Review

3D Domain swapping, protein oligomerization, and amyloid formation[★][✿]

Mariusz Jaskólski

Department of Crystallography, Faculty of Chemistry, Adam Mickiewicz University and Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland

In 3D domain swapping, first described by Eisenberg, a structural element of a monomeric protein is replaced by the same element from another subunit. This process requires partial unfolding of the closed monomers that is then followed by adhesion and reconstruction of the original fold but from elements contributed by different subunits. If the interactions are reciprocal, a closed-ended dimer will be formed, but the same phenomenon has been suggested as a mechanism for the formation of open-ended polymers as well, such as those believed to exist in amyloid fibrils. There has been a rapid progress in the study of 3D domain swapping. Oligomers higher than dimers have been found, the monomer–dimer equilibrium could be controlled by mutations in the hinge element of the chain, a single protein has been shown to form more than one domain-swapped structure, and recently, the possibility of simultaneous exchange of two structural domains by a single molecule has been demonstrated. This last discovery has an important bearing on the possibility that 3D domain swapping might be indeed an amyloidogenic mechanism. Along the same lines is the discovery that a protein of proven amyloidogenic properties, human cystatin C, is

---

Abbreviations: BS-RNase, bovine seminal RNase; BSE, bovine spongiform encephalopathy; HCC, human cystatin C; HCCAA, hereditary cystatin C amyloid angiopathy; HP-RNase, human pancreatic RNase.

capable of 3D domain swapping that leads to oligomerization. The structure of domain-swapped human cystatin C dimers explains why a naturally occurring mutant of this protein has a much higher propensity for aggregation, and also suggests how this same mechanism of 3D domain swapping could lead to an open-ended polymer that would be consistent with the cross-β structure, which is believed to be at the heart of the molecular architecture of amyloid fibrils.

THE PHENOMENON OF 3D DOMAIN SWAPPING

Discovery and definitions

Experimentally, the existence of 3D domain swapping was established, and the term introduced, relatively recently, in 1994, when Eisenberg and coworkers observed it for the first time by X-ray crystallography in diphtheria toxin (Bennett *et al.*, 1994), but a phenomenon of essentially the same character had been predicted over three decades earlier based on ingenious, and today classic, experiments with activity recovery in dimers of ribonuclease (RNase) A molecules with partly knocked-out active sites (Crestfield *et al.*, 1962; 1963). When 3D domain swapping occurs, two (or more) subunits exchange identical structural elements or "domains". Those domains could be as small as short secondary structure elements, or as large as complete functional, globular, domains. In other words, in a 3D domain-swapped oligomer a structural unit of one subunit takes the place of the identical structural unit of another subunit, and *vice versa*, leading to the recreation of the monomeric fold but from chain segments contributed by different subunits. In a protein capable of undergoing domain swapping, there must exist a flexible linker or hinge region (usually, but not always, a loop or turn segment) whose conformational changes allow the molecule to partially unfold and then find another similarly open monomer (Fig. 1). Obviously, the hinge region is the only element that has a different structure in the monomeric and 3D domain-swapped forms. The main adhesive force allowing the domain-swapped oligomer to form is the "closed interface" between the swapped domains, which

recreates the structure and interactions of the protomer. It is a powerful factor in the structure of the oligomer as it has evolved to provide stability of the monomeric molecule. The oligomeric species has, however, also a new, or "open", interface between its components
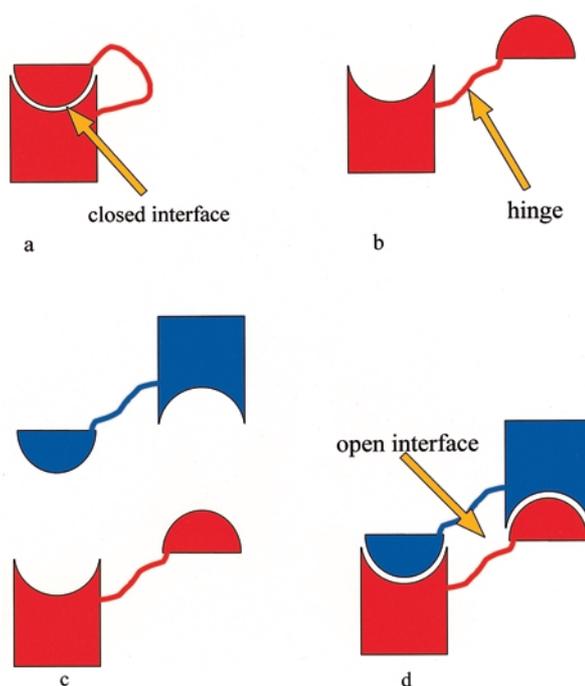

Figure 1. Cartoon illustration of dimer formation *via* 3D domain swapping.

The compact globular fold (a) is partially unfolded (b) through a conformational change at a flexible hinge region. The unfolding temporarily disrupts and exposes the closed interface, i.e. the contact area between the two domains. If sufficiently long-lived, and if present in sufficiently high concentration (c), the unfolded chains will mutually recognize their complementary interfaces and will recreate those contacts in a symmetrical, dimeric fashion (d). Through the closed interfaces, two monomeric folds are reconstructed. However, the dimer is not a simple sum of two monomeric molecules. The hinge regions in the new conformation form a new intermolecular interface that was not present in the monomer. This is the open interface.

that is not found in the monomeric form. If the oligomeric form is to be more stable than the monomers, the extra stabilization energy must come from the interactions in the open interface. It should be noted that, at best, only part of the energetic gain in the open interface will contribute to the stability of the oligomer because the rest of it will need to compensate for the entropic factor (loss of translational and rotational freedom), which always favors the monomer.

Examples

Taken rigorously, 3D domain swapping requires that the same amino-acid sequence can be observed in both the closed-monomeric form and as an intertwined oligomer. In practice the usage of the term is more liberal (Table 1), for instance tolerating sequence differences provided the folding pattern is the same, or the term could even be applied when the existence of the monomeric species is not certain at all. The classic example and the most thoroughly studied protein with 3D domain swapping properties is ribonuclease, which had been the subject of intense studies in the field of protein oligomerization even before the discovery of 3D domain swapping in diphtheria toxin. The ribonucleases, typified by RNase A, the bovine pancreatic enzyme, are a large family of monomeric proteins for which no function other than RNA hydrolysis is known. The only exception is bovine seminal RNase, BS-RNase, which is naturally dimeric and this form, in addition to allowing for allosteric regulation of the two active sites, endows the molecule with a number of unusual biological activities. BS-RNase has been found, for instance, to be immunosuppressive, antiviral, cytotoxic for tumor cells, etc. The case of BS-RNase is particularly intriguing because, genetically, it is present as a (defective) pseudogene in all ruminants except in ox (*Bos taurus*) and water buffalo (*Bubalus bubalis*), where the gene is functional. In the latter case, however, in variance

Table 1. Examples of 3D domain swapping.

The division into *bona fide* cases, where the same protein is observed in two forms, and *quasi* cases, where the monomeric fold is recreated in a 3D domain-swapped dimer by a different (although homologous) sequence, is after Eisenberg (Schlunegger *et al.*, 1997). Because of the possibility of domain-swapping control by mutations, and because of the possibility that some existing examples may have been undetected so far, this useful line of division may be somewhat fuzzy.

bona fide domain swapping
- diphtheria toxin
- RNases (A, BS-)
- CksHs (cell cycle regulation)
- CD2 (T-cell adhesion)
- staphylococcal nuclease
- λ-cro (DNA repressor)
- α-spectrin (cytoskeleton)
- antibody fragments
- NO reductase
- NOS(ox) (redox protein)
- Spo0A (development regulation)

quasi domain swapping
- crystallin
- growth factors/cytokines
- pheromone/odorant binding/transport proteins
- SH3 (signal transduction)
- glyoxalase I
- RYMV (viral capsid protein)
- human cystatin C (protease inhibitor)

with the high levels of BS-RNase in bovine seminal plasma, the protein is never expressed. Although the BS-RNase dimer has a covalent nature (two intermolecular disulfide bridges between uniquely placed cysteine residues), it exists in an equilibrium with about 2/3 of the molecules having an additional quaternary connection through an exchange of an N-terminal helix (Mazzarella *et al.*, 1993). Inspired by this observation, researches have tried to characterize domain swapped dimers of RNase A molecules, whose naturally monomeric structure had been established by Wlodawer with very high accuracy (Wlodawer *et al.*, 1982; 1988). Dimerization of RNase A

(by lyophilization from acetic acid) had been achieved early on (Crestfield *et al.*, 1962) but the crystallographic model establishing the structure of the N-terminal-domain-swapped dimer was published only in 1998 (Liu *et al.*, 1998). Surprisingly, this structure, although sharing the closed interface with dimeric

tained by one of the camps, the priming event was domain swapping triggered by a mutation or environmental change, the phenomenon of 3D domain swapping would be not only a structural curiosity but a powerful mechanism for rapid evolution of proteins from monomeric towards oligomeric forms with
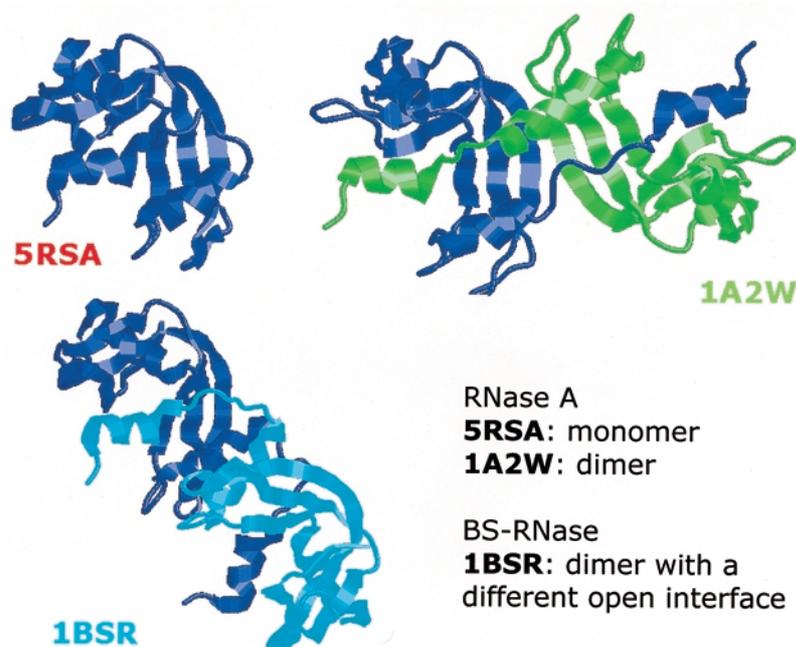

Figure 2.   Two types of 3D domain-swapped dimers with different open interfaces formed by ribonuclease.

Monomeric RNase A (PDB accession code 5RSA) is shown upper left. Using partially destabilizing conditions, it could be induced to crystallize in the domain-swapped dimeric form 1A2W (upper right), although physiologically the monomeric form dominates. The highly similar BS-RNase is known as a biological dimeric molecule with swapped domains (lower left, 1BSR). The swapped domain (and consequently the closed interface) is the same in 1A2W and 1BSR. However, the open interfaces are different.

BS-RNase (and, of course, with monomeric RNase A), had a different open interface and thus a different overall quaternary structure (Fig. 2). Since the structure of BS-RNase folded into the monomeric form is known as well (Piccoli *et al.*, 1992), both bovine ribonucleases (BS-RNase and RNase A) fulfill the requirements of Eisenberg's *bona fide* 3D domain swapping (Table 1). The special case of BS-RNase dimers interconnected through both covalent (S–S) and quaternary (domain-swap) interactions is of importance for the discussion by which of those two mechanisms ("covalent first" or "swap first") these dimers arose (D'Alessio, 1999). If, as main-

new biological properties. Recently, the RNase has surprised us again, when Liu *et al.* (2001) discovered yet another dimer, this time formed through the exchange of a C-terminal β-strand of the RNase A molecule. Another surprise came from the discovery by Park & Raines (2000) that RNase A may form dimers (in equilibrium with monomers) also in physiological conditions, and that those dimers would arise *via* N-terminal domain swapping.

Oligomers higher than dimers have been detected in several cases where 3D domain-swapped dimers had their structure confirmed by X-ray crystallography. Application of "Ockham's razor" would suggest that those

aggregates, even in the absence of a direct structural proof, are formed through domain swapping as well. For instance, Adinolfi *et al.* (1996) discovered that BS-RNase forms tetramers as well, in which the four subunits are enchained by multiple domain-swapping events. But there are also structural studies reporting higher 3D domain-swapped oligomers, as in the case of a trimeric antibody fragment with noncognate V_H-V_L domain pairs (Pei *et al.*, 1997). The antigen-binding site in antibodies is formed by the hyper-variable loop regions of V_H and V_L (heavy and light) domain pairs. In the above study, a polypeptide chain was constructed by direct fusion of a V_L domain directly to the C-terminus of a V_H domain of an unrelated antibody. The chains oligomerized into cyclic head-to-tail trimers with three Fv heads composed of V_H and V_L domains from the consecutive subunits.

In the span of about six years since 3D domain swapping was established structurally, more than two dozen cases have been characterized by X-ray crystallography of protein oligomers, mostly dimers, formed by domain swapping. Among them are proteins of very diverse biological functions including, in addition to enzymes, regulatory and signaling proteins, receptors, transport proteins, structural proteins, even a viral capsid protein (Table 1). It seems that domain swapping is a much more common phenomenon than originally believed, and that even if not always found naturally, for many proteins can be induced artificially. This reflection has a more general bearing on one of the most fundamental canons of structural biology, one sequence — one structure, viewing protein folds as iron-clad invariants uniquely determined by amino-acid sequences. As the example of RNase A tells us, a protein can even adopt several different folds. In view of the accumulating evidence, not only from the 3D domain swapping field, it may become necessary to revise those useful, but simplified assumptions.

Protein engineering for 3D domain swapping

Manipulation of protein sequences has led in several cases to control of 3D domain swapping demonstrating that our understanding of this phenomenon and of the factors governing protein folding is already quite deep. One example is the V_H-V_L construct without a linker sequence that resulted in a trimeric 3D domain swapped antibody, as described above. The trick of linker shortening or deletion has now become a standard technique in engineering proteins for 3D domain swapping.

In another study, Murray *et al.* (1995) found that when expressed as part of a fusion protein, the N-terminal domain of the lymphocyte adhesion molecule CD2 is capable of adopting a monomeric as well as 3D domain-swapped dimeric form. In the native sequence, the dimers were less abundant (15%) and represented a metastable fold since denaturation and refolding in the absence of the fusion partner converted the dimeric CD2 into monomers. Subsequently, Murray *et al.* (1998) showed that it was possible to differentially stabilize either fold by engineering the CD2 sequence, mimicking random mutagenesis events that could occur during molecular evolution.

The ribonuclease story has also its protein-engineering chapter. Human pancreatic (HP) RNase (Beintema *et al.*, 1984) is monomeric and with no special activities. By protein engineering with inspiration from the BS-RNase example, Piccoli *et al.* (1999) designed and produced a dimeric form of HP-RNase that showed a cytotoxic effect on tumor cells.

Finally, it is interesting to mention the experiments of Ogihara *et al.* (2001), who by engineering the structure of an artificial protein called 3-α-helical bundle, managed to convert it into a 3D domain-swapped dimer, exactly as designed. Those engineering experiments consisted in manipulating the loops connecting the helices in such a way that reconstruction of the bundle was only possible through a

(in the case of the dimer — reciprocal) domain swap. In another variant of this experiment, a topological modification of the protein resulted in linear polymerization (*vide* infra).

HUMAN CYSTATIN C

Properties

One of the most recent discoveries of 3D domain swapping with important physiological consequences is the case of human cystatin C (HCC) (Janowski *et al.*, 2001). In its physiological role as one of the most important extra- and transcellular cysteine protease inhibitors, monomeric HCC is present at high levels in all body fluids (Abrahamson *et al.*, 1986). In addition to inhibiting papain-like proteases through an epitope consisting of the N-terminus and two hairpin loops (L1, L2) aligned at one edge of the molecule (Fig. 3), it has been recently found to also inhibit legumain-like proteases *via* a different mechanism, probably involving residues at the opposite edge of the molecule, within the so-called back-side loops. It has now been established that wild type HCC forms part of the amyloid deposits in brain arteries of elderly patients suffering from cerebral amyloid angiopathy (Grubb, 2000). In hereditary cystatin C amyloid angiopathy (HCCAA), occurring endemically in the Icelandic population, a natural variant of HCC (Leu68Gln) forms massive amyloid deposits in brain arteries of young adults (Fig. 4) leading to lethal cerebral hemorrhage (Olafsson & Grubb, 2000). Since in both cases aggregation involves abnormal, pathological changes of protein conformation, these disorders can be classified, together with the Alzheimer's disease and the prionoses, as conformational diseases.

The cystatin fold

Before the crystal structure of human cystatin C became available, the general fold of protein inhibitors belonging to the cystatin family had been established based on the structure of a related chicken protein (Bode *et al.*, 1988; Dieckmann *et al.*, 1993; Engh *et al.*, 1993) with which HCC (consisting of 120 amino acids) shares 41% sequence identity and 62.5% similarity. The canonical features of this fold include a long $\alpha 1$ helix running across a large, five-stranded antiparallel β-sheet of the following connectivity: (N)-$\beta 1$-($\alpha 1$)-$\beta 2$-L1-$\beta 3$-(AS)-$\beta 4$-L2-$\beta 5$-(C), where AS, a broad "appending structure", is posi-
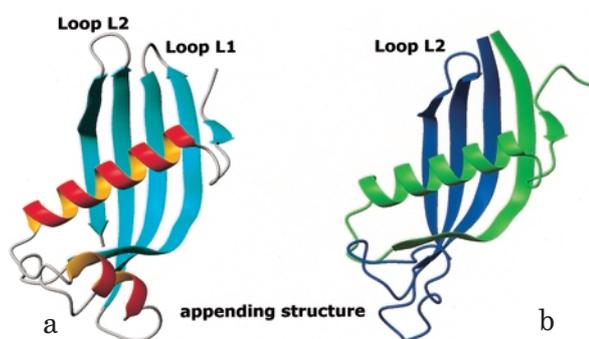

Figure 3. The cystatin fold.

The crystal structure of chicken cystatin (PDB 1CEW) (Bode *et al.*, 1988) is shown on the left (a). A five-stranded antiparallel β-sheet wraps around a long α-helix that is almost perpendicular to the general β-strand direction. The curvature of the β-sheet results from two β-bulge elements observed in the otherwise regular β structure. The first strand of the β-sheet (poorly defined in chicken cystatin structure because of N-terminal truncation) is separated in the sequence from the remaining strands by the intervening α-helix sequence. The remaining β-strands are consecutive and connected through hairpin loops (L1, L2) or through a broad loop on the opposite end of the β-sheet, known as the appending structure (AS). The helical element in the appending structure of chicken cystatin is highly uncertain because of very poor or missing electron density. The hairpin loops and the N-terminal chain are aligned in a wedge-like fashion at one end of the molecule and form the inhibitory motif that is docked in the enzyme's catalytic cleft. On the right (b), one half of the dimeric HCC molecule is shown, formed from interlaced fragments of two polypeptide chains, green and blue. Note the fidelity with which the structure of the closed monomer is reconstructed.
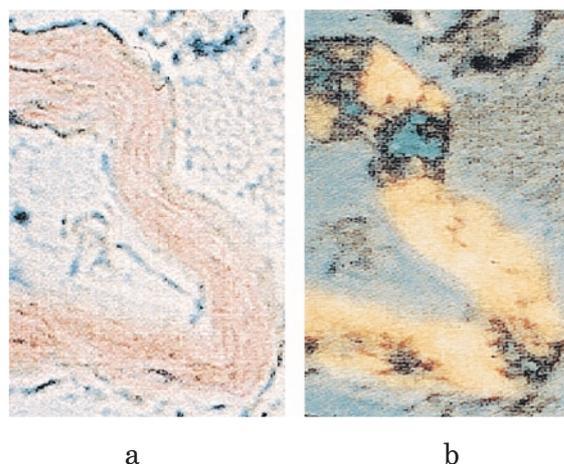
a   b

Figure 4. Amyloid deposits at ×320 magnification.

Congo Red staining of amyloid deposits in a cerebral artery of an Icelandic HCCAA patient (a). Viewed in polarized light (b), the congophilic amyloid fibrils of the same sample show yellow-green birefringence. Courtesy of Professor Anders Grubb.

tioned on the opposite ("back-side") end of the β-sheet relative to the N-terminus and loops L1 and L2 (Fig. 3). As all other type-2 cystatins (Barrett *et al.*, 1986; Rawlings & Barrett, 1990), HCC contains four characteristic disulfide-paired cysteine residues. The disulfide bridges are formed in the C-terminal half of the molecule, between Cys73 and Cys83, stabilizing the structure of the random-coil (AS) region between strands $\beta3$ and $\beta4$, and between Cys97 and Cys117, connecting the ends of the $\beta4$–$\beta5$ hairpin (Fig. 5).

The crystal structure of HCC dimer

Crystallization of human cystatin C has been reported from solutions of monomeric protein prepared by gel filtration in the final isolation step (Kozak *et al.*, 1999). The crystal structure (Janowski *et al.*, 2001) revealed, however, that the molecules formed two-fold-symmetric dimers *via* 3D domain swapping. This result is consistent with the view that local high concentration (as in the crystallization droplet) is necessary for the formation of 3D domain-swapped oligomers (Liu *et al.*, 1998). In the dimers, the monomeric fold defined by the

crystal structure of chicken cystatin is reconstructed with high fidelity but, as in all 3D domain-swapped oligomers, from parts belonging to different polypeptide chains (Figs 5, 6, 7). This confirms earlier NMR results indicating that the secondary structure elements of HCC are preserved upon dimerization (Ekiel & Abrahamson, 1996; Ekiel *et al.*, 1997). Analysis of a single polypeptide chain "extracted" from the dimeric context (Fig. 6) reveals that the monomeric molecule underwent partial unfolding through an opening movement of loop L1, one of the inhibitory elements located at the edge of the monomeric structure. This hinge movement produced an unnaturally-looking conformation, ready for swapping domains with another unfolded chain, in order to bury the exposed surfaces that are not evolved to interact with water. By changing its
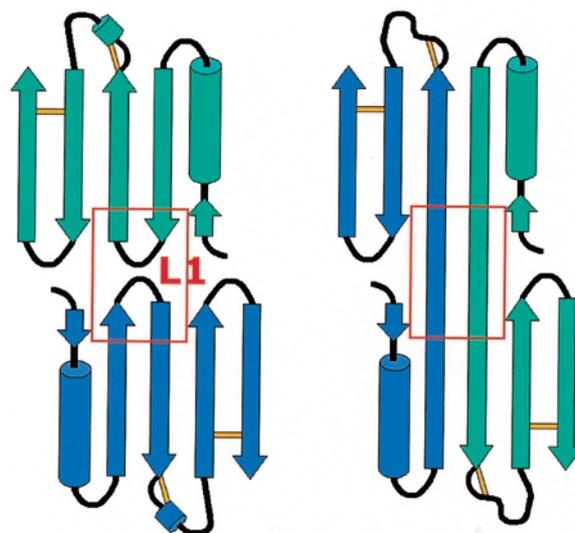

Figure 5. A diagrammatic illustration showing how a two-fold-symmetric pair of HCC molecules (left) exchange domains through a conformational change of the β-hairpin loop L1 (red box, left) known to be an inhibitory element from chicken cystatin structure.

Note that in the domain-swapped dimer (right) a very long intermolecular antiparallel β-sheet is formed, the central part of which (red box, right) is the newly created open interface that makes the dimer energetically advantageous. The yellow lines indicate the disulfide bridges, conserved in type-2 cystatins (Barrett *et al.*, 1986; Rawlings & Barrett, 1990).
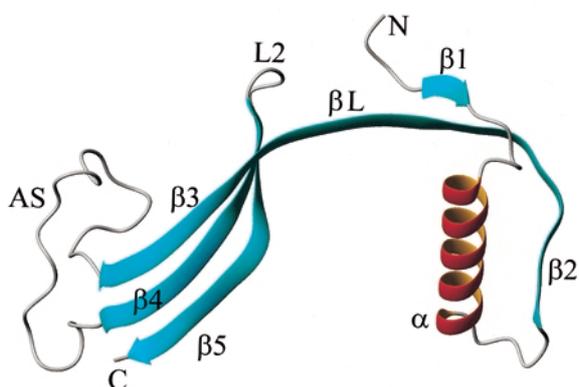
Figure 6. An HCC subunit "extracted" from the domain-swapped context of the dimer to emphasize its partially unfolded, unnatural conformation.

Standard labeling of cystatin topology is also shown. The former L1 loop (as found in chicken cystatin) serves now as a linker (or hinge) and is labeled βL.

conformation, the L1 segment now became part of a long β-strand running from the beginning of β2 to the end of β3. In addition to the monomer-type closed interface, the dimers also contain the new open interface (Fig. 7). It is formed through β-sheet interactions in the βL region (former L1) leading to the creation of an unusually long contiguous antiparallel β-sheet formed by two copies of strands Tyr42–Thr74 (β2-βL-β3) which cross from one domain to the other with as many as 34 hydrogen bonds between the main chains and extra hydrogen bonds involving side chains. The hinge element L1 separates the two domains that are exchanged in the dimer. The N-terminal domain is mostly α-helical, but includes also the N-terminal β-strand (β1), and the strand that immediately follows the α1 helix, β2. The C-terminal domain is of purely β character and consists of the remaining part of the β-sheet, strands β3–β5. This means that disruption of the monomeric fold requires not only separation of the α-helix from the β-sheet by which it is wrapped, but also tearing one of the seams of the β-sheet (β2–β3). Likewise, the closed interface rebuilt in the dimer is cemented by these same interactions. The faithfulness of the reconstruction of the fold by the swapped domains may be il-

lustrated by the presence of a tandem of β-bulges at residues Ala46 and Val49 in strand β2 that shift the β register by two residues. These β-bulges, as well as another one at Leu112 in strand β5, have their counterparts in monomeric chicken cystatin and they must be present to introduce curvature into the β-sheet that is required for its wrapping around the α-helix. Finally, it may be observed that the two disulfide bridges introducing rigidity into the fold of this small protein are both present in the C-terminal domain and in consequence not only do not interfere with the domain swapping process, but help to maintain the integrity of the C-terminal domain during the transition period when the protein is partially unfolded.

The disappearance of loop L1 in the dimeric structure and, consequently, the disruption of this functional element of the protein, agree with the observation that HCC dimers have absolutely no inhibitory effect on papain-type proteases (Abrahamson & Grubb, 1994; Ekiel & Abrahamson, 1996). On the other hand, loop 39–41, which connects helix α1 with strand β2 and contains asparagine 39 that is crucial for HCC inhibition of mammalian legumain, is not affected by dimerization. This is in agreement with the observation that
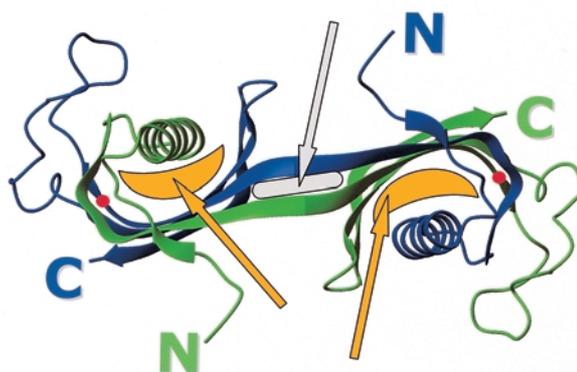

Figure 7. HCC 3D domain-swapped dimer viewed "on edge" of the β-sheets, with the open interface (grey) exposed towards the viewer.

Note the closed interfaces (yellow) between the β-sheets and the α-helices (viewed along their axes). The red dots indicate the location of leucine 68 in the β-sheet within the closed interface.

dimeric HCC is as active in inhibiting porcine legumain as the monomeric protein (Alvarez-Fernandez *et al.*, 1999).

It has to be admitted that this is not a strict *bona fide* 3D domain swapping case as defined by Eisenberg, because the structure of monomeric HCC is not precisely known. However, we know that such monomers do exist and we can be quite confident that their structure closely resembles that of the chicken analog.

The crystal structure of HCC reveals that in spite of a high solvent content (71%), there are some interesting packing interactions between the 3D domain-swapped dimers that may be of significance for further aggregation of the protein. The most conspicuous assemblies consist of eight HCC monomers, or four dimers, arranged around a crystallographic tetrad (Fig. 8a). Two four-fold-related dimers interact through a rich system of hydrogen bonds (duplicated in two copies) involving the

back-side loops of one dimer, and the solvent-accessible face of the β-sheet of the other. Additionally, the side chain of Met41 is locked in a pocket formed by residues on the $\beta3$ and $\beta4$ strands of the complementary dimer. There are as many as nine unique hydrogen bonds, five involving the $\alpha1$–$\beta2$ loop (in this number one of main-chain — main-chain type, one from a side chain to the main chain, and three between side chains), and four involving the AS loop (three between the main-chain and side-chains, one between side chains). The back-side loops interact predominantly with the convex face of the β sheet of one domain, but there are additional contacts involving the linker region and the L2 loop from the other domain. The octamer appears to have a stable structure; the total number of hydrogen bonds between the dimers is 72. The connectivity within an octamer is very interesting. The pattern is both dimeric and circular. If the four dimers, each comprising domains 1 and
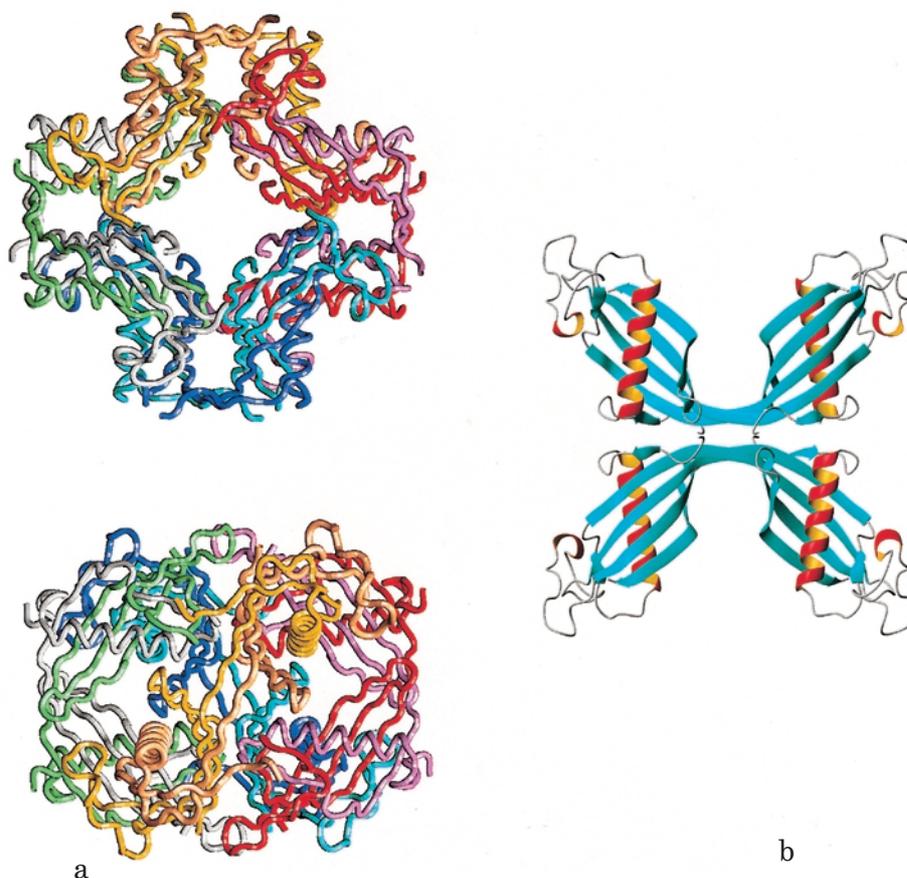

a

b

Figure 8. Aggregation of the dimers observed in the crystal structure of HCC (Janowski *et al.*, 2001).

(a) Two views of an octameric aggregate formed by four domain-swapped dimers, organized in a four-fold symmetric fashion *via* β-sheet...back-side-loops interactions. The upper view is along the four-fold axis, in the lower view the four-fold axis is vertical. (b) Dimer...dimer interactions involving hydrophobic contacts between the β-sheets of the linker regions. These interactions interconnect the octameric aggregates into an infinite three-dimensional network. The dimers' dyad is vertical. Two other two-fold axes (horizontal and along the viewing direction) operate between the dimers.

2, are labeled A, B, C, D, and the interacting elements are denoted as l (back-side loops) and b (β sheet), then the topology of the connections is as follows: Al1...Bb2 and Ab2...Bl1, Bl2...Cb1 and Bb1...Cl2, etc., and finally Dl2...Ab1 and Db1...Al2. In this fashion, all the interacting elements (l and b) are utilized and in this sense the octamers are closed, sphere-like assemblies. They are, however, interconnected *via* weaker van der Waals interactions to form an interwoven three-dimensional network. These hydrophobic interactions involve residues from the linker regions (unfolded L1 loops), which are fairly exposed in the dimer structure, from a pair of two-fold related dimers (Fig. 8b). Each set of interactions includes eight short C...C contacts involving residues Ile56, Ala58, and Val60. A single octamer has four "hydrophobic neighbors" in two perpendicular directions in one plane, but the directions involving octamers connected in one chain alternate. This leads to an intricate spatial arrangement of the octamers in which no linker regions of the dimers are left unprotected.

Implications for the L68Q mutant

Leu68 is located on the central strand β3 of the β-sheet, on its concave face covered by the α-helix (Fig. 7). In the hydrophobic core of the protein, it occupies a niche formed by the surrounding residues on the β-sheet and the hydrophobic face of the helix (Fig. 9). The closest distances in this area represent typical hydrophobic contacts. Replacement of the leucine side chain by the longer glutamine side chain, as in the naturally occurring pathological Leu68Gln variant of human cystatin C, would not only make those contacts prohibitively close but would also place the mutated hydrophilic chain in a hydrophobic environment. This would definitely destabilize the molecular α-β interface and lead to repulsive interactions expelling the α-helix, together with the intervening strand β2, from the compact molecular core and forcing the molecule to un-
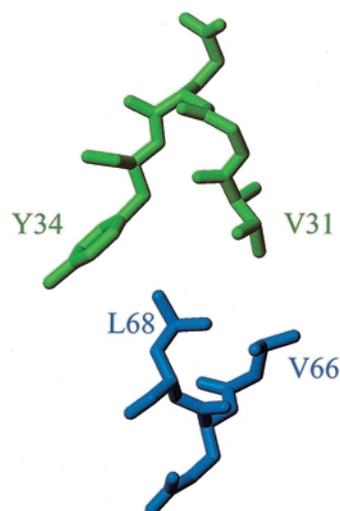

Figure 9. Leucine 68 in native HCC located in a hydrophobic cavity formed by residues at the C-terminal end of the α-helix.

In the domain-swapped dimer, this helix is contributed by the other subunit, indicated by different color (green). In monomeric human cystatin C, the interactions in this area are presumably identical. When leucine 68 is mutated to glutamine, as in the naturally occurring variant, the new residue is too big for this cavity and has an incompatible (hydrophilic) chemical character, thus destabilizing the fold.

fold into the α and β domains. This explains the increased dynamic properties of the Leu68Gln mutant compared with wild type HCC observed by NMR spectroscopy (Ekiel *et al.*, 1997; Gerhartz *et al.*, 1998). Under the assumption that the refolded dimer recreates the topology of the monomeric HCC molecule, those destabilizing effects would be similar in both cases. However, the dimeric structure may be more resistant to destruction because of the extra stabilizing contribution form the β-interactions in the linker region, or more generally in the long β2–βL–β3 region. A hydrophilic substitution at the α–β interface would be also expected to lower the energy barrier corresponding to the unfolded state through reducing the unfavorable solvent contacts of the newly exposed interface. A speculative diagram illustrating the thermodynamic relations in wild type and Leu68Gln monomer-dimer equilibria of HCC is presented in Fig. 10. The above discussion of the

effect of the Leu68Gln substitution on HCC dimerization is supported by the observation that the mutated variant forms dimers in human body fluids more easily than wild type cystatin C (Bjarnadottir *et al.*, 2001).

AMYLOID

History of discovery

The term "amyloid" has a long and colorful history (Cohen, 1986; Sunde & Blake, 1998; Sipe & Cohen, 2000). First descriptions of foreign-deposit-laden post-mortem tissues and organs appear as early as the seventeenth century. The term "amyloid", which indicated carbohydrate suggested by iodine staining (and containing a further inaccuracy connected with the confusion, at that time, between cellulose and starch) was introduced by Virchow in 1854, a few years before Friedreich and Kekulé demonstrated that amyloid deposits were of predominantly proteinaceous character. To add to the confusion, it has to be admitted that *corpora amylacea* in the brain, on the observation of which Virchow coined his term, have been recently found to indeed consist primarily of polysaccharides, and as such are not amyloids in our present use of the term. It refers to a less picturesque object, a pathological proteinaceous substance deposited extra- or intracellularly in tissues, having fibrous morphology even under light microscope, clinically leading to tissue damage, and typically connected with lethal diseases.

Amyloidogenic proteins

Amyloid deposits are formed of proteins that are otherwise normal and soluble in their physiological role. The list of proteins with confirmed amyloidogenic properties has grown in the recent years to include about 20 cases (Table 2) and it appears that as our knowledge and research tools develop, it will keep growing. In addition to such archetypal
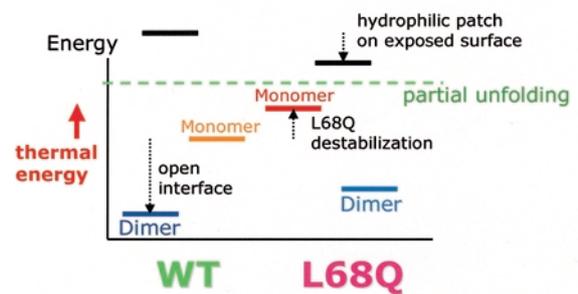

Figure 10. Hypothetical thermodynamics of wild type (left) and Leu68Gln (right) HCC.

Partial unfolding of the polypeptide chain, leading to separation of the α and β domains, requires thermal energy. In normal conditions (broken line), the wild type HCC monomer is probably sufficiently stable and the energy of the unfolded intermediate sufficiently high, to make transitions to the 3D domain-swapped dimeric form rare, even though it may be energetically favored. The energetic gain would correspond to the formation of the new open interface in the domain-swapped dimer, minus the entropic loss. The monomeric form of the Leu68Gln mutant is destabilized by the repulsive interactions at the Gln68 side chain and partial unfolding may be achieved more easily. Additionally, the hydrophilic character of Gln68 at the solvent-exposed surface would make the unfolded intermediate less unstable. In this situation, even the thermal energy at normal conditions would be sufficient to pass the barrier for the monomer→dimer transition. Once formed, the Leu68Gln HCC dimer would be sufficiently stable (and the wild type HCC dimer even more so), to make spontaneous dimer→monomer transition practically impossible. (The free energy gain of the mutant dimer over the monomer would be, as in the case of the wild type protein, a combination of the enthalpic gain from the open interface and entropic penalty for the lost degrees of freedom.)

examples as the Alzheimer amyloid β-protein or the prion protein that are perceived in this context even in popular notion because of the widely known diseases they cause, there is a whole range of cases comprising proteins with a diverse spectrum of biological functions in their normal, non-aggregated state. Among the selected examples listed in Table 2, all of which are associated with amyloid deposits in

Table 2. Examples of human amyloidogenic proteins (according to Sipe & Cohen, 2000).

Immunity/inflammation-related
- immunoglobulin fragments
- β2-microglobulin
- serum amyloid A
- lysozyme
- fibrinogen α
- human cystatin C

Endocrine system-related
- islet amyloid polypeptide
- atrial natriuretic factor
- procalcitonin
- prolactin
- insulin

Transport proteins
- transthyretin
- apolipoprotein AI
- lactoferrin

Nervous system-related
- Aβ protein precursor
- prion protein

Cell motility protein
- gelsolin

humans, there are ubiquitous or long-studied proteins that were not suspected to have such connotation, for instance, lysozyme, insulin, or transthyretin. Transthyretin, a thoroughly studied transporter of the thyroid hormone, which is now known to aggregate and cause familial amyloidotic polyneuropathy, is worthy of a special mention because it became a model case which allowed Blake and coworkers to establish the fundamental characteristics of a generic fibril structure (Sunde *et al.*, 1997). Originally, amyloid fibrils were detected in and isolated from the affected tissue but today it is possible to generate amyloid fibrils *in vitro*. The discovery that amyloid formation is not restricted to a limited number of protein sequences associated with diseases (Dobson, 1999) has significantly enlarged the field of study. From an analysis of lysozyme mutants as well as from other cases, it has been proposed that proteins could contain (in

wild type form or after mutation) "chameleon" sequences that would be equally unstable in α and β conformation thus being the triggering factors in chain unfolding and the exchangeable elements in domain exchange (Perutz, 1997; Booth *et al.*, 1997; Minor & Kim, 1996). If such labile chameleon sequences were a common situation in proteins, the potential for conformational aberrations would be much higher than is currently believed.

In vitro production of amyloid fibrils, structurally similar to those extracted from patients, usually requires partially denaturing conditions (Chiti *et al.*, 2000). There are also attempts to engineer polypeptide chains for amyloid formation, like the experiments with amyloid β-protein described by Teplow (1998) or with acylphosphatase reported by Chiti *et al.* (2000). Another mystery about amyloid is that, while it is often related to or triggered by a mutated variant of a normally stable protein, its formation can also occur in the unmutated form. This is observed, for instance, for HCC (described above) and the prion protein. In the latter case, the misfolded, or conformationally defective form of the protein is considered to be the transmissible pathogenic agent.

Amyloid criteria

After the early confusions, the introduction of modern scientific tools from the mid twentieth century, has led to the acceptance of three basic criteria that must be met by amyloid deposits, connected with their tinctorial, morphological, and structural characteristics (Sunde & Blake, 1998). Firstly, amyloids have specific tinctorial properties, i.e. are stained when treated with organic dyes, such as the bis-diazo dye Congo Red. In the test using Congo Red, amyloids are stained to give a characteristic apple green birefringence (Glenner *et al.*, 1972) when viewed in polarized light (Fig. 4). Secondly, electron micrographs of amyloid deposits show them to be

composed of uniform and straight (and hence structurally rigid) fibers (Fig. 11) with about 100 Å diameter (Cohen & Calkins, 1959; Co-

Figure 11. Electron micrograph of amyloid fibrils formed from the human prion protein.

Courtesy of Professor Paweł Liberski.

hen *et al.*, 1982). Thirdly, X-ray diffraction patterns of amyloid fibrils show them to have ordered, repeating structure, consistent with the so-called cross-β structure (Glenner, 1980a; 1980b), in which extended polypeptide chains in β-conformation are perpendicular to the fiber axis, and form (presumably antiparallel and twisted) β-sheets that are parallel to the fiber axis.

Amyloid structure

Our current knowledge about amyloid structure at the molecular level (Fig. 12) comes from fiber X-ray diffraction studies. Fiber diffraction has a long history and record of success in structural biology in having led to fundamental discoveries predating those produced by single crystal structure analysis (first clues about the structure of fibrous proteins, data for predicting secondary structures of proteins, the DNA double helix, the structure of helical viruses). It is based on the fact that fibrous material has at least one-dimensional order, along the fiber axis. In the preferred experimental setup, the fibers are oriented along their axis and set perpendicular to a monochromatic X-ray beam, ideally at a synchrotron source. The diffraction pattern that is then produced will reflect the
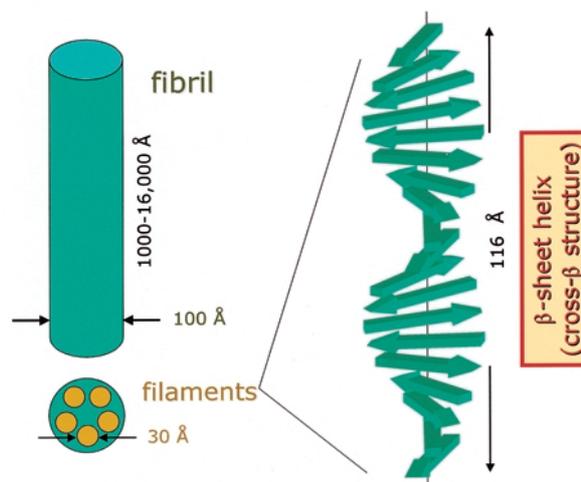

Figure 12. Fibril structure and β-sheet helix.

From electron microscopic observations, several studies have concluded that amyloid fibrils are long tubular structures with a diameter of about 100 Å (Cohen *et al.*, 1982). The fibrils are believed to be composed of thinner filaments wrapped around each other in a helical fashion. The number of protofilaments per fibril has not been firmly established. At the molecular level, the filament structure is revealed by X-ray fiber diffraction (Sunde *et al.*, 1997). The diffraction patterns are consistent with "cross-β structure" which takes the form of a (most likely antiparallel) β-sheet helix, with individual β-strands perpendicular to the helix axis, and the β-sheet face parallel to the helix axis. The repeat distance of the β-sheet helix has been estimated at 115.5 Å (Sunde & Blake, 1998). For clarity, the number of β-strands per repeat of the β-sheet helix in the diagram is arbitrary (less than found experimentally).

repetitive structural features of the fibers, such as: (i) about 4.7 Å distance between the β-strands having perpendicular arrangement to the fiber axis (from a very strong reflection in the meridional direction), (ii) about 115.5 Å repeat distance (pitch) of the helically twisted β-sheet (from high-order reflections in the meridional direction), and (iii) about 10 Å spacing of the β-sheets in the fibril (from reflections recorded in the equatorial direction) (Sunde *et al.*, 1997). In particular, synchrotron X-ray studies have suggested that the core of the transthyretin amyloid fibril is a continuous β-sheet helix (Blake & Serpell, 1996). The degree of similarity in the diffraction patterns of amyloid fibers produced from

different polypeptides is indicative of a common core structure, which must be assumed in the fibril regardless of the soluble-form properties of the constituent protein and despite the known, large differences in folding of the precursor proteins (Sunde *et al.*, 1997). The above, rather general characteristics of amyloid fibers suggest that amyloid, as a particular type of molecular structure, may be accessible to many proteins.

3D DOMAIN SWAPPING AND AMYLOIDOGENIC AGGREGATION

Current views

The formation of aggregates from monomeric proteins moved to the focus of scientific interest when it was found to be involved in the Alzheimer's disease and transmissible spongiform encephalopathies such as BSE. Recently, Eisenberg and coworkers (Liu *et al.*, 2001) have extended the conjecture of Chiti *et al.* (2000) that any protein may form amyloid if in sufficiently destabilizing conditions, to propose that every protein may undergo domain swapping at high concentration and in partially destabilizing environment.

Although 3D domain swapping has been proposed as a mechanism of amyloid fibril formation (Klafki *et al.*, 1993; Bennett *et al.*, 1995; Cohen & Prusiner, 1998), there is no direct experimental proof at present that in amyloid the protein molecules are associated *via* domain exchange. However, the striking parallels between the two phenomena provide very strong circumstantial evidence. For instance, both processes are highly selective with respect to their building blocks. Another piece of evidence comes from a recent report where, as mentioned above, the authors have elegantly engineered two 3D domain-swapped derivatives of two helical protein scaffolds, designed to undergo either 3D domain swapping dimerization or 3D domain swapping multimeric fibrous assembly (Ogihara *et al.*, 2001).

The predicted assemblies were detected by a variety of physico-chemical methods and the structure of the dimers was confirmed by X-ray crystallography. Although the structure of the fibrils that formed in the second case was not established at the molecular level, all other evidence points to 3D domain swapping as the mechanism through which those fibrils were formed. It has to be admitted, however, that, as designed, those fibers had predominantly helical structure while it is generally accepted that the underlying molecular architecture of amyloid fibrils, of both natural and artificial origin, is essentially of β character.

Thus, with the current knowledge about protein aggregation, the mechanism of 3D domain swapping seems to be a plausible and logical possibility. In clinical situations, detection of aberrant, non-physiological dimers might be a diagnostic signal warning that abnormal conformational changes are taking place and that the risk of amyloid formation is high. Indeed, it has been recently shown that dimers of the Leu68Gln mutant are present in body fluids of patients with the trait for hereditary cystatin C amyloid angiopathy leading to fatal brain hemorrhage in early adult life (Bjarnadottir *et al.*, 2001). On the other hand, closed-ended oligomerization could be also viewed as a process that is competitive with respect to infinite open-ended polymerization. Formation of symmetrical 3D domain-swapped dimers, for example, may be a dead-end on the oligomerization pathway, depleting the concentration of unfolded molecules and, at a given stage temporarily *preventing*, or rather *delaying* amyloid formation.

Amyloid formation *versus* crystallization

At favorable conditions, when the concentration of partially unfolded molecules is high and/or the environment increases their stability (and/or destabilizes the monomers), seeds of open-ended oligomers may be of sufficient life-time to sequester further monomers and grow to sufficient size when they finally be-

come stable and enter a steady growth state. The term "seed", borrowed from crystal growth theory, is very appropriate here because these two phenomena seem to be closely related. The most appealing analogy is to protein crystallization, the difference being that protein crystals are (usually) objects with three-dimensional periodicity, while amyloid fibers have regular repetition in one direction only, and thus can be termed "one-dimensional crystals". To draw the analogy further, in protein crystallization it is not sufficient to concentrate the protein ("monomers") just above the solubility limit. Crystal seeds can only form if the supersaturation is very high (labile region). If the solution is only slightly supersaturated (metastable region), crystal seeds can grow, but they cannot form spontaneously (Miers & Isaac, 1907; McPherson, 1998). Metastable solutions can be induced to produce crystals by using artificial (external) seeds ("seeding"). It is possible that the situation with amyloids might be analogous. The analogy should not be treated too superficially, however, because, for instance, protein crystals can be easily dissolved, while amyloid fibrils are very durable, practically insoluble, and sometimes even described as indestructible. Assuming 3D domain swapping in amyloid formation, the explanation of this difference is obvious. In protein crystals the contacts between individual molecules are few and tenuous, and protective hydration shells shield the molecules from aggregation into amorphous precipitate. If in a crystal structure there is a case of 3D domain swapping, the crystallographic building block is in fact the domain-swapped oligomer. In a hypothetical 3D domain-swapped amyloid fibril the situation would be diametrically different. The molecules would be intertwined using (the closed) interfaces (one or more) that evolved to produce stabilizing adhesive forces, and there would be no screening water shell. Even in the framework of this simplistic scenario, we are faced with the question, how the phenomenon of 3D domain swapping, which we

have only seen within dimers (or trimers at most), could be compatible with infinite aggregation. Two recent reports that have already been mentioned provide some insight.

Multiple exchangeable domains; lessons from RNase A

The fact that RNase A has been now found by Eisenberg and colleagues to form two types of 3D domain-swapped dimers utilizing different structural segments (N-terminal helix, C-terminal β-strand), and thus two different closed interfaces, opens a fascinating possibility (Liu *et al.*, 2001). A molecule swapping both domains, each with a different neighbor (Fig. 13), would create two (different) "sticky"
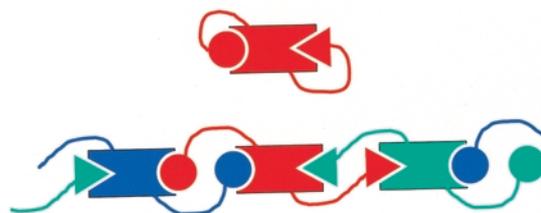

Figure 13. Schematic diagram of infinite polymerization *via* two types of closed interfaces.

A molecule capable of forming two types of dimers through the exchange of different segments, could start infinite polymerization by swapping both segments with different neighbors. This possibility exists in RNase A, which has been shown to form two types of 3D domain-swapped dimers by switching different structural elements, an N-terminal helix, or a C-terminal β-strand (Liu *et al.*, 1998; 2001).

ends and could thus start an infinite chain of swapping events leading to a polymer. This is not to say that this dual 3D domain swapping must necessarily lead only to linear polymers. In fact this type of aggregation is still compatible with closed oligomers provided they contain an even number of units. One could consider that a dimer formed through a reciprocal exchange of one domain type, for instance the triangle in Fig. 13, could close "on itself" or could pair through two domain-swapping

events (on both ends) with an identical dimer, or in fact form a cyclic oligomer with any number of triangle-swapped units. Nevertheless, the variety of possibilities opened by this discovery is very attractive from the point of view of the involvement of 3D domain swapping in amyloid formation. The implications from the major domain-swapped RNase A dimer (exchanging the C-terminal strand) for amyloidogenic mechanisms have also another aspect, as in the open interface a reinforcement reminiscent of a polar zipper was found (Liu *et al.*, 2001). Polar zippers are β-sheet-type structures formed from polyglutamine tracts (Pertuz *et al.*, 1994; Perutz, 1999). Their exceptional stability derives from the reinforcement of the main-chain β-sheet interactions by additional hydrogen bonds between the glutamine (or asparagine) side chains. Such polar zippers have been invoked by Perutz to explain the aggregation of huntingtin (and other proteins) forming neurotoxic aggregates connected genetically with expanded glutamine repeats.

Closed *versus* open association; the case of HCC

Human cystatin C is not likely to swap any other domains than the N-terminal domain seen in the currently available dimeric structure (Janowski *et al.*, 2001). In particular, the C-terminus cannot be considered because of the disulfide reinforcements in this domain. Yet HCC, and in particular its Leu68Gln variant, *is* an amyloidogenic protein. Is the domain-swapping seen in HCC a mere coincidence in this respect or does it represent the protein's fundamental ability responsible for its pathological aggregation?

As mentioned above, a hinge movement of loop L1 leads to partial unfolding of the HCC molecule into an open conformation with the α and β domains separated. Reconstruction of the α–β interactions from segments belonging to separate molecules gives rise to oligo-

merization. In the case of dimerization, the two interacting molecules recreate the two monomeric topologies in a symmetric, fireman's grip fashion, as in the crystal structure. It is, however, not very likely that such two-fold-symmetric dimers could be the first intermediate in the process of higher oligomerization. As discussed above, they are probably a suicidal trap on the oligomerization pathway, which would explain their stability and the ease with which they can be purified (Ekiel & Abrahamson, 1996; Ekiel *et al.*, 1997). Their energetic advantage may be related to the formation of the strong β-sheet interactions at the open interface, i.e. at the domain-connecting segment contributed by the two opened L1 loops. Unhampered chain-like oligomerization (Fig. 14) should start with a
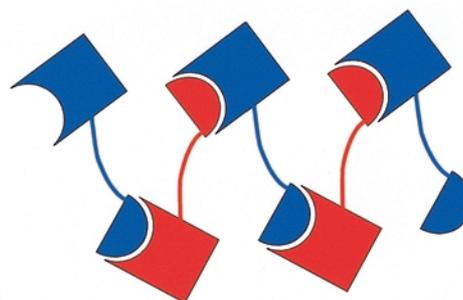

Figure 14. As an alternative to closed-ended, symmetrical dimerization (as in Fig. 1), open-ended polymerization *via* a single closed interface can also be considered.

Here, instead of producing reciprocal interactions between two molecules, the domain exchange leads from one molecule to the next, always leaving two "sticky ends" at which the polymer grows.

reconstruction, from two partially unfolded HCC monomers, of only one α–β domain, leaving the other two α and β structures available for interactions with additional monomers. One should note that partially unfolded Leu68Gln HCC monomers with largely retained secondary structure have been observed in solution by CD and NMR techniques as distinct "molten-globule"-like intermediates on the unfolding pathway (Gerhartz *et al.*, 1998).

The highly conserved sequence of the L1 loop does not suggest why it should be predisposed to destabilization. In the chicken cystatin structure (Bode *et al.*, 1988) it forms a tight five-residue β-hairpin, the central element of which, Ser56, is on the border of a generously allowed Ramachandran region. However, since this serine represents in fact a deviation from the conserved sequence, it is again not obvious that the L1 loop in monomeric HCC should be particularly unstable. The source of monomer instability may be, however, located elsewhere, for instance at the α–β interface. The experimental observation of reduced monomer stability and facilitated dimerization of the Leu68Gln mutant suggests that this interpretation may be correct. In their analysis of the different open interfaces in dimeric RNase A and BS-RNase Liu *et al.* (1998) argue that 3D domain swapping is sufficient for proteins to oligomerize but that the precise orientation of the subunits is influenced by interactions at the open interface thus lending a possibility to control the overall structure through careful mutations designed to change the nature of the open interface. These remarks are of general validity and have been illustrated above. On the other hand, the example of HCC and its Leu68Gln mutant suggests that also mutations in the closed interface have to be considered in the interplay of kinetic and thermodynamic factors controlling the formation of 3D domain-swapped oligomers. As pointed out by Perutz in a short but inspirational note (1997), mutations of "internal" residues might result in a loss in free energy of stabilization which, even if small, might lead to a disruptive, "loosening" effect on the native structure. In the case of HCC, the Leu68Gln substitution would decrease the energy necessary for the transition from the monomeric to the dimeric form by destabilizing the monomer (higher energy) and lowering the barrier of the transition state (less unfavorable interactions with solvent in the open conformation).

Although sealed by strong β-sheet hydrogen bonds, the open interface of the symmetric HCC dimers is rather small. It is thus possible that in the process of higher oligomerization a different open interface could be formed while, obviously, preserving the closed interface. It cannot be excluded that in higher oligomerization of HCC some of the packing contacts observed in the crystal structure might play a role, for instance the hydrogen-bond interactions or even the hydrophobic contacts. The van der Waals contacts are interesting because they involve the unfolded L1 loops but they are too weak and are not compatible with infinite polymerization. The large number and the nature of the hydrogen bonds operating within the octamers suggest a tempting possibility. However, although this case is energetically more plausible, these crystal-packing-type contacts can be easily replaced by interactions with water molecules. The octamers are closed units that are diffi-
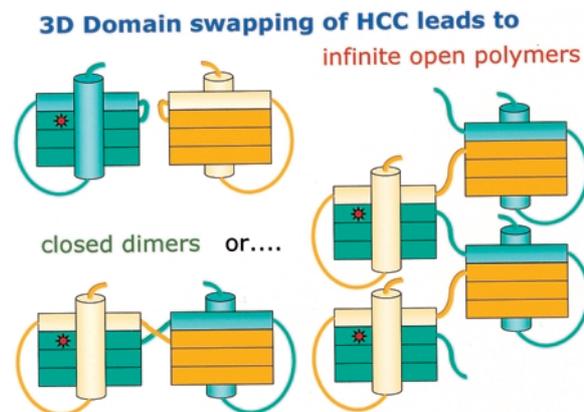

Figure 15. A cartoon illustration of human cystatin C dimerization and polymerization.

The cystatin C dimer is formed from a pair of monomers, as found in the crystal structure of 3D domain-swapped dimeric HCC (left). In an open-ended variant, the same mechanism of 3D domain swapping would lead to infinite polymerization of HCC, as in amyloid fibrils (right, hypothetical). The red asterisk marks the location of the Leu68Gln mutation, which favors dimerization and leads to massive aggregation of the mutated protein, resulting in amyloidosis, brain hemorrhage, and death of patients with this genetic defect.

cult to reconcile with repetitive and infinite aggregation, as is required for the formation of amyloid fibrils. It is possible that a system of similar β-sheet...back-side-loops interactions between HCC monomers or closed dimers could play a role in infinite aggregation. However, in spite of the relatively high number of hydrogen bonds per one such contact area (nine), it is rather doubtful whether this mode of association could produce, by itself, stable aggregates. The most likely scenario for HCC polymerization is open-ended exchange of the same domains as in the 3D domain-swapped dimer, as schematically illustrated in Fig. 15, where the molecules, instead of reciprocating the interactions, propagate the swap in a helical fashion. From the current structure it is difficult to predict, however, what would be the open interface that would stabilize the polymer.

———————

REFERENCES

Abrahamson, M., Barrett, A.J., Salvesen, G. & Grubb, A. (1986) Isolation of six cysteine proteinase inhibitors from human urine. Their physicochemical and enzyme kinetic properties and concentrations in biological fluids. *J. Biol. Chem.* **261**, 11281–11289.

Abrahamson, M. & Grubb, A. (1994) Increased body temperature accelerates aggregation of the Leu→Gln mutant cystatin C, the amyloid-forming protein in hereditary cystatin C amyloid angiopathy. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1416–1420.

Adinolfi, S., Piccoli, R., Sica, F. & Mazzarella, L. (1996) BS-RNase tetramers: An example of domain-swapped oligomers. *FEBS Lett.* **398**, 326–332.

Alvarez-Fernandez, M., Barrett, A.J., Gerhartz, B., Dando, P.M., Ni, J. & Abrahamson, M. (1999) Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. *J. Biol. Chem.* **274**, 19195–19203.

Barrett, A.J., Fritz, H., Grubb, A., Isemura, S., Jarvinen, M., Katunuma, N., Machleidt, W., Muller-Esterl, W., Sasaki, M. & Turk, V. (1986) Nomenclature and classification of the proteins homologous with the cysteine-proteinase inhibitor chicken cystatin. *Biochem. J.* **236**, 312.

Beintema, J.J., Wietzes, P., Weickmann, J.L. & Glitz, D.G. (1984) The amino acid sequence of human pancreatic ribonuclease. *Anal. Biochem.* **136**, 48–64.

Bennett, M.J., Choe, S. & Eisenberg, D.S. (1994) Domain swapping: Entangling alliances between proteins. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3127–3131.

Bennett, M.J., Schlunegger, M.P. & Eisenberg, D. (1995) 3D Domain swapping: A mechanism for oligomer assembly. *Protein Sci.* **4**, 2455–2468.

Bjarnadottir, M., Nilsson, C., Lindstrom, V., Westman, A., Davidsson, P., Thormodsson, F., Blondal, H., Gudmundsson, G. & Grubb, A. (2001) The cerebral hemorrhage-producing cystatin C variant (L68Q) in extracellular fluids. *Amyloid* **8**, 1–10.

Blake, C. & Serpell, L. (1996) Synchrotron X-ray studies suggest that the core of the transthyretin amyloid fibril is a continuous β-sheet helix. *Structure* **4**, 989–998.

Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshnikov, A., Brzin, J., Kos, J. & Turk, V. (1988) The 2.0 Å X-ray crystal structure of chicken egg white cystatin and its possible

mode of interaction with cysteine proteinases. *EMBO J.* **7**, 2593–2599.

Booth, D.R., Sunde, M., Bellotti, V., Robinson, C.V., Hutchinson, W.L., Fraser, P.E., Hawkins, P.N., Dobson, C.M., Radford, S.E., Blake, C.C. & Pepys, M.B. (1997) Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **385,** 787–793.

Chiti, F., Taddei, N., Bucciantini, M., White, P., Ramponi, G. & Dobson, C.M. (2000) Mutational analysis of the propensity for amyloid formation by a globular protein. *EMBO J.* **19**, 1441–1449.

Cohen, A.S. (1986) General introduction and a brief history of the amyloid fibril; in *Amyloidosis* (Marrink, J. & Van Rijvijk, M.H., eds.) pp. 3–19, Nijhoff, Dordrecht.

Cohen, A.S. & Calkins, E. (1959) Electron microscopic observation on a fibrous component in amyloid of diverse origins. *Nature* **183**, 1202–1203.

Cohen, A.S., Shirahamata, T. & Skinner, M. (1982) Electron microscopy of amyloid; in *Electron Microscopy of Proteins* (Harris, J.R., ed.) vol. 3, pp. 165–205, Academic Press, New York.

Cohen, F.E. & Prusiner, S.B. (1998) Pathologic conformations of prion proteins. *Annu. Rev. Biochem.* **67**, 793–819.

Crestfield, A.M., Stein, W.H. & Moore, S. (1962) On the aggregation of bovine pancreatic ribonuclease. *Arch. Biochem. Biophys.* **1** (Suppl.), 217–222.

Crestfield, A.M., Stein, W.H. & Moore, S. (1963) Properties and conformation of the histidine residues at the active site of ribonuclease. *J. Biol. Chem.* **238**, 2421–2428.

D'Alessio, G. (1999) Evolution of oligomeric proteins. *Eur. J. Biochem.* **266**, 699–708.

Dieckmann, T., Mitschang, L., Hofmann, M., Kos, J., Turk, V., Auerswald, E.A., Jaenicke, R. & Oschkinat, H. (1993) The structure of native phosphorylated chicken cystatin and of recombinant unphosphorylated variant in solution. *J. Mol. Biol.* **234**, 1048–1059.

Dobson, C.M. (1999) Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **9**, 329–332.

Ekiel, I. & Abrahamson, M. (1996) Folding-related dimerization of human cystatin C. *J. Biol. Chem.* **271**, 1314–1321.

Ekiel, I., Abrahamson, M., Fulton, D.B., Lindahl, P., Storer, A.C., Levadoux, W., Lafrance, M., Labelle, S., Pomerleau, Y., Groleau, D., LeSauteur, L. & Gehring, K. (1997) NMR structural studies of human cystatin C dimers and monomers. *J. Mol. Biol.* **271**, 266–277.

Engh, R.A., Dieckmann, T., Bode, W., Auerswald, E.A., Turk, V., Huber, R. & Oschkinat, H. (1993) Conformational variability of chicken cystatin. Comparison of structures determined by X-ray diffraction and NMR spectroscopy. *J. Mol. Biol.* **234**, 1060–1069.

Gerhartz, B., Ekiel, I. & Abrahamson, M. (1998) Two stable unfolding intermediates of the disease-causing L68Q variant of human cystatin C. *Biochemistry* **37**, 17309–17317.

Glenner, G.G. (1980a) Amyloid deposits and amyloidosis — the β-fibrilloses 1. *N. Eng. J. Med.* **302**, 1283–1292.

Glenner, G.G. (1980b) Amyloid deposits and amyloidosis — the β-fibrilloses 2. *N. Eng. J. Med.* **302**, 1333–1343.

Glenner, G.G., Eanes, E.D. & Page, D.L. (1972) The relation of the properties of congo red-stained amyloid fibrils to the β-conformation. *J. Histochem. Cytochem.* **20**, 821–826.

Grubb, A. (2000) Cystatin C — properties and use as diagnostic marker. *Adv. Clin. Chem.* **35**, 63–99.

Janowski, R., Kozak, M., Jankowska, E., Grzonka, Z., Grubb, A., Abrahamson, M. & Jaskolski, M. (2001) Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. *Nat. Struct. Biol.* **8**, 316–320.

Klafki, H.-W., Pick, A.I., Pardowitz, I., Cole, T., Awni, L.A., Barnikol, H.V., Mayer, F., Kratzin, H.D. & Hilschmann, N. (1993) Reduction of disulfide bonds in an amyloidogenic Bence Jones protein leads to formation of "amyloid-like" fibrils *in vitro. Biol. Chem. Hoppe-Seyler* **372**, 1117–1122.

Kozak, M., Jankowska, E., Janowski, R., Grzonka, Z., Grubb, A., Alvarez-Fernandez, M., Abrahamson, M. & Jaskólski, M. (1999) Expression of a selenomethionyl derivative and preliminary crystallographic studies of human cystatin C. *Acta Crystallogr.* D**55**, 1939–1942.

Liu, Y., Hart, P.J., Schlunegger, M.P. & Eisenberg, D. (1998) The crystal structure of a 3D domain-swapped dimer of RNase A at 2.1 Å resolution. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3437–3442.

Liu, Y., Gotte, G., Libonati, M. & Eisenberg, D. (2001) A domain-swapped RNase A dimer with implications for amyloid formation. *Nat. Struct. Biol.* **8**, 211–214.

Mazzarella, L., Capasso, S., Demasi, D., Di Lorenzo, G., Mattia, C.A. & Zagari, A. (1993) Bovine seminal ribonuclease: Structure at 1.9 Å resolution. *Acta Crystallogr.* D**49**, 389–402.

McPherson, A. (1998) *Crystallization of Biological Macromolecules*. Cold Spring Harbor Laboratory Press, New York.

Miers, H.A. & Isaac, F. (1907) The spontaneous crystallization of binary mixtures: Experiments on salol and betol. *Proc. Roy. Soc. Lond. A* **79**, 322.

Minor, D.L., Jr. & Kim, P.S. (1996) Context-dependent secondary structure formation of a designed protein sequence. *Nat.* **380**, 730–734.

Murray, A.J., Head, J.G., Barker, J.J. & Brady, R.L. (1998) Engineering an intertwined form of CD2 for stability and assembly. *Nature Struct. Biol.* **5**, 778–782.

Murray, A.J., Lewis, S.J., Barclay, A.N. & Brady, R.L. (1995) One sequence, two folds: A metastable structure of CD2. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7337–7341.

Ogihara, N.L., Ghirlanda, G., Bryson, J.W., Gingery, M., DeGardo, W.F. & Eisenberg, D. (2001) Design of three-dimensional domain-swapped dimers and fibrous oligomers. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1404–1409.

Olafsson, I. & Grubb, A. (2000) Hereditary cystatin C amyloid angiopathy. *Amyloid* **7**, 70–79.

Park, C. & Raines, R.T. (2000) Dimer formation by a "monomeric" protein. *Protein Sci.* **9**, 2026–2033.

Pei, X.Y., Holliger, P., Murzin, A.G. & Williams, R.L. (1997) The 2.0-Å resolution crystal structure of a trimeric antibody fragment with noncognate V_H–V_L domain pairs shows a rearrangement of V_H CDR3. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9937–9642.

Perutz, M.F. (1997) Mutations make enzymes polymerize. *Nature* **385**, 773–774.

Perutz, M.F. (1999) Glutamine repeats and neurodegenerative diseases: Molecular aspects. *Trends. Biochem. Sci.* **24**, 58–63.

Perutz, M.F., Johnson, T., Suzuki, M. & Finch, J.T. (1994) Glutamine repeats as polar zippers: Their possible role in inherited neurodegenerative diseases. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5355–5358.

Piccoli, R., Tamburrini, M., Piccialli, G., Di Donato, A., Parente, A. & D'Alessio, G. (1992) The dual-mode quaternary structure of seminal RNase. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1870–1874.

Piccoli, R., Di Gaetano, S., De Lorenzo, C., Grauso, M., Monaco, C., Spalletti-Cernia, D., Laccetti, P., Cinatl, J., Matousek, J. & D'Alessio, G. (1999) A dimeric mutant of human pancreatic ribonuclease with selective cytotoxicity toward malignant cells. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7768–7773.

Rawlings, N.D. & Barrett, A.J. (1990) Evolution of proteins of the cystatin superfamily. *J. Mol. Evol.* **30**, 60–71.

Schlunegger, M.P., Bennett, M.J. & Eisenberg, D. (1997) Oligomer formation by 3D domain swapping: A model for protein assembly and misassembly; in *Advances in Protein Chemistry* (Richards, F.M., Eisenberg, D.S. & Kim, P.S., eds.) vol. 50, pp. 61–122, Academic Press, New York.

Sipe, J.D. & Cohen, A.S. (2000) Review: history of the amyloid fibril. *J. Struct. Biol.* **190**, 88–98.

Sunde, M. & Blake, C.C.F. (1998) From the globular to the fibrous state: Protein structure and structural conversion in amyloid formation. *Quat. Rev. Biophys.* **31**, 1–39.

Sunde, M., Serpell, L.C., Bartlam, M., Fraser, P.E., Pepys, M.B. & Blake, C.C.F. (1997) Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J. Mol. Biol.* **273**, 729–739.

Teplow, D.B. (1998) Structural and kinetic features of amyloid *β*-protein fibrillogenesis. *Amyloid* **5**, 121–142.

Wlodawer, A., Bott, R. & Sjolin, L. (1982) The refined crystal structure of ribonuclease A at 2.0 Å resolution. *J. Biol. Chem.* **257**, 1325–1332.

Wlodawer, A., Svansson, L.A., Sjolin, L. & Gilliland, G.L. (1988) Structure of phosphate-free ribonuclease A refined at 1.26 Å resolution. *Biochemistry* **27**, 2705–2717.

NOTE ADDED IN PROOF

At the time this manuscript was sent for publication, a paper was published by Staniforth *et al.* (*EMBO J.* **20**, 4774–4781 (2001)) who show, using NMR spectroscopy, that human cystatin A and chicken cystatin (both closely related to human cystatin C) form dimers *via* the same 3D domain swapping mechanism described in the crystal structure of HCC. The solution dimers have the same closed interface but the conformation of the open-interface-forming region of loop L1 is different although the dimers are still two-fold symmetric. Perhaps even more exciting is the discovery, reported very recently by Knaus *et al.* (*Nat. Struct. Biol.* **8**, 770–774 (2001)), that the human prion protein is capable of dimer formation and that those dimers also arise *via* exchange of three-dimensional domains. The rapidly accumulating structural evidence strongly suggests that 3D domain swapping may indeed be involved in amyloid formation.

Knaus, K.J., Morillas, M., Swietnicki, W., Malone, M., Surewicz, W.K. & Yee, V.C. (2001) Crystal structure of the human prion protein reveals a mechanism for oligomerization. *Nat. Struct. Biol.* **8**, 770–774.

Staniforth, R.A., Giannini, S., Higgins, L.D., Conroy, M.J., Hounslow, A.M., Jerala, R., Craven, C.J. & Waltho, J.P. (2001) Three-dimensional domain swapping in the folded and molten-globule states of cystatins, an amyloid-forming structural superfamily. *EMBO J.* **20**, 4774–4781.