Structure-based substitutions for increased solubility of a designed protein

Leila K. Mosavi and Zheng-yu Peng

Department of Biochemistry, University of Connecticut Health Center, Farmington, CT 06032, USA

1To whom correspondence should be addressed.
E-mail: peng@sun.uchc.edu

Manipulation of protein solubility is important for many aspects of protein design and engineering. Previously, we designed a series of consensus ankyrin repeat proteins containing one, two, three and four identical repeats (1ANK, 2ANK, 3ANK and 4ANK). These proteins, particularly 4ANK, are intended for use as a universal scaffold on which specific binding sites can be constructed. Despite being well folded and extremely stable, 4ANK is soluble only under acidic conditions. Designing interactions with naturally occurring proteins requires the designed protein to be soluble at physiological pH. Substitution of six leucines with arginine on exposed hydrophobic patches on the surface of 4ANK resulted in increased solubility over a large pH range. Study of the pH dependence of stability demonstrated that 4ANK is one of the most stable ankyrin repeat proteins known. In addition, analogous leucine to arginine substitutions on the surface of 2ANK allowed the partially folded protein to assume a fully folded conformation. Our studies indicate that replacement of surface-exposed hydrophobic residues with positively charged residues can significantly improve protein solubility at physiological pH.

Keywords: ankyrin repeat/arginine substitution/pH dependence of stability/protein design/protein solubility

Introduction

The recent completion of whole genome sequences has resulted in a flood of new proteins and biochemical systems to be characterized. The concentration of a single protein inside the cell is usually fairly low. The transition from relatively dilute in vivo conditions to more concentrated conditions used for in vitro structure–function studies often results in aggregation, poor solubility and misfolding of proteins. Most biochemical studies, particularly quantitative binding assays, and also high-resolution structural studies, demand that the target protein to be highly soluble. Often, achieving sufficient protein solubility requires detailed and time-consuming screening of solution conditions such as buffer, pH and salt. To date, many of the biophysically well characterized proteins in the literature have been extensively studied, in part, due to their high solubility and ease of purification. There are many proteins, however, that are not amenable to biophysical studies because of low solubility and a tendency to misfold or aggregate. In addition, many industrial applications of enzymes, such as detergents, animal feed and textile products, can benefit from improved protein solubility [for a review, see Kirk et al. (Kirk et al., 2002)]. In the field of protein design, a detailed understanding of the generalized solution properties of proteins will help to produce well folded stable molecules.

Various strategies exist to alter the solubility of a protein. The most straightforward is to screen for conditions such as different buffers, salts or additives to increase solubility. This strategy has the advantage of using the native protein sequence; however, it is time consuming and the proper conditions for high solubility may never be found. Minor modifications to the protein include the addition of a fusion protein such as maltose-binding protein, protein G (B1 domain) or thioredoxin upstream or downstream of the target protein in order to improve solubility and expression (LaVallie and McCoy, 1995; Kapust and Waugh, 1999; Zhou et al., 2001). In addition, recent advances in peptide chemistry have allowed the synthesis of proteins glycosylated at specific amino acids, which results in altered solution behavior (Kochendoerfer et al., 2003). Alternatively, one can modify the amino acid sequence of the protein in order to achieve the desired properties. Directed molecular evolution combined with selection techniques using fusion proteins such as green fluorescent protein or chloramphenicol acetyltransferase as reporters for proper folding and solubility of the target protein have met with some success (Maxwell et al., 1999; Waldo et al., 1999). The drawback of such experiments is that there is no way to control the function of the target protein. Many groups have relied on rational design of substitutions to engineer increased protein solubility. Knowledge of the high-resolution structure and the functionally important residues is used to guide the design process. This requires an understanding of the role of various amino acids in protein folding, stability and solubility. Despite efforts aimed at correlating amino acid sequences with solubility (Wilkinson and Harrison, 1991), a comprehensive survey relating these properties has not been established.

Previously, we have designed a series of consensus ankyrin repeat proteins containing one, two, three and four identical repeats (Mosavi et al., 2002a). The ankyrin repeat is a 33-residue sequence motif that forms a helix–loop–helix followed by a β-hairpin positioned at a 90° angle (Sedgwick and Smerdon, 1999). The ankyrin repeat, like the tetra-tricopeptide repeat (TPR), ARM/HEAT repeat and leucine-rich repeat (LRR), is typically tandemly repeated in a protein and forms an elongated surface capable of binding other proteins. Having no specific binding partner or enzymatic function, these repeats are found in a variety of proteins including transcription factors, cell cycle regulators, cytoskeletal proteins and signal transduction molecules. Typically, a protein contains four to six ankyrin repeats, although as many as 29 repeats have been observed in a single protein (Walker et al., 2000).

We designed our consensus ankyrin repeat with the ultimate goal of using it as a generalized scaffold to engineer protein–protein interactions. Our four-repeat construct (4ANK) is folded, stable, soluble and monomeric only under acidic
conditions. In order to use this consensus ankyrin repeat sequence as a building block to create specific binding surfaces, it is important that the protein be soluble and monomeric at physiological pH. Inspection of the surface of 4ANK revealed two large hydrophobic patches. In order to increase the number of charged residues on the protein surface, we replaced six exposed leucines to arginines on the N- and C-terminal repeats of 4ANK. The resulting protein, 4ANK TALR, is soluble over a large pH range. The arginine-substituted proteins are designated ‘TALR’, which stands for ‘terminal all L to R’. Furthermore, the leucine to arginine substitutions in the same relative positions on our consensus repeat protein containing two repeats (2ANK) caused the protein to adopt a monomeric and fully folded conformation. This work demonstrates that substitution of hydrophobic residues with arginine on the protein surface is an effective way to increase protein solubility under various conditions.

Materials and methods

Gene construction

The genes for 4ANK and 2ANK were constructed as described previously (Mosavi et al., 2002a). Genes for 2ANK TALR and 4ANK TALR were constructed by ligating together separate DNA cassettes, each containing a 99-bp single ankyrin repeat gene with the desired sequence/substitutions. These cassettes were created using a recursive polymerase chain reaction (PCR) (Dillon and Rosen, 1990; Prodromou and Pearl, 1992; Chen et al., 1994) with four complementary oligonucleotides. Each cassette contains an Mlu restriction site at the 5’-end and a BssHII restriction site at the 3’-end. When cleaved, these two restriction sites have complementary overhang sequences and can be ligated to one another eliminating both cutting sites at the junction. This results in a gene containing two ankyrin repeats with a 5’-Mlu site and a 3’-BssHII site but no restriction site in between the two cassettes. This construct can then be ligated to another cassette if desired, using the restriction site at either end. The 4ANK TALR gene consists of amino acid positions 1–33 in the first three repeats and positions 1–26 in the fourth repeat, and the 2ANK TALR gene consists of positions 1–33 of the first repeat followed by positions 1–26 of the second repeat. Both genes contain an additional C-terminal tyrosine for concentration determination. The ligation product corresponding to the expected size for the desired number of repeats was cloned into the pAED4 expression vector (Staley and Kim, 1994; Mosavi et al., 2002a). The genes encoding 2ANK TALR and 4ANK TALR were identified by restriction digestion and confirmed by automated DNA sequencing analysis.

Protein expression and purification

The proteins were expressed in Escherichia coli BL21(DE3) pLysS and purified as described previously (Mosavi et al., 2002a). The predicted molecular masses for 4ANK TALR and 2ANK TALR were 13 681 and 6674 Da and the measured molecular weights, as determined by electrospray mass spectrometry, were 13 683 and 6674 Da, respectively. Protein concentration was determined by measuring the absorbance at 280 nm in 6 M guanidine hydrochloride, assuming a molar extinction coefficient of 1280 for all constructs (Edelhoch, 1967).

Circular dichroism (CD)

All CD experiments were carried out using a JASCO J-715 spectropolarimeter with a thermoelectric temperature controller. The CD spectra were recorded using 10 μM protein in a 0.1 cm pathlength cuvette with scan rate 10 nm/min, bandwidth 2 nm and response time 8 s at 20°C. Samples of 4ANK TALR contained 10 mM succinate for pH 4 and 5 or 10 mM phosphate for pH 6 and 7, 1 mM EDTA and 500 mM NaCl. Samples of 4ANK TALR contained 10 mM succinate, pH 4.25, or phosphate, pH 7.0, 50 mM NaCl and 1 mM EDTA. Samples of 2ANK TALR contained 10 mM phosphate, pH 7, 1 mM EDTA and 500 mM Na2SO4. The results shown are the averages of three scans.

Urea and guanidine denaturation studies were performed in the buffers described above with various concentrations of urea or guanidine hydrochloride. The concentrations of the urea and guanidine hydrochloride stock solution were determined by using a refractometer (Shirley, 1995). Each data point was measured in a 0.1 cm pathlength cuvette, with resolution 2 s, response time 1 s and averaged for 2 min. The chemical denaturation data were fitted with a six-parameter equation assuming a two-state folding transition (Pace, 1986):

\[
\theta(x) = \theta_u(x) + \frac{\theta_f(x) - \theta_u(x)}{1 + \exp\left[\left(-\frac{\Delta G - mx}{RT}\right)\right]}
\]

where θ represents the CD signal, x is the urea or guanidine concentration, ΔG is the free energy of unfolding and m is the cooperativity factor. θf(x) = θf1 + θf2x and θu(x) = θu1 + θu2x are the folded and unfolded baselines. For 4ANK TALR at pH 7.0, it was impossible to obtain an unfolded baseline. Therefore, an unfolded baseline was estimated using the pH 6.0 data in order to reduce the error of fitting.

Thermal denaturation studies on 2ANK TALR were performed in the same buffer except that 1 M NaCl was used instead of 500 mM Na2SO4 in order to obtain a fully reversible melt. The protein concentration was 2 μM in a 1 cm pathlength cuvette. The CD signal was monitored at 222 nm, with bandwidth 2 nm, response time 8 s and a temperature increase of 25°C/h. The forward and reverse curves were superimposable within 1°C when the sample was heated to 56°C and then allowed to cool to 4°C at the same rate (25°C/h). The thermal denaturation data were fitted with the following equation (Becktel and Schellman, 1987; Privalov, 1989):

\[
\theta(T) = \theta_u(T) + \frac{\theta_f(T) - \theta_u(T)}{1 + \exp\left(-(1/T - 1/T_m)\Delta H - (1 - T_m/T)\Delta C_p - \ln(T_m/T)\Delta C_p/R\right)}
\]

where θ represents the CD signal, Tm is the melting temperature, ΔH is the change in enthalpy and ΔCp is the change in specific heat upon unfolding. θf(T) = θf1 + θf2 T and θu(T) = θu1 + θu2 T are the folded and unfolded baselines, which were assumed to be linearly dependent on temperature.

Size-exclusion chromatography–multiple angle laser light scattering (SEC–MALLS)

SEC–MALLS studies were performed using a Varian HPLC system connected to a miniDAWN light scattering detector and an Optilab refractive index detector (both from Wyatt Technology). Buffers, described in the CD experiments, were
filtered and degassed prior to use. The system was equilibrated overnight. Samples (200 μl) were injected at 100 μM concentration and eluted at a flow rate of 0.5 ml/min for 4ANK TALR on a Pharmacia Superdex 75 size-exclusion column and 0.3 ml/min for 2ANK TALR on a Pharmacia Superdex peptide column. Scattering data were collected and analyzed using ASTRA software (Wyatt Technology). The apparent molecular weights and polydispersity factors were calculated by using the Zimm method (Folta-Stogniew and Williams, 1999) at the peak width corresponding to the peak half-height.

Results

Design of arginine substitutions

In order to choose the surface positions on 4ANK that were good candidates to improve solubility, we calculated the solvent accessibility of each residue in the 4ANK structure (PDB ID: 1N0R) using the program MOLMOL (Koradi et al., 1996). Residues L8, A9, V17, L20, L21 and A24 on the N-terminal repeat and L105, A109, L114, V117 and L121 on the C-terminal repeat are hydrophobic amino acids with solvent accessibility ranging from 43 to 10%. Mapping these residues on to the X-ray crystal structure of 4ANK showed that they formed two mostly contiguous patches on the surface of the protein (Figure 1). We chose to substitute the three leucine residues in the N-terminal repeat and the three leucine residues in the C-terminal repeat with arginine, reasoning that the arginine side-chain methylene groups could still form hydrophobic contacts with existing valine and alanine, while the positively charged guanido group would increase the solubility and pI of the protein. 4ANK TALR has substitutions L8R, L20R and L21R in the N-terminal repeat and substitutions L105R, L114R and L121R in the C-terminal repeat; 2ANK TALR has L8R, L20R and L21R in the N-terminal repeat and L39R, L48R and L55R in the C-terminal repeat. These substitutions correspond to positions L8, L20, L21 and L6, L15, L22, respectively, in our designed consensus ankyrin repeat sequence, which follows the canonical numbering scheme (Mosavi et al., 2002a).

Secondary structure and oligomerization state of 4ANK TALR

Far-UV CD scans of 4ANK TALR at pH 4–7 were compared with CD scans of 4ANK at pH 4.25 and pH 7.0 (Figure 2a). With the exception of 4ANK at pH 7.0, the spectra show well defined minima at 222 and 208 nm, characteristic of a mostly β-helical, well folded protein. Both the shape and the amplitude of the spectra of 4ANK TALR are essentially invariant at different pH values and agree well with 4ANK at
pH 4.25. In contrast, at pH 5 or higher, the CD spectra of 4ANK are distorted (Figure 2a), which may indicate the presence of soluble aggregates. The α-helical content of 4ANK TALR agrees well with that of 4ANK at pH 4.25, indicating that the arginine substitutions have not disrupted secondary structure elements in the protein.

The oligomerization state of 4ANK TALR was analyzed using SEC–MALLS studies (Table I; Figure 3). At 100 μM protein concentration and salt concentrations >500 mM, 4ANK TALR elutes as a single peak at pH 4–7 with an observed molecular weight within 5% of that expected for a monomer (13 681 Da). At low salt concentrations, 4ANK TALR tends to self-associate. This is different from the solution conditions necessary for 4ANK, which is folded and monomeric only from pH 3 to 4.25. At pH 5 and 6, 4ANK has very low solubility, and at pH 7, it forms soluble aggregates that bind to the filter used to remove particulate matter from the HPLC sample. Therefore, HPLC spectra of 4ANK at pH 7 showed little or no signal corresponding to the injected protein. Salt was only tolerated by 4ANK at more acidic pH. For example, 4ANK (50 μM) is monomeric at pH 4.0 in 500 mM NaCl but aggregates at salt concentrations >50 mM NaCl at pH 4.25 at 100 μM.

Stability of 4ANK TALR

Previously, 4ANK displayed an increase in the free energy of unfolding at higher pH, but this was difficult to analyze given the small pH range at which the protein was folded and soluble. Since 4ANK TALR is soluble at pH 4–7, we were able to characterize the pH dependence of the stability of this protein.

### Table I. SEC–MALLS data for TALR mutants under various conditions

<table>
<thead>
<tr>
<th>Sample name and conditions</th>
<th>Apparent MW (Da)</th>
<th>Polydispersity factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>4ANK TALR, pH 4</td>
<td>13850 (1.2%)</td>
<td>1.00</td>
</tr>
<tr>
<td>4ANK TALR, pH 5</td>
<td>13640 (0.3%)</td>
<td>1.00</td>
</tr>
<tr>
<td>4ANK TALR, pH 6</td>
<td>13450 (1.7%)</td>
<td>1.00</td>
</tr>
<tr>
<td>4ANK TALR, pH 7</td>
<td>14270 (4.3%)</td>
<td>1.00</td>
</tr>
<tr>
<td>2ANK TALR, pH 7, 0.5 M Na₂SO₄</td>
<td>6565 (1.6%)</td>
<td>1.00</td>
</tr>
<tr>
<td>2ANK TALR, pH 7, 1.0 M NaCl</td>
<td>6687 (0.2%)</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*The values in parentheses indicate the difference between the apparent and expected molecular weights for 4ANK TALR (13 681 Da) and 2ANK TALR (6674 Da).*

### Table II. Calculated ΔG(H₂O) and m values for 4ANK TALR guanidine hydrochloride denaturation data at various pHs and 2ANK TALR urea denaturation

<table>
<thead>
<tr>
<th>Sample name and conditions</th>
<th>ΔG(H₂O) (kcal/mol)</th>
<th>m (kcal/mol.M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4ANK TALR, pH 4</td>
<td>4.47 ± 0.31</td>
<td>2.23 ± 0.15</td>
</tr>
<tr>
<td>4ANK TALR, pH 5</td>
<td>7.60 ± 0.63</td>
<td>1.94 ± 0.16</td>
</tr>
<tr>
<td>4ANK TALR, pH 6</td>
<td>8.57 ± 0.80</td>
<td>1.69 ± 0.16</td>
</tr>
<tr>
<td>4ANK TALR, pH 7</td>
<td>12.26 ± 0.82</td>
<td>1.77 ± 0.12</td>
</tr>
<tr>
<td>2ANK TALR, pH 7b</td>
<td>0.81 ± 0.51</td>
<td>1.25 ± 0.17</td>
</tr>
</tbody>
</table>

*The ΔG(H₂O) and m values shown are the average of two independent data sets. The errors for individual data sets were obtained from a non-linear least-squares fitting program and the errors of the average were calculated using \( e_{avg} = \sqrt{e_1^2 + e_2^2} \).*

*The precision of this fit is low because of limited stability and incomplete baselines.*

---

Fig. 3. Oligomerization state of 4ANK TALR shown by SEC–MALLS. Protein was injected at 100 μM concentration in 500 mM NaCl at pH 4 (a), pH 5 (b), pH 6 (c) and pH 7 (d). The light scattering signal (solid circles) shown is from the 138° angle detector and molar mass (open squares) was calculated using ASTRA software.
Figure 2b shows the guanidine hydrochloride denaturation curves for 4ANK TALR at pH 4, 5, 6 and 7. At pH 7, even at the highest guanidine hydrochloride concentration, 4ANK TALR is not completely unfolded. The ΔG of unfolding at pH 4, 5, 6 and 7 are 4.5, 7.6, 8.6 and 12.3 kcal/mol, respectively (Table II). At neutral pH, 4ANK TALR is one of the most stable ankyrin repeat proteins characterized to date. Ankyrin repeat proteins usually show an increase in ΔG with growing number of repeats. Naturally occurring proteins such as Drosophila Notch with seven repeats have a ΔG of 8.0 kcal/mol (Zweifel and Barrick, 2001), whereas rat myotrophin, a 3.5-repeat cardiomyogenic hormone, has a ΔG of 5.1 kcal/mol (Mosavi et al., 2002b). The designed proteins created by Kohl et al., with non-identical consensus ankyrin repeats, have ΔG values ranging from 9.5 to 11.4 kcal/mol for four repeats, 9.6 to 14.8 kcal/mol for five repeats and 21.1 kcal/mol for six repeats (Kohl et al., 2003).

2ANK TALR

Previously, 2ANK appeared to be partially folded under acidic conditions (pH 3–5) and form soluble aggregates at higher pH (data not shown). With the TALR substitution, 2ANK TALR forms a fully folded monomeric protein. The far-UV CD spectra of 2ANK TALR show a folded α-helical protein with the percentage of α-helical structure in agreement with 4ANK TALR (Figure 4a). In addition, we used SEC–MALLS to verify that 2ANK TALR is a monomer in solution at 100 μM concentration (Table I). In the absence of salt, 2ANK TALR is partially structured (data not shown) and increasing concentrations of salt greatly enhance 2ANK TALR’s stability.

Thermal and chemical denaturation studies suggest that 2ANK TALR has a unique tertiary structure. Thermal denaturation of 2ANK TALR monitored by far-UV CD at 222 nm in 1 M NaCl shows a reversible, cooperative transition with a calculated midpoint of transition (Tm) of 34.7°C (Figure 4b). We monitored the unfolding of 2ANK TALR by urea denaturation in 0.5 M Na2SO4 and obtained a ΔG value of 0.8 kcal/mol (Table II). This value probably represents the lower limit of unfolding free energy because the lack of folded baseline does not allow us to fit the baseline independently. Although the Tm and ΔG values for 2ANK TALR are relatively small, this protein is folded with a packed hydrophobic core as indicated by the cooperative thermal and chemical denaturation curves. Previously, a truncated version of the tumor suppressor p16INK4A, named p16C, was the only ankyrin repeat protein containing two repeats to be characterized biophysically with a ΔG value of 1.7 kcal/mol (Zhang and Peng, 2000).

Discussion

Engineering soluble monomeric proteins is important for the field of protein biochemistry and also for many industrial applications of proteins. The use of our consensus ankyrin repeat protein, 4ANK, as a general scaffold to design protein–protein interactions, requires that it be soluble and monomeric under physiological conditions. The motivation behind our TALR substitutions was to improve the solubility of 4ANK from acidic pH to more neutral conditions for such studies. We chose to replace six surface leucine residues with arginine to increase the net positive charge of the protein. Previously, several studies have demonstrated the role of arginine in protein solubility and stability at high pH. Comparisons of subtilisin-family proteases from alkaliphilic bacteria with their less alkalophilic or mesophilic homologs show an increase in the number of arginine residues (Masui et al., 1994; Shirai et al., 1997). In addition, random mutagenesis and rational design efforts incorporating arginine residues to engineer greater alkaline stability and enzymatic activity have been successful (Cunningham and Wells, 1987; Turunen et al., 2002). In these studies, alkaline adaptation has been attributed to an increase in the isoelectric point of the proteins, either through a decrease in acidic residues such as glutamate or aspartate, replacement of lysines with arginines or an increase in the number of arginines. Arginine is also known to be effective in preventing dilution-induced and thermal unfolding–induced aggregation for a number of proteins (Shiraki et al., 2002). The results reported here agree with past experimental data. The TALR substitutions increase the theoretical pI of 4ANK from 9.3 to 10.5. 4ANK is soluble, folded and monomeric from pH 3 to 4.25 at 100 μM. However, 4ANK TALR is soluble, folded and monomeric from pH 4 to 7 at
concentrations >100 μM. The leucine to arginine substitutions replace the hydrophobic patches on the N- and C-termini of 4ANK with charged residues, resulting in greater interaction with the solvent and reduction in self-association.

**pH dependence of stability**

4ANK TALR, with its increased solubility, allowed us to examine the pH dependence of stability. Previously, 4ANK showed an increase in stability with increase in pH, although this was difficult to characterize since the protein was soluble and monomeric over only a small pH range. 4ANK TALR shows an approximately linear pH dependence of stability ranging from 4.5 kcal/mol at pH 4 to 12.3 kcal/mol at pH 7 (Table II). The classical Linderstrom-Lang model postulates that a protein surface is an average of the total side-chain charges, which are all repulsive (Linderstrom-Lang, 1924). Therefore, the protein is most stable at its isoelectric point and the stability will decrease as the pH increases or decreases from the pI. Although it appears that the stability of 4ANK TALR increases as the theoretical pI is approached, we were unable to determine the pH value at which the free energy of unfolding is maximized. It has been shown that this classical model does not apply universally to all proteins. Several studies have shown the maximum stability to be significantly different from the protein’s isoelectric point. For example, analysis of variants of RNase SA in which the net charge of the protein was altered using lysine mutations shows that, while the minimum of protein solubility coincides with the pI, the pH of maximum stability does not (Shaw et al., 2001). Similar results were found in studies of T4 lysozyme and barnase (Becktel and Bäase, 1987; Pace et al., 1992). For both proteins, the deviation in the pH of maximum stability was attributed to a single His residue in each protein with unusually high pKₐ values (Anderson et al., 1990; Pace et al., 1992). Indeed, the pH dependence of stability can be solely attributed to the presence of ionizable groups in the protein that have different pKₐ values in the folded state and unfolded state (Pace et al., 1990, 1992). The pKₐ values at low pH result from the proximity of the three arginines to the surface. It is likely that the proximity of the three arginines to one another on the surface contributes to the stabilization of the protein. At pH 4.0, 4ANK TALR is marginally stable and therefore the destabilizing effects of the charge repulsion may be particularly prominent. At higher pH values, however, it is likely these repulsive effects contribute less to overall stability, since the ΔG of unfolding is much higher.

**Salt dependence**

The ion strength of the solution affects 4ANK and 4ANK TALR differently. 4ANK, at pH 4.25, shows aggregates at NaCl concentrations >50 mM. In fact, the tolerance to salt in solution decreases as the pH of the solution increases. In the presence of >500 mM NaCl, 4ANK TALR, is a monomer at pH 4–7. However, at 50 mM NaCl, 4ANK TALR forms higher order soluble aggregates. This difference in the solution salt requirement may correspond to a shift between salting-in and salting-out effects. The aggregation of 4ANK resulting from increasing salt concentration can be attributed to the Hofmeister effect [for reviews, see Baldwin (Baldwin, 1996), Cacace et al. (Cacace et al., 1997) and Record et al. (Record et al., 1998)]. Preferential hydration of the protein, in response to ions in the solute, results in aggregation or salting-out by increasing the hydrophobic effect. If one considers the exposed hydrophobic patches on 4ANK, this explanation seems reasonable for the observed aggregation. 4ANK TALR, having six more positive charges, is a protein that may have preferential interaction with the ions in solution rather than preferential hydration (Wright et al., 2002). This can be considered as a ‘salting-in’ effect in which the solubility of the protein increases as the ionic strength of the solution increases. Although the Hofmeister effect has been discussed in detail, little is known about the balance between salting-in and salting-out effects (Arakawa and Timasheff, 1985). The mechanism of the observed self-association of 4ANK TALR at low ionic strength is not completely clear. A previous survey of protein–protein interface ‘hot spots’ found that the interfaces are enriched in arginine (Bogan and Thorn, 1998), suggesting that their presence on the surface of a protein may provoke intermolecular interactions.

Salt may also help to stabilize the overall structure of 4ANK TALR because of the high density of positive charges on the surface. It is likely that the proximity of the three arginines to one another on each terminus of the protein results in destabilizing repulsive interactions. The presence of ions in solution may partially screen out the unfavorable electrostatic effects, thereby stabilizing the protein. At pH 4.0, 4ANK TALR is marginally stable and therefore the destabilizing effects of the charge repulsion may be particularly prominent. At higher pH values, however, it is likely these repulsive effects contribute less to overall stability, since the ΔG of unfolding is much higher.

2ANK TALR

Substitution of the six leucine residues to arginine on the surface of 2ANK TALR results in a well folded, monomeric protein displaying cooperative thermal denaturation. Previously, 2ANK was partially folded at pH 5.0 and formed soluble aggregates with folded CD spectra at higher pH. 2ANK TALR, under low ionic strength conditions, was partially folded at all pH values tested. However, with the addition of 500 mM NaCl, 2ANK TALR was fully folded and monomeric at pH 7. A possible explanation for this change in the solution properties lies in the surface to volume ratio of hydrophobic residues. Owing to the identical nature of the repeats, 2ANK has approximately the same number of exposed and buried hydrophobic residues. This is in contrast to 4ANK, which has the same exposed hydrophobic patches on the surface of the protein, but contains three times more buried hydrophobic residues at the inter-repeat interfaces compared with 2ANK. Our previous studies involving the truncation of p16INK4A suggest that two repeats are the minimum folding unit of the ankyrin repeat motif (Zhang and Peng, 2000). Taken together, these results suggest that two repeats can fold only if a sufficient number of buried hydrophobic residues exist to form an energetically favorable core.

**Effect of TALR mutations on stability**

The TALR mutations, although located on the surface of the protein, have a profound effect on global stability. 4ANK is soluble and monomeric at pH 4 with 500 mM NaCl, allowing us to compare the guanidine hydrochloride denaturation curves
Increased solubility of a designed protein

of 4ANK and 4ANK TALR under these conditions. 4ANK has a $\Delta G$ of 10.1 $\pm$ 0.4 kcal/mol and $m$ value of 2.7 $\pm$ 0.1 kcal/mol M (data not shown). The stability of 4ANK is more than double that of 4ANK TALR (4.5 kcal/mol). This difference may be attributed to the destabilizing effect of unpaired ions on the surface of the protein that can result in desolvation (Yang and Honig, 1993). It is generally accepted that protein stability increases as the solution pH approaches the $pI$. Additional positively charged residues on the protein surface will increase the theoretical $pI$. Therefore, at the same pH, the difference between pH and $pI$ for 4ANK is less than that of 4ANK TALR and consequently 4ANK should be a more stable molecule. Given these results, it is likely that if 4ANK were soluble and monomeric at pH 7, the stability would be higher than that of 4ANK TALR.

In summary, we have designed substitutions on the surface of 4ANK in order to increase the solubility at physiological pH. Substitutions of arginine for leucine on the exposed hydrophobic surfaces resulted in 4ANK TALR, which is soluble, folded and monomeric from pH 4 to 7. These substitutions altered the salt requirements of the protein in solution. We have characterized the pH dependence of stability for 4ANK TALR, which, although less stable than 4ANK under the same conditions, is one of the most stable ankyrin repeat proteins characterized the pH dependence of stability for 4ANK TALR, altered the salt requirements of the protein in solution. We have double that of 4ANK TALR (4.5 kcal/mol). This difference increases as the solution pH approaches the $pI$. It is generally accepted that protein stability increases as the solution pH approaches the $pI$. Additional positively charged residues on the protein surface will increase the theoretical $pI$. Therefore, at the same pH, the difference between pH and $pI$ for 4ANK is less than that of 4ANK TALR and consequently 4ANK should be a more stable molecule. Given these results, it is likely that if 4ANK were soluble and monomeric at pH 7, the stability would be higher than that of 4ANK TALR.

Acknowledgements

We thank Ping Bai for help with cloning, Jeff Hoch for helpful discussions and Tobin Camnett, Daniel Desrosiers and Zandra Sutter for critical reading of the manuscript. This work was supported by grant GMC-103045 from the American Cancer Society.

References