Understanding the sequence determinants of conformational switching using protein design

SEEMA DALAL^{1,3} AND LYNNE REGAN^{1,2}

¹Department of Chemistry, Yale University, 266 Whitney Avenue, Bass 322, New Haven, Connecticut 06520

²Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, Bass 322, New Haven, Connecticut 06520

(RECEIVED March 7, 2000; FINAL REVISION June 8, 2000; ACCEPTED June 8, 2000)

Abstract

An important goal of protein design is to understand the forces that stabilize a particular fold in preference to alternative folds. Here, we describe an extension of earlier studies in which we successfully designed a stable, native-like helical protein that is 50% identical in sequence to a predominantly β -sheet protein, the B1 domain of *Streptococcal* IgG-binding protein G. We report the characteristics of a series of variants of our original design that have even higher sequence identity to the B1 domain. Their properties illustrate the extent to which protein stability and conformation can be modulated through careful manipulation of key amino acid residues. Our results have implications for understanding conformational change phenomena of central biological importance and in probing the malleability of the sequence/structure relationship.

Keywords: B1 domain; conformational switching; fibrils; Paracelsus Challenge; protein design; Rop

The amino acid sequence of a protein contains the information necessary to define its fold (Anfinsen, 1973). Although this observation suggests that a particular amino acid sequence specifies a single fold, examples of sequences that can adopt multiple secondary structural conformations are emerging. An early indication that the tertiary environment of a sequence is an important component in determining its conformation came from a structural database search by Kabsch and Sander (1984). They observed that identical pentapeptide sequences adopted either an α -helical or a β -sheet conformation, depending on the protein in which they were found. More recently, hexa- (Cohen et al., 1993) and heptapeptide sequences (Mezei, 1998) that adopt different conformations have been identified through similar surveys of larger structural databases.

Experimentally, it has been demonstrated that a carefully chosen 11-residue sequence can adopt either an α -helical or β -sheet conformation, depending on its placement within a small protein (Minor & Kim, 1996), the B1 domain of IgG-binding protein G (Fig. 1A). It has also been proposed that when most of the sequence of the second hairpin of the B1 domain replaces the B1 domain helix, it is forced to adopt a helical conformation (Cregut et al., 1999). Protein conformation can also be modulated by external environmental factors. In so-called “switch peptides,” the

dual nature of certain sequences is manifested upon changing solvent conditions (Mutter et al., 1991). For example, certain peptides that form monomeric α -helices can be induced to associate in a β -sheet conformation by changing the pH of the solution (Cerpa et al., 1996).

Localized dramatic changes in secondary structure conformation are physiologically important for the correct functioning of a number of proteins. An α to β conformational switch is seen in a portion of the G-protein EF-Tu upon GTP hydrolysis (Abel et al., 1996). It has been suggested that conformational switches of this nature may represent a general mechanism by which to mediate protein activation in, for example, the serpin family of serine protease inhibitors (Mottonen et al., 1992; Wei et al., 1994; Abel et al., 1996). Conformational switching has also been proposed to occur during the folding of a protein. Goto and colleagues have observed the transient population of a nonnative α -helical intermediate on the folding pathway of a predominantly β -sheet protein, β -lactoglobulin (Hamada et al., 1996).

A number of “misfolding diseases” are associated with protein conformational change. In amyloid diseases, fibril formation is a result of the conversion of soluble protein into regular aggregates. Regardless of the structure of the soluble protein, the conformation of the associating monomers is predominantly β -sheet (Harrison et al., 1997; Cohen, 1999). It has been proposed that all proteins may have the potential to form such misfolded and aggregated assemblies under appropriate conditions (Chiti et al., 1999; Gross et al., 1999).

A goal of protein design is to dissect the balance between local and long-range interactions that stabilize a particular fold in pref-

Reprint requests to: Lynne Regan, Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, Bass 322, New Haven, Connecticut 06520; e-mail: lynne@csb.yale.edu.

³Current address: Department of Cell Biology and Physiology, Washington University School of Medicine, St. Louis, Missouri 63110.

Fig. 1. Ribbon representations (Kraulis, 1991) of the folds. **A:** The B1 domain of *Streptococcal* IgG-binding protein G (Gronenborn et al., 1991). **B:** Rop (Banner et al., 1987; Eberle et al., 1990). **C:** Sequences alignment for the variants discussed. Line 1, secondary structure representation of Rop; Line 2, locations of the “a” and “d” core positions in Rop (the Rop core residues align from Line 3 with these positions); Line 3, WT Rop sequence; Line 4, Janus sequence; Line 5, Janus-61 sequence; Line 6, Janus-55 sequence; Line 7, Janus-66 sequence; Line 8, Janus-86 sequence; Line 9, core residues of the B1 domain; Line 10, wild-type B1 sequence; Line 11, secondary structure representation of the B1 domain. The color coding for the sequences is as follows: residues that are identical to Rop are shown in red, residues that are identical to the B1 domain are shown in blue, residues that are common to both Rop and the B1 domain are shown in purple, and the residues that belong to neither Rop nor the B1 domain are shown in black. The seven amino acid C-terminal tail of Rop (Gly-Asp-Asp-Gly-Glu-Asn-Leu) that extends beyond the sequences depicted for Rop, Janus, and the Janus variants is not shown. It was retained in the variants because it appears to increase the solubility of wild-type Rop (Smith et al., 1995). We have created a Janus variant that lacks this C-terminal tail. This protein is helical; however, its expression levels are considerably lower than Janus, possibly because it is more susceptible to proteolytic degradation (S. Dalal & L. Regan, unpubl. obs.).

erence to the myriad of conformational possibilities (including the unfolded state) available to a given sequence (Levitt et al., 1997). The results of such investigations provide valuable insights into how these forces are modulated in both conformational change and misfolding phenomena.

To stimulate research into the factors that specify a particular fold over alternative folds, Creamer and Rose issued the "Paracelsus Challenge" (Rose & Creamer, 1994). The challenge was to convert one protein fold to another while retaining at least 50% sequence identity to the original fold. Within these constraints, we successfully designed a protein that has 50% sequence identity to the B1 domain, a predominantly β -sheet protein (Gronenborn et al., 1991) (Fig. 1A), but which adopts a helical conformation (Dalal et al., 1997a, 1997b). The designed protein, which we named Janus, is modeled on the four-helix bundle protein Rop (Banner et al., 1987; Eberle et al., 1990) (Fig. 1B). Rop and the B1 domain were chosen as the two proteins to interconvert because not only do they have dramatically different folds, but their secondary structural content is also completely different. Circular dichroism (CD) spectroscopy can therefore be used as a sensitive method by which to follow changes in secondary structure and, hence, fold as the protein sequence is changed. This is important because it allows rapid assessment of the designs prior to their high resolution structural determination. Also, it can be applied in circumstances where the properties of the proteins preclude high resolution structural determination, e.g., low solubility.

Janus is a stable, helical dimeric protein that displays the thermodynamic properties associated with native proteins (vide infra). Key elements of the design were the optimization of the hydrophobicity and packing of the core, the incorporation of amino acids with high helix-forming propensities, the avoidance of stretches of residues with high β -sheet forming propensities, the choice of turn residues and the maintenance of an overall charge distribution similar to that of wild-type Rop (Dalal et al., 1997a).

In this paper, we describe the design and characterization of a series of Janus variants with increasing sequence identity to the B1 domain. In these variants, we change some of the key design features of Janus. The questions addressed are: How many B1 domain residues (beyond the 50% in Janus) can we incorporate while maintaining a helical fold? At what stage in the design do we induce a conformational switch from α -helix back to β -sheet? At what point do these proteins cease to be "native-like" and start populating nonnative states? We define a "native-like" protein as one that displays a cooperative and reversible thermal denaturation transition. The cooperativity of the thermal denaturation transition is a very important test of the native-like character of a designed protein because it explicitly reflects the balance of enthalpic and entropic forces by which a protein is stabilized. A number of designed proteins with significant secondary structure content, but with poorly packed hydrophobic cores, display either very broad or no thermal denaturation transitions. Such proteins have been described as being more similar to "molten globules" than to native proteins (Bryson et al., 1995). Reversibility of the thermal denaturation transition is also a key feature of a native-like protein as the observation of irreversible denaturation reflects the inability of the unfolded or partially folded protein to regain the native structure, perhaps due to trapping in an aggregated nonnative state (Lai et al., 1996).

Here, we describe the design rationale and present results for four variants of Janus that are closer in sequence identity to the B1 domain and we discuss the properties of these proteins in the context of other conformational change and association phenomena.

Results

Janus variant with an enhanced number of high β -sheet forming propensity residues (Janus-61)

In the Janus design, the goal was to stabilize the helical fold preferentially over the β -sheet fold. Janus was modeled on the four-helix bundle protein, Rop. The hydrophobic core of Rop is very regular and is comprised of eight layers, with each helix contributing one residue per layer. To specify a stable helical fold, we incorporated an idealized hydrophobic core comprised exclusively of alanine and leucine residues. Previous work had demonstrated that a Rop variant with alanine at the "a" position and leucine at the "d" position of the helical heptad repeat in all layers (except the second and the seventh layers where the packing is reversed) displayed enhanced thermal stability in comparison to wild-type Rop (Munson et al., 1994a, 1996). Another important design criterion in Janus was to avoid contiguous sequences of residues with high β -sheet forming propensity (Kim & Berg, 1993; Minor & Kim, 1994; Smith et al., 1994). However, to satisfy the terms of the challenge, it was not feasible to avoid all residues with high β -sheet forming propensity.

In an attempt to increase sequence identity to the B1 domain while maintaining a helical fold, we have designed a protein that incorporates an additional six mostly surface and boundary position residues from the B1 domain (Tyr3, Lys10, Thr16, Thr18, Thr25, Thr53) (Fig. 1C). These residues correspond to noncore residues in Janus and the "a" and "d" core positions were maintained identical to those in Janus. Most of these residues have high β -sheet forming propensities (Tyr3, Thr16, Thr18, Thr25, Thr53) and all the incorporated residues, except Lys10 and Thr25, map to the β -sheets of the B1 domain. The resulting protein is 61% identical to the B1 domain (Janus-61). Janus-61 has a far-ultraviolet circular dichroism (far-UV CD) signal typical of a helical protein (Fig. 2A) and it displays a cooperative, reversible thermal denaturation transition (Fig. 2B). Perhaps due to the increase in the number of β -sheet forming propensity residues (and the concomitant decrease in helical propensities (Chakrabarty & Baldwin, 1995)), Janus-61 is destabilized relative to Janus (Fig. 2B), but the changes made are not sufficient to disrupt either the helical fold or the native-like character of the protein.

Janus variant with reversed core packing (Janus-55)

Janus-61 is significantly destabilized relative to Janus and therefore it does not provide an ideal scaffold for incorporating additional B1 domain residues. Examination of the sequences of the B1 domain and Rop revealed that a reversed core packing with leucine at the "a" position and alanine at the "d" position for all 8 layers of the hydrophobic core would yield a protein with 55% sequence identity to the B1 domain. Previous studies had demonstrated that a Rop variant with such a core is considerably more thermally stable than a variant with the Janus core (Munson et al., 1994a, 1996). An investigation of the structural basis for the observed increase in stability of this Rop variant is in progress (M. Willis, A.T. Brünger, & L. Regan, unpubl. data). With the aim of generating a more stable scaffold upon which to incorporate further sequence identity to the B1 domain, we created Janus-55 (Fig. 1C) that has such a reversed core. Janus-55 is helical (Fig. 2A) and significantly more stable than Janus, with a T_m of 71 °C vs. a T_m of 62 °C for Janus (Fig. 2B). Hence, the strategy of reversing the

Fig. 2. Wavelength scans and thermal denaturations of Janus, Janus-55, and Janus-61. **A:** Far-UV CD spectra of Janus ($\blacktriangle$), Janus-55 ($\circ$), and Janus-61 ($\bullet$) at 25 °C. The mean residue ellipticity (MRE) of Janus-61 at 5 °C increases to 27,400 deg cm² dmol⁻¹. **B:** Thermal denaturation transitions of Janus ($\blacktriangle$), Janus-55 ($\circ$), and Janus-61 ($\bullet$) monitored at 222 nm. Note that Janus-61 may not be fully folded at 5 °C but it has been represented on the same scale as Janus and Janus-55 for ease of comparison. At 5 °C, with an MRE of 27,400 deg cm² dmol⁻¹, Janus-61 is probably close to its fully folded state.

identity of the core residues proved to be successful and Janus-55 provided a starting point for further designs.

Janus variant with reversed core packing and high β -sheet forming propensity residues (Janus-66)

We combined the reversed core design of Janus-55 with the increased number of mostly high β -sheet propensity residues of Janus-61 (Tyr3, Lys10, Thr16, Thr18, Thr25, Thr53) to create a protein that is 66% identical to the B1 domain (Janus-66; Figs. 1C, 3).

For all the proteins discussed in the paper so far, we determined the pH at which they exhibited maximum structure by monitoring their far-UV CD spectra over the pH range 1 to 8. We found that they are more structured (helical) at low pH, becoming mostly random coil at the higher pHs. All further characterizations were performed at the pH of maximum stability. We performed similar studies with Janus-66 with interesting results. At pH 2, Janus-66 is helical and its thermal denaturation transition is fully reversible (Fig. 4A). At pH 5 the protein is equally helical but at this pH the thermal denaturation transition is irreversible (Fig. 4B).

To examine this phenomenon in greater detail, we performed wavelength scans throughout the thermal melt. These revealed that a dramatic change in secondary structure, from predominantly α -helix to predominantly β -sheet, occurred between 35 and 40 °C (Fig. 4C). Once formed, the β -sheet conformation of Janus-66 is resistant to thermal denaturation. The extreme resistance of the β -sheet conformation to thermal denaturation suggested that it was likely not a monomeric protein, but a stable, soluble aggregate. This behavior is reminiscent of the conformational changes that are associated with a number of diseases in which a soluble protein forms fibrils.

We used electron microscopy (EM) as the most direct method by which to investigate the possibility of fibril formation in the β -sheet form of Janus-66. The electron micrographs we obtained confirm this hypothesis. Protofilaments and amorphous aggregates are clearly visible (Fig. 4D).

It is intriguing to consider possible explanations for the pH dependence of fibril formation. A possible explanation is that at pH 2 only lysine side chains are charged, and hence, there would be a repulsion to aggregation as the protein unfolds. At pH 5 on the other hand, both positively charged lysine side chains and negatively charged aspartate and glutamate side chains may be present. Hence, when the protein is denatured at pH 5, aggregation of the unfolded and/or partially folded conformations occurs, resulting in the formation of multimeric assemblies. When the secondary structure of the Janus-66 is monitored as a function of pH, we observe that at pH 7.2 the protein has lost a significant amount of its secondary structure (Fig. 4E). After a long incubation at pH 7.2, at room temperature, Janus-66 was found to form fibrils (Fig. 4F). These fibrils have a diameter of about 25 nm and are cleaner than those obtained through thermal unfolding at pH 5, probably due to the slower and milder conditions for their elongation. Janus-66, thus, has at least two modes of conversion from the α -helical form to the aggregated β -sheet conformation and it can either form fibrils by thermally-induced unfolding at pH 5 or through pH-induced unfolding at pH 7.2.

Janus variant designed to be a monomeric β -sheet (Janus-86)

We have demonstrated that Janus-66 can convert from a helix to an aggregated β -sheet. Starting with Janus-66, the goal was to determine the changes that would be sufficient to convert it into a monomeric β -sheet protein that adopts a fold similar to that of the B1 domain (Gronenborn et al., 1991). We made 11 amino acid changes in Janus-66 to yield Janus-86. Specifically, we incorporated Gly9, Ala26, Phe30, Trp43, Tyr45, Phe52 to complete the hydrophobic core of the B1 domain and its contacts; Gly38, Gly41, Thr49, to specify the turns; Asp22 to specify the N-cap of the helix; and Asn37 to specify the C-cap of the helix (Gronenborn et al., 1991). The incorporated glycine residues encourage termination of the helices in the helical protein. Furthermore, in making these changes we optimized cross-strand side-chain interactions both for the antiparallel (Smith & Regan, 1995) and parallel strands (Merkel et al., 1999) and destroyed the regularity of the hydrophobic core for Janus-66. Janus-86 is a monomeric protein (data not shown). Its far-UV CD signal is typical of a β -sheet protein (Fig. 5A) and it has a cooperative thermal denaturation transition (Fig. 5B). Moreover, the shape of the near-UV CD signal for Janus-86 is similar to that for the B1 domain, suggesting that the environments of the aromatic residues are conserved in both pro-

Fig. 3. Ribbon representations of the B1 domain and Rop with distribution of residues incorporated in Janus-66 highlighted. **A:** Highlighted in blue is the distribution of residues that were retained from the B1 domain in Janus-66. **B:** Highlighted in blue on the Rop backbone are the residues that were incorporated from the B1 domain in Janus-66. Highlighted in red is the distribution of residues that were incorporated from Rop in Janus-66. The positions of the residues that are common to both Rop and the B1 domain are shown in purple. **C:** Highlighted in red is the distribution of residues that were incorporated from Rop in Janus-66.

teins (data not shown) and implying conservation of the overall tertiary fold. Janus-86 has 86% sequence identity to the B1 domain and one might argue that it would have been expected to adopt a fold similar to that of the B1 domain. The main interest of Janus-86 is that it has 80% sequence identity to Janus-66. Through this design we have demonstrated that the 20% difference in sequence between Janus-66 and Janus-86 is sufficient to convert a protein that can adopt either a helical or fibrillar β -sheet conformation to a monomeric β -sheet.

Analysis: Secondary structure and fold predictions

Although all the variants discussed, except for Janus-86, are able to adopt a helical conformation, they all have very high sequence identity to the B1 domain (from 50 to 66%). These proteins therefore provide an interesting set to compare different sequence/structure analysis methods and to understand the balance between local and long-range interactions in specifying a fold.

One can consider three approaches to assigning a sequence to a particular fold. A direct search of the sequence database can be used to identify proteins with high degree of similarity to the sequence of the unknown fold. As a “rule of thumb” if two natural proteins have 30% or greater sequence identity, one has confidence that they share the same fold. A BLAST (Altschul et al., 1990) sequence database search identified Janus to be most homologous to Rop, even though it has lower sequence identity to Rop than to the B1 domain. The higher homology to Rop reflects the incorporation of a number of residues from the B1 domain into Janus that are conservative changes of residues in Rop. By contrast, the Janus variants that have even higher sequence identity to the B1 domain than does Janus, are identified to be most homologous to the B1 domain. These are obviously incorrect fold predictions for all the proteins except Janus and Janus-86. However, because these are

designed proteins that have not been subjected to normal evolutionary pressures in which the sequences have been allowed to drift, a simple comparison of sequence homology is clearly not the most appropriate method for structure prediction.

A protein’s secondary structure can be predicted from its sequence based on the intrinsic propensities of its individual amino acids to adopt a helical, sheet, or turn conformation. We used the secondary structure prediction algorithm GOR (Garnier et al., 1978) to compare the proteins. Wild-type Rop is predicted to be mainly helical, while the B1 domain is predicted to be a mixture of helix, coil, and sheets. However, in the prediction for the B1 domain, the helix is offset by five amino acids from the actual structure and stretches of coils are predicted in regions that are β -sheets. Janus, Janus-55, Janus-61, and Janus-66 are all predicted to be entirely helical. In fact, these proteins are predicted to be even more helical than Rop. This is probably due to the increase in the number of residues with high helix-forming propensities resulting from the changes (Ala and Leu) introduced by repacking the core. The Janus-86 prediction mirrors that of B1 domain prediction, again with the same inappropriate combination of helix, sheet, and coils.

A conceptually different means of structure prediction is to create a “profile” in which the relative positioning and solvent accessibility of the residues, together with their intrinsic secondary structure forming propensities, define a fold (Jones et al., 1992; Fisher & Eisenberg, 1996). We used the sequence profiling method of Eisenberg and colleagues (Fisher & Eisenberg, 1996) to identify the folds that are most compatible with the sequences of the Janus variants. The results of this search are summarized in Table 1 and the comparisons are presented in terms of a Z-score where a Z-score of greater than 5 ± 1 is considered significant.

The sequences of Janus, Janus-61, and Janus-55 are predicted to be most compatible with the Rop fold. Following Rop, the top ten

Fig. 5. Wavelength scan and thermal denaturation transitions for Janus-86 monitored at 220 nm. **A:** Far-UV CD for Janus-86. **B:** Thermal denaturation transition for Janus-86.

predicted folds for the Janus sequence are a variety of helical proteins; however, the B1 fold is not predicted. For Janus-55 and Janus-61, although not the top prediction, the B1 fold is predicted in the top ten. Thus, in spite of their high sequence identity to the B1 domain, the folds of Janus, Janus-61, and Janus-55 are correctly predicted to be helical by the profiling method. For Janus-86, the B1 domain is by far the most strongly predicted fold, and this protein has been demonstrated experimentally to adopt a monomeric β -sheet conformation. Janus and Janus-86 have high sequence homologies as well as the correct patterning of solvent accessibilities for the Rop and B1 domain folds, respectively; hence, they are correctly predicted by both the BLAST and profiling methods.

The most interesting results are seen with Janus-66. This protein is α -helical at low pH but can be converted into a fibrillar β -sheet

form by changing conditions. For the Janus-66 sequence, the #1 ranked fold prediction is the B1 domain, and the #2 ranked fold prediction is Rop. However, in this case the confidence level for both predictions is high, with very close Z-values. Thus, the “dual nature” of the Janus-66 sequence, which we observed experimentally is being detected by the profiling method.

The experimental results we present clearly demonstrate that Janus-61, Janus-55, and Janus-66 are helical proteins with high mean residue ellipticities that are similar to that of Rop and that their secondary structures are very different from the predominantly β -sheet protein, the B1 domain. The data tracks whether the proteins adopt a predominantly helical Rop-like fold or a predominantly β -sheet fold of the B1 domain and how their physical properties change. However, the limited solubility of these proteins (ranging from between 4–10 μ M) has to date precluded high resolution structural characterization that would allow us to describe their structure and packing at atomic level detail.

Discussion

We have described the properties of a series of designs that explore the sequence space between two small proteins with completely different folds. We demonstrate that it is possible, by careful selection of key amino acid residues, to manipulate the balance of long- and short-range interactions that stabilize either an α -helical or β -sheet conformation.

Our designs provide an intriguing system in which to understand conformational changes and association processes in natural proteins. Janus and Janus-55 are stable helical proteins with no indication, under the conditions tested, of conversion to either monomeric or fibrillar β -sheet forms. Janus-66 folds to a helical conformation, but is significantly less stable than these proteins. Janus-66 can be induced to form β -sheet fibrils by changing conditions. Finally, Janus-86 incorporates a sufficient number of stabilizing interactions to give monomeric β -sheet protein with no indication of α -helix or β -aggregate formation.

The results we describe parallel some of the behavior that is observed in protein misfolding phenomena, for example, with prion proteins. The cellular isoform, Prp^c, is predominantly helical although it contains a number of residues with high β -sheet forming propensity (Jackson et al., 1999). The protein has the ability to adopt two different folds: The mainly helical, monomeric, soluble Prp^c, and the β -sheet, aggregated form that is associated with disease (Cohen, 1999). It is interesting to note that individuals who are homozygous for valine (rather than methionine) at position 129 have a higher likelihood of developing Creutzfeldt Jacob disease, which is associated with the formation of β -aggregates. Valine has both a significantly higher intrinsic β -sheet forming propensity and a greater ability to form stabilizing cross-strand interactions (Smith & Regan, 1995; Merkel et al., 1999) than does methionine. Therefore, the result of this mutation may be to both stabilize the

Fig. 4 (on facing page). Far-UV CD, thermal denaturation transitions, and electron micrographs of Janus-66. **A:** Forward thermal denaturation at pH 2 monitored at 222 nm. The inset shows the forward (●) and reverse thermal (○) denaturation at pH 2. **B:** Forward (●) and reverse (○) thermal denaturation at pH 5 monitored at 222 nm. **C:** Wavelength scans as a function of temperature at pH 5: 20 °C (●), 30 °C (○), 35 °C (■), 40 °C (□), 45 °C (▲), 50 °C (△). **D:** Electron micrographs of the β -sheet conformation of Janus-66 after thermal denaturation at pH 5. **E:** Wavelength scans of Janus-66 as a function of pH: pH 2 (●), pH 3.2 (○), pH 4 (▲), pH 4.5 (△), pH 4.8 (■), pH 6.1 (□), pH 7.2 (◆). The inset shows the MRE at 222 nm as a function of pH. **F:** Electron micrograph of Janus-66 incubated at pH 7.2 at room temperature.

Table 1. Structure predictions for the variants using the sequence profiling method (Fisher & Eisenberg, 1996)

Query protein	#1 Ranked prediction	Z-score for #1 ranked prediction	Rank for either the B1 domain or Rop prediction	Z-score for either the B1 domain or Rop prediction
Janus	Rop	12.13	B1 domain not predicted ^a	N/A
Janus-61	Rop	11.78	B1 domain-#2	7.75
Janus-55	Rop	8.80	B1 domain-#4	3.17
Janus-66	B1 domain	9.07	Rop-#2	7.20
Janus-86	B1 domain	24.15	Rop-#4	4.64

^aFor Janus, a variety of helical proteins were predicted in the top ten folds and the B1 domain was not present in this list.

aggregated β -sheet form of Prpc and to destabilize the monomeric helical form. The differences in behavior we observe among Janus, Janus-55, and Janus-66 to some extent mirrors the differences observed between Prp^c and Prp^{cm129V}.

Materials and methods

Construction of clones and protein expression

The genes encoding the protein sequences were constructed from synthetic oligonucleotides using the polymerase chain reaction, cloned into T7 expression vectors pMR103 (Munson et al., 1994b) or pET11a Δ and their sequences verified by di-deoxyribonucleotide sequencing using standard procedures. Proteins were overexpressed in *E. coli* BL21 (DE3) and purified either by ion exchange chromatography and size exclusion chromatography or reversed-phase high-performance liquid chromatography. Of the proteins discussed, Janus-66 was found entirely in inclusion bodies while Janus-61, Janus-55, and Janus-86 were found both in the inclusion bodies and the soluble fraction. Janus-55 was purified from the soluble portion as well as from inclusion bodies and the purified protein displayed identical properties, regardless of which fraction it was purified from. Janus-61 and Janus-86 were purified from the soluble portion. Janus-61 is particularly susceptible to proteolytic degradation, and hence it was difficult to obtain large quantities to perform extensive characterizations. Janus-61, Janus-55, and Janus-86 were obtained from the soluble portion by sonication in 10 mM Tris·HCl pH 8 buffer and purified by ion exchange (DEAE and Q-Sepharose HP, Pharmacia, Uppsala, Sweden) and size exclusion chromatography (Superdex 75, Pharmacia). To purify Janus-55 and Janus-66 from inclusion bodies, the cell pellet was lysed by sonication in 20 mM Tris·HCl pH 8 buffer, and the inclusion bodies were separated from the contaminant proteins by centrifugation. After separating from the lysate supernatant, the pellet was washed twice with lysis buffer and then solubilized in 10 mM sodium acetate pH 5, 8 M urea. The supernatant from this extraction was injected onto a preparative-scale C18 HPLC column and purified in a linear gradient of acetonitrile and water containing 0.1% (v/v) trifluoroacetic acid. Protein pools were lyophilized to dryness. To refold the protein, urea was added to a final concentration of 8 M, and dialysed into either 10 mM glycine·HCl, pH 3.3 for Janus-55 or 10 mM sodium acetate, pH 5 or 10 mM glycine·HCl, pH 2 for Janus-66. The identity of all proteins were confirmed by amino acid analysis and MALDI mass spectrometry.

Circular dichroism spectroscopy

Experiments were performed at a protein concentration of ~ 5 – $10 \mu\text{M}$ in a 2 mm pathlength cuvette on a 62DS Circular Dichroism Spectrometer (Aviv Instruments, Lakewood, New Jersey) running Aviv Software. The pH of maximum structure was determined for each protein, and this pH was used in all further experiments unless explicitly stated in the text. Janus was studied in 10 mM sodium acetate, pH 5; Janus-55 was studied in 10 mM glycine·HCl pH 3.3; Janus-61 was studied in 10 mM glycine·HCl, pH 2; Janus-66 was studied either in 10 mM sodium acetate, pH 5 or 10 mM glycine·HCl, pH 2; and Janus-86 was studied in 10 mM Tris·HCl pH 7.5. Thermal denaturations for Janus, Janus-55, Janus-61, and Janus-66 were monitored at 222 nm and that for Janus-86 was monitored at 220 nm over the range shown for each protein, with a temperature step size of either 1 or 2 °C, 1 min equilibration time and 30 s signal averaging time.

Electron microscopy

Transmission electron microscopy (TEM) was performed on a Zeiss10 TEM operating at 80 kV. The protein samples were applied to formvar/carbon-coated grids and blotted with filter paper. The sample was then negatively stained with 1% phosphotungstic acid for about 30 s.

Acknowledgments

We would like to thank members of the Regan laboratory for helpful discussions and for comments on the manuscript. Special thanks to Dr. Marina Ramirez-Alvarado for help with the EM and to Jane Merkel for discussions about β -sheets. This work was supported by the NIH. L.R. is a Dreyfus Foundation Teacher-Scholar.

References

- Abel K, Yoder MD, Higenfeld R, Jurnak F. 1996. An alpha to beta conformational switch in EF-Tu. *Structure* 4:1153–1159.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment tool. *J Mol Biol* 215:403–410.
- Anfinsen CB. 1973. Principles that govern the folding of protein chains. *Science* 181:223–230.
- Banner D, Kokkinidis M, Tsernoglou D. 1987. Structure of the ColE1 Rop Protein at 1.7 Å Resolution. *J Mol Biol* 196:657–675.
- Bryson JW, Betz SF, Lu HS, Suich DJ, Zhou HX, O'Neil KT, DeGrado WF. 1995. Protein design: A hierarchic approach. *Science* 270:935–941.
- Cerpa R, Cohen FE, Kuntz ID. 1996. Conformational switching in designed peptides: The helix/sheet transition. *Fold Des* 1:91–101.

- Chakrabarty A, Baldwin RL. 1995. Stability of alpha helices. *Adv Protein Chem* 49:141–177.
- Chiti F, Webster P, Taddei N, Clark A, Stefani M, Ramponi G, Dobson CM. 1999. Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *Proc Natl Acad Sci* 96:3590–3594.
- Cohen BI, Presnell SR, Cohen FE. 1993. Origins of structural diversity within sequentially identical hexapeptides. *Protein Sci* 2:2134–2145.
- Cohen FE. 1999. Protein misfolding and prion diseases. *J Mol Biol* 293:313–320.
- Cregut D, Civera C, Macias MJ, Wallon G, Serrano L. 1999. A tale of two secondary structure elements: When a β -hairpin becomes an α -helix. *J Mol Biol* 292:389–401.
- Dalal S, Balasubramanian S, Regan L. 1997a. Protein alchemy: Changing β -sheet into α -helix. *Nat Struct Biol* 4:548–552.
- Dalal S, Balasubramanian S, Regan L. 1997b. Transmuting α helices and β sheets. *Fold Des* 2:R71–R79.
- Eberle W, Klaus W, Cesareni G, Sander C, Rösch P. 1990. Proton nuclear magnetic resonance assignments and secondary structure determination of the ColE1 Rop (Rom) protein. *Biochemistry* 29:7402–7407.
- Fisher D, Eisenberg D. 1996. Fold recognition using sequence-derived predictions. *Protein Sci* 5:947–955.
- Garnier J, Osguthorpe DJ, Robson B. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J Mol Biol* 120:97–120.
- Gronenborn AM, Filpula DR, Essig NZ, Achari A, Whitlow M, Wingfield PT, Clore GM. 1991. A novel highly stable fold of the immunoglobulin binding domain of *Streptococcal* protein G. *Science* 253:657–661.
- Gross M, Wilkins DK, Pitkeathly MC, Chung EW, Higham C, Clark A, Dobson CM. 1999. Formation of amyloid fibrils by peptides derived from the bacterial cold shock protein CspB. *Protein Sci* 8:1350–1357.
- Hamada D, Segawa S, Goto Y. 1996. Non-native alpha-helical intermediate in the refolding of beta-lactoglobulin, a predominantly beta-sheet protein. *Nat Struct Biol* 3:868–873.
- Harrison PM, Bamborough P, Daggett V, Prusiner SB, Cohen FE. 1997. The prion folding problem. *Curr Opin Struct Biol* 7:53–59.
- Jackson GS, Hosszu LLP, Power A, Hill AF, Kenney J, Saibil H, Craven CJ, Waltho JP, Clarke AR, Collinge J. 1999. Reversible conversion of monomeric human prion protein between native and fibrillogenic conformations. *Science* 283:1935–1937.
- Jones DT, Taylor WR, Thornton JM. 1992. A new approach to protein fold recognition. *Nature* 358:86–89.
- Kabsch W, Sander C. 1984. On the use of sequence homologies to predict protein structure: Identical pentapeptides can have completely different conformations. *Proc Natl Acad Sci* 81:1075–1078.
- Kim CA, Berg JM. 1993. Thermodynamic β -sheet propensities measured using a zinc-finger host peptide. *Nature* 362:267–270.
- Kraulis P. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 24:946–950.
- Lai Z, Colon W, Kelly JW. 1996. The acid-mediated denaturation pathway of transthyretin yields a conformational intermediate that can self-assemble into amyloid. *Biochemistry* 35:6470–6482.
- Levitt M, Gerstein M, Huang E, Subbiah S, Tsai J. 1997. Protein folding: The endgame. *Annu Rev Biochem* 66:549–579.
- Merkel JS, Sturtevant JM, Regan L. 1999. Sidechain interactions in parallel β -sheets: The energetics of cross-strand pairings. *Struct Fold Des* 7:1333–1343.
- Mezei M. 1998. Chameleon sequences in the PDB. *Protein Eng* 11:411–414.
- Minor DL Jr, Kim PS. 1994. Measurement of the β -sheet-forming propensities of amino acids. *Nature* 367:660–663.
- Minor DL Jr, Kim PS. 1996. Context-dependent secondary structure formation of a designed protein sequence. *Nature* 380:730–734.
- Mottonen J, Strand A, Symersky J, Sweet RM, Danley DE, Geohegan KF, Gerard RD, Goldsmith EJ. 1992. Structural basis of latency in plasminogen-activator inhibitor-1. *Nature* 355:270–273.
- Munson M, Balasubramanian S, Fleming KG, Nagi AD, O'Brien R, Sturtevant JM, Regan L. 1996. What makes a protein a Protein? Hydrophobic core designs that specify stability and structural properties. *Protein Sci* 5:1584–1593.
- Munson M, O'Brien R, Sturtevant JM, Regan L. 1994a. Redesigning the hydrophobic core of a four-helix-bundle protein. *Protein Sci* 3:2015–2022.
- Munson M, Predki PF, Regan L. 1994b. ColE1-compatible vectors for high-level expression of cloned DNAs From the T7 promoter. *Gene* 144:59–62.
- Mutter M, Gassmann R, Buttke U, Altmann KH. 1991. Switch peptides—pH induced α helix to β sheet transitions of bis-amphiphilic oligopeptides. *Angew Chem, Int Ed Edit* 30:1514–1516.
- Rose GD, Creamer TP. 1994. Protein folding: Predicting predicting. *Proteins* 19:502–513.
- Smith CK, Munson M, Regan L. 1995. Studying α -helix and β -sheet formation in small proteins. In: Crabb J, ed. *Techniques in protein chemistry, Vol VI*. San Diego, California: Academic Press. pp 323–332.
- Smith CK, Regan L. 1995. Guidelines for protein design: The energetics of β sheet side chain interactions. *Science* 270:980–982.
- Smith CK, Withka JM, Regan L. 1994. A thermodynamic scale for the beta-sheet forming tendencies of the amino acids. *Biochemistry* 33:5510–5517.
- Wei A, Rubin H, Cooperman BS, Christianson DW. 1994. Crystal structure of an unclaved serpin reveals the conformation of an inhibitory reactive loop. *Nat Struct Biol* 1:251–258.