31. Kabsch, W. & Sander, C. *Biopolymers* **22,** 2577–2637 (1983).
32. Kabsch, W. *Acta crystallogr.* A**34,** 827–828 (1978).
33. Mattevi, A., Schierbeek, A. J. & Hol, W. G. J. *J. molec. Biol.* (in the press).
34. Kuriyan, J. *et al. Nature* **352,** 172–174 (1991)..
35. Karplus, P. A. & Schulz, G. E. *J. molec. Biol.* **210,** 163–180 (1989)

ACKNOWLEDGEMENTS. We thank K. Fritz-Wolf for help in data collection, G. A. Petsko for discussions and K. C. Holmes for support. This work was supported by the Deutsche Forschungsgemeinschaft and the National Institutes of Health.

Convergent evolution of similar function in two structurally divergent enzymes

John Kuriyan*†, T. S. R. Krishna*†, Lim Wong*, Brian Guenther*, Arno Pahler‡§, Charles H. Williams Jr‖ & Peter Model*

* The Rockefeller University, 1230 York Avenue, New York 10021, USA
† Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York 10021, USA
‡ Department of Molecular Biophysics and Biochemistry, Columbia University, 630 W 168th Street, New York 10032, USA
‖ Department of Veterans Affairs Medical Center and Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48105, USA
§ Present address: First Department, Protein Engineering Research Institute, 6-2-3 Furuedai, Suita, Osaka 565, Japan

An example of two related enzymes that catalyse similar reactions but possess different active sites is provided by comparing the structure of *Escherichia coli* thioredoxin reductase with glutathione reductase[1]. Both are dimeric enzymes that catalyse the reduction of disulphides by pyridine nucleotides through an enzyme disulphide and a flavin[2]. Human glutathione reductase contains four structural domains within each molecule: the flavin–adenine dinucleotide (FAD)- and nicotinamide–adenine dinucleotide phosphate (NADPH)-binding domains, the 'central' domain and the C-terminal domain that provides the dimer interface and part of the active site[3,4]. Although both enzymes share the same catalytic mechanism and similar tertiary structures, their active sites do not resemble each other[5,6]. We have determined the crystal structure of *E. coli* thioredoxin reductase at 2 Å resolution, and show that thioredoxin reductase lacks the domain that provides the dimer interface in glutathione reductase, and forms a completely different dimeric structure. The catalytically active disulphides are located in different domains on opposite sides of the flavin ring system. This suggests that these enzymes diverged from an ancestral nucleotide-binding protein and acquired their disulphide reductase activities independently.

Thioredoxins are small redox-active proteins that have diverse functions. In *E. coli*, thioredoxin is reduced by NADPH-dependent thioredoxin reductase (TR; ref. 7). The three-dimensional structure of TR was determined by multiple isomorphous replacement (Fig. 1) and consists of two molecules which interact closely to form a symmetric dimer. Each molecule has three clearly delineated domains, which correspond to the FAD, NADPH and central domains of glutathione reductase (GR) (Fig. 2). The tertiary structure of TR is very similar to that of the first three domains of GR, the chief differences being a change in the orientation of the NADPH domain and the deletion of a helix (Fig. 2). Nineteen of the 21 β strands in TR correspond to elements in GR; the other two are located at the N terminus and near the redox-active disulphide that is a distinct feature of the TR sequence (Fig. 3). Likewise, five of the seven α helices and three of the four 3_{10} helices in TR align with corresponding elements in GR. A notable feature of the structural alignment is the relative displacement of the NADPH domains in the two enzymes. This corresponds to a rotation of one of these domains by 66° about the two β strands that connect it to the FAD and central domains, leaving the backbone structure relatively unperturbed (Fig. 2). On aligning the FAD/central domains and the NADPH domain separately, the root-mean-square (r.m.s.) deviation in C_α positions is 1.8 Å for the FAD/central domains and 2.0 Å for the NADPH domain for core residues comprising about 70% of the TR sequence. Including in the alignment only those C_α atoms that deviate by less than 2.5 Å results in r.m.s. deviations of 1.1 Å both in the FAD/central domains and in the NADPH domain, with 54% and 65% of the TR residues satisfying this criterion in the two parts of the structure, respectively. These results are similar to

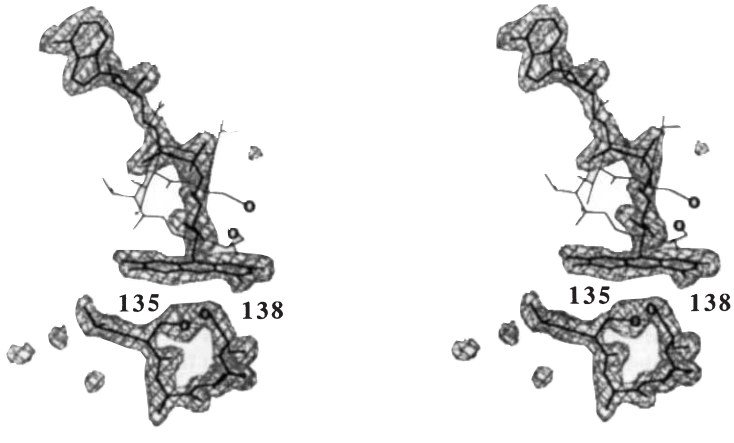

FIG. 1 Stereo diagram of electron density at 2.0 Å resolution for the FAD group and redox-active disulphide of mutant TR (Cys 138 to Ser). The model for the FAD group and residues 134–138 in TR are shown in thick lines, with the sulphur and oxygen side-chain atoms of residues 135 and 138 indicated with open circles. The N^{10} atom of the flavin ring system is at the bottom. The model for residues 57–63 in GR[1] is also shown in thin lines. The C_α atoms of the GR FAD/central domains were optimally superimposed on TR (Figs 2 and 3). This transformation results in the FAD groups of the two proteins being closely aligned (r.m.s. deviation, 0.8 Å). The resulting orientation of the active-site residues 57–63 in GR are shown in thin lines, with the redox-active sulphurs of residues 58 and 63 indicated by open circles. The electron density was generated using coefficients $(|F_o| - |F_c|) \exp(i\alpha_c)$, where F_o and F_c are the observed and calculated structure factors and α_c is the calculated phase, with the FAD group and residues 134–138 omitted from the phase calculation. The final refined structure (Fig. 2) was used for the calculation. Contour levels corresponding to 3.5 s.d. above the mean are shown. Crystals of the (Cys 135 to Ser) mutant protein[16], space group $P6_322$, with $a = b = 123.7$ Å and $c = 81.6$ Å (ref. 17), were soaked for 16–20 h in 40% polyethylene glycol ($M_r = 3{,}500$) with 0.2 M lithium sulphate and either 2 mM ethyl mercury phosphate (EMP) or 0.2 mM 1,4-diacetoxy mercury-2,3-dimethoxybutane (Baker's dimercurial). An Enraf-Nonius FAST area detector system mounted on an Elliot GX-21 rotating anode X-ray generator was used to collect data sets to 3 Å resolution, using only one each of the native and derivative crystals. The resulting native data set is 95% complete to 3 Å resolution, and has an overall $R_{sym}(I)$ of 5%. Five mercury sites for the EMP derivative and two sets of paired sites for the Baker's dimercurial were located by Patterson map and difference Fourier analysis. Refinement of heavy atom parameters[18] resulted in phasing powers, f_H/E, of 1.37 and 1.27 for EMP and Baker's dimercurial, respectively, including anomalous scattering. The phases were further improved by solvent flattening[19], increasing the mean figure of merit from 0.59 to 0.85. These phases were used to calculate a map at 3 Å resolution, from which the main chain was readily traced. Refinement at 3 Å resolution was carried out by restrained molecular dynamics using the programs X-PLOR[20,21] and FRODO[22]. The resulting structure was used directly to initiate refinement at 2 Å resolution, using data for the (Cys 138 to Ser) protein (Fig. 2).

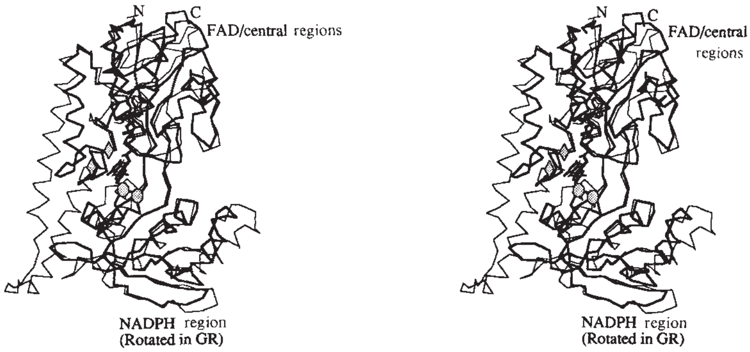

FIG. 2 Stereo diagram of the C_α atoms of TR and the FAD/central domains and the rotated NADPH domain of GR. Thick lines, TR C_α atoms (residues 1–316) and all nonhydrogen atoms of the FAD group. The last four residues are not visible in the electron density map. The active-site cysteines are indicated by shaded circles, below the flavin ring system. Thin lines, GR FAD (residues 18–158), central (291–356) and NADPH (159–290) domains, and the FAD group[1]. The active-site cysteines are indicated by shaded diamonds, above the flavin. The FAD/central (taken together) and the NADPH domains of GR were superimposed on the corresponding domains of TR by rigid-body rotations and translations (Fig. 3). Once the FAD/central domains of the two proteins were brought into alignment, the NADPH domain of GR was rotated by 66° as a rigid body in order to align it with the NADPH domain of TR. This motion can be described as a rotation about the two-stranded β-sheet separating the regions, with atoms in the hinge region (residues 160 and 289) shifting only by ~1.5 Å. The helix formed by residues 63–80 in GR is missing in TR. The structural model of TR is for a (Cys 138 to Ser) mutant. Data to 2 Å resolution were measured at Beamline X12-C, Brookhaven National Laboratory (in collaboration with R. M. Sweet). The current model for this mutant includes 2,447 non-hydrogen protein and FAD atoms and 217 water molecules, with individual temperature factors refined for all atoms. The R-factor is 17.7% for data between 6 Å and 2.0 Å, including 21,106 observed structure factors with amplitudes greater than 2σ, which corresponds to 93% of the unique data. The deviations from ideal bond lengths and angles are 0.0019 Å and 3.2°, respectively, and only three non-glycine residues have backbone dihedral angle values in disallowed regions of the Ramachandran diagram[23]. The structures of the two TR mutants are very similar, with r.m.s. deviations of backbone atoms of 0.1 Å and with the terminal sulphur and oxygen atoms of residues 135 and 138 in positions that are indistinguishable at this resolution. This suggests that these structures provide good models for that of the wild-type enzyme. Details of the structure and refinement at 2 Å resolution will be published separately (T.S.R.K., R. M. Sweet, C.H.W. and J.K., manuscript in preparation). The structural model for GR is that of Karplus and Schulz, refined at 1.54 Å resolution (Protein Databank Entry 3GRS)[1].

the deviations obtained on superimposing globin structures with comparable sequence identity[8].

The redox-active cysteines are in different locations in the two enzymes (Figs 1 and 2). But, both enzymes have sequence or structural similarities in the regions corresponding to the disulphide of the other enzyme. The sequence TSDGFF (single-letter amino-acid notation) in GR (residues 176–181) aligns with the active-site sequence TCDGFF in TR, and this hexapeptide is the one with greatest sequence similarity in the entire alignment (Fig. 3). At the site of the GR disulphide (residues 58–63), the first turn of the α helix is in structural alignment with a 3_{10} helix in TR (Figs 2 and 3). Together with the homology in secondary structural elements and the presence of several other clustered regions of relatively high sequence similarity, this

```
  EEEE EEEE    HHHHHHHHHHH      EEEEEE    333 33                                           HHHHHHHHHHHHHHH
GTTKHSKLLILGSGPAGYTAAVYAARANLQPVLITGMEKGGQL-TTT-------------------TEVENWPGDPNDL-----------TGPLLMERMHEHATKFETEIIF  81 TR
GAVASYDYLVIGGGSGGLASARRAAELGARAAVVESHKLGGTCVNVGCVPKKVMWNTAVHSEFMHDHADYGFPSCEGKFNWRVIKEKRDAYVSRLNAIYQNNLTKSHIEIIR 127 GR
      EEEE    HHHHHHHHHHHH     EEEEE       HHHHHH HHHHHHHHHHHHHHHHHH                   HHHHHHHHHHHHHHHHHHHHHHHHHH   EEEE

EEEEEEE    EEEEE   EEEEE EEEEE EEEEE        HHHHHH    EEE  HHHH3333        EEEEE  HHHHHHHHHHH      EEEEE      HHHHH
DHINKVDLQNRPFRLNGDNGEYTCDALIIATGASARYLGL---PSEEAFKGRGVSACATCDGFFYRN---QKVAVIGGGNTAVEEALYLSNIASEVHLIHRRDGFRAEKILI 187 TR
GHAAFTSDPK--PTIEVSGKKYTAPHILIATGGMPSTPHESQIPGASLGI--------TSDGFFQLEELPGRSVIVGAGYIAVEMAGILSALGSKTSLMIRHDKVLRSFDSM 229 GR
E  EEE          EEEE  EEEE   EEEEE EEEEE        33333          3333333    EEEEEEE HHHHHHHHHHHHHH  EEEEEE          HHH

HHHHHHHH     EEEEEEE EEEEEEEE  EEEEEEEEE          EEEEEE  EEEEEE EEEEE   3333     EE   EE           EEE  EEEEE 3333
KRLMDKVENG--NIILHTNRTLEEVTGDQMGVTGVRLRDTQN--SDNIESLD-VAGLFVAIGHSPNTA-IFEG--QLELE-NGYIKVQSGIHGNATQTSIPGVFAAGDVMDH 290 TR
ISTNCTEELENAGVEVLKFSQVKEVKKTL-SGLEVSMVTAVPGRLPVMTMIPDVDCLLWAIGRVPNTKDLSLNKLGIQTDDKGHIIVDE-----FQNTNVKGIYAVGDVCGK 335 GR
HHHHHHHHHHH   EEEEEE EEEEEEE   EEEE EEEE         EEEEEEEE EEEEE EEEEE   3333    EEE  EEEE           EEEE  EEEEE33333

   HHHHHHHHHHHHHHHHHHH
IYRQAITSAGTGCMAALDAERYLDGLADAK                                                                                   320 TR
-ALLTPVAIAAGRKLAHRLFEY-----------[ 121 RESIDUES OF GR INTERFACE DOMAIN NOT SHOWN ]                               356 GR
  HHHHHHHHHHHHHHHHH
```

FIG. 3 Sequence alignment of *E. coli* TR[6] and human erythrocyte GR[1] based on secondary and tertiary structure (single-letter amino-acid notation). A less extensive but similar alignment based on sequence alone has been reported[7]. Boundaries between the FAD, central and NADPH domains are shown by vertical lines; the N- and C-terminal segments constitute the FAD and central domains, with the NADPH domain in the middle. Vertical bars indicate sequence identity, and the active-site cysteines are boxed. The overall sequence identity is 21%. Secondary structure code: E, β sheet; H, α helix; 3, 3_{10} helix; the assignment for TR was done on the basis of hydrogen-bonding patterns[24], with somewhat relaxed criteria as described for GR[1]. Regions of tertiary structural similarity are shaded. The FAD/central and the NADPH domains are treated separately for the evaluation of this similarity, as their relative orientations differ (Fig. 2). A least-squares procedure was used to superimpose C_α atoms, initially by aligning the secondary structure elements. Insertions and deletions in the sequences were then determined by increasing the list of residue pairs used in the least-squares superpositioning and removing those that appeared to be structurally inequivalent. Although subjective, the results are unambiguous in most cases. Residues 47–59 of TR are underlined to indicate that they do not align with the residues below. Only four aligned segments are not considered to be structurally equivalent. All are helices that are present in both structures but play unique roles in TR and have changed their orientations substantially: residues 60–72, 183–195 and 305–320 (TR) are part of the dimer interface, and residues 135–14 include the redox-active cysteines. 137 residues in the FAD/central domains and 96 in the NADPH domain are considered structurally equivalent. This corresponds to 72% and 74% of the total in each part of the TR sequence, respectively, and the root-mean-square deviation from equivalent residues in GR is 1.8 Å and 2.0 Å, respectively.

suggests that both proteins have evolved from a common ancestor. The FAD and NADPH domains of TR and GR both contain the $\beta\alpha\beta$ nucleotide-binding motif[9] and an additional antiparallel β sheet, and may have arisen by gene duplication[10]. Previous comparison of GR and an FAD-containing monooxygenase, *p*-hydroxybenzoate hydroxylase, revealed structural similarity in their FAD domains and some common elements in their central domains[11]. Elements of divergent as well as convergent evolution were indicated, but distinguishing between the two mechanisms was difficult because of the absence of sequence identity, the lack of structural similarity in the NADPH domains and fundamental differences in the reactions catalysed[11]. By contrast, TR and GR catalyse very similar reactions that differ only in substrate, and share a sequence similarity showing relationships in all three domains of TR[6].

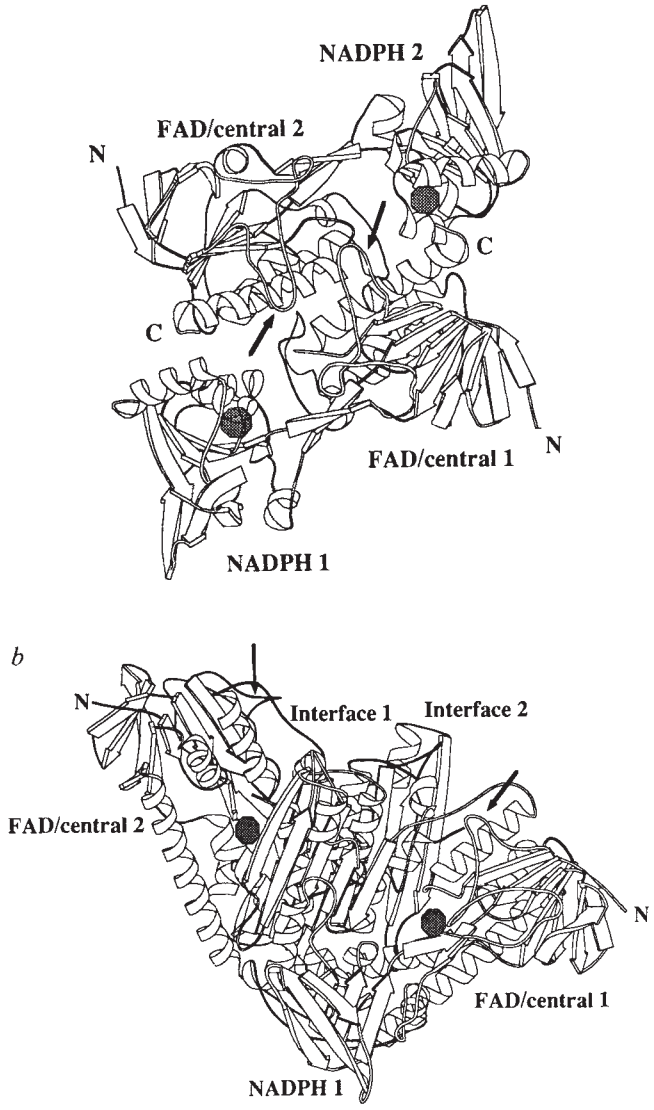

FIG. 4 Schematic diagrams of the dimeric structures of *a*, TR and *b*, GR (ref. 1). Helices are shown by ribbons, strands of β sheet by arrows[25]. Domains in different monomers within the dimers are labelled 1 and 2 and the FAD domains of monomer 1 are in the same orientation. One of the NADPH domains of GR is hidden. The approximate locations of the redox-active cysteines are shown by shaded circles. The interface domains, shown here for GR, have no counterpart in TR. The arrows indicate two symmetry related loops that in TR interact across the dimer interface and are widely separated in GR. The change in relative orientation of these loops indicates the large change in relative orientation of the monomers. The total surface area buried on dimer formation in TR is 2,950 Å^2, of which 39%, 44% and 17% are contributed by the FAD, central and NADPH domains, respectively. This is in contrast to GR, where the buried surface area is much larger (3,590 Å^2), and where the interface domain provides the largest buried surface area (57%), along with a smaller contribution[1] from the FAD domain (37%). The accessible surface areas were calculated using a probe of radius 1.6 Å (ref. 26).

In contrast to the similarity in tertiary structure, the dimeric forms of the enzyme are completely different (Fig. 4). Dimer formation in GR involves an interlocking of the interface domains, with each one making contacts with the interface and FAD domains of the other molecule and forming the deep crevices that are the binding sites for glutathione[1]. In TR, there are two shallow depressions at the dimer interface, on the same face of the enzyme. These are probably the thioredoxin binding sites, as the side chains of both reactive cysteines are exposed at each depression consistent with mechanistic studies[12]. Dimer formation in TR requires no additional structural elements other than those that are also present in GR (Fig. 4). The conformation of the FAD group and its position in the FAD/central domains is conserved between TR and GR (Fig. 2). Relative to the flavin ring system, the other components of the catalytic machinery are considerably rearranged: the redox-active disulphides are on opposite sides of the flavin ring system in the two enzymes, and the NADPH domains are rotated with respect to one another. Note that the redox active cysteines in TR stack against the flavin (Fig. 1) in a location that is coincident with that of the nicotinamide ring of NADPH in GR[13], indicating that the mode of pyridine nucleotide binding in TR is very different. In GR the redox-active cysteines and the catalytic base (His 467) are contributed by different monomers[1], whereas in TR the probable active site base (His 245)[14] is in the same molecule as the cysteines.

The process of convergent evolution that has led to the alternative dimeric structures and active sites has occurred with the preservation of the structural framework of the FAD/central and NADPH domains, confirming the intrinsically modular construction of these enzymes. Another example of the independent acquisition of similar function by proteins that have diverged from a common ancestor is the dimeric clam haemoglobin, which has subunits that resemble those of tetrameric mammalian haemoglobin[15]. The dimer interface is completely different, however, and provides for an allosteric mechanism which is unrelated to that found in mammalian haemoglobins[15]. □

Received 21 March; accepted 10 May 1991.

1. Karplus, P. A. & Schulz, G. E. *J. molec. Biol.* **195,** 701–729 (1987).
2. Williams, C. H. Jr *The Enzymes 3rd edn* **13,** 89–173 (1976).
3. *Schirmer, R. H. & Schulz, G. E. in Pyridine Nucleotide Coenzymes Part B (Coenzymes and Cofactors)* Vol. 2 (eds Dolphin, D., Poulson, R. & Avramovic, O. 333–379 (Wiley New York, 1987).
4. Thieme, R., Pai, E. F., Schirmer, R. H. & Schulz, G. E. *J. molec. Biol.* **152,** 763–782 (1981).
5. Russel, M. & Model, P. *J. biol. Chem.* **263,** 9015–9019 (1988).
6. Williams, C. H., Jr., Prongay, A. J., Lennon, B. W. & Kuriyan, J. in *Flavins and Flavoproteins* (eds Curti, B., Zannetti, G. & Ronchi, S.) (de Gruyter, Berlin, in the press).
7. Holmgren, A. *J. biol. Chem.* **264,** 13963–13966 (1989).
8. Chothia, C. & Lesk, A. M. *EMBO J.* **5,** 823–826 (1986).
9. Rossmann, M. G., A. Liljas, C. I. Brändén & Benaszak, L. J. in *The Enzymes* (ed. Boyer, P. D.) 61–102 (Academic, New York, 1975).
10. Schulz, G. E. *J. molec. Biol.* **145,** 335–347 (1980).
11. Wierenga, R. K., Drenth, J. & Schulz, G. E. *J. molec. Biol.* **167,** 725–739 (1983).
12. O'Donnell, M. E. & Williams, C. H. Jr *Biochemistry* **24,** 7617–7621 (1985).
13. Karplus, P. A. & Schulz, G. E. *J. molec. Biol.* **210,** 163–180 (1989).
14. O'Donnell, M. E. & Williams, C. H. Jr *J. biol. Chem.* **258,** 13795–13805 (1983).
15. Royer, W. E. J., Hendrickson, W. A. & Chiancone, E. *Science* **249,** 518–521 (1990).
16. Prongay, A. J., Engelke, D. R. & Williams, C. H. Jr *J. biol. Chem.* **264,** 2656–2664 (1989).
17. Kuriyan, J., Wong, L., Russel, M. & Model, P. *J. biol. Chem.* **264,** 12752–12753 (1989).
18. Terwilliger, T. C. & Eisenberg, D. *Acta crystallogr.* **A39,** 813–817 (1983).
19. Wang, B. C. *Meth. Enzym.* **115,** 90–112 (1985).
20. Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235,** 458–460 (1987).
21. Brünger, A. T. X-PLOR (*Version* 1.5) *Manual* (The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, Connecticut, 1988).
22. Jones, T. A. & Thirup, S. *EMBO J.* **5,** 819–822 (1986).
23. Ramachandran, G. N. & Sasisekharan, V. *Adv. Protein Chem.* **23,** 283–437 (1968).
24. Kabsch, W. & Sander, C. *Biopolymers* **22,** 2577–2637 (1983).
25. Priestle, J. P. *J. appl. Crystallogr.* **21,** 572–576 (1988).
26. Lee, B. K. & Richards, F. M. *J. molec. Biol.* **55,** 379–400 (1971).

ACKNOWLEDGEMENTS. We thank W. A. Hendrickson, J. W. Pflugrath, R. M. Sweet and G. A. Petsko for encouragement, advice and use of their X-ray equipment. Beamline X12-C at the National Synchrotron Light Source at Brookhaven National Laboratory is funded by the Office of Health and Environmental Research of the US Department of Energy. This work was supported by the NIH and the NSF. J.K. is a Pew Scholar in the Biomedical Sciences.