Prion proteins: evolution and preservation of secondary structure

Igor B. Kuznetsov*, Pavel S. Morozov, Yuri G. Matushkin

Laboratory of Molecular Evolution, Institute of Cytology and Genetics, Prospekt Lavrentieva 10, Novosibirsk 630090, Russia

Received 9 June 1997

Abstract Prions cause a variety of neurodegenerative disorders that seem to result from a conformational change in the prion protein (PrP). Thirty-two PrP sequences have been subjected to phylogenetic analysis followed by reconstruction of the most probable evolutionary spectrum of amino acid replacements. The replacement rates suggest that the protein does not seem to be very conservative, but in the course of evolution amino acids have only been substituted within the elements of the secondary structure by those with very similar physico-chemical properties. Analysis of the full spectrum of single-step amino acid substitutions in human PrP using secondary structure prediction algorithms shows an over-representation of substitutions that tend to destabilize α-helices.

Key words: Prion protein; Evolution; Secondary structure prediction; Phylogenetic analysis

1. Introduction

The recent epidemic of bovine spongiform encephalopathy (BSE) has attracted close attention to the prion proteins (PrP) [1]. It is suggested that the prions represent a novel class of infectious pathogens devoid of nucleic acids and cause a variety of sporadic, inherited and infectious neurodegenerative diseases: BSE, scrapie, kuru, Creutzfeld-Jakob disease (CJD) and some others [1,2]. According to available data, PrP exists in two isoforms, as the normal cellular protein (PrP^C) and the abnormal disease-related protein (PrP^{Sc}). The PrP^C secondary structure contains ≈43% of α-helix and ≈3% of β-sheet, whereas the β-sheet content in PrP^{Sc} is 30% [3,4].

On the basis of some structural findings, the following hypothesis for prion infectivity has been suggested. Exogenous prion particles containing PrP^{Sc} penetrate into the cells and act as templates for promoting the conversion of PrP^C into PrP^{Sc}. The insolubility of PrP^{Sc} makes the process irreversible, promotes the accumulation of large amounts of PrP^{Sc} and leads to cell death. Correspondingly, in the case of an inherited prion disease, PrP mutations destabilize the conformation of PrP^C and promote its conversion into PrP^{Sc} [2].

Four putative α-helices have been identified by computerized analysis of PrP sequences by secondary stricture prediction algorithms optimised for the study of globular proteins: amino acids 109–122, 129–141, 178–191 and 202–218 [5]. Most of the known point mutations associated with inherited prion diseases are clustered within or nearby these putative α-helices.

The mouse PrP domain comprising amino acid residues 121–231 has been analysed using NMR technique. The NMR-deciphered structure contains three α-helices (residues 144–154, 179–193 and 200–217) and an antiparallel β-sheet (residues 128–131 and 161–164) [6].

Due to its intriguing nature, the prion protein is a very interesting object for research, the evolution of the secondary structure elements being of special interest.

2. Materials and methods

Thirty-two sequences of PrP genes have been retrieved from the EMBL data bank. Multiple sequence alignment and phylogenetic tree reconstruction by the Neighbor-Joining (NJ) method [7] have been performed using programs from the phylogenetic analysis software package VOSTORG [8]. The average rates of nucleotide replacements have been calculated by Nei's method [9] implemented in the software package MEGA [10].

Using the software package AMS (P. Morozov, unpublished), we have reconstructed the most probable evolutionary spectrum of amino acid replacements that have occurred in all branches of the NJ-tree for the PrP sequences. Then we have built a distribution for the spectrum demonstrating how the frequency of amino acid substitutions depends on the physico-chemical distance between the original and resulting amino acids.

In order to reveal potential sites which can be of special importance for the preservation of the secondary structure, we have analyzed the influence of single-step amino acid replacements in human PrP sequence on the secondary structure predictions made with the PREDATOR program [11]. For each jth codon we have performed all possible single-step substitutions and then calculated the percentage of positions d_j predicted to be in structural state k (α-helix, β-sheet or random coil):

$$d_j = \frac{\sum_{i=1}^{m} p_{ik}}{p_{0k} * m} \quad (1)$$

P_{0k} is the number of amino acid residues in structural state k in normal human PrP; P_{ik} is the number of amino acids residues in structural state k in normal human PrP sequence with the ith amino acid substitution in the jth codon; m is the total number of non-synonymous single-step substitutions in the jth codon.

As the NMR structure has only been determined for the PrP domain comprising residues 121–231, and secondary structure prediction algorithms unambiguously predict α-helical conformation for residues 109–122, we have used a combined marked pattern of sites forming secondary structure elements. The pattern was as follows: residues 109–122 (H1) – first putative α-helix; 144–154 (H2), 179–193 (H3), 200–217 (H4) – α-helices found by NMR; 128–131 and 161–164 (S1) – β-sheet found by NMR.

3. Results

The average rates of synonymous/non-synonymous nucleotide substitutions calculated for the time of divergence of the main groups of mammals [12] are $4.63\times10^{-9}/0.60\times10^{-9}$ substitutions per site per annum. These values are close to the replacement rates calculated for the α-globin family, $3.94\times10^{-9}/0.56\times10^{-9}$ [13].

The distribution of the substitutions, selected in the course of evolution in PrP genes, with respect to the physico-chemical distance for any revealed pair of original and resulting

*Corresponding author. Fax: (7) (3832) 35-65-58.
E-mail: kuznets@bionet.nsc.ru

PII S0014-5793(97)00810-7

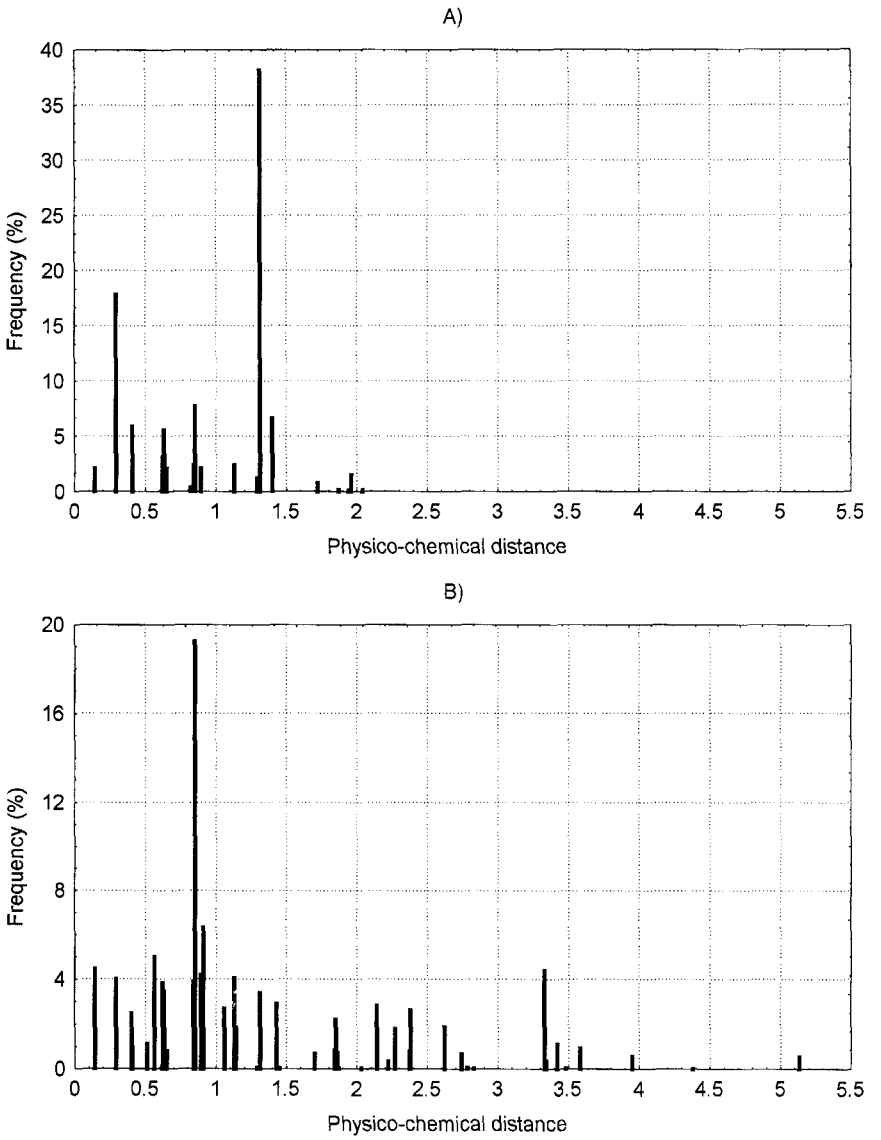

Fig. 1. Distribution of physico-chemical distances between the original and resulting amino acids for the reconstructed evolutionary spectrum of amino acid substitutions. (A) Data for sites belonging to the secondary structure elements H1–H4 and S1. (B) Data for the rest of the sequences.

amino acids is shown in Fig. 1. The matrix of physico-chemical distances was retrieved from previous data [14]. The distribution for the sites belonging to the secondary structure elements H1–H4 and S1 is shown in Fig. 1A, that for the sites not included in these elements is in Fig. 1B. One can see that the distribution shown in Fig. 1A has an over-representation of conservative amino acid replacements with a small physico-chemical difference (less than 1.5), whereas that in Fig. 1B has two peaks, one near the distance 1.0 and the other near 3.5. According to Miyata et al. [14], amino acid replacements with a distance $D \leq 1.0$ may not cause any significant change of tertiary structure over almost all variable sites in the protein over a short time interval of evolution. Thus, in the course of evolution the secondary structure elements H1–H4 and S1 had probably been undergoing strong selective pressure against amino acid substitutions with large physico-chemical differences which could distort the conformation of the polypeptide chain.

Fig. 2 shows a plot of the d_j values calculated using Eq. 1 for each amino acid position in the human PrP sequence. As can be clearly seen, the amino acid replacements in the putative helix H1 significantly decrease the number of residues predicted to be in α-helices. The synthetic peptide corresponding to the H1 helix has β-sheet conformation in aqueous solutions and can induce conformational transition of the synthetic peptide corresponding to residues 129–141 from α-helix to β-sheet [15].

If we analyse a histogram demonstrating how single-step amino acid substitutions change the total number of predicted α-helical positions and positions in β-sheets (data not shown), one could see the substitutions tending to decrease the α-helix content and increase the β-sheet content.

4. Discussion

Although there is nothing special about the rates and ratio of synonymous and non-synonymous nucleotide substitutions calculated for the total PrP (the protein does not seem to be

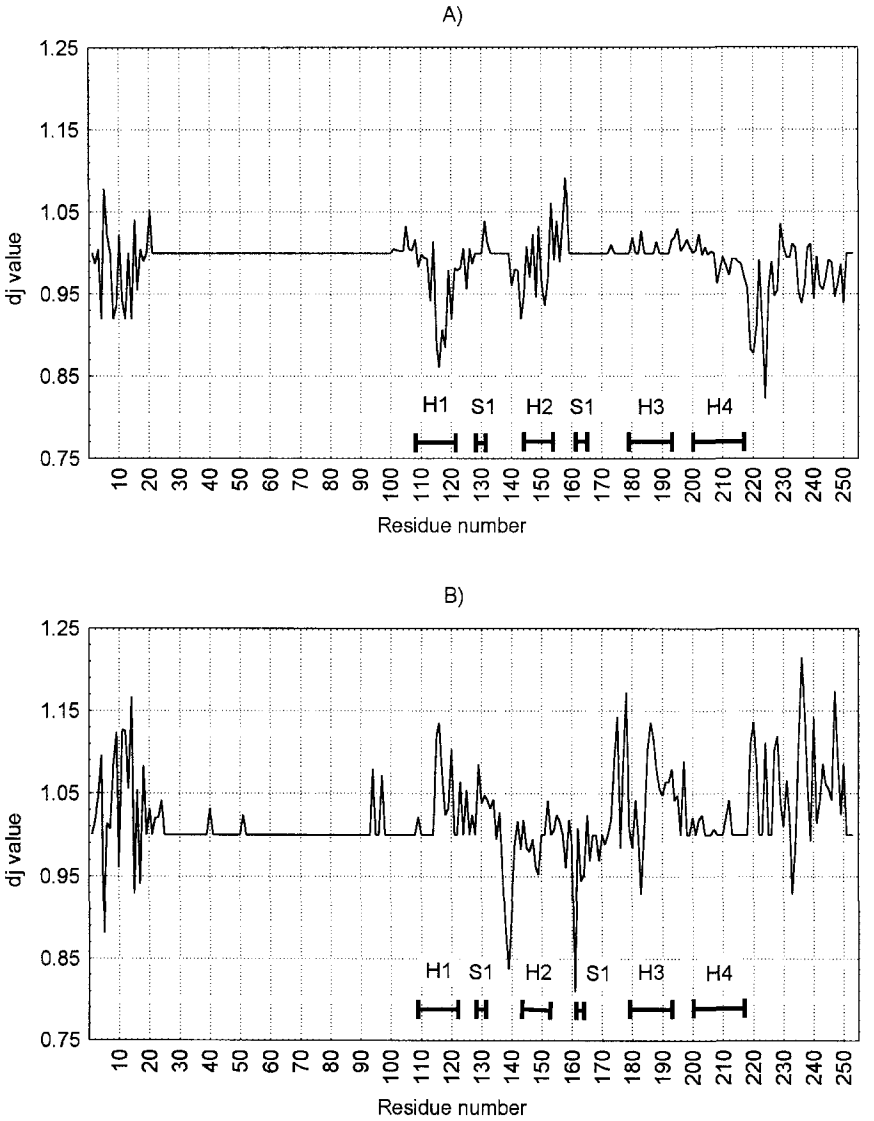

Fig. 2. Plot of the d_j values for human PrP. Note that residues 1–22 and 232–254 have been removed during biosynthesis. Data for random coil are not shown. (A) Data for α-helices. (B) Data for β-sheets.

as very conservative as, for example, the histones), most amino acid replacements have occurred in the course of evolution beyond the secondary structure elements. Furthermore, amino acids in these sites have been substituted by those having very similar physico-chemical properties. Thus it seems likely that radical amino acid replacements in the secondary structure elements are not permitted. As the exact function of PrP is not yet determined and the protein does not seem to be crucially important for survival [1], it is likely that such constraints, rather than preservation of function, prevent amino acid replacements capable of shifting conformational equilibrium toward accumulation of PrP^{Sc}.

The fact that amino acid substitutions in the secondary structure elements of PrP tend to decrease the number of predicted α-helical positions and to increase the number of positions in β-sheets may additionally favor this hypothesis. In other words, the context structure of the PrP sequence allows the protein to exist in two alternative conformations, amino acid replacements increasing the probability of conformational transition $PrP^{C} \rightarrow PrP^{Sc}$. The results on H1 site are of particular interest as the replacements there dramatically increase the number of predicted β-structures. Because the synthetic peptide corresponding to this site can convert 129–141 peptide into β-sheet conformation, it is tempting to speculate that H1 region can play a crucial role in the conversion of PrP^{C} into PrP^{Sc}.

The data obtained support the hypothesis that the possible mechanism for triggering $PrP^{C} \rightarrow PrP^{Sc}$ transition may include conversion of α-helices to β-sheets. It should be noted that the tendency to an over-representation of amino acid replacements decreasing the 'α-potential' is observed despite the relatively low sensitivity of secondary structure prediction algorithms. Therefore, it would be reasonable to anticipate that the 'in vivo' percentage of substitutions destabilising the conformation of PrP^{C} could be considerably higher.

Acknowledgements: This work was supported by the Russian State Programs Human Genome and Frontiers in Genetics.

References

[1] Prusiner, S.B. (1991) Science 252, 1515–1522.
[2] Cohen, F.E., Pan, K.-M., Huang, Z., Baldwin, M., Fletterick, R.J. and Prusiner, S.B. (1994) Science 264, 530–531.
[3] Safar, J., Roller, P.P., Gajdusek, D.C. and Gibbs Jr., C.J. (1993) J. Biol. Chem. 268, 20276–20284.
[4] Pan, K.M., Baldwin, M., Nguyen, J.M., Serban, A., Groth, D., Mehlhorn, I., Huang, Z., Fletterick, R.J. and Cohen, F.E. et al. (1993) Proc. Natl. Acad. Sci. USA 90, 10962–10966.
[5] Huang, Z., Gabriel, J.M., Baldwin, M.A., Fletterick, R.J., Prusiner, S.B. and Cohen, F.E. (1994) Proc. Natl. Acad. Sci. USA 91, 7139–7143.
[6] Rick, R., Hornemann, S., Wider, G., Billeter, M., Glockshuber, R. and Wuthrich, K. (1996) Nature 382, 180–182.
[7] Saitou, N. and Nei, M. (1987) Mol. Biol. Evol. 4, 406–425.
[8] Zharkikh, A.A., Rzhetsky, A.Y., Morosov, P.S., Sitnikova, T.L. and Krushkal, J.S. (1991) Gene 101, 251–254.
[9] Nei, M. and Gojobori, T. (1986) Mol. Biol. Evol. 3, 418–426.
[10] Kumar, S., Tamura, K. and Nei, M. (1994) Comput. Appl. Biosci. 10, 189–191.
[11] Frishman, D. and Argos, P. (1996) Protein Engineering 9, 133–142.
[12] Goodman, M. (1982) Macromolecular Sequences in Systematic and Evolutionary Biology, Plenum Press, New York, London.
[13] Li, W.-H., Wu, C.-J. and Luo, C.-C. (1985) Mol. Biol. Evol. 2, 150–174.
[14] Miyata, T., Miyazawa, S. and Yasunaga, T. (1979) J. Mol. Evol. 12, 219–236.
[15] Nguyen, J., Baldwin, M.A., Cohen, F.E. and Prusiner, S.B. (1995) Biochemistry 34, 4186–77784192.