

Conformational preferences of the full chicken prion protein in solution and its differences with respect to mammals.

Adriana Pietropaolo^a, Luca Muccioli^a,
Claudio Zannoni^{a*} and Enrico Rizzarelli^b

Abstract

We investigate the conformations of the full chicken prion protein (ChPrP1-267) in solution at neutral pH with molecular dynamics simulations. We focus on the persistence of its secondary structure motifs using a recently proposed protein chirality indicator [A. Pietropaolo et al. *Proteins*, 2008, 70, 667-677]. From this, we find a high rigidity of helix 2 (ChPrP178-195) and of the hexarepeat domain, which is turn rich, and a plasticity of the short β sheet, consistent with the available NMR structural details. We also determine the extent of solvation for each residue, revealing local minima for such structured regions. These features hint at a possible origin of the high resistance to proteolysis of the avian prion proteins and of its capability in preventing the aggregation in comparison to mammals.

Key words: avian prion; molecular dynamics; chirality; hexarepeat; solvation

Introduction

The prion protein (PrP) is a glycosyl-phosphatidylinositol-anchored surface glycoprotein, so far related to a vast class of neurodegenerative diseases, known as transmissible spongiform encephalopathies [1]. These disorders derive from a partial unfolding of the normal cellular prion protein (PrP^C), which converts to an isoform (PrP^{Sc}) that aggregates in amyloid plaques. It is worth noting that, although prion diseases seem to be spared to non mammals, PrP^C was identified in a wide range of species [2, 3]. Therefore, given the huge fundamental and medical interest of this phenomenon (see e.g. ref. [4]), the structure of PrP^C has been determined by NMR spectroscopy for many species, also to find hints on the surprising resistance to the disease of some of them, like avians. Notably, despite the low primary sequence identity to mammalian PrP (around 30% between chicken, ChPrP, and human prion, HuPrP), the molecular architecture of the globular core is preserved among the two species. In particular, both mammal and avian proteins consist of a globular domain toward the C-terminal [5, 6], involving three α helices and a short antiparallel

β -sheet, a short flexible C-terminal and a long flexible N-terminal region partially constituted by specific multiple repeats, *PHGGGWGQ* in mammals and *PHNPGY* in avians. Experimental studies indicate that the N-terminal region of PrP^C plays a regulatory role in the PrP^C-PrP^{Sc} conversion [7, 8]. Furthermore, this part of chicken prion protein is essential for the anterograde axonal transport [9] and has been described to induce acetylcholine-receptor activity [3], and to drive clathrin-coated pits endocytosis [10], supposedly because of the abundance of glycines, prolines and the possibility of forming β turn motifs. Recent simulation studies gave support to this hypothesis showing the abundance of turn structures inside the chicken hexarepeat domain [11, 12], which could to some extent lower its flexibility with respect to mammal octarepeats. Unfortunately, the available NMR structure of ChPrP [6] has been obtained not at physiological pH, but at acidic pH and is only restricted to the globular core sequence (128-242), while the assignment of the N- and C-terminal tails has been hampered until now by the absence of rigid secondary structure elements. From the computational side, although a number of Molecular Dynamics (MD) studies were carried out on PrP, [13, 14, 15, 16], such investigations again dealt only with the globular core region, excluding the N- and C-terminal domains. Several contributions dealt with the key role of copper in the biology of PrP^C. In particular, extensive studies focused on the N-terminal repeat region [17, 18, 19, 20, 21, 22], on the region connecting the unstructured termini with the globular portion [23] and as well as on the globular segments of PrP [24]. Bearing this in mind, here we take into account the effect of linking the N- and C-terminal chains to the globular ChPrP128-242 NMR structure, on the globular core itself, and on the conformational preferences of such regions inside the full avian prion protein, ChPrP1-267. We thus employ MD simulations at physiological pH and, to complement the existing analysis tools, we apply a recently proposed protein chirality index, [25], which allows for improved identification of secondary structure patterns. From this index, we are also able to investigate the persistence of a secondary structure motif inside the protein backbone. In particular we look for differences in the conformational features of avian and mammal prion protein that can be connected to their misfolding and aggregation propensity.

*E-mail: Claudio.Zannoni@unibo.it ^aDipartimento di Chimica Fisica ed Inorganica and INSTM, Università di Bologna, V.le Risorgimento, 4, I-40136 Bologna, Italy. ^b Dipartimento di Scienze Chimiche, Università di Catania V.le A. Doria 6, I-95125 Catania, Italy.

1 Materials and methods

1.1 Simulation Details

A fully atomistic molecular dynamics of the whole chicken prion protein, ChPrP1-267, was carried out in water at neutral pH. In these conditions, available pK values [26] indicate that the ionizable aminoacids such as histidines (41, 54, 60, 66, 72, 84, 104, 118, 147, 247, 267, protonated at the δ nitrogen) and tyrosines (46, 52, 58, 64, 70, 76, 82, 88, 95, 103, 135, 146, 154, 164, 169, 170, 173, 194, 235, 238) are in the neutral state; glutammic (153, 159, 193, 206, 215, 217, 219, 229, 237) and aspartic (20, 149, 152, 172, 180, 185, 250) acids are negatively charged, while lysines (25, 26, 28, 30, 107, 110, 113, 117, 201, 202, 221, 225) and arginines (3, 42, 48, 139, 150, 155, 163, 167, 171, 228, 236, 239) are positively charged; to ensure charge neutrality eight chloride ions were included in the simulation box. The simulation was run in water using the GROMACS 3.3.3 package [27], the OPLS-AA force field [28] for the protein, the SPC model for water [29], and Particle Mesh-Ewald (PME) for long range electrostatic interactions [30]. The starting configuration of ChPrP1-267 was built by linking the best representative NMR model of the globular core, pdb code 1U3M [6], to the N- and C-terminal regions, 1-127 and 243-267 respectively, both of them starting from a planar conformation with all backbone dihedrals (ϕ, ψ) set to $(-180^\circ, 180^\circ)$, and energy minimized in water before the linking. A run of about 130 ns was then performed for a cubic box containing the protein, 26560 water molecules and the counterions (in total 83733 atoms), with periodic boundary conditions (PBC), with 2 fs of time step in the isothermal isobaric ensemble (NPT), at $P=1$ atm, $T=300$ K. Temperature and pressure were controlled by Berendsen thermostat and barostat [31]; after the first 20 ns of equilibration, sufficiently long to relax local motions, the trajectory analysis was performed on a 113 ns production run, with configurations stored every 5 ps. To further ensure that linking the terminal portions does not introduce a bias, we carried out a 20 ns long run at 500 K, starting from the initial extended configuration, keeping the NMR core fixed. From this, we spun a configuration every five nanoseconds and cool it down to 300 K, running in total three 25 ns long runs.

Considerable care was taken in ensuring that the first preparation stage was effective in equilibrating the system; to this end, we have introduced a novel plausibility criterium for the conformations. We have used this criterion on our main run, as well as on three additional pre-equilibration trajectories as described in Section 1.3. Based on these tests we are satisfied that our production runs are meaningfully reporting on the full protein in solution.

1.2 Chirality calculation

To a great extent the interest in protein simulations is related to their capability to provide detailed conforma-

tional features and their evolution in time. Extracting this conformational information and its changes in time beyond the standard possibilities offered by tools like DSSP [32] and STRIDE [33] is however far from simple. In particular a relative limitation of these methods that assign a certain motif (e. g. α helix, β sheet, etc) based on hydrogen bond patterns and backbone dihedrals, resides in the individuation of poly-L-proline II helix and in “twilight” zones often classified as coils [34]. In this regard, we have recently proposed an additional tool [25], showing that it is possible to assign the motif type to a protein fragment also through the evaluation of a local chirality index, complementing the DSSP classification. In summary, the method consists in dividing up the entire backbone in fragments and computing for each of them a chirality index calculated from the backbone atoms coordinates; the instantaneous value of the index can then be compared with the characteristic values for ideal secondary structures (here reported in Table 1). This index, first proposed in [35, 36] for low molecular mass molecules [37], and suitably modified and validated for probing local backbone chirality [25], is expressed as follows, for a certain protein fragment:

$$G^{a, N_A} = \frac{4!}{3N_A^4} \sum_{\mathcal{P}_{i,j,k,l}}^{\text{all } \mathcal{P}} g_{ijkl} \quad (1)$$
$$g_{ijkl} = \begin{cases} \frac{[(\mathbf{r}_{ij} \times \mathbf{r}_{kl}) \cdot \mathbf{r}_{il}](\mathbf{r}_{ij} \cdot \mathbf{r}_{jk})(\mathbf{r}_{jk} \cdot \mathbf{r}_{kl})}{(r_{ij} r_{jk} r_{kl})^2 r_{il}} & \text{if } r_{ij}, r_{kl}, r_{il}, r_{jk} < r_c, \\ & a \leq i, j, k, l \leq N_A + a - 1 \\ 0 & \text{otherwise} \end{cases}$$

where a is the first (closest to the N-terminal) atom of a given sequence of $N_A = 15$ consecutive backbone atoms with coordinates \mathbf{r} , i, j, k, l are four atoms belonging to the sequence, $\mathcal{P}_{i,j,k,l}$ is any of the $4!$ permutations of four indexes and $r_c = 12 \text{ \AA}$ is a cutoff radius, added to avoid the computation of negligible long-range terms. We have already applied this chirality indicator to a study of the pH dependent conformation of the hexarepeat domain [12]. Here we shall employ it as an important characterization tool but also, as described in the next section, to provide a test of the overall likelihood of the generated starting secondary structure.

1.3 The protein equilibration

An unavoidable problem in simulating the evolution of complex biomolecules like proteins using molecular dynamics is the necessarily limited extension of the observation time window $\Delta\tau$. Even for much simpler low molar mass molecules, it has proved essential, in order to predict reliable observables, to follow evolutions for times longer than significant relaxation times (of the order of 10 ns) [38]. For a finite $\Delta\tau$ the choice of the starting molecular configuration itself becomes of great importance; the problem is often alleviated starting from an experimentally determined X-ray or NMR conformation. Here, however this is not fully practicable as the

complete 3D structure of the avian protein has not been reported to date, thus energy-minimized missing fragments were linked to the NMR structure of the globular core. A particular care is then necessary in assessing the plausibility of the simulated geometry.

Accordingly we considered the initial 20 ns as equilibration and excluded them from the production analysis, in order to avoid a starting configuration bias. After equilibration, we checked that volume, total energy and *end-to-end distance* fluctuated around their average value without systematic drifts. However, when, as in our case, a structure for the protein to be studied is not available, the structure in solution should at least be shown to be a plausible one, in the sense of complying with other known protein structures. In particular, since as already mentioned we focus here on the chirality of the local protein fragments along the backbone, we wish to test if the chirality indicator of the starting structure conforms to those typical of experimentally reported protein structures. To this end we have analyzed the distribution of chirality indexes for a set of 2340 NMR structures available in the PDB database ¹. This approach extends similar analysis performed with Ramachandran plots [39] or DSSP [32]. For sake of simplicity, we define a unique value of G for each aminoacid i of a protein as the average of the $G^{\alpha,15}$ values whose 15 atoms window is centered on the N, C $_{\alpha}$, C atoms of aminoacid i :

$$G_i = (G^{3(i-3)+1,15} + G^{3(i-3)+2,15} + G^{3(i-3)+3,15})/3 \quad (2)$$

This definition is clearly valid only for $i \geq 3$ and $i < N - 2$ where N is the number of aminoacids forming the protein. For each of the twenty aminoacid types AA , we then evaluated the conditional probability $P^{NMR}(G_{i+1}|G_{i,AA})$, i.e. the normalized occurrence in the dataset of having, for an aminoacid $i + 1$, a given value of chirality G_{i+1} once fixed the type of the preceding aminoacid i to AA and its chirality index to G_i . The analysis shows that typical probability values of real protein conformations, as opposed to random or non native conformations, fall in a well defined range. It is convenient to employ these probability maps in the definition of a scoring function that allows to measure how a given protein conformation complies to the dataset that originated the maps. In particular here we use a function which, once calculated the G_i values for the $i = 3, N - 2$ aminoacids of a protein chain, weighs each consecutive couple of values according to the conditional probability described above and takes the average of all the weights:

$$G_{score} = \frac{1}{N - 5} \sum_{i=3}^{N-3} P^{NMR}(G_{i+1}|G_{i,AA}) \quad (3)$$

As the chirality index is a local function of the coordinates, this score, ranging from 0 to 1, gives an information on the likelihood of the secondary structure. For the proteins of the dataset itself, the score function

has a broad non-gaussian distribution ranging approximately from 0.05 to 0.3 with a maximum at around $G_{score} = 0.13$. In addition, we have calculated the G_{score} for the frog, turtle, chicken and human prion protein from their NMR structures [6, 5] which results equal to 0.16 ± 0.01 , both for frog and turtle ones and 0.17 ± 0.01 ; 0.19 ± 0.01 , for chicken and human ones. We also calculated the G_{score} values for the MD simulation and, as it can be appreciated from Figures 1, the protein trajectories show G_{score} histograms consistent with the reference distribution extracted from the NMR protein database, revealing a not unduly biased ϕ and ψ dihedrals of the two chains bound to the globular core. We notice that this is not necessarily the case, e.g. G_{score} calculated from randomly generated chirality indexes (from a gaussian distribution centered in 0 and with $\sigma=0.3$), assuming these as unfolded structures, fall well out of the range adopted by the NMR protein database. From this, we spun a configuration every five nanoseconds and cool it down to 300 K, running in total three 25 ns long runs. We checked the time evolution of the end-to-end distance and the chirality G_{score} for the three annealings. In particular the first and the second one, converge to the equilibrium distance of the 110 ns long MD, while the third one adopts a higher distance (25 Å versus 15 Å, see Figure 2). The G_{score} time evolution (Figure 1 of Supplementary Material) shows that the third starts, especially in the C-terminal region, with very unlikely dihedral angles (low G_{score} values), equilibrating during the dynamics. This clarifies the nature of the higher end-to-end distance identified; indicating in fact that even if the end-to-end distances can be helpful in detecting the presence of an equilibrium value, this could be not sufficient to detect a correct equilibration of the system, in terms of secondary structure, as also pointed out by Courty et al. for the induced helicity in biopolymer networks [40]. To use an analogy, pulling a helical spring its end-to-end distance changes but chirality does not or to a lesser extent.

This leads us to point out the importance of chirality assessment in detecting a correct equilibration of the structures adopted during molecular dynamics, ensuring in this way the not unduly biased nature of the 110 ns long MD run we consider in the following. While it remains strictly impossible to assess that an initial conformation belongs to the “true equilibrium” basin in the free energy surface, we believe our procedure is sufficiently robust and provide a usable and viable approach.

2 Results and Discussion

After verifying the effective equilibration of the protein, we analyzed its conformational states along the production trajectories. First, we identified the average number of contacts between pairs of non consecutive aminoacids, evaluated through the C $_{\alpha}$ -C $_{\alpha}$ distances, using a cutoff of 8.5 Å[41]. In Figure 3 A) we report the contact map for the full protein; here all the secondary structures

¹<http://pdbeta.rcsb.org>.

present in the globular core can be recognized: the off-diagonal spread of contacts between strands 132-144 and 166-173 underlines the presence of a short and flexible β sheet, while the three α helices correspond to the contacts along the diagonal line. In addition, the map reveals diffuse contacts between N and C-terminal parts, in particular between residues 47-249, 51-255, 59-263 and 72-256, as suggested by the relatively short end-to-end distance in Figure 2. Also in the N-terminal part some contacts appear, notably some 3_{10} helices and turns and a β bridge. Focusing on the globular core, in Figure 3 A) we report the comparison between its contact map as obtained by our MD simulation and from the 20 NMR structures of Calzolari et al. [6]. Some differences appear: the 3_{10} helix inside the globular core, involving residues 197-199, is much more evident in MD with respect to the NMR structure, as well as the immediately following turn (around 203); on the contrary the small helix centered in 135 almost disappears. More interestingly, the contacts between residues involved in the β sheets region increase but also become more irregular if compared with the NMR structure, and a contact appears between residues 139 and 237 (at one edge of the sheet and helix 3 respectively). Both the structural irregularity of the sheet and the presence of the latter contact (termed β bulge 1) have been deemed to be designed by nature for depressing edge-to-edge interprotein dimerization [42, 43]; considering also that avian prion has a slightly shorter β sheet with respect to mammal ones [6] this is consistent with a relatively lower tendency to amyloid aggregation.

The contact between Met137 and Tyr169 or Tyr170 (and among the corresponding residues in other species prions and amyloid- β) is also believed to be important in transition metal redox chemistry linked to Alzheimer’s and prion diseases [44, 45]; for an easy electron transfer reaction between these residues it is necessary a distance inferior to 5 Å [44] between methionine sulfur and tyrosine phenolic oxygen. Even if such chemical processes are outside the scope of classical MD, it is worth noting that both 137-169 and 137-170 average distances during the simulations assume lower values with respect to the NMR structure (respectively 4.38 and 7.83 Å versus 4.62 and 10.1 Å).

In order to unambiguously assign the secondary structure and to characterize its evolution in time, we monitored it during the MD trajectory using the DSSP algorithm [32] (Figure 4 A)-E)). First we focused our attention on the N-terminal hexarepeat region, ChPrP53-88, in view of its high flexibility and its likely biological function. As previously found by us for the single [11] and the tetra-hexarepeat [12], specific residues prefer to adopt the type I β turn and 3_{10} helix structures, in this case 53-54, 64-65 for turn and 67-69 for 3_{10} helix and, in lower extent, 83-84 and 86-88 (Figure 4 B)). The presence of a short but persistent β sheet centered in residues 136-137;169-170 was confirmed by DSSP (Figure 4 C)). Interestingly, in the mouse D178N PrP pathogenic mutant this β sheet region was found, instead, to easily un-

dergo disruption [46], pointing out a role of the β sheets in the stability of the PrP fold, revealing differences in this region between mammal and avian prions.

Among the three α helices, the second (ChPrP178-195) results to be the most rigid and preserved one. This is inferred by the comparison of the time evolution of the secondary structures of the helices, reported in Figures 4C)-E), where several residues of helix 1 and, more extensively, the initial residues of helix 3, experience frequent interconversion between β turn, α and 3_{10} helix conformations; the latter secondary structure is known to be very flexible and is seldom experienced by helix 2 residues. The C-terminal part, 243-267 is substantially unordered, except residues 251-254, adopting 3_{10} helix and turn structures, and 264-265 adopting β turns and bridge conformation.

The degree of solvation of ChPrP was investigated by counting the number of water molecules inside a cylinder of radius of 12 Å, having as axis the backbone nitrogen-carbonyl carbon distance vector of the aminoacid and as height its modulus. This very intuitive measure of solvation is affected by the size of the aminoacid side chain and by the backbone dihedral angles which can vary the height of the sampling cylinder. In Figure 5, many regions of local minima are detected, centered in Val168, Val226, Arg228 and Gly197. Noteworthy, Val168 was also found in a recent theoretical study [43] to possess a tight bound water, as confirmed here by the low standard deviation of the solvation indicator. In general, poorly hydrated regions with low standard deviations are found where the secondary structure is well defined, i.e. the repeat region (turn and 3_{10} helix), helix 1 and more extensively helix 2 and helix 3 after residue 226. Besides, all three helices present high solvation sites at their N-terminal, while the long connecting loop between helices 2 and 3 is scarcely accessible to water as indicated also in [43].

Additionally, we calculated the average chirality index G for the three helices (Figures 6 A)-C)), revealing in more detail their flexibility, evidenced by the standard deviation bars, which reflect the extent of secondary structure variation. It is apparent that the three helices have different flexibility and the helix 3 has a higher propensity to changing its chirality (Figure 6 C)). Furthermore, helix 1 presents typical α -helix chirality values (cf Table 1) after residue 152, coexisting with a 3_{10} (see also Table 1), helix 2 preserves its helix structure from 180 to 192 and finally helix 3 shows an α -helix structure from residue 226 till residue 238. Such interconversions are also apparent in Figure 7, where we visualize the chirality index evolution for selected residues (Trp156, Cys186, Val223), together with the snapshots of the three helices. Here, the Val223 residue of helix 3 shows after 60 ns an interconversion from 3_{10} helix to a turn-like structure (understandable from the shift towards less negative values of the chirality index) and finally turning to an α helix. In the turn-like conformation, the side chains of Pro 198 and Val 223 get close, likely due to hydrophobic interactions between the proline ring and the aliphatic side

chain of valine. In Figure 8 we report the chirality of the different prion NMR structures, namely frog, turtle, chicken and human, which reveal a similar chirality of the three α helices and differences in the interconnecting regions of these.

In order to have an overview of the global conformation adopted by the avian prion protein in solution, the complete chirality pattern along the backbone, averaged over the trajectories, is also shown (Figure 9). As previously indicated from the DSSP assignment concerning the hexarepeat region, the N-terminal part is rich in turns and 3_{10} helices; in addition positive peaks are present, pointing to a poly-L-proline II like structure (cf Table 1), involving typically no more than three residues, i.e. residues Lys 113, Pro 178. It is worth noting that these positive peaks often coincide with highly solvated aminoacids in Figure 5, in accord with the fully solvated nature of poly-L-proline II helix.

The standard deviations of the chirality index (Figure 9) in the N- and C-terminal domains reach high values, owing to their flexibility, but it is interesting that in the hexarepeat region, such deviations become smaller, comparable to those of helix 1 and helix 3. Helix 2, instead, is the most rigid of the three helices, as the very low standard deviations suggest. Actually the stability of avian prions helix 2 was predicted by Dima and Thirumalai [47] on the basis of their high content in alanines, and related to their poor propensity to the $\alpha \rightarrow \beta$ transition. Moreover, it is worth noting that ChPrP helix 2 possesses a proline residue at the beginning of the helix (residue 178), this is clearly shown by the average chirality index along the helix 2 residues of Figure 6 B), where the first value is strongly positive, typical of poly-L-proline II dihedrals. Proline usually makes the structure stiff and it was reported to act as a fold protector, preventing non-native interactions [48, 49] and moreover, it was found that Phe¹⁹-Pro¹⁹ mutation in amyloid β 1-42 blocks the fibril formation [50], pointing out a role of proline in preventing protein aggregation.

In HuPrP helix 2, the proline residue is replaced by an histidine residue in position 177 (a putative site of interaction with signal peptidases [51]) and this appears to be consistent with the reported higher flexibility of human helix 2 with respect to chicken one [47, 52]. Kallberg et al.[53] predicted that human helix 2 is prone to favor β sheet conformation; this could explain the strong interaction between helix 2 and tetracycline [54], commonly employed to decrease the PrP^{Sc} resistance to protease K digestion, and possibly relates with the misfolding tendency. In fact, in contrast to previous models which foresaw a considerable helix conservation in prion amyloid aggregates [55, 56], the relevance of the stability of prion helices for the aggregation is supported by recent studies on human prion amyloid [57, 58], where a substantial $\alpha \rightarrow \beta$ refold of helix 2 and 3 is found in the aggregate, and on ovine prion allelic variants [59], where the ease of unwinding of helix 2 has been associated to the susceptibility to classical scrapie misfolding. Focusing again on the hexarepeat region, we recall that in

our simulation study of the tetra-hexarepeat fragment we found a hydrogen bond between the imidazole nitrogen of the first histidine and the phenolic hydrogen of the third tyrosine, determining a turn conformation in this region [12]. In the full avian prion protein, we find again such an interaction during the second half of the simulation, even if shifted of one repeat, as it involves the imidazole nitrogen of the fourth histidine and the phenolic hydrogen of the second tyrosine, as it can be appreciated from Figure 10 and from the calculated NOE contacts and distances, reported in Table 2. This side chain hydrogen bond further contributes to reduce the flexibility in the hexarepeat region. We finally report in Figure 11 a comparison between the chirality index pattern averaged on the trajectories for ChPrP53-88 and the averaged chirality index of the tetra-hexarepeat fragment dynamics, previously reported in [12]. The chirality pattern is surprisingly consistent with the periodicity of this system, that tends to adopt 3_{10} helix for residues 68-73 and 80-85 and a Type I β turn region around residue 59 and the peptide and the protein curves are largely superimposed, i.e. their secondary structures are very similar. This is encouraging for every research that aims at understanding proteins by studying only limited segments of them, and provides partial support to idea that folding happens by the initial formation of local secondary structures that subsequently assemble in tertiary and quaternary ones.

3 Conclusions

We have reported the first computational study of the conformations adopted at neutral pH by the full chicken prion protein and this has allowed us to underline some relevant differences between avian and human prion proteins. First of all, our MD simulations show that in this avian protein helix 2 is very rigid, as indicated both by chirality and DSSP analysis, while helix 1 and prevalently helix 3, do not completely maintain the α helix structure, but rather present a coexistence with 3_{10} helix and more flexible states. As for the origin of these diversities, we notice that differently from the human prion protein, in which the most flexible helix was found to be helix 2, avian helix 2 possesses a proline residue in the first position, and a higher content of alanines, that could be the cause of its high rigidity, consistent with other previous findings [48, 47].

Another important difference that is confirmed by this study is the shortness and the mobility of the avian β sheet, due to an interaction with helix 3 that prevents a perfect parallelism between the two facing strands. Also the hexarepeat region, in which a periodical conformation is adopted, presents unexpectedly higher rigidity. This is pointed out especially by the low standard deviations of the chirality indicator, which also reveals the abundance of 3_{10} helices for residues 68-73 and 80-85 and a type I β turn structure, prevalently found in residue 59. Moreover, a side chain hydrogen bond was detected

between the imidazole nitrogen of histidine 72 and the phenolic hydrogen of tyrosine 64. It is important to note that this bond was also found in simulations of the tetra-hexarepeat fragment [12] and cannot be formed by the mammal sequence because of lack of the required residues. We have further confirmed the chirality analysis with a water solvation study that reveals similar patterns for the residues. The regions with well defined secondary structure, as the repeat region and the helix ones, present local hydration minima and again helix 2 shows low values of standard deviations, while the residues with poly-L-proline II chirality values usually correspond to local solvation maxima.

We believe that the finding that a periodical structured conformation is adopted in the hexarepeat region of the avian prion protein may be correlated to its high resistance to protease [60]. Such a structured conformation of the N-terminal tail, together with the lower flexibility of ChPrP helix 2 with respect to the mammal prion analogue, and the plasticity of avian β sheet, could somehow hamper the interconversion leading to the pathogenic PrP^{S_c} isoform, explaining the rarity of prion diseases registered for avians.

Acknowledgments

We thank MIUR (PRIN 2004032851, 2005035119, PRIN2006-033492, FIRB RBNEO3PX83_001) for financial support.

References

- [1] S. B. Prusiner, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13363-13383.
- [2] F. Wopfner, G. Weidenhofer, R. Schneider, A. von Brunn, S. Gilch, T. F. Schwarz, T. Werner, H. M. Schatzl, *J. Mol. Biol.* **1999**, *289*, 1163-1178.
- [3] D. A. Harris, D. L. Falls, F. A. Johnson, G. D. Fischbach, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 7664-7668.
- [4] J. Collinge, A. R. Clarke, *Science* **2007**, *318*, 930-936.
- [5] R. Zahn, A. Liu, T. Lührs, R. Riek, C. von Schroetter, F. L. García, M. Billeter, L. Calzolari, G. Wider, K. Wüthrich, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 145-150.
- [6] L. Calzolari, D. A. Lysek, D. R. Perez, P. Guntert, K. Wüthrich, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 651-655.
- [7] M. Nunziante, S. Gilch, H. M. Schätzl, *J. Biol. Chem.* **2003**, *278*, 3726-3734.
- [8] K. N. Frankenfield, E. T. Powers, J. W. Kelly, *Protein. Sci.* **2005**, *14*, 2154-2166.
- [9] R. Butowt, P. Davies, D. R. Brown, *J. Neurosci. Res.* **2007**, *85*, 2567-2579.
- [10] S. L. Shyng, K. L. Moulder, A. Lesko, D. A. Harris, *J. Biol. Chem.* **1995**, *270*, 14793-14800.
- [11] A. Pietropaolo, L. Raiola, L. Muccioli, G. Tiberio, C. Zannoni, R. Fattorusso, C. Isernia, D. La Mendola, G. Pappalardo, E. Rizzarelli, *Chem. Phys. Lett.* **2007**, *442*, 110-118.
- [12] A. Pietropaolo, L. Muccioli, C. Zannoni, D. La Mendola, G. Maccarrone, G. Pappalardo, E. Rizzarelli, *J. Phys. Chem. B.* **2008**, *112*, 5182-5188.
- [13] E. Langella, R. Improta, V. Barone, *Biophys. J.* **2004**, *87*, 3623-3632.
- [14] A. D. Simone, G. G. Dodson, C. S. Verma, A. Zargari, F. Fraternali, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7535-7540.
- [15] M. Pappalardo, D. Milardi, D. Grasso, C. La Rosa, *New J. Chem.* **2007**, *31*, 901-905.
- [16] H. F. Ji, H. Y. Zhang, *Biochem. Biophys. Res. Comm.* **2007**, *359*, 790-794.
- [17] R. P. Bonomo, G. Impellizzeri, G. Pappalardo, E. Rizzarelli, G. Tabbí, *Chem. Eur. J.* **2000**, *6*, 4195-4202.
- [18] R. P. Bonomo, V. Cucinotta, A. Giuffrida, G. Impellizzeri, A. Magri, G. Pappalardo, E. Rizzarelli, A. M. Santoro, G. Tabbí, L. I. Vagliasindi, *Dalton Trans.* **2005**, pp. 150-158.
- [19] G. D. Natale, G. Grasso, G. Impellizzeri, D. La Mendola, G. Micera, N. Mihala, Z. Nagy, K. Osz, G. Pappalardo, V. Rigó, E. Rizzarelli, D. Sanna, I. Sóvágó, *Inorg. Chem.* **2005**, *44*, 7214-7225.
- [20] D. L. Mendola, R. P. Bonomo, G. Impellizzeri, G. Maccarrone, G. Pappalardo, A. P. A., E. Rizzarelli, V. Zito, *J. Biol. Inorg. Chem.* **2005**, *10*, 463-475.
- [21] K. Osz, Z. Nagy, G. Pappalardo, G. D. Natale, D. S. G., Micera, E. Rizzarelli, I. Sóvágó, *Chem. Eur. J.* **2007**, *13*, 7129-7143.
- [22] E. Gaggelli, H. Kozłowski, D. Valensin, G. Valensin, *Chem. Rev.* **2006**, *106*, 1995-2044.
- [23] G. Jackson, I. Murray, L. L. P. Hosszu, N. Gibbs, J. P. Waltho, A. Clarke, J. Collinge, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 8531-8535.
- [24] D. R. Brown, V. Guantieri, G. Grasso, G. Impellizzeri, G. Pappalardo, E. Rizzarelli, *J. Inorg. Biochem.* **2004**, *98*, 133-143.
- [25] A. Pietropaolo, L. Muccioli, R. Berardi, C. Zannoni, *Proteins* **2008**, *70*, 667-677.
- [26] C. P. Toseland, H. McSparron, M. N. Davies, D. R. Flower, *Nucleic Acids Res.* **2006**, *34*, D199-D203.
- [27] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, J. C. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701-1718.
- [28] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, W. L. Jorgensen, *J. Phys. Chem.* **2001**, *105*, 6474-6487.
- [29] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans, *Intermolecular Forces*, B. Pullman, editor, Reidel, Dordrecht, 1981.
- [30] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089-10092.
- [31] H. J. C. Berendsen, J. P. M. Postma, A. D. Nola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684-3690.
- [32] W. Kabsch, C. Sander, *Biopolymers* **1983**, *22*, 2577-2637.
- [33] D. Frishman, P. Argos, *Proteins* **1995**, *23*, 566-579.
- [34] J. Martin, G. Letellier, A. Marin, J.-F. Taly, A. G. de Brevern, J.-F. Gibrat, *BMC Struct. Biol.* **2005**, *5*, 17.
- [35] M. A. Osipov, B. T. Pickup, D. A. Dunmur, *Mol. Phys.* **1995**, *84*, 1193-1206.

- [36] M. Solymosi, R. J. Low, M. Grayson, M. P. Neal, *J. Chem. Phys.* **2002**, *116*, 9875-9881.
- [37] R. Berardi, G. Cainelli, P. Galletti, D. Giacomini, A. Gualandi, L. Muccioli, C. Zannoni, *J. Am. Chem. Soc.* **2005**, *127*, 10699-10706.
- [38] R. Berardi, L. Muccioli, C. Zannoni, *ChemPhysChem* **2004**, *5*, 104-111.
- [39] T. Schlick, *Molecular Modeling and Simulation*, Springer, New York, 2002.
- [40] S. Courty, J. L. Gornall, E. M. Terentjev, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13457.
- [41] M. Vendruscolo, R. Najmanovich, E. Domany, *Phys. Rev. E* **1999**, *82*, 656-659.
- [42] J. S. Richardson, D. C. Richardson, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2754-2759.
- [43] A. D. Simone, G. G. Dodson, A. Zagari, F. Fraternali, *FEBS. Lett.* **2006**, *580*, 2488-2494.
- [44] K. J. Barnham, R. Cappai, K. Beyreuther, C. L. Masters, A. F. Hill, *Trends Biochem. Sci.* **2006**, *31*, 465-472.
- [45] H. F. Ji, H. Y. Zhang, L. L. Chen, *Trends Biochem. Sci.* **2007**, *32*, 206-208.
- [46] A. Barducci, R. Chelli, P. Procacci, V. Schettino, F. L. Gervasio, M. Parrinello, *J. Am. Chem. Soc.* **2006**, *128*, 2705-2710.
- [47] R. I. Dima, D. Thirumalai, *Biophys. J.* **2002**, *83*, 1268-1280.
- [48] R. A. George, J. Heringa, *Protein Eng.* **2003**, *15*, 871-879.
- [49] W. C. Wigley, M. J. Corboy, T. D. Cutler, P. H. Thibodeau, J. Oldan, M. G. L. MG, J. Rizo, J. F. Hunt, P. J. Thomas, *Nat. Struct. Biol.* **2002**, *9*, 381-388.
- [50] S. L. Bernstein, T. Wyttenbach, A. Baumketner, J. Shea, G. Bitan, D. B. Teplow, M. T. Bowers, *J. Am. Chem. Soc.* **2005**, *127*, 2075-2084.
- [51] R. Glockshuber, S. Hornemann, M. Billeter, R. Riek, G. Wider, K. Wütrich, *FEBS Lett.* **1998**, *426*, 291-296.
- [52] M. Pappalardo, D. Milardi, C. La Rosa, C. Zannoni, E. Rizzarelli, D. Grasso, *Chem. Phys. Lett.* **2004**, *390*, 511-516.
- [53] K. Y., M. Gustafsson, B. Persson, J. Thyberg, J. Johansson, *J. Biol. Chem.* **2001**, *276*, 12945-12950.
- [54] L. Ronga, E. Langella, P. Palladino, D. Marasco, B. Tizzano, M. Saviano, C. Pedone, R. Improta, M. Ruvo, *Proteins* **2007**, *66*, 707-715.
- [55] C. Govaerts, H. Wille, S. B. Prusiner, F. E. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8342-8347.
- [56] M. L. De Marco, V. Daggett, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2293-2298.
- [57] N. J. Cobb, F. D. Sönnichsen, H. Mchaourab, W. K. Surewicz, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18946-18951.
- [58] D. N. Ganchev, N. J. Cobb, K. Surewicz, W. K. Surewicz, *Biophys. J.* **2008**, p. doi:10.1529/biophysj.108.133108.
- [59] T. J. Fitzmaurice, D. F. Burke, L. Hopkins, S. J. Yang, S. L. Yu, M.-S. Sy, A. M. Thackray, R. Bujdoso, *Biochem. J.* **2008**, *409*, 367-375.
- [60] E. M. Marcotte, D. Eisenberg, *Biochemistry* **1999**, *38*, 667-676.

Table 1: Average G values and relative standard deviations of G for ideal secondary structures, involving at least N_R residues. Each structure was built by sampling ϕ and ψ angles from a gaussian distribution, centered on the ideal ϕ and ψ values with sigma=15 degree (see reference [25]).

Structure	$\langle G \rangle$	σ_G	N_R
α helix	-.04	0.02	>3
3_{10} helix	-.07	0.01	> 3
β Turn I	-.07	0.01	2,3
β Sheets	+.00	0.01	≥ 2
PPII	+.10	0.03	>3
π helix	-.01	0.02	>3

Table 2: Tyr-His HH-NE2 and Tyr-Tyr HH-HH NOE contact distances, calculated as $\langle \frac{1}{r^6} \rangle^{-\frac{1}{6}}$, for tyrosine and histidine respectively. For sake of clarity the first, second and third order momenta, together with the skewness, of the $\frac{1}{r^6}$ distribution are reported.

Aminoacids	$r_{NOE}(\text{\AA})$	$\mu_1(10^{-4})$	$\mu_2(10^{-5})$	μ_3	Skewness
Tyr64 - His72	3.12	11	2.3	0.0	-0.69
Tyr76 - His84	4.80	0.81	0.17	0.0	-0.19
Tyr64 - Tyr82	6.37	0.15	0.0044	0.0	-0.21
Tyr64 - Tyr88	6.52	0.13	0.0027	0.0	-0.23
Tyr70 - Tyr82	7.01	0.08	0.0014	0.0	-0.23

Figure Legends

Figure 1. G_{score} distributions calculated from the values explored by ChPrP1-267, ChPrP1-127, ChPrP128-242, ChPrP243-267 during the whole simulation run, compared with the cumulative distribution calculated from the protein NMR dataset, and with a distribution obtained by sampling randomly chirality indexes from a gaussian centered at zero chirality G_i index value (equation 2). It is worth noting a significantly lower G_{score} value with respect to the globular core for the less structured N- and C-terminal region. The overall G_{score} of ChPrP1-267 appears compatible with the distribution of the protein NMR dataset.

Figure 2. Time evolution of ChPrP1-267 end-to-end C_α distances for the three annealings, all of them showing a fast decrease. Highest and lowest distance of the 110 ns long MD simulation are shown as horizontal line.

Figure 3. A) ChPrP1-267 contact map calculated from MD simulations. The contacts around residues 130-160 point out the presence of β sheets. Less spread contacts are found around residues 60-260, indicating two opposite non overlapping sequences. α helices are found along the diagonal line (approximately at 160, 200 and 240 residue number). The intense points near the diagonal line indicate the presence of turns and 3_{10} helices. B) Comparison from the contact map of the globular core (ChPrP128-242) as obtained from MD trajectories and from the NMR structures [6]. The contact map concerning the ChPrP128-242 globular core from MD simulations B) reveals the stabilization of secondary structure elements with respect to the contact map from NMR structure of B). These involve the 3_{10} helix for 197-199 residues, between helix 2 and 3; an increment of the number of residues adopting the β sheet conformation is also shown from the wider contacts around 130-160 residues.

Figure 4. Time evolution of secondary structures of ChPrP according to DSSP criteria [32]. A) Time evolution of secondary structure in the 1-52 region. Unordered states, turns, 3_{10} helix (residues 5-7) and α helix

(residues 18-21) are present. Isolated β bridges are found for residues 16-17, 34, 41.

B): Time evolution of secondary structure in the 53-127 region. In the hexarepeat region (ChPrP53-88) unordered states are present, together with turns and 3_{10} helices, which are abundant especially around residues 53-54 and 67-69, namely the second and the third repeat, and in less extent in the final one (residues 83-88).

C): Time evolution of secondary structure in the 128-177 region. Helix 1 (ChPrP150-162) conformation is retained during MD simulations, except for the initial and the final portions of the helix. β sheets (136-137, 169-170), involve also residues 138-141 and 165-167. 3_{10} helix is found prevalently in residues 172-174.

D): Time evolution of secondary structure in the 178-211 region. It is worth noting the low flexibility of the helix 2 region (ChPrP178-195), that preserves its helical structure during the MD simulation. A 3_{10} helix is found for residues 197-199, while β bridges are detected for residues 201, 204 and in less extent for residue 209.

E): Time evolution of secondary structure in the 212-267 region. In the helix 3 region (ChPrP212-242), a high flexibility emerges. This helix spans also 3_{10} helices and turn regions inside its core.

Figure 5. Average number of water molecules along the backbone. The local minima are relative to regions with well defined secondary structures, i.e. the hexarepeat (turn and 3_{10} helix), helix 1, helix 2 and helix 3 after residue 226, consistent with the chirality analysis. Histograms below each of the values are reported as standard deviations on the ensemble of trajectories.

Figure 6. Average chirality index G of the three α helices of chicken prion protein. A): Average chirality index of ChPrP helix 1 (ChPrP150-162). The values indicate the coexistence with 3_{10} helix, understandable also from the standard deviations, spanning more negative G values than the α helix, see Table 1. Error bars are reported for each of the G values as standard deviations, referring to the ensemble of trajectories. Ideal α and 3_{10} helices patterns (dotted lines) are shown as comparison. B): Average chirality index G of helix 2 (ChPrP178-195). As the standard

deviations show, the structure is not really flexible, however the α helix structure stops at residue 192 and thus not involving residues 193-195. Error bars are reported for each of the G values, as standard deviations, on the ensemble of trajectories. Ideal α and 3_{10} helices patterns (dotted lines) are shown as comparison. C): Average chirality index G of helix 3 (ChPrP212-242). Error bars are reported for each one of the G values, as standard deviations, on the ensemble of trajectories. Ideal α and 3_{10} helices patterns are shown as comparison. The large standard deviations show that this helix is very flexible, except the 227-240 region.

Figure 7. Time evolution of the chirality index, G , relative to selected residues of the three helices. Consistently with the average chirality index trend, residue 223 of helix 3 (H3, 212-242) is structured as a 3_{10} helix for the first 60 ns of the simulation, converting to an α helix between 60-75 ns, then to a bend-coil between 75-90 ns and finally turning again to an α helix. Residue 156 of Helix 1 (H1, 150-162) shows the presence of a 3_{10} helix in the first 40 ns, while residue 186 of helix 2 (H2, 178-195) shows only α helix chirality index values. Snapshots of the three helices at different simulation times are also shown. β turn is colored in red, α helix in violet and 3_{10} helix in orange.

Figure 8. Chirality index along the backbone for the different prion species, frog (pdb code 1XU0), turtle (pdb code 1U5L), chicken (pdb code 1U3M) and human (pdb code 1QM1), aligned as in ref. [6].

Figure 9. Chirality index, G , averaged among the trajectories of the ChPrP1-267. The more negative peaks ($-0.1 < G < -0.06$) underline the presence of turn regions; the three negative oscillations centered at -0.05 G values underline the presence of α helix (main residues: 156, 186, 227). 3_{10} helices are recognized from the involvement of more than three residues having its typical chirality (See Table 1). The high positive peaks, namely G values greater than 0.05, are typical of a poly-L-proline II conformation, which is present, although it involves a little extent of residues, along the backbone. Histograms below each of the G values are reported as standard

deviations on the ensemble of trajectories.

Figure 10. ChPrP1-267 typical conformation sampled from the MD simulations. The hexarepeat region, colored in green, shows the hydrogen bond between the imidazole nitrogen, N_{ϵ} , of the histidine 64 and the phenolic hydrogen of tyrosine 72. This interaction occurs also in the tetra-hexarepeat fragment, previously reported [12]. Helix 1 is shown in red, Helix 2 in purple and Helix 3 in orange. N (Met 1) and C-terminal (His 267) are shown in blue and red, respectively.

Figure 11. Chirality index, G , averaged among the trajectories of the ChPrP1-267, focused on the hexarepeat region, ChPrP53-88. The chirality index pattern of the tetra-hexarepeat fragment, previously simulated [12], is also shown by comparison. First, the index shows a periodical pattern, as found in the tetra-hexarepeat fragment and, consistently with this latter one, turn regions are frequently populated together with 3_{10} helices (68-73 and 80-85). Error bars are reported for each one of the G values as standard deviations on the ensemble of trajectories.

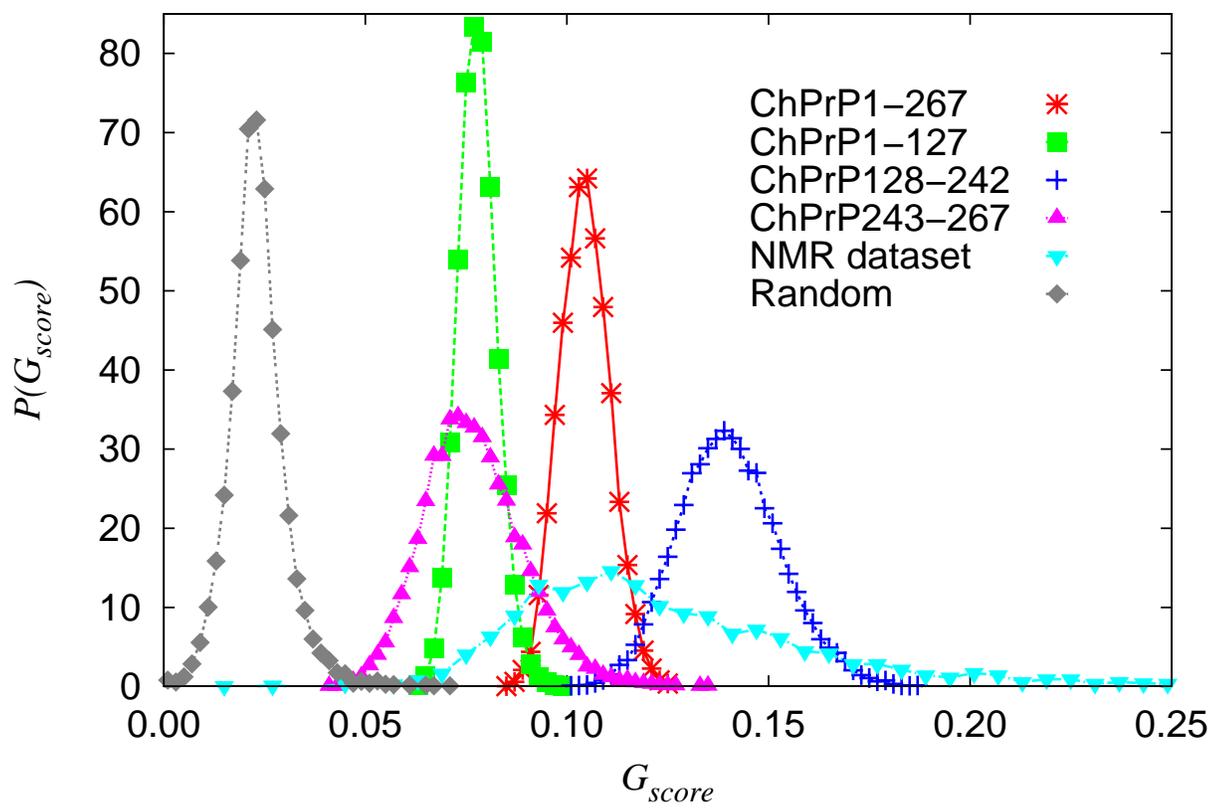


Figure 1:

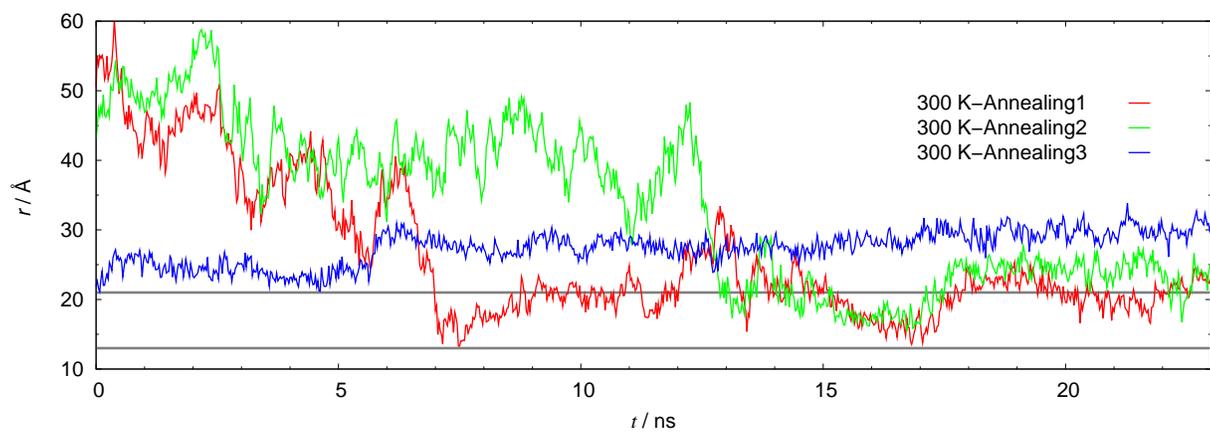
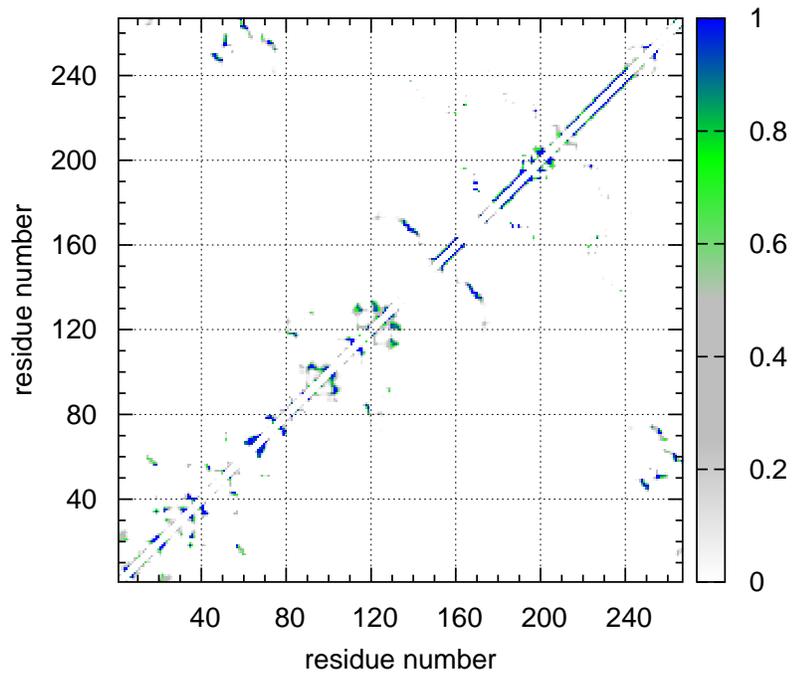
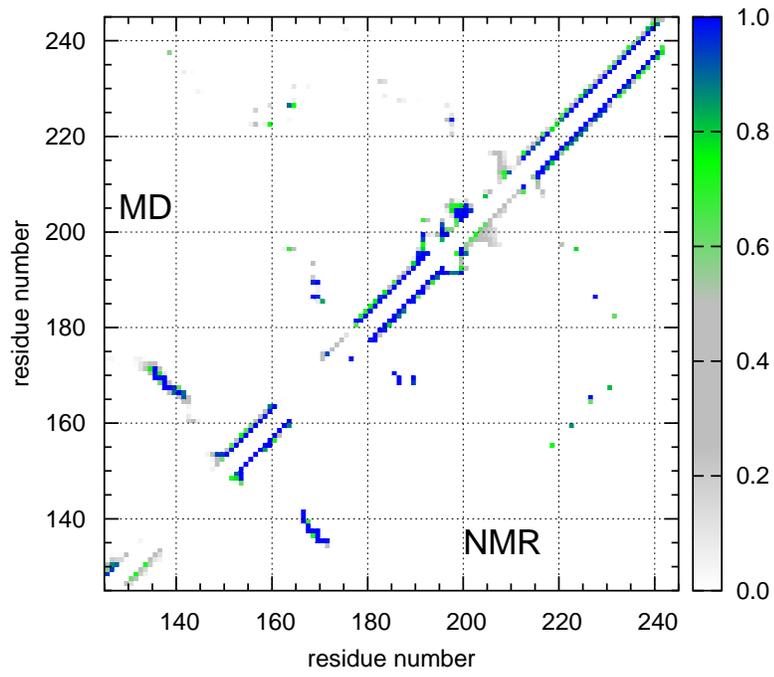


Figure 2:



A)



B)

Figure 3:

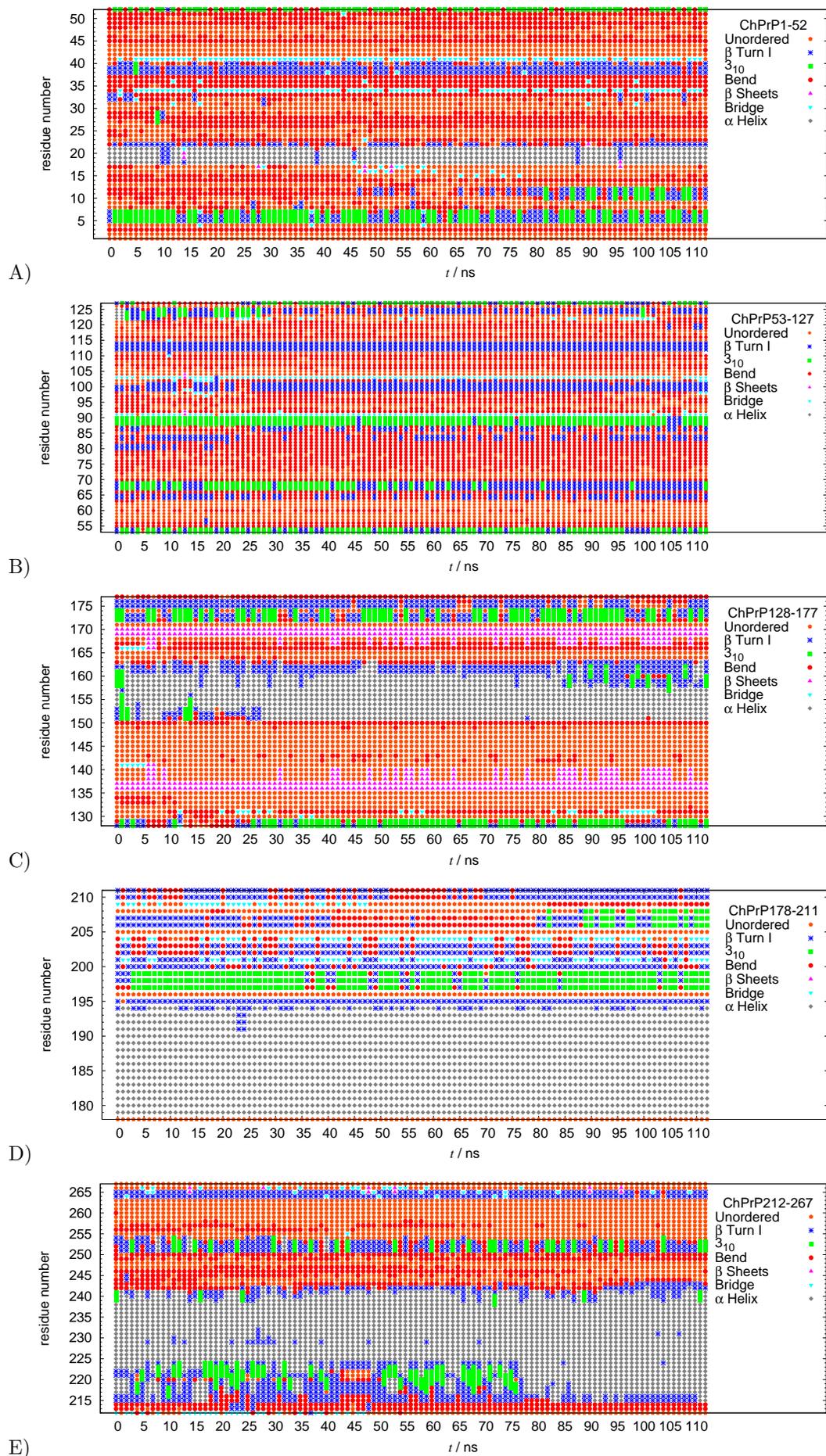


Figure 4:

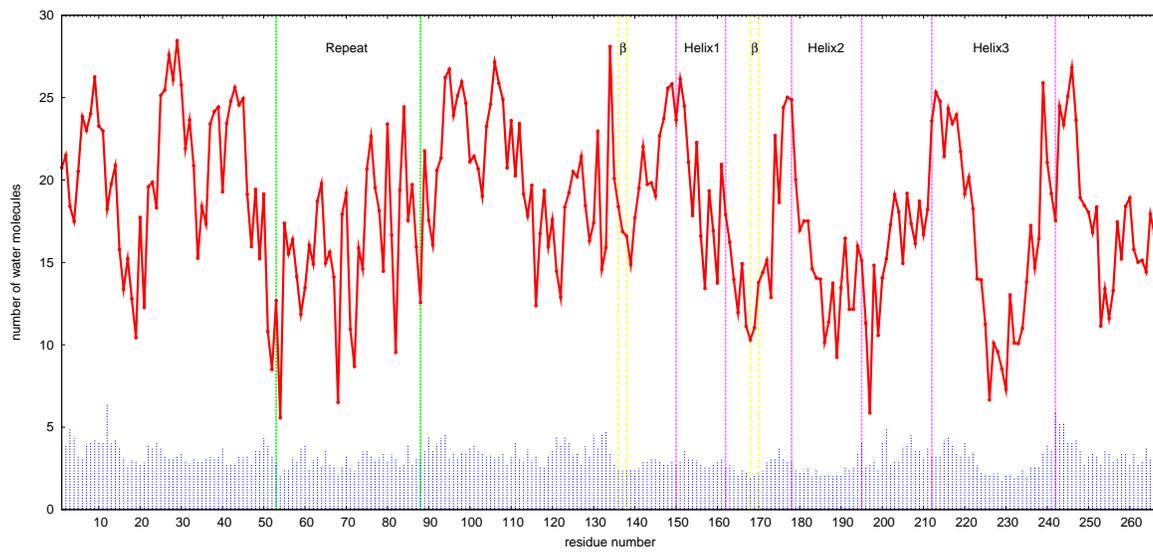


Figure 5:

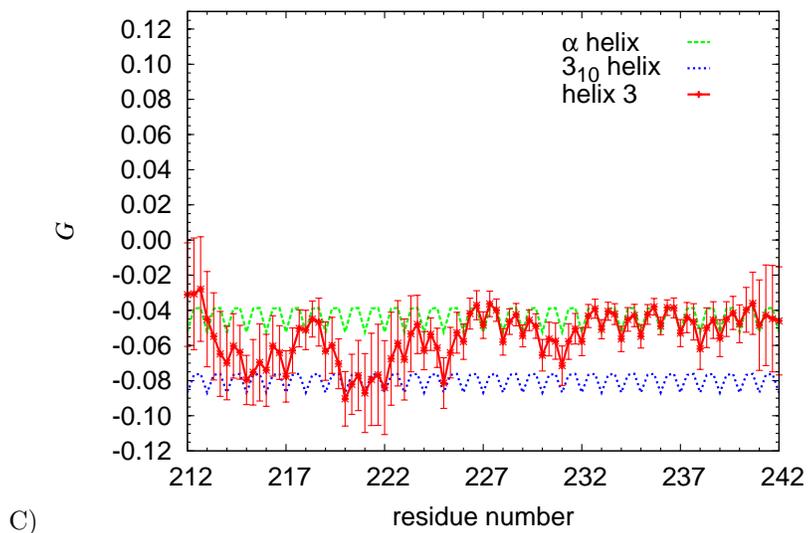
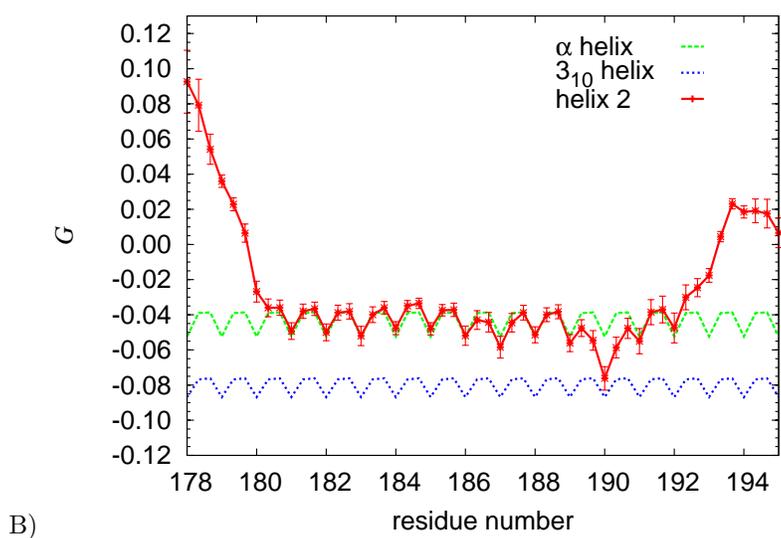
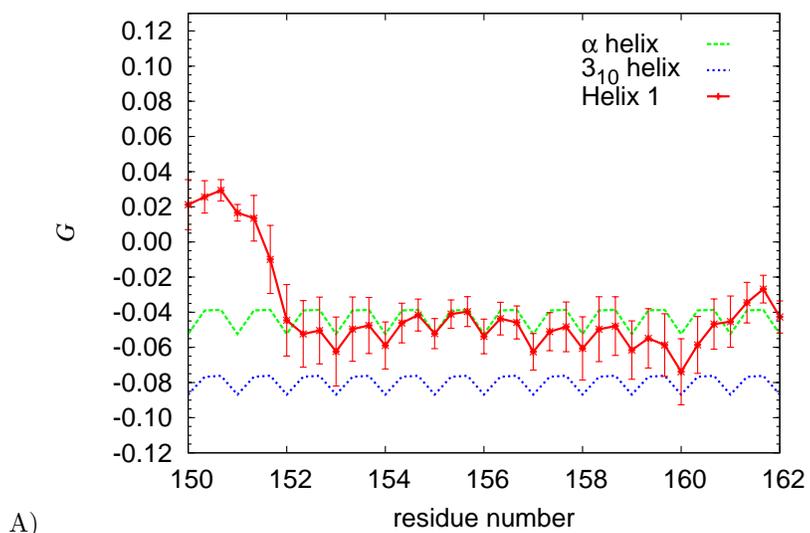


Figure 6:

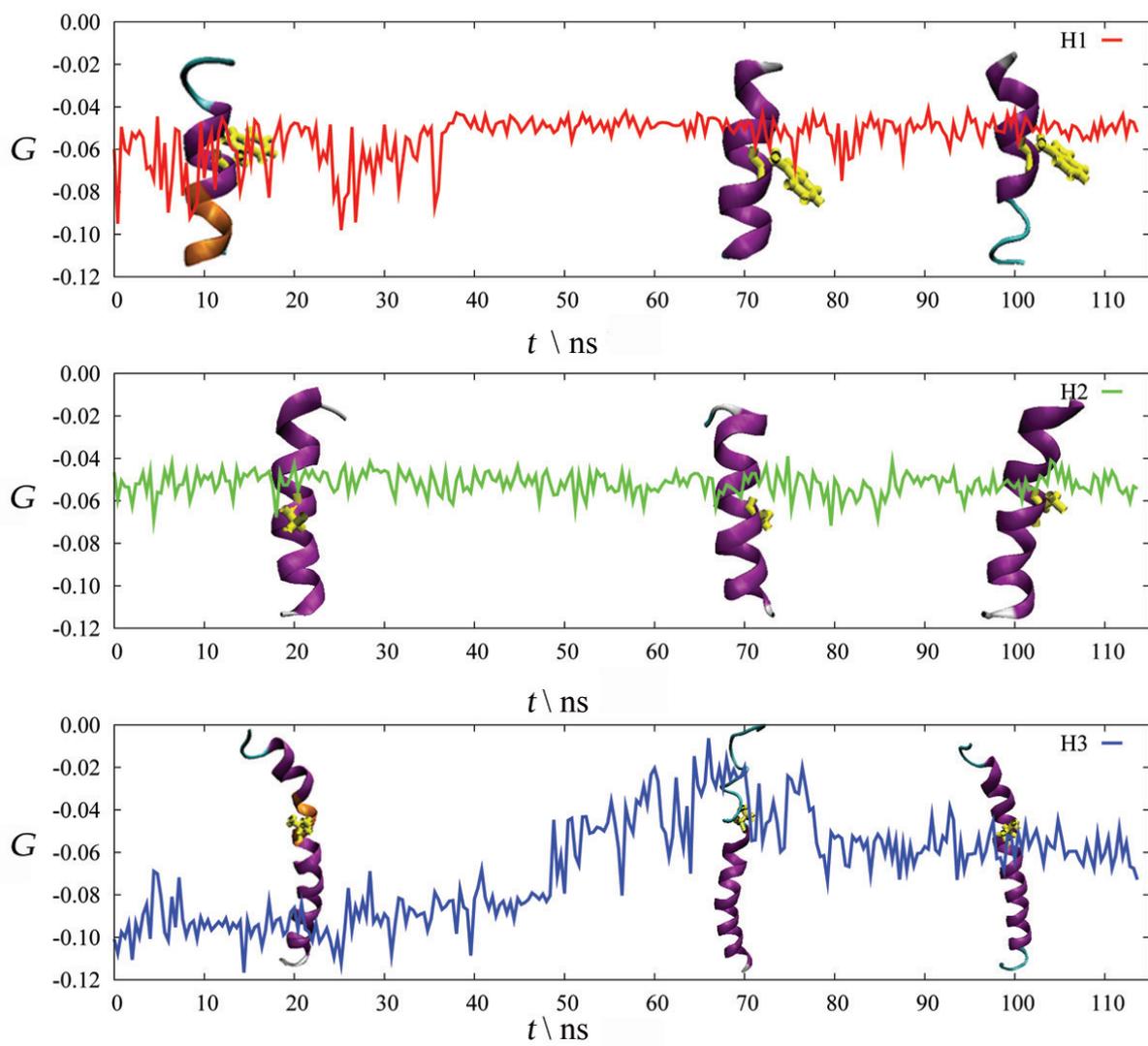


Figure 7:

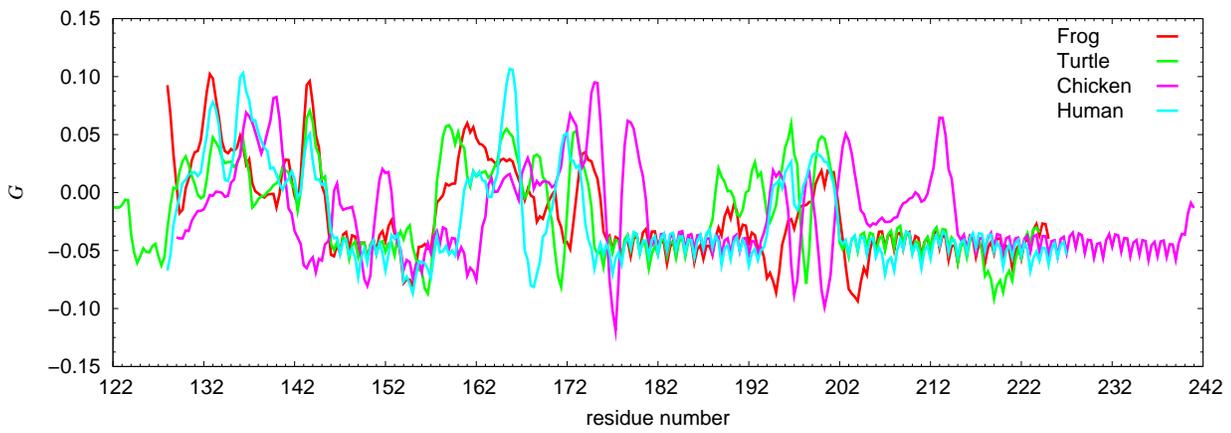


Figure 8:

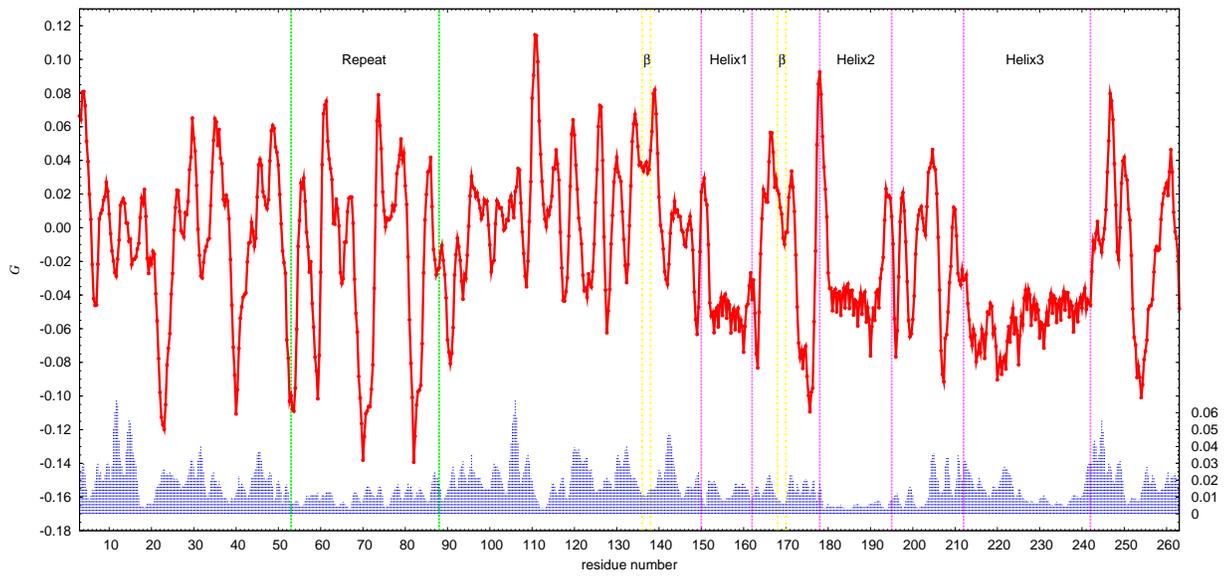


Figure 9:

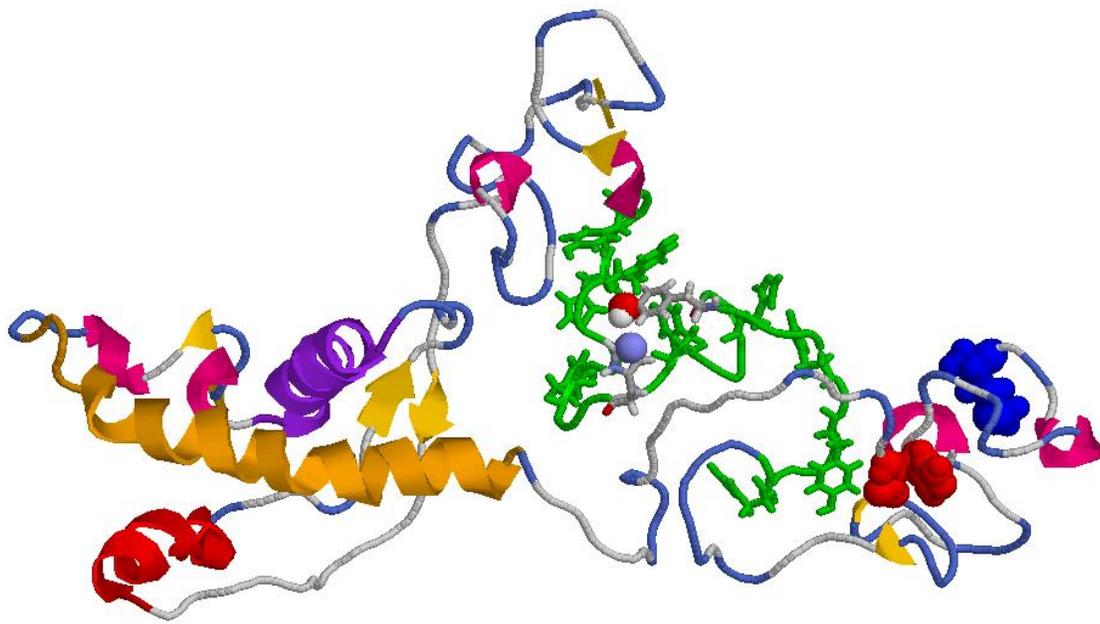


Figure 10:

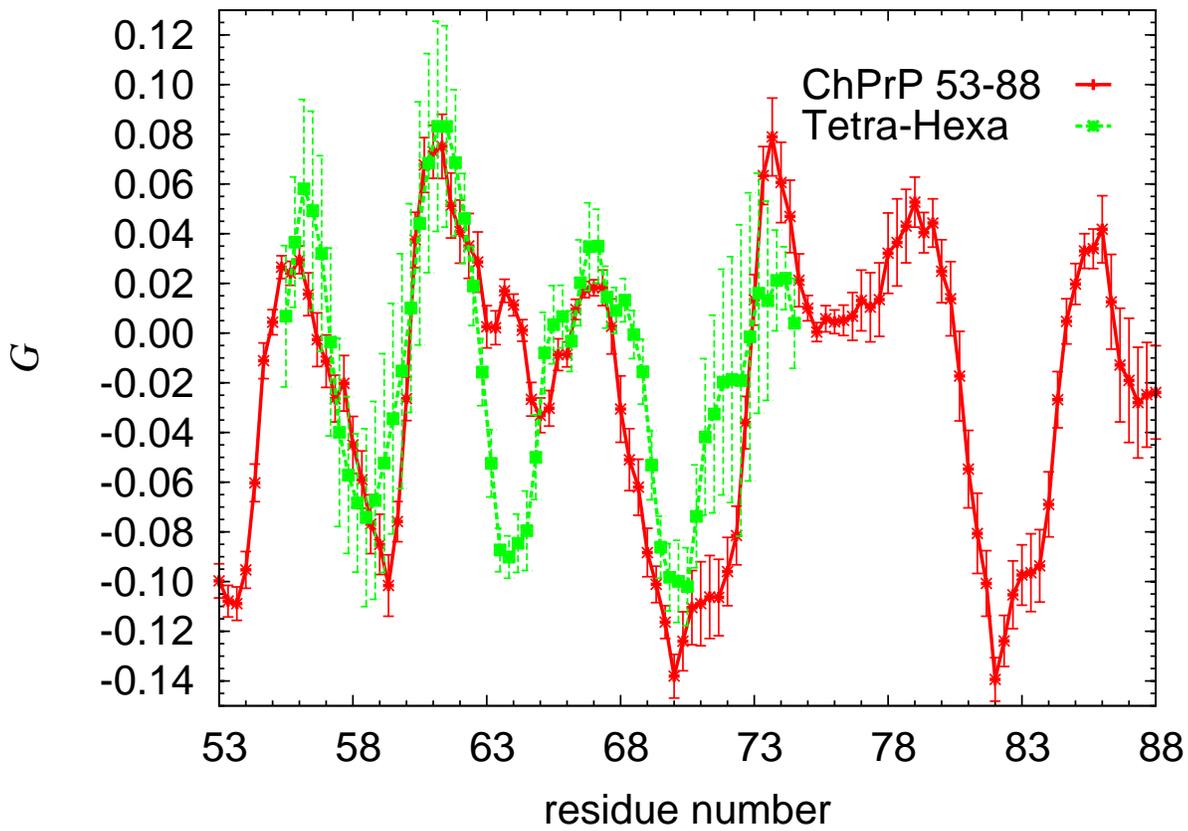
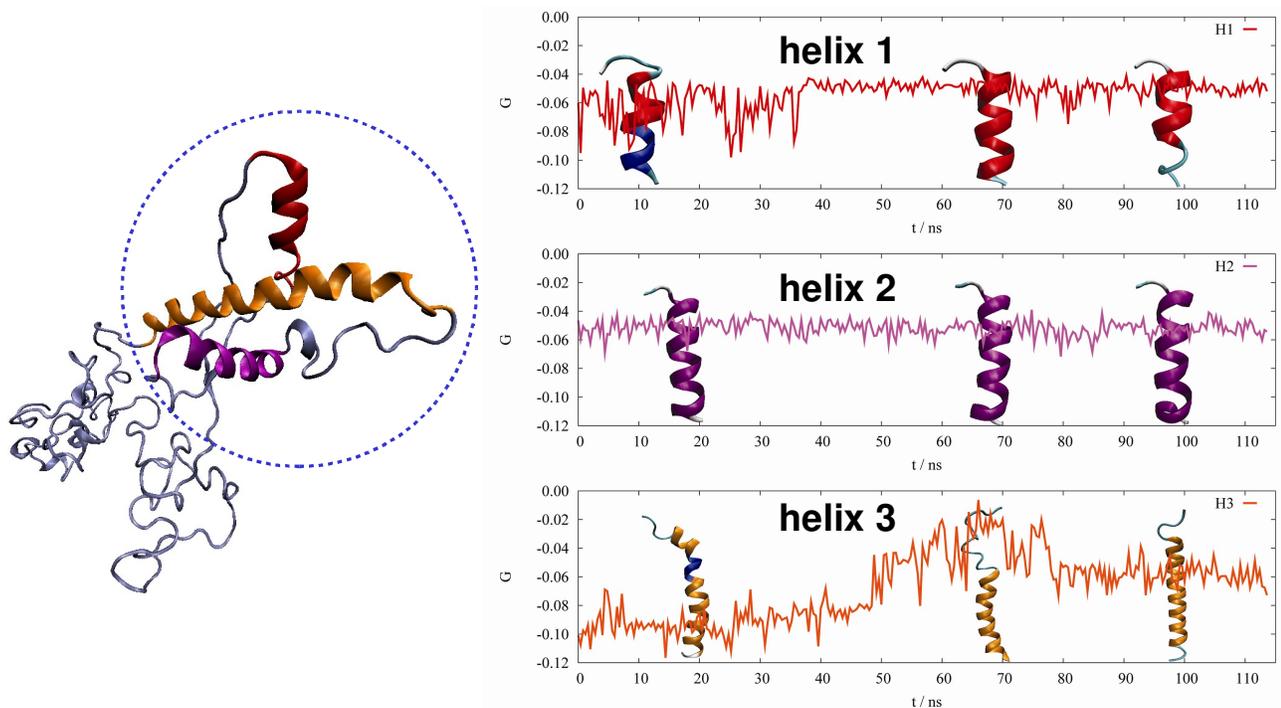


Figure 11:

4 TOC



Simulations of the chicken prion protein and chirality analysis provide possible explanations of the resistance to proteases and to the fold stability of avian prion in the globular core. The three alpha helices highlight different secondary structure propensities, being helix 2 the most stable one. α helix is shown in red for helix 1, in violet for helix 2 and in orange for helix 3.

3_{10} helix is shown in blue and turn in cyan.

5 Supplemental Materials

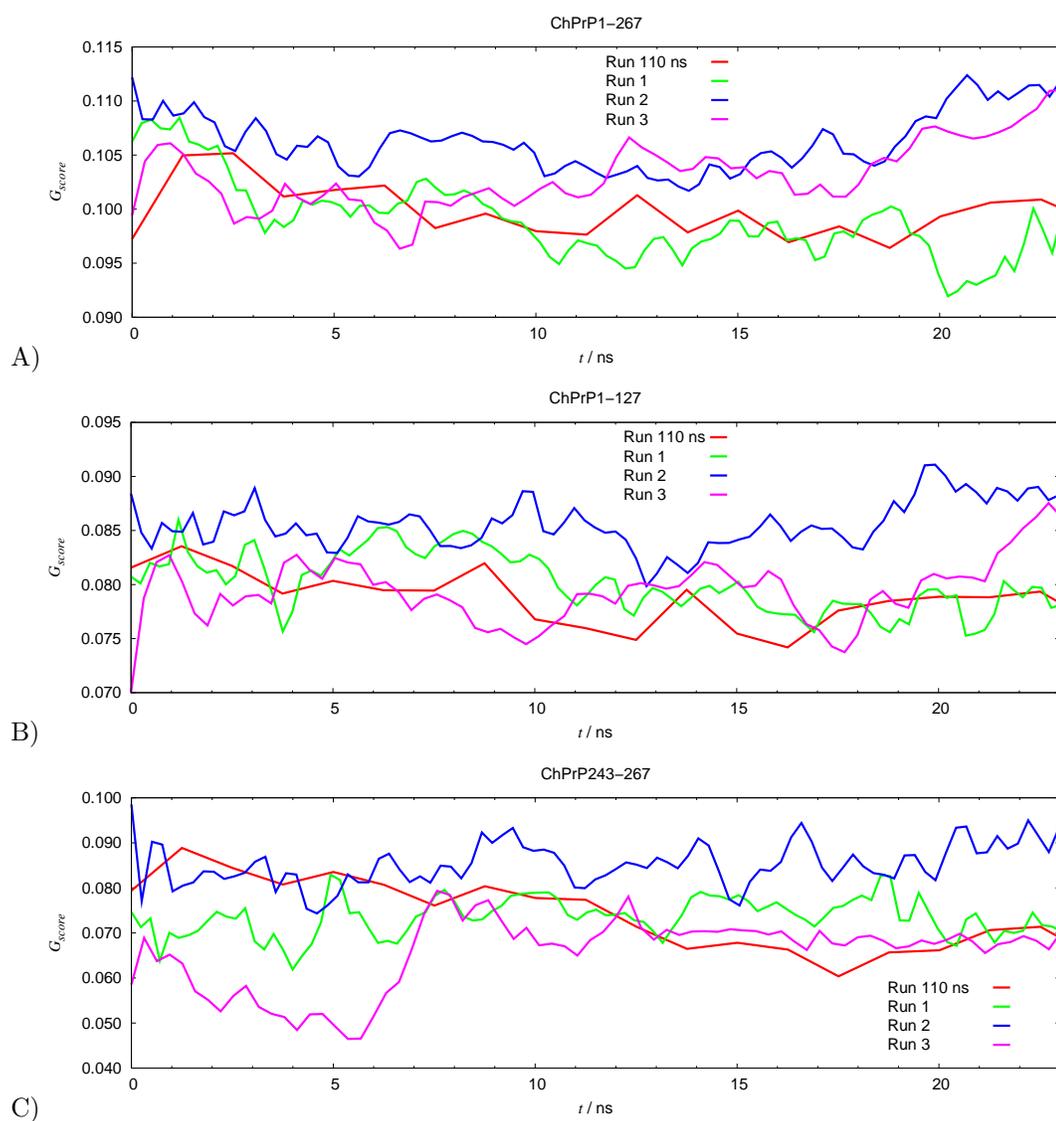


Figure 12: Time evolution of the G_{score} for ChPrP1-267 A), and the N- and C-terminal portions, ChPrP1-126 and ChPrP243-267, for the three annealings. 110 ns long MD run is reported as comparison. The first annealing tends to adopt a likely structure, while annealing 3, especially in the C-terminal portion (see C)) starts with very unfavorable dihedral angles (low G_{score} values), reflecting somehow the end-to-end distance evolution, much higher than the 110 ns long MD.