

## The role of turns in the structure of an $\alpha$ -helical protein

Adam P. Brunet\*, Enoch S. Huang†, Mary Elizabeth Huffine†, Jonathan E. Loeb†, Robert J. Weltman\* & Michael H. Hecht\*†§

Departments of \*Chemistry and †Molecular Biology, Princeton University, Princeton, New Jersey 08544-1009, USA

THE turns joining segments of secondary structure have been proposed to be key elements in dictating the folded structures of native proteins<sup>1-9</sup>. An alternative view assumes that turns play a passive role and are merely default structures that occur as a consequence of interactions between antiparallel segments of secondary structure, with chain reversal being dictated by the context surrounding the turn and not by the sequence of the turn itself<sup>10,11</sup>. The solvent-exposure of turns and their tolerance to evolutionary variance suggests that they may have little or no effect on the formation of native structures. Previous investigations have focused on various types of  $\beta$ -turns that connect antiparallel  $\beta$ -strands<sup>1-3,12,13</sup>, with comparatively little reported on the structural role of interhelical turns. Here we probe the structural importance of such a turn in an antiparallel 4-helix bundle by randomly substituting an interhelical tripeptide in cytochrome b-562 with many different amino-acid sequences. Thirty-one of the resulting substituted proteins were characterized and all of them were shown to fold into stable, native-like structures. These results suggest that this interhelical turn does not play a dominant role in determining the folded structure of this antiparallel 4-helix bundle.

Cytochrome b-562 is a small ( $M_r$  12,000) haem protein found in the periplasm of *Escherichia coli*. In the crystal structure (1.4 Å at  $R=0.164$ ) of this four-helix bundle<sup>14,15</sup>, the third and fourth helices are joined by the tripeptide Glu<sub>81</sub>-Gly<sub>82</sub>-Lys<sub>83</sub> (Fig. 1). By substituting this interhelical turn with randomly generated tripeptides, it is possible to construct 8,000 ( $20^3$ ) possible turn sequences. Among these sequences, those that can form turns compatible with the correctly folded structure can be distinguished by a simple colour assay: cells expressing mutants that fold correctly will bind haem and yield bright red periplasmic extracts (Fig. 2), whereas cells expressing mutant proteins with turns incompatible with a native-like structure will fail to bind haem and yield colourless extracts.

As the Glu<sub>81</sub>-Gly<sub>82</sub>-Lys<sub>83</sub> turn sequence is located distal to the haem site, we would expect haem binding to be affected only if local perturbations in the turn disrupt the overall native structure. If the sequence of the wild-type turn is essential in directing the 4-helix bundle into its native structure, most of the 8,000 possible sequences will prevent proper folding and produce colourless extracts. If, however, local interactions in the turn do not exert a dominant influence on the structure of the protein (that is, if the turn is passive) then most, perhaps all, of the substitutions should yield bright red samples.

We constructed a library of cytochrome mutants using cassette replacement mutagenesis<sup>16-18</sup> to randomize codons 81-83. The resulting library encodes all possible tripeptide turn sequences. From this library, 45 independent clones were characterized by DNA sequence analysis and by the red/white

§To whom correspondence should be addressed.

colour assay (Table 1). Of these, 9 contain termination codons and 31 code for mutant tripeptide sequences (the remaining 5 contain deletions). Of those mutants containing alternative tripeptides in place of the wild-type turn, all 31 sequences give rise to red extracts (Table 1; Fig. 2). The only members of the collection that yield colourless extracts are those encoding nonsense or deletion mutants. These results demonstrate that all the randomly generated tripeptide turn sequences shown in Table 1 are compatible with a stable native-like structure capable of binding haem.

To extend the results of our visual colour assay, mutant proteins with a variety of different turn sequences were purified and characterized. The secondary structures of these proteins were probed by circular dichroism (CD) spectroscopy<sup>19</sup>. Figure 3a shows that the CD spectrum of a mutant protein (Pro<sub>81</sub>-Val<sub>82</sub>-Ala<sub>83</sub>) is superimposable, within experimental error, upon the

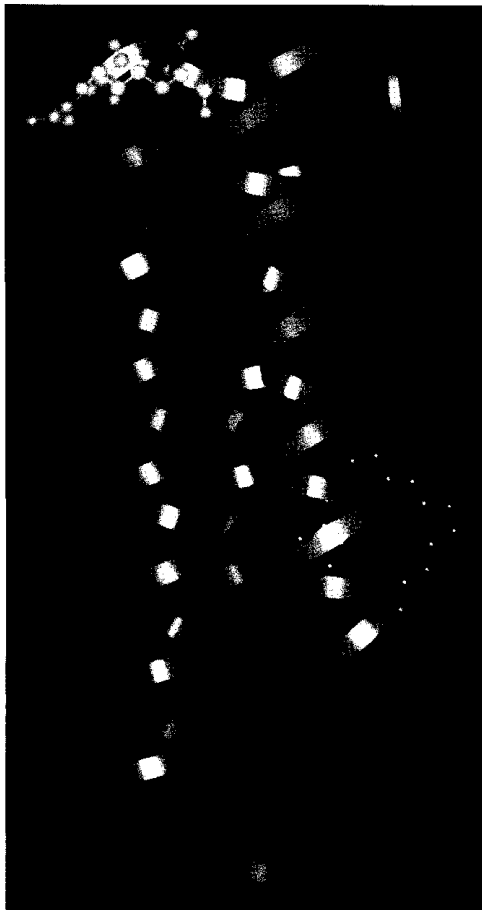


FIG. 1 Ribbon diagram of cytochrome b-562. The haem is shown in red; the sequence Glu<sub>81</sub>-Gly<sub>82</sub>-Lys<sub>83</sub>, which forms an interhelical turn connecting  $\alpha$ -helices 3 and 4, is shown in yellow. The  $\phi/\psi$  angles for these residues are  $-93.7/8.4$ ,  $78.6/9.9$  and  $-93.0/70.9$  for residues 81, 82 and 83 respectively. Coordinates are taken from file 256b in the Brookhaven Data bank<sup>34</sup>.

**METHODS.** A 53-mer having the sequence 5'-GCGGGCGGAAGCTTGC-AAATNNNNNNNGTAAAGAAGCGCAGGCTGCTGCA-3' (N represents an equimolar ratio of A, G, C and T) was annealed with a primer oligonucleotide 5'-GCAGCCTGCGCTTCT-3'. Synthesis of the complementary strand was accomplished using DNA polymerase (Sequenase version 2.0; US Biochemical). A *Pst*I sticky end was present after annealing and synthesis, and a *Hind*III sticky end was generated by restriction digestion. This semi-random cassette was then ligated into a newly constructed vector, pRW-2. This plasmid is a derivative of pEMBL-18 (ref. 35), into which the cytochrome b-562 gene<sup>36</sup>, including its natural signal sequence, is expressed under *lac* control. The *Pst*I site occurs naturally in the wild-type gene for cytochrome b-562. The *Hind*III site was introduced by changing the Leu<sub>78</sub> codon

spectrum of the wild-type protein (Glu<sub>81</sub>-Gly<sub>82</sub>-Lys<sub>83</sub>). Similar results were obtained for proteins with the turn sequences Ile<sub>81</sub>-Asp<sub>82</sub>-Leu<sub>83</sub>, Tyr<sub>81</sub>-Lys<sub>82</sub>-Leu<sub>83</sub>, and Ser<sub>81</sub>-Leu<sub>82</sub>-Ser<sub>83</sub> (data not shown). Calculations of the per cent helix (using PROSEC software; Aviv) for each of these proteins indicates a helicity within 3% of the value for the wild-type protein.

The structures of the mutant and wild-type proteins were also compared using absorption spectroscopy to probe the tertiary structure of the haem binding site. As the absorption spectrum of haem is extremely sensitive to its electronic environment, the spectrum of cytochrome b-562 tests whether varying the interhelical turn disrupts the overall protein structure<sup>20,21</sup>. As the spectrum of cytochrome b-562 is affected by the oxidation state of the iron<sup>21</sup>, spectra were measured for both the oxidized and reduced forms; both forms of the Ser<sub>81</sub>-Leu<sub>82</sub>-Ser<sub>83</sub> mutant are compared to those of the wild-type protein in Fig. 3 b and c.

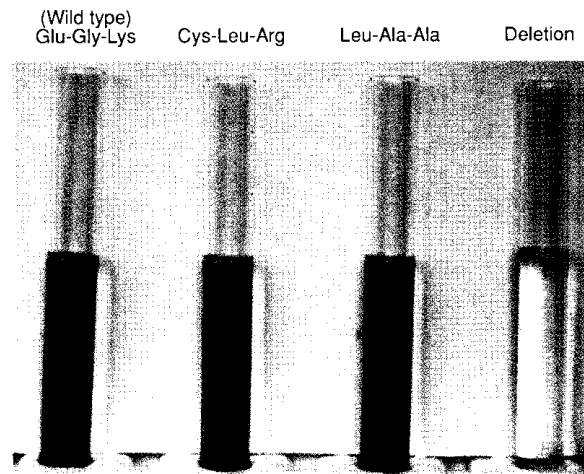


FIG. 2 Periplasmic fractions from *E. coli* cells expressing high levels of wild-type or mutant forms of cytochrome b-562. The red colour results from the accumulation of native folded haem-protein. The tube on the right represents a control experiment showing the periplasmic fraction from cells harbouring plasmid pRW-2, which contains a large deletion and a frameshift at the end of  $\alpha$ -helix 3.

**METHODS.** Cultures of *E. coli* strain MV1190 harbouring the appropriate sequence derivative of the overexpressing plasmid were grown in 2 X YT medium to  $A_{600}=1$ . IPTG was added to a concentration of  $100 \mu\text{g ml}^{-1}$  and growth was continued overnight. Cells were collected by centrifugation then washed in 50 mM Tris HCl, pH 8.0, 200 mM NaCl, and centrifuged again. The cell pellet was subjected to four cycles of freezing and thawing to disrupt the outer membrane, then taken up in water to release the periplasmic contents. Spheroplasts were removed by centrifugation, yielding a bright red supernatant with the characteristic absorption peaks of reduced cytochrome b-562 at 426 nm (the Soret band) and 562 nm. This solution was further enriched for cytochrome b-562 by adjusting the pH to 4.2 with sodium acetate (final concentration, 50 mM) and removing acid-insoluble contaminants by centrifugation. Extracts from cells harbouring the control plasmid pRW2 are colourless, indicating that background levels of cytochrome are negligible.

from CTG to CTT by site-directed mutagenesis<sup>37</sup>. But in pRW-2, the wild-type *Hind*III-*Pst*I fragment (encoding amino acids 77-91) has been replaced by a 20-bp 'dummy' fragment. In this fragment a frameshift is introduced, and a new restriction site (*Sal*I) is included. Use of the pRW-2 vector for constructing our library of turns has two advantages: (1) the true wild-type sequence does not enter the system except as one of the 8,000 possible sequences in the random library; and (2) after ligating the semi-random synthetic restriction fragment into the pRW-2 backbone, the ligation mix can be cut with *Sal*I to destroy any of the pRW-2 vector that may have re-closed without receiving a new synthetic piece.

Similar results were obtained for proteins with the turn sequences Pro<sub>81</sub>-Val<sub>82</sub>-Ala<sub>83</sub>, Ile<sub>81</sub>-Asp<sub>82</sub>-Leu<sub>83</sub> and Tyr<sub>81</sub>-Lys<sub>82</sub>-Leu<sub>83</sub> (data not shown). In both oxidation states, the spectra of all four mutants are superimposable, within experimental error, upon the spectrum of the wild-type protein.

These results show that replacing the wild-type turn by a different sequence does not significantly perturb the structure of cytochrome b-562. Although minor changes in local structure

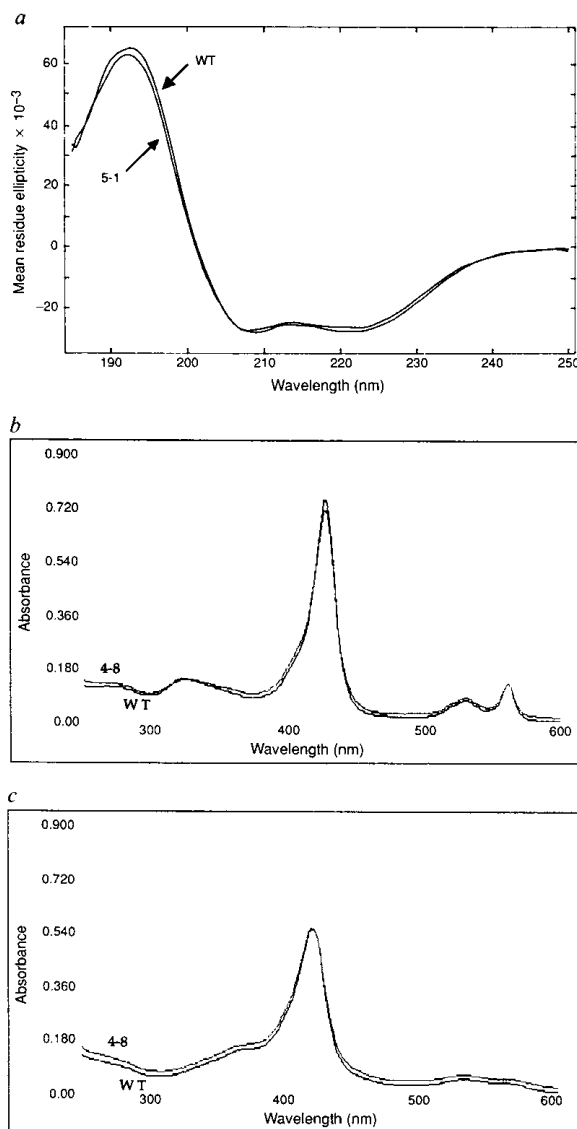


FIG. 3 *a*, Circular dichroism spectra of wild-type and mutant cytochrome b-562. The wild-type turn sequence is Glu<sub>81</sub>-Gly<sub>82</sub>-Lys<sub>83</sub>; the mutant (number 5-1) turn sequence is Pro<sub>81</sub>-Val<sub>82</sub>-Ala<sub>83</sub>. Proteins were purified by ion-exchange chromatography over an S-Sepharose fast-flow column (Pharmacia) using a salt gradient from 0–1.0 M NaCl in a buffer containing 50 mM sodium acetate, pH 4.5. Before CD spectroscopy, proteins were dialysed into 10 mM sodium phosphate, pH 7.5, 50 mM NaF. Protein solutions were 5  $\mu$ M and optical path was 1 mm; spectra were recorded using an Aviv model 62DS instrument. The data were collected in triplicate from 260 nm to 185 nm with a step size of 0.5 nm. Similar results were obtained for the mutants Ile<sub>81</sub>-Asp<sub>82</sub>-Leu<sub>83</sub>, Tyr<sub>81</sub>-Lys<sub>82</sub>-Leu<sub>83</sub> and Ser<sub>81</sub>-Leu<sub>82</sub>-Ser<sub>83</sub> (data not shown). *b* and *c*, Absorption spectra of reduced (*b*) and oxidized (*c*) forms of wild-type and mutant cytochrome b-562. The mutant 4-8 turn sequence is Ser<sub>81</sub>-Leu<sub>82</sub>-Ser<sub>83</sub>. Conditions as in *a*, except that optical path length was 1 cm and spectra were recorded in a Hewlett-Packard 8452A diode array spectrophotometer. Dithionite (final concentration 0.5  $\mu$ g ml<sup>-1</sup>) was used as reducing agent. Similar results were for mutant Ile<sub>81</sub>-Asp<sub>82</sub>-Leu<sub>83</sub>, Tyr<sub>81</sub>-Lys<sub>82</sub>-Leu<sub>83</sub> and Pro<sub>81</sub>-Val<sub>82</sub>-Ala<sub>83</sub> (data not shown).

TABLE 1 Mutations in the third interhelical turn of cytochrome b-562

#### Red mutants

Mutant	DNA sequence*	Amino-acid sequence
Wild type	GAA.GGT.AAA	Glu - Gly - Lys
1-2	AGT.TCT.AGG	Ser - Ser - Arg
1-10	TTG.TGG.CGA	Leu - Trp - Arg
2-5	AAC.CTG.CGC	Asn - Leu - Arg
2-10	CGT.ACT.AAA	Arg - Thr - Lys
3-7	TTC.TTC.AGT	Phe - Phe - Ser
3-9	CTA.TTT.ATG	Leu - Phe - Met
3-10	AGA.GGA.ATG	Arg - Gly - Met
4-3	ATT.TCT.GCG	Ile - Ser - Ala
4-4	TAT.CTT.CAC	Tyr - Leu - His
4-5	GAC.TTT.CGT	Asp - Phe - Arg
4-6	TGC.CGG.GCG	Cys - Arg - Ala
4-7	TGC.TTC.ATG	Cys - Phe - Met
4-8	TCG.CTG.TCC	Ser - Leu - Ser
4-10	TTG.GGA.CGG	Leu - Gly - Arg
5-1	CCC.GTC.GCA	Pro - Val - Ala
5-3	TGT.TTG.AGA	Cys - Leu - Arg
5-5	AGG.GGA.GTT	Arg - Gly - Val
5-7	CTT.AAC.GCG	Leu - Asn - Ala
5-9	ATT.GAC.CTT	Ile - Asp - Leu
5-10	TAT.TAC.GAC	Tyr - Tyr - Asp
6-1	AGT.AGG.GAG	Ser - Arg - Glu
6-2	TTA.AGC.ATA	Leu - Ser - Ile
6-4	TCA.TGT.TTC	Ser - Cys - Phe
6-5	CTG.GCT.GCG	Leu - Ala - Ala
6-9	TAC.AAG.TTA	Tyr - Lys - Leu
6-10	TTA.TGG.CTA	Leu - Trp - Leu
7-5	TTT.GTT.AAC	Phe - Val - Asn
7-7	GCG.GTA.AAG	Ala - Val - Lys
7-9	AGC.ATG.CGC	Ser - Met - Arg
7-10	GAA.GAT.TTT	Glu - Asp - Phe
8-7	CAT.GAA.ACT	His - Glu - Thr

#### White mutants (stop codons)

1-3	TAG.TTA.TGT	STOP - Leu - Cys
1-5	TAG.AGC.GTG	STOP - Ser - Val
3-8	GTG.TAG.AGG	Val - STOP - Arg
5-2	AAG.TAG.TAT	Lys - STOP - Tyr
6-3	TGC.TAG.GTC	Cys - STOP - Val
6-7	CCG.TAA.TGG	Pro - STOP - Trp
7-1	CGT.TGA.GGA	Arg - STOP - Gly
9-3	GTA.TGA.TCT	Val - STOP - Ser
9-9	TTT.GTA.TAG	Phe - Val - STOP

Not listed are five deletion mutants that introduced frameshifts and were therefore discarded. Phagemid DNA was prepared by standard techniques<sup>32</sup> using phage M13-K07 as helper phage. The sequencing primer is a synthetic oligonucleotide with the sequence 5'-AGTTCCTGTTACTGAGTGC-3'. DNA was sequenced by the dideoxy method<sup>33</sup> using single-stranded phagemid DNA as template. We have analysed 32 randomly chosen sequences from among the 8,000 possible tripeptides. The probability (*P*) of observing the native-like fold in all 32 of them is

$$P = \prod_{i=0}^{31} (R-i)/(8,000-i)$$

where *R* is the total number of tolerated tripeptides in the collection of 8,000. In this formula, if *R* is less than 7,286 (if less than 91% of the turn sequences give rise to native-like conformations), then the probability of randomly choosing 32/32 tolerated sequences would have been <5% (ref. 22). Thus we can say with 95% certainty that >91% of all possible tripeptides are tolerated.

around the turn are possible, the overall secondary structure of the 4-helix bundle and the specific tertiary structure around the haem must be similar in the wild-type and mutant proteins.

Our results indicate that the turn connecting helices 3 and 4 in cytochrome b-562 is not crucial in determining the folded structure of this protein. Although we cannot exclude the possibility that among 8,000 possible tripeptide sequences there exist some that prevent formation of the correct native structure, our results suggest that such sequences must be rare. A statistical analysis indicates with 95% certainty that >91% of all possible tripeptides are tolerated<sup>22</sup> (see Table 1 legend).

It is important to consider the impact of these mutations on both the kinetics of protein folding and the thermodynamics of protein stability. Our experiments cannot rule out the possibility that some of the mutant proteins may fold at rates different from the wild type, but they do show that all 31 mutants fold sufficiently rapidly to escape proteolytic degradation *in vivo*; this indicates that for cytochrome b-562 there is no obligatory folding pathway requiring a specific tripeptide in the turn connecting helices 3 and 4.

For a turn whose side chains are exposed to solvent, sequence changes might be expected to influence stability either by introducing conformational strain in the folded state<sup>23</sup> or by affecting interactions with the solvent (in either the folded or unfolded state)<sup>24,25</sup>. For example, replacement of the wild-type Gly 82 might strain the native structure, because the  $\phi$  and  $\psi$  angles of this glycine (79° and 10°) fall in a region of the Ramachandran plot<sup>26</sup> rarely observed for other residues<sup>27</sup>. However, substitution of this glycine by virtually any other amino acid can be tolerated (Table 1), indicating that the region around residue 82 can relax to accommodate turn conformations different from the wild-type.

Replacement of Glu 81 or Lys 83 might be expected to destabilize the protein by disrupting a favourable electrostatic interaction between these residues. But it is unlikely that such an interaction contributes to stability because these side chains point in opposite directions in the crystal structure. Neither does the interaction of these charged residues with the solvent contribute to the stability of the protein, because many different side chains can be tolerated at positions 81 and 83 (Table 1). Even in the extreme case in which the substituents are Ile 81 and Leu 83 the protein is stable and its  $T_m$  only slightly reduced (A.P.B., E.S.H. and M.H.H., manuscript in preparation).

We conclude that individual interhelical turns play only a minor part in determining protein structure. Although our mutant tripeptides undoubtedly differ in their intrinsic tendencies to form particular structures in solution, they can all be constrained to adopt conformations consistent with a properly folded structure. This finding argues that local effects can be overcome by long-range interactions between residues not adjacent in sequence. The tolerance of interhelical turns to alterations in sequence, even of different hydrophobicity patterns, has implications not only for our understanding of natural proteins, but also for the design of novel proteins<sup>28-31</sup> and for designing active sites into engineered proteins. □

14. Mathews, F. S., Bethge, P. H. & Czerwinski, E. W. *J. biol. Chem.* **254**, 1699–1706 (1979).
15. Lederer, F., Glatigny, A., Bethge, P. H., Bellamy, H. D. & Mathews, F. S. *J. molec. Biol.* **148**, 427–448 (1981).
16. Eisenbeis, S. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 1084–1088 (1985).
17. Wells, J. A., Vasser, M. & Powers, D. B. *Gene* **34**, 315–323 (1985).
18. Reidhaar-Olson, J. F. *et al. Meth. Enzym.* **208**, 564–586 (1991).
19. Greenfield, N. & Fasman, G. D. *Biochemistry* **8**, 4108–4116 (1969).
20. Itagaki, E. & Hager, L. P. *J. biol. Chem.* **241**, 3687–3695 (1966).
21. Fisher, M. T. *Biochemistry* **30**, 10013–10018 (1991).
22. Young, H. D. *Statistical Treatment Of Experimental Data* Ch.2 (McGraw-Hill, New York, 1962).
23. Hertzberg, O. & Moul, J. *Proteins: Structure, Function and Genetics* **11**, 223–229 (1990).
24. Shortle, D. & Meeker, A. M. *Biochemistry* **28**, 936–945 (1989).
25. Dill, K. A. & Shortle, D. *Rev. Biochem.* **60**, 795–825 (1991).
26. Ramakrishnan, C. & Ramachandran, G. N. *Biophys. J.* **5**, 909–933 (1965).
27. Richardson, J. S. & Richardson, D. C. in *Prediction of Protein Structure and the Principles of Protein Conformation* (ed. Fasman, G.) 1–98 (Plenum, New York, 1989).
28. Regan, L. & DeGrado, W. F. *Science* **241**, 976–978 (1988).
29. Fedorov, A. N. *et al. J. molec. Biol.* **225**, 927–931 (1992).
30. Hecht, M. H., Richardson, J. S., Richardson, D. C. & Ogden, R. C. *Science* **249**, 884–891 (1990).
31. Hahn, K. W., Klis, W. A. & Stewart, J. M. *Science* **248**, 1544–1547 (1990).
32. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1990).
33. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
34. Bernstein, F. *et al. J. molec. Biol.* **112**, 535–542 (1977).
35. Dent, L. & Cortese, R. *Meth. Enzym.* **155**, 111–119 (1987).
36. Nikkila, H., Gennis, R. & Sligar, S. G. *Eur. J. Biochem.* **202**, 309–313 (1991).
37. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. *Meth. Enzym.* **154**, 367–382 (1987).

ACKNOWLEDGEMENTS. We thank H. Nikkila and S. Sligar for the cytochrome b-562 gene; K. Gernert for help with initial plasmid constructions; B. Johnson for advice on protein purifications; J. Ybe for assistance with CD; and J. A. Swan and S. Subramaniam for advice on the manuscript. Oligonucleotides were prepared by the Princeton Molecular Biology Synthesis facility under the direction of M. Flocco. A.P.B. was supported by an NIH training grant; E.S.H., M.E.H., J.E.L. and R.J.W. received support from the Princeton Undergraduate Summer Research Program in the Departments of Chemistry and Molecular Biology. This work was supported by grants from the ONR and the NIH to M.H.H.

Received 17 December 1992; accepted 26 April 1993.

1. Chou, P. Y. & Fasman, G. D. *Biochemistry* **13**, 222–245 (1974).
2. Fasman, G. D. *Prediction of Protein Structure and the Principles of Protein Conformation* (Plenum, New York, 1989).
3. Sibanda, B. L., Blundell, T. L. & Thornton, J. M. *J. molec. Biol.* **206**, 759–777 (1989).
4. Skolnick, J. & Kolinski, A. *J. molec. Biol.* **212**, 787–817 (1990).
5. Sikorski, A. & Skolnick, J. *J. molec. Biol.* **212**, 819–836 (1990).
6. Chou, K. C., Maggiora, G. M. & Scheraga, H. A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7315–7319 (1992).
7. Dyson, J. H., Rance, M., Houghton, R. A., Lerner, R. A. & Wright, P. E. *J. molec. Biol.* **201**, 161–200 (1988).
8. Wright, P. E., Dyson, J. H. & Lerner, R. A. *Biochemistry* **27**, 7167–7175 (1988).
9. Milburn, P. J., Meinwald, Y. C., Takahashi, S., Ooi, T. & Scheraga, H. A. *Int. J. Pep. Prot. Res.* **31**, 311–321 (1988).
10. Kabsch, W. & Sander, C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1075–1078 (1984).
11. Argos, P. *J. molec. Biol.* **197**, 331–348 (1987).
12. Rose, G. D., Gierasch, L. M. & Smith, J. A. *Adv. Prot. Chem.* **37**, 1–109 (1985).
13. Hynes, T. R., Kautz, R. A., Goodman, M. A., Gill, J. F. & Fox, R. O. *Nature* **339**, 73–76 (1989).