Folding studies of immunoglobulin-like β-sandwich proteins suggest that they share a common folding pathway
Jane Clarke*, Ernesto Cota, Susan B Fowler and Stefan J Hamill

Background: Are folding pathways conserved in protein families? To test this explicitly and ask to what extent structure specifies folding pathways requires comparison of proteins with a common fold. Our strategy is to choose members of a highly diverse protein family with no conservation of function and little or no sequence identity, but with structures that are essentially the same. The immunoglobulin-like fold is one of the most common structural families, and is subdivided into superfamilies with no detectable evolutionary or functional relationship.

Results: We compared the folding of a number of immunoglobulin-like proteins that have a common structural core and found a strong correlation between folding rate and stability. The results suggest that the folding pathways of these immunoglobulin-like proteins share common features.

Conclusions: This study is the first to compare the folding of structurally related proteins that are members of different superfamilies. The most likely explanation for the results is that interactions that are important in defining the structure of immunoglobulin-like proteins are also used to guide folding.

Introduction
How much information can be transferred from folding studies of one protein to other members of the same family? Will folding patterns be similar, or does each protein have a unique folding mechanism dictated by its individual amino-acid sequence? Such questions can only be addressed by detailed folding studies of a number of related proteins. A major problem in the analysis of protein families is distinguishing between residues conserved for folding and residues conserved for function. Our strategy, therefore, is to study a highly diverse protein family with no conservation of function and little or no sequence identity, but with structures that are essentially the same. The immunoglobulin-like (Ig-like) β-sandwich [1,2] is one of the most common structural motifs [3]. Ig-like domains are found widely in proteins of diverse function: proteins of the extracellular matrix, muscle proteins, proteins of the immune system, cell-surface receptors and enzymes. Much structural data is available so it is possible to choose proteins that are relatively easy to study (having a tryptophan for spectroscopic observation and no disulphide cross-links, for example) and there is a wealth of statistical information on sequence variability.

Theoretical studies suggest that folding pathways are conserved in protein families: specifically, that interactions constituting the folding nucleus are conserved, although the sidechains involved in these interactions may not themselves be invariant [4,5]. To test this hypothesis and discern to what extent structure specifies folding pathway, we compare the folding of five different Ig-like proteins from two different superfamilies — three Ig domains and two fibronectin type III (fnIII) domains. Importantly, they have no common function and no discernible sequence homology so that any specific features of folding behaviour they share must be related to the only common feature, their structure. We propose, from a comparison of the folding kinetics of the wild-type proteins, that the structural constraints on proteins with this Ig-like β-sandwich fold impose a common folding mechanism.

Results and discussion
Comparison of the proteins
The Ig-like fold
Members of the Ig-like structural family, or fold, are divided into superfamilies mainly on the basis of sequence similarity [2]. Different superfamilies have no significant sequence identity, that is, they have no detectable evolutionary relationship. The immunoglobulin, fibronectin type III and cadherin superfamilies have 3495, 2103 and 609 members, respectively, in the current pfam data base [6]. All the Ig-like proteins have two antiparallel β sheets packed against each other and a similar Greek key strand topology. They differ in the number and the position of peripheral strands, but share a basic strand arrangement and hydrogen-bonding pattern (Figure 1). The strands are given letters from A to G according to a standard pattern. Thus all proteins will have a sheet containing the B and E...
strands, but this sheet may also include an A and/or a D strand. The second sheet consists of C, F and G strands, but may also include an A (or A′) strand (in parallel with G), and/or an extra C′ and even C′′ strands. There are connecting loops between the sheets. All proteins have inter-sheet loops from B to C and from E to F strands; connections for other strands vary with the number and arrangement of strands at the edge of the two β sheets (Figure 1).

The proteins
Five different proteins are used in this study, representing two different superfamilies. The structure and strand topology of these proteins are described in Figure 1. The two fnIII domains are isolated modules of two human intracellular matrix proteins — the third fnIII domain from human tenascin (TNfn3) and the tenth fnIII domain of human fibronectin (FNfn10). These have significant sequence identity, (~23%) (Table 1) and a larger number of residues that are structurally equivalent (~70%), only the A–B, C–C′, C′–D and F–G loops are different in structure (Figure 2). The conserved core tryptophan is found in the B strand. FNfn10 has eight proline residues, four of which are conserved in TNfn3, which has five prolines. All the prolines are found in the loop regions, or at the start or end of strands.

The three Ig domains are representative of two different Ig structural subsets, and are more diverse in sequence than the fnIII domains (Table 1). Two domains are isolated modules from giant muscle proteins, from human titin (TI I27) and from Caenorhabditis elegans twitchin.
(TWIg18'). These are members of the intermediate, or I-set. They are low in sequence identity (~12%), but have a larger number of structurally equivalent residues (~55%). CD2 domain 1 (CD2d1) from rat is a member of the V-set, named after the antibody variable domains. It has low sequence identity to TWIg18 and TI I27 (~10%) but has a significant structural similarity (~50% equivalent residues). All three Ig domains have a number of prolines but TWIg18 is the only domain with a cis proline (in the B–C loop). In the Ig domains the conserved tryptophan residue is found in the C strand, and CD2d1 has a second, solvent-exposed tryptophan in the A' strand.

A common structural core
Figure 2 shows the sequence alignment based on the structural superposition of all five proteins. There is a common structural core composed of residues from the B, C, E, F and G strands where the backbone atoms can be aligned in all five structures, with a pairwise backbone root mean square deviation (rmsd) ranging from 0.7 – 1.3 Å (mean = 1.0 Å) (Figure 2). Examination of these structurally equivalent residues in the three-dimensional structures shows that they form an interacting network of ten residues in the hydrophobic core of the proteins. Figure 3 shows the common core residues of TWIg18' and TNfn3. Although they are apparently unrelated in evolution, being members of different superfamilies, 'TWIg18' and TNfn3 have 24 residues (10 buried, 14 surface) that are structurally equivalent (Table 1). There is no significant residue identity in the common core of these proteins, but the pattern of inter-residue interactions is similar. These common interactions are centered around the buried residues in the B and F strands (yellow and orange in Figure 3) on opposite sheets of the sandwich. Common core residues from strands C, E and G (green, blue and red, respectively, in Figure 3) pack onto these sidechains from the B and F strands. All five proteins in our study have a pattern of hydrophobic core interactions that is the same in this region.

Stability and folding kinetics vary considerably
All the proteins have a tryptophan residue buried in the hydrophobic core; thus folding can be monitored by

<table>
<thead>
<tr>
<th>Protein</th>
<th>TI I27</th>
<th>TWIg18'</th>
<th>CD2d1</th>
<th>TNfn3</th>
<th>FNfn10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family</td>
<td>Ig I-type</td>
<td>Ig I-type</td>
<td>Ig V-type</td>
<td>fnIII</td>
<td>fnIII</td>
</tr>
<tr>
<td>Number of residues</td>
<td>89</td>
<td>93</td>
<td>98</td>
<td>92</td>
<td>96</td>
</tr>
<tr>
<td>TI I27</td>
<td>-</td>
<td>9</td>
<td>10</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>TWIg18'</td>
<td>54 (1.7)</td>
<td>-</td>
<td>14</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>CD2d1</td>
<td>47 (1.5)</td>
<td>49 (1.2)</td>
<td>-</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>TNfn3</td>
<td>22 (1.3)</td>
<td>24 (1.1)</td>
<td>25 (0.9)</td>
<td>-</td>
<td>22</td>
</tr>
<tr>
<td>FNfn10</td>
<td>22 (1.4)</td>
<td>24 (1.2)</td>
<td>25 (1.0)</td>
<td>70 (1.2)</td>
<td>-</td>
</tr>
</tbody>
</table>

Above the diagonal the table gives the pairwise sequence identity (see Figure 2). Below the diagonal the table shows the number of residues where the backbone can be superimposed, and in parentheses the backbone rmsd (in Å) for those residues.

Figure 2
Structure-based sequence alignment. The structures were aligned in a pairwise fashion using the program Insight and compared by inspection. An alignment of all five proteins was made, and regions where the structures superimposed were determined. Regions of alignment are shown as follows: red, all five (or four) proteins align; green, Ig proteins align; brown, FnIII proteins align; orange, three proteins align. See also Table 1. Residue numbers are indicated.

Table 1
Similarity of the ββ-sandwich proteins described.

<table>
<thead>
<tr>
<th>Protein</th>
<th>TI I27</th>
<th>TWIg18'</th>
<th>CD2d1</th>
<th>TNfn3</th>
<th>FNfn10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family</td>
<td>Ig I-type</td>
<td>Ig I-type</td>
<td>Ig V-type</td>
<td>fnIII</td>
<td>fnIII</td>
</tr>
<tr>
<td>Number of residues</td>
<td>89</td>
<td>93</td>
<td>98</td>
<td>92</td>
<td>96</td>
</tr>
<tr>
<td>TI I27</td>
<td>-</td>
<td>9</td>
<td>10</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>TWIg18'</td>
<td>54 (1.7)</td>
<td>-</td>
<td>14</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>CD2d1</td>
<td>47 (1.5)</td>
<td>49 (1.2)</td>
<td>-</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>TNfn3</td>
<td>22 (1.3)</td>
<td>24 (1.1)</td>
<td>25 (0.9)</td>
<td>-</td>
<td>22</td>
</tr>
<tr>
<td>FNfn10</td>
<td>22 (1.4)</td>
<td>24 (1.2)</td>
<td>25 (1.0)</td>
<td>70 (1.2)</td>
<td>-</td>
</tr>
</tbody>
</table>
changes in intrinsic fluorescence. They have no disulphide bonds, but all have a number of proline residues; only refolding phases that are independent of proline isomerization are considered here. The stability of the proteins at 0 M denaturant, \( \Delta G_{D-N} \) (where D is the denatured state and N the native state), varies significantly, from 4.0 to 9.4 kcal mol\(^{-1}\) (Table 2). All proteins fit a two-state equilibrium unfolding model, that is, at equilibrium only two states, the native and denatured states, are populated. There is a considerable range in both refolding and unfolding rates. Refolding rates vary from 2 to 240 s\(^{-1}\) and unfolding rates from \( 2 \times 10^{-4} \) to \( 2 \times 10^{-3} \) s\(^{-1}\).

The more stable proteins have marginally stable folding intermediates

To detect the presence of folding intermediates the observed folding kinetics were compared with the kinetics expected for a two-state folding mechanism. In a two-state system, the equilibrium and kinetic data are equivalent. The observed folding rate constant reflects the free-energy difference between the denatured state, D, and the transition state for folding, \( \dot{\xi} \). If a folding intermediate, I, is populated during folding, then the folding rate will reflect the free-energy difference between I and \( \dot{\xi} \). As I is more stable than D, the observed folding rate will be lower than predicted from the equilibrium \( \Delta G \) and \( k_u \). The data show evidence for the presence of a low-stability folding intermediate populated at low concentrations of denaturant in three of the proteins, FNfn10, CD2d1 and TI I27; there is a 'roll over' in the refolding kinetics at low denaturant concentrations (Figure 4). The deviation from two-state behaviour is not large, meaning that the intermediates have a comparatively low stability, relative to D. Kinetic data alone cannot distinguish on- and off-pathway intermediates. More work will be needed to characterize fully the nature and stability of these intermediates.

Folding rates correlate with stability

As we know that significant changes in stability and folding kinetics can result from single-residue changes in a single protein, it is important that we compare proteins with a wide range of stabilities and folding/unfolding rates, so that general trends, rather than small differences, are observable. The proteins here fit these general criteria. To extend the range of our analysis we include in our comparisons data from Spitzfaden et al. [7] on the ninth fnIII domain from human fibronectin (FNfn9), a significantly less stable Ig-like domain (\( \Delta G_{D-N} \approx 1.0 \) kcal mol\(^{-1}\)); this increases the range of both \( k_f \) and \( k_u \) to more than two orders of magnitude.

The logarithms of the folding rate constants vary directly with protein stability (Figure 5a). This is a unique observation. Two recent surveys of protein folding have shown that there is no general correlation between folding rates and protein stability (Figure 5b) [8,9]. Thus, this is not a general observation; neither is it the case for three other protein families where folding rates have been determined (Figure 5b). The correlation between log \( k_f \) and stability is a feature of these Ig-like proteins.

There is no such correlation between unfolding rate constants and stability for these Ig-like proteins (Figure 5c). This is in contrast to the weak general correlation between unfolding rates and stability observed when all proteins are considered (Figure 5d) [8,9].

\( \beta \) and contact order are not correlated with folding rates

Are there other folding patterns that can be discerned in this analysis? The position of the transition state for folding
on the folding coordinate, $\beta^\text{folding}$, can be assessed by the value of $m$ for folding ($m_{\text{D-N}}$) relative to $m_{\text{D-N}}$. The $m$-value is a measure of the change in solvent exposure going from one state to another. If a transition state ($\dagger$) is very native-like then $\beta^\text{folding} = m_{\text{D-N}}$ will be close to 1. These Ig-like proteins have $\beta^\text{folding} \approx 0.7$ (0.5–0.9) (Table 2). There is no correlation between $\beta^\text{folding}$ for these proteins and $k_\beta$ or $k_\text{un}$; thus the position of the transition state on the folding coordinate does not determine the relative stability of the transition state. The similarity of $\beta^\text{folding}$ values might imply similar transition-state structures in these proteins; this should be taken with caution, however, as many all-$\beta$ proteins, with very different structures, have similar $\beta^\text{folding}$ values [8,9].

In a survey of two-state proteins Plaxco et al. [9] observed that folding rates are correlated with ‘contact order’, a measure of the number of local contacts in the final structure. There is no correlation, within this set of proteins, of $k_\beta$ or $k_\text{un}$ with contact order. As these proteins have a similar structure, they all have a similar contact order (Table 2), yet they have folding and unfolding rates that vary by more than two orders of magnitude.

**Common folding behaviour suggests a common folding mechanism for Ig-like β-sandwich proteins**

By comparison of the folding of six different Ig-like proteins with a broad range of stability (1.0–9.4 kcal mol$^{-1}$), folding and unfolding rates ($0.3$–$240$ s$^{-1}$ and $5 \times 10^{-2}$–$2 \times 10^{-4}$ s$^{-1}$, respectively), it is possible to distinguish broad patterns of folding behaviour. The significant correlation between folding rates and protein stability strongly suggests that the folding processes of these structurally similar proteins share common features.

### Stability of the transition state for folding is proportional to the stability of the native state

Kinetic experiments measure the difference in free energy between the lowest-energy denatured state and the transition state for folding, $\Delta G_{\text{D-I}}$. Correlation between log $k_\beta$ and $\Delta G_{\text{D-N}}$ implies a correlation between the free energy of the transition state, $\Delta G_{\text{D-I}}$, and $\Delta G_{\text{D-N}}$. That is, the interactions that confer intrinsic stability on N also stabilize the folding transition state $\dagger$. Note that this means that unfolding rates, which depend on the energy difference between N and $\dagger$, will not correlate with stability. This is in agreement with the experimental data.

### Stability of the folding intermediate is related to the stability of the native state

The presence of a folding intermediate in some but not all of the proteins does not imply a different folding mechanism. The three most stable proteins (FNfn10, TI I27 and CD2d1) all have marginally stable folding intermediates; the least stable proteins (TNfn3 and TWIg18) do not. If the stability of I is related to the stability of the native state (N), then in the less stable proteins I becomes less stable than the denatured state (D) and folding becomes, apparently, two-state. Proteins that fold by three-state kinetics have been shown to move to a two-state folding mechanism as the folding intermediate is destabilized, relative to the denatured state, by denaturants, temperature.

### Table 2

<table>
<thead>
<tr>
<th>Conditions* (denaturant pH, temperature)</th>
<th>$\Delta G_{\text{D-I}}$ (kcal mol$^{-1}$)</th>
<th>$k_\beta$ (s$^{-1}$)</th>
<th>$k_\text{un}$ (s$^{-1}$)</th>
<th>$\beta^\text{folding}$†</th>
<th>Contact order</th>
</tr>
</thead>
<tbody>
<tr>
<td>TI I27†</td>
<td>GdmCl pH 7.4, 25°C</td>
<td>7.5</td>
<td>32</td>
<td>$4.9 \times 10^{-4}$</td>
<td>0.9</td>
</tr>
<tr>
<td>TW Ig18§</td>
<td>Urea pH 5.0, 25°C</td>
<td>4.0</td>
<td>1.5</td>
<td>$2.8 \times 10^{-4}$</td>
<td>0.7</td>
</tr>
<tr>
<td>CD2d1</td>
<td>Urea pH 5.0, 20°C</td>
<td>6.8</td>
<td>18</td>
<td>$1.7 \times 10^{-3}$</td>
<td>0.7</td>
</tr>
<tr>
<td>TNh3§</td>
<td>Urea pH 5.0, 25°C</td>
<td>5.3</td>
<td>2.9</td>
<td>$4.6 \times 10^{-4}$</td>
<td>0.7</td>
</tr>
<tr>
<td>FNfn10</td>
<td>GdmSCN pH 5.0, 25°C</td>
<td>9.4</td>
<td>240</td>
<td>$2.3 \times 10^{-4}$</td>
<td>0.5</td>
</tr>
<tr>
<td>FNfn9¶</td>
<td>GdmCl pH 4.8, 25°C</td>
<td>$-1.0$</td>
<td>0.3</td>
<td>$5 \times 10^{-2}$</td>
<td>0.7</td>
</tr>
</tbody>
</table>

*Proteins were examined at a pH optimal for individual stability. Where $20^\circ$C was used, the stability, folding and unfolding rates are altered by $<10\%$, compared with $25^\circ$C [SJH and JC, unpublished results]. *Calculated using $\beta^\text{folding} = 1- (m_{\text{D-N}}/m_{\text{D-N}})$ from unfolding data. Thus $\beta^\text{folding}$ can be determined accurately for both two-state and non-two-state systems. † Data taken from [28]. ‡ Data taken from [30]. § Calculated from kinetic data taken from [7]. GdmCl, guanidinium chloride; GdmSCN, guanidinium thiocyanate.
or mutation. In the case of barnase it has been demonstrated that the folding pathway remains essentially the same under these circumstances, but I can no longer be observed in a kinetic experiment as it is not populated in the destabilizing conditions [10,11].

Is the folding pathway determined by structural constraints?

Defining the folding pathway of a protein involves mapping how structure formation progresses during folding. Defining the structure of the transition state for folding gives information about the folding pathway. Structures with similarly structured transition states are likely to have similar folding pathways.

A number of situations could apply in investigations into the folding of structurally related proteins. If the proteins have different folding pathways, then the transition states would not be similar; that is, they would involve different regions of the structure in different proteins. In that case we should expect to see no correlation between folding rate and stability. If the proteins have similar folding pathways, that is, similar transition states, there are two possible cases. In the first, the folding nucleus involves interactions in a part of a protein that is not involved in determining the structure/stability of the fold. In this case there would be no relationship between folding rate and stability. A lack of correlation between folding rate and stability need not imply that proteins fold by different pathways. The alternative is that the proteins fold by the same pathway and the interactions important for nucleating folding and stabilizing the transition state are the same as those that define and stabilize the fold. In this case we would expect to see a correlation between folding rate and stability. This is what we observe for these Ig-like proteins. Thus the most likely explanation is that the main determinants of the folding pathway are to be found in the common elements of structure that define the fold.

As we have shown (Figures 2, 3), a common structural core is made up of a few residues from strands B, C, E, F and G. Buried residues from these strands, centred around

Figure 4

The dependence of the refolding and unfolding rates (in s⁻¹) on the concentration of denaturant. The line shows the kinetics expected for a two-state kinetic process. At low concentrations of denaturant log \( k_\text{off} \) becomes less dependent on [denaturant], and deviates from the expected curve in CD2d1, TI I27, and FNfn10. This non-linearity (which is independent of protein concentration) indicates the presence of a folding intermediate (see text). Some of the data have been described in detail elsewhere: TW Ig18′ (unfolding kinetics only) [27]; TI I27 [28]; TNfn3 [30]. Note that for FNfn10, [GdmSCN] has been adjusted to account for non-linear dependence of \( \Delta G \) on denaturant concentration (see the Materials and methods section).
strands B and F, make up a common network of interactions in the hydrophobic core of the proteins, which defines the common fold. Our results suggest that the proteins use stabilizing interactions in the common core to stabilize the folding transition state. We are now able to present a hypothesis that can be tested explicitly: that residues in this common core, centred around the B and F strands, on opposite sheets (yellow and orange residues in Figure 3), constitute a folding nucleus in these β-sandwich proteins. The folding nucleus would thus largely consist of tertiary interactions. Whether a structurally identical folding nucleus is conserved cannot be predicted. The relative importance of specific interactions in the common core may be different in the different proteins. Therefore protein engineering analysis is needed to build up a detailed picture of the individual folding nuclei to determine
whether the common features include a subset of interactions making up a specific folding nucleus.

**Comparison with theoretical and experimental results**

Shakhnovich and coworkers have proposed that a specific nucleation site for the Ig-like β-sandwich proteins will be located in the hydrophobic core, involving interactions with the residues corresponding to A18, I20, W22, L34, V70, A84 and F88 in TNfn3, that is, interactions between strands B, C E, F and G [5]. Our results are in general agreement with this proposition, although our common core does not extend to include residues A84 and W22 (Figure 3). A detailed Φ-value analysis of the proteins in our study will be necessary to test further the theoretical model as proposed, as only structural detail can distinguish the relative importance of specific interactions within the folding nucleus.

The limited experimental data available support our model. Our hypothesis suggests that peripheral interactions will have a minor role, if any, in nucleation of folding of Ig-like proteins. We have shown that TNfn3 can be stabilized by addition of two additional residues at the C terminus, and that this stabilization is reflected only by a decrease in the unfolding rate; that is, the C terminus is not involved in the formation of the folding nucleus [12]. A partially structured folding intermediate, stabilized by the addition of Na₂SO₄, has been characterised in CD²d1 [13–16]. The data are consistent with the presence of structure in this common core region in the transition state for folding. Initial Φ-value analysis of TNfn3 shows that most residues with high Φ-values correspond to residues in the putative folding nucleus described here (SJH and JC, unpublished results).

**Comparison with other protein families**

We have argued that the common interactions important in defining the structure of these Ig-like proteins are also used to guide folding. To what extent is this a common feature of proteins? Although the study presented here is the first comparison of proteins that share the same fold but are in different superfamilies, a few individual protein families have been studied. However, in all cases they have a conserved function and/or closely related sequences. A family of small all-β proteins for which there are data are the SH3 domains [17–20]. These show no correlation between folding or unfolding rates and stability (Figure 5). Two of these proteins, the SH3 domain from α-spectrin and the Sre SH3 domain have been studied in detail [17,18,21–23]. The proteins have a relatively compact folding nucleus located at one end of the structure in a loop/turn region. The interactions in the transition state are conserved, although the identity of the residues that take part in the interactions are not [24]. Clearly, the constraints on folding imposed by the SH3 structure are different from those in the Ig-like proteins. Protein families may have conserved folding pathways, but if the interactions that serve to nucleate folding are not the same as those that define the structure and stability, then no correlation between folding rates and stability will be observed. This serves to emphasise that whereas a correlation between speed of folding and stability is likely to imply a common folding nucleus, the opposite may not be true. In acyl-coenzyme A-binding proteins, which have an all-α fold, there is no apparent relationship between refolding rates and stability [25] (Figure 5b) but unfolding rates and stability are related (Figure 5d). The data set is small, but again it suggests that the all-α fold of these proteins places different structural constraints on the folding pathway.

If two Ig-like proteins had the same core structure, but different loops, then we would propose that they would have the same folding rates, as the folding nucleus would be maintained intact. Any differences in stability would be largely attributable to changes in the rate of unfolding. This is exactly the behaviour observed in another all-β family, the cold-shock proteins (Csp). Three CspB proteins with stabilities that vary from 2.7–6.3 kcal mol⁻¹ have approximately the same folding rate and all fold in a two-state manner [26], so that the differences in stability are related to differences in unfolding rates (Figure 5b,d). In these proteins the sequence identity is high, and the core structure is absolutely conserved; all the differences are peripheral, in loop regions of the structure. Possibly, the CspB proteins would have a similar pattern of folding as we observe for Ig-like proteins. However, E Shakhnovich has proposed (unpublished results) that proteins with this fold have a conserved folding nucleus that is separate from the structural core. More data are needed to explore the possibility further, but this serves to emphasize the importance of studying a diverse set of proteins if one is to discern folding patterns.

**Biological implications**

We have shown that there are common features in the folding of six different Ig-like β-sandwich proteins, with a strong correlation between folding rates and stability. The most likely explanation is that these proteins share a common folding pathway, with a folding nucleus composed of interactions between strands on both sheets, and at some considerable distance in the amino-acid sequence. This gives us a specific hypothesis to examine further, through detailed protein-engineering analysis of the structure of the transition states (and folding intermediates, where they exist), and by analysis of other proteins from the same superfamilies and from the related cadherin superfamily. It is worth stressing again that the proteins in this study are from two different superfamilies; that is, there is no evidence that they are related in evolutionary terms. Thus the ‘conservation’ of folding pathways may be a thermodynamic feature of the fold rather than an evolutionary constraint.
Materials and methods

Materials

The production of recombinant proteins has been described previously: TW Ig18 [27]; TI27 [28]; CD2d1 [29]; TNNf3 [30]; FNfn10 [31].

Stability

The stability of each protein was determined by equilibrium denaturation using urea, guanidine chloride (GdmCl) or guanidine thiocyanate (GdmSCN) as a denaturant, depending on stability, as described in [27]. For FNfn10, GdmSCN was used as a denaturant. The relationship between AG and [GdmSCN] is strongly non-linear (AR Clarke, personal communication; EC and JC, unpublished results) and so the denaturant concentration [D] was adjusted to account for this non-linearity: 
\[ [D] = [M] \times \frac{6.47}{6.47 + [M]} \]
This value was taken from unpublished data generously provided by M Pandya and AR Clarke (Bristol).

Folding kinetics

Folding and unfolding kinetics were examined using stopped-flow fluorimetry in a range where both rate constants and amplitudes were independent of protein concentration (0.5–2 μM). Unfolding was initiated by a jump into high concentrations of denaturant. Measured folding rates varied significantly, with traces collected for between 10 s and 75 min. The very slow unfolding reactions were followed by standard fluorimetry, following extrapolation of refolding rates at low denaturant concentrations. 

\[ [D] = [M] \times \frac{6.47}{6.47 + [M]} \]
This value was taken from unpublished communication; EC and JC, unpublished results)

Acknowledgements

We would like to thank Mark Bycroft, Cyrus Chothia and Eugene Shakhnovich for helpful discussions, and Maya Pandya, Tony Clarke, Eugene Shakhnovich and Leonid Mimy for sharing unpublished data. Clones for FNfn10 and CD2d1 were generously provided by Kevin Plaxco and Paul Driscoll. We acknowledge funding from the Wellcome Trust (JC, SF) and the Consejo Nacional de Ciencia y Tecnología (EC) and the MRC (SJH).

References