First publ. in: Trends in Genetics 19 (2003), 2, pp. 72-75

affect regulation of the human genome in a more global manner by creating S/MARs that form chromatin loops, and by shaping the sequence evolution of LCRs.

Acknowledgements

Galina V. Glazko was supported by research grants from NIH (GM-20293) and NASA (NCC2-1057) awarded to Masatoshi Nei. We thank Nathan J. Bowen and Wolfgang J. Miller for discussions on the relationship between TEs and S/MARs.

References

1 The Human Genome Sequencing Consortium, (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
2 Kidwell, M.G. and Lisch, D.R. (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution Int. J. Org. Evolution* 55, 1–24
3 Hickey, D.A. (1982) Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics* 101, 519–531
4 Doolittle, W.F. and Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284, 601–603
5 Orgel, L.E. and Crick, F.H. (1980) Selfish DNA: the ultimate parasite. *Nature* 284, 604–607
6 Miller, W.J. *et al.* (1999) Molecular domestication – more than a sporadic episode in evolution. *Genetica* 107, 197–207
7 Makalowski, W. (2000) Genomic scrap yard: how genomes utilize all that junk. *Gene* 259, 61–67
8 Nekrutenko, A. and Li, W.H. (2001) Transposable elements are found in a large number of human protein-coding genes. *Trends Genet.* 17, 619–621
9 Britten, R.J. (1996) DNA sequence insertion and evolutionary variation in gene regulation. *Proc. Natl Acad. Sci. U.S.A.* 93, 9374–9377
10 Brosius, J. (1999) Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica* 107, 209–238
11 Frith, M.C. *et al.* (2002) Statistical significance of clusters of motifs represented by position specific scoring matrices in nucleotide sequences. *Nucleic Acids Res.* 30, 3214–3224
12 Wingender, E. *et al.* (2001) The TRANSFAC system on gene expression regulation. *Nucleic Acids Res.* 29, 281–283
13 Purucker, M. *et al.* (1990) Structure and function of the enhancer 3′ to the human A gamma globin gene. *Nucleic Acids Res.* 18, 7407–7415
14 Li, Q. *et al.* (1999) Locus control regions: coming of age at a decade plus. *Trends Genet.* 15, 403–408
15 Hardison, R. *et al.* (1997) Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene* 205, 73–94
16 Strausberg, R. *et al.* (1999) The mammalian gene collection. *Science* 286, 455–457
17 Bode, J. *et al.* (1996) Scaffold/matrix-attached regions: topological switches with multiple regulatory functions. *Crit. Rev. Eukaryot. Gene Expr.* 6, 115–138
18 Liebich, I. *et al.* (2002) S/MARt DB: a database on scaffold/matrix attached regions. *Nucleic Acids Res.* 30, 372–374
19 van Drunen, C.M. *et al.* (1999) A bipartite sequence element associated with matrix/scaffold attachment regions. *Nucleic Acids Res.* 27, 2924–2930
20 Wyrick, J.J. *et al.* (2001) Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. *Science* 294, 2357–2360
21 Nabirochkin, S. *et al.* (1998) A nuclear matrix/scaffold attachment region co-localizes with the gypsy retrotransposon insulator sequence. *J. Biol. Chem.* 273, 2473–2479
22 Avramova, Z. *et al.* (1998) Matrix attachment regions and structural colinearity in the genomes of two grass species. *Nucleic Acids Res.* 26, 761–767

0168-9525/03/$ - see front matter. Published by Elsevier Science Ltd.
PII: S0168-9525(02)00006-9

An evolutionary basis for scrapie disease: identification of a fish prion mRNA

Eric Rivera-Milla, Claudia A.O. Stuermer and Edward Málaga-Trillo

Department of Biology, University of Konstanz, 78457 Konstanz, Germany

Infectious prion proteins cause neurodegenerative disease in mammals owing to the acquisition of an aberrant conformation. We cloned a *Fugu rubripes* gene that encodes a structurally conserved prion protein, and found rapid rates of molecular divergence among prions from different vertebrate classes, along with molecular stasis within each class. We propose that a directional trend in the evolution of prion sequence motifs associated with pathogenesis and infectivity could account for the origin of scrapie in mammals.

Prion proteins (PrP) are membrane-anchored glycoproteins of unknown function, with the unique ability to change their structure irreversibly from a normal α-helix-rich isoform (PrP^{C}) to a pathological β-sheet-rich isoform known as scrapie (PrP^{Sc}) [1]. This transformation can occur in an autocatalytic manner or with the aid of a hypothetical 'Protein X' [2], leading to the accumulation of insoluble PrP^{Sc} aggregates in the brain. These aggregates cause transmissible spongiform encephalopathies (TSE), a group of lethal, neurodegenerative diseases described only in mammals (e.g. kuru and Creutzfeldt–Jacob in humans, scrapie in sheep, and BSE or 'mad cow' disease in cattle) [3]. Transmission of prion disease between species depends on the degree of sequence similarity at specific amino acids required for interaction between the infectious PrP^{Sc} and the host's PrP^{C} molecules [4,5]. Variability at these sites can create 'host barriers', even between related species [5–7], although infection between distant species can also occur after long exposure times [4]. In fact, human fatalities during the 'mad cow crisis' resulted from the consumption of meat from cows that had been fed dietary supplements contaminated with sheep PrP^{Sc} [8]. Thus, the prion's success in infecting different host species along a food chain is an evolutionary puzzle.

Corresponding author: Edward Málaga-Trillo (edward.malaga@uni-konstanz.de).

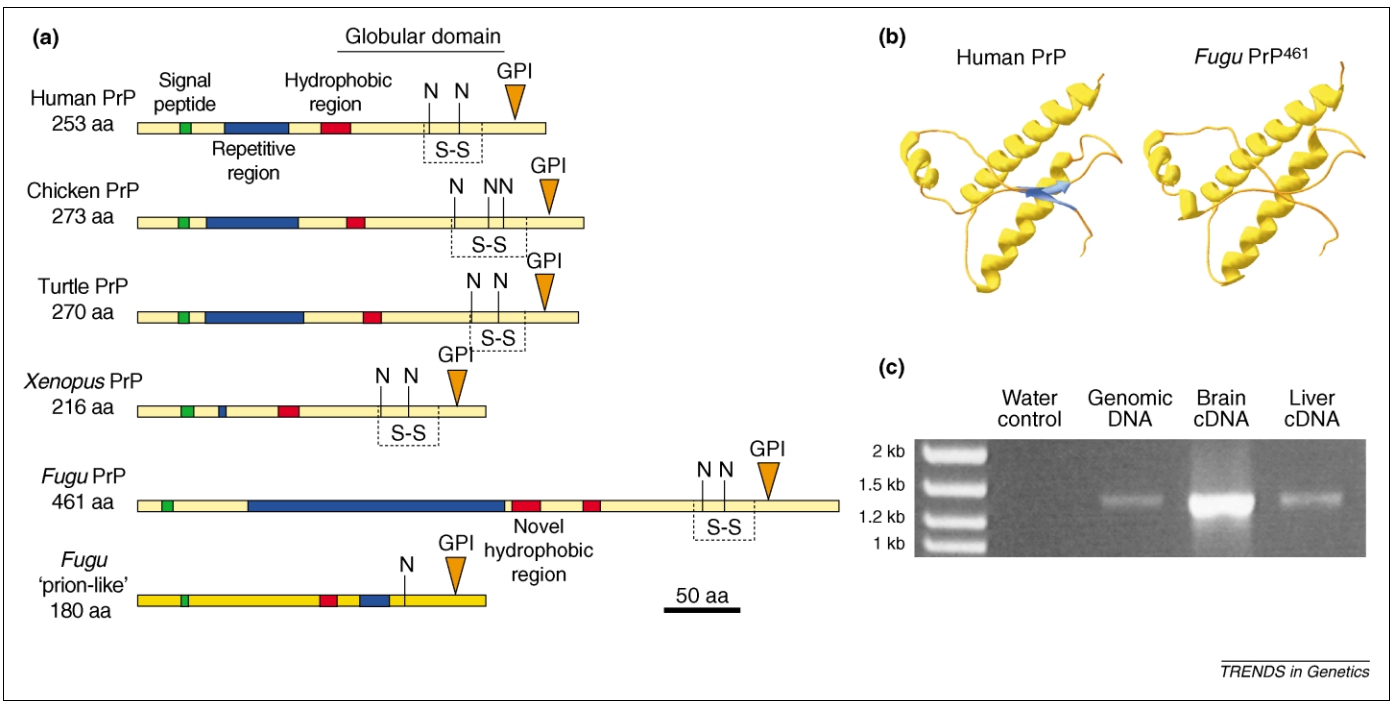

Fig. 1. Characterization of the *Fugu* prion. (a) The location and relative size of conserved structural features in the PrP polypeptide are shown for representative vertebrates, based on the structure reported for mammalian prions; domains are represented by different colored boxes; orange triangles, GPI anchor residues; N, glycosylation sites; S-S disulfide bridges. An additional hydrophobic domain unique to the fish prion is indicated immediately downstream of the repetitive region. Protein domains were predicted using SwissPDB 3.7b2, as well as DAS and HMMTOP methods (http://www.sbc.su.se/~miklos/DAS and http://www.enzim.hu/hmmtop). (b) Ribbon diagram showing the conservation of tertiary structure at the C-terminal region of human and *Fugu* PrPs. Secondary and tertiary protein structures were modeled using the ProModII program at the SWISS-MODEL Automated Protein Modeling Server (http://www.expasy.org/swissmod/SWISS-MODEL.html), based upon hamster (1QLZA.pdb), bovine (1DXOA.pdb), syrian hamster (1B10A.pdb) and mouse (1AG2.pdb) Protein Data Bank (pdb) structure files. (c) PCR amplification of PrP^{461} from *Fugu* genomic and cDNA, using primers Furu-PrPx461-5′for: 5′-CAT GAT GGG GAG GTT GTG TG-3′ and Furu-PrPx461-3′rev: 5′-GGG GCT GAT GGA TGT CAC T-3′, at an annealing temperature of 60°C. Transcripts were additionally confirmed by RACE PCR (not shown).

A novel fish prion

Genes encoding homologous prion proteins have been reported in all tetrapod groups, but not in 'lower' vertebrates or invertebrates [9]. To gain an evolutionary perspective on the origin, structure and function of prions, we searched genome databases for related loci. Using low stringency searches, we identified a 332-kilobase (kb) contig in the genome of the Japanese pufferfish *Fugu rubripes* (JGI_Scaffold_96, v3.0), containing a prion-like sequence. After correcting a sequencing mistake in the database record, which split the gene over two different reading frames, we confirmed that a single mRNA is transcribed from this locus by PCR amplifying and sequencing the corresponding open reading frame (ORF) from brain and liver cDNA (Fig. 1c).

We also identified a 1.3-kb partial sequence (FS_contig_38492) in another pufferfish genome (*Tetraodon nigroviridis*). The two DNA sequences are 85% identical, and they constitute a novel fish PrP locus present in at least two genera of the family Tetraodontidae (pufferfish). We could not find any sequence or structural homologues in invertebrate databases, but we did identify a partial genomic sequence in the Urochordate *Ciona intestinalis* (Sea squirt, JGI_clone_LQW13995) that is very similar to known prion N termini, suggesting that the PrP gene probably originated at least at the base of the chordate lineage. Phylogenetic analysis of the globular domain in representative vertebrates shows marked clustering of the sequences according to taxonomic classes, where *Fugu* PrP is the most divergent and basal to all other sequences (Fig. 2). Within avian and mammalian classes, the globular domains are highly conserved (>90%) and apparently subject to strong negative selection (low K_a/K_s ratios of ~0.1).

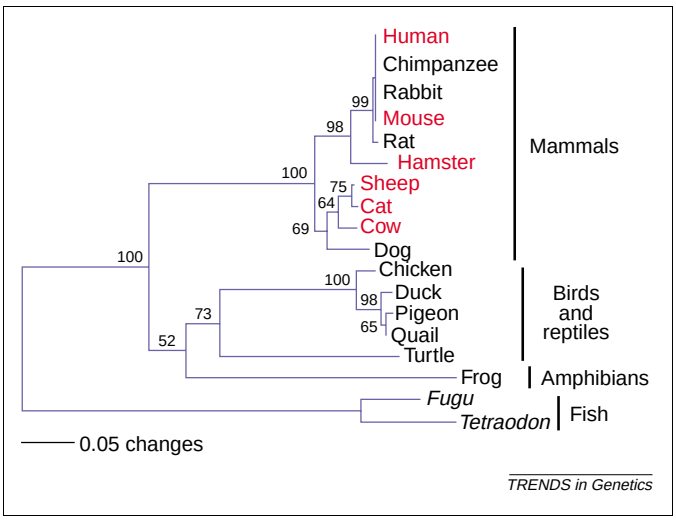

Fig. 2. Phylogenetic relationships among vertebrate prions were examined using parsimony and distance methods in PAUP v4.0d64* [19] and MEGA [20]. The neighbor-joining tree is based on genetic distances between amino acid sequences of globular domains. Numbers at the internodes are bootstrap confidence values (1000 replications). The horizontal scale bar indicates genetic distance. Taxa known to develop prion disease are shown in red. The sequence of *Fugu* PrP^{461} cDNA was deposited in GenBank (accession number AF531159). For the original *Fugu* contig sequences see http://genome.jgi-psf.org/fugu6/fugu6.home.html and [21]. All other prion-related sequences were identified at http://aluminum.jgi-psf.org/prod/bin/blast.fugu6.cgi, http://www.genoscope.cns.fr/proxy/cgi-bin/recherche_kicy.cgi, and http://www.jgi.doe.gov/programs/ciona/ciona_mainpage.html.

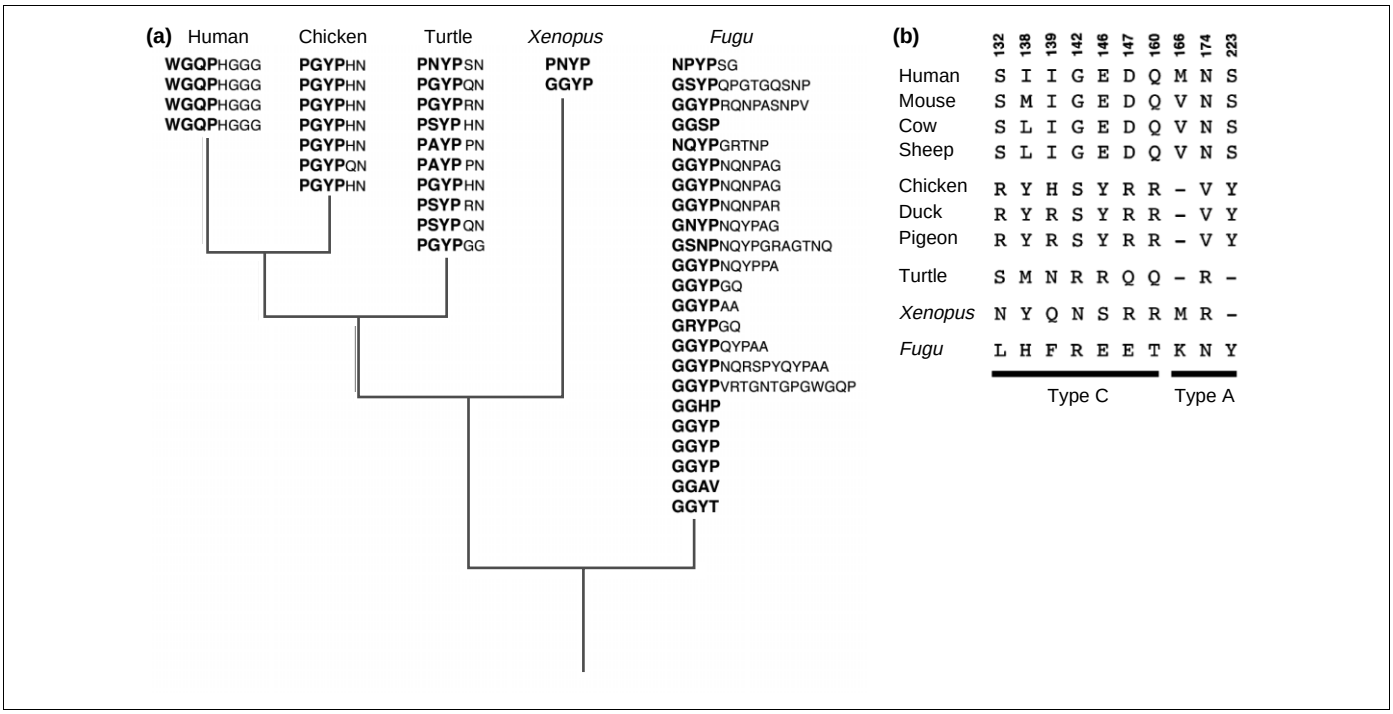

Fig. 3. Molecular evolutionary trends in prion domains. (a) Phylogram indicating the evolution in length, number and sequence variability of the N-terminal amino acid repeats in representative vertebrate prions. A comparative alignment of the internal repeats within a single prion sequence is shown for each individual species. The basic tetrapeptide unit shared by all vertebrate classes is highlighted in bold. (b) Patterns of amino acid variability at selected A and C type residues. Numbering is based on the mouse sequence. Dashes indicate deleted sites.

Sequence versus 3D structural similarities

Although 3D structure data are available only for some mammalian prion proteins [9], analysis of primary amino acid sequences indicates strong structural conservation within mammalian and avian classes alike (Fig. 1a). Each PrP has a highly flexible N terminus [10] and a stable C terminus with the potentially pathogenic globular domain [11]. Paradoxically, inter-group comparisons among mammalian, avian, reptile and amphibian polypeptides reveal a sharp decrease in sequence identity, without loss of structural similarity. Therefore, homology between distantly related prion proteins cannot be proposed based solely on sequence similarity, without the use of a minimum set of structural landmarks. Accordingly, although the 461 aa fish prion polypeptide (*Fugu* PrP^{461}) reported here is only 22% similar in primary sequence to mammalian prions, it contains the characteristic N-terminal repetitive stretch, a hydrophobic motif and a C-terminal globular domain (Fig. 1b). Moreover, 3D structural modelling predicts strong conservation of tertiary structure between *Fugu* PrP^{461} and human PrP (Fig. 1b).

Recently Suzuki *et al.* reported a 'prion-like protein' of only 180 aa in *Fugu* (PrP^{180}), which shows neither primary sequence nor structural similarity to any known prions [12]. Although it possesses a hydrophobic motif and an unrelated repetitive stretch, these elements are not in the same order as other PrPs (Fig. 1a) and it lacks the expected α-helix-rich globular domain at the C terminus. Thus, structural evidence points to *Fugu* PrP^{461}, and not PrP^{180}, as the evolutionary homologue of tetrapod prion proteins. Because the genomic environments around both *Fugu* PrP^{461} and PrP^{180} seem to have undergone multiple chromosomal rearrangements (not shown), we could not establish consistent orthologous relationships between their corresponding scaffolds and any mammalian syntenic groups. Neither could we detect linkage of fish PrP to a second prion gene (PRND), as occurs in mammals.

Molecular evolution and pathogenesis

Because the main structural domains of PrP genes are well conserved, it is possible to infer some of the changes that took place between lineages, despite the divergence between the primary sequences. These changes can, in turn, reveal functional evolutionary trends related to certain aspects of prion pathogenesis. For instance, the repetitive N-terminal domain within each vertebrate class contains a distinctive default number of degenerate repeats that share the same tetrapeptide as a basic unit (Fig. 3a). From fish to human, the repeat units within one molecule have reduced in degeneracy, possibly by gene conversion and homogenization, and have increased their size in discrete steps, reaching a maximum of eight amino acids in mammals (Fig. 3a). These mammalian 'octarepeats' (actually, two adjacent tetrapeptides) are crucial for self-assembly and aggregation of scrapie proteins, and they provide normal prions with the ability to bind copper ions [10,13–15]. Interestingly, the N-terminal domain of *Fugu* PrP^{461} is exceptionally rich in repeats (23 in total), with high levels of sequence degeneracy and ranging in size from 4 to 17 aa. Although the precise aggregation and copper-binding properties of non-mammalian PrPs remain to be examined, it is noteworthy that scrapie pathogenicity, which depends largely on the prion's ability to self-aggregate [16–18], has been observed only in animals with octarepeats at their PrP^{C} N termini (i.e. mammals).

Comparative analysis of the more conserved C-terminal

domain is equally informative. 'A-type' and 'C-type' protein–protein interaction sites have been implicated in the transformation of prions across species by mediating binding of PrP^{C} to Protein X or PrP^{Sc}, respectively [2–7]. Of nine A-type residues, only one (position 215) is conserved in all vertebrate groups, whereas the remaining eight show variability patterns that generally correlate with vertebrate class groupings (Fig. 3b). Likewise, of all C-type residues (position 96–167), the first half are conserved with exception of *Fugu* and the other half are specific for each vertebrate group (Fig. 3b). Thus, PrP^{Sc}s could fail to convert PrP^{C}s across vertebrate classes because of molecular incompatibility at specific contact sites, regardless of self-aggregation properties encoded at the N-terminal domain.

Together, the evidence presented here links discrete patterns of prion molecular evolution with important changes in their pathogenic properties. Particularly, changes at the N- and C-terminal domains could help explain why scrapie pathogenesis and transmission seem exclusive to mammals. Our discovery of a novel prion gene in fish, and possibly in a Urochordate, places the origin of prions in a common ancestor of all vertebrates, at least 550 million years ago. Moreover, our comparative analysis of PrP^{C} amino acid sequences reveals rapid rates of molecular evolution at the base of the vertebrate radiation without significant losses in protein structure, followed by a reduction in the effective substitution rates within each vertebrate class. The implications of these findings can now be tested experimentally.

Acknowledgements

E.R.-M. is a DAAD fellow. This work was supported by grants from the University of Konstanz AFF to E.M.-T. and from TSE, MWK, BW and FCI to C.A.O.S. We are grateful to the Joint Genome Institute, UK-HGMP Resource Center, Genoscope and the Whitehead Institute for Genome Research. This article is dedicated to our former colleague Joseph J.B. Jayabalan.

References

1 Prusiner, S.B. (1991) Molecular biology of prion diseases. *Science* 252, 1515–1522
2 Telling, G.C. *et al.* (1995) Prion propagation in mice expressing human and chimeric PrP transgenes implicates the interaction of cellular PrP with another protein. *Cell* 6, 79–90
3 Prusiner, S.B. (1998) Prions. *Proc. Natl Acad. Sci. U.S.A.* 95, 13363–13383
4 Billeter, M. *et al.* (1997) Prion protein NMR structure and species barrier for prion diseases. *Proc. Natl Acad. Sci. U.S.A.* 94, 7281–7285
5 Kocisko, D.A. *et al.* (1995) Species specificity in the cell-free conversion of prion protein to protease-resistant forms: a model for the scrapie species barrier. *Proc. Natl Acad. Sci. U.S.A.* 92, 3923–3927
6 Telling, G.C. *et al.* (1994) Transmission of Creutzfeldt–Jakob disease from humans to transgenic mice expressing chimeric human–mouse prion protein. *Proc. Natl Acad. Sci. U.S.A.* 91, 9936–9940
7 Horiuchi, M. *et al.* (2000) Interaction between heterologous forms of prion protein: binding, inhibition of conversion, and species barrier. *Proc. Natl Acad. Sci. U.S.A.* 97, 5836–5841
8 Aguzzi, A. *et al.* (2001) Prions: health scare and biological challenge. *Nat. Rev. Mol. Cell Biol.* 2, 118–126
9 Strumbo, B. *et al.* (2001) Molecular cloning of the cDNA coding for *Xenopus laevis* prion protein. *FEBS Lett.* 508, 170–174
10 Donne, D.G. *et al.* (1997) Structure of the recombinant full length hamster prion protein PrP(29–231): the N terminus is highly flexible. *Proc. Natl Acad. Sci. U.S.A.* 94, 13452–13457
11 Riek, R. *et al.* (1996) Prion protein NMR structure and familial human spongiform encephalopathies. *Proc. Natl Acad. Sci. U.S.A.* 95, 11667–11672
12 Suzuki, T. *et al.* (2002) cDNA sequence and tissue expression of *Fugu rubripes* prion-like protein: a candidate for the teleost orthologue of tetrapod PrPs. *Biophys. Res. Commun.* 294, 912–917
13 Murphy, R.M. (2002) Peptide aggregation in neurodegenerative disease. *Annu. Rev. Biomed. Eng.* 4, 155–174
14 Viles, J.H. *et al.* (1999) Copper binding to the prion protein: structural implications four identical cooperative binding sites. *Proc. Natl Acad. Sci. U.S.A.* 96, 2042–2047
15 Burns, C.S. *et al.* (2002) Molecular features of the copper binding sites in the octarepeats domain of the prion protein. *Biochemistry* 41, 3991–4001
16 Gazit, E. (2002) Global analysis of tandem aromatic octapeptide repeats: the significance of the aromatic-glycine motif. *Bioinformatics* 18, 880–883
17 Bamborough, P. *et al.* (1996) Prion ptotein structure and scrapie replication: theoretical, spectroscopic, and genetic investigation. *Cold Spring Harbor Symp. Quant. Biol.* 61, 495–509
18 Gazit, E. (2002) A possible role for π-stacking in the self-assembly of amyloid fibrils. *FASEB J.* 16, 77–83
19 Swofford, D.L. (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods) (Ver. 4)*, Sinauer Associates
20 Kumar, S. *et al.* (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17, 1244–1245
21 Aparicio, S. *et al.* (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297, 1301–1310

PII: S0168-9525(02)00032-X

Functional determinants of transcription factors in *Escherichia coli*: protein families and binding sites

M. Madan Babu and Sarah A. Teichmann

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

DNA-binding transcription factors regulate the expression of genes near to where they bind. These factors can be activators or repressors of transcription, or both. Thus, a fundamental question is what determines whether a transcription factor acts as an activator or a repressor? Previous research into this question found that a protein's regulatory function is determined by one or more of the following factors: protein–protein contacts, position of the DNA-binding domain in the protein primary sequence, altered DNA structure, and

Corresponding author: M. Madan Babu (madanm@mrc-lmb.cam.ac.uk).