An NMR and Molecular Dynamics investigation of the Avian Prion Hexarepeat conformational features in solution.

Adriana Pietropaolo[1], Luca Raiola[2], Luca Muccioli[1], Giustiniano Tiberio[1], Claudio Zannoni[11], Roberto Fattorusso[2], Carla Isernia[2], Diego La Mendola[3], Giuseppe Pappalardo[3], Enrico Rizzarelli[3,4]

[1] Dipartimento di Chimica Fisica e Inorganica, Università di Bologna and INSTM, v.le Risorgimento 4, 40136 Bologna (Italy)

[2] Dipartimento di Scienze Ambientali, Seconda Università di Napoli, via Vivaldi 43, 81100 Caserta (Italy)

[3] CNR-Istituto di Biostrutture e Bioimmagini Catania, v.le A. Doria 6, 95125 Catania (Italy)

[4] Dipartimento di Scienze Chimiche, Università di Catania, v.le A. Doria 6, 95125 Catania (Italy)

Abstract

The prion protein is a copper binding glycoprotein that in mammals can misfold into a pathogenic isoform leading to prion diseases, as opposed, surprisingly, to avians. The avian prion N-terminal tandem repeat is richer in prolines than the mammal one, and understanding their effect on conformation is of great biological importance. Here we succeeded in investigating the conformations of a single avian hexarepeat by means of NMR and Molecular Dynamics techniques. We found a high flexibility and a strong conformational dependence on pH: local turns are present at acidic and neutral pH, while unordered regions dominate at basic conditions.

Introduction

The prion protein (PrP^C) is a cell surface glycosyl-phosphatidylinositol (GPI) anchored glycoprotein [1,2], whose biological function has not yet been identified, even if it has been established that in mammals its conformational change gives rise to the β-sheet rich and pathogenic isoform PrP^{Sc} [1,2]. Avian species also express prion protein, but no evidence of neurodegenerative disorders have been reported among them [3-5] and chicken prion protein ($ChPrP^C$) has initially been described to induce acetylcholine-receptor activity [4]. This protein shares about 33% of primary sequence identity with the mammalian one and some essential features are conserved [3]. In particular both proteins possess: i) multiple N-glycosylated sites; ii) an amino-terminal signal sequence that is removed in the mature protein; iii) a carboxy-terminal signal that is eliminated when the mature protein is linked to GPI, and iv) an N-terminal domain featured by tandem amino acid repeats (PHNPGY in avians, PHGGGWGQ in mammals) followed by a highly conserved hydrophobic core [3,4]. Differently from mammalian PrP^C, the digestion of $ChPrP^C$ with trypsine or proteinase K produces peptide fragments stable to further proteolysis. One of these fragments comprises the 49-129 sequence, consisting of a large part of the N-terminal domain [6]. This resistance to proteolysis may suggest a compact domain of the proline/glycine rich N-terminus, although the hexarepeat amino acid sequence seems to show no tendency towards a particular structured conformation [6]. More recently, NMR results showed that the $ChPrP^C$ globular domain contains three α-helices and two short antiparallel β-sheets, with a spatial arrangement similar to that observed in mammalian PrP^C [7]. The N-terminal domain results to be flexible and unstructured and the 50-73 residues, encompassing part of the tandem hexarepeat sequence, have not yet been individually assigned by NMR [7].

Besides the wide conformational freedom of the N-terminal backbone, the significant content of proline residues may in principle bring about further complications in the structure determination of this part of protein, due to the presence of an equilibrium between *cis* and *trans* form in each Xxx-Pro peptide bond, even if the stabilization of a preferred isomer may be assisted by other residue side chains, particularly histidine and tyrosine.

In native and folded proteins the Xxx-Pro bond exists essentially in either *cis* or *trans* form because the interactions with neighbouring groups favours one of them, but an equilibrium exists in unfolded proteins. In small peptides containing proline residues, the two forms are almost

[1] Corresponding author: prof. Claudio Zannoni, Dipartimento di Chimica Fisica e Inorganica,viale Risorgimento 4, I-40136 Bologna (Italy); fax: +39 051 644702; e-mail: Claudio.Zannoni@unibo.it.

isoenergetic and are both present, with an observed increase of *trans* isomer going from polar to non-polar solvents [8-10]. In solution, polyproline peptides have been shown to exist predominantly in the PPII conformation, a left-handed helix in which all the residues are in the *trans* form. Also short peptides not containing proline are supposed to partially adopt this conformation, but it seems quite sure that protein domains having an intra-chain PXXP motif fold into a PPII helix. One of these motifs has been identified in mammalian as well as avian prion proteins and it corresponds to the region 101-104 in mouse PrP^C and 107-111 in chicken PrP^C [11]. In this context, MD studies performed at various pH provide precious information on the different conformational states populated in the 10-100 nanosecond time scale, as well as helping the interpretation of NMR r in the same conditions, when available.

With the long-term purpose of clarifying the role of the repeat regions in prion protein biological function, here we use both techniques in a pH and solvent dependent study concerning the conformation of the single hexarepeat PHNPGY (1-HexaPY), with both N- and C-termini blocked by acetylation and amidation respectively. This peptide, which is replicated seven times in the 53-94 region of the chicken prion protein, represents a starting model system for future studies concerning longer portions of avian hexarepeat and human octarepeat sequences.

Materials and methods

NMR Measurements.

The peptide under study was synthesized as previously reported [12]. All the experiments were carried out at 500 MHz on a VARIAN INOVA UNITY PLUS located at the Department of Chemical Sciences, Catania (Italy), and on a VARIAN UNITY 500 spectrometer, located at the Department of Environmental Sciences, Caserta (Italy). Spectra were processed using the VARIAN VnmrJ and XEASY [13] software. Sample solutions (5mM) were prepared in TFE-d_3/H_2O (80/20 v/v) and H_2O/D_2O 90/10 (v/v). NMR spectra for the three-dimensional structure determination were collected at 300 K and referenced to external TMS (δ = 0 ppm); the pH of the aqueous solution was

adjusted to 4.2. The dependence of the amide chemical shifts on the temperature was observed in the range 300-311 K. Furthermore, for all the solvent systems the possible occurrence of aggregation was verified by analyzing the cross peak patterns in ROESY spectra, recorded at a peptide concentration of 0.8 mM. Deuterated D_2O (99.9% relative isotopic abundance) and TFE-d_3 (99%) were purchased from Cambridge Isotope Laboratories. Mono (1D) and two dimensional (2D) spectra were accumulated with a spectral width of 6000 Hz in H_2O/D_2O and 4800 Hz in TFE/H_2O (80/20 v/v). 2D experiments DQFCOSY, TOCSY, ROESY and NOESY [14] were recorded in the phase sensitive mode using the States-Haberkorn method. Water suppression was achieved by DPFGSE sequence [15]. TOCSY, NOESY and ROESY spectra were acquired with mixing time of 70, 250 and 150 ms, respectively. Typically, 64 transients of 4K data points were collected for each of the 256 increments; the data were zero filled to 1K in ω_1. Squared shifted sine-bell functions were applied in both dimensions prior to Fourier transformation and baseline correction.

Structure Calculations.

Due to unfavourable correlation time, NOESY experiments were scarcely informative and not suitable for structure calculations. Therefore distance restraints for structure calculations were derived from the cross-peak intensities in ROESY spectra, recorded in H_2O/D_2O and TFE/H_2O (80/20 v/v). The ROESY cross peaks were manually integrated using the XEASY software [13] and converted to upper distance constraints according to an inverse sixth power peak volume-to-distance relationship for the backbone and to an inverse fourth power function for side chains, by using the CALIBA module of the CYANA program [16]. Distance constraints together with the obtained scalar coupling constants were then used by the GRIDSEARCH module, implemented in CYANA, to generate a set of allowable dihedral angles. Structure calculations, which used the torsion angle dynamics protocol of CYANA, were then started from 100 randomized conformers. The 20 conformers with the lowest CYANA target function were further refined by means of unrestrained energy minimization *in vacuo*, using the GROMOS 96 [17] force field (FF) within the program SPDB viewer [18]. Several cycles of steepest descent were repeated until the energy

difference between two successive steps was less than 10^{-3} kJ mol^{-1}. The structure analysis has been performed with the program MOLMOL [19]. Consistently with suggestions from a recent work [7] we also minimized the conformers in a water shell and in a box with TFE/H_2O 80/20 with a conjugate gradient method and tolerance of 10^{-3} kJ mol^{-1}, using ORAC 4.0 program [20] and the Amber94 FF [21].

Molecular Dynamics.

All the simulations were run in water using ORAC 4.0 [20] and the Amber94 FF [21]; a cubic sample containing one 1-HexaPY chain and 1184 water molecules with periodic boundary conditions (PBC) was studied in the isothermal-isobaric ensemble (NPT, P=1 atm, T=300 K); temperature was controlled with a Nosé-Hoover thermostat [22] and the SPC model [23] was used for water. The ESP charges of deprotonated tyrosine, not available in the FF, were calculated at HF/6-31G* level for the N-acetyl-(L)-tyrosinate amide after geometry optimization, following the Amber charge fitting philosophy. An r-RESPA multiple time-step algorithm with a potential subdivision specifically tuned for proteins [20] was used for integrating the equations of motion, using a time step of 10 fs. Due to its partial double bond character, the torsional barrier around the peptide bond Xxx-Pro is not thermally overcome [24] in the 1-100 nanosecond time scale, typical of MD simulations. Consequently, the different isomers, originated from the Asn-Pro and Ac-Pro ω torsion angles (N and C-termini were capped with acetyl and amide groups, respectively), were simulated separately. For each isomer, three different protonation states were studied to mimic the different pH conditions. Taking into consideration the pK_a of histidine and tyrosine (6.34 and 9.77 respectively [12]), we can assume in our simulations that at acidic pH, the histidine is protonated and tyrosine is in its neutral form (labelled H^+Y), while at neutral pH both histidine, in the δ form, and tyrosine are in their neutral states (labelled HY) and finally that at basic pH histidine is in the δ neutral form and tyrosine is deprotonated (labelled HY^-). One chloride ion and one sodium ion were added at acidic and basic pH conditions respectively, to ensure charge neutrality. The total

simulation time was about one hundred nanoseconds for each of the twelve simulations of the four isomers in three different protonation states. Each run was equilibrated for at least 30 ns to avoid a starting configuration bias and, after equilibration, we checked that both volume and total energy fluctuated around their average value, without systematic drifts. The trajectory analysis was performed on 73 ns-long production runs, with configurations stored every 5 ps. To assess if the force field was able to predict PPII conformations, we also carried out a 30 ns simulation of 9-mer poly-L proline in the zwitterion form (NPT, P=1 atm, T=300 K) using a cubic box with 2123 explicit water molecules and PBC. The PPII structure (ϕ= -75°, ψ= 145°) [25] was retained for the whole simulation time, with the exception of the C-terminal proline, whose average ψ angles resulted in equilibrium between -30° and 150° values.

During the 1-Hexapy simulations, the presence of hydrogen bonds between carbonyl oxygens and amide hydrogens belonging to different amino acids was estimated following the DSSP criteria [26]. To investigate their persistence we also calculated the variation of a given O-H distance and the angle between the C-O and N-H vectors (τ) as a function of time, on one hand to understand the orientation of the two atoms involved in the hydrogen bond and on the other to calculate the τ angle/O-H distance/energy surface so as to decide if a hydrogen bond was present. The hydrogen bond energy used in this work is purely electrostatic and it is expressed as $E_{HB}=u\ q_1q_2(-1/r_{CN}+1/r_{CH}-1/r_{OH}+1/r_{NO})$ [26], where q_1 and q_2 are respectively 0.42*e* and 0.20*e*, with *e* the unit electron charge, u=332 the conversion factor from e^2/Å to kcal/mol and r_{CN}, r_{CH}, r_{OH} and r_{NO} the inter-residue distances in Å; a hydrogen bond is considered to be present when $E_{HB} \leq$ -0.5 kcal/mol, $\tau \leq 60°$ and $r_{OH} \leq 3$ Å.

Results and Discussion

NMR Analysis and Structure Determination.

The conformational behaviour in solution of the 1-HexaPY hexapeptide was studied by proton NMR spectroscopy in two different solvents: water (H_2O/D_2O 90/10), because of its biological

relevance and trifluoroethanol (TFE/H_2O 80/20 v/v). The TFE/H2O 80/20 mixture, often used as a membrane-mimicking solvent [see, e.g. 27] and as a hydrogen bond promoting solvent, here can provide indications on the possible folded states of the peptide, which may not appear in water, and of the conformation of the hexarepeat region when inserted in a biological membrane. Identification of the complete spin systems in both solvents was readily accomplished by the homonuclear J-correlated 2D techniques TOCSY and DQF-COSY. Since these were not sufficient to determine their position in the sequence, sequential NOE connectivities between backbone protons were employed to unambiguously assign all the amino acids and their position in the peptide chain. Complete chemical shift assignments in H_2O/D_2O and TFE/H_2O (80/20 v/v) are reported in Table 1 and 2 respectively. $^3J_{NH\alpha CH}$ and $^3J_{\alpha CH\beta CH}$ coupling constants and temperature coefficients are reported in Table 3.

A key feature of 1-HexaPY is the presence of two proline residues in the sequence; the first one is located at the beginning of the chain, the other at a central position in the Ac-Pro^1-His^2-Asn^3-Pro^4-Gly^5-Tyr^6-NH_2 sequence: consequently Pro^4 plays a major role in determining the distribution of conformers in solution. Starting with the TFE/H_2O solvent, we find that the hexapeptide shows a well resolved 1D ^1H spectrum, characterized by a good dispersion of the proton resonances and a predominant set of sharp peaks for amide protons. Its features indicate the existence of two main families of conformers: a dominant one representing 95% of the population, and a minor one covering less than 5%. Two observable sets of resonances, presumably due to the *cis-trans* isomers about the Xxx-Pro bond and with relative populations 83% and 17%, are clearly present also in the proton spectrum of 1-HexaPY in H_2O/D_2O. The hexapeptide assumes a *trans* conformation for the Ac-Pro and Asn-Pro peptide bonds in TFE/H_2O, as indicated by the strong NOEs between the H^δ of Pro^1 and Ac, and between the H^δ of Pro^4 and H^α of Asn^3; also in H_2O/D_2O, both the Xxx-Pro bonds of the major conformer family assume a *trans* conformation as confirmed by the presence of two NOE contacts between the H^δ of Pro^1 and Ac, and between the H^δ of Pro^4 and H^α of Asn^3 (see Tables A and B of Supplementary Materials). The presence of a less populated conformer group is

presumably due to a *cis* arrangement of the Asn^3-Pro^4 peptide bond since an Ac-Pro^1 NOE contact is not observed in the ROESY spectrum and, moreover, the H^α resonance of Asn^3 of the minor conformer is obscured by the water signal. All amide protons present a linear and negative dependence of their chemical shifts on temperature; the backbone amide protons of residues His^2 and Tyr^6 in TFE/H_2O and Tyr^6 in H_2O/D_2O show low temperature gradients in both solutions (Table 3). In particular, the value obtained for the Tyr^6 was the lowest and identical in both solvent systems indicating that this amide proton has a good tendency to be solvent shielded likely by its aromatic side chain, as the up-field shifted chemical shift of Tyr^6 amide proton also suggests. The scalar coupling constants, $^3J_{NH\alpha CH}$, for residues His^2, Asn^3 and Tyr^6 show values in the range 6-8 Hz (Table 3), not evidencing the formation of a canonical folded conformation. On the basis of $^3J_{\alpha CH\beta CH}$ coupling constants and αCH-βCH, NH-βCH NOEs, the preferred conformations of the side chains were identified [28] as a *trans* arrangement for all the amino acids except for Asn^3 which, only in TFE/H_2O, appears to prefer a *gauche* conformation. These data were confirmed by an analysis of rotamer populations [29] reported in Table C of the Supplementary Materials. After a careful analysis of the ROESY spectra, 50 proton-proton NOE cross peaks were assigned and integrated for 1-HexaPY in TFE/H_2O (80/20 v/v), and 37 in H_2O/D_2O, for the principal conformer family. A list of the relevant distances derived from the integration of the ROESY peaks is reported in Tables A and B of the Supplementary Materials; it should be noted that in both solvents a correlation between the NH of Gly^5 and the NH of Tyr^6, corresponding to a calculated distance of about 3 Å, is clearly observed. Furthermore, only in TFE/H_2O a medium range NOE value could be assigned as a weak correlation between the Pro^4 αCH and the Tyr^6 amide proton. Four $^3J_{NH\alpha H}$ and ten $^3J_{\alpha H\beta H}$ coupling constants were extracted from the ^1H monodimensional spectra of 1-HexaPY either in TFE/H_2O (80/20 v/v) and H_2O/D_2O (Table 3). Stereospecific assignments for His^2 (βCH_2 protons) and Asn^3 (βCH_2 and δCH_2 protons) of 1-HexaPY in TFE/H_2O (80/20 v/v) were derived from the input data, using the GRIDSEARCH module of CYANA software. The final input files for

the structure calculation contained 26 meaningful distance constraints (16 intraresidue and 10 short-range) and 13 angle constraints for the 1-HexaPY in H_2O/D_2O, and concerning the study in TFE/H_2O (80/20 v/v), 26 meaningful distance constraints (15 intraresidue, 10 short- and 1 medium-range) and 23 angle constraints. These constraints were then used to generate a total of 100 structures and among them the 20 structures with the lowest target function values were selected and energy minimized. The best 20 CYANA conformers of 1-HexaPY in H_2O/D_2O (target function 0.77) and TFE/H_2O (80/20 v/v) (target function 1.05) were then minimized in *vacuo* and in a solvent shell; a stick drawing of the representative structures is reported in Figure 1 of the Supplementary Materials. In TFE/H2O (80/20 v/v), the NMR structure of 1-HexaPY is well defined (RMSD 0.16 in the 1-4 region; RMSD= 0.38 in the region 1-5) and the average dihedral angles of the 20 NMR structures minimized in vacuum and in the proper solvent are reported in Tables 4 and 5. In this solvent system Pro^4 assumes ϕ and ψ angles close to the PPII conformation, while Pro^1, His^2, Asn^3 conformational angles only reflect a propensity towards this elongated conformation.

The NMR structure in H_2O/D_2O is reasonably defined (RMSD=0.43 in the 1-4 region; RMSD= 0.59 in the region 1-5) and elements of PPII conformation are present (Figure 1, a and b), which include the Asn^3 and Pro^4 residues; the presence of Gly^5 causes the break of this secondary structure and allows Tyr^6 to experience an higher conformational freedom.

Molecular Dynamics simulations

The NMR study discussed above gives a fairly clear picture on the averaged conformation adopted by the 1-HexaPY peptide chain. However, for technical difficulties the study was limited to acidic pH values, thus allowing only a partial interpretation of other available data, such as the CD spectra, previously reported by some of us [12], that cover a wide range of pH values. In order to get a more complete understanding at molecular level and in particular to further investigate the pH dependence of 1-HexaPY conformational states in aqueous solution, a fully atomistic Molecular Dynamics (MD) study was then carried out in water.

It is known that the *trans* conformation of proline tends to stabilize type I β turn (ϕ_{i+1}= -60°, ψ_{i+1}= -30°, ϕ_{i+2} = -90°, ψ_{i+2} = 0°) [30] and PPII structures (ϕ= -75°, ψ = 145°) [25]; conversely the *cis* conformation of proline stabilizes a special type of turn, devoid of intrachain hydrogen bonds, called VI [31] (VIa1 ϕ_{i+1}= -60°, ψ_{i+1}= 120°, ϕ_{i+2}= -90°, ψ_{i+2}= 0°, VIa2 ϕ_{i+1}= -120, ψ_{i+1}= 120°, ϕ_{i+2}= -60°, ψ_{i+2}= 0°, VIb ϕ_{i+1}= -135°, ψ_{i+1}= 135°, ϕ_{i+2}= -75°, ψ_{i+2}= 160°). To claim the presence of such conformers, the existence of consecutive amino acid residues with the typical angles is required, thus we calculated the percentage of type I β turn and PPII considering both 2, 3, 4 consecutive residues (see Table D of the Supplementary Materials), using the standard tolerance interval of ± 30° for all the dihedrals. During the MD simulations, no more than two consecutive residues with type I β turn ϕ and ψ dihedrals were found, indicating a high conformational flexibility of the hexarepeat. The time evolution of the conformers (Figure 2), calculated taking into account only two residues, proves the fast flipping of backbone dihedrals, showing also a prevalence of angles typical of β turn and unordered structures.

The analysis of i-i+3 hydrogen (H) bonds and of intrapeptide contacts (evaluated as $C_{\alpha i}$-$C_{\alpha i+3}$ distances inferior to 7 Å [32]), revealed the pivotal role of asparagine and of the conformation of prolines on the peptide structure: in fact, H-bonds and contacts always involve asparagine, despite its tendency to adopt PPII dihedrals registered both by NMR and MD, and they exist only in regions containing *trans* prolines (cf H-bond percentages in Table 6). In particular we found a NPGY β turn fulfilling the DSSP criteria [26], and a second turn region was detected, albeit in lower extent, in the AcPHN sequence. The existence of turn structures is indicated also by $C_{\alpha i}$-$C_{\alpha i+3}$ distances, quite large for PHNP and HNPG segments, and short in the NPGY C-terminal sequence (Table E of the Supplementary Materials). Summarising, the peptide shows a tendency to adopt β turn conformations, which on the other hand exist only in a fraction of the trajectory and quickly interconvert with unordered ones.

pH-dependent features. We notice first that the Ramachandran maps of all the isomeric forms, calculated from the simulations, present strong variations as a function of pH, the most striking feature being the presence of distinct regions of local minima (see Tables F-H of the Supplementary Materials). For an easier comparison with the NMR results in water, in Figure 3 we report the experimental dihedrals and the MD distribution obtained by the *trans-trans* simulation at acidic pH. The two maps show a nice agreement, in particular in highlighting that the Asn^3 and Pro^4 residues assume polyproline II dihedrals.

Besides unordered structures, type I β turn results to be the principal conformer at pH lower than tyrosine pKa (9,77 [12], see Table D of the Supplementary Materials). The turn conformation in the NPGY and Ac-PHN regions is adopted mainly when tyrosine is not deprotonated and the involved proline is in *trans* conformation, as in the *trans-trans* isomer (Figures 1b, 1c, 4a, 4b) and in lower amounts, in the *cis-trans* isomer at neutral and acidic pH conditions. The occurrence of such turn conformations decreases as pH increases, thus shifting the equilibrium toward less ordered structures. The conformational change with pH is likely driven by the interaction between the phenolate oxygen of tyrosine (available only at basic pH) and the side chain amide hydrogen of asparagine. This interaction determines a tilt of the glycine residue, breaking the turn structure (Figure 4 a,b) in favour of unordered ones (Figure 4c). The conformational change therefore can be also tracked following the variations of glycine (ϕ,ψ) distribution angles minima from -97,1 in the H^+Y and in the HY form, to 137,-15 in the HY^- form (see Tables F-H of the Supplementary Materials), and more evidently from the distribution of $C_{\alpha i}$-$C_{\alpha i+3}$ distances in the NPGY region (Figure 5). This result agrees with the CD spectra blue shift for the minimum around 205 nm at pH 10 and 298 K previously reported [12], and explains the driving force of the phenomenon.

The presence of the NPGY turn at acidic pH, visible also in the NMR minimized structures (both in vacuum and in water, Figure 1b), is compatible with the ROESY correlation between the NH of Gly^5 and the NH of Tyr^6. As a final test of the agreement between the MD and the NMR structures at acidic pH we report in Table 7 the average backbone RMSD values of the 20 NMR structures

divided in two groups (disordered and turn as in Figure 1 [a] and [b]) with respect to the MD configurations sampled each 2 ns. Group [b] show the lowest RMSDs, confirming again the presence of the turn conformation in the NPGY sequence.

Conclusions

In this study we have applied NMR and MD simulations to the investigation of the conformational behaviour of avian prion hexarepeat Ac-PHNPGY-NH_2. NMR experiments indicated that the *trans-trans* isomer is the predominant species in H_2O/D_2O at pH 4.2 (83%), followed presumably by the *trans-cis* one (17%), while in the less polar TFE/H_2O environment the relative ratio is further increased to 95 to 5. From the analysis of the best 20 NMR structures minimized in vacuum and water, an averaged conformation with characteristic polyproline II dihedrals emerged for the Asn^3 and Pro^4 residues. The presence of such dihedrals inside a small region of the peptide is not related to a more extended sequence, and the local conformation does not propagate along the backbone, which is found instead to have more frequently either an unordered or a turn-like structure in the NPGY region (Figure 1, a and b). MD simulations of the *trans-trans* isomer in water succeeded in explaining this behaviour on the basis of the hydrogen bond distribution, hard to detect from NMR investigations of peptides. In particular, we found an i-i+3 turn structure in the NPGY region at acidic and physiological pH with Asn^3 and Pro^4 angles typical of polyproline II. Moreover, MD calculations were performed for all four isomers in water at acidic, neutral and basic pH conditions, showing in all the cases a fast interconversion among the accessible conformations. An interesting dependence of the peptide shape on the tyrosine residue deprotonation was also detected (Figure 4, 5), mainly for the *trans-trans* isomer. The deprotonation causes a shortening of the distance between the phenolate oxygen of tyrosine and the amide side chain hydrogens of asparagine thus tilting the glycine residue. This appears to be the driving force for the increase of unordered structures at basic pH explaining the relative blue shift in the CD spectrum [12].

The presence of turns inside the hexarepeat fragment has also been suggested by secondary structure prediction algorithms [33], and it is believed to play a crucial role in the endocytosis

processes, as a turn conformation is adopted by the NPXY internalization signal of the LDL receptor [34], [35]. The present combined NMR and MD study provides a further evidence of turn conformations in hexarepeats and represents a first step in understanding the structural features of the N-terminal region of the avian prion protein.

Acknowledgements. We thank MIUR (PRIN 2004032851, 2005035119, FIRB RBNEO3PX83_001 and grant no. 196 D.M. 1105/2002) for financial support.

References

1 . S. B. Prusiner, Science 278 (1997) 245.

2. S. B. Prusiner, M. R. Scott, S. J. DeArmond, F. E. Cohen, Cell 93 (1998) 337.

3. F. Wopfner, G. Weidenhofer, R. Schneider, A. von Brunn, S. Gilch, T. F. Schwarz, T. Werner, H. M. Schatzl, J. Mol. Biol. 289 (1999) 1163.

4. D. A. Harris, D. L. Falls, F. A. Johnson, G. D. Fischbach, Proc. Natl. Acad. Sci. USA 88 (1991) 7664.

5. J. M. Gabriel, B. Oesch, H. Kretzschamer, M. Scott, S. B. Prusiner, Proc. Natl. Acad. Sci. USA 89 (1992) 9097.

6. E. M. Marcotte, D. Eisenberg, Biochemistry 38 (1999) 667.

7. L. Calzolai, D. A. Lysek, D. R. Perez, P. Guntert, K. Wüthrich, Proc. Natl. Acad. Sci. USA 102 (2005) 651.

8. A. Radzicka, S. A. Acheson, R. Wolfenden, Bioorg. Chem. 20 (1992) 382.

9. J. S. Jhon, Y. K. Kang, J. Phys. Chem. A 103 (1999) 5436.

10. C. Benzi, R. Improta, G. Scalmani, V. Barone, J. Comput. Chem. 23 (2002) 341.

11. D. A. Lysek, K. Wüthrich, Biochemistry 43 (2004) 10393.

12. D. La Mendola, R. P. Bonomo, G. Impellizzeri, G. Maccarrone, G. Pappalardo, E. Rizzarelli, A. Pietropaolo, V. Zito, J. Biol. Inorg. Chem. 10 (2005) 463.

13. C. Bartels, T. Xia, M. Billeter, K. Wüthrich, J. Biomol. NMR 5 (1995) 1.

14. J. Cavanagh, W. J. Fairbrother, A. G. Palmer III , N. J. Skelton, Protein NMR Spectroscopy, Principles and Practice, Academic Press, San Diego 1996.

15. T.L. Hwang, A. J. Shaka, J. Magn. Reson. A 112 (1995) 275.

16. T. Herman, P. Guntert, K. Wüthrich, J. Mol Biol. 319 (2002) 209.

17. W. F. Van Gunsteren, S. R. Billeter, A. A. Eising, P. H Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, I. G. Tironi, Biomolecular simulation: the GROMOS96 manual and user guide VdF: Hochschulverlag AG an der ETH Zurich and BIOMOS Zurich, Gronigen 1996.

18. N. Guex, M.C. Peitsch, Electrophoresis 18 (1997) 2714.

19. R. Koradi, M. Billeter, K. Wüthrich, J. Mol. Graph. 14 (1996) 51.

20. P. Procacci, E. Paci, T. A. Darden, M. Marchi, J. Comput. Chem. 18 (1997) 1848.

21. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, J. Am. Chem. Soc. 117 (1995) 5179.

22. D. Frenkel, B. Smit, Understanding Molecular Simulation, 2nd edition. Academic Press, Orlando, 2001.

23. H. J. C. Berendsen, J. P. M Postma, W. F. van Gunsteren, J. Hermans, in B. Pullman, editor, Intermolecular Forces, Reidel, Dordrecht, 1981, p. 331.

24. C. Dugave, L. Demange, Chem. Rev. 103 (2003) 2475.

25. A. A Adzhubei, M. J. E. Stemnberg, J. Mol. Biol. 229 (1993) 472.

26. W. Kabsch, C. Sander, Biopolymers 22 (1983) 2577.

27. E. Strandberg and A.S. Ulrich, Concepts in Magn. Reson. A, 23A (2004) 89

28. G. Wagner, W. Braun, F. T. Havel, T. Schaumann, N. Go, K. Wüthrich, J. Mol. Biol. 196 (1987) 611.

29. O. Jardetzky, G. C. K. Roberts, NMR in Molecular Biology, Academic Press, New York, 1981.

30. G. Hutchinson, J. M. Thornton, Protein Science 3 (1994) 2207.

31. Y. Che, G. R. Marshall, Biopolymers 81(2006) 392.

32. G. D. Rose, L. M. Gierasch, J. A. Smith, Adv. Protein Chem. 37(1985) 1.

33. J. F. Bazan, R. J. Fletterick, M. P. McKinley, S. B. Prusiner, Protein Eng. 1 (1987) 125.

34. A. Bansal, L. M. Gierasch, Cell 67 (1991) 1195.

35. J. P. Paccaud, W. Reith, B. Johansson, K. E. Magnusson, B. Mach, J. L. Carpenter, J. Biol. Chem. 268 (1993) 23191.

Table 1. Proton chemical shifts (ppm) for the major conformer family a) and the minor conformer family b) of 1-HexaPY at 300 K in H_2O/D_2O (90/10, v/v). The values for the Acetyl and $CONH_2$ groups are 2.10 ppm and 7.55/7.05 ppm a), 2.10 ppm and 7.55/7.05 ppm b).

a)

AA	NH	αCH	βCH	γCH	Others
Pro[1]	/	4.31	2.20/1.91	1.79	δCH_2 3.61
His[2]	8.58	4.71	3.24/3.13	/	H(2) 8.59 H(4) 7.27
Asn[3]	8.47	4.95	2.82/2.66	/	γNH_2 7.60/6.93
Pro[4]	/	4.41	2.25/2.02	1.95	δCH_2 3.85/3.73
Gly[5]	8.41	3.91/3.84	/	/	/
Tyr[6]	7.91	4.52	3.05/2.95	/	H(2,6) 7.13 H(3,5) 6.84

b)

AA	NH	αCH	βCH	γCH	Others
Pro	/	4.47	2.34/1.92	1.72	δCH_2 3.45
His[2]	8.76	4.70	3.20/3.14	/	H(2) 8.61 H(4) 7.29
Asn[3]	8.63	4.77	2.82/2.65	/	γNH_2 7.60/6.95
Pro[4]	/	4.20	2.15/1.85	1.65	δCH_2 3.55
Gly[5]	8.49	3.90	/	/	/
Tyr[6]	8.00	4.47	3.08/2.95	/	H(2,6) 7.13 H(3,5) 6.80

Table 2. Proton chemical shifts (ppm) for 1-HexaPY at 300 K in TFE/H_2O (80/20, v/v). The values for the Acetyl and $CONH_2$ groups are 2.07 ppm and 7.14/6.54 ppm, respectively.

1

AA	NH	αCH	βCH	γCH	Others
Pro [1]	/	4.30	2.18/1.88	1.96	δCH_2 3.60/3.52
His[2]	7.93	4.68	3.29/3.13	/	H(2) 8.43 H(4) 7.21
Asn[3]	8.16	4.92	2.81/2.68	/	γNH_2 7.30/6.48
Pro [4]	/	4.41	2.25/1.95	2.00	δCH_γ 3.79/3.73

Table 3. Temperature coefficients (ppb/K) and ^3J coupling constants (Hz) for 1-HexaPY in TFE/H_2O (80/20, v/v) and in H_2O/D_2O (90/10, v/v). In H_2O/D_2O data are reported only for the major conformer family.

	TFE/H_2O (80/20)				H_2O/D_2O (90/10)			
AA	Δδ/Δt	$^3J_{NH\alpha C}$ H	$^3J_{\alpha CH\beta C}$ H	$^3J_{\alpha CH\beta' C}$ H	Δδ/Δt	$^3J_{NH\alpha C}$ H	$^3J_{\alpha CH\beta C}$ H	$^3J_{\alpha CH\beta' C}$ H
Pro	-	-	5.3	8.5	-	-	5.1	8.5
His	- 4.9	7.8	5.5	7.5	- 5.6	7.8	6.1	8.6
Asn	- 6.7	6.8	7.4	6.3	- 6.5	7.0	6.3	8.1
Asn $NH_2\gamma$	-7.0/-6.2				-4.6/-4.6			
Pro	-	-	5.1	8.5	-	-	5.5	8.5

Table 4. Average dihedral angles obtained from the 20 NMR structures minimized in *vacuo* for 1-HexaPY in H_2O/D_2O (90/10, v/v) a) and in TFE/H_2O (80/20, v/v) (b).

a)

AA		ϕ			ψ	
Pro[1]	-75	±	1	80	±	43
His[2]	-92	±	78	98	±	17
Asn[3]	-73	±	7	142	±	4
Pro[4]	-73	±	5	124	±	30
Gly[5]	7	±	88	0	±	78
Tyr[6]	-93	±	26	94	±	50

b)

AA		ϕ			ψ	
Pro[1]	-75	±	1	-177	±	1
His[2]	-54	±	1	173	±	1
Asn[3]	-52	±	1	125	±	1
Pro[4]	-75	±	1	126	±	30
Gly[5]	-79	±	87	-1.5	±	15
Tyr[6]	-100	±	28	65	±	67

1

Table 5. Average dihedral angles obtained from the 20 NMR structures minimized in solvent for 1-HexaPY in water a) and in TFE/H20 80/20 b)

a)

AA		ϕ			ψ	
Pro[1]	-71	±	10	103	±	60
His[2]	-90	±	45	89	±	28
Asn[3]	-65	±	21	142	±	9
Pro[4]	-68	±	7	138	±	45
Gly[5]	-173	±	70	1	±	56
Tyr[6]	-111	±	40	67	±	64

b)

AA		ϕ			ψ	
Pro[1]	-69	±	7	155	±	42
His[2]	-87	±	28	148	±	33
Asn[3]	-39	±	51	117	±	18
Pro[4]	-72	±	8	135	±	40
Gly[5]	-62	±	52	20	±	30
Tyr[6]	-124	±	33	110	±	66

Table 6: pH dependent MD hydrogen bond percentages for 1-HexaPY isomers found between i and i+3 residues.

	H^+Y			
H-bond $_{i-i+3}$	**trans-trans**	**cis-trans**	**trans-cis**	**cis-cis**
$ACE_{C=O}$- ASN_{N-H}	76%	0%	57%	0%
$ASN_{C=O}$-TYR_{N-H}	68%	21%	0%	0%
	HY			
$ACE_{C=O}$-ASN_{N-H}	84%	0%	38%	0%

Table 7: Average backbone RMSD values (Å) of the two clusters among the 20 NMR structures, [a] and [b] as in Figure 1, from the MD configurations sampled every 2 ns. Cluster [b] shows the lowest RMSDs, singling out the C-terminal turn conformation. The accuracy of NMR and MD ensembles is 1.17 Å and 0.59 respectively. The average RMSD between NMR structures and the mean of MD configurations is 1.87 Å.

Structures [a]	*Structures [b]*
2.72	1.37
2.46	1.11
2.24	1.08
2.10	1.26
2.10	1.15
2.34	1.88
2.54	0.88
2.11	1.58
2.41	1.17
-	1.17
-	0.89

Figure captions

Figure 1: From left to right: 1-HexaPY *trans-trans* conformers obtained from CYANA structures minimized in a water shell at acidic pH: disordered [a] and turn-like [b]. 1-HexaPY *trans-trans* [c] isomer obtained from MD simulations in the H^+Y form sampled every 2.5 ns. 1-HexaPY *trans-trans* MD isomer [c] shows the turn conformer, underlined also in the NMR structures [b] (top: N-terminus, bottom: C-terminus).

Figure 2: Time evolution of 1-HexaPY conformers for each of the four isomers (trans-trans [a], cis-trans [b], trans-cis [c], cis-cis [d]) in the three different protonation states obtained from MD simulations. The calculation is based on at least two consecutive residues with the characteristic dihedral angles concerning type I β turn and three consecutive residues for PPII.

Figure 3: Superposition of the backbone dihedrals of the 20 NMR structures minimized in water and the MD (ϕ, ψ) distributions of the 1-HexaPY obtained by the *trans-trans* simulations at acidic pH.

Figure 4: 1-HexaPY *trans-trans* typical conformations in the H^+Y [a], HY [b] and HY^- [c] form obtained from MD simulations. It is worth to note the short distance between the phenolate oxygen of tyrosine and the amide side chain hydrogen of asparagine residue (blue line) only in the HY^- form, where tyrosine is deprotonated. In the H^+Y and HY forms, instead, a turn structure is stabilized by an intra-chain $ASN_{C=O}$-TYR_{NH} H-bond (yellow line). The orange line shows the distance between the NH of Gly^5 and the NH of Tyr^6, corresponding to a calculated distance of about 3 Å.

Figure 5: $C_{\alpha i}$-$C_{\alpha i+3}$ distance distribution concerning the NPGY region of 1-HexaPY *trans-trans* in the three different protonation states. It can be noticed the short distance in the HY^+ and HY forms and the larger distance distribution in the HY^- form.

Figure 1

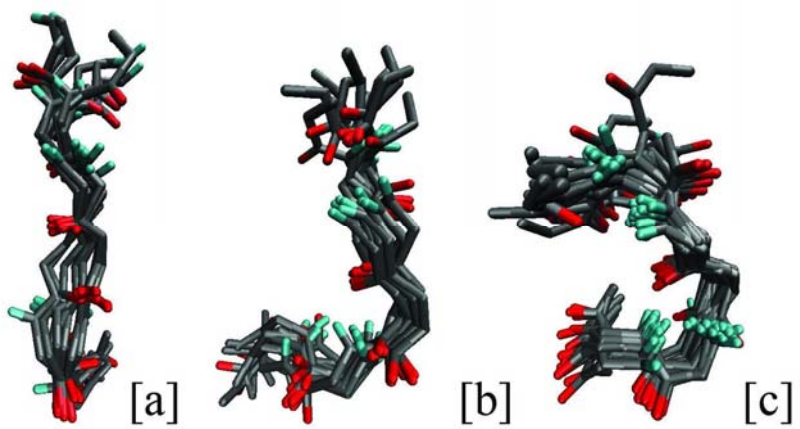

Figure 2

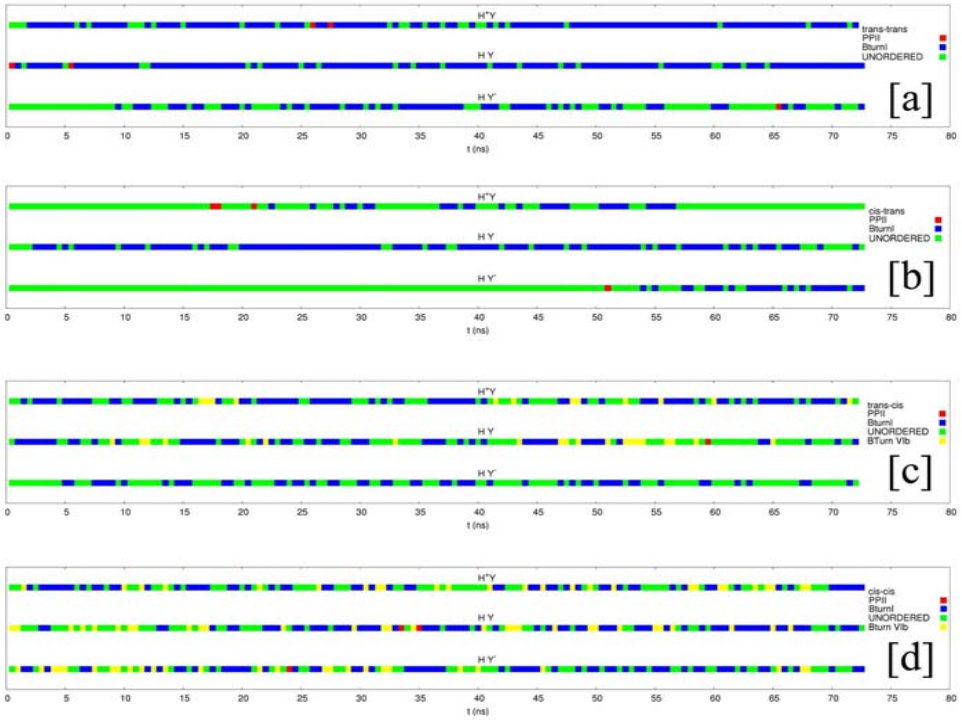

Figure 3

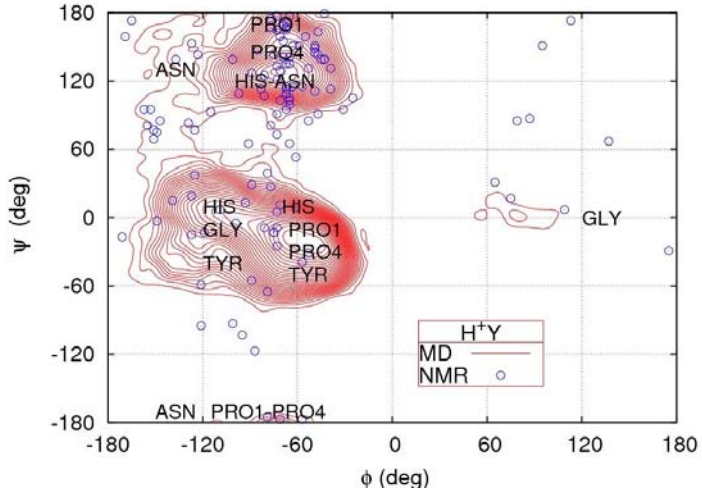

Figure 4

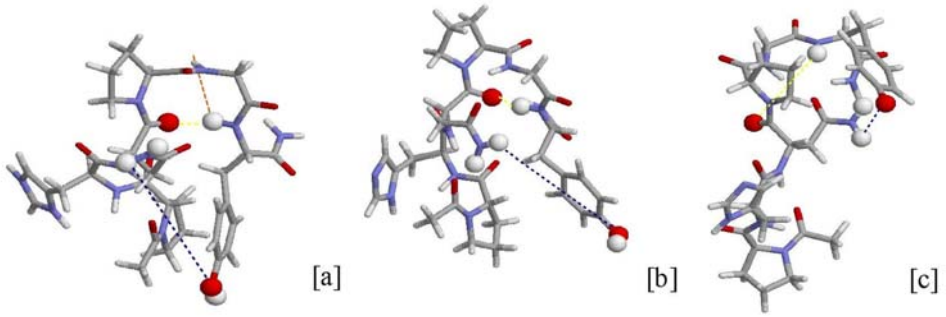

Figure 5