Pacific Symposium on Biocomputing 4:566-577 (1999)

ONE SEQUENCE, FOUR FOLDS: TRANSITIONS BETWEEN AN ENSEMBLE OF METASTABLE FOLDS FOR THE N-TERMINAL DOMAIN OF CD2

R.B. SESSIONS, M.V. HAYES, A.J. MURRAY, A.R. CLARKE & R.L. BRADY
Department of Biochemistry, University of Bristol, Bristol BS8 1TD U.K.

Recombinant forms of the N-terminal domain of the cell adhesion receptor CD2 adopt a variety of folds by exchange of β-sheets between adjacent polypeptide chains. Although these interdigitated forms are normally metastable, we have used site-directed mutagenesis to alter the kinetics of formation and relative stabilities of these states, leading to spontaneous formation of monomeric, dimeric, trimeric and tetrameric intertwined folded states. A characteristic feature of these fold-disorder-alternative fold transitions is the independence of each domain folding event, as deduced from kinetic analysis of folding data. Structures for fully interdigitated trimeric and tetrameric forms have been modelled, consistent with both the crystallographic and kinetic data. Although the biological role of these alternative folded states remains unclear, these structures form a remarkable demonstration of the fluidity of structure generated from a single polypeptide chain.

1. CD2, an unusual member of the immunoglobulin superfamily

Alternative folded states have only rarely been described for proteins. This derives in part from the limitation that most of our techniques for protein structure determination can usually only successfully be applied to compact and stable protein folds. Although energetic considerations imply a direct link between these prevalent stable forms and biological function, a great deal of biology arises from the transient existence of protein conformations of only limited stability. The simplest examples are local conformational changes near or in the active sites of enzymes during catalysis. However, far more dramatic reordering events [e.g. those shown by calmodulin and haemaglutinin] have also been described. Central to all of these events is the ability of a polypeptide chain to undergo a transition between at least two states, each sufficiently long-lived under the prevailing conditions to permit a desired biological function. The instability of the majority of these alternative forms ensures their transient existence, and hence inaccessibility to characterisation. In this paper we describe the application of a variety of biophysical approaches to studying transitions between a family of altered folding states of a single domain from the cell adhesion molecule, CD2.

CD2 is found on the surface of T-lymphocytes, where it serves to anchor clusters of cells through interaction with its ligands (CD48 in humans, CD58 in rodents) present on the opposing cell surface. The extracellular portions of CD2 and its ligands comprise tandem arrangements of two immunoglobulin superfamily (IgSF) domains, a relatively simple protein fold comprising a sandwich of two β-sheets. These prevalent and widespread domains are invariably stabilised by the inclusion of a disulphide bond joining the two sheets, but

Pacific Symposium on Biocomputing 4:566-577 (1999)

unusually in CD2 and in each of its ligands the IgSF domain most distant from the cell surface lacks this characteristic bond. These N-terminal domains form the active surface leading to cell adhesion. Although genetic studies indicate the intersheet disulphide appears to have been lost in the evolution of these domains, the relationship between this alteration and biological function is not clear.

The dominant fold of the N-terminal domain of CD2 has been defined in a range of crystallographic [1,2] and NMR [3] studies. These confirmed the expected variable (V-) type IgSF topology (Fig 1a). Consistent head-to-head packing of molecules in a variety of crystal lattices (Fig. 1a) is believed to mimic cell surface interactions between CD2 and its structurally homologous ligands. The proposed active surface has been confirmed through mutagenesis studies [4,5] and is dominated by the presence of many hydrophilic and charged residues. Ligand-receptor interactions are believed to be predominantly electrostatic, hence correlating with the measured relatively low ($< 10^{-3}$ M) affinity [6].

2. CD2 can adopt an alternative fold

In previous studies we found that when the N terminal domain of rat (CD2 D1) was expressed in tandem with glutathione S-transferase (GST) (i.e. a fusion protein in which the second IgSF domain of CD2 was replaced by GST) a small fraction (15%) of the recombinant protein was found to adopt an alternative fold as an intertwined dimer [7]. Although only metastable (see below), this unconventionally folded form was of sufficient stability to be crystallised. The crystal structure (Fig. 1b) revealed two polypeptide chains intertwined head-to-tail to form a dimer of two "pseudo" domains, to each of which both polypeptide chains contributed strands and which resembled most of the IgSF fold described for the prevalent, monomeric form of the protein. Each pseudo domain is formed from strands A,B, C and C' from one polypeptide chain, and strands D, E, F and G

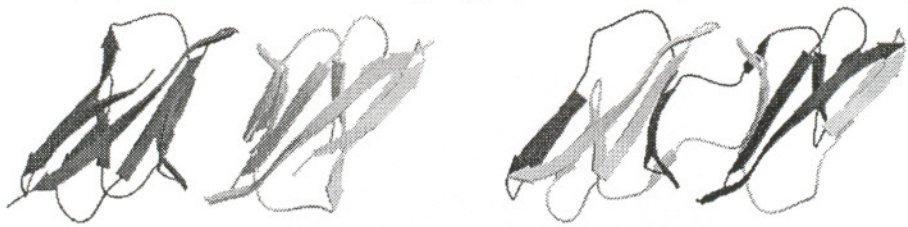

Figure 1: (a - left) Crystal structure of the monomeric form of CD2 D1, showing head-to-head packing of two monomers in the crystal lattice. (b - right) Crystal structure of the intertwined dimer of CD2 D1. One polypeptide chain is shown in dark shading, the other in light. Drawn with MOLSCRIPT [17].

from the other chain (standard V-type IgSF nomenclature). The unravelled C" strand forms a hinge or linker between the two lobes, which enclose a substantial hydrophilic core in the centre of the fold. Buried hydrophilic cores are rare in proteins, and in this case this region is likely to be crucial to the observed metastability of the dimer. This intertwined form of CD2 represents an extreme example of a "3D domain-swapped" protein [reviewed in Refs 8-10] and hence a potential molecular fossil illustrating protein evolution.

Initially the formation of this dimer was considered to be an artefact of the expression system, as GST is itself a dimer and consequently ensures the two CD2 polypeptide chains are kept in close proximity, hence increasing the probability of interactions. This was supported by the observation that the unfolding and refolding of the isolated CD2 dimer produced only the monomeric fold. Additionally, after removal of the GST, samples of intertwined dimer would slowly revert to the monomeric fold over a period of several weeks, illustrating the metastable nature of the dimeric fold. However, the anchoring effect of the GST dimer is reminiscent of events at the cell membrane surface, where interactions between coupled extra- or intra-cellular domains frequently tether molecules in close proximity. This effect is further enhanced by the essentially two-dimensional restriction in movement within the membrane. Fortuitously, this fusion protein has enabled the study of less stable folds for CD2 normally inaccessible to study in solution, but perhaps more prevalent and functionally important at the cell surface.

3. Gel filtration allows discrimination between folded states

Gel filtration provides a simple technique for identifying and quantitating the existence of differentially folded states of CD2. These states result from interdigitation of β-strands, leading to the assembly of polypeptide chains to form higher order oligomers. Face-to-face interactions between monomers of CD2 (Fig. 1a) are known to be of very low affinity (less than 10^{-3} M, [6]) and hence

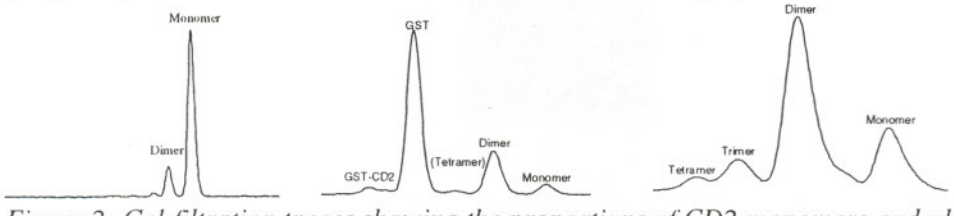

Figure 2: Gel-filtration traces showing the proportions of CD2 monomers and oligomers in recombinant preparations. (a) [left] is wild-type, (b) [middle] is the deletion (Δ46Δ47) mutant, and (c)[right] the deletion mutant after refolding in the absence of GST and equilibration for 100 hrs

oligomers observed in solution must result from alternatively folded forms - as confirmed by the crystallographic analysis of the intertwined dimer. Fractionation of the cleaved fusion protein product by gel filtration (Fig 2a) allows the separation of intertwined dimeric fold CD2 from the monomeric form. This simple technique has permitted a range of experiments probing the effect of molecular changes on the transition between the two folded states of CD2.

4. Mutagenesis as a tool for the study of folding transitions

Using site-directed mutagenesis we have incorporated chemical changes to the CD2 polypeptide in order to alter the relative stability of each folded form. From the crystal structure of the interdigitated dimer we designed amino acid changes that would differentially stabilise either fold. Mutation of residues in the buried, hydrophilic core of the dimer lead to drastic changes in the relative proportion of

Pacific Symposium on Biocomputing 4:566-577 (1999)

monomer and dimer produced (Table 1, details in [11]). This site is characterised by an intricate network of electrostatic and hydrogen bonds in which virtually every possible interaction that can be made is observed. Although in all cases single-site mutations were sufficient to regulate the monomer-to-dimer transition, the largest effects were observed where mutations altered the hydrophobicity of either the buried core, or its equivalent exposed surface in the monomer. From these studies we conclude that entropic considerations are the dominant factor governing partitioning between the two folded states, which parallels experience with protein folding in general. Not surprisingly, transitions between folded states of the same protein appear to be regulated by the same factors that determine protein folding.

The major topological change between the monomeric and dimeric forms of CD2 is the C" strand region, which unravels to form a hinge beween the two pseudo domains in the dimeric form. Removal of two residues from this segment of the polypeptide chain had a dramatic effect on the proportions of monomer and dimer obtained (Table 1). Modelling studies suggested this shortened region would introduce strain to the C" strand of the monomer. As expected, CD2 now predominantly forms intertwined dimers, although note there is also some evidence for further assembly to a tetramer (Fig 2b). The marked differences in folding behaviour between this protein and the wild-type form of CD2 led us to examine their kinetics of folding in more detail.

Table 1: Effect of site-directed mutations on monomer-dimer transitions

Mutation	Position in polypeptide sequence	Expected effect of mutation	% monomer	% dimer
Wild type	-	-	85	15
Glu -> Arg	29 - interface	Loss of salt bridge	94	6
Lys -> Glu	43 - interface	Loss of salt bridge	97	3
Tyr -> Ser	81 - interface	Loss of salt bridge	93	7
Arg -> Ala	87 - interface	Increased hydrophobicity	48	52
Ser -> Glu	36 - interface	Addition of salt bridge	85	15
Phe -> Tyr	49 - interface	Addition of H-bond	98	2
Tyr -> Arg	81 - interface	H-bond to salt bridge	78	22
Delete Met	46 - hinge region	Destabilise monomer	85	15
Del Met+Lys	46+47 - hinge region	Destabilise monomer	15	85

5. Destabilisation of the monomeric fold induces a transition to dimer

A single tryptophan residue (Trp 32), which is buried within the core of each monomer or dimeric pseudo domain, provides a convenient fluorescence probe for monitoring folding and unfolding events. The equilibrium distributions [F]/[U] in water ($K_{F/U(w))}$) for the wild type and deletion mutant CD2 were determined by equilibrium denaturation methods (Table 2, details in Ref 10). A value for $K_{F/U(w)}$ of ~4300 was derived from the data showing that, in comparison to the wild type

Pacific Symposium on Biocomputing 4:566-577 (1999)

monomer ($K_{F/U(w)}$ ~ 1 x 10^6), the folded state is about 3 kcal/mol less stable (see Table 2). The dimerisation constants for wild type and mutant dimer formation within the fusion protein are 2 x 10^{11} and 1 x10^8 respectively. The increase in dimer formation observed for the mutant CD2 is explained by the fact that a pair of folded monomers within the fusion protein is destabilised by 6 kcal/mol (2 x 3 kcal/mol) whereas the dimer is only destabilised by 4 kcal/mol; thus there is a net shift towards dimer formation of 2 kcal/mol. The relative differences, and not the absolute value, of the free energy drives the transition.

Equilibrium and rate constants show that the deletion destabilises the monomer chiefly by increasing the unfolding rate (k_{F-I}). Neither the rate-limiting folding step (k_{I-F}) nor the stability of the intermediate state ($K_{I/U}$) are strongly influenced by the mutation, a result which is consistent with the deletion lying outside the key folding nucleus which, in a separate study using deuterium-exchange NMR data [13], one of us (ARC) has found to be formed by the B, C, E and F strands.

Table 2. Kinetic and equilibrium folding behaviour of wild type and Δ46 Δ 47 mutant CD2

	Folding parameters of mutant monomer	Folding parameters of wild type monomer	Folding parameters of mutant dimer	Folding parameters of wild type dimer
$k_{I-F(w)}(s^{-1})$	20.42 ± 3.43	7.20 ± 0.58	ND	ND
$k_{F-I(w)}(s^{-1})$	0.029 ± 0.009	(2.0 ± 0) x 10^{-4}	(1.2±0.3) x 10^{-3}	(1.9±0. 6) x 10^{-3}
$K_{I/U}$	6.50 ± 4.6	28.23 ± 4.78	ND	ND
$K_{F/I}$	716	36000	ND	ND
$K_{F/U}$	4300	1 x 10^6	ND	ND
ΔG_{eq} (kcal mol $^{-1}$)	-4.97	-8.22	-10.96	-15.40
K_{Dim} (in situ)			1.0 x 10^8 (unimolecular)	1.8 x 10^{11} (unimolecular)
$m_f(M^{-1})$	-2.21 ± 0.14	-2.75 ± 0.06	-2.73 ± 0.12	-2.44 ± 0.12

6. Crystal structure of the deletion dimer reveals domain rotations

The structure of the stabilised deletion mutant dimer has been determined in several crystal forms, and is shown in Fig 3a (details of crystallographic determinations are reported in [11]). This intertwined dimer has the same overall topology as the wild-type dimer, but differs significantly in the relative positioning of each of the pseudo domains. When one domain is superimposed, the other is seen to swivel by about 90° relative to its position in the wild type structure. This domain rearrangement produces an important effect: the buried hydrophilic interface observed in the wild-type structure is now open and exposed to solvent.

Figure 3: Cα traces showing the crystal structures of (a, left) the intertwined dimer form of wild-type D1 of CD2; with the hydrophilic interface buried; (b, centre) dimer of the Δ45Δ46 deletion mutant, with the interface now open; and (c, right) under high salt conditions, the same mutant crystallises as a tetramer (dimer-of-dimers). The arrangement of domains is approximately tetrahedral with a quasi-continuous central β-barrel. Drawn with MOLSCRIPT [17].

When crystallised under high-salt conditions [11], deletion mutant dimers further assemble to tetramers (Fig 3b). In this case, the hydrophilic surface becomes buried again as initially observed in the wild type dimer structure, but this time between two complementary halves from independent dimers. Note that the 90° relative rotation of each pseudo domain allows a complementary half from another dimer to bind in a similar orientation to that of the contiguous pseudo-domain of the wild type dimer. The delicate balance between different states of assembly is coordinated and simplified through the expoitation of complementary and symmetrical structures. Simple mutagenesis of the original protein in an appropriate site has resulted in sufficient stabilisation of alternative folds for us to define crystal structures for three different forms: a monomer, an intertwined dimer, and a tetrameric dimer of intertwined dimers.

7. Isolated monomers also spontaneously form higher-order structures via partially disordered states

The increased stability of the mutant dimer, relative to its monomeric counterpart, and unexpected further assembly to tetramers, prompted a re-appraisal of the requirement for GST in dimer formation. Cleaved protein was unfolded in 6 M GuHCl, refolded in buffer to produce the monomer. A 1 mM solution of this monomeric material was then allowed to equilibrate and the sample subjected to gel filtration to separate species of differing molecular weights. After 100 hrs a stable distribution of protein peaks was established as shown in Fig 2c. Not only does this result show that the dimer can form spontaneously in the absence of GST, but there are also well-defined peaks corresponding to trimer and tetramer. Hence, at elevated protein concentrations the assembly reaction progresses from monomer to dimer to trimer and finally to tetramer. The relative proportions of each oligomer can be used to calculate concentrations of each species at equilibrium giving values of 150 μM, 365 μM, 33 μM and 5μM, respectively, for the monomer, dimer, trimer and tetramer. Therefore the effective dissociation constants for the incremental addition of single chains are 5.9×10^{-5} M (monomer -> dimer), 2.0×10^{-3} M (dimer -> trimer) and 1.0×10^{-3} M^{-1} (trimer -> tetramer). These values demonstrate that the addition

of the first chain is most favourable as it releases strain in two monomers whereas in subsequent steps only one monomer opens up and adds to the complex. Note that the time course of assembly is inconsistent with simple diffusion-controlled association of monomers, supporting the notion that the dimers, trimers and tetramers observed must all be interdigitated species which must unfold partially before the assembly process.

8. Structure of trimers and tetramers can be modelled

The surprising observation that not only the intertwined dimers could spontaneously form, but also that trimers and tetramers appeared over time, led us to speculate on the molecular structures of these higher-order oligomers. Unlike the dimers, where their abundance and stability allowed us to determine the molecular structure via crystallography, the limited quantities of the trimeric and tetrameric species have so far prevented direct structural determination. However, the structure of an assembled tetramer was already available from the crystal structure of the dimer of intertwined dimers (Fig 3c). The positioning of the chains in each dimer is such that it is relatively straightforward to form a fully intertwined tetramer (Fig 4) by breaking the peptide bonds between residues 50 and 51 in chains B and C and rejoining the N-terminal half of chain B to the C-terminal half of chain C, likewise the N-terminal half of chain C to the C-terminal half of chain B. This model leaves each pseudo-domain in an almost identical position to that observed in the dimer of intertwined dimers, and it is feasible that both forms could have a very similar overall topology. Indeed, the electron density maps of the tetrameric species show low-occupancy density in this direction consistent with a small percentage of the molecules within the

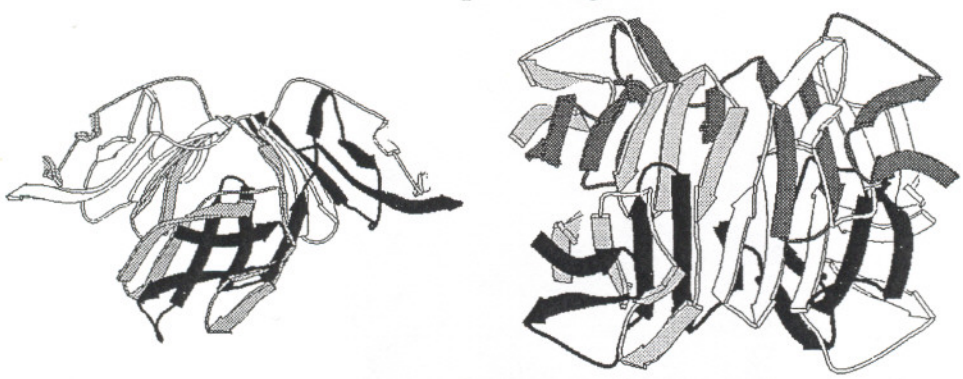

Fig 4: Modelled structures of (left) a fully-intertwined trimer, and (right) fully-intertwined tetramer of CD2. These models are derived from the crystal structure of the dimer-of-intertwined dimers tetramer, in each case preserving the quasi-continuous β-barrel in the protein core. Drawn with MOLSCRIPT[17]

crystal lattice being intertwined tetramers. Hence, the crystal structure has provided an indirect route to the structure of the fully intertwined tetramer.

By contrast, we have no crystallographic evidence for a trimeric form of CD2. However, two important principles emerge from a comparison of the existing crystal structures. Firstly, domain rotations are both feasible and energetically accessible. Secondly, oligomers maximise intra-domain interactions by

centralising the hydrophilic faces to form a buried core. It seems very likely that trimers will exploit these same principles. We have modelled an intertwined trimer by rearranging three pseudo-domains around a similar central core (Fig 4). In the trimer the pseudo domains are comprised of complementary segments from chains A & B, B & C, and A & C respectively.

After soaking with an 8 Å layer of water, both the trimer and tetramer model structures were energy minimised while tethering the backbone atoms to their initial positions. The tethering force was reduced during this procedure from 100 to a final value of 0.5 kcal/Å. The energy minimised structures (average derivative < 0.02 kcal/Å) were analysed using Procheck and found to be of similar or better quality compared with the original crystal structure. Structural manipulations were carried out using InsightII (95) and energy minimisations using the cvff forcefield implemented in Discover (2.95).

Although mutation of the native peptide sequence has been essential in order to visualise this ensemble of structures, it seems feasible that these higher-order interdigitated assemblies could also form with the native sequence. The relative energetic differences between dimeric and monomeric forms regulate equilibrium concentrations of these folds, but nevertheless the increased flexibility associated with the longer C" strand in the wild-type sequence should ensure at least transient formation of these structures is possible.

9. Nature of disordered intermediate states deduced from the independent folding of pseudo-domains

Formation of interdigitated dimers, trimers and tetramers could occur either by simultaneous assembly of unfolded chains in a single step, or via intermediates in which chains are associated but folding of each pseudo-domain is achieved separately. Through analysis of the kinetics of unfolding of monomeric and dimeic forms of CD2 we have been able to distinguish between these possibilities. Stopped-flow fluorescence analyses of samples of both monomer and dimer rapidly mixed with aliquots of guanidinium chloride have been used to determine unfolding rates for each species (Table 2, experimental details in Ref 10). The deduced folding rates obtained for dimer formation ($4 \times 10^8 s^{-1}$ for the wild-type and $1 \times 10^5 s^{-1}$ for the deletion mutant) are unrealistically high, being many orders of magnitude greater than the folding rate of the CD2 monomer. This dilemma is resolved if the behaviour of each of the intertwined pseudo-domains is considered independent, whereupon these two separate folding steps contribute to the overall U~U <-> D equilibrium such that:

$$K_{Dim} = k_{f(app)}^{\ 2} / k_U^{\ 2} \qquad \text{and} \qquad k_{f(app)} = (k_U^{\ 2} \cdot K_{Dim})^{1/2}$$

Re-interpretation of the data now gives an apparent folding rate of ~900 s^{-1} for the wild-type dimer and ~10 s^{-1} for the mutant. As this rate includes a term for the stability of any intermediate state in the folding of a pseudo-domain (i.e. $k_{f(app)} = K_{I/U} \times k_f$), these values are highly plausible given that the rate-limiting folding

Pacific Symposium on Biocomputing 4:566-577 (1999)

steps are 7 s^{-1} and 20 s^{-1} for the wild-type and mutant monomers, respectively. Hence these data support a two-step folding mechanism in which the folding of pseudo-domains within the dimer is independent.

This conclusion is reinforced by an analysis of the m-values for each state in the system, which indicate its sensitivity to denaturant and the degree of exposure of hydrocarbon to solvent. They provide a measure of the degree of packing of each state in the folding reaction - the more negative the value, the more exposed the state with respect to the folded molecule for which, by convention, $m_F = 0$. A comparison of the absolute m_t values for the monomers with those for the dimers (m_t represents the level of solvation of the structure as it passes from the folded state to the transition state for unfolding) shows that the values for the dimeric species are close to those determined for the monomers. If attainment of this state in the dimeric species required the co-ordinated unfolding of both pseudo-domains, we would expect the magnitude of m_t to be two fold greater than that seen for the monomeric state. These results provide further evidence that the pseudo-domains in the dimer behave independently.

When considered with the interdigitated topologies observed for the dimers, trimers and tetramers of CD2, we can now deduce that the protein must transiently exist in states where at least one domain is folded, and another becomes disordered to permit incorporation of another polypeptide chain and hence assembly (Fig 5). The symmetrical nature of the oligomers ensures equal probability of any of the domains unfolding, hence maximising the likelihood of constructuve assembly taking place.

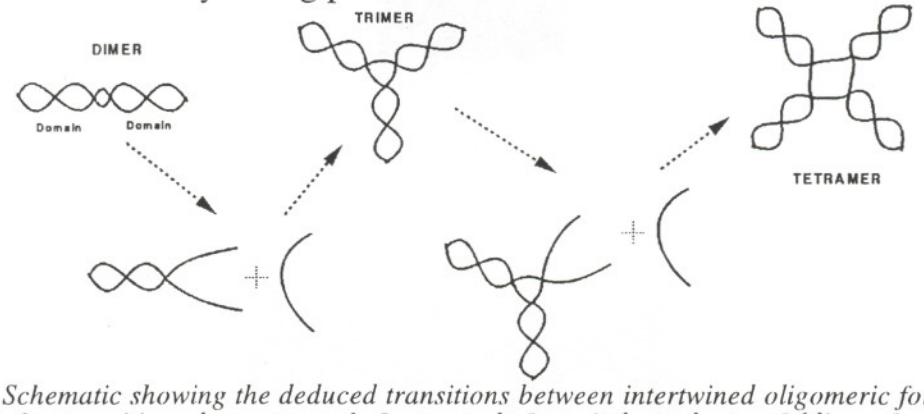

Figure 5. Schematic showing the deduced transitions between intertwined oligomeric forms of CD2 D1. Note that the transitions between each form result from independent unfolding of a single pseudo-domain, and subsequent incorporation of an additional polypeptide chain.

10. Conclusions - Determinants of order-disorder transitions leading to alternative folds

CD2 D1 is an unusual polypeptide sequence that adopts not one, but a variety of folds. These folds are variations on the universal IgSF β-sandwich motif, with oligomeric forms resulting from interdigitation of strands from adjacent, complementary polypeptide chains. Transitions between these states must involve a variety of intermediates characterised by the presence of unfolded, disordered

polypeptide segments. Although these oligomeric forms are of limited stability, the dimeric form was sufficiently stable to enable its structure to be determined by crystallographic means. Further stabilisation of this dimer through site-directed mutagenesis has enabled us to observe and study higher order interdigitated assemblies, and using a broad range of techniques to probe the factors governing the partition of material between these alternative folding pathways.

Although this ensemble of structures for CD2 essentially represents variations on the same folding theme, this simplified family of structures has proven amenable to rigorous structural analysis and allows us to derive some general principles that may have universal application to protein structural transitions.

(i) Transitions can exploit symmetry - Access to a variety of conformational states for any protein will be regulated firstly and foremost by energetic considerations. Significant populations of alternative folded states will only be observed when these forms are of overall similar free energy. In the case of CD2, this is achieved through a family of structures assembled from the same basic fold. By swapping equivalent segments of polypeptide between complementary domain halves, minimal disruption to the energetics of the native domain fold occurs. This same principle has been proposed to have been exploited during the evolution of protein oligomers [8,9]. CD2 also maximises complementarity through the use of symmetry. Precisely half of each domain fold is swapped in each intertwined domain. This permits assembly to dimers, trimers, and tetramers, with the relative energies of each being regulated only by changes in the interface between the domains or in the linker region. The dramatic changes in equilibrium associated with the deletion mutant are probably accentuated by the location of the mutation site in the centre of the polypeptide chain (residues 45 and 46 out of 99).

A further consequence of the inherent symmetry in these structures is that transitions between ordered and disordered forms are of equal likelihood for each pseudo-domain. Simultaneous assembly of four polypeptide chains is statistically unlikely to be productive. Independent folding of each domain segment is crucial to the ability to form an ensemble of structures, and is likely to be characteristic of many proteins undergoing fold-disorder-alternative fold transitions.

(ii) Transitions can be initiated by destabilisation of one conformation - Although the intertwined form of wild type CD2 is fundamentally metastable, the proportion adopting this fold can be increased through the introduction of strain (as in the deletion mutant) to the monomeric "ground" state. The difference between, and not absolute values of, the energies of the alternative conformations determine the predominant species. In this case the relative increase in energy of the ground state derives from a change in the amino acid sequence (through mutagenesis), but in other systems might arise from complex formation with another biomolecule.

(iii) Transitions are regulated by the same factors that determine protein folding - This is a direct consequence of the energetic considerations above. Our mutagenesis studies on CD2 highlighted the dominance of entropic considerations

Pacific Symposium on Biocomputing 4:566-577 (1999)

in determining the energy, and hence prevalence, of a particular fold. Intertwined forms of CD2 are highly dependent on the delicate balance of electrostatic and hydrogen bonds across the hydrophilic interface. These are not unexpected findings, but support computational approaches to conformational transitions based on our understanding of protein stability from existing structures.

(iv) Domain rearrangements can facilitate transitions - As seen in the structure of the deletion dimer, domain movements in intertwined forms of CD2 can minimise energy differences between folded forms by alleviating steric hindrance which might otherwise prevent domain formation, and enable the maintainance of an interface between domains. This proves crucial to the interdigitated forms of CD2, where small changes even in single hydrogen bonds across the hydrophilic interface significantly alters the partitioning between monomeric and dimeric forms. In order to exploit rearrangements of this kind, a protein must comprise at least two structural elements joined by a flexible linker region. These need not be whole protein domains, but might be segments capable of insertion to complete a domain as exemplified in CD2. Nonetheless, we expect functional conformational transitions to be more prevalent in multidomain, symmetrical proteins.

11. Why does CD2 adopt more than one structure?

It is not immediately obvious that the ability of CD2 to adopt a variety of assembled forms is relevant for its biological function. There are parallels with the eye lens proteins, the β- and γ-crystallins, where insertions and deletions of amino acid fragments noted between different types of these proteins appear to regulate assembly to form lens tissues of differing refractive index [14]. Although for CD2 different folds have only been observed for engineered and recombinant forms of the protein, the apparent intentional loss in evolution of the inter-sheet disulphide that enables these transitions to occur suggests this may be required to achieve a particular biological function. One possibility is regulation of adhesive activity. Intertwined forms of the protein bury the active protein surface, hence switching off its ability to combine with ligands. This could be desirable not only at the cell surface, but also during its translocation to the cell membrane after synthesis.

The propensity of the immunoglobulin domain to oligomerise upon removal of residues from the central C" strand highlights the possible associated loss of function that could accompany such transitions in either disease states or evolution. Analyses of genome sequence data show a low abundance of all-β-sheet proteins within each genome (5-7% of proteins, [15]), perhaps reflecting the susceptibility of such folds to intertwining and assembly. This restriction also explains the almost universal presence of inter-sheet disulphide bonds in extracellular β-sandwich folds. IgSF domains also occur in intracellular cytoskeletal proteins such as telokin, twitchin and titin, and where intersheet disulphides do not form. As titin, which comprises extremely long concatenations of β-sandwich domains, has been proposed to regularly unfold and refold during

muscular extension [16], the question arises as to whether alternate folding states might form when the many similar domains within the titin molecule refold in close proximity.

Acknowledgments

This work was supported by grants to ARC and RLB from the BBSRC, MRC and Wellcome Trust.

References

1. E.Y. Jones, S.J. Davies, A.F. Williams, K. Harlos & D.I. Stuart, *Nature* **360**, 232-239 (1992).
2. D.L. Bodian, E.Y. Jones, K. Harlos, D.I. Stuart & S.J. Davis, *Structure* **2**, 755-766 (1994).
3. P.C. Driscol, G.J. Cyster, I.D. Campbell & A.F. Williams, *Nature* **353**, 762-765 (1991).
4. C. Somoza, P.C. Driscoll, J.G. Cyster & A.F. Williams, *J.Exp.Med*. **178**, 549-558 (1993).
5. A.R.N. Arulanandam, J.M. Withka, D.F. Wyss, G. Wagner, A. Kister, P. Pallai, M.A. Recny & E.L. Reinherz, *Proc. Natl. Acad. Sci. USA.* **90,** 11613-11617 (1993).
6. P.A. van der Merwe, P.N. McNamee, S.J. Davies & A.N. Barclay, *EMBO J*. **12**, 4945-4954 (1993).
7. A.J. Murray, S.J. Lewis, A.N. Barclay & R.L. Brady, *Proc. Natl. Acad. Sci. USA.* **92**: 7337-7341 (1995).
8. M.J. Bennet, M.P. Schlunegger & D. Eisenberg, *Protein Science* **4**: 2455-2468 (1995).
9. M.P. Schlunegger, M.J. Bennet & D. Eisenberg, *Advances in Protein Chemistry* **50**, 61-122 (1997).
10. D. Xu, C.J. Tsai & R. Nussinov *Protein Science* **7**, 533-544 (1998).
11. A.J. Murray, J.G. Head, J.J. Barker & R.L. Brady, *Nature Struct. Biol.* 5, 778-782 (1998).
12. M.V. Hayes, R.B. Sessions, R.L. Brady & A.R. Clarke, *Manuscript in preparation.*
13. M.J. Parker, C.E. Dempsey, L.L.P. Hosszu, J.P. Waltho & A.R. Clarke, *Nature Struct. Biol.* 5, 194-198 (1998).
14. B.V. Norledge, S. Trinkl, R. Jaenicke & C. Slingsby, *Protein Science* **6**, 1612-1620 (1997).
15. D. Frishman & H.W. Mewes, *Nature Struct. Biol*. **4**, 626-628 (1997).
16. H.P. Erickson, *Proc. Natl. Acad. Sci. USA.* **91**: 10114-10118 (1994).
17. P.J. Kraulis, *J. Appl. Cryst*. 24, 946-950 (1991).