J Mol Model (2005) 11: 468–473
DOI 10.1007/s00894-005-0239-8

ORIGINAL PAPER

David P Molloy · Beining Chen

Predicted consequences of site-directed mutagenesis and the impact of species variation on prion protein misfolding through the N-terminal domain

Received: 20 May 2004 / Accepted: 11 January 2005 / Published online: 21 July 2005
© Springer-Verlag 2005

Abstract Variant Creutzfeldt-Jacob disease (vCJD) is considered to afflict humans through the acquisition of variant isomers and misfolding of the normal cellular prion polypeptide, PrP^C. Although the exact mechanism of the misfolding is not been yet clearly understood, this paper provides four additional pieces of evidence in support of the hypothesis that misfolding within PrP^C involves N-terminal residues, up to and including Asn^{178}. Structural predictions for N-terminal residues between Leu^4 and Gly^{124} revealed that Leu^4–Leu^{19} might adopt a helical conformation. Furthermore, measurement of C_α distance variations, as determined from available NMR solution structures of wild type, as well as the biologically significant Val^{166}, Asn^{170} and Lys^{220} variants of PrP^C, revealed previously unreported global and local conformational differences may occur in PrP^C as a result of these amino-acid substitutions. Notably, three regions, His^{140}–Tyr^{150} and Met^{166}–Phe^{175} showed deviations greater than 3 Å in their C_α-coordinates (cf wild type) indicating that the majority of the N-terminal domain is likely to contribute to the misfolding of PrP^C. Minor variations in the orientation of amino acids Thr^{193}–Glu^{200}, located towards the C terminus of the protein, were also noted. This most likely indicates the presence of a hinge mechanism, inherent to a Helix-Loop-helix (HLH) motif formed by amino acids within α2, LIII and α3, in order to accommodate reorientation of the motif in response to misalignment of the N-terminal domain. An unexpected 3 Å deviation from the coordinates of the wild type polypeptide, absent from either Val^{166}, Asn^{170} variants was observed over the region Arg^{154}–Tyr^{155} within the Val^{166} form of PrP^C. This may contribute to the explanation as to why patients carrying the Val^{166} isoform of PrP^C may be more susceptible to vCJD.

Keywords vCJD · Misfolding · Bioinformatics · Structure prediction

D. P Molloy (✉)
Research Office, Thames Valley University, Ealing, London, W5 2BS, UK
E-mail: David.Molloy@tvu.ac.uk
Tel.: +44-20-82312328

B. Chen
Department of Chemistry, University of Sheffield, Sheffield, S3 7HF, UK

Introduction

Neurodegenerative diseases, for example variant Crutzfeld-Jacob disease (vCJD), pertain to abnormalities of protein folding and progressive deposition of amyloid protein in a β-sheet conformation [1, 2]. The pathology of prion diseases is, therefore, known to involve the refolding of normal cellular α-helical PrP^C into a pathological isomer, PrP^{SC} that contains a high component of β-strand [3] and is proteinase K (PK) resistant [4].

Recent NMR spectroscopic investigations performed on samples of the fibrous state of the prion demonstrated the existence of a $PrP^C \leftrightarrow PrP^{SC}$ conformational switch in vitro [5]. Similarly, FTIR spectroscopic techniques reveal that the globular domain (residues 90–230) of PrP^C adopts a 42% α-helical and 3.5% β-sheet conformation, whilst PrP^{SC} contains 30% α-helix and 43% β-sheet [6]. Although PrP^C has been isolated from samples of peripheral tissues and leukocytes [7] and more recently from within blood plasma samples, as demonstrated through use of the Conformation Dependent Immunoassay (CDI; [8]), the distribution of PrP^C within human tissues remains uncertain [7, 9–12]. The potential for misalignment of the protein backbone and aggregation is, therefore, of concern.

Given that the dramatic tertiary conformational switch within PrP^C is likely to arise from uniquely substituted amino acids in the primary sequence through single point mutation at the nucleic acid level, we have examined the global conformation of wild type human PrPC (huPrP^C) through computational analysis and measurements of C_α-coordinate distance variation using previously published NMR spectroscopic data. We

suggest that flexibility within the N-terminal domain might be central to the $PrP^{C} \leftrightarrow PrP^{SC}$ transition.

Materials and methods

Sequences and three-dimensional structures of proteins used in this study

The primary sequence for PrP^{C} from *Homo sapiens* used in this study was identified from searches of and uploaded from the SWISS-PROT database (accession number P04156 [13–15]).

The average from NMR conformational ensembles for wild type (1QMO; [16]), Met^{166} Val, Ser^{170} Asn and Arg^{220} Lys (1EIJ, 1EIS and 1EIW, respectively; [17]) and Glu^{200} Lys (1FCK [17]) isoforms of *Homo sapiens* PrP^{C} protein were uploaded from the protein data bank (www.rscb.org).

Structure predictions

Predictions of secondary structural elements within PrP^{C} were performed using the highly accurate PSIpred algorithm available at http://bioinf.cs.ucl.ac.uk/psipred [18], which predicts the secondary structure type for each residue in an amino-acid sequence. PSIpred allows submission of a protein sequence upon which an array of predictions can be made and the results of which are returned by e-mail (http://bioinf.cs.ucl.ac.uk/psipred). PSIpred uses an input file containing a sequence of amino acids for each protein using the single-letter abbreviation and produces a results output showing a prediction for individual amino acids as either: 'H' for a helix element, 'E' for a beta strand element, or C for a random element with of between 79 and 85% reliability. Structural predictions for the Val^{166}, Asn^{170} and Lys^{220} variants of PrP^{C} were performed using input files containing wild type polypeptide sequence substituted manually with the appropriate single letter abbreviation corresponding to the mutation.

Molecular modeling

Molecular modeling was carried out using Swiss pdb viewer [19], allowing several PDB files to be displayed and analyzed simultaneously. Proteins were typically superposed after structural alignment using the autofit and iterative magic fit subroutines, in order to minimize the root mean square deviations (RMS). Distance variation in C_{α}-coordinates between wild type and variants of $huPrP^{C}$ were then determined iteratively.

Results and discussion

Within worldwide human populations, the 127 incidences of vCJD account for only a miniscule percentage of the total population, with susceptibility of individuals to the disease believed to occur primarily through acquisition of an aberrant form of the prion protein, PrP^{SC}. Thus, single point mutations, especially in human PrP are considered of importance, as from a total of 26 variants identified to date, the 13 illustrated (Fig. 1) are believed to be associated with, or are in part responsible, for vCJD in humans [1].

Secondary structure prediction analysis of PrP^{C}

As an initial approach, secondary structural elements within full-length wild type $huPrP^{C}$ (Fig. 2a) were investigated using the prediction algorithm, PSIpred. The findings were on the whole comparable to the NMR solution structures for the globular domains of $HuPrP^{C}$ [16, 17] of varying sequence lengths, including amino acids Gly^{124}–Asn^{231} (Fig. 2b). It was noted from the PSIpred prediction that residues Gly^{20}–Gly^{124} adopt a predominately random conformation consistent with inferences derived from the NMR spectroscopic investigation [16].

It was of interest to note that amino acids in the region Leu^{4}–Leu^{19} are predicted to adopt a α-helix conformation (Fig. 2a). This finding provides valuable insights into potential conformations present within the N terminus of PrP^{C}. Whilst NMR spectroscopic inves-

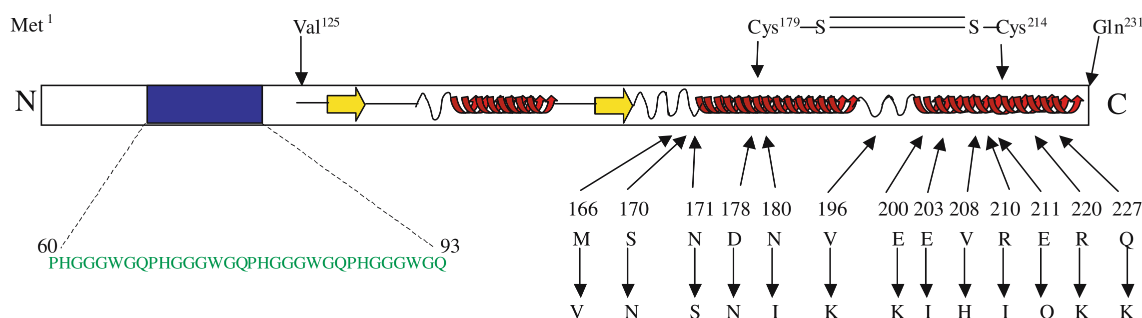

Fig. 1 Schematic representation of the primary structure of PrP^{C}. The sequence of PrP^{C} (Met^{1}–Gln^{231}) from human is illustrated with structural elements corresponding to β-strand and α-helix within the globular fold (Val^{125}–Gln^{231}). The octapeptide repeat signal (a.a.s Pro^{60}–Gln^{93}) within the N-terminal domain presented (*green typset*). The *arrows below* the PrP^{C} protein represent the biologically significant mutations associated with the development of scrapie

a)

```
Conf: 960589999999987433202488766677877788675324222736211222111222222221111111133444685878787887
Pred: CCCHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
  AA: MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWG
                10        20        30        40        50        60        70        80        90

Conf: 788656778788878774245432220122003543012244414745416971466677665301786017701688788762011201
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHCCCEEECCCHHHHHHHHHHHCCCCCEEEEEECCCCCCCCCCCCCCCC
AA:   QGGGTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNNFVHDCV
               100       110       120       130       140       150       160       170       180

Conf: 01688754368766862124458999999999987789999988763057 4
Pred: CCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCC
  AA: NITIKQHTVTTTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGS
               190       200       210       220       230
```

b)

Fig. 2 Structural properties of hPrP^C. **a** PSIpred prediction for wild type hPrP^C. The experimental output from the PSIpred algorithm is illustrated for the 253 amino acids present in wild type hPrP^C using the single letter abbreviation (*AA*) in *green typeset*. The predicted secondary structure (*Pred*) for individual amino acids is presented above the amino-acid sequence by the assignment for helix (*H*), strand (*E*) or random coil (*C*). Statistical analyses for structure predictions assigned to individual amino acids (*Conf*) are shown above the prediction being assigned values between 0 (*low*) and 9 (*high*) statistical correlation, respectively. **b** Three-dimensional structure of the globular fold of wild type hPrP^C. The backbone structure for wild type PrP^C is illustrated as a protein cartoon whereby amino acids adopting roman numerical loops coil (*L*), α-helices (α) and β-strands (β) are presented in *blue*, *red* and *yellow*, respectively. The disulphide bridge (*yellow line*) formed Cys^{179}–Cys^{214} is shown surrounded by van der Waal spheres

tigations on recombinant forms of PrP^C residues Lys^{23}–Ser^{230} are available, data for the sequence Lys^{23}–Gly^{123} are absent from the PDB coordinates and this region is deemed to adopt a random conformation [16]. Equivalent findings were obtained from structure-prediction studies performed using the primary sequences from the Val^{166}, Asn^{170} and Arg^{220} PrP^C isoforms (data not shown). This finding indicates that the data are on the whole comparable between the experimentally derived studies elsewhere [16, 17] and thus reflecting high accuracy, yet uniqueness of the PSIpred algorithm [18].

It is not expected that the subtle changes in the three-dimensional structure as a function of amino acid substitution within PrP^C would be detected in secondary structure predictions. The results obtained (see above) are, however, of further interest as recent FTIR spectroscopic investigations have revealed that the globular domain (residues 90–230) of PrP^C adopts a 42% α-helical and 3.5% β-sheet content, whilst PrP^{SC} contains 30% α-helix and 43% β-sheet [6].

Further examination of the three-dimensional structure of PrP^C ([16, 17] Fig. 2b) reveals that the C terminus of PrP^C is stabilized through a disulfide bridge Cys^{179}–Cys^{214} (Fig. 2b) and a substantial number of hydrogen bonds. It is therefore plausible that reorganization of the C terminus is of less concern than unwinding and refolding of the N-terminal domain.

Computational analysis of the three-dimensional structure of PrP^C s

Through the analysis of secondary conformation (see above), residues between Leu^4–Leu^{19} at the N terminus of PrP^C are predicted to adopt an α-helical conformation. The unwinding of which, combined with local and global conformational changes as a consequence of amino acid substitution, is likely to indicate firstly that the protein might display greater flexibility than previously considered. Secondly, the conformational change

$PrP^C \leftrightarrow PrP^{Sc}$ involves the N terminus of the protein. Further computational studies were, therefore, performed to investigate the impact of mutation within the available PDB coordinates of the PrP^C isoforms wild type, Val^{166}, Asn^{170} and Lys^{220} over the region Gly^{124}–Ser^{230} (Fig. 3).

Significant variations were observed over the regions His^{140}–Tyr^{150}, Asn^{153}–Tyr^{157}, Val^{166}–Phe^{175} and Thr^{193}–Glu^{200} of the Val^{166} variant (cf wild type $huPrP^C$; Fig. 3); His^{140}–Tyr^{150}, Tyr^{162}–Tyr^{178} and Thr^{190}–Glu^{200} for the $huPrP^C$ variant Asn^{170}, as well as His^{140}–Tyr^{150}, Pro^{165}–Tyr^{172} and Thr^{190}–Glu^{200} for the Lys^{220} variant (Fig. 3). With the exception of the region Asn^{153}–Tyr^{157} within the Val^{166} isoform (Fig. 3), little or no significant variation from the wild type structure was observed for each of the three PrPC variants between residues Arg^{151}–Tyr^{162} and Cys^{179}–Thr^{190}. This suggests that variation from the wild type protein structure in these variants of $huPrP^C$ is small and unlikely to be of significance to misfolding of PrP^C.

A radical deviation, unique to the Val^{166} isoform, of 3.0 Å from the wild type $huPrP^C$ structure was observed over the region His^{155} and Arg^{156} (Fig. 3). This finding is of interest as this region spans helices α1 and α2 within the wild type polypeptide [16] and consequently, may contribute further to misfolding likely to occur in the conformational switch of Val^{166} $PrP^C \leftrightarrow Val^{166}$ PrP^{Sc}. Interestingly, this region also lies outside of the vicinity of amino acids 129 and 178, where the perturbations are thought to be the largest for the conversion $PrP^C \leftrightarrow PrP^{Sc}$.

Recent data have suggested that the development of vCJD in *Mesocricetus auratus*, in which PrP^C carries Asn at position 170 [12], is comparatively rapid with respect to *Mus muculus* strains [12]. It is considered that this arises primarily owing to the presence of a more extensive C-terminal α-helix, α3 in $hPrP^C$ (cf $huPrP^C$). In addition, the $Ser^{170} \rightarrow$ Asn substitution within $huPrP^C$, where α3 is less extensive, is considered of importance to the onset of vCJD in *Homo sapiens* (Fig. 1; [17]). The present comparisons of wild type and Asn^{170} variant of huPrP vary with that previously reported [12–14] and here, reveal that regions His^{140}–Tyr^{150}, Arg^{164}–Phe^{175} and Thr^{193}–Glu^{200} show notable variations in their averaged structures. This variation is more extensive over the region formed by L2 and α2 with respect to counterpart regions in either Val^{166} or Lys^{220} isoforms.

Examination of these three-dimensional structures indicates that Asn at position 170 within PrP^C is unlikely to adopt any specific orientation and hydrogen-bonding pattern consistent with high mobility observed for many polypeptides. However, the extent of structural variation in this region of Asn^{170} $huPrP^C$ is significantly greater than for either Val^{166} or Lys^{220} isoforms and extends between residues Glu^{168}–Val^{175}. A further region of significant variation for Asn^{170} $huPrP^C$ from the wild type protein was observed over the region Ile^{139}–Arg^{150}. Thus, the influence of the Ser^{170} Asn substitution upon misfolding of $huPrP^C$ and as a consequence, the development of vCJD is most likely mediated through the first 178 amino acids of PrP^C, rather than residues within the C terminus.

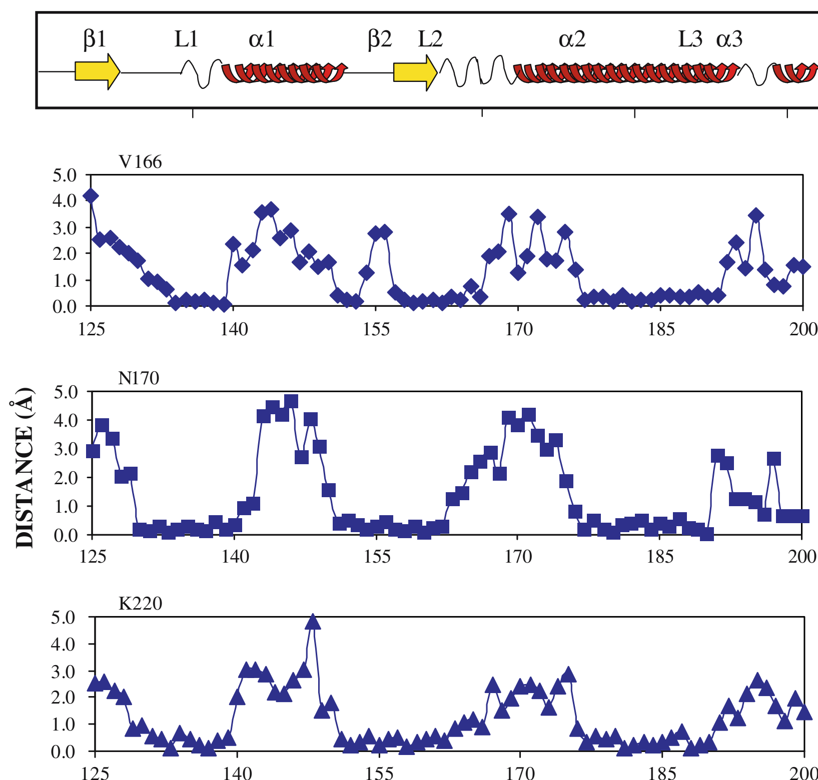

Fig. 3 Distance variations between wild type and mutant PrP^C s. The variations in distance (Å) between the C_α backbone of Val^{166}, Asn^{170} and Lys^{220} variants compared to wild type $hPrP^C$ are illustrated over Leu^{125}–Glu^{200}. All coordinates for the energy-minimized average of conformational ensembles (n = 20) were uploaded from the Protein Data Bank (www.rcsb.org) and superposed in Swiss pdb viewer (see Materials and methods). In the *upper most panel* a schematic for the secondary structural elements of alpha helices, beta strands, loop regions and random conformations within the global fold of wild type PrP^C is presented

Interestingly, a deviation from wild type form of the protein was observed for all three variants over the region Thr^{190}–Glu^{200}. Within wild type $huPrP^{C}$, this region forms a helix-loop-helix (HLH) motif involving the N terminus of α2, LIII and the N terminus of α3 and mediates the orientation of α2 with respect to α3 through a disulphide bridge Cys^{179}–Cys^{214} (Fig. 3). Flexibility over the region Thr^{190}–Glu^{200} described herein, as well as the presence of the disulphide bridge Cys^{179}–Cys^{214}, regardless of Val^{166}, Asn^{170} or Lys^{220} substitution [16], indicates that the HLH motif remains intact within the three-dimensional conformation of these variants. Thus, the observed variations of 2.5–3.5 Å from the wild type structure presumably involve only a minor unwinding of the C terminus of α2 and N terminus of α3 and changes to the backbone conformation of LIII. In addition, the orientation of α2 with respect to α3 and maintenance of the HLH motif within the corresponding PrP^{Sc} state of each variant may only involve minor perturbations. This suggests that the HLH motif undergoes only minor structural changes as a result of more major conformational changes that occur within the N-terminal domain during misfolding, although confirmation of this will have to await further study.

The observation concerning the unfolding of PrP^{C} as mediated through the N-terminal domain is further exemplified within the Lys^{220} isoform [17], where the corresponding regions His^{140}–Tyr^{150}, Arg^{164}–Phe^{175} also show significant deviations from the wild type structure and provide more evidence for conformational flexibility within the N-terminal domain of PrP^{C}.

The plausibility of unwinding C-terminal α-helices of PrP^{C} has been investigated using the peptide mimetic approach [20–22]. Although the transition between α-helix and random coil conformations is believed to occur via a 3_{10}-helix intermediate in which the carbonyl oxygen acceptor in a CO_i–$NH_{i,i+3}$ hydrogen bond, migrates to form a CO_i–$NH_{i,\ i+4}$ hydrogen bond, forming an isolated turn [23]. Such a mechanism would be expected to occur within small synthetic peptides [20, 21] in the absence of disulfide bridges. Interestingly, alanine-rich synthetic peptides equivalent to PrP residues His^{106}–Val^{122} and Met^{109}–Val^{122} within the N-terminal domain have been shown to adopt staggered lattice arrangements [24]. Furthermore, the fact that His^{140}–Tyr^{150}, Arg^{164}–Phe^{175} and Thr^{191}–Glu^{200} within the Lys^{220} isomer match closely the counterpart regions within the Val^{166} and Asn^{170} isomers of $HuPrP^{C}$ as being regions of highest variation from the wild type structure, which would suggest a these proteins share a common mechanism of misfolding. Whilst there are likely to be slight differences in averaging processes to derive NMR structural data, deviations over the amino-acid Thr^{190}–Phe^{200} region within PrP^{C} are similar between the proteins and stabilization of the C-terminal helices α2 and α3 would mitigate against this region as a site for unfolding of PrP^{C}. Thus, regions N-terminal to this can be considered as possessing the potential to misfold in all three variants of PrP^{C} studied.

Conclusions

Four additional pieces of evidence to support the hypothesis for misfolding within PrP^{C} as occurring in N-terminal residues [24] up to, and including Asn^{178}, were obtained in this study.

First, structural prediction studies upon the amino acids within the N terminus (a.a.s Leu^{4}–Gly^{124}) revealed that residues between Leu^{4}–Leu^{19} might adopt a α-helical conformation.

Second, superposition of the structures for wild type and three isoforms of $huPrP^{C}$ showed significant variations in the C_{α}-distance coordinates over the N-terminal region Gly^{124}–Asp^{178} with only minor variations in distances of amino acids between Cys^{179}–Ser^{230}. Thus, it is expected that variations within the C-terminal domain, stabilized through a single disulphide bridge Cys^{179}–Cys^{214} would be minimal and therefore, misfolding most likely involves the more flexible N-terminus.

Third, there are three regions within PrP^{C} ; His^{140}–Tyr^{150}, Met^{166}–Phe^{175} and Thr^{193}–Glu^{200} that regardless of the isoform of PrP^{C} studied, 16 residues have a greater than 3′ variation from the wild type structure. This finding provides further evidence that the N terminus is more flexible and consequently may be prone to significant structural changes such as unwinding and refolding upon the conversion of PrP^{C} to PrP^{Sc}.

Forth, within the Val^{166} isoform, an additional component of structural variation observed over amino acids Arg^{154}–Tyr^{155}, which was absent from other structures analyzed may explain why patients with this mutation may be more susceptible to CJD.

Acknowledgements DPM is most grateful for the provision of facilities by TVU and dedicates this paper to his late aunt, Dorothy Molloy. BC is indebted to the Department of Health (UK) for financial assistance.

References

1. Pitschke M, Prior R, Haupt M, Riesner D (1998) Nat Med 4:832–834
2. Bolton DC, McKinley MP, Prusiner SB (1982) Science 218:309–1311
3. Brown DR, Hafiz F, Glasssmith LL, Wong BS, Jones IM, Clive C, Haswell SJ (2000) EMBO J 19:1180–1186
4. Horiuchi M, Caughey B (1999) EMBO J 18:3193–3203
5. Collinge J, Whittington MA, Sidle KCL, Smith CJ, Palmer MS, Clarke AR, Jefferys JGR (1994) Nature 370:295–297
6. Pan K, Baldwin M, Nguyen J, Gasset M, Serban A, Groth D, Mehlhorn I, Huang Z, Fletterick RJ, Cohen FE, Prusiner SB (1993) Proc Natl Acad Sci USA 90:10962–10966
7. Pauly PC, Harris DA (1998) J Biol Chem 273:33107–33110
8. Bellon A, Seyfert-Brandt W, Lang W, Barron H, Vey M (2003) J Gen Virol 84:1921–1925

9. Alper T, Cramp WA, Haig DA, Clarke MC (1967) Nature 214:764–766
10. Kellings K, Prusiner SB, Riesner D (1994) Philos Trans R Soc Lond B 343:425–430
11. Purdey M (2000) Med Hypotheses 54:278–306
12. Priola SA, Chabry J, Chan K (2001) J Virol 75:4673–4680
13. Locht C, Chesebro B, Race R, Keith JM (1986) Proc Natl Acad Sci USA 83:6372–6376
14. Turk E, Teplow DB, Hood LE, Prusiner SB (1998) Eur J Biochem 176:21–30
15. Kretzschmar HA, Stowring LE, Westaway D, Stubblebine WH, Prusiner SB, Dearmond SJ (1986) DNA 5:315–324
16. Zahn R, Liu A, Luhrs T, Riek R, von Schroetter C, Lopez Garcia F, Billeter M, Calzolai L, Wider G, Wuthrich K (2000) Proc Natl Acad Sci USA 97:145–150
17. Calzolai L, Lysek DA, Güntert P, von Schroetter C, Riek R, Zahn R, Wüthrich K (2000) Proc Natl Acad Sci USA 97:8340–8345
18. Jones DT (1999) J Mol Biol 292:195–202
19. Guex N, Peitsch MC (1997) Electrophoresis 18:2714–2723
20. Molloy DP, Milner AE, Yakub IK, Chinnadurai G, Gallimore PH, Grand RJA (1998) J Biol Chem 273:20867–20876
21. Molloy DP, Smith KJ, Milner AE, Gallimore PH, Grand RJA (1999) J Biol Chem 274:3503–3512
22. Gasset M, Baldwin MA, Llyod DH, Gabriel J-M, Holtzman DM, Cohen FE, Fletterick R, Prusiner SB (1992) Proc Natl Acad Sci USA 89:10940–10944
23. Tirado-Rives J, Jorgensen WL (1991) Biochemistry 30:3864–3871
24. Inouye H, Kirschner DA (2003) Fibre Ref Rev 11:102–112