CMLS, Cell. Mol. Life Sci. 59 (2002) 1366–1376
1420-682X/02/081366-11

CMLS Cellular and Molecular Life Sciences

Research Article

Structural features of prions explored by sequence analysis I. Sequence data

J.-P. Mornon*, K. Prat, F. Dupuis and I. Callebaut

Systèmes Moléculaires & Biologie Structurale, LMCP, CNRS UMR 7590, Universités Paris 6 et Paris 7, case 115, 4 place Jussieu, 75252 Paris Cedex 05 (France), Fax +33 1 44 27 37 85, e-mail: mornon@lmcp.jussieu.fr

Received 24 June 2002; accepted 28 June 2002

Abstract. Animal prion proteins (PrPs) form at the sequence level a very homogenous and 'closed' family. Therefore, few of their structural and functional features can be gleaned from sequence comparison as is now possible on a wide scale for other protein families. To detect putatively related proteins (at the structural and/or functional level), we used a battery of sequence analysis tools. This analysis resulted in (i) the identification of a putative 'prion-like' domain within the envelope of foamy retroviruses, (ii) the detection of putative similarities between prions and an interferon-inducible membrane protein, and (iii) the proposal that of the TATA-box-binding protein is a structural scaffold, which might allow understanding of a key event leading to the structural conversion from PrP^{C} (normal cellular prion structure) towards PrP^{Sc} (pathogenic structure).

Key words. Prion; HCA; foamy virus; interferon-inducible membrane protein; TATA-box-binding protein; fusion peptide.

Introduction

Animal prion proteins (PrPs) [for a review see ref. 1] constitute at the level of their amino acid sequences a remarkably homogenous family, exhibiting only weak divergence (40% sequence identity between the most remote members known, human and chicken). Moreover, to our knowledge, prion sequences have not been reported to share obvious similarities with other protein families, as detectable by efficient classical lexical analysis (e.g. PSI-BLAST [2]), although some hypotheses have been proposed [see e.g. ref. 3]). Consequently, the apparent absence of potentially related proteins within the known genomes precludes useful data to further our understanding of prion proteins at the structural and functional levels. In this context, the human downstream prion-like Doppel protein constitutes an exception, sharing only 21% sequence identity with human prion along its globular domain. Prion proteins are indeed structured into two main domains: the first, non-globular, includes repeated sequences, whereas the second shares features of globular domains. Only this last globular domain is present in Doppel, lacking the N-terminal repeats of prion proteins. The normal cellular prion protein PrP^{C} is thought to accomplish a three-dimensional (3D) conversion to be transformed into a pathogenic PrP^{Sc} protein leading to fibril formation [4]. This conversion is accompanied by an increase in the β strand content and, conversely, a decrease in the α helix content [4], but the conformation of PrP^{Sc} is not yet known, although several hypotheses have been proposed [see e.g. refs 5, 6]. The monomeric 3D structures of the globular domain of several PrP^{C} proteins have been experimentally determined from nuclear magnetic resonance (NMR) studies [see e.g. refs 7–9]. More recently, a dimeric crystal form has also been investigated [10], showing a canonical 'swapping' co-operative effect between monomers and the formation of an additional β strand with respect to the previously known monomeric PrP^{C} structure. The prion-like Doppel protein 3D structure determined by NMR showed, as expected from se-

* Corresponding author.

quence data, a fold similar to the monomeric PrP^C [11]. However, no experimental structural data at the atomic level for isolated full PrP^{Sc} or PrP^{Sc} fibers have yet been reported, although the very recent electron crystallography study of two-dimensional truncated PrP^{Sc} crystals [6] gave new insights in this field. Moreover, recent structural analysis of a crystal form of the globular domain of the yeast Ure2 prion protein has added new data to our overall knowledge [12], although being a priori not correlated to the PrP^C fold.

We previously reported that the sequence of the PrP^C globular domain which, at first glance, appears as conventional as any other globular domain, possesses in its N terminus a peculiar peptide (ranging from amino acids ~110 to 135). The global amino acid composition of this peptide indeed shares similarities with those of fusion peptides from viral envelope glycoproteins and of reactive loops of serine protease inhibitors (serpins) [13]. These fusion peptides and serpin reactive loops are characterized by a high content of small amino acids (alanine, glycine, threonine), are intrinsically flexible [14–19], and are associated with large 3D conversions of their support proteins, leading to more stable structures that allow the achievement of their respective biological functions. Similar structural mobility is expected between the PrP^C and PrP^{Sc} forms. Another example of such dramatic flexibility associated with non-charged sequence segments rich in small amino acids is furnished by the structure of the transmembrane stem region of *Staphylococcus aureus* α-hemolysin [20] and LukF proteins [21].

According to these observations, the 106–126 PrP^C peptide has been shown to be flexible [22, 23] and neurotoxic [24]. As the path of conversion of PrP^C toward PrP^{Sc} necessarily implies a substantial conversion of its 3D structure, we hypothesized that the occurrence of a potentially flexible 'fusion-like' peptide in PrP^C could be directly associated with this conversion, as is the case for the particular peptides mentioned above.

A complementary approach in the analysis of the $PrP^C \rightarrow PrP^{Sc}$ conversion concerns the detection of sequence segments which may jump from α to β or coil conformations and, vice versa, from one 3D protein state to another activated one. Comparing predicted secondary structures (based on accurate multiple sequence alignments) and observed secondary structures [automatically attributed from Protein Data Bank (PDB) 3D coordinates], we detected 'discordant' peptides in a large bank of 3D structures. These are characterized by predicted structures different from those that are experimentally observed and are consequently good candidates for structural mobility. We identified, among others, the already known conformationally mobile segments of hemagglutinin [25, 26] and serpins [18, 19]. Three peptides within the prion protein were also identified: the above-described 110–135 'fusion-like' peptide, and peptides covering helix H2 and the very C-terminal end of helix H3. For the well-documented examples of proteins for which the two states of 3D structures are known, predictions from sequences match the final and more stable state, reached after structural conversion. Therefore, these observations suggest that from PrP^C to PrP^{Sc}, within the prion core, helix H2 may jump from an α-helical structure to two or three β strands, as already suggested by us after hydrophobic cluster analysis (HCA) [27]. This hypothesis was recently confirmed for this third hypothetical strand by the 3D crystal structure of a dimeric human prion, in which a short β strand substitutes an α-helical segment (end of helix H2) of the monomeric form [10].

Starting from these data and hypotheses, we propose here new insights into prion structures, mainly based on compilation of and human expertise with sequence data. We performed an in-depth analysis of the prion sequences to gain more information on their structural features, in particular those involved in the conversion into a pathogenic structure, and to suggest new lines and hypotheses for further investigations. We used various approaches and tools of sequence analysis, in particular those that are well adapted to the detection of remote relationships (e.g. HCA). In addition, this latter approach allows easy visualization of the protein sequence 'texture,' which can be defined as the global composition and organization of regular secondary structures participating in the core of protein structures [27]. The predicted secondary structures of part of the PrP^{Sc} form were also searched through the PDB, gathering all known experimental structures, to identify possible 3D structure templates which could be locally similar to the pathogenic PrP^{Sc} structure. This approach allowed the construction of a model of the PrP^{Sc} structure which may be in overall agreement with experimental data, and which will be reported in a forthcoming paper.

Materials and methods

Protein sequences were searched against databases using two programs: BLAST and PSI-BLAST [2], running at the National Center for Biological Information, USA (NCBI), and FASTA [28], running at the European Bioinformatics Institute, UK (EBI). Sequences were submitted to human expertise through the HCA approach [27, 29], which offers the possibility to add information about secondary structures to the lexical analysis of the considered sequences. The HCA-processed sequence is handled on a duplicated helical net (helix is in 3D a 1D object with a constant curvature and of constant properties along its path as the circle is in 2D) in which hydrophobic amino acids (V, I, L, F, M, Y, W) are contoured. The thus defined hydrophobic clusters [i.e. hydrophobic amino acids that are separated from each other by at least four non-hy-

drophobic residues (connectivity distance linked to the use of an α-helical support)] were shown to mainly correspond to the internal faces of regular secondary structures (α helices or β strands) [30]. The maintenance of a similar structure between divergent sequences is very often associated with conservation of hydrophobic cluster features that participate in the protein core, together with some sequence similarities. This conservation often helps and/or allows the alignment procedure for highly divergent sequences (typically in the 10–20% sequence identity range, below the so-called 'twilight zone' of 25–30% identity). The sensitivity of this approach, combined with profile-based lexical tools, has often been successfully used to identify new domains and/or to link orphan sequences to particular structural and functional families.

To assess the significance of HCA-deduced sequence alignments, we calculated a reliability index (RI), using the home-made Mscore software. This index measures the distance between the alignment of two actual sequences and the best alignments obtained from randomized versions of one of the two compared sequences. It is calculated on the basis of the consideration of Z-scores, which represent the differences between the observed scores and the mean score of a distribution of scores calculated from the alignment of one of the two compared sequences with 1000 randomized versions of the other. These Z-score values are expressed in standard deviation units of the random distribution. The RI is then defined by dividing the product of the three observed Z-scores [calculated for sequence identity, sequence similarity (using the Blosum 62 matrix] and hydrophobic match) by the product of the best Z-scores calculated from randomized versions of sequence 2. Therefore, an RI value greater than 1 indicates that the observed alignment is more likely than the best one observed when comparing sequence 1 with 1000 randomized versions of sequence 2.

Sequence motifs were searched using Prosite [31]. Atomic co-ordinates were taken from the PDB [32] and protein structures were manipulated using Swiss-PDBViewer [33, 34]. Secondary-structure motifs were searched in the PDB using the home-made MotifF program. Complementing the above tool panel, a new and as yet unpublished search method was tested, based on the processing of HCA cluster data, which are insensitive to the level of similarity between compared sequences. The thus detected sequences were then sorted manually.

Results and discussion

Envelope proteins of foamy viruses include a sequence sharing striking similarities with those of human prion and Doppel globular domains

A first search with the PSI-BLAST screening program using the human Doppel (1–176) sequence as query against the non-redundant database at NCBI (August 2000, 585,701 sequences) led to the observation, just after a long list of animal prion proteins, of marginal similarities with the envelope protein sequence of human foamy virus (also named spumaretrovirus) on a 130-amino-acid overlap (24% identity, 7% insertion-deletions, Expected-value 123). The analysis of the HCA 2D transpositions of both sequences (fig. 1A, lanes B and C) revealed good compatibility between several hydrophobic clusters all along the compared sequences, associated with patches of sequence identity, which constitute typical features of distant but related sequences [27]. The statistics calculated on the basis of the HCA-deduced alignment with human Doppel and prion proteins (fig. 1B) strengthened this putative relationship. Sequence identity between human prion (138–229 segment) and the related Doppel protein is 21% with an RI of 12.6. The RI value is 4.4 between human Doppel and foamy virus envelope sequences (17% sequence identity), and 2.2 between human prion and foamy virus envelope sequences (16% sequence identity). When alignments include the upstream 'fusion-like' peptide region (the considered sequence of human prion ranging from amino acids 98 to acid 229), these data are: prion/Doppel 20% identity, R.I 18.4; Doppel/envelope 17% identity, RI 14.1; prion/envelope 14%, RI 3.4. The similarities between Doppel and envelope signal peptides reinforce this hypothesis (fig. 1A).

Thus, a 'prion-like' domain might be included in the envelope glycoprotein of foamy viruses. However, this envelope protein does not possess, as does Doppel, repeats or a 'fusion-like' peptide. A central segment (boxed in green), around an invariant proline, is particularly well conserved between these three proteins (prion-Doppel-envelope), perhaps due to structural and/or functional constraints. The long C-terminal helix of human prion and human Doppel (helix H3) appears to be well conserved in the envelope protein (its corresponding hydrophobic cluster is very typical of helices). In the N-terminal part of the envelope protein and Doppel sequences, the signal peptide, as well as a downstream W-N-(K,R) peptide, exhibit good conservation.

As viruses may integrate foreign domains within their architecture to accomplish specific functions [see e.g. refs 35, 36], the identification of a sequence related to human Doppel and prion proteins in the envelope of a foamy retrovirus may not be, in itself, really surprising. However, our observation can shed new light on the unconventional behavior of these non-pathogenic retroviruses [37]. Indeed, among other features, foamy viruses do not encode a nucleocapsid protein, in contrast to other retroviruses and the majority of RNA viruses [38]. A speculative hypothesis could therefore be proposed that this particular feature may be linked to the recently observed capacity of prions to mimic the nucleocapsid [39, 40].

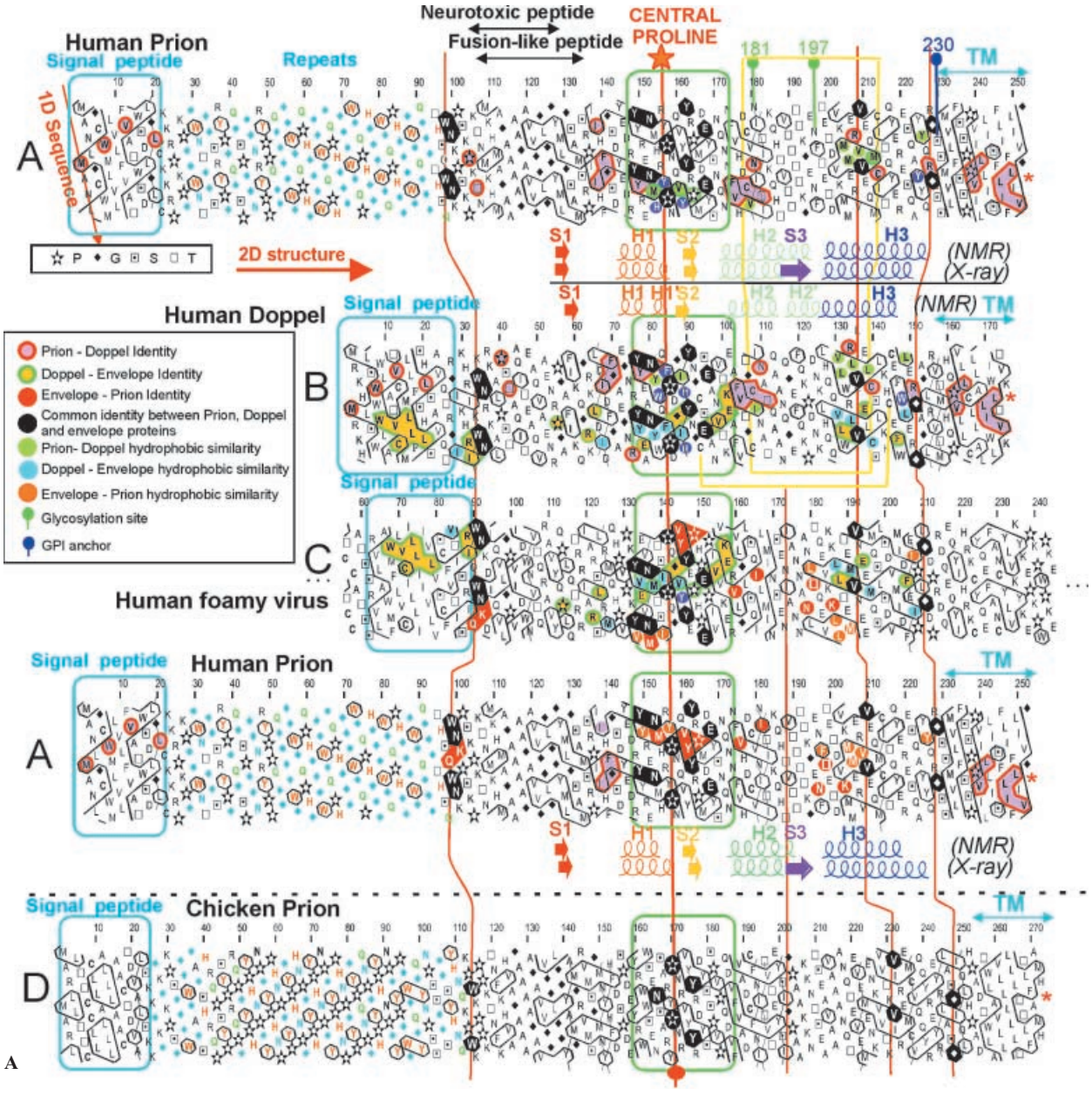
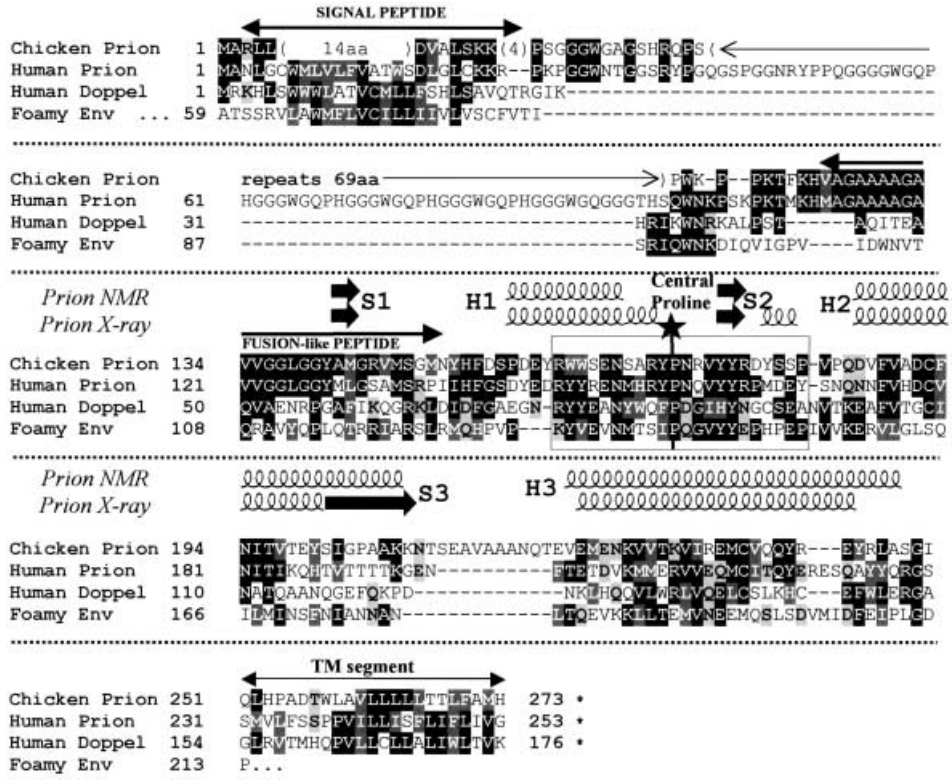


Figure 1. Comparison of the sequences of animal prions and envelope proteins from foamy viruses. (*A*) HCA alignment of the entire sequences (before post-translational modifications) of human prion (lanes A; SwissProt accession number P04156), human Doppel (lane B; TrEMBL accession number Q9NTM4), of part of human foamy virus (or spumaretrovirus) envelope (lane C; SwissProt accession number P14351), and of the entire sequence of chicken prion (lane D; SwissProt accession number P27177). The beginning of lane A indicates the way to read HCA plots as well as the correspondence for the four symbols used in these plots. Additional guidelines to interpret the plots are given in the Materials and methods section and in Callebaut et al. [27]. The pairwise sequence identities and similarities are shown with different colors, which are detailed in the box on the left. The glycosylation sites and the glycosyl-phosphatidylinositol (GPI) anchor of human prion are also indicated. Putative transmembrane-like segments (TM) are shown with arrows. Disulfide bridges are indicated with yellow lines connecting cysteine residues. Cluster correspondences are indicated with vertical lines. Proline 158, named 'central proline,' constitutes a striking anchor point for this comparison. Centered on the 'central proline,' the green box constitutes a particularly well-conserved segment between human prion, human Doppel and human foamy virus envelope, with nearly no insertion-deletion. Reported between lanes A and B are the experimentally observed regular secondary structures of human prion [monomeric (NMR; pdb identifier 1qlx) and dimeric (X-ray; pdb identifier 1i4m) forms] and mouse Doppel (pdb identifier 1i17), whose sequence has been aligned with its human ortholog. H stands for helix, S for strand. Within repeats, glycine (G, dark diamond) and asparagine (N), which otherwise mainly populate left-handed helical conformations, are colored light blue, whereas glutamine and aromatic amino acids are colored green and orange, respectively. Red stars indicate the C-terminal ends of the sequences. (*B*) Linear alignment of the sequences of prion, Doppel and foamy virus envelope proteins, as deduced from the HCA alignment shown in *A*. Secondary structures of human prion, from NMR (pdb identifier 1qlx) and X-ray (pdb identifier 1i4m) studies, are shown above the sequences (H stands for helix, S for strand). Sequence identities are reported white on a black background. Sequence similarities are shaded gray (white and black letters for hydrophobic and other amino acids, respectively).



B

Figure 1 (continued)

Perhaps the foamy virus envelope proteins use a 'prion-like' domain to 'mimic' a missing nucleocapsid.

We would like to re-emphasise the striking similarity between the amino acid composition of the prion 110–135 segment and those of viral fusion peptides and reactive-center loops of serpins [13, 27, 41]. This prion segment appears to meet the global amino acid composition of canonical fusion-like and mobile peptides [e.g. the human prion 119–137 segment **G**A**V**VGGLGG**Y**ML-GSA**M**SR**P** globally matches quite well the 356–373 reactive loop of human plasminogen activator inhibitor **G**T**V**ASSSTAVIVSAR**M**A**P**). As stated in the introduction, viral fusion peptides and 'fusion-like' peptides of serpins, rich in small amino acids (glycine, alanine, threonine), are intrinsically flexible structures probably fluctuating between α, β, and coil conformations, a dynamic behavior obviously linked to their respective functions. Both are also associated with a dramatic structural conversion of their supporting protein, as illustrated by the X-ray structures of influenza hemagglutinin protein envelope before and after activation [25, 26] and by those of native, latent, and activated serpins [18, 19]. The presence within prion sequences of such intrinsically mobile peptides may be, therefore, similarly associated with the overall structural conversion of these proteins, particularly from PrP^{C} to PrP^{Sc}, as suggested below. However, this hypothesis remains to be further characterized.

In this context, we note the reported possibility of a connection between endogenous retrovirus, prion diseases, and accelerated senescence [42].

An intriguing similarity was also detected using FASTA between the 'central proline' region of human prion (fig. 1 A) and the V3 loop of a human immunodeficiency virus (HIV) envelope glycoprotein (human prion: **R**EN**M**H**R**Y**P**N**Q**V**Y**Y**R**P**M**D** / HIV: **R**TS**I**H**I**G**P**G**Q**E-F**YRP**G**D**). Although the significance of such similarity might be questionable, of note is that, within the 3D structure of the HIV envelope surface unit [43], this peptide has not been seen in the electronic density due to its over-large flexibility. This putative similarity lies within the most conserved segment between prion, Doppel, and foamy virus envelope proteins (fig. 1 A) which noticeably included helix H1, otherwise corresponding to one of the most hydrophilic helices detected so far in the PDB [44]. However, this helix is strongly stabilized by a high density of $i \rightarrow i+4$ side chain interactions.

Prion and Doppel proteins share sequence and 2D 'texture' similarities with an interferon-inducible membrane protein

The above-reported similarity between the sequences of prion and foamy virus envelope proteins was detected within the statistically non-significant PSI-BLAST results, using HCA to sift all the potential candidates. How-

ever, no obvious similarity with other proteins was highlighted in this way, even when including foamy envelope sequences in profiles used to screen databases. We then used as a screening tool a new procedure, based on a search for similar hydrophobic clusters. The rationale of this approach is that regular secondary structures, and thus hydrophobic clusters that mainly correspond to their internal faces, are more conserved than sequences. Searching for cluster similarities, instead of only sequence similarities, should thus greatly enhance the sensitivity of such a procedure in detecting remote relationships.

Using as query a short segment of human Doppel, centered around the 'central proline' [from amino acids 70 to 100, the region which is more conserved than other sequences in prion, Doppel, and foamy virus (green boxes in fig. 1A)], our attention was especially drawn, within 20 candidates automatically selected by this procedure, to the sequence of an interferon-inducible 17-kDa membrane protein, implicated at the cell surface in relaying growth inhibitory and aggregation signals [45] (TrEMBL accession number Q9R175). Along this short segment, the interferon-inducible membrane protein shares only 3% sequence identity with the Doppel protein used as query, but has in common with Doppel and prion proteins a similar HCA texture spread out all along the sequence of their globular domains, which are of similar size (fig. 2). The resulting alignment leads to 15% sequence identity along ~100 amino acids (with Doppel) and 14% sequence identity along ~70 amino acids (with prion). We note that some membrane proteins that are significantly related to the interferon-inducible membrane protein (e.g. the human NG5 protein) possess upstream of their globular domain long stretches of pseudo-repeated sequences (data not shown). However, no 3D structure has yet been solved for this family of proteins, which could be used to assess the observed similarities at the 2D level.

Several studies have provided evidence for the involvement of prion as well as Doppel in signal transduction pathways [see e.g. refs 46–48]. In this context, the significance of the similarities reported here, if they are not fortuitous, remains to be established, and could open a new avenue to explore the normal function of PrP^{C}.

The TATA-box-binding protein fulfils some expected features of the PrP^{Sc} structure

The structural conversion from the cellular prion form PrP^{C} to the pathogenic PrP^{Sc} (scrapie) implies $\alpha \rightarrow \beta$ transitions [4], some of which may affect the globular domain (amino acids ~110–230) and possibly the N-terminal repeats. The pathogenic form remains, however, unknown, despite numerous investigations. Several putative models for PrP^{Sc} structure have been considered [e.g. refs 5, 6]. We have proposed [27] that helix H2 may be involved in an $\alpha \rightarrow \beta$ conversion. Indeed, mammalian prion proteins include within the segment ranging from amino acids 170 to 200 (human prion numbering) sequence stretches typical of β strands and uncommon in helices (fig. 1A, lane A): FVH, CVNITI, TVTTTT (which may correspond to a succession of β strands). These are, however, observed in the monomeric PrP^{C} structure as mainly corresponding to helix H2. The foamy virus envelope sequence (fig. 1A, lane C) also suggests the presence at this location of extended structures (IVV, VLGL, ILMI) as does the Doppel sequence (VT, FVT, CINAT). The last putative strand of this segment appears more elusive and only the human prion sequence shows a clear tendency for a β conformation (TVTTTT in human prion but QAANQG in Doppel and SIGPAA in chicken prion). Remarkably, precisely this sequence was observed in a β strand conformation in the human prion crystal structure [10]. In contrast, prion helix H3 sequence appears to fit the standard of α-helical structures in all prion, Doppel, and foamy virus envelope proteins. The shapes of helix H3-associated hydrophobic clusters are typical of amphiphilic helices. Moreover, PrP helix H3 appears to be particularly stable [49].

Therefore, as a working hypothesis, we postulated that the second half of the transformed globular domain of PrP^{Sc} may be structured as follows: an invariant β strand (S2 one), two or three β strands (replacing helix H2 and neighboring sequences), and a long C-terminal helix (invariant H3 helix). We searched in PDB for proteins possessing such a succession of regular secondary structures using a home-made tool (MotifF software) to identify folds which may fit part of the locally predicted PrP^{Sc} structure. Indeed, assuming the reality of this predicted succession of secondary structures, there is about one chance in two that this fold has already been described at the experimental level, because the number of distinct folds used by nature to build proteins is expected to be only ~10^3 [see e.g. refs 50–52] and about half of them (~600) are now present in PDB [53]. However, because from one structure to another within the same fold family, loops may sometimes include extra regular secondary structures, the succession of the sought-after secondary structures (e.g. $\beta\beta\beta\beta\alpha$) needed to be extended to a much larger ensemble [e.g. β,(α),$\beta\beta\beta\alpha$]. MotifF performs such a complex combinatory exploration, allowing only one extra regular secondary structure per set to avoid a dramatic increase of possibilities. To improve accuracy, MotifF distinguishes three sizes for each α or β structure (short, medium, long). The resulting screening procedure is consequently expected to be redundant but more efficient. It was applied to a pre-processed bank of 6485 protein structures, using as queries several secondary structure sequences [for example β-β-β-α, β-β-β-β-α, β-X-β-β-(β)-X-α, where X is an α or β structure and (β) means that this structure may be absent]. More than 3000 hits were recorded and then manually sorted using their HCA signature as a ridding tool. Among them, a TATA-box-

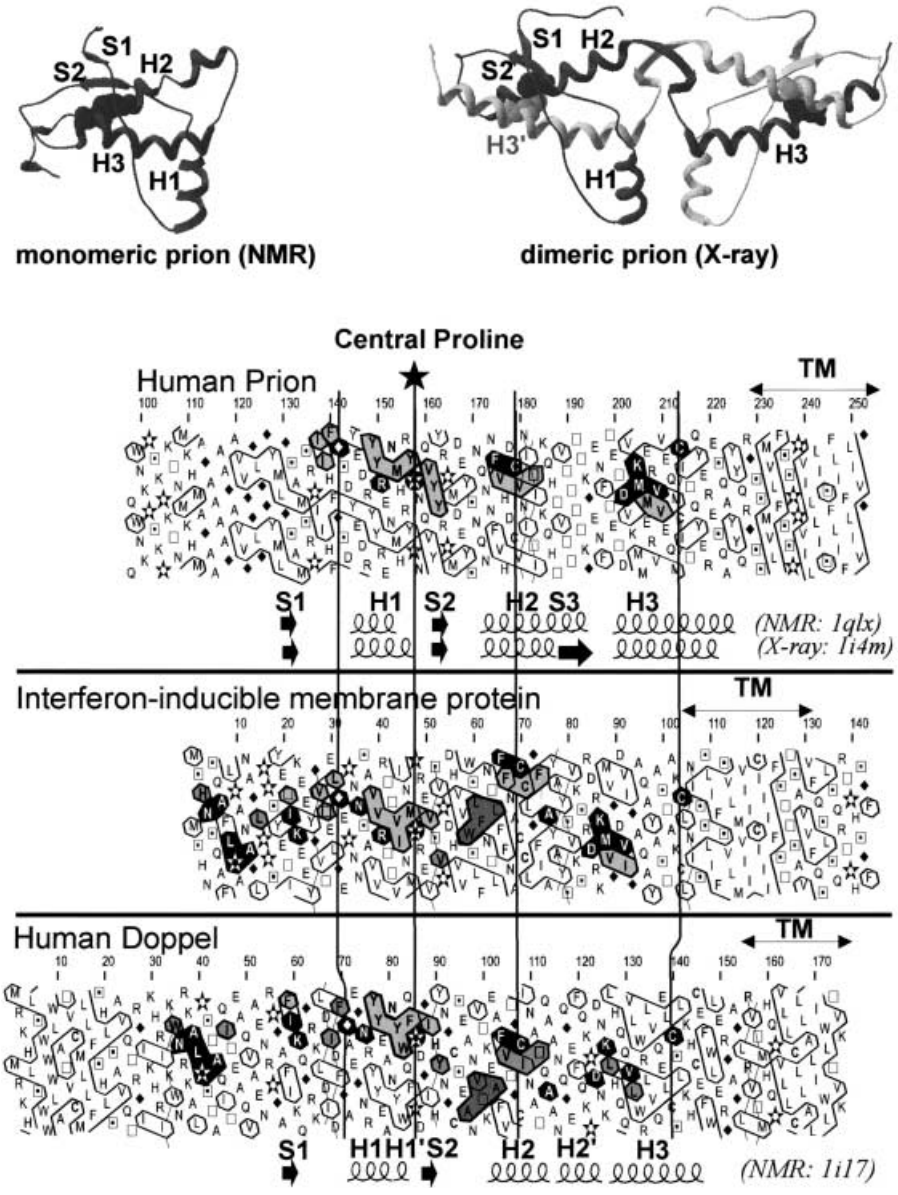


Figure 2. HCA comparison of the sequence of an interferon-inducible membrane protein with those of human prion and Doppel proteins. Accession numbers of human prion are P04156 (SwissProt), human Doppel Q9NTM4 (TrEMBL), and interferon-inducible membrane protein Q9R175 (TrEMBL). Putative transmembrane-like (TM) segments are shown with arrows. The secondary structures of human prion, as defined from NMR (pdb identifier 1qlx) and X-ray (pdb identifier 1i4m) are shown below its sequence (H stands for helix, S for strand). To recall structural data, ribbon representations of the 3D structures of monomeric and dimeric forms of human prion (pdb identifiers 1qlx and 1i4m, respectively) are shown above its sequence. Sequence identities are shown with white letters on a black background, main hydrophobic cluster similarities are shaded gray. Vertical lines indicate putative cluster correspondences. The presence of a cluster of alanine residues upstream of the transmembrane segment of the interferon-inducible membrane protein suggests that this region is helical and therefore might be the counterpart to the long helix H3.

binding protein (TBP, PDB identifier: 1cdw) meets, much better than the other candidates, the main features we expected for the PrP^{Sc} structure. Figure 3 A illustrates the HCA comparison between the sequences of human prion and this protein. The TATA-box-binding protein is built from three domains: an N-terminal non-globular domain, followed by two similar globular domains (domains 1 and 2). The N-terminal domain (not known at the 3D level) possesses a long stretch of glutamine (Q) repeats as well as, immediately before the first globular domain, several PXXP, MXXM and AXXA motifs (X being any amino acid), which are also present in human prion. This segment also exhibits five PXZPXZ repeats (X = M, I, A, Z = T, S) reminiscent of the numerous PHNPGY repeats of chicken prion (Fig. 1 A, lane D). The globular domains possess exactly the set of secondary structures we hypothesized for PrP^{Sc} after the putative conversion of the prion helix H2 region into three successive β strands (the

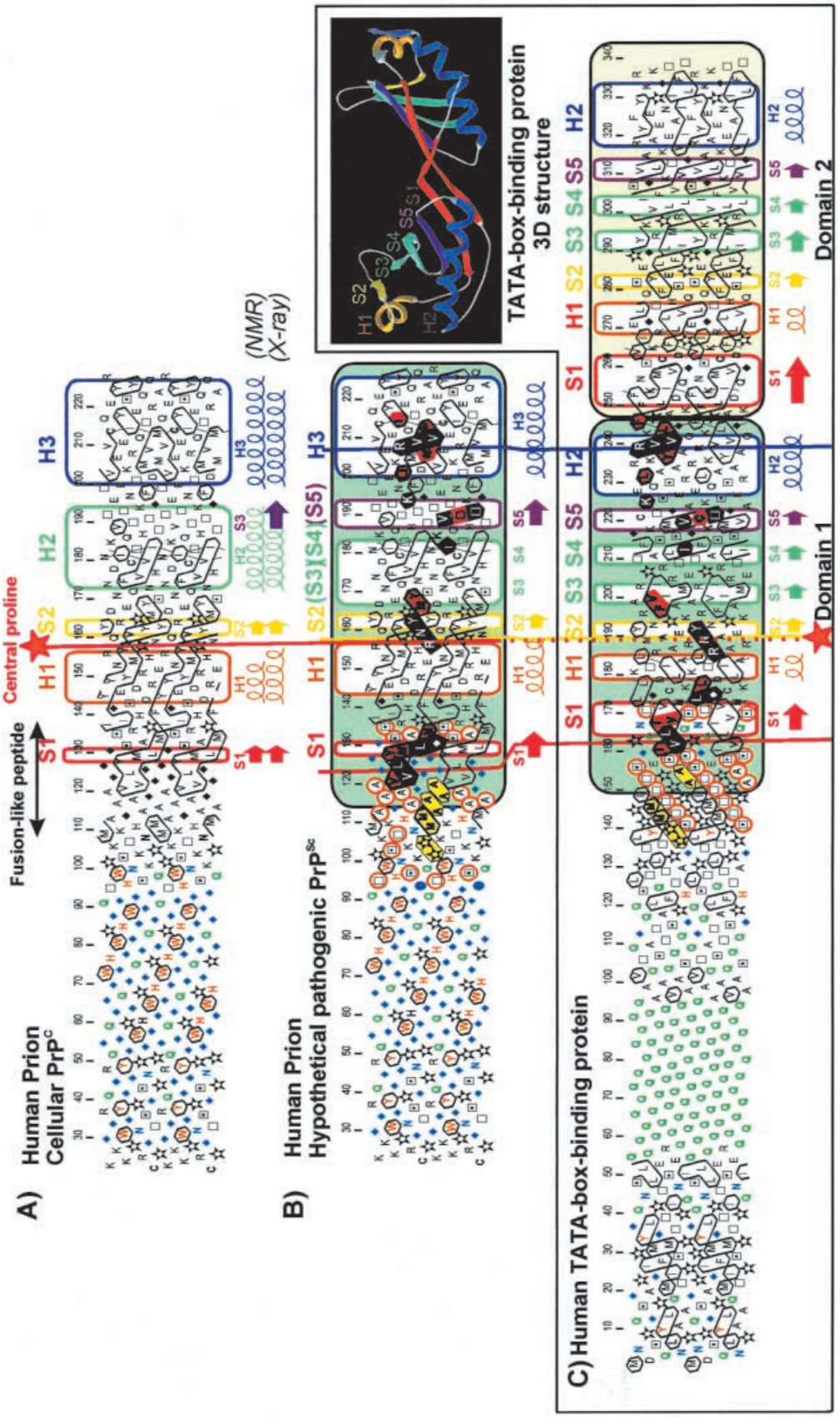
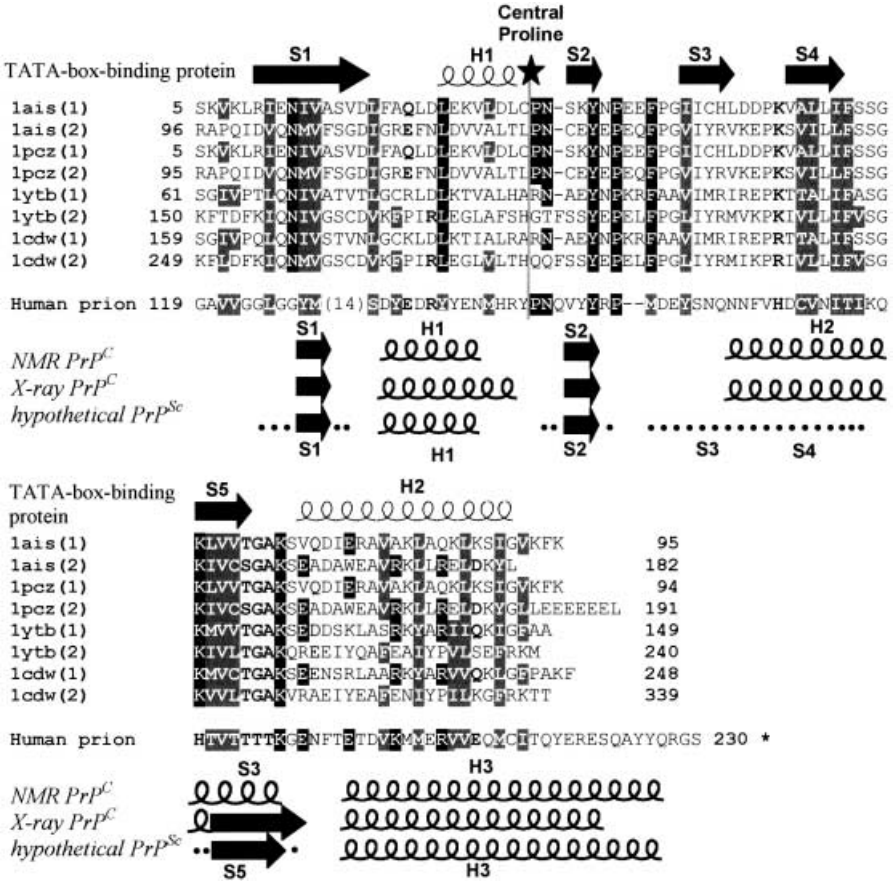


Figure 3. Comparison of the overall sequence texture of human prion and TATA-box-binding protein. (*A*) Comparison of the HCA plots. A) Human prion secondary structures, as observed from NMR (PDB identifier 1q1x) and X-ray (PDB identifier 1i4m) studies. B) Putative PrP^{Sc} secondary structures (see text). C) Secondary structures of the human TATA-box-binding protein (PDB identifier 1cdw), colored similarly to panel B. A ribbon representation of its 3D structure is shown right of lane B. Repeats are colored similarly to figure 1 A. In addition, before globular domains, alanine, serine, and threonine are circled orange. (*B*) Sequence similarities, deduced from the HCA-based alignment of observed secondary structures for the TATA-box-binding protein family with those expected for human prion (PrP^{Sc}). Sequence identities are reported white on a black background. Sequence similarities are shaded gray (white and black letters for hydrophobic and other amino acids, respectively). This comparison results in ~15 % sequence identity between 1cdw (1) of the TATA-box-binding protein family and human prion. However, this does not imply that these two proteins share a common ancestor, but merely highlights potential structural similarity of their C-terminal parts (after the 'central proline'). The illustrated members of the TATA-box-binding protein family are designated with their PDB identifiers, accompanied by domain numbering (1 or 2). Dots represent the region in which the putative strands S3 and S4 should lie, as well as possible extensions of the strands S1, S2, and S5.



B

Figure 3 (continued)

third one being recently confirmed by the dimeric X-ray structure of prion, as mentioned above [10]). Figure 3 moreover shows that several stretches of sequence similarities accompany this secondary-structure match. The typical 'central proline' PN motif of animal prion proteins is also found in members of the TATA-box-binding protein family (fig. 3B). Consequently, these data suggest that the TATA-box-binding protein may share, to some extent, structural features with prions and particularly with PrP^{Sc}.

The TATA-box-binding protein is required by the three eukaryotic RNA polymerases for correct initiation of transcription of ribosomal messenger, small nuclear, and tranfer RNAs. The crystal structure of human TATA-box-binding protein in complex with the DNA TATA element [54] revealed that the two successive and similar globular domains are wrapped around DNA, through a ten-β-stranded curved antiparallel sheet, which is zipped in its center by two long S1 strands (fig. 3A). Outside, a short helix H1 and a long helix H2 cover this sheet. Part of both β strands S1 swaps within the other globular domain reinforcing the cohesion of this architecture (fig. 3A).

The TATA-box-binding protein has a succession of regular secondary structures similar to those we expected for PrP^{Sc} from its sequence analysis, and thus may feature the fold adopted by the PrP^{Sc} structure. We therefore wish to design a coherent model of the globular domain of the human PrP^{Sc} prion, based on the TATA-box-binding protein structural organization, and how it may assemble itself to, together with the N-terminal repeats, constitute fibrils through $\alpha \rightarrow \beta$ conversion and possibly swapping. This study will be discussed in a forthcoming paper.

Acknowledgements. The authors thank Jérôme Hennetin for scanning databases with hydrophobic cluster signatures and Richard Eudes for the comparison of predicted and observed secondary structures of prion sequences.

1 Prusiner S. B. (1998) Prions. Proc. Natl. Acad. Sci. USA **95:** 13363–13383
2 Altschul S. F., Madden T. L., Schaffer A. A., Zhang J., Zhang Z., Miller W. et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402
3 Glockshuber R., Hornemann S., Billeter M., Riek R., Wider G. and Wüthrich K. (1998) Prion protein structural features indicate possible relations to signal peptidases. FEBS Lett. **426:** 291–296
4 Pan K. M., Baldwin M., Nguyen J., Gasset M., Serban A., Groth D. et al. (1993) Conversion of alpha-helices into beta-sheets

features in the formation of the scrapie prion proteins. Proc. Natl. Acad. Sci. USA **90:** 10962–10966
5 Huang Z., Prusiner S. B. and Cohen F. E. (1995) Scrapie prions: a three-dimensional model of an infectious fragment. Fold. Des. **1:** 13–19
6 Wille H., Michelitsch M. D., Guenebaut V., Supattapone S., Serban A., Cohen F. E. et al. (2002) Structural studies of the scrapie prion protein by electron crystallography. Proc. Natl. Acad. Sci. USA **99:** 3563–3568
7 Riek R., Hornemann S., Wider G., Billeter M., Glockshuber R. and Wüthrich K. (1996) NMR structure of the mouse prion protein domain PrP(121–321). Nature **382:** 180–182
8 Riek R., Hornemann S., Wider G., Glockshuber R. and Wüthrich K. (1997) NMR characterization of the full-length recombinant murine prion protein, mPrP(23–231). FEBS Lett. **413:** 282–288
9 Zahn R., Liu A., Luhrs T., Riek R., Schroetter C. von, Lopez Garcia F. et al. (2000) NMR solution structure of the human prion protein. Proc. Natl. Acad. Sci. USA **97:** 145–150
10 Knaus K. J., Morillas M., Swietnicki W., Malone M., Surewicz W. K. and Yee V. C. (2001) Crystal structure of the human prion protein reveals a mechanism for oligomerization. Nat. Struct. Biol. **8:** 770–774
11 Mo H., Moore R. C., Cohen F. E., Westaway D., Prusiner S. B., Wright P. E. et al. (2001) Two different neurodegenerative diseases caused by proteins with similar structures. Proc. Natl. Acad. Sci. USA **98:** 2352–2357
12 Bousset L., Belrhali H., Janin J., Melki R. and Morera S. (2001) Structure of the globular region of the prion protein Ure2 from the yeast *Saccharomyces cerevisiae*. Structure **9:** 39–46
13 Callebaut I., Tasso A., Brasseur R., Burny A., Portetelle D. and Mornon J.-P. (1994) Common prevalence of alanine and glycine in mobile reactive centre loops of serpins and viral fusion peptides: do prions possess a fusion peptide? J. Comput. Aided Mol. Des. **8:** 175–191
14 Davies S. M. A., Kelly S. M., Price N. C. and Bradshaw J. P. (1998) Structural plasticity of the feline leukaemia virus fusion peptide: a circular dichroism study. FEBS Lett. **425:** 415–418
15 Pécheur E.-I., Martin I., Bienvenüe A., Ruysschaert J.-M. and Hoekstra D. (2000) Protein-induced fusion can be modulated by target membrane lipids through a structural switch at the level of the fusion peptide. J. Biol. Chem. **275:** 3936–3942
16 Chang D.-K., Cheng S.-F., Trivedi V. D. and Tyang S.-H. (2000) The amino-terminal region of the fusion peptide of influenza. J. Biol. Chem. **275:** 19150–19168
17 Han X., Bushweller J. H., Cafiso D. S. and Tamm L. K. (2001) Membrane structure and fusion-triggering conformational change of the fusion domain form influenza hemagglutinin. Nat. Struct. Biol. **8:** 715–720
18 Ye S. and Goldsmith E. J. (2001) Serpins and other covalent protease inhibitors. Curr. Opin. Struct. Biol. **11:** 740–745
19 Whisstock J., Skinner R. and Lesk A. M. (1998) An atlas of serpin conformations. Trends Biochem. Sci. **23:** 63–67
20 Song L., Hobaugh M. R., Shustak C., Cheley S., Bayley H. and Gouaux J. E. (1996) Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. Science **274:** 1859–1866
21 Olson R., Nariya H., Yokota K., Kamio Y. and Gouaux E. (1999) Crystal structure of staphylococcal LukF delineates conformational changes accompanying formation of a transmembrane channel. Nat. Struct. Biol. **6:** 134–140
22 De Gioia L., Selvaggini C., Ghibaudi E., Diomede L., Buigiani O., Forloni G. et al. (1994) Conformational polymorphism of the amyloidogenic and neurotoxic peptide homologous to the residues 106–126 of the prion peptide. J. Biol. Chem. **269:** 7859–7862
23 Levy Y., Hanan E., Solomon B. and Becker O. M. (2001) Helix-coil transition of Pr106-126: molecular dynamic study. Proteins **45:** 382–396
24 Forloni G., Angeretti N., Chiesa R., Monzani E., Salmona M., Bugiani O. et al. (1993) Neurotoxicity of a prion protein fragment. Nature **362:** 543–546
25 Wilson I. A., Skehel J. J. and Wiley D. C. (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature **289:** 366–373
26 Bullough P. A., Hughson F. M., Skehel J. J. and Wiley D. C. (1994) Structure of influenza haemagglutinin at the pH of membrane fusion. Nature **371:** 37–43
27 Callebaut I., Labesse G., Durand P., Poupon A., Canard L., Chomilier J. et al. (1997) Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. Cell. Mol. Life Sci. **53:** 621–645
28 Pearson W. R. and Lipman D. J. (1988) Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:** 2444–2448
29 Gaboriaud C., Bissery V., Benchetrit T. and Mornon J.-P. (1987) Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. FEBS Lett. **224:** 149–155
30 Woodcock S., Mornon J.-P. and Henrissat B. (1992) Detection of secondary structure elements in proteins by hydrophobic cluster analysis. Protein Eng. **5:** 629–635
31 Hofmann K., Bucher P., Falquet L. and Bairoch A. (1999) The PROSITE database, its status in 1999. Nucleic Acids Res. **27:** 215–219
32 Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H. et al. (2000) The Protein Data Bank. Nucleic Acids Res. **28:** 235–242
33 Guex N. and Peitsch M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis **18:** 2714–2723
34 Kaplan W. and Littlejohn T. G. (2001) Swiss-PDB Viewer (Deep View). Brief Bioinform. **2:** 195–197
35 Vigon I., Mornon J.-P., Cocault L., Mitjavila M. T., Tambourin P., Gisselbrecht S. et al. (1992) Molecular cloning and characterization of MPL, the human homolog of the v-mpl oncogene: identification of a member of the hematopoietic growth factor receptor superfamily. Proc. Natl. Acad. Sci. USA **89:** 5640–5644
36 Murphy P. M. (2001) Viral exploitation and subversion of the immune system through chemokine mimicry. Nat. Immunol. **2:** 116–122
37 Linial M. L. (2000) Why aren't foamy viruses pathogenic? Trends Microbiol **8:** 284–289
38 Linial M. L. (1999) Foamy viruses are unconventional retroviruses. J. Virol. **73:** 1747–1755
39 Gabus C., Auxilien S., Pechoux C., Dormont D., Swietnicki W., Morillas M. et al. (2001) The prion protein has DNA strand transfer properties similar to retroviral nucleocapsid protein. J. Mol. Biol. **307:** 1011–1021
40 Gabus C., Derrington E., Leblanc P., Chnaiderman J., Dormont D., Swietnicki W. et al. (2001) The prion protein has RNA binding and chaperoning properties characteristic of nucleocapsid protein NCP7 of HIV-1. J. Biol. Chem. **276:** 19301–19309
41 Dominguez del Angel V., Dupuis F., Mornon J.-P. and Callebaut I. (2002) Viral fusion peptides and identification of membrane-interacting segments. Biochem. Biophys. Res. Commun. **293:** 1153–1160
42 Carp R. I. (2000) An endogenous retrovirus and exogenous scrapie in a mouse model of aging. Trends Biochem. Sci. **8:** 39–42
43 Kwong P. D., Wyatt R., Robinson J., Sweet R. W., Sodroski J. and Hendrickson W. A. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature **393:** 648–659
44 Morrissey M. P. and Shakhnovich E. I. (1999) Evidence for the role of PrP(C) helix 1 in the hydrophilic seeding of prion aggregates. Proc. Natl. Acad. Sci. USA **96:** 11293–11298

45 Deblandre G. A., Marinx O. P., Evans S. S., Majjaj S., Leo O., Caput D. et al. (1995) Expression cloning of an interferon-inducible 17-kDa membrane protein implicated in the control of cell growth. J. Biol. Chem. **270:** 23860–23866
46 Martins V. R., Linden R., Prado M. A. M., Walz R., Sakamoto A. C., Izquierdo I. et al. (2002) Cellular prion protein: on the road for functions. FEBS Lett. **512:** 25–28
47 Westaway D. and Carlson G. A. (2002) Mammalian prion proteins: enigma, variation and vaccination. Trends Biochem. Sci. **27:** 301–307
48 Spielhaupter C. and Schätzl H. M. (2001) PrP^C directly interacts with proteins involved in signaling pathways. J. Biol. Chem. **276:** 44604–44612
49 Hosszu L. L. P., Baxter N. J., Jackson G. S., Power A., Clarke A. R., Waltho J. P. et al. (1999) Structural mobility of the human prion protein probed by backbone hydrogen exchange. Nat. Struct. Biol. **6:** 740–743
50 Govindarajan S., Recabarren R. and Goldstein R. A. (1999) Estimating the total number of protein folds. Proteins **35:** 408–414
51 Wang Z. X. (1998) A re-estimation for the total numbers of protein folds and superfamilies. Protein Eng. **11:** 621–626
52 Zhang C. and DeLisi C. (1998) Estimating the number of protein folds. J. Mol. Biol. **284:** 1301–1305
53 Lo Conte L., Brenner S. E., Hubbard T. J., Chothia C. and Murzin A. G. (2002) SCOP database in 2002: refinements accommodate structural genomics. Nucleic Acids Res. **30:** 264–267
54 Nikolov D. B., Chen H., Halay E. D., Hoffman A., Roeder R. G. and Burley S. K. (1996) Crystal structure of a human TATA box-binding protein/TATA element complex. Proc. Natl. Acad. Sci. USA **93:** 4862–4867