Published online 31 October 2008 *Nucleic Acids Research, 2009, Vol. 37, Database issue* ***D185–D190***
doi:10.1093/nar/gkn865

Minimotif miner 2nd release: a database and web system for motif search

Sanguthevar Rajasekaran[1,*], Sudha Balla[3], Patrick Gradie[1], Michael R. Gryk[2], Krishna Kadaveru[2], Vamsi Kundeti[1], Mark W. Maciejewski[2], Tian Mi[1], Nicholas Rubino[1], Jay Vyas[2] and Martin R. Schiller[2]

[1]Department of Computer Science and Engineering, University of Connecticut, Storrs, CT 06029-2155, [2]Department of Molecular, Microbial, and Structural Biology, Biological System Modeling Group, University of Connecticut Health Center, 263 Farmington Ave. Farmington, CT 06030-3305 and [3]Memorial Sloan-Kettering Cancer Center, NY 10021, USA

Received August 15, 2008; Accepted October 16, 2008

ABSTRACT

Minimotif Miner (MnM) consists of a minimotif database and a web-based application that enables prediction of motif-based functions in user-supplied protein queries. We have revised MnM by expanding the database more than 10-fold to approximately 5000 motifs and standardized the motif function definitions. The web-application user interface has been redeveloped with new features including improved navigation, screencast-driven help, support for alias names and expanded SNP analysis. A sample analysis of prion shows how MnM 2 can be used. Weblink: http://mnm.engr.uconn.edu, weblink for version 1 is http://sms.engr.uconn.edu.

INTRODUCTION

Protein sequence homology analysis has proven effective in inferring protein function, most notably by facilitating the identification of similar protein domains in different genes and organisms. The numerous resources for protein domain analysis include SMART, ProSite, ProRule, InterPro, Blocks, eBLOCKs, Prints, CoPS, pFAM, CDART and CDD (1–11). The function of a domain can be inferred from previously characterized proteins and subsequently confirmed in the uncharacterized proteins.

As protein domains are highly conserved throughout evolution, it is logical to expect that their binding partners or substrates would be conserved as well. Conserved binding or substrate motifs provide complementary information about protein function. These contiguous motifs are restricted to a single secondary structure element, typically consist of fewer than 15 amino acids, and are termed '*minimotifs*' to distinguish them from the longer, more complex motifs which serve as domain signatures. For example, the Pro-X-X-Pro sequence in proteins forms a polyproline type II left-handed helix, which binds to a hydrophobic surface of SH3 domains. Identifying a putative Pro-X-X-Pro minimotif within a protein can be equally as insightful as identifying a putative SH3 domain.

Minimotifs are the pattern signatures that define the targets of domain and are not signatures for the domains, themselves. While there are many resources for analyzing domains, far fewer resources exist for minimotifs. Rather short contiguous functional motifs are generally catalogued by functional groupings and dispersed over a collection of specialized databases such as MEROPS, Phosphobase and PDZbase (12–14). Minimotif Miner (MnM) contains a broader set of minimotifs allowing analysis of many different types of minimotifs in a single query (15). Through minimotif prediction, MnM provides the means for identifying new aspects of protein function, regulation and generating new hypotheses concerning the causes of disease (16,17).

MnM WEB SITE

There are numerous individual databases and search algorithms for identifying different types of minimotifs, most often categorized by a single function (e.g. prediction of phosphorylation sites). This approach is tremendously limiting as locating and querying each individual database with proteins of interest is not practical. Thus, biologists are not aware of the many potential functions in the proteins they study. To address this problem we have

*To whom correspondence should be addressed. Tel: +860 486 2428; Fax: +860 486 4817; Email: rajasek@engr.uconn.edu
Correspondence may also be addressed to Martin R. Schiller. Tel: +860 6794610; Fax: +860 679 1726; Email: schiller@nso.uchc.edu

built MnM, a database of functional minimotifs and an associated web-based application to enable querying of the database. The first version of MnM released in 2006, had 462 short functional minimotifs (15). These minimotifs were obtained by manually searching the biological literature and including minimotifs from other specialized databases.

The MnM database and webtool are complementary to other major systems for minimotif prediction. Eukaryotic Linear Motif Resource (ELM) and Scansite have a more limited set of minimotifs. ELM provides a more detailed annotation for each motif and Scansite uses experimental data to derive position-specific scoring matrices rather than using consensus sequence definitions (18,19). For automated annotation of minimotifs, Rigoutsos and colleagues (20) developed a Biodictionary of amino acid patterns and their annotations. MnM is synergistic with these exisiting resources having several unique features. A novel aspect of MnM is that minimotifs identified in a protein query can be ranked in terms of their likelihood of being functional using three independent scoring metrics. These metrics are based on frequencies of occurrence of the motif, evolutionary conservation of the motif and the probabilities of the motif occurring on the protein surface. While these metrics each have limitations, they allow users to rank candidate motifs. Other aspects that distinguish MnM from other short motif databases are the relatively larger number of motifs. The long-term goal is for MnM to be a comprehensive database of short contiguous functional motifs.

In this article, we summarize the revision of old and addition of new functions on the MnM website, and the growth of the MnM database to more than 5000 minimotifs, and provide an analysis of prion to show how MnM 2 can be used.

USER INTERFACE SEARCH PAGE IMPROVEMENTS

Improved navigation

The MnM 2 website search page has been redesigned to better organize different types of data and improve site navigation. MnM 2 now has a title bar that has pulldown lists of links to databases, funding, help, minimotif links, domain links, homology links, people working on the MnM projects, and publications on MnM, citing MnM and other publications related to minimotif analysis.

Enhanced help

In addition to a revised file of a sample analysis provided in the original MnM, we now provide a series of screencast tutorials. These tutorials include an overview, motif sequence definitions and tutorials for the search and results pages. Tutorials for the search page include: finding a RefSeq accession number, basic MnM search, restricting the search by subcellular localization and searching proteomes for a user-defined motif. Tutorial for the results page include: SNP analysis, homologous protein analysis, interpreting motif table, using and interpreting the frequency score, the surface prediction score and homology conservation scores.

Input accepts protein names, protein name aliases and alias accession numbers

In the first release of MnM, a user could only query using a protein's RefSeq accession number or a protein sequence. In version 2, we have added the ability to query using protein names, aliases and accession aliases. Inputs are queried against the MnM database for the accession number. If the accession number is not found, the database is queried for a protein name alias and subsequently an accession alias. Information for aliases was derived from the Entrez Gene database. If neither is found, a message is displayed.

Automated species selection

The original version of MnM required users to select from one of 10 species for RefSeq accession number entries. For proteins selected by name or accession number, the species is now retrieved by first identifying a RefSeq record, which has species as an attribute. This enforces correct species choice which was lacking in the original MnM. In MnM 2, the user still needs to select a species for pasted protein sequence entries. However, now the species for all completed proteomes are listed. An auto-fill feature utilizes AJAX to provide a list of choices in a pulldown menu as the user types in the species name. This species selection is used to calculate a frequency score which can be used to rank predicted motifs.

UPDATED MnM 2 RESULTS PAGE

Expanded single nucleotide polymorphism (SNP) analysis

In the original MnM, a function allowed mapping of SNPs from the dbSNP database onto the protein sequence in the Protein Sequence Window (21). When SNPs were loaded a new MnM search identified minimotifs in the new protein sequence, thus identified putative minimotifs introduced by SNPs. However, this SNP analysis was limited in several ways; (i) the effect of SNPs could only be analyzed as a group of all SNPs present in dbSNP, (ii) minimotifs that were eliminated by an SNP were not assessed and (iii) users could not readily identify SNPs that affect motifs, without an exhaustive comparison search through the Motif Tables with and without SNPs selected.

In MnM 2, we have revised the SNP function to address these limitations. SNPs can now be analyzed individually or in any combination. The SNPs in Protein Sequence window can now be dynamically changed by clicking on the SNP residue. Each SNP can be changed independently with the new SNP residue highlighted in green and the sequence of the original SNP position highlighted in blue. After changing one or more SNPs, the user can select the 'View motifs from new SNPs' button which will produce a new table that shows minimotifs that were introduced (colored green) or eliminated (colored red) by the group of selected SNPs.

The computation behind SNP minimotif search was done as follows. When a user queries a protein with selected SNPs, both the new and the old sequences are sent to the request handler. The new sequence is searched

for the minimotifs in the database and these motifs are stored in a list. The list of motifs from the old sequence is then compared with the new sequence and two more lists are populated. These lists are made up of the new motifs found due to the change in the sequence and the lost motifs from the old sequence due to the change. Each position of the new or lost motifs is also recorded and presented to the user in a table.

Formatted output

The original release of MnM only allowed for a printer-friendly output format that was not readily imported by other programs without building a parser. In version 2, support has been added for output as an excel file to give the user more flexibility in the usage of their query results.

Grouping of related motifs

As the MnM database grows the number of hits for a given protein query is expanding. To minimize redundancy we are grouping related motifs in the motif table. We are now grouping motifs based on a common subset of motif attributes in the database (required posttranslational modification, activity, subactivity, target domain, domains site, multidomain and reference). These groups can have one or more consensus sequences. An expansion arrow can be used to see more detailed information about members of a minimotif group.

Motif filtering

A new set of filters allows users to focus the minimotifs in the output. Since consensus sequences are an interpretation of experimental data, we have tagged minimotifs in the MnM 2 database as consensus sequences or instances. The MnM 2 website now gives the user flexibility in analyzing consensus sequence, instances or both using checkboxes in a pulldown menu. Selecting instances generally increases the stringency of motif prediction by limiting motif predictions to only exact sequences in proteins of known function. Other filter categories include motif activities such as binds, posttranslational modifications, trafficking, etc.

Frequency scoring for species with complete genomes

In the original MnM release, statistics on motifs for 10 proteomes were updated manually. These statistics consist of each motif's probability of occurrence in a proteome, expected count in a proteome, actual count in a proteome and an enrichment factor (computed by dividing the actual count by the expected count). There are now over 6000 genomes that have been sequenced and we have calculated motif frequency statistics for these genomes as previously described (15). The other improvement in this function is that species choice for frequency score is now enforced (see 'Grouping of related motifs' section). Since MnM 2 now contains approximately 5000 minimotifs, we have built an automated script that updates motif statistics when one or more new motifs are added to MnM 2.

Reformatted motif table

The motif table contains information about each motif prediction in a protein query. We have reformatted this table to accommodate new changes in MnM 2 and present motif-related information in a clearer format. Motifs are now presented in groups of related instances and consensus sequences. The minimotif sequence is hyperlinked to a reference for each motif. Annotations use standardized semantics and a set of syntax rules. As in the original version, frequency and surface prediction scores are shown. Evolutionary conservation can be used to rank motifs using the 'View Homologous Proteins' function. An advanced motif table can be selected which provides additional frequency information previously displayed in the original version on MnM.

Aliases

Aliases for protein names are listed in the Protein Details Table.

GROWTH OF MINIMOTIF DATABASE

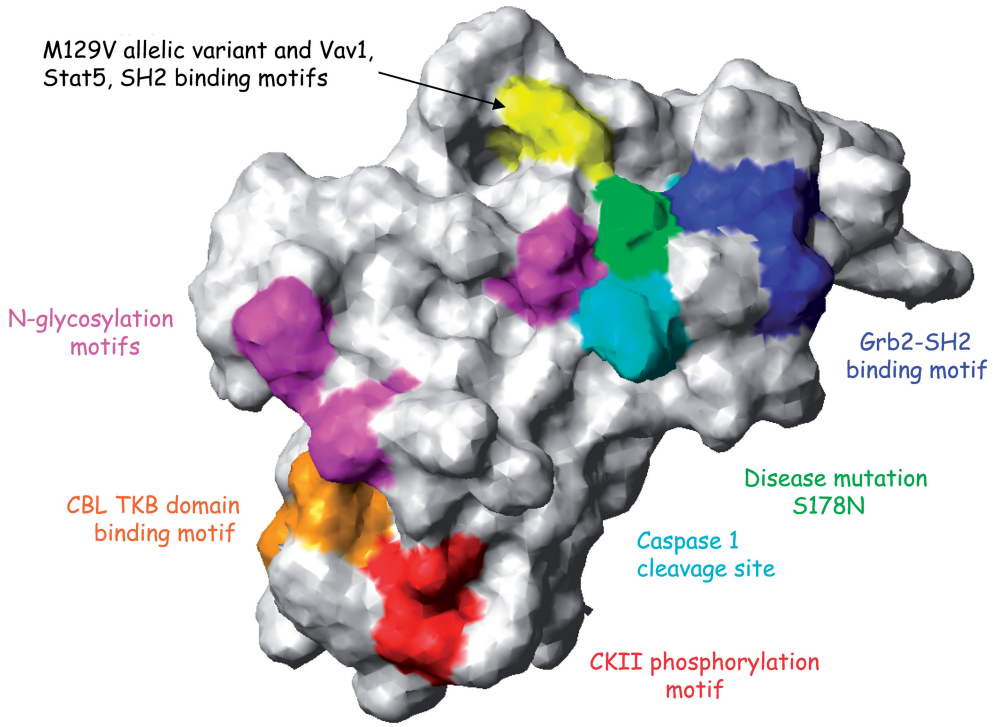
Since the first release of MnM (15), the number of minimotif sequences has grown approximately 11-fold to 5089 sequences (Table 1). The source of these minimotifs has been primary literature with the exception of several hundred minimotifs imported from PDZbase (14). To identify new motifs several sets of keywords were used to search PubMed. Typical words were 'motif', 'peptide', 'site', etc. Papers were read by an expert, who then inserted the minimotif into the database. The majority of the growth was due to new motif entries; however, another reason for the increase in the number of motif entries arises because some previous annotations had motifs that bound to more than one different protein. We now consider a single motif entry to describe a single binding protein.

Complete entries in the first release of MnM had a motif sequence, annotation, identifier, cellular compartment and a reference source. For an entry to be complete in MnM 2, the motif annotation has been replaced with a motif sequence and a corresponding source protein (and accession number), an activity, and a target, which can be a protein, nucleic acid, lipid or other small molecule.

Table 1. Growth of minimotif entries in MnM

Category	MnM	MnM 2
Total		
Motif sequences	462	5089
Consensus sequences	312	858
Instance sequences	44	4229
Post-translational modifications	116	663
Binding	162	4689
Trafficking	34	195
Unique		
Motif sequences	312	2224
Motif proteins	<312	1211
Motif targets	<312	687
References	178	800

A



B

MOTIFS ELIMINATED BY SNP	MOTIFS INTRODUCED BY SNP	POSITION	ANNOTATION
	Y[TLVFIC]	128	This consensus motif in Growth hormone receptor binds the SH2 domain of Stat5; Tyr must be phosphorylated
YM		128	This instance motif in peptide binds the SH2 domain of Vav1; Tyr must be phosphorylated
[FWYL]x[HAT]D[ACFGHILMNSTVWY]		175	This consensus motif in pro-Interleukin 1 beta is proteolyzed by Caspase 1; Cleaves after Asp; no modification required

Figure 1. Minimotifs predicted by MnM 2 painted onto the surface of prion. (**A**) Surface plot created from human prion coordinates in PDB accession number 1QM0. The 129V variant that favors Creutzfeldt–Jakob disease eliminates a potential Vav1 SH2 binding motif (yellow) that is present in the 129M variant that favors fatal familial insomnia. Also, the D178N disease-associated mutation eliminates a potential Caspase 1 cleavage site (green/cyan). cCBL TKB domain binding motif (orange), CKII phosphorylation site (red), N-glycosylation motif (magenta) and Grb2 SH2 binding motif (Blue). Several motifs predicted on the opposing face or on missing fragments of the protein are not shown. (**B**) Output from SNP analysis using MnM 2 website. Minimotifs with red text are eliminated by an SNP and those with green text are introduced by an SNP. SNPs analyzed were M129V and D178V.

For a protein target, support for corresponding information for a target region such as a protein domain has been added. This alteration enables us to integrate motifs, activities and motif targets with other biological databases. Inclusion of data in the MnM database is still based on the requirement that the motif sequence and its activity are published.

We also now designate whether a minimotif sequence is a consensus or an instance of a protein or peptide. The original MnM release contained 312 consensus sequences and has grown approximately 3-fold to 858 sequences and we have started annotation of verified instances of minimotifs which has grown approximately 100-fold to 4229 peptide sequences. Most of the new minimotifs are binding motifs which have grown approximately 29-fold and the numbers of post-translational modifications have grown approximately 5-fold. We have now broken up several previous annotations that were for either multiple minimotif sequence sources or multiple targets into separate entries, artificially inflating the number of entries, but properly segregating information, reducing ambiguities and allowing the database to be mined in new ways. The number of references has grown approximately 5-fold which likely, more accurately reflects the growth of the database over the past 2 years.

EXAMPLE ANALYSIS OF PRION WITH MnM 2

A sample analysis of prion (NP_898902) using MnM 2 is provided to demonstrate how minimotif analysis can be used. Fifty-two potentially functional minimotifs in prion were identified using MnM 2. At least five predicted minimotifs had already been experimentally demonstrated. These include a cryptic nuclear localization, tyrosine phosphorylation, N-glycosylation, and Casein Kinase II and PKC phosphorylation motifs (22–24).

Five additional minimotifs seem to be closely related to known prion functions and might be inferred from published experiments. Mdm2 is involved in prion-induced cell death, but has never been shown to bind prion protein as predicted by the MdM2 binding motif at residues 12–19 (25). A cABL kinase inhibitor inhibits prion signaling and conversion to its pathogenic form, but prion is not known to be phosphorylated by cABL (26). Prion has a potential

cABL phosphorylation site at residue 162. Prion contains several potential Erk phosphorylation sites and activates Erk, but is not known to be phosphorylated by Erk (27). Although the CBL ubiquitination protein is not implicated in prion function, ubiquitination is known to be involved in prion turnover. Thus, the CBL binding motif at residues 147–152 may play a role in prion ubiquitination (28). Prion protein binds the SH3 domain of Grb2, but is not known to bind the SH2 domain of Grb2, although it does contain a Grb2/SH2 domain binding motif (29).

In the case of prion, as well as other nonsynonymous missense mutations in disease, MnM 2 can be used to generate new disease-causing hypotheses. For prion, the D178N mutation is associated with disease. The D178N mutation eliminates a potential Caspase 1 cleavage site. Allelic variation of the 129 position determines whether individuals get Creutzfeldt–Jakob disease (V129) or fatal familial insomnia (M129) (30). These residues are juxtaposed on the protein surface (Figure 1). Several other minimotifs (N-glycosylation and Grb2-SH2 binding) surrounding these residues may be responsive to mutation and/or allelic variation. Furthermore, the M129 variant has a putative Vav1 SH2 and no Stat5 SH2 binding minimotifs; whereas, the presence of these putative motifs is switched in the V129 variant. This MnM analysis suggests that these proteins, through their interaction with prion may be involved in these diseases.

This analysis illustrates how MnM 2 can be used in combination with the known biology of the protein, SNP analysis, and plotting motifs onto the surface of the protein structure to develop new hypothesis for the roles of proteins in disease.

FUNDING

National Institutes of Health (AI078708, GM079689); National Science Foundation (ITR-0326155). Funding for open access charge: GM079689.

Conflict of interest statement. None declared.

REFERENCES

1. Falquet,L., Pagni,M., Bucher,P., Hulo,N., Sigrist,C.J., Hofmann,K. and Bairoch,A. (2002) The PROSITE database, its status in 2002. *Nucleic Acids Res*., **30**, 235–238.
2. de Castro,E., Sigrist,C.J.A., Gattiker,A., Bulliard,V., Langendijk-Genevaux,P.S., Gasteiger,E., Bairoch,A. and Hulo,N. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res*., **34**, W362–W365.
3. Su,Q.J.J., Lu,L., Saxonov,S. and Brutlag,D.L. (2005) eBLOCKs: enumerating conserved protein blocks to achieve maximal sensitivity and specificity. *Nucleic Acids Res*., **33**, D178–D182.
4. Prakash,T., Khandelwal,M., Dasgupta,D., Dash,D. and Brahmachari,S.K. (2004) CoPS: comprehensive peptide signature database. *Bioinformatics*, **20**, 2886–2888.
5. Attwood,T.K., Avison,H., Beck,M.E., Bewley,M., Bleasby,A.J., Brewster,F., Cooper,P., Degtyarenko,K., Geddes,A.J., Flower,D.R. *et al*. (1997) The PRINTS database of protein fingerprints: a novel information resource for computational molecular biology. *J. Chem. Inf. Comput. Sci*., **37**, 417–424.
6. Sammut,S.J., Finn,R.D. and Bateman,A. (2008) Pfam 10 years on: 10 000 families and still growing. *Brief. Bioinform*., **9**, 210–219.
7. Geer,L.Y., Domrachev,M., Lipman,D.J. and Bryant,S.H. (2002) CDART: protein homology by domain architecture. *Genome Res*., **12**, 1619–1623.
8. Henikoff,J.G., Greene,E.A., Pietrokovski,S. and Henikoff,S. (2000) Increased coverage of protein families with the Blocks Database servers. *Nucleic Acids Res*., **28**, 228–230.
9. Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Birney,E., Biswas,M., Bucher,P., Cerutti,T., Corpet,F., Croning,M.D.R. *et al*. (2001) The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res*., **29**, 37–40.
10. Marchler-Bauer,A., Anderson,J.B., DeWeese-Scott,C., Fedorova,N.D., Geer,L.Y., He,S., Hurwitz,D.I., Jackson,J.D., Jacobs,A.R., Lanczycki,C.J. *et al*. (2003) CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res*., **31**, 383–387.
11. Letunic,I., Goodstadt,L., Dickens,N.J., Doerks,T., Schultz,J., Mott,R., Ciccarelli,F., Copley,R.R., Ponting,C.P. and Bork,P. (2002) Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res*., **30**, 242–244.
12. Kreegipuu,A., Blom,N. and Brunak,S. (1999) PhosphoBase, a database of phosphorylation sites: release 2.0. *Nucleic Acids Res*., **27**, 237–239.
13. Rawlings,N.D., Morton,F.R. and Barrett,A.J. (2006) MEROPS: the peptidase database. *Nucleic Acids Res*., **34**, D270–D272.
14. Beuming,T., Skrabanek,L., Niv,M.Y., Mukherjee,P. and Weinstein,H. (2005) PDZBase: a protein-protein interaction database for PDZ-domains. *Bioinformatics*, **21**, 827–828.
15. Balla,S., Thapar,V., Luong,T., Faghri,T., Huang,C.H., Rajasekaran,S., del Campo,J.J., Shin,J.H., Mohler,W.A., Maciejewski,M.W. *et al*. (2006) Minimotif Miner, a tool for investigating protein function. *Nat. Methods*, **3**, 175–177.
16. Schiller,M.R. (2007) Minimotif Miner: a computation tool to investigate protein function, disease, and genetic diversity. In Coligan,J.E., Dunn,B.M., Speicher,D.W. and Winkler,H. (eds), *Current Protocols in Protein Science*, John Wiley & Sons Inc., Hoboken, NJ, pp. 2.12.1–2.12.14.
17. Kadaveru,K., Vyas,J. and Schiller,M.R. (2008) Viral infection and human disease- insights from minimotifs. *Front Biosci*., **13**, 6455–6471.
18. Obenauer,J.C., Cantley,L.C. and Yaffe,M.B. (2003) Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res*., **31**, 3635–3641.
19. Puntervoll,P., Linding,R., Gemund,C., Chabanis-Davidson,S., Mattingsdal,M., Cameron,S., Martin,D.M.A., Ausiello,G., Brannetti,B., Costantini,A. *et al*. (2003) ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res*., **31**, 3625–3630.
20. Rigoutsos,I., Floratos,A., Parida,L. and Platt,D. (2002) Dictionary-driven protein annotation. *Nucleic Acids Res*., **30**, 3901–3916.
21. Wheeler,D.L., Chappey,C., Lash,A.E., Leipe,D.D., Madden,T.L., Schuler,G.D., Tatusova,T.A. and Rapp,B.A. (2000) Database resources of the national center for biotechnology information. *Nucleic Acids Res*., **28**, 10–14.
22. Negro,A., Meggio,F., Bertoli,A., Battistutta,R., Sorgato,M.C. and Pinna,L.A. (2000) Susceptibility of the prion protein to enzymic phosphorylation. *Biochem. Biophys. Res. Comm*., **271**, 337–341.
23. Walmsley,A.R., Zeng,F.N. and Hooper,N.M. (2001) Membrane topology influences N-glycosylation of the prion protein. *EMBO J*, **20**, 703–712.
24. Gu,Y.P., Hinnerwisch,J., Fredricks,R., Kalepu,S., Mishra,R.S. and Singh,N. (2003) Identification of cryptic nuclear localization signals in the prion protein. *Neurobiol. Dis*., **12**, 133–149.
25. Paitel,E., Fahraeus,R. and Checler,F. (2003) Cellular prion protein sensitizes neurons to apoptotic stimuli through Mdm2-regulated and p53-dependent caspase 3-like activation. *J. Biol. Chem*., **278**, 10061–10066.
26. Ertmer,A., Gilch,S., Yun,S.W., Flechsig,E., Klebl,B., Stein-Gerlach,M., Klein,M.A. and Schatzl,H.M. (2004) The tyrosine kinase inhibitor STI571 induces cellular clearance of PrPSc in prion-infected cells. *J. Biol. Chem*., **279**, 41918–41927.

27. Monnet,C., Gavard,J., Mege,R.M. and Sobel,A. (2004) Clustering of cellular prion protein induces ERK1/2 and stathmin phosphorylation in GT1-7 neuronal cells. *FEBS Lett.*, **576**, 114–118.
28. Yedidia,Y., Horonchik,L., Tzaban,S., Yanai,A. and Taraboulos,A. (2001) Proteasomes and ubiquitin are involved in the turnover of the wild-type prion protein. *EMBO J.*, **20**, 5383–5391.
29. Leadbeater,C., McIver,L., Campopiano,D.J., Webster,S.P., Baxter,R.L., Kelly,S.M., Price,N.C., Lysek,D.A., Noble,M.A., Chapman,S.K. *et al.* (2000) Probing the NADPH-binding site of Escherichia coli flavodoxin oxidoreductase. *Biochem. J.*, **352**, 257–266.
30. Goldfarb,L.G., Petersen,R.B., Tabaton,M., Brown,P., LeBlanc,A.C., Montagna,P., Cortelli,P., Julien,J., Vital,C., Pendelbury,W.W. *et al.* (1992) Fatal familial insomnia and familial Creutzfeldt-Jakob disease: disease phenotype determined by a DNA polymorphism. *Science*, **258**, 806–808.