Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 1104–1107, February 1985
Biophysics

Influence of solvent accessibility and intermolecular contacts on atomic mobilities in hemerythrins

(*B* values/crystal structures/oxygen carriers/protein flexibility/x-ray diffraction)

STEVEN SHERIFF*†, WAYNE A. HENDRICKSON*‡§, RONALD E. STENKAMP¶||, LARRY C. SIEKER¶||, AND LYLE H. JENSEN¶||

*Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, DC 20375; and Departments of ¶Biological Structure and ||Biochemistry, University of Washington, Seattle, WA 98195

Communicated by Michael G. Rossmann, October 17, 1984

ABSTRACT **Thermal factor parameters (*B* values) have been compared from the refined crystal structures of the myohemerythrin from *Themiste zostericola* and of the octameric hemerythrin from *Themiste dyscrita*. These *B* values, which are directly related to atomic mobilities, are found to correlate rather closely with the solvent accessible areas within the respective crystals. Although protomeric units of the two molecules have exceptionally similar three-dimensional structures, there are marked differences between the patterns of relative atomic mobilities along the polypeptide chains. The differences correspond to lattice and oligomer contacts. An adjustment of the *B* values based on the fraction of accessible area occluded by contacts yields values that correlate well between the independent subunits and that should pertain more closely to those for the protomer free in solution.**

That proteins are internally dynamic molecules is demonstrated by various spectroscopic experiments (1–3), by reaction kinetics (4, 5), by theoretical simulations (6, 7), and by crystal diffraction studies (8–11). This investigation of mobility in hemerythrins is based on standard crystallographic experiments. These provide a comprehensive, albeit averaged, picture as the diffraction data are sensitive to all of the individual atomic mobilities. Detailed refinements yield relatively accurate positional coordinates and a measure of the distribution of atomic displacements about each atomic centroid. In isotropic analyses this measure is the *B* value, or "thermal parameter," which is simply related by $B = 8\pi^2\overline{u^2}$ to the unidirectional mean-square displacement, $\overline{u^2}$, averaged over the measurement time and the entire lattice. These displacements reflect mobility due to harmonic thermal vibrations, dynamic disordering between discrete conformers, and static variations frozen into the crystal lattice.

We have analyzed the *B* values in our well-refined structures of myohemerythrin and an octameric hemerythrin and find strong correlations with solvent-accessible area (12) in the crystal. Even turns of α-helices are readily distinguished by variation in *B* values between internal and external surfaces. However, the patterns of relative mobilities along the polypeptide chains of these two similar molecules differ markedly. Differences arise especially at lattice and oligomer contacts. We have devised a procedure that relates hindrance of mobility to the occlusion of surface accessibility caused by intermolecular contacts. Thus, we "correct" *B* values to ones that might pertain to hemerythrin protomers free in solution.

Hemerythrins are a class of oxygen carriers found in four invertebrate phyla. These proteins bind oxygen and certain other anionic ligands at a dimeric iron center coordinated directly by functional side chains (13, 14). Hemerythrins occur as monomers (myohemerythrin), dimers, trimers, tetramers, and octamers. The folding of the protein backbone has been determined for myohemerythrin (15), a trimeric hemerythrin (16), and two octameric hemerythrins (17, 18) and is a 4-α-helical bundle (19) with an NH_2-terminal arm of nonrepetitive secondary structure (Fig. 1 *a* and *b*). The sequences of *Themiste zostericola* myohemerythrin (21) and *Themiste dyscrita* octameric hemerythrin (24) show 46% identity. There is a 5-residue deletion in hemerythrin between residues 89 and 95 of the 118-residue myohemerythrin chain. An optimal superposition of atomic models for a hemerythrin

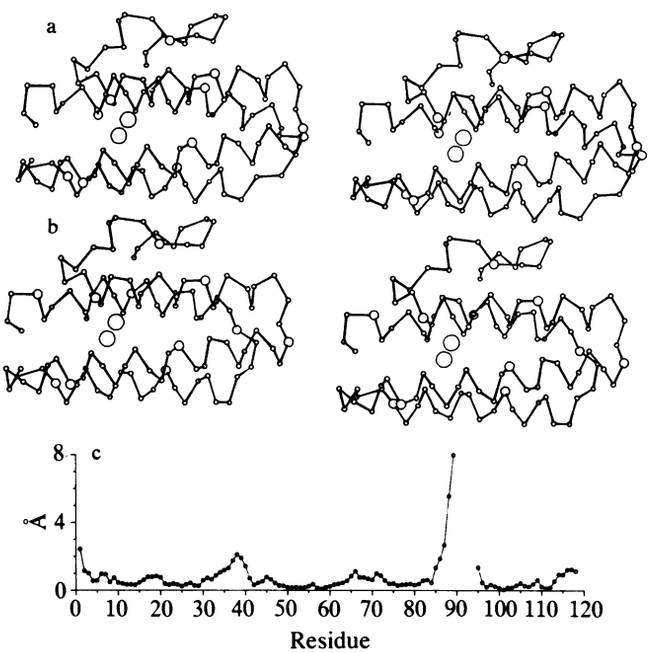

FIG. 1. Three-dimensional structure of hemerythrins in ORTEP stereodrawings of the Cα backbone. (*a*) Myohemerythrin; (*b*) hemerythrin subunit IA; (*c*) rms difference in Å between main-chain atoms vs. residue number for myohemerythrin and hemerythrin subunit IA. Main-chain atoms include N, Cα, C, and O. Optimal superposition of the molecules was obtained by using a program written by R. B. Honzatko (personal communication), which searches for matches after beginning with a small starting set. The transformation matrix in this program is calculated by the Kabsch (20) algorithm. This superposition shows that previous sequence alignments (21–23) that placed the deletion between residues 90 and 96 were misaligned by 1 residue.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

†Present address: 112 Fleetwood Terrace, Silver Spring, MD 20910.
‡Present address: Department of Biochemistry and Molecular Biophysics, Columbia University, 630 West 168th Street, New York, NY 10032.
§To whom reprint requests should be addressed.

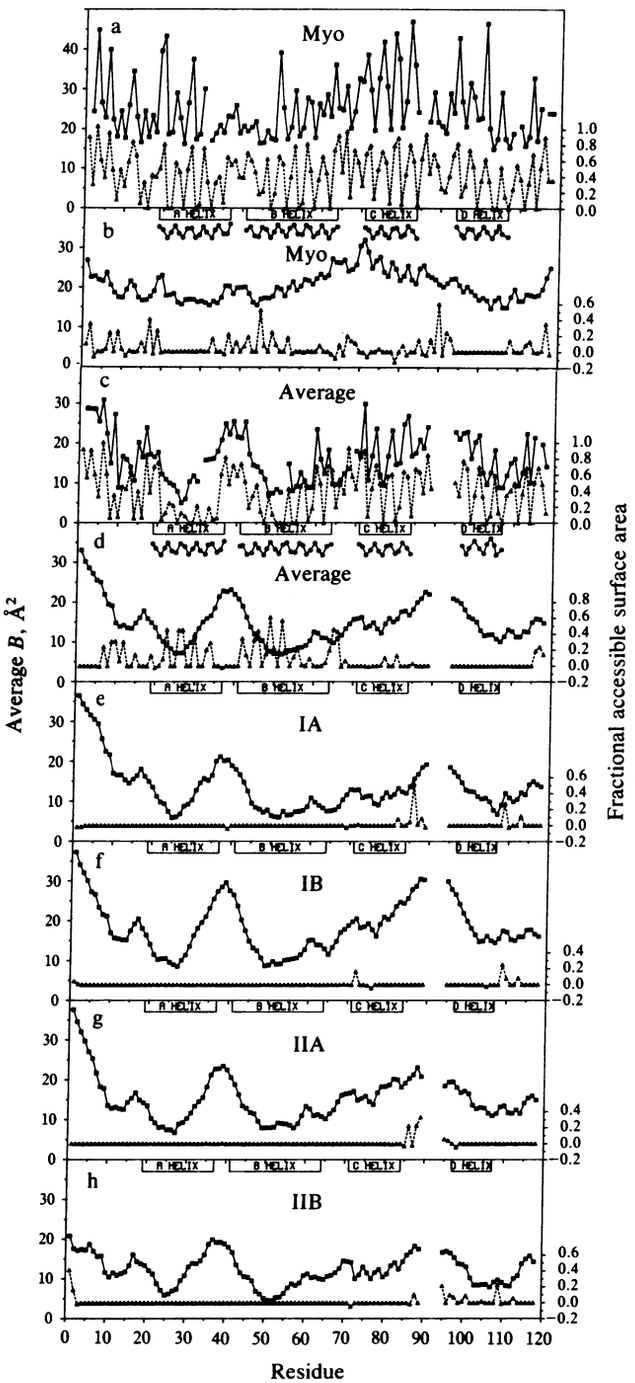

FIG. 2. Average *B* value and fractional accessible surface area vs. residue number of myohemerythrin and hemerythrin. Average *B* value (■—■); fractional surface area (▲--▲); projection of α-helices with external surface pointing up (●—●). (*a*) Myohemerythrin (Myo) average side-chain *B* values and total fractional surface area; (*b*) myohemerythrin average main-chain *B* values and lattice-occluded fractional surface area; (*c*) hemerythrin (average for all four subunits) average side-chain *B* values and total (main- and side-chain) fractional surface area; (*d*) hemerythrin (average for all four subunits) average main-chain *B* values and oligomer-occluded fractional surface area; (*e–h*) hemerythrin subunit main-chain *B* values and lattice-occluded fractional surface area: (*e*) subunit IA, (*f*) subunit IB, (*g*) subunit IIA, and (*h*) subunit IIB. The average main-chain *B* values for each of the hemerythrin subunits vary from 12.2 Å² to 18.3 Å² and that for myohemerythrin is 20.6 Å². Main-chain atoms are defined in the legend for Fig. 1. Surface accessibility calculations used Connolly's MS program (30) and the following parameters: probe sphere radius, 1.70 Å; surface point density, 10; van der Waals radii: carbonyl carbon, 1.70 Å; all other carbon atoms, 1.85 Å; nitrogen, 1.55 Å; oxygen, 1.50 Å, sulfur, 1.75 Å; and iron, 2.00 Å. The values plotted are fractional surface areas for the entire residue. Negative values of fractional occluded surface area are an artifact of the algorithm whereby it can happen that a contacting atom adds more new shared surface area than it occludes.

subunit and myohemerythrin has a rms difference of 0.79 Å between all main-chain atoms in 111 matched residues (Fig. 1*c*).

Structure Refinements

Both proteins were refined by stereochemically restrained least-squares methods (25). The initial myohemerythrin model was derived from an interpretation of a 5.5-Å resolution electron-density map (22). Productive refinement followed use of model-resolved anomalous phasing (26), which led to a definitive 2.8-Å resolution map (unpublished results in collaboration with J. L. Smith). The refinement at high resolution benefited from the application of an overall anisotropic ΔB to account for the anisotropic diffraction with d_{min} = 1.7 Å in *a* and *c* and 1.3 Å in *b*. The current *R* value is 0.159 with a rms deviation from ideal bond lengths of 0.017 Å and a mean magnitude of difference in *B* between bonded main-chain atoms of 0.95 Å². Tests of the effects of restraints on the *B* values showed the pattern of main-chain average *B* values to be remarkably resistant to change, although release of restraints caused excursions of the average side-chain *B* values to increase and *R* to decrease and *vice versa* for tightening of restraints. Seven residues (12, 15, 23, 37, 66, 75, and 92) were modeled as discretely disordered between two conformers; thus, *B* values for these side chains are smaller than would pertain to a single-conformation model.

The initial model for the octamer from *T. dyscrita* was fitted to a 2.8-Å resolution map of the acid methemerythrin (27). This model was refined at 2.0-Å resolution (28) and used as the starting point for refinement of the azidomethemerythrin structure described here. The four subunits in the asymmetric unit of this crystal structure were refined independently throughout. The current model (29) has *R* = 0.175 at 2.0-Å resolution with a rms deviation from ideal bond lengths of 0.026 Å and a mean magnitude of difference in *B* between bonded main-chain atoms of 0.75 Å².

Surface Accessibility Calculations

Surface accessibility was calculated with program MS from Connolly (30). In order to obtain the surface area as it occurs in the crystal or octamer we included those atoms from neighboring molecules or subunits that were within 8 Å of the molecular unit for which the surface was being determined. Lattice- or oligomer-occluded surface areas were obtained from differences between the accessible areas for an isolated molecule or protomer and that for the same structure in the presence of its neighbors. The exposed fraction of the total surface area was determined for each residue. This fractional accessible surface area was evaluated by normalizing the accessible area for a residue in a particular environmental context by the area of that residue, X, as found in a Gly-X-Gly tripeptide in the native conformation (31). Under the conditions employed (probe sphere of radius 1.7 Å and standard van der Waals radii) there is a small cavity in the hemerythrin protomer. We ignore this internal surface here.

Correlation of *B* Value and Surface Accessibility

Plots of average *B* and fractional surface accessibility are shown in Fig. 2 for myohemerythrin (Fig. 2 *a* and *b*) and for

the average (Fig. 2 *c* and *d*) and four individual subunits (Fig. 2 *e–h*) of hemerythrin. The patterns of *B* values, and thus atomic mobilities, for main-chain atoms (Fig. 2 *b* and *d–h*) are distinctive and rather smoothly varying, whereas those for the side chains (Fig. 2 *a* and *c*) have very sharp features. Distributions of fractional accessible surface area (Fig. 2 *a* and *c*) correlate strongly with these features in the *B*-value distributions. The correlation coefficient for comparison of fractional accessibilities in the crystal environment with *B* values for myohemerythrin is 0.71 overall and for the averaged hemerythrin subunit it is 0.65 overall. Correlations are greater for side chains than for main chains—0.74 vs. 0.48 for myohemerythrin and a range from 0.61 to 0.68 vs. 0.57 to 0.61 for the four independent hemerythrin subunits. There is a striking periodicity in *B* values for the side chains of all helices in myohemerythrin, which reflects the sidedness of surface exposure in these helices. A somewhat moderated effect carries over to main-chain mobilities—especially in helices B and C. Hemerythrin *B* values also show periodicity for the exposed helices, C and D and part of B.

The plots of occluded surface accessibility also shown in Fig. 2 indicate a strong influence of intermolecular contacts on atomic mobilities. For example, the most noticeable differences in the otherwise similar *B*-value patterns in the four hemerythrin subunits are associated with the lattice contacts evident in fractional occluded surface areas (Fig. 2 *e–h*). The uniquely low *B* values near the NH_2 terminus of IIB reflect the presence there of a lattice contact, whereas the exceptionally high *B* values at the CD corner of IB parallel an absence of contacts in this vicinity only for this subunit. It seems likely that lattice interactions also hinder mobilities in myohemerythrin. Notably, regions with extensive lattice occlusions of surface area (Fig. 2*b*) have correspondingly low *B* values in relation to those of hemerythrin. Contacts between subunits in the octamer also appear to influence mobilities; there are extensive occluded surfaces (Fig. 2*d*) along the A and B helices where the *B*-value plots have deep depressions.

Adjustment of *B* Values

At first sight the dissimilarity in atomic mobility patterns for myohemerythrin and hemerythrin (Fig. 2 *b* and *d*) is somewhat surprising since the three-dimensional structures are so similar (Fig. 1). The apparent influence of intermolecular contacts noted above offers a possible explanation. In order to test this hypothesis we sought to adjust *B* values for the effect of lattice or oligomer contacts by using the correlation between occluded surface accessibility and hindered atomic mobility. A single contact, such as at the NH_2 terminus of subunit IIB, is also expected to restrict the mobility for several adjacent residues. Thus, there is need to spread the impact of contacts over several residues, and we adjust the average *B* value of residue *i* for lattice contacts according to

$$B'_i = B_i + s \sum_j w_{ij}\, \Delta A_j^{\text{lattice}} \qquad [1]$$

and for oligomer contacts according to

$$B''_i = B'_i + s' \sum_j w'_{ij}\, \Delta A_j^{\text{octamer}}, \qquad [2]$$

where ΔA_j is the fractional surface area of residue *j* that is occluded by contacts, *s* is the factor that correlates area with mobility, and w_{ij} is a weighting function that distributes the effect of contacts to neighbors. We have tried two forms of weighting: a ramp function with $w_{ij} = (1 - |i - j|/n)$ for $i - n < j < i + n$ and $w_{ij} = 0$ outside the number of residues *n* over which the influence is spread; and a Gaussian function with $w_{ij} = \exp[-(i - j)^2/2\, m^2]$. The ramp function proved superior and we report only these results.

Parameters *s*, *n*, *s'*, and *n'* for Eqs. **1** and **2** were chosen to optimize the similarity of main-chain *B*-value distributions in the five different molecular units. This was accomplished through a least-squares minimization of the function

$$\Phi = \sum_{i=1}^{N} (1/4) \sum_{k=1}^{4} (\overline{B'_{\text{dys},i}} - B'_{\text{dys},k,i})^2 + \sum_{i=1}^{N} (\overline{B''_{\text{dys},i}} - B'_{\text{myo},i})^2 \qquad [3]$$

over all *N* residues with allowance for up to three pairs of *s* and *n* parameters. These are, respectively, for adjustments from myohemerythrin lattice contacts to produce B'_{myo}, from *T. dyscrita* hemerythrin lattice contacts to produce $B'_{\text{dys},k}$, and from hemerythrin octamer contacts to produce $\overline{B''_{\text{dys}}}$ averaged over the four crystallographically independent *T. dyscrita* subunits. In addition, B_i was modified by a fixed increment to bring the overall average *B* for each chain to a common average value. Initial *s* values were based on the linear regression of *B* on fractional surface accessibility ($s = 12.92$ Å^2) and various initial *n* values were tried. Fittings were judged by normalized residual magnitudes and by correlation coefficients: C_1 for the average of the six pairwise comparisons of *B* values in the main chain of the four hemerythrin subunits and C_2 for the comparison of the average hemerythrin subunit with myohemerythrin.

The lowest residuals were obtained with six variable parameters (Table 1, case 2) that refined to give $C_1 = 0.90$ and $C_2 = 0.64$ as compared to $C_1 = 0.88$ and $C_2 = 0.24$ for the unmodified data. The adjusted values are displayed in Fig. 3 together with the original main-chain *B* values. The new distributions are obviously more similar and, in addition, they are physically sensible as the regions with highest adjusted *B* values—the NH_2 and COOH termini and the AB, BC, and CD corners—are more exposed than are the helices. Three variable refinements with all *s* values fixed at 12.92 gave similar results ($C_2 = 0.62$; Table 1, case 3), but the fit with a single pair of variables was greatly degraded ($C_2 = 0.35$; Table 1, case 4). It was essential to distinguish between *myo* and *dys* parameters for good results, but not between *dys* lattice and octamer parameters (s, $n_{\text{myo}} = s$, n_{dys} yields $C_2 = 0.55$, whereas s, $n_{\text{dys}} = s$, n'_{dys} yields $C_2 = 0.63$; Table 1, cases 5 and 6).

Table 1. Correlation coefficients and parameters for different sets of *s* and *n*

Case	C_1	C_2	s_{myo}	s_{dys}	s'_{dys}	n_{myo}	n_{dys}	n'_{dys}
1: No adjustment	0.88	0.24	—	—	—	—	—	—
2: Vary all parameters	0.90	0.64	10.3	8.2	6.8	11.6	7.0	8.8
3: $s_{\text{myo}} = s_{\text{dys}} = s'_{\text{dys}} = 12.92$	0.89	0.62	12.9	12.9	12.9	9.5	5.4	4.5
4: $s_{\text{myo}} = s_{\text{dys}} = s'_{\text{dys}}$; $n_{\text{myo}} = n_{\text{dys}} = n'_{\text{dys}}$	0.88	0.35	6.6	6.6	6.6	4.2	4.2	4.2
5: $s_{\text{myo}} = s_{\text{dys}}$; $n_{\text{myo}} = n_{\text{dys}}$	0.88	0.55	6.6	6.6	5.8	14.3	14.3	7.8
6: $s_{\text{dys}} = s'_{\text{dys}}$; $n_{\text{dys}} = n'_{\text{dys}}$	0.90	0.63	10.1	6.8	6.8	11.6	8.6	8.6

C_1 is the average of the correlation coefficients for the six pairwise comparisons of *B* values in the four hemerythrin subunits; C_2 is the correlation coefficient of the average hemerythrin subunit with myohemerythrin; *s*, *n*, *s'*, and *n'* are defined in Eqs. **1** and **2**.

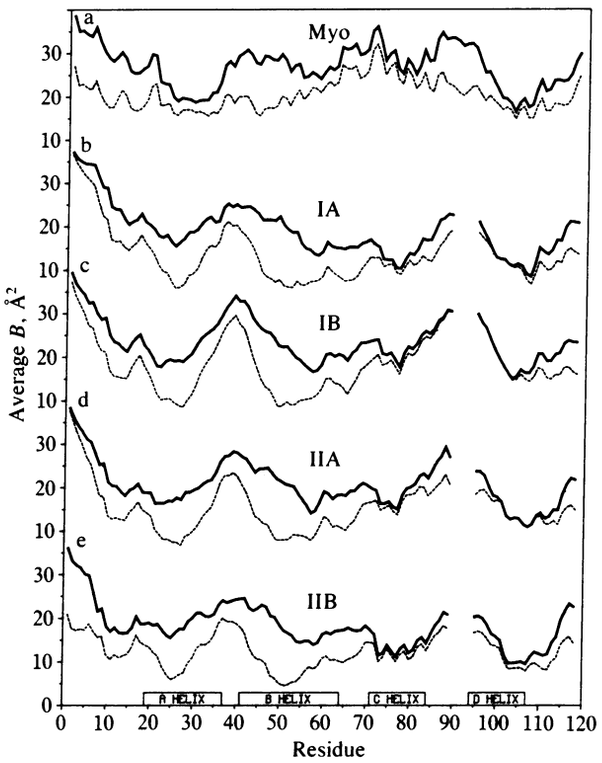

FIG. 3. Adjusted and unadjusted average main-chain B values for myohemerythrin and hemerythrin. Adjusted values (—); unadjusted values (----). (*a*) Myohemerythrin; (*b*) hemerythrin subunit IA; (*c*) hemerythrin subunit IB; (*d*) hemerythrin subunit IIA; (*e*) hemerythrin subunit IIB. The parameters used in Eqs. **1** and **2** to generate the adjusted B-value plots were $s_{myo} = 10.3$; $s_{dys} = 8.2$; $s'_{dys} = 6.8$; $b_{myo} = 11.6$; $n_{dys} = 7.00$; $n'_{dys} = 8.8$. Main-chain atoms are defined in the legend for Fig. 1.

DISCUSSION

The values taken by parameters of this model should be related to physical properties of the crystalline proteins such as local flexibility and concerted deformability. Thus, for example, the strength (s, s') and interaction (n, n') parameters may be lower for *dys* than *myo* because the strong and pervasive octamer contacts dampen the repercussions of individual lattice and octamer contacts. However, it should be kept in mind that the procedure described here is based on statistical correlations rather than fundamental theory. Accordingly, some adjustments in atomic mobilities could be misleading. For example, a region of intrinsically low mobility might wrongly be enhanced if involved in intermolecular contacts. Moreover, the influence of contacts is less clear from other work: (*i*) the correlation coefficient between main-chain atomic displacements in two lysozyme species is 0.61 (8) and (*ii*) we find from deposited data (32) that main-chain B values in dihydrofolate reductases (33) correlate at 0.80 between two copies of the *Escherichia coli* enzyme and at 0.59 between the *Lactobacillus casei* and average *E. coli* structures (compare values of 0.88 and 0.24 from unmodified hemerythrin data in comparable matches). Contacts may nevertheless be influential in these cases; we see here the specific influence of lattice contacts even in the midst of the high correlation of 0.88.

This analysis of diffraction data from hemerythrins points to a clear correlation between solvent accessibility and atomic mobility. Others also find such correlations by crystallography (34) or by molecular dynamics simulation (35). This suggests generality and gives us added confidence in the validity and utility of methods proposed here to extrapolate from atomic mobilities found in a crystalline macromolecule to those expected for a protomer free in solution. This is important for the interpretation of solution experiments such as those in recent antigenicity studies (36).

We thank John Konnert for helpful advice and Jim Griffin for access to computing facilities at the Naval Research Laboratory. S.S. was supported by a National Research Council Research Associateship. The work was supported in part by National Institutes of Health Grant GM-10828 to L.H.J.

1. Wagner, G. & Wuthrich, K. (1978) *Nature (London)* **275,** 247–248.
2. Parak, F., Knapp, E. W. & Kucheida, D. (1982) *J. Mol. Biol.* **161,** 177–194.
3. Lakowicz, J. R. & Weber, G. (1973) *Biochemistry* **12,** 4171–4179.
4. Englander, S. W. & Kallenbach, N. R. (1983) *Q. Rev. Biophys.* **16,** 521–655.
5. Beece, D., Eisenstein, L., Frauenfelder, H., Good, D., Marden, M. C., Reinisch, L., Reynolds, A. H., Sorensen, L. B. & Yue, K. T. (1980) *Biochemistry* **19,** 5147–5157.
6. Karplus, M. & McCammon, J. (1983) *Annu. Rev. Biochem.* **53,** 263–300.
7. Levitt, M. (1983) *J. Mol. Biol.* **168,** 621–657.
8. Artymiuk, P. J., Blake, C. C. F., Grace, D. E. P., Oatley, S. J., Phillips, D. C. & Sternberg, M. J. E. (1979) *Nature (London)* **280,** 563–568.
9. Hartmann, H., Parak, F., Steigemann, W., Petsko, G. A., Ringe Ponzi, D. & Frauenfelder, H. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4967–4971.
10. Kossiakoff, A. A. (1982) *Nature (London)* **296,** 713–721.
11. Konnert, J. H. & Hendrickson, W. A. (1980) *Acta Crystallogr. Sect. A* **36,** 344–350.
12. Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55,** 379–400.
13. Stenkamp, R. E., Sieker, L. C., Jensen, L. H. & Sanders-Loehr, J. (1981) *Nature (London)* **291,** 263–264.
14. Sheriff, S., Hendrickson, W. A. & Smith, J. L. (1983) *Life Chem. Rep.*, Suppl. 1, 305–308.
15. Hendrickson, W. A., Klippenstein, G. L. & Ward, K. B. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 2160–2164.
16. Smith, J. L., Hendrickson, W. A. & Addison, A. W. (1983) *Nature (London)* **303,** 86–88.
17. Stenkamp, R. E., Sieker, L. C., Jensen, L. H. & Sanders-Loehr, J. (1976) *J. Mol. Biol.* **100,** 23–24.
18. Ward, K. B., Hendrickson, W. A. & Klippenstein, G. L. (1975) *Nature (London)* **257,** 818–821.
19. Weber, P. C. & Salemme, F. R. (1980) *Nature (London)* **287,** 82–84.
20. Kabsch, W. (1976) *Acta Crystallogr. Sect. A* **32,** 922–923.
21. Klippenstein, G. L., Cote, J. L. & Ludlam, S. E. (1976) *Biochemistry* **15,** 1128–1136.
22. Hendrickson, W. A. & Ward, K. B. (1977) *J. Biol. Chem.* **252,** 3012–3018.
23. Loehr, J. S. & Loehr, T. M. (1979) *Adv. Inorg. Biochem.* **1,** 235–252.
24. Loehr, J. S., Lammers, P. J., Brimhall, B. & Hermodson, M. A. (1978) *J. Biol. Chem.* **253,** 5726–5731.
25. Hendrickson, W. A. & Konnert, J. H. (1980) in *Computing in Crystallography*, eds. Diamond, R., Ramaseshan, S. & Venkatsen, K. (Indian Academy of Sciences, Bangalore), pp. 13.01–13.23.
26. Smith, J. L. & Hendrickson, W. A. (1982) in *Computational Crystallography*, ed. Sayre, D. (Clarendon, Oxford), pp. 209–222.
27. Stenkamp, R. E., Sieker, L. C., Jensen, L. H. & McQueen, J. E., Jr. (1978) *Biochemistry* **17,** 2499–2504.
28. Stenkamp, R. E., Sieker, L. C. & Jensen, L. H. (1982) *Acta Crystallogr. Sect. B* **38,** 784–792.
29. Stenkamp, R. E., Sieker, L. C. & Jensen, L. H. (1983) *Acta Crystallogr. Sect. B* **39,** 697–703.
30. Connolly, M. L. (1983) *J. Appl. Crystallogr.* **16,** 548–558.
31. Shrake, A. & Rupley, J. A. (1973) *J. Mol. Biol.* **79,** 351–371.
32. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542.
33. Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C. & Kraut, J. (1982) *J. Biol. Chem.* **257,** 13650–13662.
34. Higuchi, Y., Kusunoki, M., Matsuura, Y., Yasuoka, N. & Kakudo, M. (1984) *J. Mol. Biol.* **172,** 109–139.
35. Harvey, S. C., Prabhakaran, M., Mao, B. & McCammon, J. A., (1984) *Science* **223,** 1189–1191.
36. Tainer, J. A., Getzoff, E. D., Alexander, H., Houghten, R. A., Olson, A. J., Lerner, R. A. & Hendrickson, W. A. (1984) *Nature (London)* **312,** 127–134.