Accommodation of Insertions in Helices: The Mutation in Hemoglobin Catonsville (Pro 37α-Glu-Thr 38α) Generates a $3_{10} \rightarrow \alpha$ Bulge†

Jeffrey S. Kavanaugh,‡ Winston F. Moo-Penn,§ and Arthur Arnone*,‡

Department of Biochemistry, University of Iowa, Iowa City, Iowa 52242, and Division of Host Factors, Center for Infectious Diseases, Centers for Disease Control, Atlanta, Georgia 30333

Received August 5, 1992; Revised Manuscript Received January 4, 1993

ABSTRACT: Hemoglobin Catonsville is a mutation of human hemoglobin (an $\alpha_2\beta_2$ tetramer) in which a glutamate residue is inserted into the first turn of a highly conserved 3_{10} helix (the C helix) of each α subunit. In theory, amino acid insertions (or deletions) in protein helices can be accommodated via two distinct mechanisms. One, termed the register shift mechanism, preserves the geometry of the helix while requiring all of the residues on one flank of the insertion site to rotate by 100° in the case of an α helix or by 120° in the case of a 3_{10} helix. The other, termed the bulge (or indentation) mechanism, distorts the local geometry of the helix but does not alter the helix register. High-resolution X-ray diffraction analysis of deoxyhemoglobin Catonsville shows that the inserted residue is accommodated as a bulge, demonstrating that this is a viable mechanism. (In contrast, no such evidence is yet available for the register shift mechanism.) More specifically, the insertion converts one turn of the C helix from 3_{10} geometry to α helix-like geometry, raising the possibility that a common mechanism for accommodating insertions and deletions within helices may involve localized interconversions between 3_{10}, α, and π helical structures.

While amino acid substitutions have been used extensively to investigate the structural and functional properties of proteins (Matthews, 1987; Alber, 1989; Shortle, 1989), amino acid insertions and deletions have received much less attention (Sondek & Shortle, 1990; Sondek & Shortle 1992). Despite the rarity of naturally occurring insertion or deletion mutations within helices or β strands of related proteins (Pascarella & Argos, 1992), several studies have demonstrated that it is possible for proteins to accommodate the addition or deletion of residues in elements of secondary structure without gross loss of function (Starzyk et al., 1989; Sondek & Shortle, 1990; Sondek & Shortle, 1992; Marti et al., 1992). However, the structural consequences of this class of mutation have not been characterized in great detail. In this article we report a 1.7-Å X-ray crystal structure of deoxyhemoglobin Catonsville (Moo-Penn et al., 1989), an unstable hemoglobin mutant that contains a glutamate insertion between the second (C2) and third (C3) residues (Pro 37α and Thr 38α) of the α subunit C helix. A detailed structure/function analysis of hemoglobin Catonsville describing the structural basis for its high oxygen affinity and low cooperativity (Moo-Penn et al., 1989) will be presented elsewhere (J. S. Kavanaugh, P. H. Rogers, W. F. Moo-Penn, and A. Arnone, manuscript in preparation). Here we focus solely on the C helix and the mechanism by which the inserted glutamate is accommodated. The deoxyhemoglobin Catonsville structure reveals a potentially general stereochemical mechanism for accommodating insertions and deletions within helices that involves interconversions of 3_{10}, α, and π helical structures.

MATERIALS AND METHODS

Crystals of deoxyhemoglobin Catonsville, grown according to the procedure of Perutz (1968), were isomorphous with deoxyhemoglobin A crystals (space group symmetry $P2_1$ with $a = 63.2$ Å, $b = 83.6$ Å, $c = 53.8$ Å, and $\beta = 99.4°$). Diffraction data for deoxyhemoglobin Catonsville, including Friedel pairs, were collected with two crystals (approximate dimensions 1.2 × 0.8 × 0.6 mm^3) on a Rigaku AFC6 diffractometer fitted with a San Diego Multiwire Systems area detector. The data (248 787 measurements after merging Bijvoets) were scaled and merged using the method described by Howard et al. (1985), yielding 53 172 independent reflections. The data are 88% complete out of a resolution of 1.7 Å, and the average agreement of the intensities of symmetry-related reflections is 4.2%. Four crystals of approximate dimensions 1.2 × 1.0 × 0.7 mm^3 were used to collect 1.5-Å resolution diffraction data for wild-type deoxyhemoglobin (deoxyhemoglobin A). A total of 828 088 measurements were made of 84 664 independent reflections. The deoxyhemoglobin A diffraction data are 95% complete out to a resolution of 1.5 Å, and the average agreement of the intensities of symmetry-related reflections is 4.4%.

The deoxyhemoglobin Catonsville structure and the isomorphous crystal structure of deoxyhemoglobin A were determined using the published 1.7-Å structure of human deoxyhemoglobin (Fermi et al., 1984) as a starting model. Manual corrections to the atomic models were made with the computer graphics program TOM/FRODO (Cambillau, 1989; Jones, 1985) and then refined with the restrained least-squares program PROLSQ (Hendrickson, 1985). Water molecules were added to the atomic models as described by Kavanaugh et al. (1992), except that a temperature factor cutoff of 50 Å^2 was used. The final crystallographic R[1] value for the deoxyhemoglobin Catonsville model, which contains 199 water molecules and 2 sulfate molecules, is 15.8% for diffraction data with magnitudes greater than 2σ between 6.0- and 1.7-Å resolution (49 525 reflections). The refined deoxyhemoglobin A model contains 197 water molecules and 2 sulfate molecules and has a crystallographic R value of 16.4% for diffraction data with magnitudes greater than 2σ

† This work was supported by grants to A.A. from the National Institutes of Health (HL-40453 and GM-40852) and an Iowa Cardiovascular Center Institutional Research Fellowship to J.S.K.

‡ University of Iowa.

§ Center for Infectious Diseases.

[1] Abbreviations: R value = $\Sigma||F_o| - |F_c||/\Sigma|F_o|$; rms, root mean square.

0006-2960/93/0432-2509$04.00/0

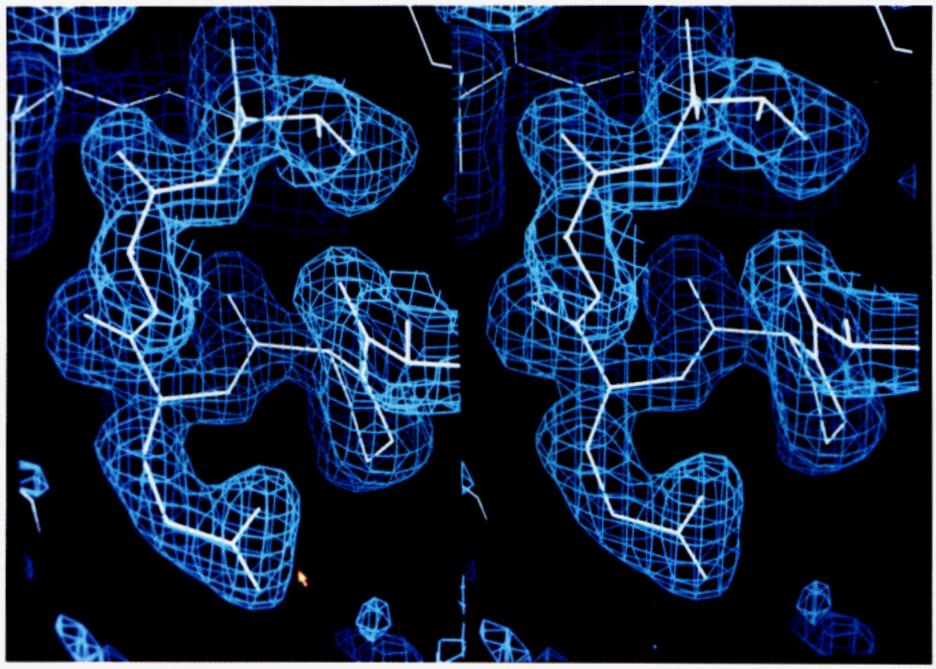

FIGURE 1: Stereo diagram showing the fit of the inserted glutamate residue (red arrow) to a $F_o - F_c$ electron density map. The map was calculated with diffraction data between 6.0- and 1.7-Å resolution, and residues Phe 36α through Phe 43α were omitted from the atomic model. The contouring is at 2 times the rms density of the map.

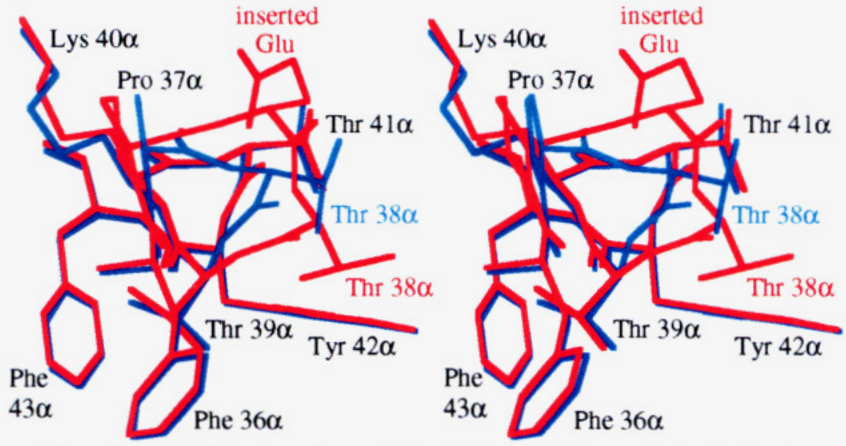

FIGURE 2: Stereo diagram of the α chain C helix in deoxyhemoglobins A (blue) and Catonsville (red) after least-squares superposition of the mutant and wild-type tetramers. Main-chain atoms were superimposed as described previously (Kavanaugh et al., 1992), using the method of Kabsch (1976) as implemented in the program BMFIT (Yuen & Nyburg, 1979). The view is down the length of the helical axes. Shown are residues Phe 36(C1) through Phe 43(CD1).

between 6.0- and 1.4-Å resolution (80 200 reflections). Both deoxyhemoglobin structures have good stereochemistry, with rms deviations from ideal bond lengths of 0.014 Å and rms deviations from ideal bond angles of 1.5°. The average B values are 19.4 and 20.0 Å^2 for the deoxyhemoglobin Catonsville and deoxyhemoglobin A models, respectively.

RESULTS

As shown in Figure 1, the deoxyhemoglobin Catonsville electron density image is completely consistent with the insertion of a glutamate residue between Pro 37α and Thr 38α. Comparison of the refined structures of the mutant and wild-type deoxyhemoglobins (Figure 2) clearly shows that the inserted glutamate residue is accommodated through the formation of a bulge that protrudes from the side of the helix. The bulge—which includes Pro 37α, the inserted glutamate, and Thr 38α—involves a slight conformational change of the proline and a large displacement of Thr 38α. The mutation does not perturb the other C helix residues, whose positions are identical within the accuracy of the refined coordinates. Specifically, the positions of the residues that pack against the heme and the B and G helices, Phe 36 (C1), Thr 39 (C4), and Phe 43 (CD1), are left completely unperturbed by the insertion.

Analysis of the parameters that characterize helix geometry reveals that the first turn of the C helix (a two-turn 3_{10} helix in deoxyhemoglobin A) adopts the geometry of a distorted α helix in deoxyhemoglobin Catonsville (Tables I and II). Specifically, the average ϕ/ψ angles for α and 3_{10} helices are $-62°/-41°$ and $-71°/-18°$, respectively (Barlow & Thornton, 1988). However, since the standard deviation of these mean angles is about 7° (Barlow & Thornton, 1988), only the ψ angle effectively discriminates between the two types of helices. For residues 37α–39α, the average ψ angle is –21.2° for deoxyhemoglobin A (Table I), consistent with the normal 3_{10} character of the C helix. In contrast, the 37α–39α peptide in deoxyhemoglobin Catonsville (a four-residue peptide that includes the inserted glutamate) has an average ψ angle of –43.8°, a value very close to the α helix average.

Table I: Torsion Angles for the α Chain C Helix

	Hb A				Hb Catonsville			
	α1		α2		α1		α2	
residue	φ/ψ (deg)	ω (deg)	φ/ψ (deg)	ω (deg)	φ/ψ (deg)	ω (deg)	φ/ψ (deg)	ω (deg)
Phe 36α (C1)	−131/76	−176	−131/79	−174	−129/64	−177	−130/64	−175
Pro37α (C2)	−56/−25	178	−57/−24	180	−67/−38	−177	−69/−41	−175
inserted Glu					−57/−49	180	−51/−49	179
Thr 38α (C3)	−63/−19	179	−64/−23	−179	−89/−47	−171	−90/−54	−171
Thr 39α (C4)	−70/−18	180	−68/−18	180	−76/−34	−178	−63/−38	−177
Lys 40α (C5)	−66/−20	180	−68/−19	−179	−69/−22	−179	−68/−19	−178
Thr 41α (C6)	−61/−21	179	−63/−18	180	−59/−17	180	−65/−19	−179
Tyr 42α (C7)	−91/−4	180	−90/−4	−177	−97/−5	−178	−93/−4	−178
Phe 43α (CD1)	−123/45	−179	−131/54	−179	−126/47	180	−130/50	−179
mean[a] φ(σ)/ψ(σ)	−63.0(7.0)/−20.7(3.8)		−63.0(5.6)/−21.7(3.2)		−72.3(13.6)/−42.0(7.2)		−68.3(16.3)/−45.5(7.3)	

[a] Average φ/ψ angles and (standard deviations) are in degrees for C helix residues Pro 37α through Thr 39α, including the inserted glutamate residue in deoxyhemoglobin Catonsville. Average φ/ψ angles for α helices and 3_{10} helices in known protein crystal structures are −62°/−41° and −71°/−18°, respectively (Barlow & Thornton, 1988).

Table II: Hydrogen-Bonding Geometry for the α Chain C Helix

	Hb A				Hb Catonsville			
	α1		α2		α1		α2	
residue pair	distance[a] (Å)	angle[b] (deg)	distance[a] (Å)	angle[b] (deg)	distance[a] (Å)	angle[b] (deg)	distance (Å)	angle[b] (deg)
F 36 → T 39	2.92	137	2.97	135				
P 37 → K 40	2.99	123	2.97	124				
T 38 → T 41	3.16	116	3.14	119				
T 39 → Y 42	3.24	120	3.33	121				
K 40 → F 43	2.90	120	3.03	120				
F 36 → T 39[c]					3.28	152	3.37	156
P 37 → K 40[c]					3.17	135	3.26	138
T 39 → Y 42					3.28	120	3.45	121
K 40 → F 43					2.83	121	3.10	119

[a] Distances from the carbonyl O of the first residue to the N of the second residue. [b] Hydrogen bond angles in degrees for the C═O···N interaction. [c] Since the inserted glutamate residue is not numbered, these hydrogen bonds are of the $n \rightarrow n + 4$ type.

The hydrogen-bonding pattern of backbone CO and NH groups provides another gauge of helix character. While the CO group of the *n*th residue forms a hydrogen bond with the NH group of the $n + 3$ residue in a 3_{10} helix, an $n \rightarrow n + 4$ interaction forms in an α helix. In deoxyhemoglobin A, residues 36α–43α clearly show the $n \rightarrow n + 3$ hydrogen-bonding pattern of a 3_{10} helix (Table II). Furthermore, the C═O···N hydrogen bond angles for these residues are in good agreement with the average value of 128° found in 3_{10} helices (Toniolo & Benedetti, 1991). The hydrogen-bonding pattern for the corresponding peptide in deoxyhemoglobin Catonsville is more complex (Table II). Residues 36α and 37α contribute to $n \rightarrow n + 4$ hydrogen bonds, and the inserted glutamate residue and Thr 38α fail to form backbone hydrogen bonds that are shorter than 4.0 Å, while residues 39α and 40α maintain the $n \rightarrow n + 3$ hydrogen-bonding pattern of a 3_{10} helix. The hydrogen bond angles associated with residues 36α and 37α increase an average of 16° toward the mean value of 156° observed for α helices (Toniolo & Benedetti, 1991). The resulting hybrid C helix in deoxyhemoglobin Catonsville is obviously distorted, in that two hydrogen bonds are missing, two hydrogen bonds (those associated with residues 36α and 37α) are slightly elongated, and one residue (Thr 38α) has slightly irregular conformational angles (Table II). Nevertheless, when taken together, these parameters clearly indicated that the insertion of a glutamate residue between Pro 37α and Thr 38α creates a "$3_{10} \rightarrow \alpha$ bulge", where the first turn of the C helix undergoes a transition from 3_{10} helix geometry to α helix geometry.

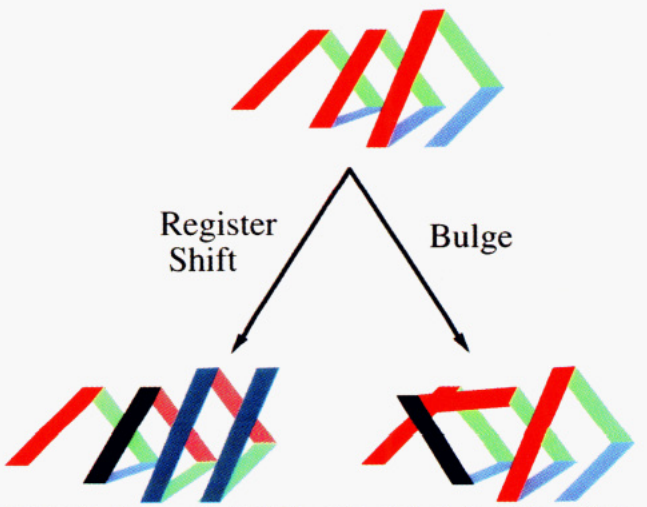

FIGURE 3: Schematic depiction of a stylized 3_{10} helix and the structural consequences of accommodating an inserted residue (black segment) via the register shift and bulge models. The three sides of the 3_{10} helix are color coded (red in front, green in back, and blue on the bottom) to illustrate changes in helix registration.

DISCUSSION

Two modes for accommodating an insertion in a helix, termed "register shift" and "bulge" (Figure 3), have been discussed by Sondek and Shortle (1990, 1992). In the register shift model, all of the amino acids preceding (or following) the insertion shift their positions within the helix by one, and the extra residue is accommodated by enlarging a loop or turn adjacent to the helix. In essence, this type of response is

equivalent to a series of single amino acid substitutions that are either preceded or followed by an insertion within an adjoining solvent-exposed loop. While this preserves the geometry of the helix, it necessarily results in new packing interactions between the helix and the remainder of the protein. The structural implications of the bulge model are just the opposite—the packing interactions are preserved, but the geometry of the helix is distorted by forming a bulge analogous to the β bulge found in β sheets (Richardson et al., 1978).

The insertion-induced perturbations in deoxyhemoglobin Catonsville clearly conform to the bulge model with a mechanism that interconverts different types of protein helices. In this connection, it is interesting to note that interconversions between 3_{10} and α helical conformations have been observed in crystal structures of small (8–16 residues) peptides (Karle et al., 1988, 1990, 1992). This stereochemical mechanism is consistent with the fact that the backbone conformational angles of 3_{10}, α, and π helices are all located in the same low-energy region of ϕ/ψ conformation space (the region that ranges from $-60°$ to $-100°$ in ϕ and $-15°$ to $-70°$ in ψ) and consistent with molecular dynamics simulations which show that the unfolding of α helices takes place via an $\alpha \rightarrow 3_{10}$ pathway (Tirado-Rives & Jorgensen, 1991). (Consistent with such a low-energy pathway is the fact that hemoglobin Catonsville makes up 24% of the patient's total hemoglobin (Moo-Penn et al., 1989), suggesting that the decrease in stability due to the insertion is probably not very large.) On the other hand, it seems energetically unlikely that an insertion would induce a series of "effective mutations" as required by the register shift model. This should be particularly unfavorable because the amphipathic nature of most helices would cause hydrophilic residues to exchange position with hydrophobic ones, disrupting the packing with core residues. Of the types of helices that could potentially accommodate an insertion through a register shift, the globin C helix might have been considered a good candidate because it is short, not highly amphipathic, and its 3_{10} geometry is strictly conserved in all oligomeric and monomeric globin protein structures reported to date (Lesk & Chothia, 1980). Conceivably, the very strong requirement for 3_{10} stereochemistry could have forced a shift in the register of the C helix in hemoglobin Catonsville. The fact that this did not occur suggests that insertion-induced register shifts occur very rarely, if at all, in helices.

Whereas the mutation in deoxyhemoglobin Catonsville is an example of a $3_{10} \rightarrow \alpha$ bulge, the common occurrence of 3_{10} turns at the ends of α helices (Barlow & Thornton, 1988) shows that terminal $\alpha \rightarrow 3_{10}$ "dents" are also energetically possible. This raises the question of whether or not an insertion or deletion can be accommodated as a bulge or dent in the central portion of an α helix. While a dent in the interior of an α helix has not been observed to date, Keefe and co-workers (Keefe et al., 1992; Keefe & Lattman, 1992) have observed a bulge in a Staphylococcal nuclease mutant (Sondek & Shortle, 1990, 1992) containing a glycine insertion that is located in the middle of this enzyme's COOH-terminal α helix between residues Arg 126 and Lys 127. It appears from their proposed atomic model (Figures 2 and 3 in Keefe et al. (1992)) that this insertion forms an $\alpha \rightarrow \pi$ bulge; His 124, Leu 125, and Arg 126 appear to make $n \rightarrow n + 5$ hydrogen bonds characteristic of π helices, the CO group of the inserted Gly residue does not contribute a hydrogen bond to the helix, and the rest of the helix maintains an α helical hydrogen-bonding pattern. Their work clearly demonstrates that amino acid insertions in the middle of α helices can also be accommodated by forming bulges. On the other hand, Pascarella and Argos (1992) have analyzed a large number of insertions/deletions (so-called "indels") in homologous protein structures and found that all of the indels occurring in helices are located within four residues of the helix termini. Together, these data suggest that, while it is possible to accommodate insertions in the interior of helices by forming bulges, insertions at helix termini are energetically less disruptive.

Given our observations for hemoglobin Catonsville and those of Keefe et al. (1992) for Staphylococcal nuclease, researchers should remain open to the possibility that insertions (or deletions) in helices may result in bulges (or dents). The tacit assumption often has been that insertions induce changes in helix register. For example, in a recent publication describing genetically engineered mutants containing either single amino acid insertions or deletions or both an insertion and a deletion within the C helix of bacteriorhodopsin (Marti et al., 1992), the authors operate under the assumption that these mutations result in "predetermined displacements of amino acids in helices" and interpret all of their results in terms of a register shift model. They conclude that they are able to distinguish between a "backrotation of the preceding residues by 100°" and a "forward rotation of all residues subsequent to the insertion" on the basis of functional data alone. Our observations indicate that in the absence of structural information at atomic resolution such a detailed interpretation of functional data is not possible.

ACKNOWLEDGMENT

We thank Dr. Isabella L. Karle for her comments on this paper.

REFERENCES

Alber, T. (1989) *Annu. Rev. Biochem. 58*, 765–798.
Barlow, D. J., & Thornton, J. M. (1988) *J. Mol. Biol. 201*, 601–619.
Cambillau, C. (1989) in *Silicon Graphics Geometry Partners Directory, spring volume*, p 61, Silicon Graphics, Mountain View, CA.
Fermi, G., Perutz, M. F., Shaanan, B., & Fourme, R. (1984) *J. Mol. Biol. 175*, 159–174.
Hendrickson, W. A. (1985) *Methods Enzymol. 115*, 252–270.
Howard, A. J., Nielsen, C., & Xuong, N. H. (1985) *Methods Enzymol. 114*, 452–472.
Jones, T. A. (1985) *Methods Enzymol. 115*, 157–171.
Kabsch, W. (1976) *Acta Crystallogr. A32*, 922–923.
Karle, I. L., Flippen-Anderson, J., Uma, K., & Balaram, P. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 299–303.
Karle, I. L., Flippen-Anderson, J., Uma, K., Sukumar, M., & Balaram, P. (1990) *J. Am. Chem. Soc. 112*, 9350–9356.
Karle, I. L., Flippen-Anderson, J., Uma, K., Gurunath, R., & Balaram, P. (1992) (Abstract) Facile Transition Between 3_{10}- and α- Helix: Structures of 8, 9, and 10 Residue Peptides, in *American Crystallographic Association 1992 Annual Meeting*, Series 2, Vol. 20 (ISSN 0569-4221), p 87, American Crystallographic Association, Buffalo, NY.
Kavanaugh, J. S., Rogers, P. H., Case, D. A., & Arnone, A. (1992) *Biochemistry 31*, 4111–4121.
Keefe, L. J., & Lattman, E. E. (1992) (Abstract) The α Aneurism: a Structural Motif Common to a Diverse Set of Proteins, in *American Crystallographic Association 1992 Annual Meeting*, Series 2, Vol. 20 (ISSN 0569-4221), p 88, American Crystallographic Association, Buffalo, NY.
Keefe, L. J., Lattman, E. E., Wolkow, C., Woods, A., Chevrier, M., & Cotter, R. J. (1992) *J. Appl. Crystallogr. 25*, 205–210.
Lesk, A. M., & Chothia, C. (1980) *J. Mol. Biol. 136*, 225–270.
Marti, T., Otto, H., Rosselet, S. J., Heyn, M. P., & Khorana, H. G. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 1219–1223.
Matthews, B. W. (1987) *Biochemistry 26*, 6885–6888.

Moo-Penn, W. F., Swan, D. C., Hine, T. K., Baine, R. M., Jue, D. L., Benson, J. M., Johnson, M. H., Virshup, D. M., & Zinkham, W. H. (1989) *J. Biol. Chem. 264*, 21454–21457.

Pascarella, S., & Argos, P. (1992) *J. Mol. Biol. 224*, 461–471.

Perutz, M. F. (1968) *J. Cryst. Growth 2*, 54–56.

Richardson, J. S., Getzoff, E. D., & Richardson, D. C. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 2574–2578.

Shortle, D. (1989) *J. Biol. Chem. 264*, 5315–5318.

Sondek, J., & Shortle, D. (1990) *Proteins 7*, 299–305.

Sondek, J., & Shortle, D. (1992) *Proteins 13*, 132–140.

Starzyk, R. M., Burbaum, J. J., & Schimmel, P. (1989) *Biochemistry 28*, 8479–8484.

Tirado-Rives, J., & Jorgensen, W. L. (1991) *Biochemistry 30*, 3864–3871.

Toniolo, C., & Benedetti, E. (1991) *TIBS 16*, 350–353.

Yuen, P. S., & Nyburg, S. C. (1979) *J. Appl. Crystallogr. 12*, 258.