

# 3D domain swapping – implications for conformational disorders and ways of control

MARIUSZ JASKÓLSKI

Department of Crystallography, Faculty of Chemistry, Adam Mickiewicz University in Poznań, Poznań, Poland  
 Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland  
 e-mail: mariuszj@amu.edu.pl

## Abstract

3D Domain swapping is a mechanism of protein aggregation, in which a structural element of a protein fold is replaced by an identical element from another subunit. Some proteins, for instance RNase A, can assume many domain-swapped forms, thus undermining the dogma, “one sequence – one structure” in a particularly spectacular way. Completed in a mutual fashion, it is a mechanism of protein oligomerization. In an open-ended fashion, 3D domain swapping could be a mechanism of amyloid fibril formation. In another mechanism, possibly operating together with domain swapping, a specific sequence, such as a glutamine expansion, could form a  $\beta$ -spine of the fibril in a motif called steric zipper. The first connection between 3D domain swapping and amyloidogenicity was established in human cystatin C (HCC), the second - in the prion protein (PrP). In both cases, a disulfide bridge (natural in PrP, engineered in HCC) can be used for redox control of 3D domain swapping and to demonstrate that it is indeed involved in amyloid fibril formation. HCC has a naturally occurring L68Q mutant with drastically increased propensity for aggregation. The L68Q mutation occurs at the closed interface, or protein core. Mutations in other areas, such as the flexible hinge (especially deletions and insertions) can also be used to control 3D domain swapping and aggregation. Paradoxically, 3D domain swapping can also be used by Nature for prevention of nucleation processes that lead to high-order aggregates or crystals. Such a situation exists in the eye lens, where despite astronomical concentration of crystallins, the solution remains clear. One of the Nature’s tricks to achieve polydispersity is to use a palindromic sequence for the swapped domain, thereby frustrating the growth of aggregates by constantly changing the interaction topology.

**Key words:** protein aggregation, mutagenesis, amyloid, fibril, cross- $\beta$  structure, steric zipper, amyloyme, ribonuclease A, human cystatin C, prion protein, crystallins

## Introduction

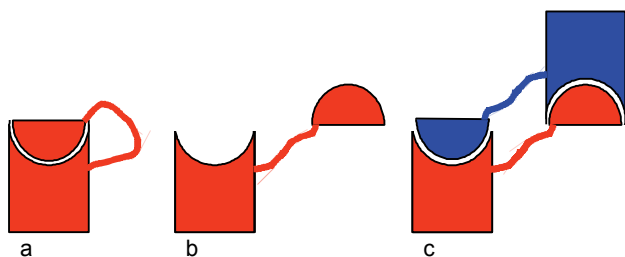
In his Nobel-winning hypothesis, the biochemist Christian Anfinsen postulated that the native, biologically active conformation of a polypeptide chain represents a kinetically accessible unique free energy minimum (Anfinsen, 1973). This hypothesis is often simplified in the form of Anfinsen’s Dogma, “one sequence – one structure”. It is a very useful dogma, helping us to make sense of the structural principles of protein function, but there is growing evidence that – at least in its trivialized form – it is not always true. We now know many examples of protein chains that can fold in more than one way. There are even general mechanisms underscoring such cases, one of them being the phenomenon of three-dimensional (3D) domain swapping. It is quite ironic that the flagship of 3D domain swapping is

ribonuclease A (RNase A), the very protein used by Anfinsen to work out his dogma.

## Definition of terms

Officially the phenomenon of 3D domain swapping was discovered, and the term coined, by Bennett et al. (1994) based on crystallographic studies of diphtheria toxin. However, its existence had been deduced much earlier from the observation that enzymatic activity of bovine pancreatic RNase A (an enzyme with two catalytic histidine residues, an N-terminal one and a C-terminal one) could be recovered in mixtures of two inactive mutant proteins with partly knocked-out active sites (Crestfield et al., 1962, 1963). This prediction was later confirmed experimentally in a crystallographic study (Liu et al., 1998), which showed a dimeric molecule with two

monomer-like folding units, each formed from structural elements contributed by both polypeptide chains. Specifically, the N-terminal helix, which in monomeric RNase A is nested in the fold of the C-terminal part of the molecule, switched places with the N-terminal helix from the other molecule and docked in the “wrong” C-terminal domain. The mutual character of this exchange has led to a symmetric, or closed dimer. This is the general scenario of 3D domain swapping: a protein domain breaks its contact with other structural elements and its place is taken by the same domain from another molecule, leading to an interlaced dimer or higher-order oligomer (Fig. 1). The swapped domain could be a complete globular domain but in most cases (including RNase A) it can be just a simple secondary-structure element, such as a terminal  $\alpha$ -helix or  $\beta$ -chain. It is obvious that a protein chain capable of 3D domain swapping must undergo at least partial unfolding, which breaks, and transiently exposes, an intramolecular interface between the domains. This closed interface is then recreated in an aberrant fashion from fragments of different subunits. Therefore, the closed interface is preserved in the oligomer. However, the oligomer also contains a novel feature, the open interface between the subunits, which is absent in the monomer. It is obvious from this description of 3D domain swapping that the protein chain must be capable of assuming two conformations. One is called “closed” (in monomeric form), the other – “open” (in the oligomeric form). The closed and open conformations are almost the same and differ in only one area, a loop or a similar “hinge” element that links the swapped domain with the rest of the polypeptide, and whose movement leads to the transition.



**Fig. 1.** In preparation for 3D domain swapping, a protein molecule (a) must undergo partial unfolding (b) via a conformational change of a flexible hinge element, to expose the closed interface, characteristic for the separated domains. Two unfolded molecules can mutually recognize the complementary surfaces, recreating the interface in a symmetrical, dimeric fashion (c). The 3D domain swapped dimer is not a simple sum of the two monomeric molecules, as it contains a unique new feature, the open interface

The open interface can have more than one form. For instance, a closely related bovine seminal ribonuclease (BS-RNase) forms native 3D domain-swapped dimers (Mazzarella et al., 1993). Although the swapped domain is the same as in the above RNase A case (N-terminal  $\alpha$ -helix), the overall organization of the dimers is different. However, RNase A has provided even more astonishing examples. First, Liu et al. (2001) showed that the protein can also form dimers with a 3D swap of the C-terminal  $\beta$ -strand. Later, the same C-terminal swap was found in circular trimers (Liu et al., 2002). An even more exciting case was inspired by the discovery of a linear trimer, in which the central molecule has exchanged different domains with its neighbors, creating a possibility for an infinite, open-ended polymerization (Liu et al., 2002).

Although at first regarded as a marginal scientific curiosity, 3D domain swapping has been now documented in more than 50 proteins. It appears that the number of cases could be much higher if the right conditions were found, including pH change, mild denaturation, or other suitable environmental factor. In some cases, a slight modification of the sequence, even a point mutation, can drastically influence the ability of a protein to undergo 3D domain swapping. In a strict definition given by Eisenberg (Bennett et al., 1994), bona fide 3D domain swapping requires experimental observation of the same protein chain in a monomeric fold and as an oligomer with exchanged domains. In many cases, described as quasi 3D domain swapping, we know the structure of the oligomer but have only indirect evidence about the existence of the monomeric form, or the monomer is a slight variant of the protein. For example, the case of human cystatin C (HCC) belongs to the latter category, because the protein has been crystallized in several forms but invariably as domain-swapped dimers (Janowski et al., 2001, 2004, 2005). It is, however, clear that monomeric HCC exists because only in this form the protein can function as a physiological inhibitor of pain-like cysteine proteases.

### Examples of possible roles of 3D domain swapping

The increasing frequency of 3D domain swapping has prompted questions if it might be of biological significance. One thing is certain, namely that it is an ideal, and very elegant, mechanism of protein oligomerization.

It could be utilized, for example, for the stabilization of the building blocks of virus coats. Another possibility exists for enzymatic proteins, which could utilize 3D domain swapping for functional regulation, in particular in allostery, which involves multiple active sites. 3D domain swapping could also rescue protein function after a deleterious mutation, as was illustrated with the case of RNase A. So far, however, it appears that 3D domain swapping is predominantly connected with molecular pathology, through its involvement in conformational disorders, of which amyloid aggregation is perhaps most important.

### 3D Domain swapping of human cystatin C

3D Domain swapping was proposed initially as an intellectually attractive mechanism for amyloid aggregation (Klafki et al., 1993; Bennet et al., 1995; Cohen and Prusiner, 1998) but unquestionable experimental evidence is not easy to find. The first experimental hint linking 3D domain swapping and amyloid came from crystallographic studies of HCC, which showed a dimeric assembly formed through a domain switch involving the N-terminal part ( $\beta 1$ - $\alpha$ - $\beta 2$  element) of the protein. In its monomeric form, the 120-residue human cystatin C plays a protective role in all body fluids, with a particularly high concentration in the cerebrospinal fluid. There is, however, a naturally occurring variant of HCC, found endemically in Icelandic population, where a single-residue mutation, L68Q, leads to a dramatic change of the protein's behavior. L68Q HCC is not only spontaneously converted to a dimeric form, but very easily aggregates, forming insoluble amyloid deposits in the vasculature of the central nervous system, leading to lethal cerebral hemorrhages in young adult life. The disease, endemic to Iceland, is known as Hereditary Cystatin C Amyloid Angiopathy (HCCAA) (Olafson and Grubb, 2000). Evidently, the amyloidogenic properties of L68Q HCC are greatly increased. In addition, in a cruel thermodynamic experiment by Nature, the condition of HCCAA patients is apparently worsened if they suffer from bouts of fever. It has to be stressed that in controlled *in vitro* conditions it is also possible to form amyloid fibrils from wt HCC. One of the tricks is to use elevated temperature. There are also reports suggesting that wt HCC may be involved in amyloid fibril formation in other deposition diseases.

Because of the absence of structural information about monomeric HCC, the structure of chicken cystatin had to be used as a template. That model, consisting of

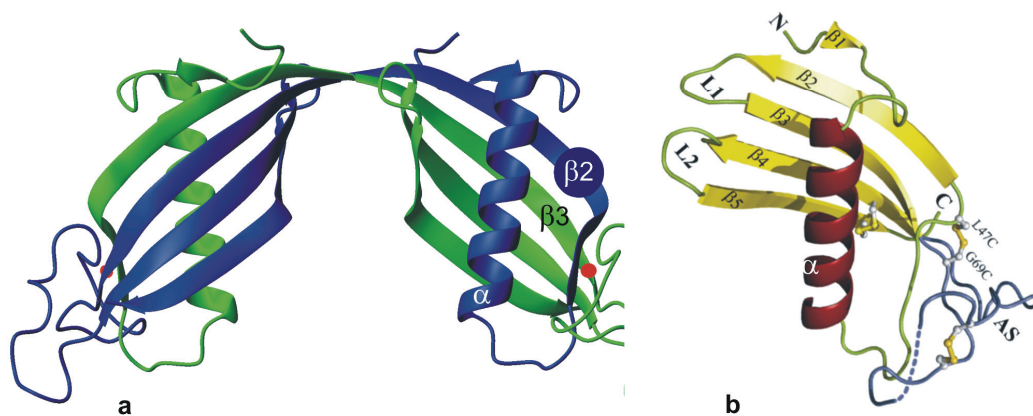
a five-stranded antiparallel  $\beta$ -sheet gripped around a long  $\alpha$ -helix, has the following topology: N- $\beta 1$ - $\alpha$ - $\beta 2$ -L1- $\beta 3$ -AS- $\beta 4$ -L2- $\beta 5$ -C, with the enzyme-binding epitope formed by the N-terminus and two  $\beta$ -hairpin loops L1 and L2, all aligned in a wedge-like fashion at one end of the molecule. The opposite end consists of a system of poorly ordered "back-side" loops designated as the appending structure (AS) – Figure 2.

The open HCC molecule, which is interlaced with another identical subunit in the 3D domain-swapped dimer, arises by a conformational change of loop L1, which effectively becomes part of a very long  $\beta$ -chain extending from the N-terminus of  $\beta 2$  to the C-terminus of  $\beta 3$ . In the dimer, two copies of this extended  $\beta$  structure face each other, creating an extremely long antiparallel  $\beta$ -sheet, with as many as 34 intermolecular main-chain hydrogen bonds (Fig. 2). Part of this structure is the newly created open interface. Considering also the effect of cooperativity, the consequence of the open interface on the dimer stability could be quite appreciable. The fact that in dimeric HCC the L1 loop has disappeared explains why in this form the protein has no inhibitory effect on papain-like proteases. The conformation of the 3D domain-swapped dimer of HCC clearly showed (Janowski et al., 2001) two chicken cystatin-type folding units, confirming that the closed interface, and indeed the entire fold, of (unknown for a number of years) monomeric HCC has been faithfully recreated. The open monomer, however, if extracted from its dimeric context, looks very bizarre and certainly cannot represent a stable protein fold. We can safely accept that the protein chain assumes this conformation only transiently, on its transition from the monomeric to the oligomeric form. The open conformation cannot be stable in aqueous environment because it has the hydrophobic core of the closed interface exposed to solvent.

3D Domain-swapped HCC has been characterized in several crystal forms. They differ to a variable degree with regard to the conformation of the "loop L1" hinge. Although this does affect the mutual orientation of the two cystatin-like folding units, the internal organization of these units is unchanged, once again underscoring the correctness, and stability, of the recreated fold.

### Molecular background of L68Q HCC amyloidosis

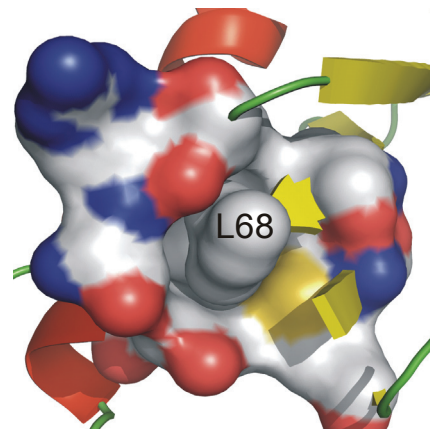
It would be most instructive to analyze the environment of leucine 68 in the monomeric fold of HCC to investigate the reasons why its mutation to glutamine has



**Fig. 2.** Dimeric human cystatin C (HCC) with 3D swapped domains (a). The red dot indicates the location of L68. Its substitution by glutamine leads to severe cerebral amyloidosis and early death of persons carrying this mutation. The monomeric fold of HCC (b) is now known thanks to the stabilizing L47C/G69C double cysteine mutation (stab1), which introduces an extra S-S bond tethering the two  $\beta$ -strands ( $\beta 2$  and  $\beta 3$ ) that must separate on 3D domain swapping. The papain-binding epitope of this inhibitory molecule is formed by the N-terminus and loops L1 and L2. Since the dimer of HCC (a) is formed upon a conformational change of loop L1, it has no inhibitory activity. AS (appending structure) denotes a system of poorly ordered back-side loops

such a dramatic effect on the protein's behavior, and in consequence on the prognosis for the carriers of this mutation. L68 is located on strand  $\beta 3$ , where its side chain protrudes from the concave face of the  $\beta$ -sheet into the hydrophobic protein core at the closed interface formed between the  $\alpha$ -helix and the  $\beta$ -sheet. Assuming that the closed interface is faithfully recreated upon 3D domain swapping, it should be even possible to conclude about the role of the amino acid residue in position 68 from the crystal structure of dimeric HCC. We see the leucine side chain nearly perfectly nested in a hydrophobic cavity (Fig. 3) that is formed by the surrounding residues from the  $\beta$ -sheet and, importantly, by hydrophobic residues (V31, Y34) at the N-terminal end of the  $\alpha$ -helix. An attempt to replace L68 with glutamine, would force into the cavity a side chain that is too big for it and has a completely incompatible, highly polar (amide) group at its end. This would highly destabilize the protein core, leading in consequence to its disintegration in the vicinity of Q68. In addition, the too-long side chain would be pushing on the N-terminal end of the  $\alpha$ -helix, forcing it through a lever effect to separate from the  $\beta$ -sheet. This is what essentially happens in the act of partial unfolding that must precede the domain-swapping event. It is therefore quite logical that L68Q HCC dimerizes much more readily than the wild-type protein. It should be noted that glutamine in position 68 would not only destabilize the closed interface but, upon its separation, would make the exposed closed interface more com-

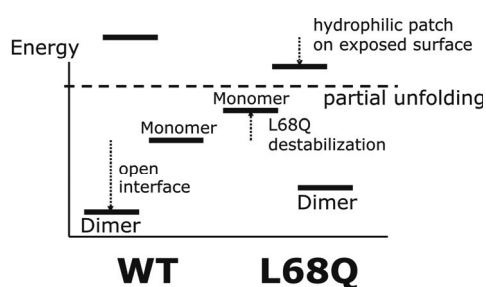
patible with aqueous environment because of the hydrophilic character of the Q68 side chain.



**Fig. 3.** L68 in a hydrophobic cavity formed at the closed interface of HCC. The view is from the concave face of the  $\beta$ -sheet toward the  $\alpha$ -helix. Figure prepared by Dr. Robert Kolodziejczyk

### Thermodynamics of HCC 3D domain swapping

The above considerations can be used to construct a speculative diagram illustrating the thermodynamic relations in HCC dimer-formation equilibria (Jaskolski, 2001) – Figure 4. The energy required for unfolding of wt HCC is probably quite high and such events are not very frequent in normal conditions. An unfolded (open) monomer could refold back regenerating the monomeric molecule, or it could “react” with another open monomer to form a 3D domain-swapped dimer.



**Fig. 4.** Speculative thermodynamics of wild-type (left) and L68Q (right) HCC unfolding and dimerization. The free energy levels are only qualitative, meant to illustrate the relative relations among the various forms of the protein chain.

For explanations, see text

The dimer would be much more stable than the monomer because of the extra stabilization energy contributed by the newly formed open interface. (It should be noted, however, that the entropic contribution of the dimerization process would be always unfavorable, thus increasing the free energy of the oligomeric species.) In this scenario, the increased energy gap between the dimeric form and the open monomer would be sufficiently large to effectively exclude the possibility of dissociation of the dimers in normal conditions. In other words, the dimeric state would be a sink (or a “black hole”) from which there would be no return to the monomeric species. The situation with the L68Q mutant would be quite different. The energy of the closed monomer must be higher because of the destabilization introduced by the incompatible Q68 side chain in the hydrophobic core. On the other hand, the open Q68 HCC monomer would have lower energy than the corresponding wt form because of the favorable solvent interactions of the hydrophilic Q68 side chain exposed to water. In consequence, the energy required for the closed→open transition could be quite low, leading to a relatively easy crossing of this barrier on the path from the closed monomer to the dimer. A 3D domain-swapped dimer would be again stabilized by virtue of the open interface. Although it would be less stable than the wt dimer on account of the destabilization introduced by the Q68 side chain, its energetic separation from the open monomer could be sufficiently large to effectively eliminate transitions to the monomeric form.

### Cysteine mutations stabilizing monomeric HCC

The observation of HCC in crystal structures exclusively in the dimeric form has been quite frustrating. The

need to structurally characterize HCC in its monomeric form was dictated by a desire to confirm the predictions based on the dimeric fold and to precisely map the enzyme-binding epitope as a template for rational design of effective inhibitors of papain-like proteases (Grzonka et al., 2001). Ultimately, one would also like to test our approaches aimed at stabilizing the monomeric fold of HCC, in an effort to stop the progression of HCCAA. If 3D domain swapping is indeed at the root of all the pathological aggregation, then its abrogation should stabilize the monomeric form of HCC and stop its aggregation. 3D Domain swapping could be prevented if the structural elements of the protein that get separated upon unfolding could be tethered by a covalent bond. The simplest way of introducing this measure is a new S-S bond between strategically placed cysteine residues. With this goal in mind, two cysteine double mutants were designed. One, L47C/G69C termed stab1, in which a disulfide bond would link strands  $\beta$ 2 and  $\beta$ 3, and one, F29C/M110C termed stab2, in which an S-S tether would rivet the  $\alpha$ -helix to the  $\beta$ -sheet. It was very gratifying to see in the MS spectra of tryptic peptide maps of recombinant HCC stab1 and stab2 proteins that the new disulfide bridges are indeed formed as desired.

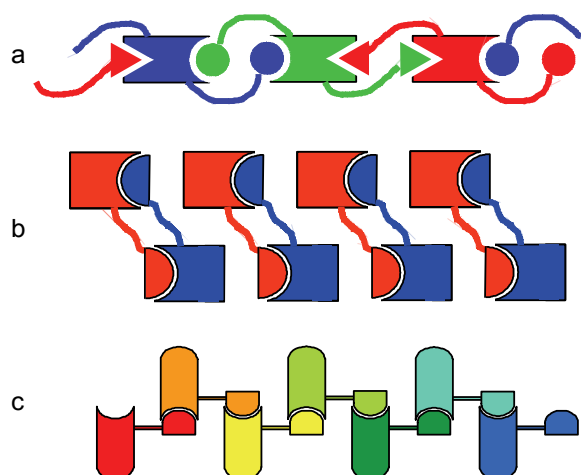
The HCC stab1 protein could be crystallized and its structure has been solved (Kolodziejczyk et al., 2010) revealing, as expected, a monomeric protein with a fold exactly as in the folding units of the domain-swapped species (Fig. 2), without any distortion in the area of the new disulfide bond. In particular, the location of the side chain of L68 in a hydrophobic cavity was fully confirmed. In addition, comparisons of HCC stab1 with other monomeric cystatins, in particular with chicken cystatin, used previously as models for monomeric HCC, revealed that the existing picture of the N/L1/L2 enzyme-binding epitope was not quite accurate.

The stab mutations were introduced not only into the wild-type protein but also to the L68Q form. As expected, the stab mutations have completely suppressed dimerization of the modified proteins. For example, wt HCC is largely converted into dimeric form upon a week-long incubation at mildly denaturing conditions (0.5 M guanidinium chloride). HCC stab1 and stab2 remain monomeric even on much longer incubation in similar conditions. L68Q HCC shows a quick transition into the dimeric state even in physiological conditions but this transformation is completely blocked with the stab1 and stab2 mutations.

In a similar way, the stab mutations prevented the formation of HCC amyloid fibrils. One can produce artificial fibrils even from wt HCC, by incubating it for prolonged periods of time at elevated temperature (48 °C) at low pH (2.0) with vigorous stirring. Those fibrils can be quantified by measuring the change in thioflavine T fluorescence intensity. This assay shows very minimal amounts of amyloid fibrils formed by stab1 and stab2 HCC even after one month of incubation (Nilsson et al., 2004).

### 3D Domain swapping and amyloid fibril formation of HCC

From the above experiment it is obvious that 3D domain swapping must be involved in HCC amyloid fibril formation since its abrogation also suppresses the fibril formation process. However, this experiment does not provide a clear picture of *how* the domain swapping is involved at the molecular level.



**Fig. 5.** Possible scenarios for 3D domain swapping-driven amyloid aggregation. A protein, such as RNase A, capable of swapping different structural elements at its two termini, could start forming an amyloid fibril by initiating two different domain swapping acts with its two immediate neighbors (a). 3D domain swapping could be utilized only for the construction of dimeric building blocks, which would subsequently aggregate to form a fiber via a different mechanism, for instance by intermolecular  $\beta$ -sheet stacking (b). Intellectually the most satisfying is however the open-ended runaway model, in which molecule A swaps a domain into protein B core, which swaps the same domain in protein C core, etc. (c). It is now clear that HCC forms amyloid fibrils using the propagated model (c)

What we need to explain here is how a phenomenon, which we know only from symmetrical, closed-ended situations, can generate an open-ended, linear polymer.

At least three mechanisms could be proposed. In one of them, the protein would be exchanging different domains with its immediate neighbors (Fig. 5a). While this is a possibility with RNase A, it is not likely in the case of HCC where, despite numerous studies, only one mode of domain exchange has been seen.

In another scenario, 3D domain swapping would be only utilized for the creation of dimeric molecules, which would then aggregate as building blocks of the ultimate fibril using a different mechanism, for example intermolecular  $\beta$ -sheet formation, which has been actually seen several times in the crystal structures of dimeric HCC (Fig. 5b). While it is consistent with the experimental facts, this scenario is intellectually less satisfactory because it invokes two different mechanisms (3D domain swapping and intermolecular  $\beta$ -sheet association) to explain one phenomenon (fibril formation).

In the third scenario, 3D domain swapping would be operating in an open-ended, run-away fashion (Fig. 5c). Here, molecule A inserts its N-terminal domain in the core of molecule B, which inserts its N-terminal domain in molecule C, etc. The two ends of a linear fibril growing in this way would be sticky, i.e. the fibril could continue growing at both ends simultaneously. What would be the priming event that would divert the protein from oligomerization into inert dimers to a path leading to infinite polymerization? Probably, high prevalence of open monomers, which from time to time could dimerize in an open-ended fashion instead of as the inert closed form. If through a quick accretion of new monomeric units at the sticky ends of the growing monomer, it would be extended to a length that would make mutual association of its ends unlikely – a nucleus, or seed, for fibrillation would be formed. Alternatively, a seed could be added externally, as has been well described with other amyloidogenic proteins. The two phenomena involving seeding or nucleation described above, make the fibril growth process very similar to protein crystallization, which also involves a phase of nucleation and of nucleus growth.

To resolve the ambiguity of HCC fibril formation, which could proceed according to scenario number 2 or 3, a set of experiments has been conducted with control of redox conditions (Wahlbom et al., 2007). First, we note that it is possible to detect the presence of high-order oligomers using SDS-PAGE electrophoresis. For instance, if HCC aggregates formed on high-tempera-

ture, low-pH incubation are loaded onto the stacking gel, they will practically stay there because their size is too big for migration in the resolving gel. In effect, a heavy smear at the top of the gel will be seen, corresponding to molecular masses above 200 kDa. However, if instead of 0.1%, a higher concentration (2%) of SDS is used, the high-mass smear is dissolved and only a band corresponding to monomeric HCC is detected. This indicates that the oligomers of native HCC are not cemented with covalent bonds because they can be disrupted upon protein denaturation (caused by elevated concentration of SDS).

Now, similar high-mass smears can also be obtained for the stab proteins. This statement contradicts what has been said before, where we argued that introduction of the stab mutations abrogates HCC oligomerization. The argument was true, but for the implicitly assumed oxidizing conditions, under which the stabilizing S-S bonds would be formed. However, if we incubate stab HCC under reducing conditions (e.g. in the presence of DTT), we will be able in fact to generate, and detect, high-order oligomers, just as with native HCC! If before applying the stab HCC oligomers to the SDA-PAGE gel, we remove DTT, i.e. convert the sample to oxidizing conditions, we will not be able to dissolve the aggregates even at very high concentration of SDS. This means that the oligomers are now cemented with covalent bonds. The explanation is that the intramolecular S-S bridges, originally disrupted by the addition of DTT, have been re-connected in the aggregates in intermolecular context, i.e. linking separate molecules. Considering the positions of the Cys mutations, this is only possible between domains that have been exchanged in a 3D domain swapping event. The correctness of this argument is confirmed by an additional test, in which the dissolving experiment (addition of 2% SDS) is conducted in the presence of DTT, i.e. under reducing conditions. In this test, the aggregates are dissolved exactly as those formed from native HCC.

### **Control of 3D domain swapping through environment and sequence factors**

We have already discussed several ways in which 3D domain swapping could be controlled. There are basically two possibilities: environmental conditions or mutations. The environmental factors that can be used to control (usually promote) 3D domain swapping include low pH, controlled heating or mild denaturant, all of which

facilitate gentle denaturation, which in turn facilitates partial unfolding, considered to be a prerequisite of 3D domain swapping.

The mutations that affect 3D domain swapping can occur in the hinge region or in a different fragment of the protein sequence. A mutation of a hinge residue could lead to interactions that are unfavorable in the loop conformation and can be relaxed only via a conformational change in this area. This would be compared to a “loaded spring” effect. The loaded spring would drive the protein into 3D domain swapping. A more intuitive strategy involves manipulation of the length of the hinge loop. If it is made too short (deletion), the chain may not be able to properly fold as monomer and the only rescue would be in a 3D domain-swapped dimer (or other oligomer). Here, the shorter hinge loop promoted 3D domain swapping via an energetic effect because a closed monomer would be energetically expensive. It is possible, however, to promote 3D domain swapping also by lengthening the hinge loop (insertion). In this approach, we are lengthening the tether in order to endow the protein chain with so much conformational freedom that the complementary domains would recognize each other in an intramolecular context only with low probability. In this case, the 3D domain swapping association is driven by a kinetic effect.

Mutations outside of the hinge region can be introduced at the open interface or at the closed interface. The former case is usually connected with energetic effect: we engineer a stronger interaction at the new interface, which is present in the oligomer but not in the monomer. The effect of a mutation in the closed interface is rather of a kinetic nature because the consequences of the mutation are the same for the monomer and dimer. It must be, therefore, connected with the ease with which the transition can be completed. This kind of mutational control of aggregation can be seen in the L68Q variant of HCC.

In addition to the cases described above, there will be of course also other possibilities, such as those described for stab HCC, where the domain-swapping process is influenced by alteration of the covalent topology of the polypeptide chain.

We can illustrate the considerations concerning mutational control of 3D domain swapping with a work reported by Murray et al. (1998), in which they used a fusion construct of the lymphocyte adhesion protein CD2 as a test field for their mutagenesis experiments. In this

particular construct, the CD2 protein exists in an equilibrium with about 15% of 3D domain-swapped dimers. The mutations were introduced either directly in the hinge region (residues 45-50) or in other areas that were somehow involved in interactions with the hinge segment. One set of mutations, introduced chiefly in order to create new salt bridges, was meant to stabilize the unswapped monomer. Those mutations were largely successful. For instance, the E29R mutant formed only about 2% of dimer. Conversely, removal of a salt bridge (R87A), as expected, destabilized the monomer (>50% dimer). There were also mutations designed to stabilize the dimer. The most successful one, deletion of the M46K47 fragment, converted the protein to nearly purely dimeric form (90%). Other manipulations with this goal in mind, were, however, spectacular failures. For instance the F49Y mutation meant to stabilize the dimer, in fact eliminated this form almost completely. The mixed results of Murray et al. demonstrate that we know already quite a lot about the relationship between protein structure and behavior, but that there are still deeper aspects of these phenomena that require further studies.

### 3D Domain swapping and the prion protein

A very interesting case, providing another link between 3D domain swapping and amyloid, was reported by Knaus et al. (2001) for the prion protein. Its 118-226 fragment, known from NMR studies to exist in the PrP<sup>C</sup> form of the protein as a monomer built of three helices (A, B, C) arranged in a triangular manner (Zahn et al., 2000; Zhang et al., 2000), was found by X-ray crystallography to form a dimer with helix C swapped between the folding units. After HCC, PrP was the second case of an amyloidogenic protein with domain-swapped structure, strongly supporting the view that the two phenomena may be connected, i.e. that 3D domain swapping could be the molecular mechanism of amyloid fibril formation. The case of PrP is particularly intriguing because helix C at the closed interface is linked to helix B by a disulfide bond. This means that upon 3D domain swapping the intramolecular S-S bond must be broken and then re-formed in intermolecular context. This unusual requirement for changing redox condition plus NMR studies of hydrogen-exchange which suggested that the B-C region of PrP is the most stable one, led Nicholson et al. (2002) to question the proposal linking 3D

domain swapping with amyloid aggregation of PrP. However, in a subsequent series of ingenious experiments, Lee and Eisenberg (2003) presented actually quite convincing arguments supporting the domain-swapping hypothesis.

The experiments started with oxidized monomeric PrP<sup>C</sup>. An addition of a reducing (DTT) and denaturing (guanidinium chloride) agents converted the protein into open monomer. By removal of the denaturant with simultaneous removal of DTT (return to oxidizing conditions), the monomers are converted into S-S cross-linked dimers and also to microscopic worm-like oligomers. When used as seeds, those aggregates can 'infect' PrP<sup>C</sup> used in the redox cycle, leading to fully grown amyloid fibrils.

### Amyloid criteria

Fibrillar morphology is one of the criteria used to define a protein deposit as amyloid. Tinctorial properties are the second criterion. In popular language, this means that amyloid samples can be stained using organic dyes, such as thioflavine T, or the better known Congo Red (CR). A true amyloid sample will not only turn red after CR treatment, but the color will change to apple-green in polarized light.



**Fig. 6.** A diagrammatic model of  $\beta$ -sheet helix, the purported motif of cross- $\beta$  signature. In this model, an infinite  $\beta$ -sheet along the fiber axis is formed by  $\beta$ -chains of consecutive molecules. The individual  $\beta$ -chains are perpendicular to the fiber axis. They could be antiparallel (as in this scheme) or parallel. The sheet has a helical twist, thus its name

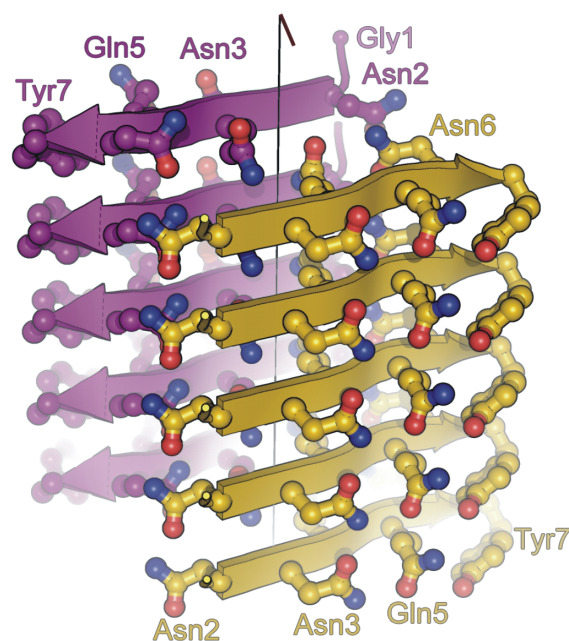


This effect, typical of highly organized crystal structures, strongly suggests that the protein chains within amyloid fibrils must also be very well ordered. The third criterion is furnished by amyloid fiber X-ray diffraction which produces a pattern with an extremely strong meridional (i.e. lying on a line parallel to the fiber axis) reflection with 4.8 Å Bragg spacing. This Bragg spacing and the intensity are highly suggestive of a progression of parallel (or antiparallel)  $\beta$ -strands forming an infinite intermolecular  $\beta$ -sheet. It is typically assumed that this structure would have a helical twist (thus the name " $\beta$ -sheet helix") although all the individual  $\beta$ -strands would remain perpendicular to the fiber axis, with the  $\beta$ -sheet propagation direction along this axis (Fig. 6). The characteristic meridional reflection and its underlying molecular structure are often referred to as cross- $\beta$  signature and cross- $\beta$  structure. The diffraction pattern of amyloid fibers usually also contains an outstanding equatorial reflection with about 10 Å Bragg spacing, interpreted as an indication that several  $\beta$ -sheet helices pack together in parallel, with a 10 Å separation, meaning that they must writhe in synchrony along the fiber axis.

### Steric zipper $\beta$ -spine

In search for the exact atomic architecture of the cross- $\beta$  structure, Eisenberg noticed that the GNNQQNY sequence motif found in the so-called yeast prion protein Sup35 appears to be particularly amyloidogenic, and that similar hexa-heptapeptide sequences (for instance, polyglutamine tracts) could be found also in other amyloidogenic proteins. Moreover, he managed to crystallize the GNNQQNY peptide (and a number of similar peptides) and showed that the crystal structure indeed contains a highly regular arrangement of the GNNQQNY peptides in an extended conformation, which form a pair of parallel  $\beta$ -sheets that have opposite directionality and an axial shift equal 1/2 of the inter-strand distance (confirmed to be 4.8 Å) within a sheet (Sawaya et al., 2007). There are several features that make this  $\beta$ -spine motif, termed the steric zipper, very unusual, and also very stable. (i) The  $\beta$ -chains propagate endlessly through the entire length of the crystal; (ii) the network of main-chain N-H...O=C hydrogen bonds stabilizing the  $\beta$ -sheet is reinforced by similarly arranged hydrogen bonds between the amide groups of the Asn side chains (and similarly Gln side

chains) that are endlessly stacked one above the other in the direction of the fiber axis; (iii) the inter-sheet space is filled with extremely densely packed, interdigitated side chains that protrude from both sheets into this interface, and that exclude all water molecules; for this reason this space is called the dry interface. The authors argue that analogously to the crystallization of the isolated peptides, these sequences, when present in a suitable protein context, are able to recognize each other and spontaneously form a similar  $\beta$ -spine in protein aggregates (Eisenberg et al., 2006).



**Fig. 7.** Steric zipper with the sequence GNNQQNY. Figure courtesy of Prof. David Eisenberg and Dr. Michael Sawaya

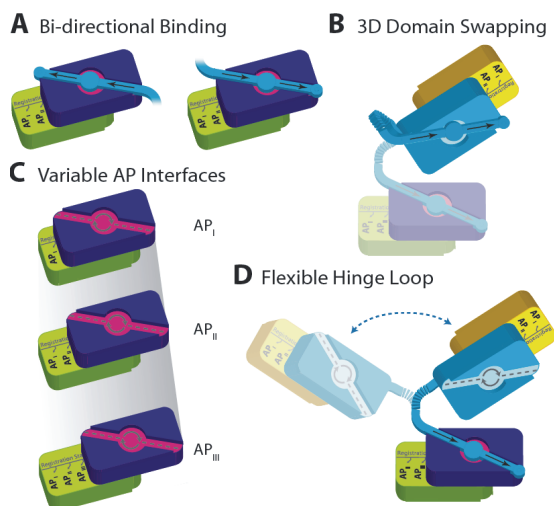
Considering various options for gain-of-interaction models of amyloid fibrils, Nelson and Eisenberg (2006) propose (i) direct stacking, (ii) cross- $\beta$  spine, (iii) 3D domain swapping, and (iv) 3D domain swapping with cross- $\beta$  spine. The primitive model (i) is not very convincing, while the dual-interaction model (iv) is the most appealing. It has been confirmed experimentally in RNase A with a genetically inserted glutamine expansion. The peripheral RNase A sequences form a familiar open-ended domain-swapped structure, while in the middle of the construction, the polyQ tracts are stacked into a  $\beta$ -spine.

Eisenberg and co-workers went even further, searching the genomes of *E. coli*, *S. cerevisiae*, and *H. sapiens* for the occurrence of steric zipper-forming peptides, and concluded that they are found in almost all pro-

teins but are usually buried to prevent aggregation (Goldschmidt et al., 2009). They estimated the energetic gain of steric zipper formation at 23 kcal/mol and demonstrated that introduction or disruption of steric-zipper sequences via genetic engineering provides a practical method to control amyloid formation in the laboratory.

### 3D Domain swapping and crystallin polydispersity

Finally, it may be illuminating to show that Nature uses 3D domain swapping not only to create harmful aggregates, but also to prevent high-order protein aggregation and crystallization in physiological milieus. The example comes from the vertebrate eye lens, in which concentration of its constituent crystallins can exceed 1 g/ml (fish eye). How are crystallin aggregation and crystallization, which would cause opacity, prevented? First, there are several crystallins (e.g.  $\alpha$ A,  $\alpha$ B) with slightly different sequences, so that misincorporations of a similar but actually wrong molecule would frustrate the growth of a nucleus. One has to understand that the lens milieu does contain crystallin oligomers; however, they are sufficiently small and polydisperse so that they do not cause transparency problems.



**Fig. 8.** In the lens of the vertebrate eye, crystallization and aggregation of  $\alpha$ -crystallins (which are present at astronomical concentrations) is frustrated by polydispersity, which is achieved not only by several different amino acid sequences of  $\alpha$ -crystallins, but also by: (A, B) 3D domain swapping of a palindromic C-terminal tail in two orientations, (C) a register shift at another intermolecular interface (AP), formed via  $\beta$ -sheet interactions, (D) variable conformation of a very flexible hinge element, allowing 3D domain swapping with different neighbors within an oligomer. Figure courtesy of Prof. David Eisenberg and Dr. Arthur Laganowsky

The second mechanism noted by Laganowsky et al. (2010) involves an intermolecular  $\beta$ -type interface, which can be formed with different register shifts (Fig. 8). Thirdly, there is another aggregation interface, formed via 3D domain swapping of a C-terminal tail. Amazingly, this tail has a palindromic sequence, ERTIPITRE, and can bind in its docking site in two orientations! Obviously, the situation is more complicated because even a palindromic oligopeptide does not have full twofold symmetry. Nevertheless, the interactions of the tail in the two orientations are nearly the same, thus very seriously frustrating high-order aggregation which would require exact repetition of the interaction motif. Finally, the hinge loop connecting the C-terminal tail with the body of the protein is long and very flexible, allowing for swapping of the tail with different partners within a growing oligomer.

### Conclusions

In this chapter, we have discussed selective topics focused on 3D domain swapping as a mechanism of protein oligomerization and aggregation. We have discussed the possible connection between 3D domain swapping and amyloid fibril formation. We have demonstrated that protein engineering can be used to control protein aggregation occurring via 3D domain swapping. We have also discussed selected aspects of the possible significance of 3D domain swapping in nature, illustrating its negative, but also positive physiological consequences.

### Acknowledgement

Publication of this work was supported by an EU-funded grant Nr 02.01.00-30-182/09 from the Ministry of Science and Higher Education under the Operational Programme *Innovative Economy*.

### References

- Anfinsen C.B. (1973) *Principles that govern the folding of protein chains*. Science 181: 223-230.
- Bennett M.J., Choe S., Eisenberg D.S. (1994) *Domain swapping. Entangling alliances between proteins*. Proc. Natl. Acad. Sci. USA 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.
- Cohen F.E., Prusiner S.B. (1998) *Pathologic conformations of prion proteins*. Ann. Rev. Biochem. 67: 793-819.

- Crestfield A.M., Stein W.H., Moore S. (1992) *On the aggregation of bovine pancreatic ribonuclease*. Arch. Biochem. Biophys. 1 (Suppl): 217-222.
- Crestfield A.M., Stein W.H., Moore S. (1993) *Properties and conformation of the histidine residues at the active site of ribonuclease*. J. Biol. Chem. 238: 2421-2428.
- Eisenberg D., Nelson R., Sawaya M.R., Balbirnie M., Sambashivan S., Ivanova M.I., Madsen M.O., Riekel Ch. (2006) *The structural biology of protein aggregation diseases: fundamental questions and some answers*. Acc. Chem. Res. 39: 568-575.
- Goldschmidt L., Teng P.K., Riek R., Eisenberg D. (2009) *Identifying the amyloids, proteins capable of forming amyloid-like fibrils*. Proc. Natl. Acad. Sci. USA 107: 3487-3492.
- Grzonka Z., Jankowska E., Kasprzykowski F., Kasprzykowska R., Łankiewicz L., Wiczek W., Wiczerzak E., Ciarkowski J., Drabik P., Janowski R., Kozak M., Jaskolski M., Grubb A. (2001) *Structural studies of cysteine proteases and their inhibitors*. Acta Biochim. Polon. 48: 1-20.
- 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*. Nature Struct. Biol. 8: 316-320.
- Janowski R., Abrahamson M., Grubb A., Jaskolski M. (2004) *3D Domain-Swapped Dimers of N-Truncated Human Cystatin C*. J. Mol. Biol. 341: 151-160.
- Janowski R., Kozak M., Abrahamson M., Grubb A., Jaskolski M. (2005) *3D Domain-swapped human cystatin C with amyloidlike intermolecular  $\beta$ -sheets*. Prot. Struct. Funct. Bioinform. 61: 570-578.
- Jaskolski M. (2001) *3D Domain swapping, protein oligomerization, and amyloid formation*. Acta Biochim. Polon. 48: 807-827.
- 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.
- Kolodziejczyk R., Michalska K., Hernandez-Santoyo A., Wahlbom M., Grubb A., Jaskolski M. (2010) *Crystal structure of human cystatin C stabilized against amyloid formation*. FEBS J. 277: 1726-1737.
- Knaus K.J., Morillas M., Świątnicki W., Malone M., Surewicz W.K., Mee V.C. (2001) *Crystal structure of the human prion protein reveals a mechanism for oligomerization*. Nat. Struct. Biol. 8: 770-774.
- Laganowsky A., Benesch J.L., Landau M., Ding L., Sawaya M.R., Cascio D., Huang Q., Robinson C.V., Horowitz J., Eisenberg D. (2010) *Crystal structures of truncated  $\alpha$ -phaA and  $\alpha$ -phaB crystallins reveal structural mechanisms of polydispersity important for eye lens function*. Protein Sci. 19: 1031-1043.
- Lee S., Eisenberg D. (2003) *Seeded conversion of recombinant prion protein to a disulfide-bonded by a reduction-oxidation process*. Nat. Struct. Biol. 10: 725-730.
- 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. USA 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.
- Liu Y., Eisenberg D. (2002) *3D domain swapping: as domains continue to swap*. Protein Sci. 11: 1285-1299.
- 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 Cryst. D49: 389-402.
- 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. USA 92: 7337-7341.
- Nelson R., Eisenberg D. (2006) *Recent atomic models of amyloid fibril structure*. Curr. Opin. Struct. Biol. 16: 260-265.
- Nicholson E.M., Mo H., Prusiner S.B., Cohen F.E., Marqusee S. (2002) *Differences between the prion protein and its homolog Doppel: a partially structured state with implications for scrapie formation*. J. Mol. Biol. 316: 807-815.
- Nilsson M., Wang X., Rodziejewicz-Motowidlo S., Janowski R., Lindstrom V., Onnerfjord P., Westermark G., Grzonka Z., Jaskolski M., Grubb A. (2004) *Prevention of domain swapping inhibits dimerization and amyloid fibril formation of cystatin C: use of engineered disulfide bridges, antibodies and carboxymethylpapain to stabilize the monomeric form of cystatin C*. J. Biol. Chem. 279: 24236-24245.
- Olafson I., Grubb A. (2000) *Hereditary cystatin C amyloid angiopathy*. Amyloid 7: 70-79.
- Sawaya M.R., Sambashivan S., Nelson R., Ivanova M.I., Sievers S.A., Apostol M.I., Thompson M.J., Balbirnie M., Wiltzius J.J.W., McFarlane H.T., Madsen A.O., Riekel Ch., Eisenberg D. (2007) *Atomic structures of amyloid cross- $\beta$  spines reveal varied steric zippers*. Nature 447: 453-457.
- Wahlbom M., Wang X., Lindstrom V., Carlemalm E., Jaskolski M., Grubb A. (2007) *Fibrillogenic oligomers of human cystatin C are formed by propagated domain swapping*. J. Biol. Chem. 282: 18318-18326.
- Zahn R., Liu A., Lurs T., Riek R., von Schroetter C., Lopez Garcia F., Billeter M., Calzolari L.G., Wüthrich K. (2000) *NMR solution structure of the human prion protein*. Proc. Natl. Acad. Sci. USA 97: 145-150.
- Zhang Y., Świątnicki W., Zagorski M.G., Surewicz W.K., Sonnichsen F.D. (2000) *Solution structure of the E200K variant of human prion protein. Implications for the mechanism of pathogenesis in familial prion diseases*. J. Biol. Chem. 275: 33650-33654.