Amino-acid substitutions in a surface turn modulate protein stability

Paul F. Predki[1], Vishal Agrawal[2], Axel T. Brünger[2] and Lynne Regan

A surface turn position in a four-helix bundle protein, Rop, was selected to investigate the role of turns in protein structure and stability. Although all twenty amino acids can be substituted at this position to generate a correctly folded protein, they produce an unusually large range of thermodynamic stabilities. Moreover, the majority of substitutions give rise to proteins with enhanced thermal stability compared to that of the wild type. By introducing the same twenty mutations at this position, but in a simplified context, we were able to deconvolute intrinsic preferences from local environmental effects. The intrinsic preferences can be explained on the basis of preferred backbone dihedral angles, but local environmental context can significantly modify these effects.

Department of Molecular Biophysics and Biochemistry Yale University, New Haven, Connecticut 06520, USA

[1]Present address: CuraGen Corporation Branford, Connecticut 06405,USA
[2]Howard Hughes Medical Institute Yale University, New Haven, Connecticut 06520, USA

Correspondence should be addressed to L.R.

The role of turns in determining the structure and stability of proteins is not well understood. On one hand, certain turns appear simply to connect elements of protein secondary structure, and considerable variability in the identity of the turn residues is permitted[1–4]. On the other hand, the identity of the amino acid at specific turn positions can be important, and may exert significant influence on the ability of a turn to form[5–9]. These two extremes are not necessarily mutually exclusive. Here we address in detail the role of a specific turn in determining the structure and stability of a protein.

Our model system is the protein Rop[10,11], a four-helix bundle formed by the association of two identical helix-turn-helix monomers (Fig. 1*a*). The two helices within each monomer are connected by a tight, two-residue turn. The focus of our studies is Asp 30, the first residue of this turn and one of the few residues in the protein with non-helical backbone dihedral angles. Asp 30 is highly solvent exposed, while its neighbouring residues, Leu 29 and Ala 31, pack into the hydrophobic core of the protein. The local structural environment of the turn in wild-type Rop is depicted in Fig. 1*b*.

Effect of mutagenesis

The first stage of our investigation involved mutating position 30 to each of the 19 other amino acids. As Rop is almost entirely helical, circular dichroism (CD) provides a sensitive probe of the overall integrity of the protein's secondary structure. Each position 30 mutant has a CD spectrum that is nearly identical (±5% at 222 nm) to that of the wild-type protein (mean residue ellipticity 3.2×10^4 deg cm^{-1} mol^{-1} at 222 nm), indicating that the position 30 mutations cause no major structural perturbations. Because Asp 30 is not involved in binding Rop's natural RNA ligand[12], the affinity of a position 30 mutant for RNA serves as an additional sensitive probe of the protein's structural integrity. All 19 position 30 mutants bind RNA with affinities that are comparable to that of wild-type Rop (Fig. 2).

The results outlined above provided the first indication that the position 30 variants fold into structures that are similar to that of wild-type Rop. From this evidence alone, one would conclude that turn position 30 is very tolerant of amino-acid substitutions, and that any amino acid can function at this position equally well. The interpretation of these results changes dramatically, however, when the stabilities of the position 30 variants are compared.

We used CD to monitor the thermal denaturation transitions of the position 30 mutants and calculated the effect of each mutation on protein stability. Of the nineteen position 30 substitutions, twelve give rise to proteins that are more stable than the wild type. Furthermore, there is a substantial (2.8 kcal mol^{-1}) difference between the most and least stable mutant (excluding Asp30Pro) (Table 1). This range is comparable, if not slightly greater, than those measured for α-helical and β-sheet forming propensities at solvent-exposed sites[13–20]. These results indicate that there are distinct thermodynamic preferences for the identity of the amino acid at position 30.

Crystallographic analysis

Before proceeding with further analyses, we selected the most stable of the position 30 mutants, Asp30Gly, for high-resolution structural analysis by X-ray crystallography (Table 2). Fig. 3*a* shows the electron density

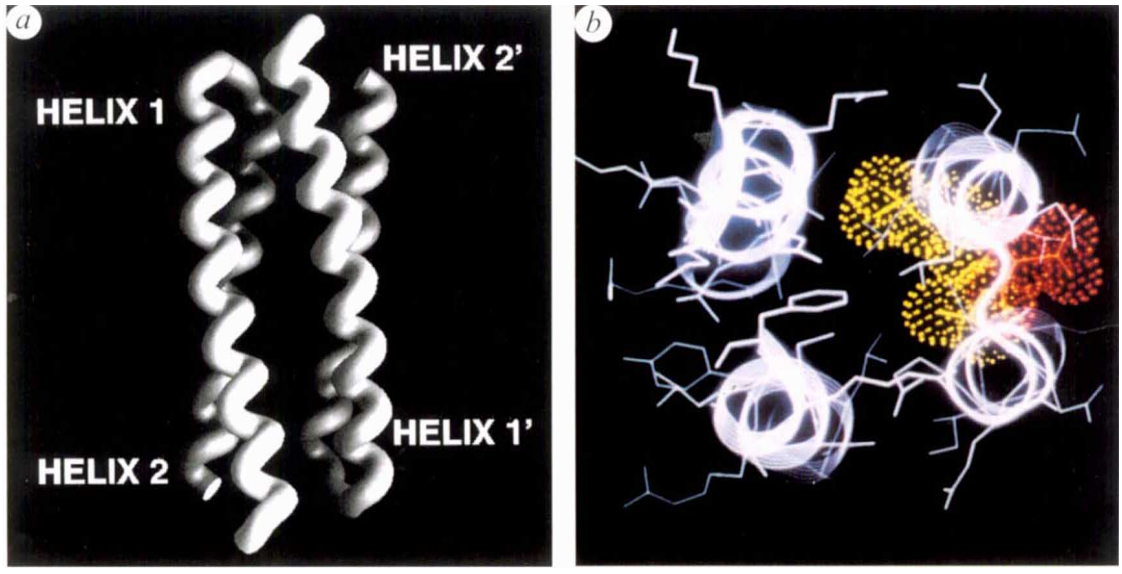

Fig. 1 Structure of Rop and the turn. *a*, Backbone representation of the structure of wild-type Rop[10]. Helices 1 and 2 of the first monomer and 1′ and 2′ of the second are indicated. The short unstructured C-terminal tail is not depicted. Created using the program GRASP[32]. *b*, Closeup of the turn. The aspartate residue at position 30 is shown in red. Neighbouring residues leucine 29 and alanine 31 are shown in yellow. The van der Waals radii of residues 29–31 are indicated by appropriately coloured dots. Other residues are shown in white. The protein backbone is depicted by ribbons. Creating using the program Insight II (Biosym Technologies).

and the model of the turn of the Asp30Gly mutant. The backbone dihedral angles of position 30 are similar to those of the wild-type protein, with averaged values of ϕ=78°, ψ=11°. Fig. 3*b* shows a superposition of the backbone of wild-type Rop (yellow) and the Asp 30Gly mutant (red). The r.m.s. deviation of residues 29–31 backbone atoms is 0.08 Å. This close structural similarity led us to propose that the exceptional stability of the Asp30Gly mutant may be explained in terms of the ease with which Gly residues in particular can adopt the left-handed helical backbone dihedral angles observed in position 30[21].

We therefore considered the relative stability of all the different position 30 mutants in terms of their ability to adopt the ϕ-ψ angles of residue 30. To this end, we performed a survey of a nonhomologous subset of high-resolution protein structures. For each amino acid in this database, the probability of adopting the backbone dihedral angles observed at position 30 in wild-type Rop was calculated. In an approach similar to one taken earlier by Muoz and Serrano[22], we then compared the experimental stabilization free energy for each turn substitution with the statistical preference for that residue to adopt the ϕ-ψ angles of residue 30, ±15°. With the possible exception of glycine and proline, due to their exceptionally high and low respective amounts of conformational freedom, the dihedral angles resulting from substitutions at position 30 are expected to fall close to that of the wild-type values. However, the correlation between the statistical preferences and the experimentally observed stabilization free energies was low (Fig. 4*a*), suggesting that significant local environmental effects may be superimposed upon the intrinsic effects of backbone dihedral preferences.

Mutagenesis in an altered structural context

In an attempt to deconvolute intrinsic and context-dependent effects, we made the same 19 substitutions at position 30, but in a simplified local environment, created by mutating the two nearest side chains, Asp 28 and Glu 32, to alanine. All twenty of these altered context position 30 mutants retained the overall protein fold, as suggested by the similarity of their CD spectra

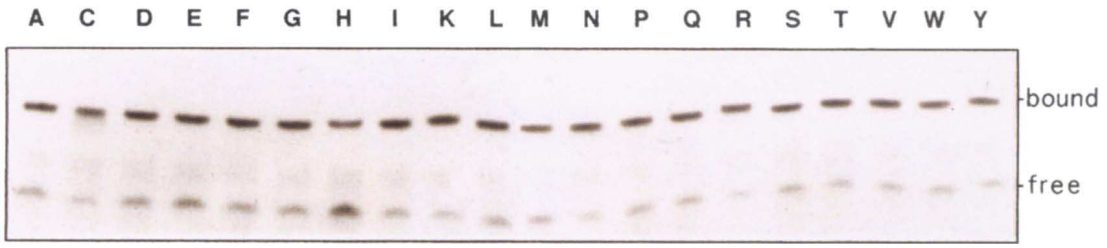

Fig. 2 Electromobility shift assay of the wild-type and 19 position 30 Rop mutants. The identity of each mutant is indicated above each lane. The locations of free and protein-bound forms of the RNA are indicated on the right hand side.

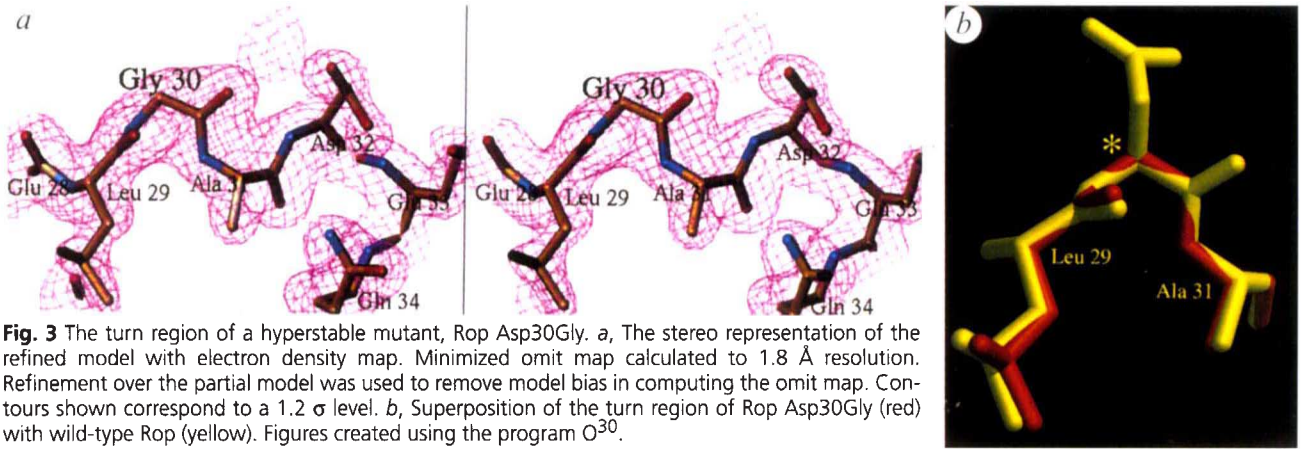

Fig. 3 The turn region of a hyperstable mutant, Rop Asp30Gly. *a*, The stereo representation of the refined model with electron density map. Minimized omit map calculated to 1.8 Å resolution. Refinement over the partial model was used to remove model bias in computing the omit map. Contours shown correspond to a 1.2 σ level. *b*, Superposition of the turn region of Rop Asp30Gly (red) with wild-type Rop (yellow). Figures created using the program O[30].

and RNA-binding affinities to those of wild-type Rop (data not shown).

As expected, a comparison between the protein database and the 20 new $\Delta\Delta G$ energies reveals an improved correlation (new R=0.78, *versus* R=0.57, data not shown). The new experimental stabilization free energies were compared with the stability of proteins with corresponding mutations in the wild-type context. The rationale for making this cross-correlation was as follows: if there is no interaction between the side chains of residues 28 and 32 and that of position 30, then the stabilities in the two environments will correlate exactly. Any change in stability would reflect the inherent change in stability caused by the Glu28Ala and Asp32Ala mutations, and would be identical for all mutants. Conversely, if there are significant interactions between position 30 side chains and side chains of residues 28 and 32, then the effect of each position 30 mutation will be different.

We found that ten of the position 30 substitutions all show a similar increase in Tm in the altered context (2±1 °C). These amino acids fall on the diagonal in the correlation plot, Fig. 4*b*. This slight increase presumably reflects the stabilizing effect of the position 28 and 32 mutations, as expected for the introduction of alanine in an α-helix. By contrast, the other position 30 variants display a range of effects that include both increases and decreases in stability, presumably reflecting interactions of these position 30 amino acids with the side chains at positions 28 and 32 in wild-type Rop.

The correlation plot (Fig. 4*b*) therefore identifies two subsets of mutations. The residues that fall off the diagonal have uncorrelated effects in the different environments, and represent those mutants for which context effects dominate stability. For the subset which fall on the diagonal, however, intrinsic preferences play the major role. This information was used to re-examine the apparently poor relationship between the statistical backbone dihedral preferences and the experimentally observed stabilization free energies (Fig. 4*a*). Fig. 4*a* is replotted in Fig. 4*c* using only those amino acids which fall on the diagonal in Fig. 4*b*. A dramatically improved correlation is observed (R=0.90). The observation of a good correlation between the statistical backbone dihedral preferences and the experimentally observed stabilization free energies is only seen for this subset of mutants. Thus, we are able to demonstrate that the intrinsic preferences of these amino acids at position 30 are dominated by their backbone dihedral angle preferences.

The important role we observe for these dihedral preferences in protein stability is likely an key feature of many turns, especially those with constrained backbone dihedral angles. However, local context can significantly modify these effects. It is expected that the cross-correlation mutagenesis approach we used to demonstrate the importance of both intrinsic and context-dependent effects should be a generally applicable method of deconvoluting thermodynamic preferences.

Table 1 Melting temperature and $\Delta\Delta G_{folding}$ for wild-type and mutant Rop proteins.

Amino acid	T_m(°C)	$\Delta\Delta G_{342}$(kcal mol^{-1})
Gly	80.3	-2.0
Gln	79.1	-1.8
Glu	74.9	-1.0
Ser	74.7	-1.0
Lys	74.6	-0.9
His	74.5	-0.9
Cys	74.0	-0.8
Arg	74.0	-0.8
Asn	73.8	-0.8
Met	72.7	-0.6
Ala	71.2	-0.3
Tyr	70.4	-0.2
Asp	68.7	0.0
Leu	68.4	0.1
Phe	68.3	0.1
Thr	66.6	0.4
Val	66.6	0.4
Trp	65.9	0.4
Ile	63.4	0.8
Pro	58.9	1.6

The amino acid in position 30 is indicated in the left column. Experimental error yields $\Delta\Delta G_{folding}$ values accurate within ~±0.15 kcal mol^{-1} . $\Delta\Delta G$ is relative to the stability of Rop (Asp 30).

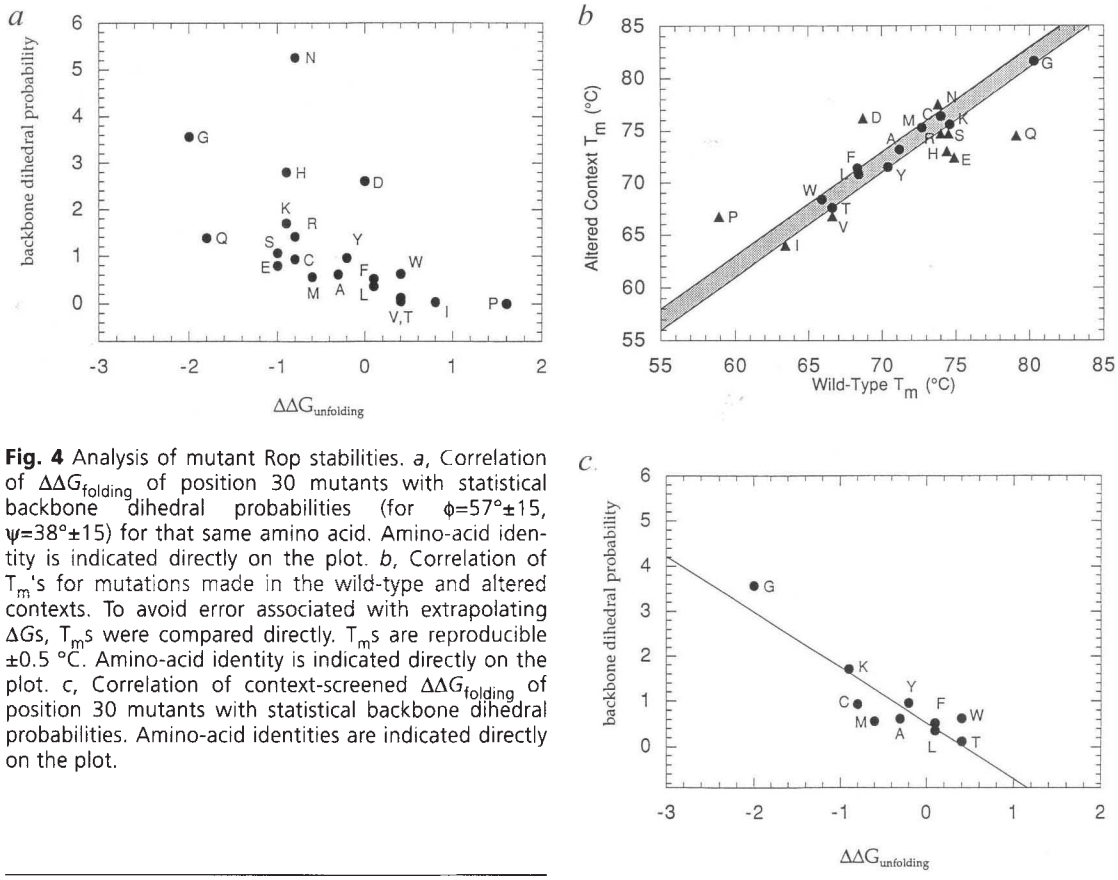

Fig. 4 Analysis of mutant Rop stabilities. *a*, Correlation of $\Delta\Delta G_{folding}$ of position 30 mutants with statistical backbone dihedral probabilities (for ϕ=57°±15, ψ=38°±15) for that same amino acid. Amino-acid identity is indicated directly on the plot. *b*, Correlation of T_m's for mutations made in the wild-type and altered contexts. To avoid error associated with extrapolating ΔGs, T_ms were compared directly. T_ms are reproducible ±0.5 °C. Amino-acid identity is indicated directly on the plot. *c*, Correlation of context-screened $\Delta\Delta G_{folding}$ of position 30 mutants with statistical backbone dihedral probabilities. Amino-acid identities are indicated directly on the plot.

Methods

Protein purification and RNA binding assay. Protein purification, electromobility shift assays and K_d determination were performed as described previously[12,23]. Samples (5 µg) of each protein were incubated with a trace amount of ^{32}P-labelled RNA. Within limits of error, apparent mutant K_ds are identical to that of the wild-type protein (w.t. $K_d = 2\times10^{-6}$ M). Estimated error in K_d values is ±10%.

Thermodynamic analysis. Thermodynamic calculations were performed as previously described[12]. The enthalpy of unfolding was determined by van't Hoff plots and free energies of unfolding were calculated at an intermediate temperature of 342 K to decrease errors associated with extrapolation. An estimate of ΔC_p (the difference in heat capacity between the native and denatured states) of 1.0 kcal K^{-1} mol^{-1} was determined from the slope of a plot of $\Delta H_{unfolding}$ *versus* T_m for wild-type protein.

Statistical analysis. A database of protein structures with less than 50% homology to each other (available from authors by request) was used for statistical determination of dihedral angles frequencies. The program WHATIF[24] was used to determine both ϕ and ψ angles for 52,743 amino acids within this 177 protein database. This dihedral angle database was then used to calculate the percent probabilities of each amino acid occupying the backbone dihedral angles found in position 30 in wild-type Rop (ϕ=57±15°, ψ=38±15°). This probability is defined as the number of times each amino acid is found in the database with backbone dihedral angles of ϕ=57±15° and ψ=38±15°, divided by the number of occurrences of that amino acid (excluding terminal amino acids), times 100.

Rop Asp30Gly structure solution. The crystal structure of Rop Asp30Gly was determined using the direct rotation function[25], Patterson correlation refinement[26], and the correlation coefficient translation function[27] as implemented in the program X-PLOR[28]. Details of the structure solution process will be published elsewhere. The crystal structure of wild-type Rop[10] served as the search model for molecular replacement. Molecular replacement indicated three Rop protomers in the asymmetric unit, two of which are related by a non-crystallographic two-fold relationship forming a Rop dimer. The third Rop protomer forms another Rop dimer through a crystallographic symmetry mate. The positions and orientations of the correctly placed protomers were adjusted by rigid-body refinement. The correctness of the molecular replacement solution was checked by computing a

Table 2 Rop Asp30Gly crystallographic data summary

Space group	I422
Cell axes (Å)	*a*=*b*=95.117 *c*=85.561
Resolution range (Å)	20.0–1.82
Highest resolution shell (Å)	1.87–1.82
Number of reflections	17,635
Average redundancy	7.9
Completeness (%)[1]	98.9 (99.4)
R_{sym}[1,2]	0.049 (0.241)
Data with $I>2\sigma_I$ (%)[1]	92.3 (75.1)

[1]Values in parenthesis are for the highest resolution shell.
[2]$R_{sym}(|F|)=\Sigma_h\Sigma_i||F_i(h)|-<|F(h)|>|\Sigma_h<|F(h)|>$ where $F_i(h)$ is the amplitude of the i^{th} measurement and $<|F(h)|>$ is the weighted mean of all measurements of $|F(h)|$.

complete omit map and by computing an anomalous difference map confirming the positions of nearly all sulphur atoms (methionine and cysteine) in the crystal structure. Throughout the refinement, annealed omit maps were used to reduce model bias[29]. Iterative molecular modelling using the program O[30] was performed with relatively minor corrections from the wild-type Rop carbon backbone. Ordered water positions (91) were identified by selecting well-separated and approximately spherical peaks at 1.5 σ above the average density of the experimental map, with the protein masked out and within hydrogen bonding distance to polar or charged atoms. Overall *B*-factor refinement, overall anisotropic *B*-factor refinement, conjugate gradient positional refinement, and several independent slow-cooling simulated annealing cycles[31], interspersed with rounds of model building, were performed. The free *R*-value was monitored using a test set containing a random selection of 10% of the diffraction data. Non-crystallographic symmetry constraints or restraints were not imposed at any point in the refinement. All refinements were carried out with X-PLOR[28]. The coordinates will be deposited in the Brookhaven Protein Data Bank.

Received 28 September; accepted 21 November 1995.

Table 3 Statistics of model refinement

Resolution range	50.0–1.82
Completeness	98.9
R-factor[1]	0.224
Free *R*-factor	0.227
Number of:	
protein residues	176
protein non-H atoms	1469
water molecules	91
Overall *B*-factor	41.2
Average *B* main chain (Å^2)	38.3
Average *B* side chain (Å^2)	43.8
Average *B* water (Å^2)	46.1
R.m.s. bond lengths (Å)	0.005
R.m.s. bond angles (°)	1.15
R.m.s. dihedral angles (°)	17.0
R.m.s. improper angles (°)	0.84
Residues in disallowed ϕ–ψ regions (%)	0.6

[1] $R = \Sigma||F_{obs}|-|F_{calc}||/\Sigma|F_{obs}|$

Acknowledgements
Thanks to Regan lab members for critical reading of the manuscript and helpful discussions. We thank Dushyant Pathak for invaluable help during crystallization of Rop Asp30 Gly. L.R. is a National Science Foundation National Young Investigator and a Dreyfus Teacher Scholar. P.F.P. was a Medical Research Council (Canada) post-doctoral fellow. This work was supported by the NIH.

1. Brunet, A.P. *et al.* The role of turns in the structure of an α-helical protein. *Nature* **364**, 355–358 (1993).
2. Castagnoli, L., Vetriani, C. & Cesareni, G. Linking an easily detectable phenotype to the folding of a common structural motif: Selection of rare turn mutations that prevent the folding of Rop. *J. molec. Biol.* **237**, 378–387 (1994).
3. Vlassi, M. *et al.* Restored heptad pattern continuity does not alter the folding of four-α-helix bundle. *Nature struct. Biology* **1**, 706–716 (1994).
4. Predki, P.F. & Regan, L. Redesigning the topology of a four-helix-bundle protein: Monomeric Rop. *Biochemistry* **34**, 9834–9839 (1995).
5. Wright, P.E., Dyson, J.H. & Lerner, R.A. Conformation of peptide fragments of proteins in aqueous solution: Implications for initiation of protein folding. *Biochemistry* **27**, 7167–7175 (1988).
6. Milburn, P.J., Konishi, Y., Meinwald, Y.C. & Scheraga, H.A. Chain Reversals in model peptides: Studies of cystine-containing cyclic peptides I. Conformational free energies of cyclization of hexapeptides of sequence. Ac-Cys-X-Pro-Gly-Y-Cys-NHMe *J. Am. Chem. Soc.* **109**, 4486–4496 (1987).
7. Dyson, H.J., Rance, M., Houghten, R.A., Lerner, R.A. & Wright, P.E. Folding of immunogenic peptide fragments of proteins in water solution I. Sequence requirements for the formation of a reverse turn. *J. molec. Biol.* **201**, 161–200 (1988).
8. Fasman, G.D. *Prediction of protein structure and the principles of protein conformation* (Plenum, New York, 1989).
9. Muoz, V., Blanco, F.J. & Serrano, L. The hydrophobic-staple motif and a role for loop-residues in α-helix stability and protein folding. *Nature struct. Biology* **2**, 380–385 (1995).
10. Banner, D.W., Kokkinidis, M. & Tsernoglou, D. Structure of the ColE1 Rop protein at 1.7 Å resolution. *J. molec. Biol.* **196**, 657–675 (1987).
11. Polisky, B. Col-E-1 replication control circuitry sense from antisense. *Cell* **55**, 929–932 (1988).
12. Predki, P.F., Nayak, L.M., Gottlieb, M.B.C. & Regan, L. Dissecting RNA-protein interactions: RNA-RNA recognition by Rop. *Cell* **80**, 41–50 (1995).
13. O-Neil, K.T., & DeGrado, W.F. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* **250**, 646–651 (1990).
14. Horovitz, A., Matthews, J.M. & Fersht, A.R. α-helix stability in proteins II. Factors that influence stability at an internal position. *J. molec. Biol.* **227**, 560–568 (1992).
15. Blaber, M. *et al.* Determination of α-helix propensity within the context of a folded protein: Sites 44 and 131 in bacteriophage T4 lysozyme. *J. molec. Biol.* **235**, 600–624 (1994).
16. Kim, C.A. & Berg. J.M. Thermodynamic β-sheet propensities measured using a zinc-finger host peptide. *Nature* **362**, 267–270 (1993).
17. Minor, D.L. Jr. & Kim. P.S. Measurement of the β-sheet-forming propensities of amino acids. *Nature* **367**, 660–663 (1994).
18. Smith, C.K., Withka, J.M. & Regan. L. A thermodynamic scale for the β-sheet forming tendencies of the amino acids. *Biochemistry* **33**, 5510–5517 (1994).
19. Minor, D.L. Jr., & Kim, P.S. Context is a major determinant of β-sheet propensity. *Nature* **371**, 264–267 (1994).
20. Smith, C.K. & Regan, L. Guidelines for protein design: energetics of β-sheet side-chain interactions. *Science* **270**, 980–982 (1995).
21. Ramachandran, G.N. & Sasisekharan, V. Conformation of polypepties and proteins. *Adv. Prot. Chem.* **23**, 283–437 (1968).
22. Munoz, V., & Serrano, L. Intrinsic secondary structure propensities of the amino acids, using statistical phi-psi matrices: Comparison with experimental scales. *Proteins* **20**, 301–311 (1994).
23. Munson, M., O'Brien, R., Sturtevant, J.M. & Regan, L. Redesigning the hydrophobic core of a four-helix-bundle protein. *Prot. Sci.* **3**, 2015–2022 (1994).
24. Vriend, G. WHAT IF: A molecular modelling and drug design package. *J. molec. Graphics* **8**, 52–56 (1990).
25. DeLano, W.L. & Brünger, A.T. The direct rotation function: Rotational patterson correlation search applied to molecular replacement. *Acta Crystallogr.* D, **in the press**.
26. Brünger, A.T. Extension of molecular replacement: A new search strategy based on patterson correlation refinement. *Acta Crystallogr.* A**46**, 46–57 (1990).
27. Fuginaga, M. & Read, R.J. Experiences with a new translation-function program. *J. appl. Crystallogr.* **20**, 517–521 (1987).
28. Brünger, A.T. X-PLOR. Version 3.1 A System for X-ray crystallography and NMR. Yale University Press, New Haven, CT (1992).
29. Hodel, A., Kim, S.-H. & Brnger, A.T. Model bias in macromolecular crystal structures. *Acta Crystallogr.* A**48**, 851–859 (1992).
30. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* A**47**, 110–119 (1991).
31. Brünger, A.T., Krukowski, A. & Erickson, J. Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr* A**46**, 585 (1990).
32. Nicholls, A., Sharp, K.A., & Honig, B. Protein folding and association. Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 282–293 (1991).