COMMUNICATION

Similarity between the C-terminal domain of the prion protein and chimpanzee cyromegalovirus glycoprotein UL9

Igor B.Kuznetsov and S.Rackovsky

Department of Biomathematical Sciences, Mount Sinai School of Medicine, Box 1023, One Gustave L. Levy Place, New York, NY 10029, USA

Introduction

Prion diseases are a group of fatal neurodegenerative disorders associated with structural conversion of a normal, mostly $\alpha$-helical cellular prion protein, PrP$^C$, into a pathogenic $\beta$-sheet-rich conformation, PrP$^Sc$. The structure of PrP$^Sc$ is well studied, whereas the insolubility of PrP$^Sc$ makes the characterization of its structure problematic. No proteins similar to PrP, except for its paralog with the same fold, PrP-Doppel, are known. However, PrP-Doppel does not undergo a structural transition into a $\beta$-sheet-rich conformation. Structural information from proteins that share a weak but significant sequence similarity with PrP may be used to gain additional insights into the conformation of PrP$^Sc$. We construct a sequence profile corresponding to the structured domain of PrP and use this profile to search the SWISS-PROT and TrEMBL databases. We identify a significant sequence similarity between PrP and chimpanzee cyromegalovirus glycoprotein UL9. This glycoprotein scores higher than all PrP-Doppel sequences. Fold recognition methods assign a mainly-$\beta$ fold to UL9. Owing to the observed sequence similarity with PrP and a putative mainly-$\beta$ fold, the UL9 glycoprotein may represent a potential target for experimental structure determination aimed at obtaining a structural template for PrP$^Sc$ modeling.

Keywords: alignment/conformational transition/Doppel/sequence profile

Methods and results

We used sequence profile search software (Gribkov and Veretnik, 1996) from the GCG Package [Wisconsin Package Version 10; Accelrys (GCG), 9685 Scranton Road, San Diego, CA 92121, USA] for PrP sequences corresponding to the structured C-terminal domain of the human PrP (residues 128–231) was used to construct the sequence profile. The disordered N-terminal part of PrP has very low sequence complexity and was excluded in order not to bias the results. We used 57 non-identical PrP sequences for which the complete sequence of the C-terminal domain is available for sequence profile construction (Table I). The multiple alignment was obtained by means of the GCG PILEUP program using the BLOSUM50 similarity matrix (Henikoff and Henikoff, 1992), a gap initiation penalty of –16
and a gap extension penalty of ±4. The sequence profile was calculated using the GCG PROFILEMAKE program and the BLOSUM50 similarity matrix with the default logarithmic weighting. The profile was searched against the SP-TrEMBL database (release 22.0) (Bairoch and Apweiler, 2000) using the GCG PROFILESEARCH program with default local alignment options. The profile was able to identify all prion proteins as well as Doppel sequences among the highest scoring hits.

The highest scoring (z-score = 11.76) non-PrP and non-Doppel sequence is the chimpanzee cytomegalovirus (CMV) glycoprotein UL9 (Table I). UL9 is the only non-PrP sequence which scores higher than all Doppel proteins. It should be noted that the prion protein and Doppel share the same fold and ~25% sequence identity. The next highest scoring non-PrP and non-Doppel sequence (TrEMBL accession No. Q9DFV7) has a z-score of 8.63, which is lower than the z-scores of all but one...
The CMV is a member of the herpesvirus group. It has been proposed as the most prevalent infectious agent causing neurological dysfunction in the developing brain, and therefore has a high affinity for developing brain cells (van den Pol et al., 2002). Normal cellular prion protein is also expressed in brain cells, and accumulation of abnormal, insoluble PrPSc leads to neurological dysfunctions as well. Since PrP is located on the outer membrane and undergoes endocytosis, it has been hypothesized that PrP may function as a receptor protein participating in signal transduction (Prusiner et al., 1998). The observed sequence similarity between the prion protein and CMV glycoprotein UL9 is likely to reflect structural similarities. However, the structure and function of UL9 are not known. In the absence of an obvious homologous template, potentially matching folds can be identified using fold recognition techniques.

We used the mGenThreader fold recognition server (McGuffin and Jones, 2003), which has been shown to have the lowest rate of false positive predictions among all automated fold recognition servers (Bujnicki et al., 2001), to make predictions for UL9 protein. It should be noted that all highest scoring templates (E-value from 0.03 to 0.06) belong to mainly-β proteins involved in substrate binding: immunoglobulin antigen-binding domains (PDB i.d. 8fab, 12e8, 1a3l, 32c2, ligt) and T-cell receptors (PDB i.d. 1tc, 1hxm, 1bec). A different fold recognition method, SAM_T02 (Karplus et al., 2001), also assigns highest scoring hits for UL9 to immunoglobulin-antigen-binding domains and T-cell receptors. The same two methods do not find any significant matches for the C-terminal domain of the prion protein, except for the match between PrP and Doppel. The evidence of a putative mainly-β fold of the UL9 protein and its sequence similarity with the prion protein, which undergoes a conformational transition into mainly-β conformation, identify UL9 as a potential target for experimental structure determination aimed at obtaining a template for modeling the structure of PrPSc. Further progress in structural and functional annotation of UL9 may help understand the function of PrP and what type of substrate it binds.

Acknowledgements

This work was supported by grant number 1R01 LM06789 from the National Library of Medicine of the National Institutes of Health. I.B.K. is supported by NSF EPSCoR.

References

