Chimia 55 (2001) 281–285

ISSN 0009–4293

Protein Folding and Assembly

Rudi Glockshuber*

Abstract: One of the central dogmas in biochemistry is the view that the biologically active, three-dimensional structure of a protein is unique and exclusively determined by its amino acid sequence, and that the active conformation of a protein represents its state of lowest free energy in aqueous solution. Despite a large number of novel experiments supporting this view, including an exponentially increasing number of solved three-dimensional protein structures, it is still impossible to predict the tertiary structure of a protein from knowledge of its amino acid sequence alone.

Towards the goal of identifying general principles underlying the mechanism of protein folding *in vitro* and *in vivo*, we are pursuing several projects that are briefly described in this article: (1) Circular permutation of proteins as a tool to study protein folding, (2) Catalysis of disulfide bond formation during protein folding, (3) Assembly of adhesive type 1 pili from *Escherichia coli* strains, and (4) Structure and folding of the mammalian prion protein.

Keywords: Catalysis of disulfide bond formation · Prion proteins · Protein assembly · Protein folding · Type 1 pili

During the last 20 years, the field of protein folding has gained enormous general attention in the life sciences. First, this is because large quantities of biologically active, recombinant proteins are required for both modern diagnostic and therapeutic medicine and structural biology. As many recombinant proteins can only be produced as unfolded protein aggregates (inclusion bodies) and thus have to be refolded *in vitro*, there is a strong need for the development of functional protein expression systems and protein refolding protocols. Besides these practical aspects, numerous human diseases, including Alzheimer's disease and prion diseases, are associated with protein misfolding and formation of amyloid aggregates. Therefore, investigations on the mechanisms underlying protein folding and protein misfolding are of key interest to many areas in biology and medicine. The most important recent advances in the field of protein folding [1] involve i) the view that multiple pathways are available for a polypeptide chain to fold to its native conformation, ii) the discovery of molecular chaperones which assist the protein folding process by preventing irreversible aggregation of unfolded and partially folded polypeptide chains, and iii) the prion hypothesis, suggesting the existence of self-propagating, exclusively proteinaceous infectious agents. Our group is addressing basic questions on the mechanism of protein folding and assembly within the following research projects:

*Correspondence: Prof. Dr. R. Glockshuber
Institut für Molekularbiologie und Biophysik
Eidgenössische Technische Hochschule
Hönggerberg
CH-8093 Zürich
Tel.: +41 1 633 68 19
Fax: +41 1 633 10 36
E-Mail: rudi@mol.biol.ethz.ch

1. Circular Permutation of Proteins: A Tool to Study Protein Folding

One of the key questions about protein folding is whether polypeptide chains require unique nucleation sites to fold to the native state. To identify such nucleation sites, we are applying circular permutation, *i.e.* connection of the natural termini of a protein by a linker peptide and introduction of new termini by cleavage of the polypeptide chain at a different site. The rationale behind this approach is that folding nuclei formed by sequentially adjacent polypeptide segments may no longer be formed when the corresponding segments are far apart in a circularly permuted protein variant.

Using genetic engineering in conjunction with screening or selection procedures at the bacterial colony level, we have performed the first systematic random circular permutation analysis of a protein using the periplasmic disulfide oxidoreductase DsbA from *E. coli* as a model system [2]. DsbA is a monomeric two-domain protein of 189 residues with known three-dimensional structure. After linkage of the natural termini, each regular secondary structure and each loop region of the protein was disrupted by introduction of new termini. Functional analysis of about 70 different permuted variants showed that only about 30% of the polypeptide chain of DsbA may not be disrupted by new termini without loss of folding competence (Fig. 1a). Surprisingly, novel termini were even tolerated within many regular secondary structure elements [2]. All purified, biologically active permuted DsbA variants showed the same tertiary structure and spectroscopic properties as the wild type protein. In contrast, all permuted variants with new termini within the 'forbidden' regions, comprising four α-helices of DsbA (Fig. 1a), proved to be catalytically inactive and showed a significantly lowered secondary structure content compared to the wild type. This demonstrates that random circular permutation can be used as a general method to identify segments in a protein that are essential for folding and stability. As some of the circularly permuted DsbA variants that we have generated are catalytically more active than the wild type protein [2], circu-

lar permutation may also become a useful tool for generating proteins with improved functions.

In addition, we have applied a series of rational circular permutation experiments to study the folding of the green fluorescent protein (GFP) from *Aeqora victoria*. This monomeric 238 residue protein has the unique property that it spontaneously forms a *p*-hydroxybenzylidene-imidazolidone chromophore during folding through main chain cyclization and subsequent air oxidation [3]. In the three-dimensional structure of native GFP, the chromophore is located within an 11-stranded β-barrel (Fig. 1b) and solvent inaccessible. All these properties have made GFP one of the most important reporter molecules available for studying biological processes such as transcription, protein targeting and protein/ligand interactions inside the living cell [3]. Analysis of 20 different circularly permuted GFP variants produced in *E. coli* showed that GFP is much less tolerant toward circular permutations than DsbA [4]. Specifically, any disruption of a strand from the β-barrel by introduction of new termini causes folding incompetence. These data indicate that the 11-stranded β-barrel of GFP represents an extremely cooperative unit which requires all interactions for stability (Fig. 1c).

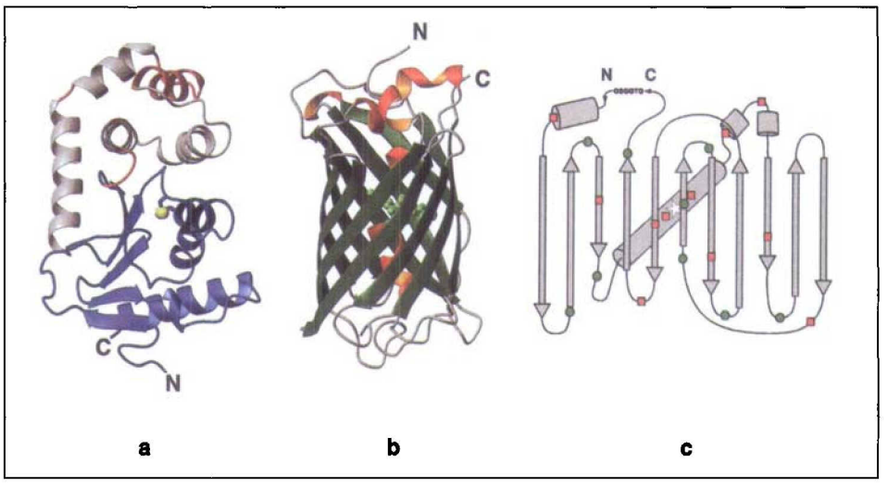

Fig. 1: Results of circular permutation studies on the disulfide oxidoreductase DsbA from *E. coli* (a) and the green fluorescent protein (GFP) from *A. victoria* (b). (a) The disulfide oxidoreductase DsbA from *E. coli* is required for disulfide bond formation in the bacterial periplasm. The monomeric protein consists of 189 residues. The ribbon diagram shows that DsbA is composed of a thioredoxin-like domain (blue) and an α-helical domain (top, red) which is inserted into the thioredoxin motif. The sulphur atoms of the active-site disulfide bond are indicated in yellow. Random circular permutation of DsbA revealed that four α-helices (grey) may not be disrupted by introduction of new termini in circularly permuted variants without loss of folding competence. The natural termini were connected by a 5-residue linker peptide in all permuted variants. (b) Ribbon representation of the three-dimensional structure of GFP (238 residues, monomer). The chromophore in the centre of the 11-stranded β-barrel is shown in a ball-and-stick representation. (c) Rational circular permutation of GFP. The sequential arrangement of regular secondary structures is shown, as well as the hexapeptide linker used to connect the natural termini. The red and green symbols indicate the positions of newly introduced termini in permuted GFP variants. Green circles: variants which retained folding competence during expression in *E. coli*. Red squares: variants which were either proteolytically degraded in the *E.coli* cytoplasm, or soluble and nonfluorescent, or insoluble in the cytoplasm.

2. Catalysis of Oxidative Protein Folding by Bacterial Disulfide Oxidoreductases

Disulfide bond formation generally constitutes the rate-limiting step during folding of secretory proteins. The reason is that a disulfide bond cannot form automatically when two cysteine residues come close during the folding process, because disulfide bond formation is an oxidation process which requires an external oxidant. In addition to the formation of disulfide bonds, the rearrangement of nonnative disulfide bonds, which are particularly likely to be formed in proteins with multiple disulfide bonds, may also limit the rate of folding. In the living cell, the efficient formation of disulfide bonds and the rapid isomerization of incorrect disulfide bonds is catalysed by enzymes belonging to the disulfide oxidoreductase family. These proteins share the thioredoxin fold and possess a reactive disulfide bond with the consensus sequence Cys-Xaa-Xaa-Cys. Fig. 2 summarizes the present knowledge on the catalysis of disulfide bond formation in the periplasm of *E. coli*. Specifically, different enzymes are responsible for oxidation of folding proteins and isomerization of disulfide bonds, *i.e.* DsbA and DsbC/DsbG, respectively [5]. We are characterizing the catalytic properties of bacterial members of this enzyme family, in particular those of DsbA, an enzyme with extraordinary biophysical properties. DsbA is a very strong oxidant, undergoes the fastest disulfide exchange reactions known so far, and randomly transfers its reactive disulfide bond to folding polypeptides [6][7]. We have performed extensive mutagenesis experiments on DsbA to investigate the molecular basis of its extreme reactivity. Specifically, we found that the Xaa-Xaa dipeptide within the active-site sequence strongly influences the redox potential of the enzyme. Replacement of this sequence by the Xaa-Xaa sequences of more reducing members of the thioredoxin family yields more reducing DsbA variants [8]. Analogous results were obtained with Xaa-Xaa variants of cytoplasmic thioredoxin, the most reducing member of the enzyme family [9]. By using these variants of thioredoxin and DsbA we found that the rates by which these enzymes undergo disulfide exchange reactions with their substrate proteins *in vivo* determines their catalytic function, rather than the recycling of their reactive redox form [10][11]. We are presently extending our mechanistic studies on the function of DsbC and are working on bacterial expression systems that provide high yields of correctly folded proteins with disulfide bonds in the periplasm through coexpression of DsbA and DsbC.

3. Assembly of Type 1 Pili from *Escherichia coli*

Type 1 pili of *Escherichia coli* are large, heterooligomeric protein complexes anchored to the outer bacterial membrane and required for the attachment of the bacteria to host cell surfaces. The pili are critical virulence factors of uropathogenic *E. coli* strains and mediate the long-term survival of the bacteria within host cells. Type 1 pili are thus targets for the development of new antibiotics against infections of the urinary tract. Such drugs may either prevent pilus assembly or abolish their ability to bind to host cell surfaces [12].

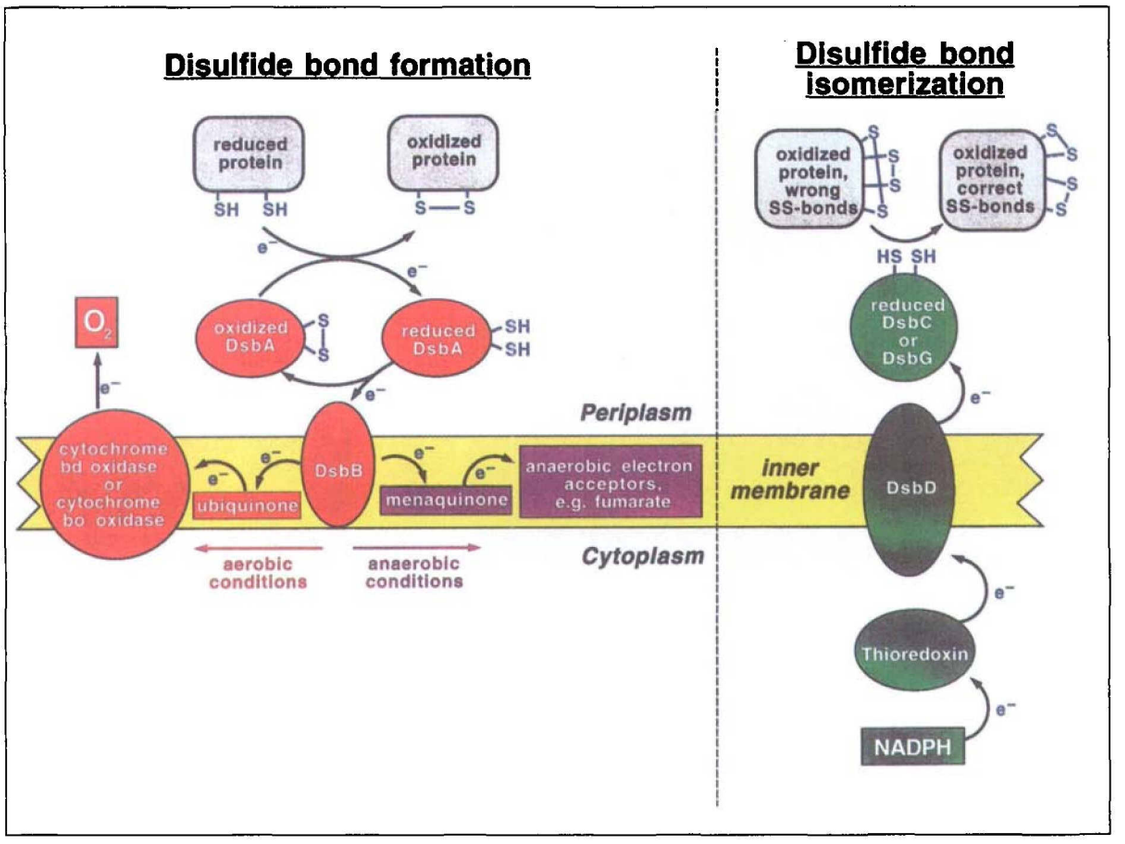

Fig. 2: Catalysis of disulfide bond formation and disulfide bond isomerization during protein folding in the periplasm of *E. coli*. Left: The disulfide oxidoreductase DsbA randomly donates its reactive disulfide bond to folding polypeptides in an extremely rapid reaction and becomes recycled as an oxidant by disulfide exchange with the inner membrane protein DsbB. DsbB in turn becomes reoxidised either by molecular oxygen under aerobic conditions, or by alternate electron acceptors under anaerobic conditions. Right: nonnative disulfide bonds in proteins with multiple disulfide bonds in the native state, introduced by DsbA, are rearranged to the native disulfide conformation by the enzymes DsbC and DsbG. To be catalytically active, DsbC and DsbG must be kept in the reduced state in the periplasm. This is accomplished by a specific electron transport chain from NADPH to thioredoxin in the cytoplasm, and from thioredoxin through the inner membrane protein DsbB to DsbC and DsbG.

A single *E. coli* cell may bear up to 500 type 1 pili, each being 7 nm wide and 0.5 to 2 μm long. Each pilus consists exclusively of protein, with an average number of about 1000 subunits per pilus. About 98% of the pilus subunits are comprised of the main structural pilus subunit FimA, while the residual 2% of the subunits are composed of three hologous subunits, namely FimF, FimG, and the mannose-binding subunit FimH (Fig. 3a). Assembly of the pilus *in vivo* requires the assembly factor FimC, a soluble, periplasmic chaperone protein that is not a component of the pilus. In collaboration with the group of Prof. K. Wüthrich (ETH Zürich), we have solved the three-dimensional structure of FimC in solution by NMR spectroscopy [13]. FimC is a monomeric, 205-residue protein that consists of two domains with an immunoglobulin-like fold. FimC forms stoichiometric complexes with the individual subunits and delivers them to the growing pilus which is anchored to the outer membrane *via* the assembly platform protein FimD.

Using the 51 kDa complex between FimC and the intact, 28 kDa mannose-binding subunit FimH as a model for pilus chaperone/subunit interactions, we have characterized the complete binding surface of FimC for FimH with ^{15}N and ^{1}H chemical shift mapping by applying transverse-relaxation-optimised NMR spectroscopy (TROSY). The binding surface for FimH is almost entirely formed by residues from the N-terminal FimC domain and is much larger than the peptide binding surface observed in the PapD/peptide complexes. This large contact area indicates that FimC, unlike other chaperones, specifically binds to folded forms of the pilus subunits [14][15]. This was confirmed by the elucidation of the X-ray structure of the FimC/FimH complex (Fig. 3b) [16]. We propose that the main function of FimC is the prevention of premature pilus assembly in the periplasm and the correct delivery of the subunits to FimD in the outer membrane. We are presently investigating the folding and stability of individual type 1 pilus subunits in the presence and absence of FimC, the reconstitution of type 1 pili *in vitro*, and the influence of the outer membrane protein FimD on dissociation of FimC/subunit complexes, the last step before subunit incorporation into the growing pilus.

4. Folding and Three-dimensional Structure of the Recombinant Murine Prion Protein

According to the 'protein only' hypothesis, mammalian prion diseases are believed to be caused by PrP^{Sc}, the abnormal, oligomeric form of the monomeric cellular prion protein of the host, PrP^{C} [17]. Many experimental data indicate that PrP^{C} and PrP^{Sc} have identical covalent structures and differ exclusively in their tertiary structures and oligomeriza-

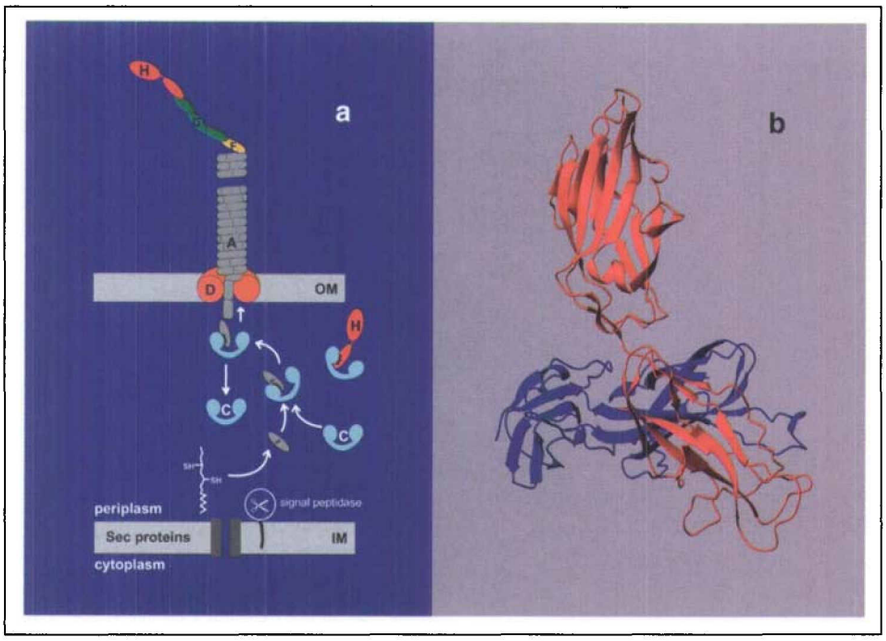

Fig. 3: Chaperone-mediated type 1 pilus assembly in *E. coli*. (a) Model of type 1 pilus assembly: the pilus is anchored to the outer membrane protein FimD, which also forms a pore in the outer membrane allowing translocation of folded pilus subunits. Type 1 pili are 1–2 μm long and consist of up to 3000 protein subunits. These are the main structural pilus subunit FimA (grey, 98% of all subunits) and the minor subunits FimF (yellow), FimG (green) and the mannose-binding adhesin FimH (red) at the pilus tip. FimA, FimF, FimG, and the C-terminal domain of FimH share a pilin domain fold and are supposed to be recognized similarly and specifically by the pilus chaperone FimC (light blue). (b) Ribbon representation of the X-ray structure of the complex between the type 1 pilus chaperone FimC (blue) and the type 1 pilus adhesin FimH (red) [16]. FimC only interacts with the C-teminal pilin domain of FimH.

tion states. PrP^C is a 210-residue cell surface glycoprotein which is attached to the cell membrane by a C-terminal glycosylphosphatidyl inositol (GPI) anchor. PrP^{Sc} is believed to propagate by forcing PrP^C from noninfected cells to adopt its own conformation in an unknown mechanism [17].

In 1996, we found that the recombinant carboxy-terminal domain of the mouse prion protein (residues 121-231) is an autonomous folding unit. In contrast to model predictions, its three-dimensional structure, which was solved in collaboration with the group of Prof. K. Wüthrich in our institute, contains a two-stranded β-sheet and three α-helices (Fig. 4) [18][19]. In collaboration with Prof. K. Wüthrich, we have also solved the three-dimensional solution structure of the recombinant full-length prion protein from the mouse, PrP(23-231). Its entire N-terminal segment 23-125 proved to be flexibly disordered in solution, while the C-terminal residues 126-231 adopt the same structure as in the isolated C-terminal domain PrP(121-231). All structural data and the monomeric state of recombinant PrP(23-231) indicate that the tertiary structure of the recombinant protein is identical to that of natural PrP^C. As the segment 90-231 is protease-resistant in the oligomeric scrapie form PrP^{Sc}, the structure of PrP(23-231) suggests that the minimal structural change that occurs during the conversion of the cellular prion protein to the scrapie form is that the segment 90-125 becomes structured or buried in PrP^{Sc} [20].

Further studies on the pH-dependence of denaturant-induced unfolding transitions of PrP(121-231) revealed that the folding mechanism of the protein changes from a two-state transition at physiological pH to a three-state transition at acidic pH. Specifically, PrP(121-231) forms an unfolding intermediate at pH 4.0 which shows strongly increased β-sheet content, reminiscent of PrP^{Sc}. As PrP^{Sc} is known to accumulate in endosomes of prion-infected cells where acidic pH values are prevalent, the acid-induced unfolding intermediate of the prion protein may represent a first intermediate state in the pathway of PrP^{Sc} formation [21].

One of the strongest arguments in favour of the 'protein only' hypothesis is the fact that all forms of inherited human prion diseases known to date are linked with mutations in the gene encoding the human prion protein (Fig. 4b). We have investigated the influence of the point mutations within the structured C-terminal domain of the recombinant prion protein to test whether these mutation destabilize PrP^C and thus possibly facilitate the spontaneous formation of prions in inherited human prion diseases [22][23]. We found that some, but not all of the amino acid replacements are destabilizing (Fig. 4b) [23]. This indicates that additional mechanisms may underlie the generation of prions in certain familial prion diseases, *e.g.* faster kinetics of PrP^{Sc} formation or higher stability of PrP^{Sc}. Importantly, the stability of the PrP(121-231) variants also did not correlate with the corresponding disease phenotype, which may be the Gerstmann-Sträussler-Scheincker Syndrome (GSS), familial Creutzfeldt-Jakob disease (CJD) or fatal familial insomnia (FFI) (Fig. 4) [23].

Finally, we have found that the C-terminal domain of the prion protein, PrP(121-231), undergoes one of the fastest protein folding reactions known so far and folds without kinetic intermediates. The deduced rate constant of folding of PrP(121-231) at pH 7.0 and 4 °C is 170 μs [24]. This almost certainly excludes one of the proposed models for prion amyloid formation, namely the recruitment of kinetic PrP folding intermediates as precursors of PrP^{Sc}.

Received: March 9, 2001

[1] S.E. Radford, *Trends Biochem Sci.* **2000**, *25*, 611.
[2] J. Hennecke, P. Sebbel., R. Glockshuber *J. Mol. Biol.* **1999**, *286*, 1197.
[3] R.Y. Tsien, *Annu. Rev. Biochem.* **1998**, *67*, 509.
[4] S. Topell, J. Hennecke, R. Glockshuber, *FEBS Lett* **1999**, *457*, 283.
[5] R. Glockshuber, *Nature* **1999**, *402*, 30.
[6] M. Wunderlich, A. Otto, R. Seckler, R. Glockshuber, *Biochemistry* **1993**, *32*, 12251.
[7] M. Wunderlich, A. Otto, K. Maskos, M. Mücke, R. Seckler, R. Glockshuber, *J. Mol. Biol.* **1995**, *247*, 28.
[8] M. Huber-Wunderlich, R. Glockshuber, *Folding & Design* **1998**, *3*, 161.
[9] E. Mössner, M. Huber-Wunderlich, R. Glockshuber, *Protein Sci.* **1998**, *7*, 1233.

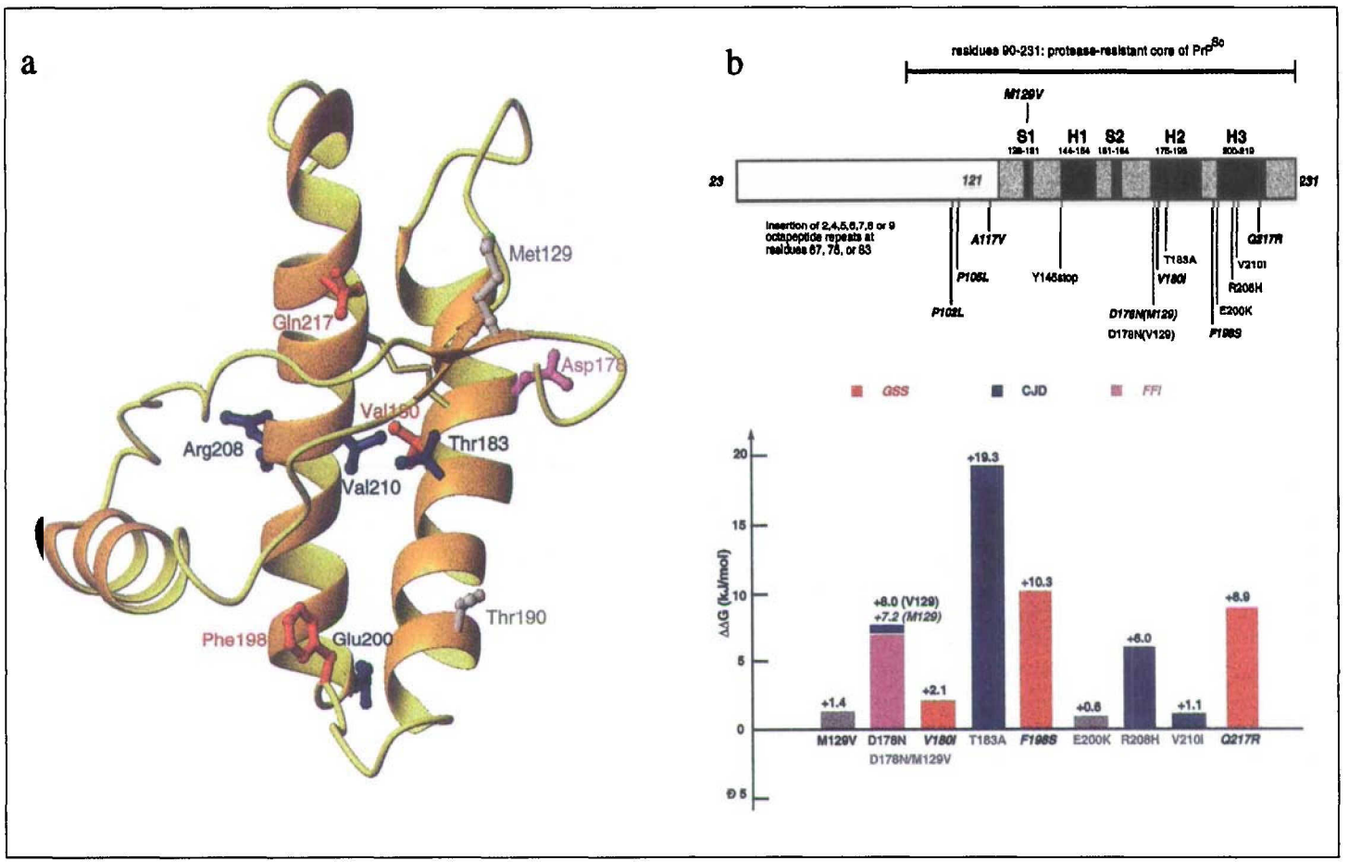

Fig. 4: Consequences of amino acid replacements related to inherited human prion diseases for the thermodynamic stability of the recombinant prion protein: (a) Ribbon diagram of the structured, C-terminal part of the murine prion protein, PrP(121-231) [22], showing the side chains of residues that are exchanged in inherited human prion diseases. Residues that are replaced in inherited Creutzfeldt Jakob diseases (CJDs) are indicated in blue, and residues that were found to be replaced in patients with Gerstmann-Sträussler-Scheincker syndrome (GSS) are shown in red. The human polymorphism at residue 129 (Met or Val), depicted in grey, determines the phenotype of familial the disease associated with the mutation Asp178Asn (pink), which is either fatal familial insomnia (FFI) (Met129/Asn178) or inherited CJD (Val129/Asn178). The single disulfide bond in PrP(1212-231) connecting helix 2 and helix 3 is also indicated. (b) Influence of amino acid replacements on the thermodynamic stability of PrP(121-231) at pH 7.0 and 22 °C [23]. The upper panel indicates the amino acid replacements in mature human PrP that are linked with inherited prion diseases in humans. Mature PrP^C consists of an unstructured N-terminal segment (residues 23-120), and a structured C-terminal domain (residues 121-231) (see (a)) consisting of a two-stranded β-sheet (S1, S2) and 3 α-helices (H1-3). The segment 90-231 in all subunits of oligomeric PrP^{Sc} exhibits resistance to proteinase K, while monomeric PrP^C is rapidly degraded. This indicates that PrP^{Sc} is an ordered oligomer in which the segment 90-120 most likely adopts a defined three-dimensional structure. Most of the point mutations linked with familial prion diseases are located within the structured C-terminal domain of PrP^C. The lower panel shows the difference between the free energy of unfolding of wild type murine PrP(121-231) and the corresponding recombinant variant (ΔΔG). The difference is defined such that a destabilization relative to the wild type yields positive ΔΔG values

[10] S. Jonda, M. Huber-Wunderlich, R. Glockshuber, E. Mössner, *EMBO J.* **1999**, *18*, 3271.
[11] E. Mössner, M. Huber-Wunderlich, A. Rietsch, J. Beckwith, R. Glockshuber, F. Åslund, *J. Biol. Chem.* **1999**, *274*, 25254.
[12] G.E. Soto, S.J. Hultgren, *J. Bacteriol.* **1999**, *181*, 1059.
[13] M. Pellecchia, P. Güntert, R. Glockshuber, K. Wüthrich, *Nature Struct. Biol* **1998**, *5*, 885.
[14] M. Pellecchia, P. Sebbel, U. Hermanns, K. Wüthrich, R. Glockshuber, *Nature Struct. Biol.* **1999**, *6*, 336.
[15] U. Hermanns, P. Sebbel, V. Eggli, R. Glockshuber, *Biochemistry* **2000**, 39, 11564.
[16] D. Choudhury, A. Thompson, V. Stojanoff, S. Langermann, J. Pinkner, S.J. Hultgren, S.E. Knight, *Science* **1999**, *285*, 1061.
[17] S.B. Prusiner, *Science* **1997**, *278*, 245.
[18] S. Hornemann, R. Glockshuber, *J. Mol. Biol.* **1996**, *261*, 614.
[19] R. Riek, S. Hornemann, G. Wider, M. Billeter, R. Glockshuber, K. Wüthrich, *Nature* **1996**, *382*, 180.
[20] R. Riek, S. Hornemann, G. Wider, R. Glockshuber, K. Wüthrich, *FEBS Lett.* **1997**, *413*, 282.
[21] S. Hornemann, R. Glockshuber, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6010.
[22] R. Riek, G. Wider, M. Billeter, S. Hornemann, R. Glockshuber, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **1998**, 95, 11667.
[23] S. Liemann, R. Glockshuber, R., *Biochemistry* **1999**, *38*, 3258.
[24] G. Wildegger, S. Liemann, R. Glockshuber, R., *Nature Struct. Biol.* **1999**, *6*, 550.