

Conservation of mechanism, variation of rate: folding kinetics of three homologous four-helix bundle proteins

Seema Dalal^{1,2,5}, Denis Canet³, Stephen E. Kaiser⁶,
Christopher M. Dobson⁴ and Lynne Regan^{1,2,7}

Departments of ¹Chemistry, ²Molecular Biophysics and Biochemistry, Yale University, Bass 322, 266 Whitney Avenue, New Haven, CT 06520-8114, USA, ³Oxford Centre for Molecular Sciences, New Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QT, UK and ⁴Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

⁵Present address: Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

⁶Present address: Stanford University, School of Medicine, Stanford CA94306

⁷To whom correspondence should be addressed.
E-mail: lynne.regan@yale.edu

The amino acid sequence of a protein determines both its final folded structure and the folding mechanism by which this structure is attained. The differences in folding behaviour between homologous proteins provide direct insights into the factors that influence both thermodynamic and kinetic properties. Here, we present a comprehensive thermodynamic and kinetic analysis of three homologous homodimeric four-helix bundle proteins. Previous studies with one member of this family, Rop, revealed that both its folding and unfolding behaviour were interesting and unusual: Rop folds ($k_f^0 = 29 \text{ s}^{-1}$) and unfolds ($k_u^0 = 6 \times 10^{-7} \text{ s}^{-1}$) extremely slowly for a protein of its size that contains neither prolines nor disulphides in its folded structure. The homologues we discuss have significantly different stabilities and rates of folding and unfolding. However, the rate of protein folding directly correlates with stability for these homologous proteins: proteins with higher stability fold faster. Moreover, in spite of possessing differing thermodynamic and kinetic properties, the proteins all share a similar folding and unfolding mechanism. We discuss the properties of these naturally occurring Rop homologues in relation to previously characterized designed variants of Rop.

Keywords: four-helix bundle/homologues/protein folding/protein stability/RNA-binding

Introduction

Proteins fold with differing mechanisms and a wide range of time constants that vary from microseconds to hours. Theoretical studies have implicated factors such as protein size (Wolynes, 1997), topology (Gross, 1996), proline content and thermodynamic stability (Sali, 1994 #901; Wright *et al.*, 2004) as factors which may influence the large range of kinetic behaviours observed. It has also been

suggested that for single domain proteins, there is a significant correlation between the rate of folding and the topological complexity (contact order) of the protein (Plaxco *et al.*, 1998; Ivankov *et al.*, 2003; Dobson, 2003).

Kinetic and thermodynamic studies have been performed on several families of homologous proteins to understand better the role a protein's sequence plays in acquiring the properties of a particular fold. These include studies on the all β -protein families of the fibronectin-like modules (Clarke *et al.*, 1997; Plaxco *et al.*, 1997), the intracellular lipid binding protein family (Burns *et al.*, 1998), SH3 domains (Martinez and Serrano, 1999; Riddle *et al.*, 1999), the all α -helical acyl-CoA binding proteins (Kragelund *et al.*, 1996) and the helical immunity proteins, Im7 and Im9 (Ferguson *et al.*, 1999), and the mixed α/β proteins acylphosphatase and ADA2 h (Chiti *et al.*, 1999). In some cases, all members of a family have been shown to fold with a similar mechanism (Kragelund *et al.*, 1996; Martinez and Serrano, 1999; Riddle *et al.*, 1999) whereas in others, different family members have been proposed to fold via significantly different mechanisms (Burns *et al.*, 1998; Ferguson *et al.*, 1999).

Here, we present a study of the thermodynamic and kinetic properties of the small (63 amino acids per monomer) homodimeric, four-helix bundle protein Rop and its homologues. Despite its regular structure and simple topology, Rop has interesting kinetic properties: it folds and unfolds much more slowly than would be expected for a protein of its size that contains neither prolines nor disulphide bonds (Munson *et al.*, 1996, 1997; Nagi and Regan, 1997; Nagi *et al.*, 1999). Rop takes seconds to minutes to fold and hours to days to unfold. In contrast, other small dimeric proteins such as P22 Arc repressor, *trp* aporepressor and small leucine zippers fold in milliseconds or less and unfold in seconds to minutes (Gittelman *et al.*, 1990; Milla and Sauer, 1994; Wendt *et al.*, 1995; Zitzewitz *et al.*, 1995; Sauer, 1996).

To explore further the basis of the unusual folding kinetics of Rop we identified its natural homologues in a BLAST search of the non-redundant protein database (Fig. 1) (Altschul *et al.*, 1990). Rop is encoded by the *Escherichia coli* plasmid ColE1. All of the homologue sequences also map to extrachromosomal locations in gram negative bacteria, belonging to the *Enterobacteriaceae* family (Woese, 1990). The Rop homologues are found on plasmid DNA in *Enterobacter cloacae* (EC), *Yersinia pestis* (YP), *Klebsiella oxytoca* (KO), *Enterobacter agglomerans* (EA) and *Proteus vulgaris* (PV). We refer to the homologues by the initials denoted in brackets. Of the proteins identified, EA and PV are the least homologous to Rop, with 51% and 48% identity to Rop, respectively. The experimental work presented here was performed on EA (Mikiewicz *et al.*, 1997) and PV (Calvin Koons and Blumenthal, 1995) and their RNA-binding, thermodynamic and kinetic properties are discussed in relation to each other and to Rop.

Abbreviations: GuHCl, Guanidine hydrochloride; DTT, dithiothreitol; UV, ultraviolet; CD spectroscopy, circular dichroism spectroscopy.

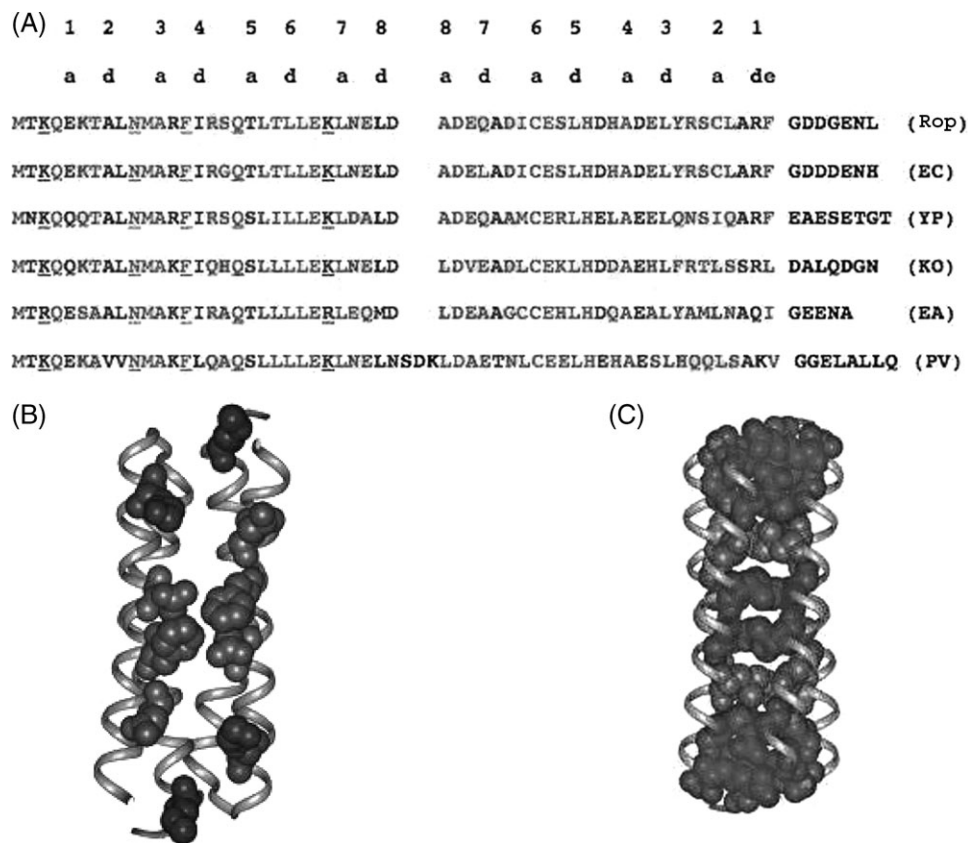


Fig. 1. Key elements in the sequence of the homologues. (A) Sequence alignments of Rop and its homologues. The top line shows the layer number of the core, the second line shows the position of the 'a' and 'd' position residues. The RNA-binding residues are underlined. The colour coding for the sequences are as follows: in green are the residues that are identical in all six homologues; in blue are the residues that are highly conserved across the homologues (either 5/6 have the same residue identity, or the differences are conservative); in red are the residues positions that are variable; and in black are the residues that align with the unstructured C-terminal tail of Rop. For PV, the residues that form the five-residue turn (NSDKL) are also shown in black. The sequences are shown in order of decreasing homology from the Rop sequence. (B) The RNA-binding residues (Lys3, Asn10, Phe14, Gln18, Lys25) on helix 1 and the complementary set on helix 1'. Shown in green are the RNA-binding residues that are identical across all the homologues and shown in blue are the residue positions that vary from Lys to Arg in EA. (C) The 'a' and 'd' heptad-repeat positions for helical structure for all eight layers of the hydrophobic core of Rop. The colour-coding of the layers is as follows: shown in green are the residues that form the core layers that are conserved across all six homologues; in blue are the residues that form the core layers that are less conserved across these homologues.

Results

RNA-binding activity

The function of Rop is to regulate the copy number of ColE1 plasmids in *E. coli*. Regulation is achieved by binding to a complex of two complementary RNA molecules, RNA I and RNA II (Polisky, 1988), and the binding of Rop facilitates the formation of the RNA I–RNA II complex. The two RNA molecules interact through their complementary loops to form a 'kiss complex' (Marino *et al.*, 1995; Lee *et al.*, 1998). It has been proposed that Rop recognizes the overall kiss complex structure, rather than specific elements in the RNA sequence (Comolli *et al.*, 1998). *In vitro*, the binding of Rop to the kiss complex can be detected in an electromobility shift assay with a truncated RNA kiss complex (Gregorian, 1995; Predki *et al.*, 1995) (Fig. 2). The kiss complex sequences used in the *in vitro* binding assay of Rop are shown in Fig. 2A along with the equivalent sequences identified in the regulatory regions of plasmid DNA from EA (Mikiewicz *et al.*, 1997) and PV (Calvin Koons and Blumenthal, 1995). A comparison of these RNA sequences suggests that EA and PV have RNAs that potentially adopt a kiss complex structure similar to that recognized by Rop.

Alanine scanning mutagenesis studies in Rop identified five residues as critical for RNA binding (Predki *et al.*, 1995). These residues are Lys3, Asn10, Phe14, Gln18, Lys25 in helix 1 and the complementary set of residues in helix 1', which form a narrow stripe down the positively charged 1/1' face of the protein (Fig. 1B). A comparison of the RNA-binding residue positions of Rop with the homologues demonstrates that these residues are identical across all of the homologues except for EA where conservative Lys to Arg substitutions occur at positions 3 and 25.

Because of the similar predicted secondary structures for the different RNAs, and the conservation of residues required for RNA binding, we predicted that EA and PV also function by binding regulatory RNA molecules in a similar fashion to Rop. To test the ability of EA and PV to bind RNA, an electromobility shift assay of EA, PV and Rop with the RNA sequences recognized by Rop, was performed. All three proteins bind the RNA complex resulting in a lower mobility complex compared with the free kiss complex (Fig. 2B). The similar RNA-binding properties of Rop and its homologues suggest that the relative orientation of the helices in all three proteins is similar, and that they likely share the same topology. However, the affinity of EA and PV for Rop's RNA is

is $PV < Rop < EA$ (Fig. 3A inset and Table I). The relative stability (ΔG) by GuHCl denaturation also increases in the order $PV < Rop < EA$ (Fig. 3B and Table I).

Folding mechanism for Rop

In addition to CD, fluorescence spectroscopy also provides a sensitive technique by which to monitor changes in structure. Each monomer of Rop has a single fluorophore, Tyr49, which is buried within the folded protein. The fluorescence intensity of Tyr49 changes depending on its chemical environment, and its fluorescence intensity can therefore be used as a probe of tertiary structure to monitor folding and unfolding. EA also has a Tyr at position 49 and therefore it is possible to use fluorescence to monitor its folding and unfolding as well. PV however has a histidine residue at this position and therefore it was necessary to monitor its folding and unfolding kinetics exclusively by stopped-flow CD.

Detailed kinetic refolding studies have led to the proposal that Rop folds through a dimeric intermediate (Munson *et al.*, 1997; Nagi *et al.*, 1999; Brockwell and Redford, 2007). At low GuHCl concentrations, the dimeric intermediate is stable and biphasic folding kinetics are observed. The fast rate is strongly dependent on protein concentration, whereas the slow rate has little dependence on protein concentration. This observation suggests that the fast phase involves a bimolecular association leading to the formation

of the dimeric intermediate. The slow phase is modeled as a structural rearrangement of the dimeric intermediate to yield the native protein.

At high concentrations of GuHCl, however, the dimeric intermediate is destabilized and therefore the observed kinetics are slow and monophasic.

The unfolding kinetics for Rop are monophasic at all concentrations of GuHCl, with no detectable unfolding intermediates.

Refolding of EA and PV

We investigated the refolding of EA by both stopped-flow CD and stopped-flow fluorescence spectroscopy. EA follows biphasic refolding kinetics up to 3 M GuHCl (Fig. 4A and B), becoming monophasic at concentrations of 3.5 M GuHCl and higher (Fig. 4C and D). When refolding the protein at low GuHCl concentrations (0–1.5 M), the fast phase is too rapid to be detected, whereas refolding at concentrations greater than 1.5 M GuHCl, results in a detectable fast phase. The refolding data obtained from CD and fluorescence are coincident with each other (within error), indicating that the acquisition of secondary and tertiary structure is coincident. The refolding rate constant at 0 M GuHCl for EA is estimated to be ~ 15 times faster than Rop (Table II).

PV refolding monitored by stopped-flow CD follows biphasic refolding kinetics until 1.0 M GuHCl, and when it

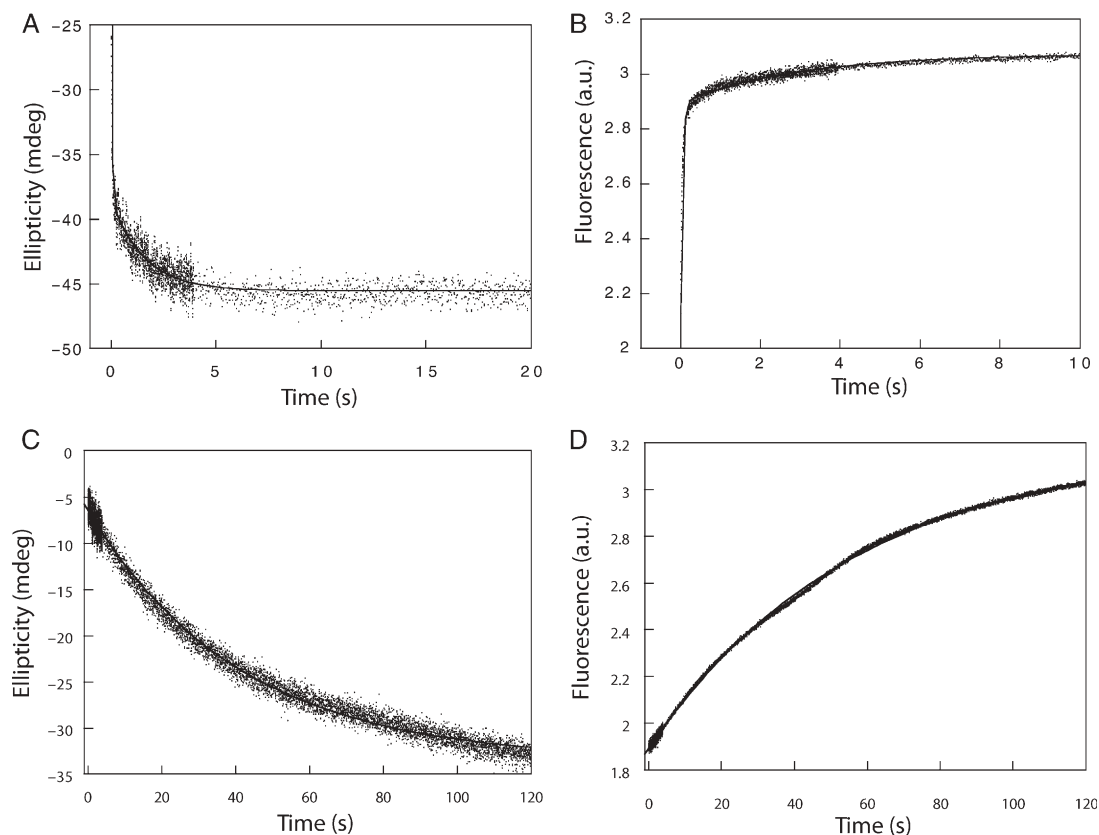


Fig. 4. Representative kinetic traces for the refolding of EA with fits to biphasic and monophasic transitions. (A) EA refolding monitored by stopped-flow CD spectroscopy when refolding to 2 M GuHCl. The trace is fit to a double exponential function. (B) EA refolding monitored by stopped flow fluorescence spectroscopy when refolding to 2.25 M GuHCl. The trace is fit to a double exponential function. (C) EA refolding monitored by stopped-flow CD spectroscopy when refolding to 3.5 M GuHCl. The trace is fit to a single exponential function. (D) EA refolding monitored by stopped-flow fluorescence when refolding to 3.5 M GuHCl. The trace is fit to a single exponential function.

Table II. Kinetic parameters for the homologues

Protein	k_f^0 (s ⁻¹)	k_s^0 (s ⁻¹)	k_u^0 (s ⁻¹)	k_f^0 , rel	k_s^0 , rel	k_u^0 , rel
PV	14	1.4	9.9×10^{-3}	1	1	1
Rop	29	0.5	6×10^{-7}	2.1	0.4	6×10^{-5}
EA	437	18.8	6×10^{-6}	31.2	13.4	6×10^{-4}

The homologues are listed in the order of increasing folding rate constants. k_f^0 and k_s^0 refer the rate constants at 0 M GuHCl for the fast and slow phases of refolding, respectively, and k_u^0 refers to the unfolding rate constant at 0 M GuHCl. The rate constants relative to those of PV's are also shown. The folding rate constants for PV and EA were determined at a concentration of 20 μ M dimeric protein, whereas the refolding rate constant for Rop was determined at a concentration of 100 μ M dimeric protein.

is refolded at higher concentrations of GuHCl, its refolding is monophasic (Fig. 5A and B). The refolding rates for PV are significantly slower than those for EA, and consequently there is very little missing amplitude even under conditions where this protein refolds the fastest, that is, when refolding to the lowest concentrations of GuHCl. When the refolding rate constants for the fast and slow phases of PV are calculated at 0 M GuHCl, they are slower than those of Rop (Fig. 6B and Table II).

Hence, the order in increasing refolding rates for the proteins is PV < Rop < EA.

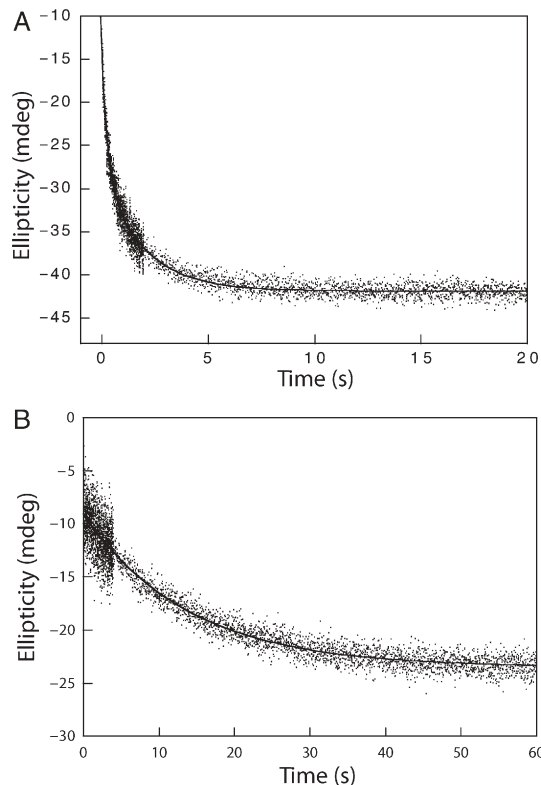


Fig. 5. Representative kinetic traces for the refolding of PV with fits to biphasic and monophasic transitions. (A) PV refolding monitored by stopped-flow CD when refolding to 0.4 M GuHCl. The trace is fit to a double exponential function. (B) PV refolding monitored by stopped-flow CD when refolding to 1.2 M GuHCl. The trace is fit to a single exponential.

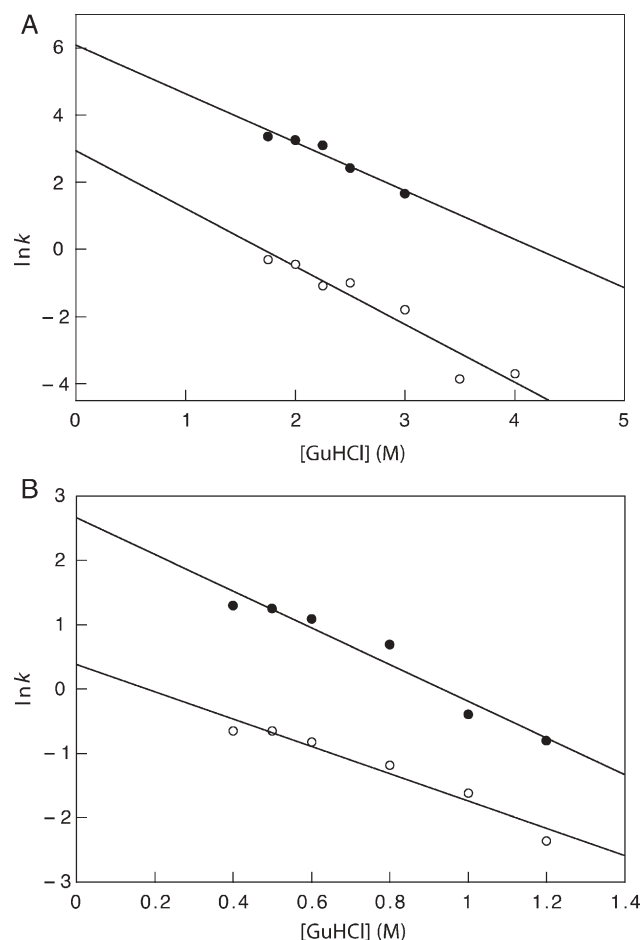


Fig. 6. $\ln k$ versus GuHCl plots for EA and PV. The closed circles and the open circles represent the fast and slow phases of folding, respectively. (A) $\ln k$ versus GuHCl for EA. The data are the average of the data obtained from fluorescence and CD spectroscopy. (B) $\ln k$ versus GuHCl for PV.

Protein concentration dependence for the refolding of EA and PV

To investigate whether the rate determining step in the folding process is unimolecular or multimolecular, we studied the dependence of the refolding rates for EA and PV on protein concentration. The plot of protein concentration versus the rate constant, k , shows that the fast phases for both EA and PV have a significant dependence on protein concentration, whereas the slow phases have a much lower dependence on protein concentration (Fig. 7). This observation is similar to the behaviour seen for Rop suggesting that the fast phase in EA and PV may represent the formation of a dimeric intermediate, as proposed for Rop (Munson *et al.*, 1996; Nagi *et al.*, 1999).

The protein concentration dependence of the refolding rates for EA measured by CD and fluorescence correlate well with each other (Fig. 7A and B). A comparison of the theoretical diffusion limited folding rate obtained by assuming that a doubling of the protein concentration results in a doubling of the refolding rate with those obtained experimentally for EA and PV (Fig. 7) demonstrates that the structure formation for both of these proteins is determined predominantly by the bimolecular collision event.

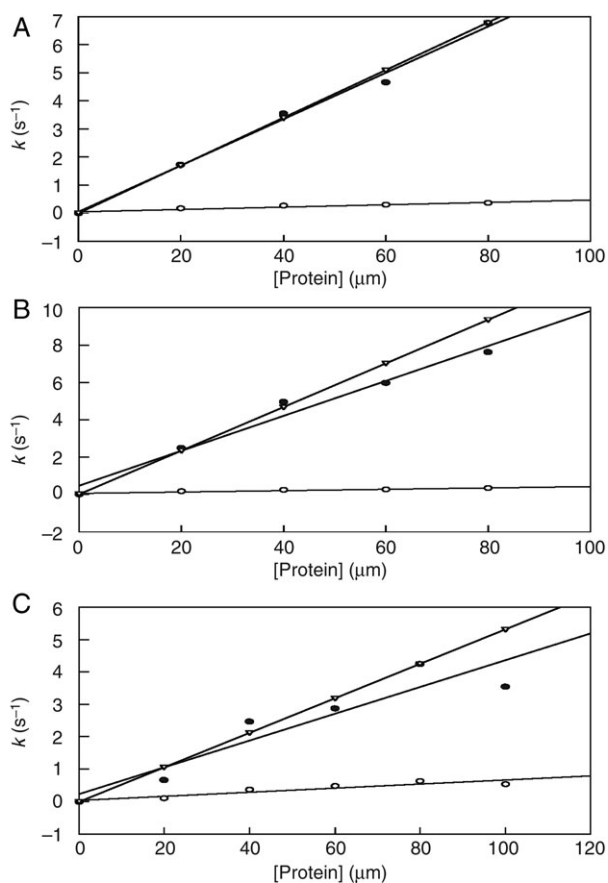


Fig. 7. Dependence of the refolding rates of EA and PV on protein concentration. The fast phase is shown by filled circles, the slow phase is shown by open circles. For comparison, the theoretical rates obtained by assuming a diffusion limited association of the monomers are shown in open triangles. (A) Refolding rate for EA as a function of protein concentration at 3 M GuHCl followed by stopped flow CD. (B) Refolding rate for EA as a function of protein concentration at 3 M GuHCl followed by stopped-flow fluorescence. (C) Refolding rate for PV as a function of protein concentration at 1 M GuHCl followed by stopped-flow CD.

Salt concentration dependence for the refolding of EA and PV

It was observed that the equilibrium stabilities of Rop, EA and PV are dependent on NaCl concentration—increasing NaCl concentration stabilizes the protein. Therefore, we probed the effect of increasing NaCl concentrations on the refolding rates for both proteins. The fast phase for both proteins has a significant dependence on NaCl concentration, whereby increasing NaCl concentration increases its rate. The rate of the slow phase is independent of NaCl concentration (Fig. 8). These observations suggest that an increase in the NaCl concentration facilitates the formation of the dimeric intermediate (see Discussion).

Unfolding of EA and PV

The unfolding kinetics of Rop are monophasic at all concentrations of GuHCl, and there are no detectable intermediates along the unfolding pathway. Rop unfolds extremely slowly, with a rate constant at 0 M GuHCl of $6 \times 10^{-7} \text{ s}^{-1}$. The unfolding of EA and PV are also monophasic at all concentrations of GuHCl. The slow unfolding of EA can be monitored by manual mixing. The rate of unfolding for EA

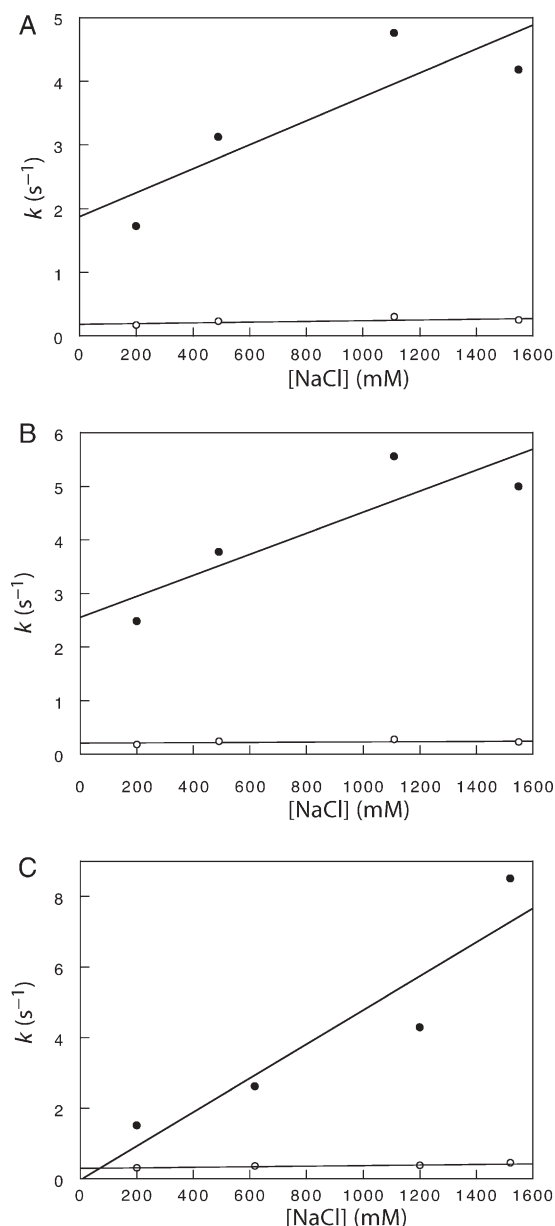


Fig. 8. Dependence of the refolding rates of EA and PV on NaCl concentration. The fast phase is shown by closed circles and the slow phase is shown by open circles. (A) Refolding rate for EA as a function of NaCl concentration at 3 M GuHCl followed by stopped-flow CD. (B) Refolding rate for EA as a function of NaCl concentration at 3 M GuHCl followed by stopped-flow fluorescence. (C) Refolding rate for PV as a function of NaCl concentration at 1 M GuHCl followed by stopped flow CD.

when extrapolated to 0 M GuHCl is $6 \times 10^{-6} \text{ s}^{-1}$ (Fig. 9, Table II). The unfolding rate of PV is faster than those for Rop and EA and was therefore monitored by stopped flow methods. The rate constant for PV when extrapolated to 0 M GuHCl is $9.9 \times 10^{-3} \text{ s}^{-1}$ (Fig. 9, Table II). Hence, the unfolding of EA is 10 times faster than that of Rop and the unfolding of PV is 1.7×10^4 -fold faster than that of Rop (Table II).

Discussion

We present a comparative study of a group of homologous four-helix bundle proteins. The high-resolution structure of

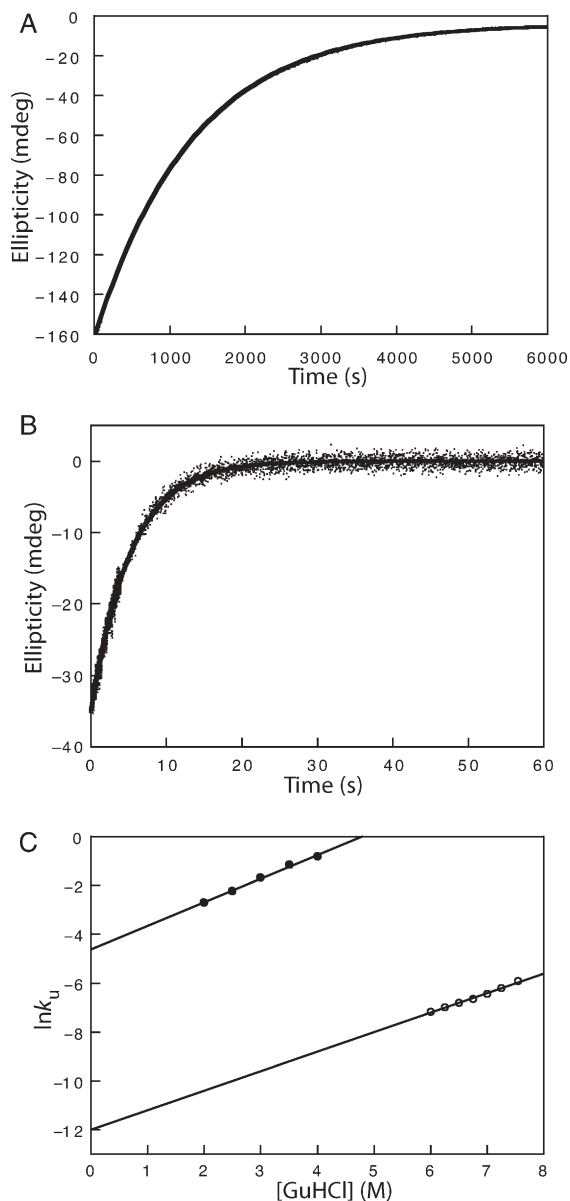


Fig. 9. Unfolding kinetics of EA and PV. (A) Representative kinetic trace for the unfolding of EA in 6 M GuHCl monitored by manual mixing CD. The data are fit to a single exponential. (B) Representative kinetic trace for the unfolding of PV in 3 M GuHCl monitored by stopped-flow CD. The data are fit to a single exponential. (C) $\ln k_u$ versus GuHCl for EA (open circles) and PV (filled circles).

one of these homologues, Rop, is known (Banner *et al.*, 1987; Eberle *et al.*, 1990). The other two homologues, EA and PV, are also highly helical and bind to the RNA complex that is recognized by Rop. This suggests that all three proteins share the same topology and similar *in vivo* functions.

The stabilities and folding rates of all three proteins are directly correlated. The rank order of stability for the three proteins is $PV < Rop < EA$. The rank order of folding rates is $PV < Rop < EA$ and the rank order of unfolding rates is $Rop < EA < PV$. It is therefore apparent that PV is destabilized relative to Rop and EA by a combination of factors: its folding rate is decreased and its unfolding rate is increased. EA is stabilized relative to Rop because of a faster folding

rate which more than compensates for the 10-fold increase in its unfolding rate.

The effect of systematically increasing the length of the loop connection between helix 1 and 2 in Rop on its stability and folding/unfolding kinetics has been previously studied (Nagi and Regan, 1997; Nagi *et al.*, 1999). Rop has a tight two-residue connection (Asp-Ala) between its helices. This sequence was systematically replaced to give 10 variants with between 1 and 10 glycine residues. It was observed that as loop length was increased, equilibrium stability of the protein decreased, the folding rate decreased and the unfolding rate increased. PV has a loop length of five amino acids (when aligned as shown in Fig. 1A) compared with two amino acids for Rop and EA and thus provided a natural protein in which to investigate the effect of increasing loop length on stability and kinetics. The trends of the kinetic and thermodynamic properties of PV mirror those of the longer loop variants of Rop (Nagi and Regan, 1997; Nagi *et al.*, 1999). The folding rate of PV is decreased relative to EA and Rop, presumably due to a stabilization of the unfolded ensemble resulting from the greater entropy allowed in the longer loop and/or an increase in the energy of the transition state relative to EA and Rop caused by constraining the longer loop in the helix-turn-helix monomer. The increase in unfolding rate is likely due to the decrease in stability of the native state of PV relative to Rop and EA.

Rop has also been used as a model protein to study the effects of systematically repacking the hydrophobic core with combinations of aliphatic residues (Munson *et al.*, 1994, 1996). The Rop core is very regular and is composed of eight layers with each helix of the four-helix bundle contributing one residue per layer (Fig. 1C). The repacked cores are composed entirely of two hydrophobic residues (e.g. alanine and leucine) packed in regular layers and have more 'idealized' cores compared with wild-type Rop which has a number of polar residues in its core (Fig. 1A). In particular, when Rop's core is repacked with Ala at the 'a' position, and Leu at the 'd' position of the hydrophobic core for all layers except for second and seventh layers where the packing is reversed (Ala₂Leu₂-8-rev), it results in a protein with increased stability compared with wild-type Rop while retaining the RNA-binding ability of the protein. EA is more similar in its hydrophobic core to the idealized variant Ala₂Leu₂-8-rev, and EA's second and seventh layers are identical to it. Its properties parallel what is observed with the repacked Rop core variants where each additional repacked layer increases the equilibrium stability of the variants, while increasing both the folding and unfolding rates (Munson *et al.*, 1996, 1997). The enhanced stability of EA along with its faster folding and unfolding rates suggests that the transition state is stabilized for EA relative to Rop. This is perhaps due to the presence of more favourable core packing interactions in its transition state compared with that of Rop. Although we discuss the properties of the homologues in terms of their core packing interactions and the length of loop connections between elements of secondary structure, it is important to emphasize that other differences in the sequences may also play a significant role.

The dependence of the refolding rates on protein concentration is similar for Rop, EA and PV. This suggests that they all follow a similar folding mechanism that proceeds through the formation of a dimeric intermediate. For EA and PV, the

fast phase has a significant dependence on NaCl concentration, such that higher NaCl concentrations increase its rate. The rate of the slow phase is independent of NaCl concentration (Fig. 8). The dependence on NaCl concentration of the fast phase for refolding suggests that increasing NaCl concentration facilitates the formation of the dimeric intermediate, possibly by facilitating the hydrophobic collapse step or by shielding the repulsion of the negative charge as the 2/2' face forms.

Our results from this study on a dimeric protein family are similar to the results from a number of studies on monomeric protein families (Kragelund *et al.*, 1996; Clarke, 1997; Plaxco *et al.*, 1997; Martinez and Serrano, 1999; Riddle *et al.*, 1999; Chiti *et al.*, 1999). We have demonstrated that even though members of a homologous protein family may have differing thermodynamic and kinetic properties, they all follow a similar folding and unfolding mechanism.

We have shown that this group of Rop homologues provide a representative set in which to relate the equilibrium stabilities to the kinetic properties of a family of protein variants. The properties of the homologues are discussed in relation to the properties of the designed proteins in which Rop was used as a model system. Studies on families of homologous proteins clearly demonstrate how small changes at the sequence level can dramatically influence the kinetics and stability of a protein while maintaining the overall three-dimensional fold and function.

Materials and methods

Database search to identify the Rop homologues

The BLAST search was performed using the entire sequence of wild-type Rop protein and searching against a non-redundant GenBank CDS translations + PDB + Swiss Prot + PIR + PRF database at www.ncbi.nlm.nih.gov/BLAST/.

Chemicals

GuHCl, ultra pure, was purchased from American Bioanalytical (Natick, MA, USA).

Cloning and protein purification

The genes encoding Rop, EA and PV sequences were constructed into the vector pMR103 (Munson *et al.*, 1994) using overlapping DNA templates in a polymerase chain reaction. A glycine residue was incorporated as the residue following the initiator methionine in order to ensure uniform processing of the N-terminal methionine by N-terminal methionine aminopeptidase (Smith *et al.*, 1995). The sequences of the genes were confirmed by dideoxynucleotide sequencing. All proteins were expressed and purified as described previously for Rop, using anion exchange chromatography (Munson *et al.*, 1994, 1996). The purified proteins were concentrated using CentriPrep 3.5 (Amicon) concentrators and dialyzed against 100 mM Na-phosphate (pH 7), 200 mM NaCl (phosphate buffer), for use in experiments. The identity of the proteins was confirmed by amino acid analysis and MALDI mass spectrometry.

Equilibrium studies

All CD experiments were performed on an AVIV 62DS spectropolarimeter (Aviv Instruments, Lakewood, NJ, USA). Experiments were performed using 20 μ M dimeric protein in phosphate buffer, in a 2 mm pathlength cuvette, unless explicitly stated. The protein concentration determined initially by amino acid analysis was used to calculate the MRE of the proteins. To determine the protein concentration for use in experiments, the CD ellipticity at 222 nm was measured and the MRE was used to calculate the protein concentration. We estimate the error in protein concentration determination to be approximately $\pm 5\%$. Far-UV wavelength scans were obtained from 260 to 200 nm with an averaging time of 5 s. Protein unfolding was monitored by following change in ellipticity at 222 nm either as a function of temperature or GuHCl concentration in phosphate buffer supplemented with 1 mM DTT. The thermal denaturation was performed in increments of 1°C, with a signal averaging time of 1 min and an equilibration time at each temperature of 1 min.

To ensure complete equilibration for the chemical denaturation, samples were incubated in varying concentrations of phosphate buffered GuHCl for points along the equilibrium denaturation curve, for ~ 48 h. The concentrations of the stock GuHCl and samples used in denaturation experiment were determined by measuring the refractive index with an Atago R5000 refractometer (Atago, Japan) (Pace, 1986).

The GuHCl denaturation curves were analyzed to derive the free energy and m -values. To obtain these values, the baselines were corrected manually and the K_d values were obtained using $K_d = 2C_T(F_U)^2/(F_N)$, where C_T = total concentration of dimeric protein, F_U = fraction unfolded and F_N = fraction native. Using these K_d values, the ΔG at each concentration of GuHCl was calculated using the relationship $\Delta G = -RT \ln K_d$. In order to calculate the free energy value in the absence of denaturant, the graph of ΔG versus GuHCl is fit to a straight line. The intercept of this line is ΔG^0 and the slope of this line gives the m -value.

Kinetic studies

The kinetics of refolding were followed by stopped-flow CD and stopped-flow fluorescence. The unfolding kinetics of PV were followed by stopped-flow CD and the unfolding kinetics of EA were monitored by CD using manual mixing methods. EA and GuHCl were mixed to initiate unfolding. In manual mixing experiments, typically 20–25 s elapsed between the mixing of solutions, transfer of the solution to a cuvette and transfer of the cuvette to the CD instrument. The loss of secondary structure for EA was monitored as a function of time at 222 nm, with an averaging time and time constant of 3 s.

Stopped-flow methods

Prior to performing the refolding experiments, PV was denatured in 4 M GuHCl and EA was denatured in 6.5 M GuHCl in phosphate buffer, supplemented with 10 mM DTT. All the kinetic experiments were performed in the presence of 10 mM DTT to ensure complete reduction of the cysteine residues in the proteins. The samples were equilibrated overnight in denaturant. Unfolding and refolding experiments were carried out using a Bio-Logic (Claix, France) SFM3 mixer with a cell path length of 1.5 mm and a Bio-Logic

PMS 400 detection system. The theoretical dead time based on the flow rate was 5.4 ms. To enable the CD and fluorescence signals to be monitored simultaneously, the excitation wavelength was set at 220 nm and the fluorescence emission was detected at 305 nm using a monochromator set at the emission wavelength. To improve the signal-to-noise ratio of the fluorescence experiment, EA refolding was repeated under exactly the same conditions but by exciting at 275 nm. Three syringes were used in the experiments: for refolding, the first contained GuHCl buffered in phosphate at the highest desired concentration of GuHCl, the second contained native phosphate buffer and the third contained the unfolded protein at $\times 10$ the final concentration. To monitor the unfolding for PV, the three syringes were as follows: the first one contained GuHCl buffered in phosphate buffer, the second contained native phosphate buffer and the third one contained protein in native phosphate buffer at $\times 10$ the final concentration. In general, varying volumes of the components in syringes 1 and 2 were mixed to get the desired final concentration of GuHCl while maintaining a constant protein concentration of 20 μM dimer, except in experiments that probed protein concentration. Data acquisition was performed using two time bases (every 2 ms for the first 4 s and thereafter every 10 ms) in order to record a higher number of points in the initial times following mixing. Hence, the signals show a higher noise level, because they are less filtered, at the start of the kinetic profiles.

Analysis of kinetic data

Stopped-flow CD and manual mixing CD data were fit to either a single or double exponential growth or decay function.

For the growth function,

$$S(t) = S_0 - A(1 - e^{-kt})$$

or

$$S(t) = S_0 - A_1(1 - e^{-k_1t}) - A_2(1 - e^{-k_2t})$$

for the decay function,

$$S(t) = S_\infty + A(e^{-kt})$$

or

$$S(t) = S_\infty + A_1(e^{-k_1t}) + A_2(e^{-k_2t})$$

where $S(t)$ is the signal at time t , S_0 the signal at time $t = 0$, S_∞ the signal at time $t = \infty$, k , k_1 and k_2 the rates of the single phase or of the fast and slow phases, respectively, and A , A_1 and A_2 the signal amplitudes associated with those phases. To determine whether the data fit better to the single exponential or double exponential, fits to both were performed and when an obvious improvement in residual plots was observed, two exponentials were used.

Data sets are the average between 5 and 8 separate mixing events. The average rate constants obtained at different GuHCl concentrations were used to determine the rate constants, k^0 , in the absence of denaturant by linear extrapolation of the plot of $\ln k$ versus $[\text{GuHCl}]$ to 0 M GuHCl concentration. To generate this line for the proteins at least five different concentrations of GuHCl were used.

The unfolding data were used to plot $\ln k_u$ versus $[\text{GuHCl}]$ for at least five different GuHCl concentrations and a k_u^0 was obtained by extrapolation to 0 M GuHCl. The standard deviation for the rate constants derived from individual fits is approximately $\pm 1\%$ and the standard deviation for the average rate constants is approximately $\pm 3\%$. Taking into account the error in individual rate constants, along with the uncertainties in protein and denaturant concentrations, we estimate that the reported rate constants have an associated error of $\pm 15\%$.

RNA-binding studies

RNA I and RNA II were generated by T7 RNA polymerase transcription of DNA templates. The RNA I and RNA II transcripts were heated separately to 85°C for 2 min and then quickly cooled on ice for ~ 5 min to facilitate the formation of the kiss complex preferentially over the duplex. RNA I and RNA II were then mixed and the protein and incubation buffer (final concentration of 25 mM *tris*-borate, 5 mM MgCl_2 and 100 mM NaCl) were added and incubated on ice for ~ 30 min. Reactions were loaded onto an 18% (36:1 crosslink) native polyacrylamide gel in TBM buffer (25 mM *tris*-borate, 5 mM MgCl_2), and run at 4°C and 100 V to separate the kiss complex from the protein-RNA complex. Bands were visualized by staining with ethidium bromide.

Acknowledgements

We would like to thank the members of the Regan laboratory for critical reading of the manuscript.

Funding

National Institute of Health (GM 49146-01A1); NATO Collaborative Travel Grant. The work was also supported by the Oxford Centre for Molecular Sciences, with funding from the UK BBSRC, EPSRC and MRC.

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
- Banner, D.W., Kokkinidis, M. and Tsernoglou, D. (1987) *J. Mol. Biol.*, **196**, 657–675.
- Brockwell, D.J. and Radford, S.E. (2007) *Curr Opin Struct Biol.*, **17**, 30–37.
- Burns, L.L., Dalessio, P.M. and Ropson, I.J. (1998) *Proteins*, **33**, 107–118.
- Calvin Koons, M.D. and Blumenthal, R.M. (1995) *Gene*, **157**, 73–79.
- Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G. and Dobson, C.M. (1999) *Proc. Natl Acad. Sci. USA*, **96**, 3590–3594.
- Clarke, J., Hamill, S.J. and Johnson, C.M. (1997) *J. Mol. Biol.*, **270**, 771–718.
- Comolli, L.R., Pelton, J.G. and Tinoco, I. Jr (1998) *Nucleic Acids Res.*, **26**, 4688–4695.
- Dobson, C.M. (2003) *Nature*, **426**, 884–890.
- Eberle, W., Klaus, W., Cesareni, G., Sander, C. and Rosch, P. (1990) *Biochemistry*, **29**, 7402–7407.
- Ferguson, N., Capaldi, A.P., James, R., Kleanthous, C. and Radford, S.E. (1999) *J. Mol. Biol.*, **286**, 1597–1608.
- Gittelman, M.S. and Matthews, C.R. (1990) *Biochemistry*, **29**, 7011–7020.
- Gregorian, R.S. Jr and Crothers, D.M. (1995) *J. Mol. Biol.*, **248**, 968–984.
- Gross, M. (1996) *FEBS Lett.*, **390**, 249–252.
- Ivankov, D.N., Garbuzynskiy, S.O., Alm, E., Plaxco, K.W., Baker, D. and Finkelstein, A.V. (2003) *Protein Sci.*, **12**, 2057–2062.
- Kragelund, B.B., Hojrup, P., Jensen, M.S., Schjerling, C.K., Juul, E., Knudsen, J. and Poulsen, F.M. (1996) *J. Mol. Biol.*, **256**, 187–200.
- Lee, A.J. and Crothers, D.M. (1998) *Structure*, **6**, 993–1005.
- Marino, J.P., Gregorian, R.S. Jr, Csankovszki, G. and Crothers, D.M. (1995) *Science*, **268**, 1448–1454.
- Martinez, J.C. and Serrano, L. (1999) *Nat. Struct. Biol.*, **6**, 1010–1016.

- Mikiewicz,D., Wrobel,B., Wegrzyn,G. and Plucienniczak,A. (1997) *Plasmid*, **38**, 210–219.
- Milla,M.E. and Sauer,R.T. (1994) *Biochemistry*, **33**, 1125–1133.
- Munson,M., Predki,P.F. and Regan,L. (1994) *Gene*, **144**, 59–62.
- Munson,M., Balasubramanian,S., Fleming,K.G., Nagi,A.D., O'Brien,R., Sturtevant,J.M. and Regan,L. (1996) *Protein Sci.*, **5**, 1584–1593.
- Munson,M., Anderson,K. and Regan,L. (1997) *Fold. Des.*, **2**, 77–87.
- Nagi,A.D. and Regan,L. (1997) *Fold. Des.*, **2**, 67–75.
- Nagi,A.D., Anderson,K.S. and Regan,L. (1999) *J. Mol. Biol.*, **286**, 257–265.
- Pace,C.N. (1986) *Methods Enzymol.*, **131**, 266–280.
- Plaxco,K.W., Spitzfaden,C., Campbell,I.D. and Dobson,C.M. (1997) *J. Mol. Biol.*, **270**, 763–770.
- Plaxco,K.W., Simons,K.T. and Baker,D. (1998) *J. Mol. Biol.*, **277**, 985–994.
- Polisky,B. (1988) *Cell*, **55**, 929–932.
- Predki,P.F., Nyak,L.M., Gottlieb,M.B.C. and Regan,L. (1995) *Cell*, **80**, 41–50.
- Riddle,D.S., Grantcharova,V.P., Santiago,J.V., Alm,E., Ruczinski,I. and Baker,D. (1999) *Nat. Struct. Biol.*, **6**, 1016–1024.
- Sali,A., Shakhnovich,E. and Karplus,M. (1994) *Nature*, **369**, 248–251.
- Sauer,R.T., Milla,M.E., Waldburger,C.D., Brown,B.M. and Schildbach,J.F. (1996) *FASEB J.*, **10**, 42–48.
- Smith,C.K., Munson,M. and Regan,L. (1995) *Techniques in Protein Chemistry VI*. Academic Press, pp. 323–332. .
- Wendt,H., Berger,C., Baici,A., Thomas,R.M. and Bosshard,H.R. (1995) *Biochemistry*, **34**, 4097–4107.
- Woese,C.R., Kandler,O. and Wheelis,M.L. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 4576–4579.
- Wolynes,P.G. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 6170–6175.
- Wright,C.F., Christodoulou,J., Dobson,C.M. and Clarke,J. (2004) *Protein Eng Des Sel.*, **17**, 443–453.
- Zitzewitz,J.A., Bilsel,O., Luo,J., Jones,B.E. and Matthews,C.R. (1995) *Biochemistry*, **34**, 12812–12819.

**Received December 6, 2007; revised December 6, 2007;
accepted December 11, 2007**

Edited by Regina Murphy