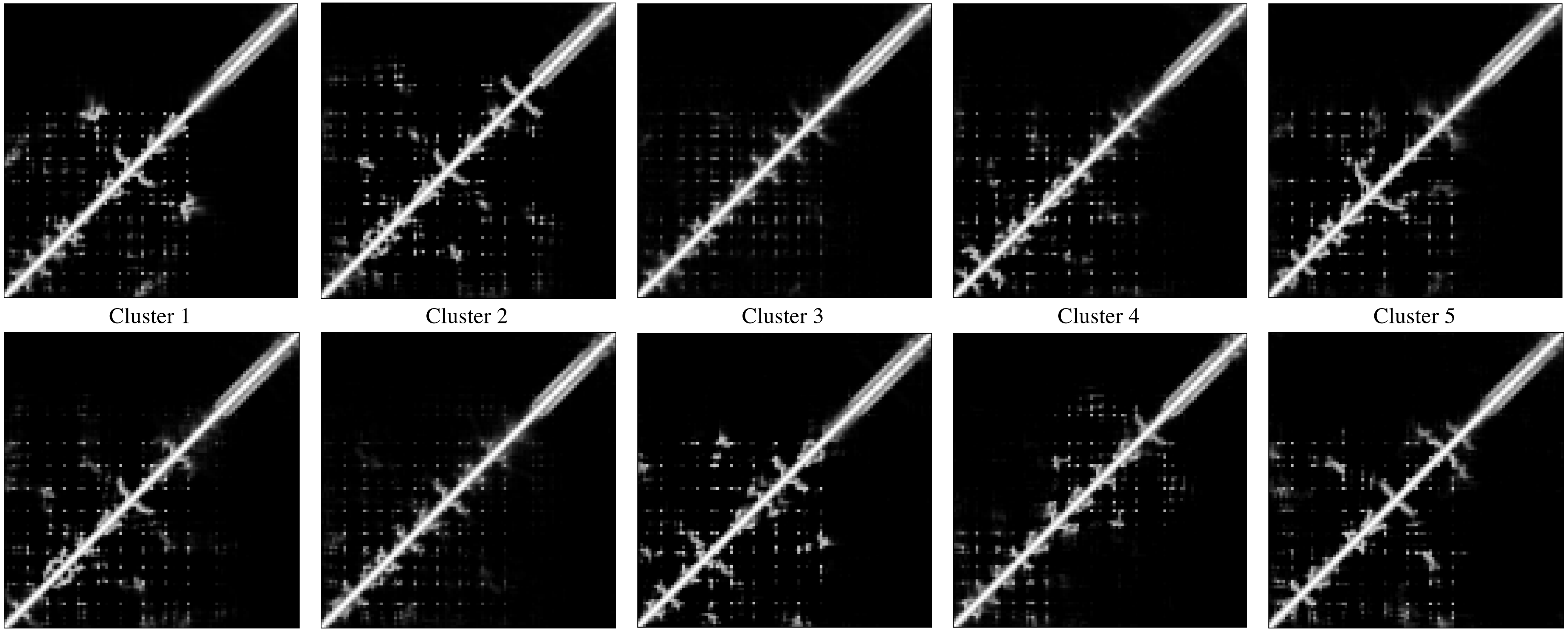
Figure A3.1.1-3: Contact maps of N-term_MoPrP_P102L.



Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5

Cluster 6 Cluster 7 Cluster 8 Cluster 9 Cluster 10

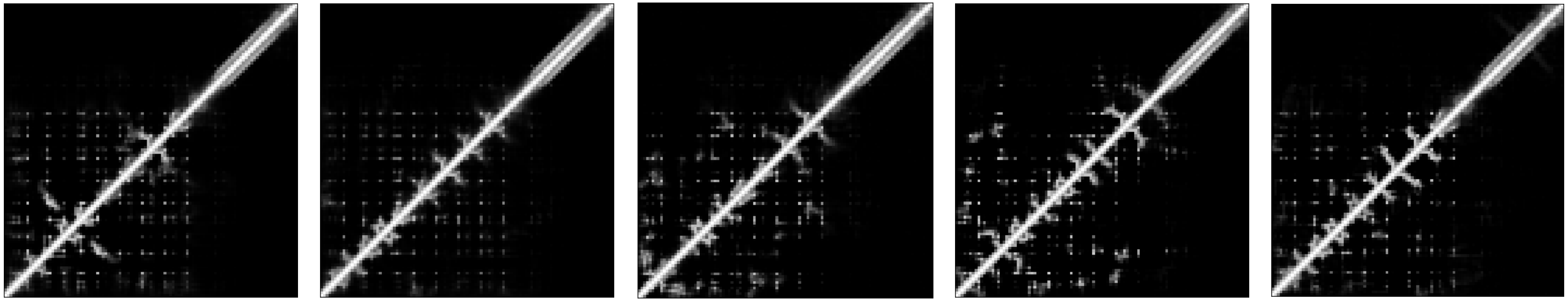
Figure A3.1.1-4: Contact maps of N-term_MoPrP_P105L.



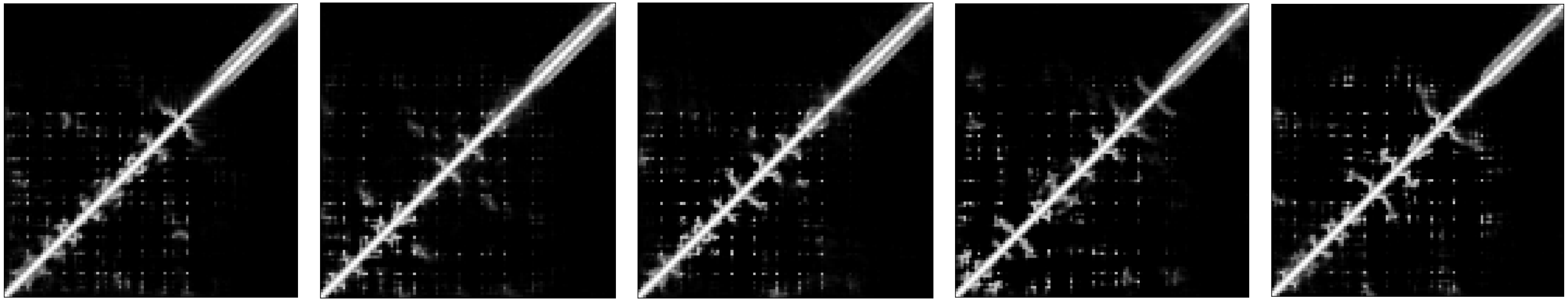
Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5

Cluster 6 Cluster 7 Cluster 8 Cluster 9 Cluster 10

Figure A3.1.1-5: Contact maps of N-term_MoPrP_P105S.

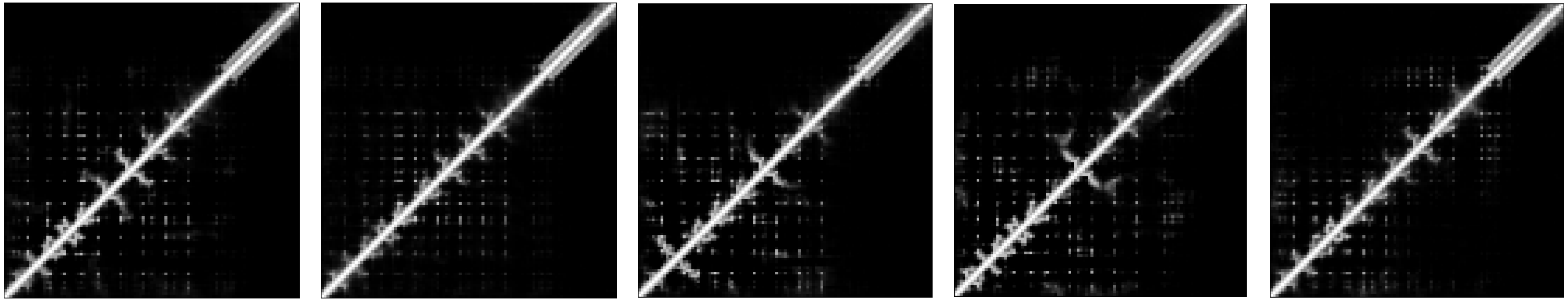


Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5

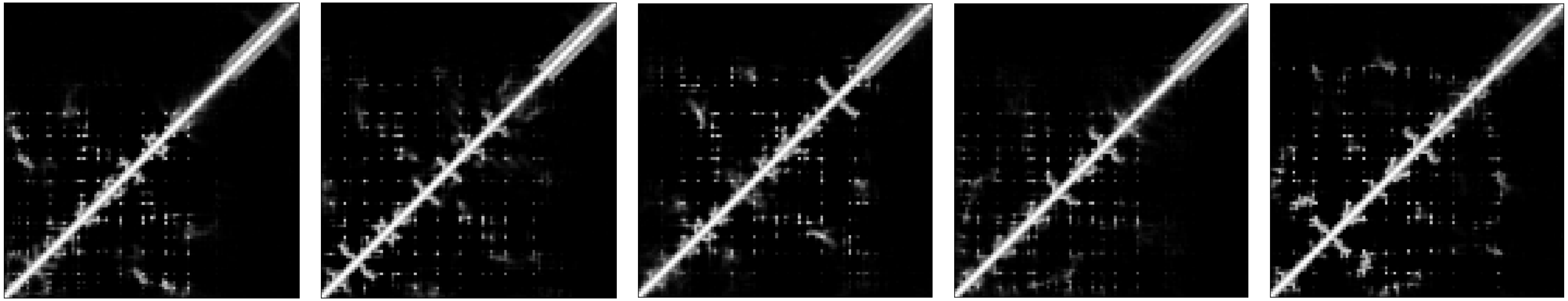


Cluster 6 Cluster 7 Cluster 8 Cluster 9 Cluster 10

Figure A3.1.1-6: Contact maps of N-term_MoPrP_P105T.

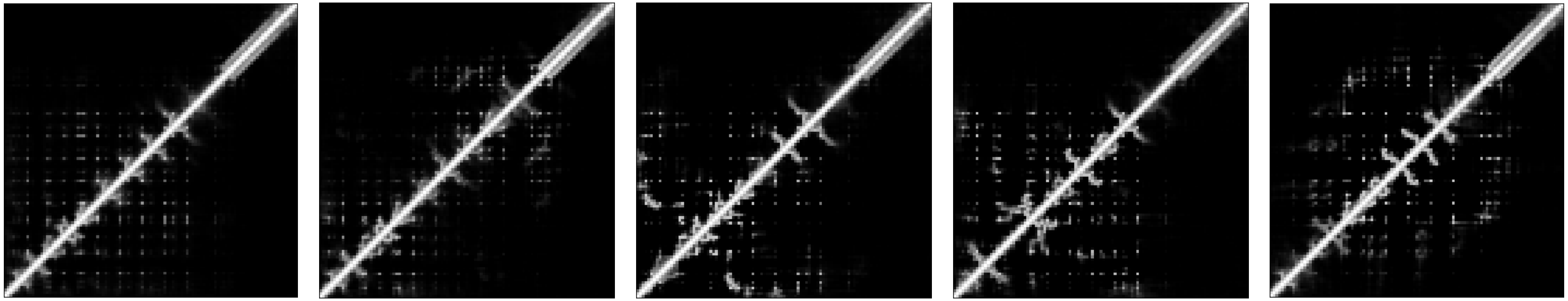


Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5

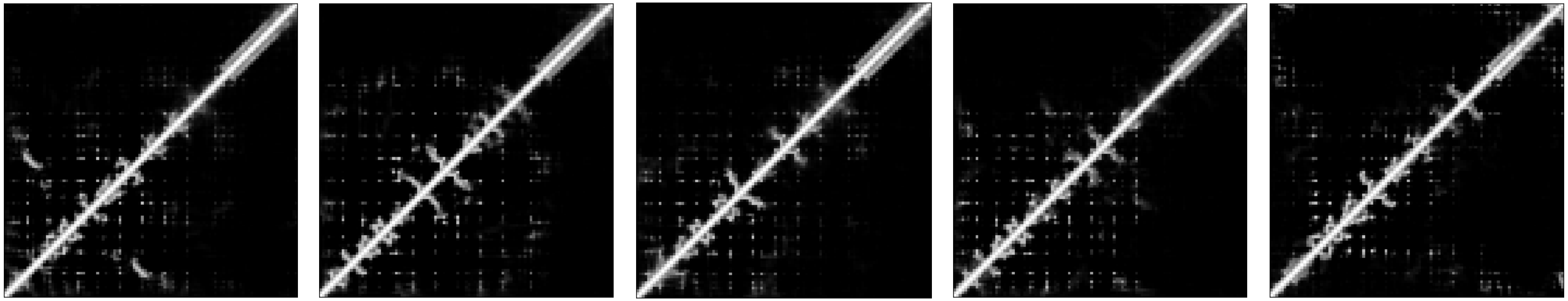


Cluster 6 Cluster 7 Cluster 8 Cluster 9 Cluster 10

Figure A3.1.1-7: Contact maps of N-term_MoPrP_G114V.



Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5



Cluster 6 Cluster 7 Cluster 8 Cluster 9 Cluster 10

Figure A3.1.1-8: Contact maps of N-term_MoPrP_A117V.

Table A3.1.1: Clustering based on contact maps of N-term_MoPrP and PMs conformations sampled from the production trajectories.

Cluster	Population	R_g (Å)[i]	Population	R_g (Å)	Population	R_g (Å)	Population	R_g (Å)
	N-term_MoPrP		N-term_MoPrP_Q52P		N-term_MoPrP_P102L		N-term_MoPrP_P105L	
1	17.1%	19.8±3.0	34.3%	16.5±1.3	19.0%	19.8±2.7	19.7%	17.8±2.5
2	13.1%	19.3±3.0	7.6%	19.6±2.7	13.5%	19.8±2.8	13.9%	20.2±2.9
3	12.2%	16.0±1.3	6.1%	15.8±0.8	11.7%	17.2±2.5	8.5%	20.29±3.10
4	5.6%	15.7±1.2	5.4%	20.1±2.6	10.1%	20.5±2.8	6.7%	20.0±2.4
5	5.5%	19.2±2.8	4.8%	19.4±2.9	5.4%	21.0±3.2	5.8%	20.2±2.9
6	5.4%	18.8±2.4	4.3%	14.7±1.5	3.0%	20.1±3.0	4.3%	20.0±3.0
7	3.7%	19.2±2.8	3.2%	16.8±2.2	2.2%	20.5±2.5	3.6%	20.0±2.9
8	2.9%	19.1±3.0	2.9%	16.0±1.7	2.1%	20.2±3.2	3.6%	18.6±2.6
9	2.6%	18.2±2.3	2.7%	16.6±1.4	2.0%	20.0±2.6	3.5%	17.0±1.3
10	2.6%	18.9±2.5	2.0%	16.1±1.2	1.8%	19.6±2.8	3.3%	20.9±2.7
Total	70.5%		73.1%		70.7%		72.7%	
	N-term_MoPrP_P105S		N-term_MoPrP_P105T		N-term_MoPrP_G114V		N-term_MoPrP_A117V	
1	42.4%	20.6±2.5	22.6%	20.1±3.0	16.0%	19.0±3.2	21.8%	19.8±3.0
2	5.0%	16.6±1.1	18.8%	19.9±3.0	10.7%	19.6±3.0	10.0%	18.4±2.4
3	3.9%	20.9±2.5	7.0%	19.9±3.1	9.2%	19.9±3.1	8.7%	19.7±3.2
4	3.7%	20.3±2.8	5.6%	18.6±2.4	8.2%	17.9±2.5	6.8%	19.5±3.2
5	3.5%	20.8±2.8	4.1%	19.7±3.0	7.9%	18.9±2.4	6.0%	17.2±1.4
6	3.3%	19.7±2.6	3.4%	20.0±2.8	5.6%	19.9±2.8	4.9%	19.2±3.5
7	3.0%	21.0±3.0	3.0%	19.0±2.2	4.1%	18.4±3.4	4.2%	17.7±2.2
8	20.%	21.2±2.5	2.4%	19.6±2.5	3.5%	16.4±1.2	3.6%	19.1±2.9
9	1.9%	19.1±1.9	2.2%	18.8±3.0	2.7%	19.6±2.7	2.8%	19.8±3.1
10	1.9%	20.8±2.5	2.0%	17.4±1.4	2.6%	17.3±3.1	2.7%	17.2±2.2
Total	70.5%		71.0%		70.3%		71.2%	

[i] Standard deviations (SDs) of R_g are indicated by "±".

Table A3.2: Four different hydrophobicity scales for amino acid residues in proteins. From the top to the bottom the hydrophobicity decreases gradually.

	Tanford[426]	Kauzmann[427]	Biswas et al.[428]	Charton *et al.*[429]
Most hydrophobic	Ile	Cys	Gly, Leu, Ile	Cys
	Val		Val, Ala	Ile
		Phe, Ile		Val
	Leu	Val	Phe	Leu, Phe
		Leu, Met, Trp	Cys	Met
	Phe		Met	Ala, Gly, Trp
	Cys			
	Met, Ala	His	Thr, Ser	
		Tyr	Trp, Tyr	His, Ser
	Gly	Ala		Thr
	Thr, Ser	Gly		Pro
	Trp, Tyr	Thr		Tyr
	Pro			Asn
			Asp, Lys, Gln	Asp
	His	Ser	Glu, His	Gln, Glu
	Asn, Gln	Pro, Arg	Asp	
	Asp, Glu	Asn		
	Lys	Gln, Asp, Glu		
				Arg
Most hydrophilic	Arg	Lys	Arg	Lys

Table A3.3: Sequence Identity (in %) between HuPrP and PrP of Mo, SHa and bovine (Bv) species. The residues involved in PMs in the N-term are fully conserved.

	Domain (residues in HuPrP)		
	Full-length (23-230)	N-term (23-124)	OR (51-91)
MoPrP	90	93	90
SHaPrP	90	98	98
BoPrP	90	89	100[i]

[i] BoPrP contains one more OR repeat than HuPrP but the residues corresponding to HuPrP OR are fully conserved.

Table A3.4: Average content of transient SS elements (in %[i]) in the N-term obtained by our simulation of the full-length MoPrP.

Residues	α-helix	β-sheet	β-bridge	β-turn	Bend	3_{10}-helix	π-helix	Coil
N-term (23-128)	19±8	8±6	7±3	27±5	11±4	4±3	0±2	24±5
23-30	4±3	3±2	6±3	17±5	10±3	2±2	1±1	56±9
31-50	8±4	7±6	12±5	32±6	12±4	5±3	0±1	24±5
51-91 (OR)	6±2	13±6	9±3	38±6	12±3	5±3	0±2	18±5
90-99	4±2	15±6	5±3	22±4	17±5	3±2	0±2	34±8
100-118	65±10	2±2	1±0	6±3	7±3	2±4	0±1	17±6
119-126	32±8	4±6	2±3	32±8	9±4	4±3	1±1	16±5

[i] SDs are indicated by "±".

List of Publications

Publications related to this thesis

The work reported here are published in the following articles. The "*" indicates that these authors contributed equally to the article.

- **Cong**, **X.**, Bongarzone, S., Giachin, G., Rossetti, G., Carloni, P. and Legname, G., Dominant-negative effects in prion diseases: insights from molecular dynamics simulations on mouse prion protein chimeras, *Journal of Biomolecular Structure and Dynamics* 31, p. 829-840 (2013).
- Rossetti, G.*, **Cong, X.***, Caliandro, R., Legname, G. and Carloni, P., Common Structural Traits across Pathogenic Mutants of the Human Prion Protein and Their Implications for Familial Prion Diseases, *Journal of Molecular Biology* 411, p. 700-712 (2011).
- **Cong**, **X.***, Casiraghi, N.*, Rossetti, G., Sandipan, M. Giachin, G., Legname, G. and Carloni, P., Role of prion disease-linked mutations in the intrinsically disordered N-terminal domain of the prion protein, *Journal of Chemical Theory and Computation*, in press.

Other publications

- Tan, J.*, **Cong, X.***, Liu, M., Chen, W., Wang, C., Lee, J. and Liang, X., Therapeutic Strategies Underpinning the Development of Novel Techniques for the Treatment of HIV Infection, *Drug Design Today* 15, p. 186-197 (2010).
- **Cong**, **X.**, Tan, J., Liu, M., Chen, W. and Wang, C., Computational Study of Binding Mode for N-substituted Pyrrole Derivatives to HIV-1 gp41, *Progress in Biochemistry and Biophysics* 37, p. 904-915 (2010).
- Liu, M., **Cong**, **X.**, Li, P., Tan, J., Chen, W., Wang, C. and Liu, C., Study on the inhibitory mechanism and binding mode of the hydroxycoumarin compound NSC158393 to HIV-1 integrase by molecular modeling, *Biopolymers,* 91, p. 700-709 (2009).
- Zhuo, Y., Kong, R., **Cong**, **X.**, Chen, W. and Wang, C., Three-dimensional QSAR analyses of 1,3,4-trisubstituted pyrrolidine-based CCR5 receptor inhibitors, *European Journal of Medicinal Chemistry*, 43, p. 2724-2734 (2008).

Bibliography

1. Prusiner, S. B. (1998). Prions. *Proc Natl Acad Sci U S A* **95**, 13363-83.
2. Caughey, B. (2003). Prion protein conversions: insight into mechanisms, TSE transmission barriers and strains. *British Medical Bulletin* **66**, 109-120.
3. Diaz-Espinoza, R. & Soto, C. (2010). Generation of prions in vitro and the protein-only hypothesis. *Prion* **4**, 1-7.
4. Prusiner, S. B. (2012). A Unifying Role for Prions in Neurodegenerative Diseases. *Science* **336**, 1511-1513.
5. Ashe, K. H. & Aguzzi, A. (2013). Prions, prionoids and pathogenic proteins in Alzheimer disease. *Prion* **7**, 55-9.
6. Mougenot, A. L., Nicot, S., Bencsik, A., Morignat, E., Verchere, J., Lakhdar, L., Legastelois, S. & Baron, T. (2012). Prion-like acceleration of a synucleinopathy in a transgenic mouse model. *Neurobiology of Aging* **33**, 2225-2228.
7. World Health Organization & Alzheimer's Disease International. (2012). *Dementia : a public health priority*, Geneva.
8. Bishop, M. T., Pennington, C., Heath, C. A., Will, R. G. & Knight, R. S. (2009). PRNP variation in UK sporadic and variant Creutzfeldt Jakob disease highlights genetic risk factors and a novel non-synonymous polymorphism. *BMC Med Genet* **10**, 146.
9. Bossers, A., Schreuder, B. E., Muileman, I. H., Belt, P. B. & Smits, M. A. (1996). PrP genotype contributes to determining survival times of sheep with natural scrapie. *Journal of General Virology* **77 (Pt 10)**, 2669-73.
10. Sigurdson, C. J., Nilsson, K. P. R., Hornemann, S., Manco, G., Fernandez-Borges, N., Schwarz, P., Castilla, J., Wuthrich, K. & Aguzzi, A. (2010). A molecular switch controls interspecies prion disease transmission in mice. *Journal of Clinical Investigation* **120**, 2590-2599.
11. Sigurdson, C. J., Nilsson, K. P. R., Hornemann, S., Heikenwalder, M., Manco, G., Schwarz, P., Ott, D., Rulicke, T., Liberski, P. P., Julius, C., Falsig, J., Stitz, L., Wuthrich, K. & Aguzzi, A. (2009). De novo generation of a transmissible spongiform encephalopathy by mouse transgenesis. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 304-309.

12. Sigurdson, C. J., Joshi-Barr, S., Bett, C., Winson, O., Manco, G., Schwarz, P., Rulicke, T., Nilsson, K. P. R., Margalith, I., Raeber, A., Peretz, D., Hornemann, S., Wuthrich, K. & Aguzzi, A. (2011). Spongiform Encephalopathy in Transgenic Mice Expressing a Point Mutation in the beta 2-alpha 2 Loop of the Prion Protein. *Journal of Neuroscience* **31**, 13840-13847.
13. Kaneko, K., Zulianello, L., Scott, M., Cooper, C. M., Wallace, A. C., James, T. L., Cohen, F. E. & Prusiner, S. B. (1997). 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**, 10069-74.
14. Beland, M. & Roucou, X. (2012). The prion protein unstructured N-terminal region is a broad-spectrum molecular sensor with diverse and contrasting potential functions. *Journal of Neurochemistry* **120**, 853-868.
15. Giachin, G., Biljan, I., Ilc, G., Plavec, J. & Legname, G. (2013). Probing Early Misfolding Events in Prion Protein Mutants by NMR Spectroscopy. *Molecules* **18**, 9451-9476.
16. van der Kamp, M. W. & Daggett, V. (2010). Pathogenic Mutations in the Hydrophobic Core of the Human Prion Protein Can Promote Structural Instability and Misfolding. *Journal of Molecular Biology* **404**, 732-748.
17. Rossetti, G., Giachin, G., Legname, G. & Carloni, P. (2010). Structural facets of disease-linked human prion protein mutants: A molecular dynamic study. *Proteins-Structure Function and Bioinformatics* **78**, 3270-3280.
18. Irback, A. & Mohanty, S. (2006). PROFASI: A Monte Carlo simulation package for protein folding and aggregation. *Journal of Computational Chemistry* **27**, 1548-55.
19. Jonsson, S. A., Mohanty, S. & Irback, A. (2012). Distinct phases of free alpha-synuclein--a Monte Carlo study. *Proteins-Structure Function and Bioinformatics* **80**, 2169-77.
20. Aguzzi, A., Baumann, F. & Bremer, J. (2008). The prion's elusive reason for being. *Annual Review of Neuroscience* **31**, 439-477.
21. Gajdusek, D. C., Gibbs, C. J. & Alpers, M. (1966). Experimental Transmission of a Kuru-Like Syndrome to Chimpanzees. *Nature* **209**, 794-&.

22. Hill, A. F., Desbruslais, M., Joiner, S., Sidle, K. C. L., Gowland, I., Collinge, J., Doey, L. J. & Lantos, P. (1997). The same prion strain causes vCJD and BSE. *Nature* **389**, 448-450.

23. Bruce, M. E., Will, R. G., Ironside, J. W., McConnell, I., Drummond, D., Suttie, A., McCardle, L., Chree, A., Hope, J., Birkett, C., Cousens, S., Fraser, H. & Bostock, C. J. (1997). Transmissions to mice indicate that 'new variant' CJD is caused by the BSE agent. *Nature* **389**, 498-501.

24. Collinge, J., Palmer, M. S., Sidle, K. C. L., Gowland, I., Medori, R., Ironside, J. & Lantos, P. (1995). Transmission of Fatal Familial Insomnia to Laboratory-Animals. *Lancet* **346**, 569-570.

25. Hsiao, K. K., Scott, M., Foster, D., Groth, D. F., Dearmond, S. J. & Prusiner, S. B. (1990). Spontaneous Neurodegeneration in Transgenic Mice with Mutant Prion Protein. *Science* **250**, 1587-1590.

26. Lasmezas, C. I., Deslys, J. P., Robain, O., Jaegly, A., Beringue, V., Peyrin, J. M., Fournier, J. G., Hauw, J. J., Rossier, J. & Dormont, D. (1997). Transmission of the BSE agent to mice in the absence of detectable abnormal prion protein. *Science* **275**, 402-405.

27. Bueler, H., Aguzzi, A., Sailer, A., Greiner, R. A., Autenried, P., Aguet, M. & Weissmann, C. (1993). Mice devoid of PrP are resistant to scrapie. *Cell* **73**, 1339-47.

28. Sailer, A., Bueler, H., Fischer, M., Aguzzi, A. & Weissmann, C. (1994). No Propagation of Prions in Mice Devoid of Prp. *Cell* **77**, 967-968.

29. Brandner, S., Isenmann, S., Raeber, A., Fischer, M., Sailer, A., Kobayashi, Y., Marino, S., Weissmann, C. & Aguzzi, A. (1996). Normal host prion protein necessary for scrapie-induced neurotoxicity. *Nature* **379**, 339-343.

30. Manson, J. C., Clarke, A. R., Mcbride, P. A., Mcconnell, I. & Hope, J. (1994). Prp Gene Dosage Determines the Timing but Not the Final Intensity or Distribution of Lesions in Scrapie Pathology. *Neurodegeneration* **3**, 331-340.

31. Collinge, J. & Clarke, A. R. (2007). A general model of prion strains and their pathogenicity. *Science* **318**, 930-936.

32. Schmitt-Ulms, G., Legname, G., Baldwin, M. A., Ball, H. L., Bradon, N., Bosque, P. J., Crossin, K. L., Edelman, G. M., DeArmond, S. J., Cohen, F. E. & Prusiner, S.

B. (2001). Binding of neural cell adhesion molecules (N-CAMs) to the cellular prion protein. *Journal of Molecular Biology* **314**, 1209-1225.

33. Chesebro, B., Trifilo, M., Race, R., Meade-White, K., Teng, C., LaCasse, R., Raymond, L., Favara, C., Baron, G., Priola, S., Caughey, B., Masliah, E. & Oldstone, M. (2005). Anchorless prion protein results in infectious amyloid disease without clinical scrapie. *Science* **308**, 1435-1439.
34. Solforosi, L., Criado, J. R., McGavern, D. B., Wirz, S., Sanchez-Alavez, M., Sugama, S., DeGiorgio, L. A., Volpe, B. T., Wiseman, E., Abalos, G., Masliah, E., Gilden, D., Oldstone, M. B., Conti, B. & Williamson, R. A. (2004). Cross-linking cellular prion protein triggers neuronal apoptosis in vivo. *Science* **303**, 1514-1516.
35. Sandberg, M. K., Al-Doujaily, H., Sharps, B., Clarke, A. R. & Collinge, J. (2011). Prion propagation and toxicity in vivo occur in two distinct mechanistic phases. *Nature* **470**, 540-542.
36. Aguzzi, A. & Falsig, J. (2012). PERSPECTIVE Prion propagation, toxicity and degradation. *Nature Neuroscience* **15**, 936-939.
37. van Rheede, T., Smolenaars, M. M., Madsen, O. & de Jong, W. W. (2003). Molecular evolution of the mammalian prion protein. *Mol Biol Evol* **20**, 111-21.
38. Chakrabarti, O., Ashok, A. & Hegde, R. S. (2009). Prion protein biosynthesis and its emerging role in neurodegeneration. *Trends in Biochemical Sciences* **34**, 287-295.
39. Brown, D. R. (2001). Prion and prejudice: normal protein and the synapse. *Trends in Neurosciences* **24**, 85-90.
40. Rudd, P. M., Wormald, M. R., Wing, D. R., Prusiner, S. B. & Dwek, R. A. (2001). Prion glycoprotein: structure, dynamics, and roles for the sugars. *Biochemistry* **40**, 3759-66.
41. Prusiner, S. B. (1998). Prions. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 13363-13383.
42. Ott, C. M., Akhavan, A. & Lingappa, V. R. (2007). Specific features of the prion protein transmembrane domain regulate nascent chain orientation. *Journal of Biological Chemistry* **282**, 11163-11171.

43. Hegde, R. S., Mastrianni, J. A., Scott, M. R., DeFea, K. A., Tremblay, P., Torchia, M., DeArmond, S. J., Prusiner, S. B. & Lingappa, V. R. (1998). A transmembrane form of the prion protein in neurodegenerative disease. *Science* **279**, 827-834.

44. Zanusso, G., Petersen, R. B., Jin, T. C., Jing, Y., Kanoush, R., Ferrari, S., Gambetti, P. & Singh, N. (1999). Proteasomal degradation and N-terminal protease resistance of the codon 145 mutant prion protein. *Journal of Biological Chemistry* **274**, 23396-23404.

45. Jin, T. C., Gu, Y. P., Zanusso, G., Sy, M. S., Kumar, A., Cohen, M., Gambetti, P. & Singh, N. (2000). The chaperone protein BiP binds to a mutant prion protein and mediates its degradation by the proteasome. *Journal of Biological Chemistry* **275**, 38699-38704.

46. Stewart, R. S. & Harris, D. A. (2003). Mutational analysis of topological determinants in prion protein (PrP) and measurement of transmembrane and cytosolic PrP during prion infection. *Journal of Biological Chemistry* **278**, 45960-45968.

47. Hegde, R. S., Tremblay, P., Groth, D., DeArmond, S. J., Prusiner, S. B. & Lingappa, V. R. (1999). Transmissible and genetic prion diseases share a common pathway of neurodegeneration. *Nature* **402**, 822-6.

48. Ma, J. Y., Wollmann, R. & Lindquist, S. (2002). Neurotoxicity and neurodegeneration when PrP accumulates in the cytosol. *Science* **298**, 1781-1785.

49. Ma, J. Y. & Lindquist, S. (2001). Wild-type PrP and a mutant associated with prion disease are subject to retrograde transport and proteasome degradation. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 14955-14960.

50. Chakrabarti, O. & Hegde, R. S. (2009). Functional Depletion of Mahogunin by Cytosolically Exposed Prion Protein Contributes to Neurodegeneration. *Cell* **137**, 1136-1147.

51. Aguzzi, A. & Steele, A. D. (2009). Prion Topology and Toxicity. *Cell* **137**, 994-996.

52. Hegde, R. S., Voigt, S. & Lingappa, V. R. (1998). Regulation of protein topology by trans-acting factors at the endoplasmic reticulum. *Mol Cell* **2**, 85-91.

53. Yost, C. S., Lopez, C. D., Prusiner, S. B., Myers, R. M. & Lingappa, V. R. (1990). Non-hydrophobic extracytoplasmic determinant of stop transfer in the prion protein. *Nature* **343**, 669-72.

54. Lopez, C. D., Yost, C. S., Prusiner, S. B., Myers, R. M. & Lingappa, V. R. (1990). Unusual topogenic sequence directs prion protein biogenesis. *Science* **248**, 226-9.

55. Zahn, R., Liu, A. Z., Luhrs, T., Riek, R., von Schroetter, C., Garcia, F. L., Billeter, M., Calzolai, L., Wider, G. & Wuthrich, K. (2000). NMR solution structure of the human prion protein. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 145-150.

56. Calzolai, L. & Zahn, R. (2003). Influence of pH on NMR structure and stability of the human prion protein globular domain. *Journal of Biological Chemistry* **278**, 35592-6.

57. Riek, R., Wider, G., Billeter, M., Hornemann, S., Glockshuber, R. & Wuthrich, K. (1998). Prion protein NMR structure and familial human spongiform encephalopathies. *Proc Natl Acad Sci U S A* **95**, 11667-72.

58. Gossert, A. D., Bonjour, S., Lysek, D. A., Fiorito, F. & Wuthrich, K. (2005). Prion protein NMR structures of elk and of mouse/elk hybrids. *Proc Natl Acad Sci U S A* **102**, 646-50.

59. James, T. L., Liu, H., Ulyanov, N. B., FarrJones, S., Zhang, H., Donne, D. G., Kaneko, K., Groth, D., Mehlhorn, I., Prusiner, S. B. & Cohen, F. E. (1997). Solution structure of a 142-residue recombinant prion protein corresponding to the infectious fragment of the scrapie isoform. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 10086-10091.

60. Lysek, D. A., Schorn, C., Nivon, L. G., Esteve-Moya, V., Christen, B., Calzolai, L., von Schroetter, C., Fiorito, F., Herrmann, T., Guntert, P. & Wuthrich, K. (2005). Prion protein NMR structures of cats, dogs, pigs, and sheep. *Proc Natl Acad Sci U S A* **102**, 640-5.

61. Garcia, F. L., Zahn, R., Riek, R. & Wuthrich, K. (2000). NMR structure of the bovine prion protein. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 8334-8339.

62. Hornemann, S., Schorn, C. & Wuthrich, K. (2004). NMR structure of the bovine prion protein isolated from healthy calf brains. *Embo Reports* **5**, 1159-1164.

63. Snow, A. D., Wight, T. N., Nochlin, D., Koike, Y., Kimata, K., Dearmond, S. J. & Prusiner, S. B. (1990). Immunolocalization of Heparan-Sulfate Proteoglycans to the Prion Protein Amyloid Plaques of Gerstmann-Straussler Syndrome, Creutzfeldt-Jakob Disease and Scrapie. *Laboratory Investigation* **63**, 601-611.

64. Leach, S. P., Salman, M. D. & Hamar, D. (2006). Trace elements and prion diseases: a review of the interactions of copper, manganese and zinc with the prion protein. *Anim Health Res Rev* **7**, 97-105.

65. Qin, K., Yang, Y., Mastrangelo, P. & Westaway, D. (2002). Mapping Cu(II) binding sites in prion proteins by diethyl pyrocarbonate modification and matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometric footprinting. *J Biol Chem* **277**, 1981-90.

66. Walter, E. D., Stevens, D. J., Visconte, M. P. & Millhauser, G. L. (2007). The prion protein is a combined zinc and copper binding protein: Zn2+ alters the distribution of Cu2+ coordination modes. *Journal of the American Chemical Society* **129**, 15440-+.

67. Watt, N. T. & Hooper, N. M. (2003). The prion protein and neuronal zinc homeostasis. *Trends Biochem Sci* **28**, 406-10.

68. Pushie, M. J., Pickering, I. J., Martin, G. R., Tsutsui, S., Jirik, F. R. & George, G. N. (2011). Prion protein expression level alters regional copper, iron and zinc content in the mouse brain. *Metallomics* **3**, 206-214.

69. Spevacek, A. R., Evans, E. G. B., Miller, J. L., Meyer, H. C., Pelton, J. G. & Millhauser, G. L. (2013). Zinc Drives a Tertiary Fold in the Prion Protein with Familial Disease Mutation Sites at the Interface. *Structure* **21**, 236-246.

70. Taubner, L. M., Bienkiewicz, E. A., Copie, V. & Caughey, B. (2010). Structure of the Flexible Amino-Terminal Domain of Prion Protein Bound to a Sulfated Glycan. *Journal of Molecular Biology* **395**, 475-490.

71. Kardos, J., Kovacs, I., Hajos, F., Kalman, M. & Simonyi, M. (1989). Nerve-Endings from Rat-Brain Tissue Release Copper Upon Depolarization - a Possible Role in Regulating Neuronal Excitability. *Neuroscience Letters* **103**, 139-144.

72. Wong, B. S., Chen, S. G., Colucci, M., Xie, Z. L., Pan, T., Liu, T., Li, R. L., Gambetti, P., Sy, M. S. & Brown, D. R. (2001). Aberrant metal binding by prion protein in human prion disease. *Journal of Neurochemistry* **78**, 1400-1408.

73. Wong, B. S., Brown, D. R., Pan, T., Whiteman, M., Liu, T., Bu, X. D., Li, R. L., Gambetti, P., Olesik, J., Rubenstein, R. & Sy, M. S. (2001). Oxidative impairment in scrapie-infected mice is associated with brain metals perturbations and altered antioxidant activities. *Journal of Neurochemistry* **79**, 689-698.

74. Thackray, A. M., Knight, R., Haswell, S. J., Bujdoso, R. & Brown, D. R. (2002). Metal imbalance and compromised antioxidant function are early changes in prion disease. *Biochemical Journal* **362**, 253-258.

75. Hijazi, N., Shaked, Y., Rosenmann, H., Ben-Hur, T. & Gabizon, R. (2003). Copper binding to PrPC may inhibit prion disease propagation. *Brain Research* **993**, 192-200.

76. Mitteregger, G., Korte, S., Shakarami, M., Herms, J. & Kretzschmar, H. A. (2009). Role of copper and manganese in prion disease progression. *Brain Research* **1292**, 155-164.

77. Quaglio, E., Chiesa, R. & Harris, D. A. (2001). Copper converts the cellular prion protein into a protease-resistant species that is distinct from scrapie isoform. *Journal of Biological Chemistry* **276**, 11432-11438.

78. Sigurdsson, E. M., Brown, D. R., Alim, M. A., Scholtzova, H., Carp, R., Meeker, H. C., Prelli, F., Frangione, B. & Wisniewski, T. (2003). Copper chelation delays the onset of prion disease. *Journal of Biological Chemistry* **278**, 46199-46202.

79. Nishida, Y. (2011). The chemical process of oxidative stress by copper(II) and iron(III) ions in several neurodegenerative disorders. *Monatshefte Fur Chemie* **142**, 375-384.

80. Klamt, F., Dal-Pizzol, F., Conte da Frota, M. L., Jr., Walz, R., Andrades, M. E., da Silva, E. G., Brentani, R. R., Izquierdo, I. & Fonseca Moreira, J. C. (2001). Imbalance of antioxidant defense in mice lacking cellular prion protein. *Free Radic Biol Med* **30**, 1137-44.

81. Encalada, S. E., Moya, K. L., Lehmann, S. & Zahn, R. (2008). The role of the prion protein in the molecular basis for synaptic plasticity and nervous system development. *Journal of Molecular Neuroscience* **34**, 9-15.

82. Herms, J., Tings, T., Gall, S., Madlung, A., Giese, A., Siebert, H., Schurmann, P., Windl, O., Brose, N. & Kretzschmar, H. (1999). Evidence of presynaptic location and function of the prion protein. *J Neurosci* **19**, 8866-75.

83. Pauly, P. C. & Harris, D. A. (1998). Copper stimulates endocytosis of the prion protein. *Journal of Biological Chemistry* **273**, 33107-33110.

84. Kuczius, T., Buschmann, A., Zhang, W., Karch, H., Becker, K., Peters, G. & Groschup, M. H. (2004). Cellular prion protein acquires resistance to proteolytic degradation following copper ion binding. *Biological Chemistry* **385**, 739-747.

85. Vassallo, N. & Herms, J. (2003). Cellular prion protein function in copper homeostasis and redox signalling at the synapse. *Journal of Neurochemistry* **86**, 538-544.

86. Kiachopoulos, S., Heske, J., Tatzelt, J. & Winklhofer, K. F. (2004). Misfolding of the prion protein at the plasma membrane induces endocytosis, intracellular retention and degradation. *Traffic* **5**, 426-436.

87. Orem, N. R., Geoghegan, J. C., Deleault, N. R., Kascsak, R. & Supattapone, S. (2006). Copper(II) ions potently inhibit purified PrPres amplification. *Journal of Neurochemistry* **96**, 1409-1415.

88. Brown, D. R. (1999). Prion protein expression aids cellular uptake and veratridine-induced release of copper. *Journal of Neuroscience Research* **58**, 717-725.

89. Kozlowski, H., Luczkowski, M. & Remelli, M. (2010). Prion proteins and copper ions. Biological and chemical controversies. *Dalton Trans* **39**, 6371-85.

90. Klewpatinond, M., Davies, P., Bowen, S., Brown, D. R. & Viles, J. H. (2008). Deconvoluting the Cu(2+) binding modes of full-length prion protein. *Journal of Biological Chemistry* **283**, 1870-1881.

91. Cereghetti, G. M., Schweiger, A., Glockshuber, R. & Van Doorslaer, S. (2001). Electron paramagnetic resonance evidence for binding of Cu(2+) to the C-terminal domain of the murine prion protein. *Biophysical Journal* **81**, 516-25.

92. Kramer, M. L., Kratzin, H. D., Schmidt, B., Romer, A., Windl, O., Liemann, S., Hornemann, S. & Kretzschmar, H. (2001). Prion protein binds copper within the physiological concentration range. *Journal of Biological Chemistry* **276**, 16711-16719.

93. Qin, K. F., Yang, Y., Mastrangelo, P. & Westaway, D. (2002). Mapping Cu(II) binding sites in prion proteins by diethyl pyrocarbonate modification and matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometric footprinting. *Journal of Biological Chemistry* **277**, 1981-1990.

94. Burns, C. S., Aronoff-Spencer, E., Legname, G., Prusiner, S. B., Antholine, W. E., Gerfen, G. J., Peisach, J. & Millhauser, G. L. (2003). Copper coordination in the full-length, recombinant prion protein. *Biochemistry* **42**, 6794-6803.

95. Burns, C. S., Aronoff-Spencer, E., Dunham, C. M., Lario, P., Avdievich, N. I., Antholine, W. E., Olmstead, M. M., Vrielink, A., Gerfen, G. J., Peisach, J., Scott, W. G. & Millhauser, G. L. (2002). Molecular features of the copper binding sites in the octarepeat domain of the prion protein. *Biochemistry* **41**, 3991-4001.

96. Klewpatinond, M. & Viles, J. H. (2007). Fragment length influences affinity for Cu(2+) and Ni(2+) binding to His(96) or His(111) of the prion protein and spectroscopic evidence for a multiple histidine binding only at low pH. *Biochemical Journal* **404**, 393-402.

97. Gogolla, N., Galimberti, I., DePaola, V. & Caroni, P. (2006). Preparation of organotypic hippocampal slice cultures for long-term live imaging. *Nature Protocols* **1**, 1165-1171.

98. Davies, P., Marken, F., Salter, S. & Brown, D. R. (2009). Thermodynamic and Voltammetric Characterization of the Metal Binding to the Prion Protein: Insights into pH Dependence and Redox Chemistry. *Biochemistry* **48**, 2610-2619.

99. Berti, F., Gaggelli, E., Guerrini, R., Janicka, A., Kozlowski, H., Legowska, A., Miecznikowska, H., Migliorini, C., Pogni, R., Remelli, M., Rolka, K., Valensin, D. & Valensin, G. (2007). Structural and dynamic characterization of copper(11) binding of the human prion protein outside the octarepeat region. *Chemistry-a European Journal* **13**, 1991-2001.

100. Osz, K., Nagy, Z., Pappalardo, G., Di Natale, G., Sanna, D., Micera, G., Rizzarelli, E. & Sovago, I. (2007). Copper(II) interaction with prion peptide fragments encompassing histidine residues within and outside the octarepeat domain: speciation, stability constants and binding details. *Chemistry* **13**, 7129-43.

101. Hasnain, S. S., Murphy, L. M., Strange, R. W., Grossmann, J. G., Clarke, A. R., Jackson, G. S. & Collinge, J. (2001). XAFS study of the high-affinity copper-binding site of human PrP(91-231) and its low-resolution structure in solution. *Journal of Molecular Biology* **311**, 467-73.

102. D'Angelo, P., Della Longa, S., Arcovito, A., Mancini, G., Zitolo, A., Chillemi, G., Giachin, G., Legname, G. & Benetti, F. (2012). Effects of the Pathological Q212P

Mutation on Human Prion Protein Non-Octarepeat Copper-Binding Site. *Biochemistry*.

103. Hajj, G. N. M., Lopes, M. H., Mercadante, A. F., Veiga, S. S., da Silveira, R. B., Santos, T. G., Ribeiro, K. C. B., Juliano, M. A., Jacchieri, S. G., Zanata, S. M. & Martins, V. R. (2007). Cellular prion protein interaction with vitronectin supports axonal growth and is compensated by integrins. *Journal of Cell Science* **120**, 1915-1926.
104. Hajj, G. N., Lopes, M. H., Mercadante, A. F., Veiga, S. S., da Silveira, R. B., Santos, T. G., Ribeiro, K. C., Juliano, M. A., Jacchieri, S. G., Zanata, S. M. & Martins, V. R. (2007). Cellular prion protein interaction with vitronectin supports axonal growth and is compensated by integrins. *Journal of Cell Science* **120**, 1915-26.
105. Zanata, S. M., Lopes, M. H., Mercadante, A. F., Hajj, G. N. M., Chiarini, L. B., Nomizo, R., Freitas, A. R. O., Cabral, A. L. B., Lee, K. S., Juliano, M. A., de Oliveira, E., Jachieri, S. G., Burlingame, A., Huang, L., Linden, R., Brentani, R. R. & Martins, V. R. (2002). Stress-inducible protein 1 is a cell surface ligand for cellular prion that triggers neuroprotection. *Embo Journal* **21**, 3307-3316.
106. Lauren, J., Gimbel, D. A., Nygaard, H. B., Gilbert, J. W. & Strittmatter, S. M. (2009). Cellular prion protein mediates impairment of synaptic plasticity by amyloid-beta oligomers. *Nature* **457**, 1128-U84.
107. Nicoll, A. J., Panico, S., Freir, D. B., Wright, D., Terry, C., Risse, E., Herron, C. E., O'Malley, T., Wadsworth, J. D., Farrow, M. A., Walsh, D. M., Saibil, H. R. & Collinge, J. (2013). Amyloid-beta nanotubes are associated with prion protein-dependent synaptotoxicity. *Nat Commun* **4**, 2416.
108. Chen, S. G., Yadav, S. P. & Surewicz, W. K. (2010). Interaction between Human Prion Protein and Amyloid-beta (A beta) Oligomers ROLE OF N-TERMINAL RESIDUES. *Journal of Biological Chemistry* **285**, 26377-26383.
109. Parkyn, C. J., Vermeulen, E. G., Mootoosamy, R. C., Sunyach, C., Jacobsen, C., Oxvig, C., Moestrup, S., Liu, Q., Bu, G., Jen, A. & Morris, R. J. (2008). LRP1 controls biosynthetic and endocytic trafficking of neuronal prion protein. *Journal of Cell Science* **121**, 773-83.

110. Santuccione, A., Sytnyk, V., Leshchyns'ka, I. & Schachner, M. (2005). Prion protein recruits its neuronal receptor NCAM to lipid rafts to activate p59fyn and to enhance neurite outgrowth. *J Cell Biol* **169**, 341-54.

111. Brandner, S., Raeber, A., Sailer, A., Blattler, T., Fischer, M., Weissmann, C. & Aguzzi, A. (1996). Normal host prion protein (PrPC) is required for scrapie spread within the central nervous system. *Proc Natl Acad Sci U S A* **93**, 13148-51.

112. Diaz-Espinoza, R. & Soto, C. (2012). High-resolution structure of infectious prion protein: the final frontier. *Nature Structural & Molecular Biology* **19**, 370-377.

113. DeMarco, M. L. & Daggett, V. (2004). From conversion to aggregation: Protofibril formation of the prion protein. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 2293-2298.

114. Wille, H., Bian, W., McDonald, M., Kendall, A., Colby, D. W., Bloch, L., Ollesch, J., Borovinskiy, A. L., Cohen, F. E., Prusiner, S. B. & Stubbs, G. (2009). Natural and synthetic prion structure from X-ray fiber diffraction. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 16990-16995.

115. Govaerts, C., Wille, H., Prusiner, S. B. & Cohen, F. E. (2004). Evidence for assembly of prions with left-handed beta 3-helices into trimers. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 8342-8347.

116. Cobb, N. J., Sonnichsen, F. D., Mchaourab, H. & Surewicz, W. K. (2007). Molecular architecture of human prion protein amyloid: A parallel, in-register beta-structure. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 18946-18951.

117. Solforosi, L., Milani, M., Mancini, N., Clementi, M. & Burioni, R. (2013). A closer look at prion strains Characterization and important implications. *Prion* **7**, 99-108.

118. Safar, J., Wille, H., Itrri, V., Groth, D., Serban, H., Torchia, M., Cohen, F. E. & Prusiner, S. B. (1998). Eight prion strains have PrPSc molecules with different conformations. *Nature Medicine* **4**, 1157-1165.

119. Fraser, H. & Dickinson, A. G. (1973). Scrapie in mice. Agent-strain differences in the distribution and intensity of grey matter vacuolation. *J Comp Pathol* **83**, 29-40.

120. Collinge, J., Sidle, K. C. L., Meads, J., Ironside, J. & Hill, A. F. (1996). Molecular analysis of prion strain variation and the aetiology of 'new variant' CJD. *Nature* **383**, 685-690.

121. Parchi, P., Castellani, R., Capellari, S., Ghetti, B., Young, K., Chen, S. G., Farlow, M., Dickson, D. W., Sima, A. A. F., Trojanowski, J. Q., Petersen, R. B. & Gambetti, P. (1996). Molecular basis of phenotypic variability in sporadic Creutzfeldt-Jakob disease. *Annals of Neurology* **39**, 767-778.

122. Khalili-Shirazi, A., Summers, L., Linehan, J., Mallinson, G., Anstee, D., Hawke, S., Jackson, G. S. & Collinge, J. (2005). PrP glycoforms are associated in a strain-specific ratio in native PrPSc. *J Gen Virol* **86**, 2635-44.

123. Wadsworth, J. D., Hill, A. F., Joiner, S., Jackson, G. S., Clarke, A. R. & Collinge, J. (1999). Strain-specific prion-protein conformation determined by metal ions. *Nat Cell Biol* **1**, 55-9.

124. Kuczius, T. & Groschup, M. H. (1999). Differences in proteinase K resistance and neuronal deposition of abnormal prion proteins characterize bovine spongiform encephalopathy (BSE) and scrapie strains. *Mol Med* **5**, 406-18.

125. Bessen, R. A. & Marsh, R. F. (1992). Biochemical and physical properties of the prion protein from two strains of the transmissible mink encephalopathy agent. *J Virol* **66**, 2096-101.

126. Telling, G. C., Parchi, P., DeArmond, S. J., Cortelli, P., Montagna, P., Gabizon, R., Mastrianni, J., Lugaresi, E., Gambetti, P. & Prusiner, S. B. (1996). Evidence for the conformation of the pathologic isoform of the prion protein enciphering and propagating prion diversity. *Science* **274**, 2079-82.

127. Gajdusek, D. C. & Gibbs, C. J., Jr. (1968). Slow, latent and temperate virus infections of the central nervous system. *Res Publ Assoc Res Nerv Ment Dis* **44**, 254-80.

128. Bruce, M., Chree, A., Mcconnell, I., Foster, J., Pearson, G. & Fraser, H. (1994). Transmission of Bovine Spongiform Encephalopathy and Scrapie to Mice - Strain Variation and the Species Barrier. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **343**, 405-411.

129. Telling, G. C., Scott, M., Hsiao, K. K., Foster, D., Yang, S. L., Torchia, M., Sidle, K. C. L., Collinge, J., Dearmond, S. J. & Prusiner, S. B. (1994). Transmission of Creutzfeldt-Jakob-Disease from Humans to Transgenic Mice Expressing Chimeric Human-Mouse Prion Protein. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 9936-9940.

130. Shibuya, S., Higuchi, J., Shin, R. W., Tateishi, J. & Kitamoto, T. (1998). Codon 219 Lys allele of PRNP is not found in sporadic Creutzfeldt-Jakob disease. *Ann Neurol* **43**, 826-8.

131. Soldevila, M., Calafell, F., Andres, A. M., Yague, J., Helgason, A., Stefansson, K. & Bertranpetit, J. (2003). Prion susceptibility and protective alleles exhibit marked geographic differences. *Hum Mutat* **22**, 104-5.

132. Sabuncu, E., Petit, S., Le Dur, A., Lan Lai, T., Vilotte, J. L., Laude, H. & Vilette, D. (2003). PrP polymorphisms tightly control sheep prion replication in cultured cells. *Journal of Virology* **77**, 2696-700.

133. Baylis, M. & Goldmann, W. (2004). The genetics of scrapie in sheep and goats. *Current Molecular Medicine* **4**, 385-396.

134. Collinge, J. (1999). Variant Creutzfeldt-Jakob disease. *Lancet* **354**, 317-323.

135. Donne, D. G., Viles, J. H., Groth, D., Mehlhorn, I., James, T. L., Cohen, F. E., Prusiner, S. B., Wright, P. E. & Dyson, H. J. (1997). Structure of the recombinant full-length hamster prion protein PrP(29-231): The N terminus is highly flexible. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 13452-13457.

136. Gossert, A. D., Bonjour, S., Lysek, D. A., Fiorito, F. & Wuthrich, K. (2005). Prion protein NMR structures of elk and of mouse/elk hybrids. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 646-650.

137. Christen, B., Perez, D. R., Hornemann, S. & Wuthrich, K. (2008). 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**, 306-312.

138. Christen, B., Hornemann, S., Damberger, F. F. & Wuthrich, K. (2009). Prion Protein NMR Structure from Tammar Wallaby (Macropus eugenii) Shows that the beta 2-alpha 2 Loop Is Modulated by Long-Range Sequence Effects. *Journal of Molecular Biology* **389**, 833-845.

139. Wen, Y., Li, J., Yao, W. M., Xiong, M. Q., Hong, J., Peng, Y., Xiao, G. F. & Lin, D. H. (2010). Unique Structural Characteristics of the Rabbit Prion Protein. *Journal of Biological Chemistry* **285**, 31682-31693.

140. Perez, D. R., Damberger, F. F. & Wuthrich, K. (2010). Horse Prion Protein NMR Structure and Comparisons with Related Variants of the Mouse Prion Protein. *Journal of Molecular Biology* **400**, 121-128.

141. Christen, B., Hornemann, S., Damberger, F. F. & Wuthrich, K. (2012). Prion protein mPrP[F175A](121-231): structure and stability in solution. *Journal of Molecular Biology* **423**, 496-502.

142. Race, R., Meade-White, K., Raines, A., Raymond, G. J., Caughey, B. & Chesebro, B. (2002). Subclinical scrapie infection in a resistant species: Persistence, replication, and adaptation of infectivity during four passages. *Journal of Infectious Diseases* **186**, S166-S170.

143. Mastrianni, J. A. (2010). The genetics of prion diseases. *Genet Med* **12**, 187-95.

144. Apetri, A. C., Vanik, D. L. & Surewicz, W. K. (2005). Polymorphism at residue 129 modulates the conformational conversion of the D178N variant of human prion protein 90-231. *Biochemistry* **44**, 15880-15888.

145. Apetri, A. C., Surewicz, K. & Surewicz, W. K. (2004). The effect of disease-associated mutations on the folding pathway of human prion protein. *J Biol Chem* **279**, 18008-14.

146. Swietnicki, W., Petersen, R. B., Gambetti, P. & Surewicz, W. K. (1998). Familial mutations and the thermodynamic stability of the recombinant human prion protein. *Journal of Biological Chemistry* **273**, 31048-31052.

147. Ilc, G., Giachin, G., Jaremko, M., Jaremko, L., Benetti, F., Plavec, J., Zhukov, I. & Legname, G. (2010). NMR structure of the human prion protein with the pathological Q212P mutation reveals unique structural features. *PLoS One* **5**, e11715.

148. Biljan, I., Ilc, G., Giachin, G., Plavec, J. & Legname, G. (2012). Structural Rearrangements at Physiological pH: Nuclear Magnetic Resonance Insights from the V210I Human Prion Protein Mutant. *Biochemistry* **51**, 7465-7474.

149. Biljan, I., Ilc, G., Giachin, G., Raspadori, A., Zhukov, I., Plavec, J. & Legname, G. (2011). Toward the Molecular Basis of Inherited Prion Diseases: NMR Structure of the Human Prion Protein with V210I Mutation. *Journal of Molecular Biology* **412**, 660-673.

150. Rossetti, G., Cong, X. J., Caliandro, R., Legname, G. & Carloni, P. (2011). Common Structural Traits across Pathogenic Mutants of the Human Prion Protein and Their Implications for Familial Prion Diseases. *Journal of Molecular Biology* **411**, 700-712.

151. Bamdad, K. & Naderi-Manesh, H. (2007). 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**, 719-724.

152. Meli, M., Gasset, M. & Colombo, G. (2011). Dynamic Diagnosis of Familial Prion Diseases Supports the beta 2-alpha 2 Loop as a Universal Interference Target. *PLoS One* **6**.

153. Liemann, S. & Glockshuber, R. (1999). Influence of amino acid substitutions related to inherited human prion diseases on the thermodynamic stability of the cellular prion protein. *Biochemistry* **38**, 3258-3267.

154. Schelzke, G., Stoeck, K., Eigenbrod, S., Grasbon-Frodl, E., Raddatz, L. M., Ponto, C., Kretzschmar, H. A. & Zerr, I. (2013). Report about Four Novel Mutations in the Prion Protein Gene. *Dementia and Geriatric Cognitive Disorders* **35**, 229-237.

155. Liu, Z., Jia, L., Piao, Y., Lu, D., Wang, F., Lv, H., Lu, Y. & Jia, J. (2010). Creutzfeldt-Jakob disease with PRNP G114V mutation in a Chinese family. *Acta Neurologica Scandinavica* **121**, 377-383.

156. van der Kamp, M. W. & Daggett, V. (2009). The consequences of pathogenic mutations to the human prion protein. *Protein Engineering Design & Selection* **22**, 461-468.

157. Barron, R. M., Thomson, V., King, D., Shaw, J., Melton, D. W. & Manson, J. C. (2003). Transmission of murine scrapie to P101L transgenic mice. *J Gen Virol* **84**, 3165-72.

158. Barron, R. M., Thomson, V., Jamieson, E., Melton, D. W., Ironside, J., Will, R. & Manson, J. C. (2001). Changing a single amino acid in the N-terminus of murine PrP alters TSE incubation time across three species barriers. *EMBO J* **20**, 5070-8.

159. Kovacs, G. G., Trabattoni, G., Hainfellner, J. A., Ironside, J. W., Knight, R. S. G. & Budka, H. (2002). Mutations of the prion protein gene - Phenotypic spectrum. *Journal of Neurology* **249**, 1567-1582.

160. Jones, E. M., Surewicz, K. & Surewicz, W. K. (2006). Role of N-terminal familial mutations in prion protein fibrillization and prion amyloid propagation in vitro. *J Biol Chem* **281**, 8190-6.

161. Lutz, J., Brabeck, C., Niemann, H. H., Kloz, U., Korth, C., Lingappa, V. R. & Burkle, A. (2010). Microdeletions within the hydrophobic core region of cellular prion protein alter its topology and metabolism. *Biochemical and Biophysical Research Communications* **393**, 439-444.

162. Mishra, R. S., Gu, Y. P., Bose, S., Verghese, S., Kalepu, S. & Singh, N. (2002). Cell surface accumulation of a truncated transmembrane prion protein in Gerstmann-Straussler-Scheinker disease P102L. *Journal of Biological Chemistry* **277**, 24554-24561.

163. Jansen, C., Parchi, P., Capellari, S., Vermeij, A. J., Corrado, P., Baas, F., Strammiello, R., van Gool, W. A., van Swieten, J. C. & Rozemuller, A. J. M. (2010). Prion protein amyloidosis with divergent phenotype associated with two novel nonsense mutations in PRNP. *Acta Neuropathologica* **119**, 189-197.

164. Revesz, T., Holton, J. L., Lashley, T., Plant, G., Frangione, B., Rostagno, A. & Ghiso, J. (2009). Genetics and molecular pathogenesis of sporadic and hereditary cerebral amyloid angiopathies (vol 118, pg 115, 2009). *Acta Neuropathologica* **118**, 321-321.

165. Beck, J. A., Mead, S., Campbell, T. A., Dickinson, A., Wientjens, D. P. M. W., Croes, E. A., Van Duijn, C. M. & Collinge, J. (2001). Two-octapeptide repeat deletion of prion protein associated with rapidly progressive dementia. *Neurology* **57**, 354-356.

166. Mead, S., Whitfield, J., Poulter, M., Shah, P., Uphill, J., Campbell, T., Al-Dujaily, H., Hummerich, H., Beck, J., Mein, C. A., Verzilli, C., Whittaker, J., Alpers, M. P. & Collinge, J. (2009). A Novel Protective Prion Protein Variant that Colocalizes with Kuru Exposure. *New England Journal of Medicine* **361**, 2056-2065.

167. Henzler-Wildman, K. & Kern, D. (2007). Dynamic personalities of proteins. *Nature* **450**, 964-972.

168. Malaney, P., Pathak, R. R., Xue, B., Uversky, V. N. & Dave, V. (2013). Intrinsic Disorder in PTEN and its Interactome Confers Structural Plasticity and Functional Versatility. *Sci Rep* **3**, 2035.

169. Schlick, T., Collepardo-Guevara, R., Halvorsen, L. A., Jung, S. & Xiao, X. (2011). Biomolecular modeling and simulation: a field coming of age. *Quarterly Reviews of Biophysics* **44**, 191-228.

170. Adcock, S. A. & McCammon, J. A. (2006). Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chemical Reviews* **106**, 1589-1615.

171. van Gunsteren, W. F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Geerke, D. P., Glattli, A., Hunenberger, P. H., Kastenholz, M. A., Ostenbrink, C., Schenk, M., Trzesniak, D., van der Vegt, N. F. A. & Yu, H. B. (2006). Biomolecular modeling: Goals, problems, perspectives. *Angewandte Chemie-International Edition* **45**, 4064-4092.

172. Mccammon, J. A., Gelin, B. R. & Karplus, M. (1977). Dynamics of Folded Proteins. *Nature* **267**, 585-590.

173. Arkhipov, A., Shan, Y. B., Das, R., Endres, N. F., Eastwood, M. P., Wemmer, D. E., Kuriyan, J. & Shaw, D. E. (2013). Architecture and Membrane Interactions of the EGF Receptor. *Cell* **152**, 557-569.

174. Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. (2011). How Fast-Folding Proteins Fold. *Science* **334**, 517-520.

175. Mackerell, A. D. (2004). Empirical force fields for biological macromolecules: Overview and issues. *Journal of Computational Chemistry* **25**, 1584-1604.

176. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995). A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society* **117**, 5179-5197.

177. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* **4**, 187-217.

178. Jorgensen, W. L. & Tiradorives, J. (1988). The Opls Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin. *Journal of the American Chemical Society* **110**, 1657-1666.

179. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). Stereochemistry of Polypeptide Chain Configurations. *Journal of Molecular Biology* **7**, 95-&.

180. Verlet, L. (1967). Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **159**, 98-&.

181. Hockney, R. W., Goel, S. P. & Eastwood, J. W. (1974). Quiet High-Resolution Computer Models of a Plasma. *Journal of Computational Physics* **14**, 148-158.

182. Van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **26**, 1701-1718.

183. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327-341.

184. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. (1997). LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463-1472.

185. Lague, P., Pastor, R. W. & Brooks, B. R. (2004). Pressure-based long-range correction for Lennard-Jones interactions in molecular dynamics simulations: Application to alkanes and interfaces. *Journal of Physical Chemistry B* **108**, 363-368.

186. Allen, M. P. & Tildesley, D. J. (1987). *Computer simulation of liquids*. Oxford science publications, Clarendon Pr., Oxford.

187. Ewald, P. P. (1921). Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **369**, 253-287.

188. Darden, T., York, D. & Pedersen, L. (1993). Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics* **98**, 10089-10092.

189. Smith, E. R. (1981). Electrostatic Energy in Ionic Crystals. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **375**, 475-505.

190. Deleeuw, S. W., Perram, J. W. & Smith, E. R. (1980). Simulation of Electrostatic Systems in Periodic Boundary-Conditions .1. Lattice Sums and Dielectric-

Constants. *Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences* **373**, 27-56.

191. Hockney, R. W. & Eastwood, J. W. (1981). *Computer Simulation using particles*, McGraw-Hill, New York, NY.
192. Kittel, C. (1971). *Introduction to solid-state physics*. 4. ed edit, Wiley, New York [u.a.].
193. Nose, S. (1984). A Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Molecular Physics* **52**, 255-268.
194. Hoover, W. G. (1985). Canonical Dynamics - Equilibrium Phase-Space Distributions. *Physical Review A* **31**, 1695-1697.
195. Andersen, H. C. (1980). Molecular-Dynamics Simulations at Constant Pressure and-or Temperature. *Journal of Chemical Physics* **72**, 2384-2393.
196. Parrinello, M. & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52**, 7182-7190.
197. Lazaridis, T. & Karplus, M. (2003). Thermodynamics of protein folding: a microscopic view. *Biophysical Chemistry* **100**, 367-395.
198. Feig, M. & Brooks, C. L. (2004). Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Current Opinion in Structural Biology* **14**, 217-224.
199. Onufriev, A. (2008). Chapter 7 Implicit Solvent Models in Molecular Dynamics Simulations: A Brief Overview. In *Annu Rep Comput Chem* (Ralph, A. W. & David, C. S., eds.), Vol. Volume 4, pp. 125-137. Elsevier.
200. Lin, J. H., Baker, N. A. & McCammon, J. A. (2002). Bridging implicit and explicit solvent approaches for membrane electrostatics. *Biophysical Journal* **83**, 1374-1379.
201. Edwards, S., Corry, B., Kuyucak, S. & Chung, S. H. (2002). Continuum electrostatics fails to describe ion permeation in the gramicidin channel. *Biophysical Journal* **83**, 1348-1360.
202. Simonson, T., Carlsson, J. & Case, D. A. (2004). Proton Binding to Proteins: pKa Calculations with Explicit and Implicit Solvent Models. *Journal of the American Chemical Society* **126**, 4167-4180.

203. Levy, Y. & Onuchic, J. N. (2006). Water mediation in protein folding and molecular recognition. *Annual Review of Biophysics and Biomolecular Structure* **35**, 389-415.

204. Levy, R. M. & Gallicchio, E. (1998). Computer simulations with explicit solvent: Recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects. *Annual Review of Physical Chemistry* **49**, 531-567.

205. Shea, J. E., Onuchic, J. N. & Brooks, C. L. (2002). Probing the folding free energy landscape of the src-SH3 protein domain. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 16064-16068.

206. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **79**, 926-935.

207. Mahoney, M. W. & Jorgensen, W. L. (2000). A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *Journal of Chemical Physics* **112**, 8910-8922.

208. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. (1987). The Missing Term in Effective Pair Potentials. *Journal of Physical Chemistry* **91**, 6269-6271.

209. Horn, H. W., Swope, W. C., Pitera, J. W., Madura, J. D., Dick, T. J., Hura, G. L. & Head-Gordon, T. (2004). Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *Journal of Chemical Physics* **120**, 9665-9678.

210. Kiss, P. T. & Baranyai, A. (2009). Clusters of classical water models. *Journal of Chemical Physics* **131**.

211. Vega, C. & Abascal, J. L. F. (2005). Relation between the melting temperature and the temperature of maximum density for the most common models of water. *Journal of Chemical Physics* **123**.

212. Frenkel, D. & Smit, B. (2008). *Understanding molecular simulation : from algorithms to applications*. 2. edit. Computational science series, Acad. Press, San Diego, Calif. [u.a.].

213. Northrup, S. H. & Mccammon, J. A. (1980). Simulation Methods for Protein-Structure Fluctuations. *Biopolymers* **19**, 1001-1016.

214. Jonsson, S. A. E., Mohanty, S. & Irback, A. (2011). Accelerating atomic-level protein simulations by flat-histogram techniques. *Journal of Chemical Physics* **135**.

215. Dinner, A. R. (2000). Local deformations of polymers with nonplanar rigid main-chain internal coordinates. *Journal of Computational Chemistry* **21**, 1132-1144.

216. Ulmschneider, J. P. & Jorgensen, W. L. (2003). Monte Carlo backbone sampling for polypeptides with variable bond angles and dihedral angles using concerted rotations and a Gaussian bias. *Journal of Chemical Physics* **118**, 4261-4271.

217. Vitalis, A. & Pappu, R. V. (2009). Methods for Monte Carlo simulations of biomacromolecules. *Annu Rep Comput Chem* **5**, 49-76.

218. Hu, J., Ma, A. & Dinner, A. R. (2006). Monte Carlo simulations of biomolecules: The MC module in CHARMM. *Journal of Computational Chemistry* **27**, 203-216.

219. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics* **21**, 1087-1092.

220. Irback, A., Mitternacht, S. & Mohanty, S. (2009). An effective all-atom potential for proteins. *PMC Biophys* **2**, 2.

221. Brändén, C. & Tooze, J. (1991). *Introduction to protein structure*, Garland Publ., New York, NY [u.a.].

222. Hovmoller, S., Zhou, T. & Ohlson, T. (2002). Conformations of amino acids in proteins. *Acta Crystallographica Section D-Biological Crystallography* **58**, 768-776.

223. Tsai, J., Taylor, R., Chothia, C. & Gerstein, M. (1999). The packing density in proteins: Standard radii and volumes. *Journal of Molecular Biology* **290**, 253-266.

224. Favrin, G., Irback, A. & Sjunnesson, F. (2001). Monte Carlo update for chain molecules: Biased Gaussian steps in torsional space. *Journal of Chemical Physics* **114**, 8154-8158.

225. Lal, M. (1969). Monte Carlo Computer Simulation of Chain Molecules .I. *Molecular Physics* **17**, 57-&.

226. Madras, N. & Sokal, A. D. (1988). The Pivot Algorithm - a Highly Efficient Monte-Carlo Method for the Self-Avoiding Walk. *Journal of Statistical Physics* **50**, 109-186.

227. Lyubartsev, A. P., Martsinovski, A. A., Shevkunov, S. V. & Vorontsovvelyaminov, P. N. (1992). New Approach to Monte-Carlo Calculation of the Free-Energy - Method of Expanded Ensembles. *Journal of Chemical Physics* **96**, 1776-1783.

228. Hukushima, K. & Nemoto, K. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *Journal of the Physical Society of Japan* **65**, 1604.

229. Zuckerman, D. M. (2011). Equilibrium sampling in biomolecular simulations. *Annu Rev Biophys* **40**, 41-62.

230. Roitberg, A., Symposium on Enhanced Sampling Techniques in Molecular Dynamics and Monte Carlo Simulations & American Chemical Society Division of Computers in Chemistry. (2004). *Special issue: conformational sampling : [... Symposium on "Enhanced Sampling Techniques in Molecular Dynamics and Monte Carlo Simulations" held within the Computers in Chemistry Division of the American Chemical Society, in their 223rd annual meeting ... in Orlando, FL, in April 2002].* Journal of molecular graphics and modelling, Elsevier, New York, NY.

231. Tai, K. (2004). Conformational sampling for the impatient. *Biophysical Chemistry* **107**, 213-20.

232. Smith, L. J., Daura, X. & van Gunsteren, W. F. (2002). Assessing equilibration and convergence in biomolecular simulations. *Proteins-Structure Function and Bioinformatics* **48**, 487-96.

233. Lei, H. & Duan, Y. (2007). Improved sampling methods for molecular simulation. *Curr Opin Struct Biol* **17**, 187-91.

234. Berne, B. J. & Straub, J. E. (1997). Novel methods of sampling phase space in the simulation of biological systems. *Curr Opin Struct Biol* **7**, 181-9.

235. Hansmann, U. H. & Okamoto, Y. (1999). New Monte Carlo algorithms for protein folding. *Curr Opin Struct Biol* **9**, 177-83.

236. Zhou, R. (2007). Replica exchange molecular dynamics method for protein folding simulation. *Methods Mol Biol* **350**, 205-23.

237. Okabe, T., Kawata, M., Okamoto, Y. & Mikami, M. (2001). Replica-exchange Monte Carlo method for the isobaric-isothermal ensemble. *Chemical Physics Letters* **335**, 435-439.

238. Rathore, N., Chopra, M. & de Pablo, J. J. (2005). Optimal allocation of replicas in parallel tempering simulations. *Journal of Chemical Physics* **122**.

239. Kofke, D. A. (2002). On the acceptance probability of replica-exchange Monte Carlo trials. *Journal of Chemical Physics* **117**, 6911-6914.

240. Patriksson, A. & van der Spoel, D. (2008). A temperature predictor for parallel tempering simulations. *Physical Chemistry Chemical Physics* **10**, 2073-2077.

241. Meyer, R. K., McKinley, M. P., Bowman, K. A., Braunfeld, M. B., Barry, R. A. & Prusiner, S. B. (1986). Separation and properties of cellular and scrapie prion proteins. *Proc Natl Acad Sci USA* **83**, 2310-4.

242. Oesch, B., Westaway, D., Wälchli, M., McKinley, M., Kent, S., Aebersold, R., Barry, R., Tempst, P., Teplow, D., Hood, L., Prusiner, S. & Weissmann, C. (1985). A cellular gene encodes scrapie PrP 27-30 protein. *Cell* **40**, 735-746.

243. Pan, K. M., Baldwin, M., Nguyen, J., Gasset, M., Serban, A., Groth, D., Mehlhorn, I., Huang, Z., Fletterick, R. J. & Cohen, F. E. (1993). Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *Proc Natl Acad Sci USA* **90**, 10962-6.

244. Nazabal, A., Hornemann, S., Aguzzi, A. & Zenobi, R. (2009). Hydrogen/deuterium exchange mass spectrometry identifies two highly protected regions in recombinant full-length prion protein amyloid fibrils. *J Mass Spectrom* **44**, 965-977.

245. Govaerts, C., Wille, H., Prusiner, S. & Cohen, F. (2004). Evidence for assembly of prions with left-handed beta 3-helices into trimers. *Proc Natl Acad Sci USA* **101**, 8342-8347.

246. DeMarco, M. L. & Daggett, V. (2004). From conversion to aggregation: protofibril formation of the prion protein. *Proc Natl Acad Sci USA* **101**, 2293-2298.

247. Sim, V. & Caughey, B. (2009). Ultrastructures and strain comparison of under-glycosylated scrapie prion fibrils. *Neurobiol Aging* **30**, 2031-2042.

248. Stoehr, J., Weinmann, N., Wille, H., Kaimann, T., Nagel-Steger, L., Birkmann, E., Panza, G., Prusiner, S. B., Eigen, M. & Riesner, D. (2008). Mechanisms of prion protein assembly into amyloid. *Proc Natl Acad Sci USA* **105**, 2409-2414.

249. Zahn, R., Liu, A. Z., Luhrs, T., Riek, R., von Schroetter, C., Garcia, F. L., Billeter, M., Calzolai, L., Wider, G. & Wuthrich, K. (2000). NMR solution structure of the human prion protein. *Proc Natl Acad Sci USA* **97**, 145-150.

250. Calzolai, L. & Zahn, R. (2003). Influence of pH on NMR structure and stability of the human prion protein globular domain. *J Biol Chem* **278**, 35592-35596.

251. Costantini, S. & Facchiano, A. M. (2009). Human prion protein helices: studying their stability by molecular dynamics simulations. *Protein Pept Lett* **16**, 1057-62.

252. Ji, H.-F. & Zhang, H.-Y. (2007). A comparative molecular dynamics study on thermostability of human and chicken prion proteins. *Biochem Biophys Res Commun* **359**, 790-4.

253. Blinov, N., Berjanskii, M., Wishart, D. S. & Stepanova, M. (2009). Structural domains and main-chain flexibility in prion proteins. *Biochemistry-Us* **48**, 1488-97.

254. Pappalardo, M., Milardi, D., Grasso, D. & La Rosa, C. (2007). Steered molecular dynamics studies reveal different unfolding pathways of prions from mammalian and non-mammalian species. *New J Chem* **31**, 901-905.

255. Kuwata, K., Kamatari, Y. O., Akasaka, K. & James, T. L. (2004). Slow conformational dynamics in the hamster prion protein. *Biochemistry* **43**, 4439-4446.

256. Kuwata, K., Li, H., Yamada, H., Legname, G., Prusiner, S. B., Akasaka, K. & James, T. L. (2002). Locally disordered conformer of the hamster prion protein: A crucial intermediate to PrPSc? *Biochemistry* **41**, 12277-12283.

257. Lu, X., Wintrode, P. L. & Surewicz, W. K. (2007). Beta-sheet core of human prion protein amyloid fibrils as determined by hydrogen/deuterium exchange. *Proc Natl Acad Sci USA* **104**, 1510-1515.

258. Adrover, M., Pauwels, K., Pringent, S., De Chiara, C., Xu, Z., Chapuis, C., Pastore, A. & Rezaei, H. (2010). Prion fibrillization is mediated by a native structural element which comprises the helices H2 and H3. *J Biol Chem*, 1-20.

259. Bae, S. H., Legname, G., Serban, A., Prusiner, S. B., Wright, P. E. & Dyson, H. J. (2009). Prion Proteins with Pathogenic and Protective Mutations Show Similar Structure and Dynamics. *Biochemistry* **48**, 8120-8128.

260. Zhang, Y. B., Swietnicki, W., Zagorski, M. G., Surewicz, W. K. & Sonnichsen, F. D. (2000). Solution structure of the E200K variant of human prion protein - Implications for the mechanism of pathogenesis in familial prion diseases. *J Biol Chem* **275**, 33650-33654.

261. Chebaro, Y. & Derreumaux, P. (2009). The Conversion of Helix H2 to beta-Sheet Is Accelerated in the Monomer and Dimer of the Prion Protein upon T183A Mutation. *J Phys Chem B* **113**, 6942-6948.

262. Chen, W., van der Kamp, M. W. & Daggett, V. (2010). Diverse effects on the native beta-sheet of the human prion protein due to disease-associated mutations. *Biochemistry* **49**, 9874-81.

263. Rossetti, G., Giachin, G., Legname, G. & Carloni, P. (2010). Structural facets of disease-linked human prion protein mutants: A molecular dynamic study. *Proteins Struct Funct Bioinf* **78**, 3270-3280.

264. Hirschberger, T., Stork, M., Schropp, B., Winklhofer, K. F., Tatzelt, J. & Tavan, P. (2006). Structural instability of the prion protein upon M205S/R mutations revealed by molecular dynamics simulations. *Biophys J* **90**, 3908-18.

265. Santini, S., Claude, J.-B., Audic, S. & Derreumaux, P. (2003). Impact of the tail and mutations G131V and M129V on prion protein flexibility. *Proteins* **51**, 258-65.

266. Santini, S. & Derreumaux, P. (2004). 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**, 951-60.

267. Apetri, A. C. & Surewicz, K. (2004). The effect of disease-associated mutations on the folding pathway of human prion protein. *J Biol Chem* **279**, 18008-18014.

268. Zhang, Y., Dai, L., Iwamoto, M. & Ouyang, Z. (2006). Molecular dynamics study on the conformational transition of prion induced by the point mutation: F198S. *Thin Solid Films* **499**, 224-228.

269. Bamdad, K. & Naderi-Manesh, H. (2007). Contribution of a putative salt bridge and backbone dynamics in the structural instability of human prion protein upon R208H mutation. *Biochem Biophys Res Commun* **364**, 719-24.

270. Sekijima, M., Motono, C., Yamasaki, S., Kaneko, K. & Akiyama, Y. (2003). Molecular dynamics simulation of dimeric and monomeric forms of human prion protein: Insight into dynamics and properties. *Biophys J* **85**, 1176-1185.

271. Levy, Y. & Becker, O. M. (2002). Conformational polymorphism of wild-type and mutant prion proteins: Energy landscape analysis. *Proteins* **47**, 458-68.

272. Alonso, D. O., DeArmond, S. J., Cohen, F. E. & Daggett, V. (2001). Mapping the early steps in the pH-induced conformational conversion of the prion protein. *Proc Natl Acad Sci USA* **98**, 2985-2989.

273. Bujdoso, R., Burke, D. F. & Thackray, A. M. (2005). Structural differences between allelic variants of the ovine prion protein revealed by molecular dynamics simulations. *Proteins* **61**, 840-9.

274. Alzualde, A., Indakoetxea, B., Ferrer, I., Moreno, F., Barandiaran, M., Gorostidi, A., Estanga, A., Ruiz, I., Calero, M., van Leeuwen, F. W., Atares, B., Juste, R.,

Rodriguez-Martinez, A. B. & Lopez de Munain, A. (2010). A novel PRNP Y218N mutation in Gerstmann-Straussler-Scheinker disease with neurofibrillary degeneration. *J Neuropathol Exp Neurol* **69**, 789-800.

275. Kong, Q., Surewicz, W. K., Petersen, R. B., Zou, W., Chen, S. G., Gambetti, P., Parchi, P., Capellari, S., Goldfarb, L., Montagna, P., Lugaresi, E., Piccardo, P. & Ghetti, B. (2004). *Inherited Prion Diseases*. 2nd Ed. edit. 2004, 14, Cold Spring Harbor.

276. Surewicz, K., Jones, E. M. & Apetri, A. C. (2006). The emerging principles of mammalian prion propagation and transmissibility barriers: Insight from studies in vitro. *Accounts Chem Res* **39**, 654-662.

277. Swietnicki, W., Petersen, R. B., Gambetti, P. & Surewicz, W. K. (1998). Familial mutations and the thermodynamic stability of the recombinant human prion protein. *J Biol Chem* **273**, 31048-52.

278. Horiuchi, M. & Caughey, B. (1999). Prion protein interconversions and the transmissible spongiform encephalopathies. *Structure* **7**, R231-40.

279. Lee, S., Antony, L., Hartmann, R., Knaus, K. J., Surewicz, K. & Yee, V. C. (2009). Conformational diversity in prion protein variants influences intermolecular &beta;-sheet formation. *EMBO J*, 1-12.

280. Zhang, Y., Dai, L. R., Iwamoto, M. & Ou-Yang, Z. C. (2006). Molecular dynamics study on the conformational transition of prion induced by the point mutation: F198S. *Thin Solid Films* **499**, 224-228.

281. van der Kamp, M. W. & Daggett, V. (2010). Pathogenic Mutations in the Hydrophobic Core of the Human Prion Protein Can Promote Structural Instability and Misfolding. *J Mol Biol* **404**, 732-748.

282. Barducci, A., Chelli, R., Procacci, P., Schettino, V., Gervasio, F. L. & Parrinello, M. (2006). Metadynamics simulation of prion protein: beta-structure stability and the early stages of misfolding. *J Am Chem Soc* **128**, 2705-10.

283. Zhang, J. (2010). Studies on the structural stability of rabbit prion probed by molecular dynamics simulations of its wild-type and mutants. *J Theor Biol* **264**, 119-22.

284. Shamsir, M. S. & Dalby, A. R. (2005). One gene, two diseases and three conformations: molecular dynamics simulations of mutants of human prion protein at room temperature and elevated temperatures. *Proteins* **59**, 275 - 290.

285. Sakudo, A., Xue, G., Kawashita, N., Ano, Y., Takagi, T., Shintani, H., Tanaka, Y., Onodera, T. & Ikuta, K. (2010). Structure of the prion protein and its gene: an analysis using bioinformatics and computer simulation. *Curr Protein Pept Sci* **11**, 166-79.

286. Kovacs, G. G., Puopolo, M., Ladogana, A., Pocchiari, M., Budka, H., van Duijn, C., Collins, S. J., Boyd, A., Giulivi, A., Coulthart, M., Delasnerie-Laupretre, N., Brandel, J. P., Zerr, I., Kretzschmar, H. A., de Pedro-Cuesta, J., Calero-Lara, M., Glatzel, M., Aguzzi, A., Bishop, M., Knight, R., Belay, G., Will, R., Mitrova, E. & Eurocjd. (2005). Genetic prion disease: the EUROCJD experience. *Hum Genet* **118**, 166-74.

287. Capellari, S., Strammiello, R., Saverioni, D., Kretzschmar, H. & Parchi, P. (2010). Genetic Creutzfeldt–Jakob disease and fatal familial insomnia: insights into phenotypic variability and disease pathogenesis. *Acta Neuropathol.* **121**, 21-37.

288. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B. & Woods, R. J. (2005). The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668-1688.

289. Wang, J., Cieplak, P. & Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem* **21**, 1049-1074.

290. Biljan, I., Giachin, G., Ilc, G., Zhukov, I., Plavec, J. & Legname, G. (2012). Structural basis for the protective effect of the human prion protein carrying the dominant-negative E219K polymorphism. *Biochemical Journal* **446**, 243-251.

291. Sutcliffe, M. J. (1993). Representing an Ensemble of Nmr-Derived Protein Structures by a Single Structure. *Protein Science* **2**, 936-944.

292. Micheletti, C., Seno, F. & Maritan, A. (2000). Recurrent oligomers in proteins: An optimal scheme reconciling accurate and concise backbone representations in automated folding and design studies. *Proteins-Structure Function and Bioinformatics* **40**, 662-674.

293. Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**, 2714-2723.

294. Campos, S. R. R., Machuqueiro, M. & Baptista, A. M. (2010). Constant-pH Molecular Dynamics Simulations Reveal a beta-Rich Form of the Human Prion Protein. *Journal of Physical Chemistry B* **114**, 12692-12700.

295. Aaqvist, J. (1990). Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* **94**, 8021-8024.

296. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R. & Klein, M. (1983). Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926-935.

297. Darden, T., York, D. & Pedersen, L. (1993). Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089-10092.

298. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* **4**, 435-447.

299. Nose, S. & Klein, M. (1983). Constant pressure molecular dynamics for molecular systems. *Mol Phys* **50**, 1055-1076.

300. Andersen, H. (1980). Molecular dynamics simulations at constant pressure and/or temperature. *J Chem Phys* **72**, 2384-2393.

301. Parrinello, M. & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* **52**, 7182-7190.

302. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun* **91**, 43-56.

303. Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W. F. & Mark, A. E. (1999). Peptide folding: When simulation meets experiment. *Angewandte Chemie-International Edition* **38**, 236-240.

304. Fisher, N. (1993). *Statistical Analysis of Circular Data*, Cambridge University Press.

305. Caliandro, R., Rossetti, G. & Carloni, P. (2012). Local Fluctuations and Conformational Transitions in Proteins. *Journal of Chemical Theory and Computation* **8**, 4775-4785.

306. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph* **14**, 33.

307. Yong, Y., Dai, L. R., Iwamoto, M. & Ou-Yang, Z. C. (2006). Molecular dynamics study on the conformational transition of prion induced by the point mutation: F198S. *Thin Solid Films* **499**, 224-228.

308. Dima, R. & Thirumalai, D. (2002). Exploring the propensities of helices in PrPc to form beta sheet using NMR structures and sequence alignments. *Biophys J* **83**, 1268-1280.

309. Dima, R. I. & Thirumalai, D. (2004). Probing the instabilities in the dynamics of helical fragments from mouse PrPC. *Proc Natl Acad Sci USA* **101**, 15335-40.

310. Kallberg, Y., Gustafsson, M., Persson, B., Thyberg, J. & Johansson, J. (2001). Prediction of amyloid fibril-forming proteins. *J Biol Chem* **276**, 12945-50.

311. Kamtekar, S. & Hecht, M. H. (1995). Protein Motifs. 7. The four-helix bundle: what determines a fold? *Faseb J* **9**, 1013-22.

312. Xian, W., Connolly, P. J., Oslin, M., Hausrath, A. C. & Osterhout, J. J. (2006). Fundamental processes of protein folding: Measuring the energetic balance between helix formation and hydrophobic interactions. *Protein Sci* **15**, 2062-2070.

313. Chakroun, N., Prigent, S., Dreiss, C. A., Noinville, S., Chapuis, C., Fraternali, F. & Rezaei, H. (2010). The oligomerization properties of prion protein are restricted to the H2H3 domain. *Faseb J* **24**, 3222-31.

314. Eghiaian, F., Daubenfeld, T., Quenet, Y., van Audenhaege, M., Bouin, A.-P., van der Rest, G., Grosclaude, J. & Rezaei, H. (2007). Diversity in prion protein oligomerization pathways results from domain expansion as revealed by hydrogen/deuterium exchange and disulfide linkage. *Proc Natl Acad Sci USA* **104**, 7414-9.

315. Korth, C., Stierli, B., Streit, P., Moser, M., Schaller, O., Fischer, R., SchulzSchaeffer, W., Kretzschmar, H., Raeber, A., Braun, U., Ehrensperger, F., Hornemann, S., Glockshuber, R., Riek, R., Billeter, M., Wuthrich, K. & Oesch, B. (1997). Prion (PrPSc)-specific epitope defined by a monoclonal antibody. *Nature* **390**, 74-77.

316. Kaneko, K., Zulianello, L., Scott, M., Cooper, C. M., Wallace, A. C., James, T. L., Cohen, F. E. & Prusiner, S. B. (1997). Evidence for protein X binding to a

discontinuous epitope on the cellular prion protein during scrapie prion propagation. *Proc Natl Acad Sci USA* **94**, 10069-74.

317. Bolognesi, M. L., Ai Tran, H. N., Staderini, M., Monaco, A., López-Cobeñas, A., Bongarzone, S., Biarnés, X., López-Alvarado, P., Cabezas, N., Caramelli, M., Carloni, P., Menéndez, J. C. & Legname, G. (2010). Discovery of a Class of Diketopiperazines as Antiprion Compounds. *ChemMedChem* **5**, 1324-1334.

318. Bongarzone, S., Tran, H. N. A., Cavalli, A., Roberti, M., Carloni, P., Legname, G. & Bolognesi, M. L. (2010). Parallel Synthesis, Evaluation, and Preliminary Structure-Activity Relationship of 2,5-Diamino-1,4-benzoquinones as a Novel Class of Bivalent Anti-Prion Compound. *J Med Chem* **53**, 8197-201.

319. Corsaro, A., Thellung, S., Bucciarelli, T., Scotti, L., Chiovitti, K., Villa, V., D'Arrigo, C., Aceto, A. & Florio, T. (2010). High hydrophobic amino acid exposure is responsible of the neurotoxic effects induced by E200K or D202N disease-related mutations of the human prion protein. *Int J Biochem Cell Biol* **In Press, Corrected Proof**.

320. Nicoll, A. J., Trevitt, C. R., Tattum, M. H., Risse, E., Quarterman, E., Ibarra, A. A., Wright, C., Jackson, G. S., Sessions, R. B., Farrow, M., Waltho, J. P., Clarke, A. R. & Collinge, J. (2010). Pharmacological chaperone for the structured domain of human prion protein. *Proc Natl Acad Sci USA* **107**, 17610-5.

321. Heske, J., Heller, U., Winklhofer, K. F. & Tatzelt, J. (2004). The C-terminal globular domain of the prion protein is necessary and sufficient for import into the endoplasmic reticulum. *Journal of Biological Chemistry* **279**, 5435-5443.

322. Mishra, R. S., Bose, S., Gu, Y., Li, R. & Singh, N. (2003). Aggresome formation by mutant prion proteins: the unfolding role of proteasomes in familial prion disorders. *J Alzheimers Dis* **5**, 15-23.

323. Ashok, A. & Hegde, R. S. (2009). Selective Processing and Metabolism of Disease-Causing Mutant Prion Proteins. *Plos Pathogens* **5**.

324. Peretz, D., Williamson, R. A., Kaneko, K., Vergara, J., Leclerc, E., Schmitt-Ulms, G., Mehlhorn, I. R., Legname, G., Wormald, M. R., Rudd, P. M., Dwek, R. A., Burton, D. R. & Prusiner, S. B. (2001). Antibodies inhibit prion propagation and clear cell cultures of prion infectivity. *Nature* **412**, 739-43.

325. Antonyuk, S. V., Trevitt, C. R., Strange, R. W., Jackson, G. S., Sangar, D., Batchelor, M., Cooper, S., Fraser, C., Jones, S., Georgiou, T., Khalili-Shirazi, A., Clarke, A. R., Hasnain, S. S. & Collinge, J. (2009). Crystal structure of human prion protein bound to a therapeutic antibody. *Proc Natl Acad Sci USA* **106**, 2554-2558.

326. Kuwata, K., Nishida, N., Matsumoto, T., Kamatari, Y. O., Hosokawa-Muto, J., Kodama, K., Nakamura, H. K., Kimura, K., Kawasaki, M., Takakura, Y., Shirabe, S., Takata, J., Kataoka, Y. & Katamine, S. (2007). Hot spots in prion protein for pathogenic conversion. *Proc Natl Acad Sci USA* **104**, 11921-6.

327. Yamamoto, N. & Kuwata, K. (2009). Regulating the Conformation of Prion Protein through Ligand Binding. *J Phys Chem B* **113**, 12853-12856.

328. Dossena, S., Imeri, L., Mangieri, M., Garofoli, A., Ferrari, L., Senatore, A., Restelli, E., Baiducci, C., Fiordaliso, F., Salio, M., Bianchi, S., Fioriti, L., Morbin, M., Pincherle, A., Marcon, G., Villani, F., Carli, M., Tagliavini, F., Forloni, G. & Chiesa, R. (2008). Mutant Prion Protein Expression Causes Motor and Memory Deficits and Abnormal Sleep Patterns in a Transgenic Mouse Model. *Neuron* **60**, 598-609.

329. Jackson, W. S., Borkowski, A. W., Faas, H., Steele, A. D., King, O. D., Watson, N., Jasanoff, A. & Lindquist, S. (2009). Spontaneous Generation of Prion Infectivity in Fatal Familial Insomnia Knockin Mice. *Neuron* **63**, 438-450.

330. Friedman-Levi, Y., Meiner, Z., Canello, T., Frid, K., Kovacs, G. G., Budka, H., Avrahami, D. & Gabizon, R. (2011). Fatal Prion Disease in a Mouse Model of Genetic E200K Creutzfeldt-Jakob Disease. *Plos Pathogens* **7**.

331. Goldmann, W., Hunter, N., Smith, G., Foster, J. & Hope, J. (1994). Prp Genotype and Agent Effects in Scrapie - Change in Allelic Interaction with Different Isolates of Agent in Sheep, a Natural Host of Scrapie. *Journal of General Virology* **75**, 989-995.

332. Westaway, D., Zuliani, V., Cooper, C. M., Dacosta, M., Neuman, S., Jenny, A. L., Detwiler, L. & Prusiner, S. B. (1994). Homozygosity for Prion Protein Alleles Encoding Glutamine-171 Renders Sheep Susceptible to Natural Scrapie. *Genes & Development* **8**, 959-969.

333. Fink, J. K., Peacock, M. L., Warren, J. T., Roses, A. D. & Prusiner, S. B. (1994). Detecting Prion Protein Gene-Mutations by Denaturing Gradient Gel-Electrophoresis. *Human Mutation* **4**, 42-50.

334. Perrier, V., Kaneko, K., Safar, J., Vergara, J., Tremblay, P., DeArmond, S. J., Cohen, F. E., Prusiner, S. B. & Wallace, A. C. (2002). Dominant-negative inhibition of prion replication in transgenic mice. *Proc Natl Acad Sci U S A* **99**, 13079-84.

335. Mead, S., Stumpf, M. P. H., Whitfield, J., Beck, J. A., Poulter, M., Campbell, T., Uphill, J. B., Goldstein, D., Alpers, M., Fisher, E. M. C. & Collinge, J. (2003). Balancing selection at the prion protein gene consistent with prehistoric kurulike epidemics. *Science* **300**, 640-643.

336. Riek, R., Hornemann, S., Wider, G., Billeter, M., Glockshuber, R. & Wuthrich, K. (1996). NMR structure of the mouse prion protein domain PrP(121-231). *Nature* **382**, 180-2.

337. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O. & Shaw, D. E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins-Structure Function and Bioinformatics* **78**, 1950-1958.

338. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* **4**, 435-447.

339. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics & Modelling* **14**, 33-38.

340. Kabsch, W. & Sander, C. (1983). Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **22**, 2577-2637.

341. Riek, R., Wider, G., Billeter, M., Hornemann, S., Glockshuber, R. & Wuthrich, K. (1998). Prion protein NMR structure and familial human spongiform encephalopathies. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 11667-11672.

342. Biljan, I., Giachin, G., Ilc, G., Zhukov, I., Plavec, J. & Legname, G. (2012). Structural basis for the protective effect of the human prion protein carrying the dominant-negative E219K polymorphism. *Biochem J.*

343. Capellari, S., Strammiello, R., Saverioni, D., Kretzschmar, H. & Parchi, P. (2011). Genetic Creutzfeldt-Jakob disease and fatal familial insomnia: insights into phenotypic variability and disease pathogenesis. *Acta Neuropathologica* **121**, 21-37.

344. Chebaro, Y. & Derreumaux, P. (2009). The Conversion of Helix H2 to beta-Sheet Is Accelerated in the Monomer and Dimer of the Prion Protein upon T183A Mutation. *Journal of Physical Chemistry B* **113**, 6942-6948.

345. Santini, S., Claude, J. B., Audic, S. & Derreumaux, P. (2003). Impact of the tail and mutations G131V and M129V on prion protein flexibility. *Proteins-Structure Function and Genetics* **51**, 258-265.

346. Santini, S. & Derreumaux, P. (2004). Helix H1 of the prion protein is rather stable against environmental perturbations: molecular dynamics of mutation and deletion variants of PrP(90-231). *Cellular and Molecular Life Sciences* **61**, 951-960.

347. Sekijima, M., Motono, C., Yamasaki, S., Kaneko, K. & Akiyama, Y. (2003). Molecular dynamics simulation of dimeric and monomeric forms of human prion protein: Insight into dynamics and properties. *Biophysical Journal* **85**, 1176-1185.

348. Alonso, D. O. V., DeArmond, S. J., Cohen, F. E. & Daggett, V. (2001). Mapping the early steps in the pH-induced conformational conversion of the prion protein. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 2985-2989.

349. Lu, X. J., Wintrode, P. L. & Surewicz, W. K. (2007). beta-Sheet core of human prion protein amyloid fibrils as determined by hydrogen/deuterium exchange. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 1510-1515.

350. Adrover, M., Pauwels, K., Prigent, S., de Chiara, C., Xu, Z., Chapuis, C., Pastore, A. & Rezaei, H. (2010). Prion Fibrillization Is Mediated by a Native Structural Element That Comprises Helices H2 and H3. *Journal of Biological Chemistry* **285**, 21004-21012.

351. Kallberg, Y., Gustafsson, M., Persson, B., Thyberg, J. & Johansson, J. (2001). Prediction of amyloid fibril-forming proteins. *Journal of Biological Chemistry* **276**, 12945-12950.

352. Chakroun, N., Prigent, S., Dreiss, C. A., Noinville, S., Chapuis, C., Fraternali, F. & Rezaei, H. (2010). The oligomerization properties of prion protein are restricted to the H2H3 domain. *Faseb Journal* **24**, 3222-3231.

353. Lu, X., Wintrode, P. L. & Surewicz, W. K. (2007). Beta-sheet core of human prion protein amyloid fibrils as determined by hydrogen/deuterium exchange. *Proc Natl Acad Sci U S A* **104**, 1510-5.

354. Cobb, N. J., Sonnichsen, F. D., McHaourab, H. & Surewicz, W. K. (2007). Molecular architecture of human prion protein amyloid: a parallel, in-register beta-structure. *Proc Natl Acad Sci U S A* **104**, 18946-51.

355. Tycko, R., Savtchenko, R., Ostapchenko, V. G., Makarava, N. & Baskakov, I. V. (2010). The alpha-helical C-terminal domain of full-length recombinant PrP converts to an in-register parallel beta-sheet structure in PrP fibrils: evidence from solid state nuclear magnetic resonance. *Biochemistry* **49**, 9488-97.

356. Helmus, J. J., Surewicz, K., Nadaud, P. S., Surewicz, W. K. & Jaroniec, C. P. (2008). Molecular conformation and dynamics of the Y145Stop variant of human prion protein in amyloid fibrils. *Proc Natl Acad Sci U S A* **105**, 6284-9.

357. Lim, K. H., Nguyen, T. N., Damo, S. M., Mazur, T., Ball, H. L., Prusiner, S. B., Pines, A. & Wemmer, D. E. (2006). Solid-state NMR structural studies of the fibril form of a mutant mouse prion peptide PrP89-143(P101L). *Solid State Nucl Magn Reson* **29**, 183-90.

358. Damo, S. M., Phillips, A. H., Young, A. L., Li, S., Woods, V. L., Jr. & Wemmer, D. E. (2010). Probing the conformation of a prion protein fibril with hydrogen exchange. *J Biol Chem* **285**, 32303-11.

359. Prusiner, S. B. (2001). Shattuck lecture - Neurodegenerative diseases and prions. *New England Journal of Medicine* **344**, 1516-1526.

360. Cohen, F. E. & Kelly, J. W. (2003). Therapeutic approaches to protein-misfolding diseases. *Nature* **426**, 905-909.

361. Nekooki-Machida, Y., Kurosawa, M., Nukina, N., Ito, K., Oda, T. & Tanaka, M. (2009). Distinct conformations of in vitro and in vivo amyloids of huntingtin-exon1 show different cytotoxicity. *Proc Natl Acad Sci U S A* **106**, 9679-84.

362. Tessier, P. M. & Lindquist, S. (2007). Prion recognition elements govern nucleation, strain specificity and species barriers. *Nature* **447**, 556-+.

363. Legname, G., Nguyen, H. O. B., Baskakov, I. V., Cohen, F. E., DeArmond, S. J. & Prusiner, S. B. (2006). Strain-specified characteristics of mouse synthetic prions (vol 102, pg 2168, 2005). *Proceedings of the National Academy of Sciences of the United States of America* **103**, 14642-14642.

364. Perez, D. R., Damberger, F. F. & Wuthrich, K. (2010). Horse Prion Protein NMR Structure and Comparisons with Related Variants of the Mouse Prion Protein (vol 400, pg 121, 2010). *Journal of Molecular Biology* **402**, 929-930.

365. James, T. L., Liu, H., Ulyanov, N. B., Farr-Jones, S., Zhang, H., Donne, D. G., Kaneko, K., Groth, D., Mehlhorn, I., Prusiner, S. B. & Cohen, F. E. (1997). Solution structure of a 142-residue recombinant prion protein corresponding to the infectious fragment of the scrapie isoform. *Proc Natl Acad Sci U S A* **94**, 10086-91.

366. Wen, Y., Li, J., Yao, W., Xiong, M., Hong, J., Peng, Y., Xiao, G. & Lin, D. (2010). Unique structural characteristics of the rabbit prion protein. *J Biol Chem* **285**, 31682-93.

367. Christen, B., Hornemann, S., Damberger, F. F. & Wuthrich, K. (2009). Prion protein NMR structure from tammar wallaby (Macropus eugenii) shows that the beta2-alpha2 loop is modulated by long-range sequence effects. *Journal of Molecular Biology* **389**, 833-45.

368. Agrimi, U., Nonno, R., Dell'Omo, G., Di Bari, M. A., Conte, M., Chiappini, B., Esposito, E., Di Guardo, G., Windl, O., Vaccari, G. & Lipp, H. P. (2008). Prion protein amino acid determinants of differential susceptibility and molecular feature of prion strains in mice and voles. *Plos Pathogens* **4**.

369. Damberger, F. F., Christen, B., Perez, D. R., Hornemann, S. & Wuthrich, K. (2011). Cellular prion protein conformation and function. *Proc Natl Acad Sci U S A* **108**, 17308-13.

370. Damberger, F. F., Christen, B., Perez, D. R., Hornemann, S. & Wuthrich, K. (2011). Cellular prion protein conformation and function. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 17308-17313.

371. Telling, G. C., Scott, M., Mastrianni, J., Gabizon, R., Torchia, M., Cohen, F. E., DeArmond, S. J. & Prusiner, S. B. (1995). Prion propagation in mice expressing human and chimeric PrP transgenes implicates the interaction of cellular PrP with another protein. *Cell* **83**, 79-90.

372. Deleault, N. R., Lucassen, R. W. & Supattapone, S. (2003). RNA molecules stimulate prion protein conversion. *Nature* **425**, 717-720.

373. Deleault, N. R., Piro, J. R., Walsh, D. J., Wang, F., Ma, J., Geoghegan, J. C. & Supattapone, S. (2012). Isolation of phosphatidylethanolamine as a solitary cofactor for prion formation in the absence of nucleic acids. *Proc Natl Acad Sci U S A* **109**, 8546-51.

374. Deleault, N. R., Walsh, D. J., Piro, J. R., Wang, F., Wang, X., Ma, J., Rees, J. R. & Supattapone, S. (2012). Cofactor molecules maintain infectious conformation and restrict strain properties in purified prions. *Proc Natl Acad Sci U S A* **109**, E1938-46.

375. Geoghegan, J. C., Miller, M. B., Kwak, A. H., Harris, B. T. & Supattapone, S. (2009). Trans-Dominant Inhibition of Prion Propagation In Vitro Is Not Mediated by an Accessory Cofactor. *Plos Pathogens* **5**.

376. Hizume, M., Kobayashi, A., Teruya, K., Ohashi, H., Ironside, J. W., Mohri, S. & Kitamoto, T. (2009). Human Prion Protein (PrP) 219K Is Converted to PrP(Sc) but Shows Heterozygous Inhibition in Variant Creutzfeldt-Jakob Disease Infection. *Journal of Biological Chemistry* **284**, 3603-3609.

377. Lee, C. I., Yang, Q., Perrier, V. & Baskakov, I. V. (2007). The dominant-negative effect of the Q218K variant of the prion protein does not require protein X. *Protein Science* **16**, 2166-73.

378. Geoghegan, J. C., Miller, M. B., Kwak, A. H., Harris, B. T. & Supattapone, S. (2009). Trans-dominant inhibition of prion propagation in vitro is not mediated by an accessory cofactor. *Plos Pathogens* **5**, e1000535.

379. Kobayashi, A., Hizume, M., Teruya, K., Mohri, S. & Kitamoto, T. (2009). Heterozygous inhibition in prion infection The stone fence model. *Prion* **3**, 27-30.

380. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**, 403-410.

381. Silva, J. L., Vieira, T. C. R. G., Gomes, M. P. B., Rangel, L. P., Scapin, S. M. N. & Cordeiro, Y. (2011). Experimental approaches to the interaction of the prion protein with nucleic acids and glycosaminoglycans: Modulators of the pathogenic conversion. *Methods* **53**, 306-317.

382. Leal, S. S., Botelho, H. M. & Gomes, C. M. (2012). Metal ions as modulators of protein conformation and misfolding in neurodegeneration. *Coordination Chemistry Reviews* **256**, 2253-2270.

383. Migliorini, C., Porciatti, E., Luczkowski, M. & Valensin, D. (2012). Structural characterization of Cu2+, Ni2+ and Zn2+ binding sites of model peptides associated with neurodegenerative diseases. *Coordination Chemistry Reviews* **256**, 352-368.

384. Wong, C., Xiong, L. W., Horiuchi, M., Raymond, L., Wehrly, K., Chesebro, B. & Caughey, B. (2001). Sulfated glycans and elevated temperature stimulate PrPSc-dependent cell-free formation of protease-resistant prion protein. *Embo Journal* **20**, 377-386.

385. Hegde, R. S. & Kang, S. W. (2008). The concept of translocational regulation. *Journal of Cell Biology* **182**, 225-232.

386. Hara, H., Okemoto-Nakamura, Y., Shinkai-Ouchi, F., Hanada, K., Yamakawa, Y. & Hagiwara, K. (2012). Mouse Prion Protein (PrP) Segment 100 to 104 Regulates Conversion of PrPC to PrPSc in Prion-Infected Neuroblastoma Cells. *Journal of Virology* **86**, 5626-5636.

387. Abalos, G. C., Cruite, J. T., Bellon, A., Hemmers, S., Akagi, J., Mastrianni, J. A., Williamson, R. A. & Solforosi, L. (2008). Identifying key components of the PrPC-PrPSc replicative interface. *J Biol Chem* **283**, 34021-8.

388. Zhao, H. X., Klingeborn, M., Simonsson, M. & Linne, T. (2006). Proteolytic cleavage and shedding of the bovine prion protein in two cell culture systems. *Virus Research* **115**, 43-55.

389. Li, D. W., Mohanty, S., Irback, A. & Huo, S. (2008). Formation and growth of oligomers: a Monte Carlo study of an amyloid tau fragment. *PLoS Comput Biol* **4**, e1000238.

390. Degioia, L., Selvaggini, C., Ghibaudi, E., Diomede, L., Bugiani, O., Forloni, G., Tagliavini, F. & Salmona, M. (1994). Conformational Polymorphism of the Amyloidogenic and Neurotoxic Peptide Homologous to Residues-106-126 of the Prion Protein. *Journal of Biological Chemistry* **269**, 7859-7862.

391. Miura, T., Yoda, M., Takaku, N., Hirose, T. & Takeuchi, H. (2007). Clustered negative charges on the lipid membrane surface induce beta-sheet formation of prion protein fragment 106-126. *Biochemistry* **46**, 11589-11597.

392. Satheeshkumar, K. S. & Jayakumar, R. (2003). Conformational polymorphism of the amyloidogenic peptide homologous to residues 113-127 of the prion protein. *Biophysical Journal* **85**, 473-483.

393. Zahn, R. (2003). The octapeptide repeats in mammalian prion protein constitute a pH-dependent folding and aggregation site. *Journal of Molecular Biology* **334**, 477-488.

394. Soding, J., Biegert, A. & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research* **33**, W244-W248.

395. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U. & Sali, A. (2006). Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* **Chapter 5**, Unit 5 6.

396. Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993). Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography* **26**, 283-291.

397. Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K. & Uversky, V. N. (2010). PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **1804**, 996-1010.

398. McGuffin, L. J., Bryson, K. & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404-405.

399. Cole, C., Barber, J. D. & Barton, G. J. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Research* **36**, W197-W201.

400. Pollastri, G. & McLysaght, A. (2005). Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics* **21**, 1719-1720.

401. Kallberg, M., Wang, H. P., Wang, S., Peng, J., Wang, Z. Y., Lu, H. & Xu, J. B. (2012). Template-based protein structure modeling using the RaptorX web server. *Nature Protocols* **7**, 1511-1522.

402. Blaszczyk, M., Jamroz, M., Kmiecik, S. & Kolinski, A. (2013). CABS-fold: server for the de novo and consensus-based prediction of protein structure. *Nucleic Acids Research* **41**, W406-W411.

403. Ilc, G., Giachin, G., Jaremko, M., Jaremko, L., Benetti, F., Plavec, J., Zhukov, I. & Legname, G. (2010). NMR Structure of the Human Prion Protein with the Pathological Q212P Mutation Reveals Unique Structural Features. *Plos One* **5**.

404. Jacques, D. A., Langley, D. B., Kuramitsu, S., Yokoyama, S., Trewhella, J. & Guss, J. M. (2011). The structure of TTHA0988 from Thermus thermophilus, a KipI-KipA homologue incorrectly annotated as an allophanate hydrolase. *Acta Crystallographica Section D-Biological Crystallography* **67**, 105-111.

405. Acharya, K. R., Shapiro, R., Riordan, J. F. & Vallee, B. L. (1995). Crystal-Structure of Bovine Angiogenin at 1.5-Angstrom Resolution. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 2949-2953.

406. Yamada, H., Tamada, T., Kosaka, M., Miyata, K., Fujiki, S., Tano, M., Moriya, M., Yamanishi, M., Honjo, E., Tada, H., Ino, T., Yamaguchi, H., Futami, J., Seno, M., Nomoto, T., Hirata, T., Yoshimura, M. & Kuroki, R. (2007). 'Crystal lattice engineering,' an approach to engineer protein crystal contacts by creating intermolecular symmetry: Crystallization and structure determination of a mutant human RNase 1 with a hydrophobic interface of leucines. *Protein Science* **16**, 1389-1397.

407. Li, Z. X., Huang, Y., Ge, J., Fan, H., Zhou, X. H., Li, S. T., Bartlam, M., Wang, H. H. & Rao, Z. (2007). The crystal structure of MCAT from Mycobacterium tuberculosis reveals three new catalytic models. *Journal of Molecular Biology* **371**, 1075-1083.

408. Hukushima, K. & Nemoto, K. (1996). Exchange Monte Carlo method and application to spin glass simulations. *Journal of the Physical Society of Japan* **65**, 1604-1608.

409. Johansen, A. M. & Evers, L. (2008). Monte Carlo Methods. *Lecture Notes*, 5-8.

410. Jain, A. K. & Dubes, R. C. (1988). *Algorithms for clustering data*, Prentice-Hall, Inc.

411. Sneath, P. H. & Sokal, R. R. (1973). *Numerical taxonomy. The principles and practice of numerical classification*.

412. Cappai, R., Stewart, L., Jobling, M. F., Thyer, J. M., White, A. R., Beyreuther, K., Collins, S. J., Masters, C. L. & Barrow, C. J. (1999). Familial prion disease

mutation alters the secondary structure of recombinant mouse prion protein: implications for the mechanism of prion formation. *Biochemistry* **38**, 3280-4.

413. D'Angelo, P., Della Longa, S., Arcovito, A., Mancini, G., Zitolo, A., Chillemi, G., Giachin, G., Legname, G. & Benetti, F. (2012). Effects of the pathological Q212P mutation on human prion protein non-octarepeat copper-binding site. *Biochemistry* **51**, 6068-79.
414. You, H. T., Tsutsui, S., Hameed, S., Kannanayakal, T. J., Chen, L. N., Xia, P., Engbers, J. D. T., Lipton, S. A., Stys, P. K. & Zamponi, G. W. (2012). A beta neurotoxicity depends on interactions between copper ions, prion protein, and N-methyl-D-aspartate receptors. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 1737-1742.
415. Viles, J. H. (2012). Metal ions and amyloid fiber formation in neurodegenerative diseases. Copper, zinc and iron in Alzheimer's, Parkinson's and prion diseases. *Coordination Chemistry Reviews* **256**, 2271-2284.
416. Hornemann, S., von Schroetter, C., Damberger, F. F. & Wuthrich, K. (2009). Prion Protein-Detergent Micelle Interactions Studied by NMR in Solution. *Journal of Biological Chemistry* **284**, 22713-22721.
417. Kanyo, Z. F., Pan, K. M., Williamson, R. A., Burton, D. R., Prusiner, S. B., Fletterick, R. J. & Cohen, F. E. (1999). Antibody binding defines a structure for an epitope that participates in the PrPC -> PrPSc conformational change. *Journal of Molecular Biology* **293**, 855-863.
418. Luginbuhl, B., Kanyo, Z., Jones, R. M., Fletterick, R. J., Prusiner, S. B., Cohen, F. E., Williamson, R. A., Burton, D. R. & Pluckthun, A. (2006). Directed evolution of an anti-prion protein scFv fragment to an affinity of 1 pM and its structural interpretation. *Journal of Molecular Biology* **363**, 75-97.
419. Swayampakula, M., Baral, P. K., Aguzzi, A., Kav, N. N. & James, M. N. (2013). The crystal structure of an octapeptide repeat of the Prion protein in complex with a fab fragment of the POM2 antibody. *Protein Science*.
420. Csermely, P., Palotai, R. & Nussinov, R. (2010). Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochemical Sciences* **35**, 539-546.

421. Favrin, G., Irback, A. & Mohanty, S. (2004). Oligomerization of amyloid Abeta16-22 peptides using hydrogen bonds and hydrophobicity forces. *Biophys J* **87**, 3657-64.

422. Irback, A. & Mohanty, S. (2005). Folding thermodynamics of peptides. *Biophys J* **88**, 1560-9.

423. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B. & Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **26**, 1668-1688.

424. Wang, J., Cieplak, P. & Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry* **21**, 1049-1074.

425. Oostenbrink, C., Villa, A., Mark, A. E. & van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry* **25**, 1656-76.

426. Tanford, C. (1980). *<<The>> hydrophobic effect: Formation of micelles and biological membranes*. 2. edit, Wiley, New York u. a.

427. Kauzmann, W. (1959). Some Factors in the Interpretation of Protein Denaturation. *Advances in Protein Chemistry* **14**, 1-63.

428. Biswas, K. M., DeVido, D. R. & Dorsey, J. G. (2003). Evaluation of methods for measuring amino acid hydrophobicities and interactions. *Journal of Chromatography A* **1000**, 637-655.

429. Charton, M. & Charton, B. I. (1982). The Structural Dependence of Amino-Acid Hydrophobicity Parameters. *Journal of Theoretical Biology* **99**, 629-644.