Structural Heterogeneity in Protein Crystals†

Janet L. Smith,*,‡,§ Wayne A. Hendrickson,‡,§ Richard B. Honzatko,§,∥ and Steven Sheriff§,⊥
Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, and Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, D.C. 20375
Received December 3, 1985; Revised Manuscript Received March 5, 1986

ABSTRACT: Extensive conformational heterogeneity is reported in highly refined crystallographic models for the proteins crambin, erabutoxin, myohemerythrin, and lamprey hemoglobin. From 6% to 13% of the amino acid side chains of these four proteins are seen in multiple, discrete conformations. Most common are flexible side chains on the molecular surface, but structural heterogeneity occasionally extends to buried side chains or to the polypeptide backbone. A few instances of sequence heterogeneity are also very clear. Numerous solvent sites are multiplets, and at high resolution, multiple, mutually exclusive solvent networks are observed. The proteins have been studied with X-ray diffraction data extending to spacings of from 0.945 to 2.0 Å. The extensive heterogeneity observed here provides detailed, accurate structures for conformational substates of these molecules and sets a lower bound on the number of substates accessible to each protein molecule in solution. Electron density is missing or very weak for only a few side chains in these protein crystals, revealing a strong preference for discrete over continuous conformational perturbations. The results at very high resolution further suggest that even rather small conformational fluctuations produce discrete substates and that unresolved conformers are accommodated in increased atomic thermal parameters.

Dynamic properties are important components of the structure of protein molecules and, in many cases, must be important contributors to biological function. Local conformational fluctuations are one type of dynamic behavior. The wealth of attractive nonbonded and hydrogen-bonded interactions that confer on globular protein molecules their remarkable stability in aqueous solution also facilitates these conformational fluctuations. Through the many conformational degrees of freedom available to a polypeptide chain, local changes in structure are possible that affect only a small subset of the many weak interactions in the folded molecule. This allows the protein molecule considerable local flexibility without appreciably perturbing its overall structural stability. Precise structural information on conformational fluctuations is important in understanding the dynamic structure of protein molecules.

Protein flexibility has been considered on various levels from thermal vibrations of individual atoms to the unfolding and refolding of part of the polypeptide chain. A review of the various experimental approaches that have been applied to this problem is presented by Karplus and McCammon (1981). Solution studies of hydrogen-exchange kinetics (Englander & Kallenbach, 1983), particularly those using NMR spectroscopy as a probe (Wagner, 1983; Weiss et al., 1984), have proven very powerful in identifying flexible groups in proteins. Molecular dynamics simulations have predicted specific conformational variations in protein molecules through mostly local fluctuations over a short time (Karplus & McCammon, 1983; Levitt et al., 1985). Several sorts of crystallographic observations bear on the dynamic properties of proteins. Individual atomic or group flexibility has been described on the basis of the refined thermal parameters of crystallographic models (Hendrickson, 1985; Artymiuk et al., 1979; Frauenfelder et al., 1979). Flexibility of entire domains has been proposed for those regions of crystalline proteins that lack electron density altogether or that differ in different crystal lattices (Bennett & Huber, 1984). Conformational change in response to ligand binding has also been observed crystallographically (Steitz et al., 1983). Very few experimental results have been reported that offer specific structural information on conformational variability in proteins (see Discussion). Here we describe the crystallographic observation of discrete conformational isomers of protein molecules.

Each of us has recently refined a protein model determined by X-ray crystallography, and in all four proteins we have observed numerous examples of discrete heterogeneity. This paper describes the structures of the observed substates. Several amino acid side chains in each protein exist in multiple conformations, and there is one example of conformational heterogeneity in the polypeptide backbone. Where resolution permits we also see discrete disorder in the solvent structure; this is often correlated with side-chain heterogeneity. During this work we have developed an approach to modeling heterogeneous protein conformers and multiple solvent sites. Our results indicate that discrete structural substates are a general phenomenon in protein crystals. The proteins we have studied are crambin at 0.945-Å resolution (Hendrickson, unpublished results; Hendrickson & Teeter, 1981), erabutoxin at 1.4-Å resolution (Smith, Corfield, Hendrickson, and Low, unpublished results; Bourne et al., 1985, preliminary refinement), myohemerythrin at 1.7/1.3-Å resolution (Sheriff, Hendrickson, and Smith, unpublished results; Hendrickson et al., 1975), and lamprey hemoglobin at 2.0-Å resolution (Honzatko et al., 1985).

Several terms—including flexibility, structural fluctuation, conformational heterogeneity, presence of conformational substates, conformational variability, disorder, conformational perturbation, and thermal motion—have been used to describe the dynamic behavior of protein molecules ranging in scale from thermal vibration of atoms to concerted motion of folded domains. Flexibility is used here as a general term applicable to any internal motion of a protein molecule. We distinguish

† This work was supported in part by Grant PCM-8409658 from the National Science Foundation.
* Address correspondence to this author at Columbia University.
‡ Columbia University.
§ Naval Research Laboratory.
∥ Present address: Department of Biochemistry and Biophysics, Iowa State University, Ames, IA 50011.
⊥ Present address: Laboratory of Molecular Biology, National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases, Bethesda, MD 20205.

0006-2960/86/0425-5018$01.50/0

Table I: Summary of Refinements of Protein Models

protein	atoms in model	amino acid residues	resolution limit (Å)	R factor[a]	mean B for protein atoms (Å^2)	rms deviation from ideal covalent bond lengths (Å)	mean $\|\Delta B\|$ between bonded main-chain atoms (Å^2)
crambin	729	46	0.945	0.111[b]	7[b]	0.015	0.67
erabutoxin	686	62	1.4	0.141	14	0.016	0.90
myohemerythrin	1177	118	1.7/1.3[c]	0.159	23[c]	0.017	0.95
lamprey hemoglobin	1480	149	2.0	0.142	16	0.014	0.95

[a] $R = \sum||F_o| - |F_c||/\sum|F_o|$. [b] For crambin, the R value corresponds to a model with anisotropic thermal factors. The B values tabulated here are the isotropic equivalents. [c] The diffraction data for myohemerythrin fall off anisotropically, extending to 1.7-Å spacings in the a and c directions and to 1.3-Å spacings in the b direction. Although individual isotropic thermal parameters were refined, an overall anisotropic correction was separately refined and applied to these values (Sheriff & Hendrickson, 1986). This resulted in average B values of 28.7, 13.9, and 26.2 Å^2 in the a, b, and c directions, respectively. The same procedure applied to erabutoxin resulted in values of 13, 14, and 15 Å^2.

among three sets of terms according to their differing crystallographic implications. Thermal motion or atomic vibration are among the small-scale atomic displacements modeled by crystallographic thermal parameters. Conformational heterogeneity, discrete disorder, or the existence of conformational substates refers to discrete conformational isomers and is the subject of this paper. The term disorder without modification describes structural regions for which no model can be constructed since electron density is either extremely weak or nonexistent. These classifications imply specific physical phenomena but actually are more indicative of the state of a particular crystallographic analysis. For example, regions of a crystalline protein molecule that are represented by rather featureless density before refinement often become well-defined by electron density as errors in phasing are eliminated. Multiple conformers that had been incorporated in larger thermal parameters may become resolved as an analysis proceeds from lower to higher resolution. In general, thermal parameters describe displacements more complex than harmonic oscillation of atoms about mean positions, and conformational isomers may be a part of this at any resolution. Finally, the temperature of a crystallographic experiment affects the dynamic properties of the crystalline molecule.

Experimental Procedures

All crystallographic models were refined by the stereochemically restrained least-squares method (Hendrickson & Konnert, 1980; Konnert, 1976) using data measured at room temperature by automated diffractometry. This refinement method incorporates well-established stereochemical information along with X-ray diffraction data as "observations" against which the atomic model is refined by least-squares fitting. The results of such a refinement are judged both by the crystallographic residuals and by the agreement of the model with known stereochemistry. These two criteria are balanced during refinement by the relative weighting of the diffraction and stereochemical observations. The least-squares weights for the stereochemical observations are the inverses of the target variances for each class of stereochemical features ($w = 1/\sigma^2$). Values for target variances were as described previously (Hendrickson, 1981) and generally were not changed during the course of refinement. Least-squares weights applied to the diffracted structure amplitudes were a simple function of diffraction angle (Hendrickson, 1981) and were adjusted to maximize convergence of the crystallographic residual while keeping variances of the stereochemical features of the models close to their target values. For example, the weights on the structure amplitudes were kept low enough so that the root mean square (rms) deviation from ideality in covalent bonds was 0.02 Å or less throughout the refinements. Similarly, when individual atomic thermal parameters were refined, these were restrained so that the root mean square difference in thermal parameters between bonded atoms in the main chain was less than about 1.5 Å^2. The refinement of crambin and most of the erabutoxin calculations were done on a Texas Instruments Advanced Scientific Computer, a vectorizing supercomputer. The lamprey hemoglobin refinement was done on a Digital Equipment VAX 11/780 with an attached Floating Point Systems 120B array processor. Both computers were used during the myohemerythrin refinement. The final erabutoxin refinement was done on a VAX 11/780. Refinement statistics for each of the proteins are summarized in Table I.

Conformational heterogeneity generally was not observed until final corrections were being made to the protein models and the solvent structures were being defined. R values at this stage were typically near 0.20. In the course of this work we developed a set of programs for analyzing peaks in $|F_o| - |F_c|$ difference Fourier syntheses, which proved very helpful in describing both discrete disorder and solvent structure. One program searched maps for peaks (or holes), interpolated to find peak centers and heights, sorted peaks by size, and stored a list of sorted peaks with crystallographic equivalents labeled and grouped together. Peak lists were then used by subsequent programs to find where the peaks contacted the protein/solvent models and each other. This information allowed the immediate assignment of some peaks as solvent molecules. Other peaks were examined visually before interpretation. Programs for searching intra- and intermolecular nonbonded and hydrogen-bonded contacts were used extensively in this process. Considerable use was made of the density-fitting program FRODO (Jones, 1978, 1982) on an MMS-X interactive graphics system (Barry et al., 1976) at the Naval Research Laboratory during the refinements of erabutoxin, myohemerythrin, and lamprey hemoglobin. Later an Evans & Sutherland Multi-Picture System at Columbia University was used to examine models of all four proteins. The powerful geometric features of FRODO were especially useful for modeling conformational heterogeneity and for interpreting $|F_o| - |F_c|$ difference electron density. A fast link between the graphics system and main frame computer made this a convenient process. Interactive graphics were not available during most of the crambin refinement, and sections of electron density contoured on paper were studied.

Significant difference electron density was interpreted as heterogeneity, or discrete disorder, when it was too near the existing model to represent solvent. Putative atomic sites were considered too close to be simultaneously occupied if their separation was less than about 2.3 Å for potential hydrogen bonds and about 3.0 Å for potential van der Waals contacts. Heterogeneity was incorporated into the model when a single molecular model did not fit the observed density, provided that

the alternative model was stereochemically reasonable. Stereochemical reasonableness included such features as absence of short contacts with nonheterogeneous parts of the model, reasonable conformational torsion angles, and where appropriate, formation of alternative hydrogen bonds or salt bridges. Heterogeneous models were refined after adjustment of the occupancies of affected atoms. Initially each stereochemical restraint needed to refine every alternative conformer was specified explicitly in PROTIN, the program that determines stereochemical restraints. Later the program was modified to include automatically all stereochemical restraints for groups designated as alternatives. These included any atoms that could not exist simultaneously with the designated "primary" model, whether such atoms represented conformational or compositional alternatives. Of course, restraints on nonbonded contacts between primary and alternative conformers were not applied.

Our approach to refining discrete disorder in protein models determined at less than atomic resolution was to refine individual atomic thermal parameters and to fix atomic occupancies such that thermal parameters for atoms in alternative conformers refined to approximately equal values. The sum of occupancies for alternative positions of a single atom were unity or less, and all atoms of a single conformer were given the same occupancy. It is surely not valid in all cases to demand that thermal parameters be nearly the same for alternative conformers of the same group. However, the extent of our diffraction data did not permit refinement of both occupancy and thermal parameters, which are highly correlated. Further, the atomic displacements modeled by thermal parameters probably are similar for alternative conformers of a given group in most cases. This approach is partly historical since we have not yet modified our refinement program to refine group occupancies.

RESULTS

The discretely disordered residues that we have modeled in each of the four proteins are listed in Table II along with the relevant conformational torsion angles, occupancy and thermal parameters, and chemical circumstances for each. These represent a substantial fraction of the amino acid side chains in each protein, with crambin (15%) and erabutoxin (13%) having more observable heterogeneity than myohemerythrin (6%) or lamprey hemoglobin (7%). The differences in frequency of heterogeneity are probably not a reflection of the true situation but are more likely due to specifics in each case affecting observability of heterogeneity.

Crambin. Crystals of crambin diffract exceptionally well (Teeter & Hendrickson, 1979), and the observed heterogeneity was very clear in electron density maps. The alternative conformers for Phe-13 are a case in point. When refinement was carried out against 1.5-Å diffraction data, this residue appeared very well ordered (Figure 1a), even with individual anisotropic thermal parameters in the model. The isotropic equivalents of B values in the ring ranged from 5.6 Å^2 for C_γ to 8.2 Å^2 for $C_{\epsilon 1}$. However, when data to 1.0-Å spacings were included in the analysis, several large peaks in an $|F_o| - |F_c|$ synthesis (Figure 1c) appeared around the phenyl ring, clearly showing two conformers for this residue. The separation between equivalent atoms in the two confomers is very small (0.68 Å for $C_{\epsilon 1}$'s, 0.61 Å for $C_{\epsilon 2}$'s, and 0.72 Å for C_ζ's), and the side-chain torsion angles χ_1 and χ_2 differ by only 11° and 3°, respectively. The flexibility of Phe-13 appears to be facilitated by an internal cavity of radius approximately 1.4 Å adjacent to the phenyl ring. While this cavity provides space within which the Phe-13 ring can move, attractive interactions

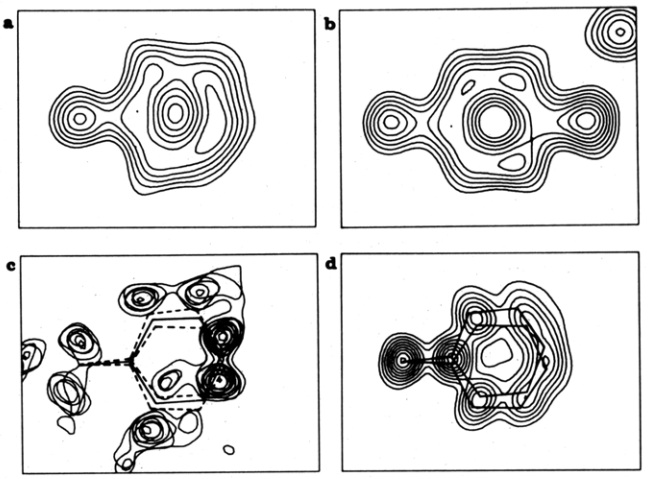

FIGURE 1: Conformational heterogeneity in Phe-13 of crambin. (a) $2|F_o| - |F_c|$ density at 1.5-Å resolution in the plane of the phenyl ring. (b) Equivalent density in the plane of the nonheterogeneous Tyr-44 phenolic ring. (c) $|F_o| - |F_c|$ density at 1.0-Å resolution for Phe-13. The refined single-conformer model is shown with solid bonds and the two-conformer model fit to this density in dashed bonds. (d) $2|F_o| - |F_c|$ density at 0.945 Å with the refined two-conformer model.

with nearby groups doubtless stabilize the observed conformers. The side chain of Thr-2 also lines this small, hydrophobic cavity and it, too, exists in multiple conformations.

Crambin crystals include molecules variable in amino acid sequence at residue positions 22 (Ser/Pro) and 25 (Ile/Leu). This compositional heterogeneity was discovered in the course of model refinement and later confirmed chemically (Teeter et al., 1981). It is not known from chemical or crystallographic evidence whether there are two, three, or four sequence isomers of crambin. In the crystal there is further heterogeneity at these positions since both Ser-22 and Ile-25 have conformational isomers. Crambin is the only one of the four proteins where such three-way disorder is observed. The sequence heterogeneity at Ser/Pro-22 apparently causes conformational heterogeneity in Tyr-29 of an adjacent molecule. The phenolic hydroxyl group of Tyr-29 makes a lattice contact with the NH of Ser-22 through a water molecule. This interaction cannot occur with Pro-22, causing the Tyr-29 side chain to occupy another position and the water molecule to be removed. Ile/Leu-25 is one turn away from Ser/Pro-22 in an α-helix of crambin, and Tyr-29, in both its conformations, makes contacts shorter than 4 Å with Leu-25 and one conformer of Ile-25.

Two conformers were introduced at Pro-36 in crambin when a flat ring resulted from refinement with a single conformer. Two models were constructed with the C_γ methylene groups puckered on either side of the plane, and each was assigned half-occupancy. The strain of the flat proline ring thus relieved, the two-conformer model was refined. The opposite puckers were maintained in refinement, and the isotropic equivalents of thermal parameters of the affected atoms were reduced by 0.5 Å^2 to 0.7 Å^2 compared with the values in the single-conformer model. It is possible that the electron density calculated for the C_γ methylene hydrogens in the single-conformer model was sufficient to prevent the appearance of difference electron density for the puckered C_γ atoms.

Erabutoxin. Eight external side chains in erabutoxin are seen in multiple conformations. All of these side chains participate in more than one set of hydrogen bonds or salt bridges. Many are involved in intermolecular lattice contacts in at least one of their conformations. An example of this is the correlated heterogeneity of Arg-33 and the C-terminal Asn-62 (Figure 2). In one case a strong intermolecular salt

Table II: Heterogeneous Side-Chain Conformers

residue	side-chain torsions (deg)	mean thermal parameter (Å^2)	partial occupancy	environment[a]
crambin				
Thr-2	62	7.2	0.60	C
	−85	7.7	0.20	
	−46	7.7	0.20	
Ile-7	−62, −170	8.3	0.50	L
	−84, −60	11.1	0.50	
Phe-13	178, −90	6.0	0.50	C
	−171, −87	5.6	0.50	
Ser/Ser/Pro-22	72	8.5	0.30	Q, L
	151	12.0	0.30	
	−30, 37, −30	6.7	0.40	
Ile/Ile/Leu-25	−83, −71	5.4	0.30	Q, L
	−63, 176	6.2	0.30	
	−69, −66	5.7	0.40	
Tyr-29	−172, 55	10.3	0.60	L, Q
	170, 57	9.2	0.40	
Pro-36	−27, −37, 32	9.3	0.50	P
	−16, 26, −26	9.4	0.50	
erabutoxin				
Ser-9	−75	19	0.50	S
	65	21	0.50	
Ser-18	−21	16	0.50	S
	55	8	0.50	
Arg-33	−64, −179, −52, 149, 179	15	0.60	L, H
	−45, 173, 72, −152, 177	17	0.40	
Glu-38	173, −156, 117	13	0.50	L
	−78, −163, 175	15	0.50	
Lys-51	−65, 169, −153, −158	15	0.50	S
	−68, −175, −77, 169	13	0.50	
Glu-56	−80, −152, 36	25	0.50	L
	−69, 164, −12	20	0.50	
Ser-57	81	16	0.50	S
	−37	21	0.50	
Asn-62	40, −131	23	0.60	L, H
	179, 92	16	0.40	
myohemerythrin				
Glu-12	−173, 89, 165	36	0.50	L, H
	−70, −174, −16	36	0.50	
Arg-15	−155, 168, −102, −64, 4	25	0.60	H
	−155, 165, −108, 168, −4	26	0.40	
Glu-23	−76, −57, −55	31	0.60	S
	−76, 172, −8	33	0.40	
Arg-37	−69, −160, −65, 135, −11	23	0.50	L, H
	−60, −78, 82, 63, 177	24	0.50	
Lys-66	−58, 160, −166, 131	31	0.50	L
	−7, 161, −81, −170	33	0.50	
Lys-75	−39, −86, −154, 146	44	0.50	S
	−39, −86, 110, −104	42	0.50	
Asp-92	83, 132	29	0.50	H
	57, −76	29	0.50	
lamprey hemoglobin				
Thr/Asx-29	60	16	0.50	Q, L, H
	−46, 99	16	0.50	
Tyr-30	56, −73	21	0.60	L
	69, −80	22	0.40	
Leu-38	−65, −171	11	0.50	C
	−138, 44	13	0.50	
Glu-63	−74, 175, 4	25	0.60	H
	−104, −173, 111	19	0.40	
Lys-93	−163, −162, −134, −113	22	0.50	S
	−162, 159, −120, −122	24	0.50	
Lys-97	−66, 154, 178, 74	27	0.50	S
	−103, 173, −120, 141	27	0.50	
Arg-99	−88, −170, −67, −88, −2	14	0.50	L
	−88, −146, −141, −97, 5	12	0.50	
Lys-107	−75, 85, −162, −116	33	0.50	L
	−75, 45, 172, 106	34	0.50	
Glu-114	62, 178, −78	18	0.50	L
	143, −160, 174	19	0.50	
Tyr-115	−72, 112	16	0.50	S
	−62, 114	16	0.50	
Met-140	−55, −178, 179	13	0.50	S
	−55, −178, 61	13	0.50	

[a] Codes for side-chain environments: C = buried side chain near cavity, H = hydrogen bonds or salt bridges formed to other protein atoms, L = involved in direct protein–protein lattice contact, Q = sequence heterogeneity, P = oppositely puckered proline conformers, and S = side chain in solvent.

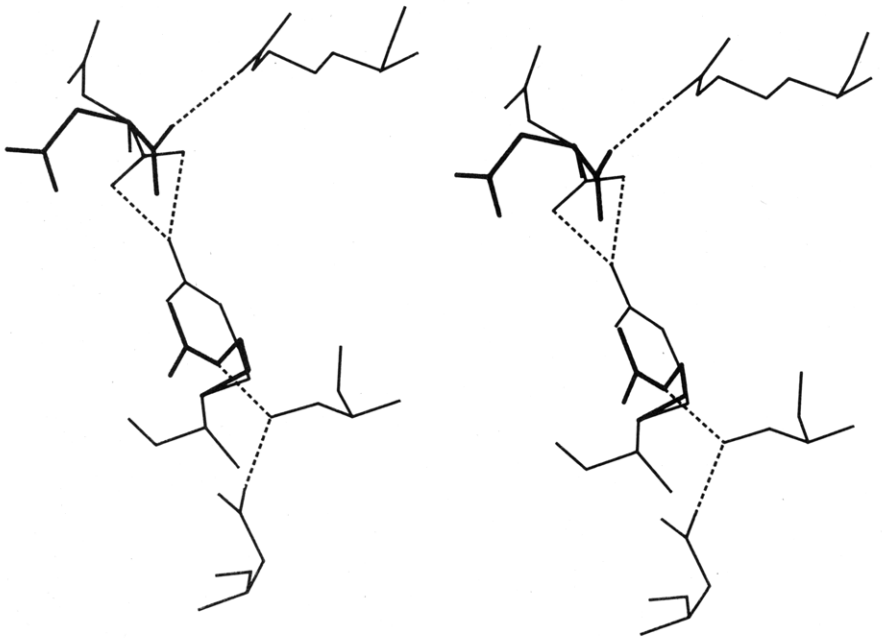

FIGURE 2: Stereo drawing of correlated heterogeneities at a lattice contact in erabutoxin B. Conformers 1 are shown in heavy lines, conformers 2 and nonheterogeneous groups in light lines, and hydrogen bonds and salt bridges in dashed lines. Asn-62 (top left) is the C-terminal residue. Its carboxyl group (conformer 1) forms an intramolecular salt bridge to Arg-39 (upper right) while conformer 1 of Arg-33 (center) bridges to Asp-31 (bottom) through the side chain of Ser-53 (lower right). Conformers 2 of Asn-62 and Arg-33 salt bridge to each other. Asn-62, Arg-39, and Ser-53 are in one molecule, and Arg-33 and Asp-31 are from a neighboring molecule in the crystal.

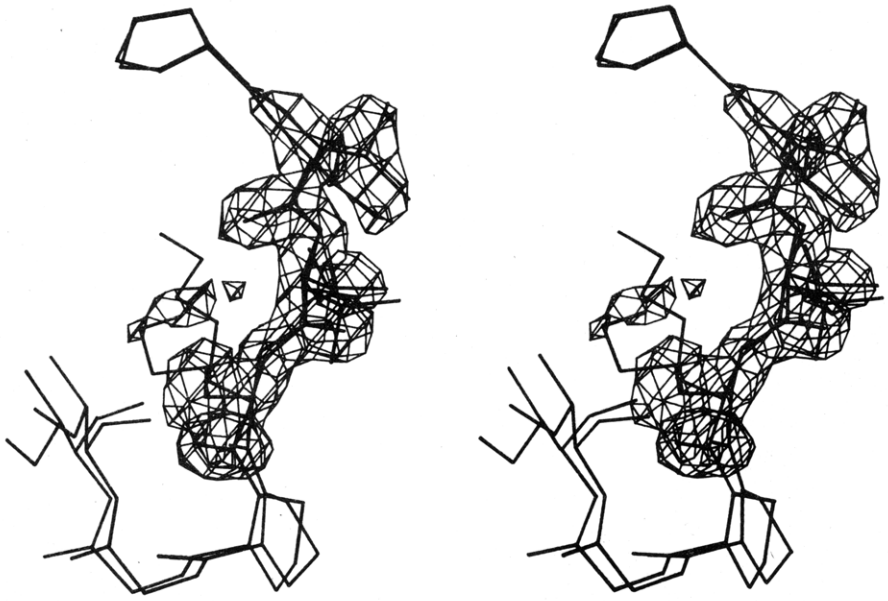

FIGURE 3: Stereo drawing showing two possible conformers for the heterogeneous backbone stretch in erabutoxin B. From the top, the sequence is Pro-44-Thr-45-Val-46-Lys-47-Pro-48-Gly-49-Ile-50. $|F_o| - |F_c|$ electron density calculated with the Thr-45-Val-46-Lys-47 tripeptide removed from the model is superimposed.

bridge forms between the Arg-33 side chain and the C-terminal carboxyl group, and in the other both charged groups form intramolecular salt bridges: Arg-33 to Asp-31 and the C-terminus to Arg-39. Similarly, side chains of Glu-38, Lys-47, and symmetry-related Glu-56 are grouped together in the crystal and appear to be correlated in their heterogeneity, but the correlation cannot be as clearly defined because the contacts are mediated by solvent molecules and the thermal parameters are large. Side-chain heterogeneity in erabutoxin was confirmed by removing all discretely disordered side chains, refining several cycles, and verifying the return of electron density for removed groups.

In addition to the eight individual, heterogeneous side chains in erabutoxin, the hexapeptide Thr-45-Val-46-Lys-47-Pro-48-Gly-49-Ile-50 exists in the crystal in at least two conformations. Except for the ends of the Val-46 and Lys-47 side chains, electron density is strong and continuous through this region but has several persistent features incompatible with a single conformer. Thr-45 and Val-46 each have large density features near their carbonyl groups; the Pro-48 ring has density consistent with pucker in both directions and refines to be flat when modeled as a single conformer. Density for the Lys-47-Pro-48 peptide is large and strong but rather featureless, and the Gly-49-Ile-50 peptide has density suggesting a large flexibility of the carbonyl bond perpendicular to the peptide plane. We have modeled this region as the two conformers shown in Figure 3. In one conformer there is a hydrogen bond between the Lys-47 C=O and the Ile-50 NH to form a tight turn. In the other conformer the turn does not involve a hydrogen bond, and the Lys-47-Pro-48 and Gly-49-Ile-50

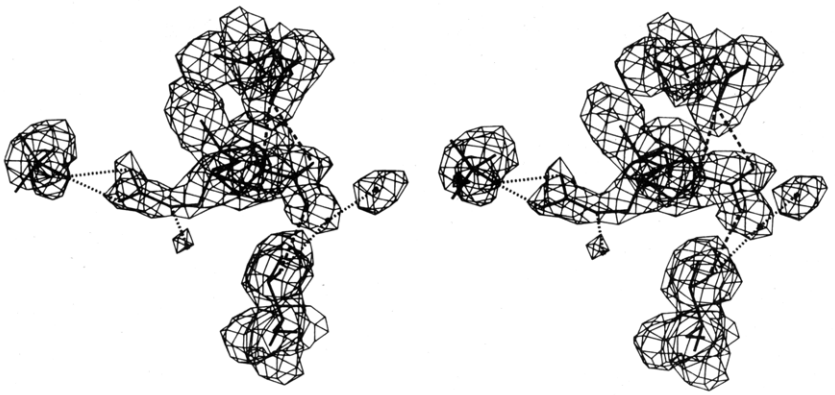

FIGURE 4: Stereo drawing of the widely separated Arg-37 conformers in myohemerythrin. In conformer 1 (pointing right) the guanidinium group salt bridges to Asp-34 (top), which is one turn earlier in the A helix, and hydrogen bonds to Thr-52 (bottom) of another molecule. Conformer 2 (pointing left) salt bridges to a sulfate ion (left), while a water molecule occupies the site vacated by conformer 1 and hydrogen bonds to Thr-52.

peptides are rotated slightly. Partially occupied solvent sites compatible with one of the peptide conformers fill the electron density features near the Thr-45 and Val-46 carbonyl groups. Pro-48 is puckered in opposite directions in the two conformers. Interpretation of this part of the erabutoxin electron density maps has been particularly difficult throughout refinement (Bourne et al., 1985) and has been the subject of much refinement experimentation in the course of this work. It is probably impossible to sort out the heterogeneity accurately at 1.4-Å resolution. It is quite likely that this hexapeptide exists in more than two conformations in the crystal since so many flexible torsion angles are involved and since there are neither hydrogen bonds tying it to the rest of the protein nor direct lattice contacts with other protein molecules. The situation is further complicated by the solvent molecules that certainly alternate in sites with flexible parts of the protein. Nevertheless, our current interpretation is both consistent with the electron density and stereochemically reasonable.

Myohemerythrin. All of the conformational heterogeneity observed in the myohemerythrin crystal involves charged side chains on the surface of the molecule. Three of the seven heterogeneous side chains (Glu-12, Arg-37, and Lys-66) form lattice contacts to nonheterogeneous groups on other protein molecules through salt bridges or hydrogen bonds in one of the observed conformers. The heterogeneities in Glu-12 and Arg-15 are correlated in that one conformer of Glu-12 is too near one conformer of Arg-15 for them to be present simultaneously. These side chains have compatible conformations where they salt bridge to each other, but when Arg-15 is in its alternative position, Glu-12 must move. The alternative for Glu-12 forms a salt bridge with a symmetry-related Lys-113. Four of the seven discretely disordered side chains in myohemerythrin alternate in sites with solvent molecules, based on stereochemical exclusion. For example, Arg-37 (Figure 4) forms a salt bridge either to Asp-34 or to a sulfate ion. When the sulfate bridge is in place, a water molecule occupies the alternative site for Arg-37 and replaces its hydrogen bond to a symmetry-related Thr-52 as well.

The modeled heterogeneity in myohemerythrin was verified near the end of the refinement process by removing atoms with multiple sites, refining for a few cycles, and then examining difference electron density maps for the return of density for both conformers.

Lamprey Hemoglobin. Potentially heterogeneous side chains in lamprey hemoglobin were modeled first in one and then in the other conformation. Only if difference density returned for both omitted conformers was heterogeneity included in the model. Significant difference density within 2 Å of the protein was modeled in this fashion. In several cases, side chains that had been incorrectly modeled initially were corrected by this procedure. For two of the eleven residues observed to be heterogeneous (Arg-99 and Lys-107), the two conformers are truly resolved in electron density maps at 2.0-Å resolution. Eight other cases of heterogeneity that have been modeled account for an anisotropic spread of electron density that could not be accounted for by either conformer alone. The case of sequence heterogeneity at Thr/Asx-29 was modeled on the basis of information from the chemical sequence (Li & Riggs, 1970; Hombrados et al., 1983) and was not suggested strongly by electron density. Both side chains make chemically reasonable contacts. The Thr side chain hydrogen bonds to the main-chain carbonyl oxygen of residue 25, one turn earlier in the A helix. The Asp or Asn side chain forms salt bridges or hydrogen bonds to Arg-76 and to a symmetry-related Lys-54, suggesting that it is probably a negatively charged Asp residue.

Most of the side chains observed to be heterogeneous in lamprey hemoglobin are on the outer surface of the molecule and have multiple hydrogen bonding or salt bridging interactions available to them, as described above for residue 29. Also of interest are Gln-114 and Tyr-115, residues whose heterogeneity appears to be correlated since their planar side chains are nearly parallel and separated by 3.5–3.7 Å. Four of the external, heterogeneous side chains in lamprey hemoglobin are involved in crystal lattice contacts in one of their conformations.

Leu-38 is an unusual example of an internal residue modeled in two conformations. The side-chain torsions are such that the alternative C_γ atoms point in very different directions from the single C_β atomic site, while the alternative pairs of C_δ atoms are quite close to one another. This model accounted for persistent electron density at the alternative C_γ site.

Heterogeneous Solvent Structure. In all four protein crystals we find several pairs of solvent sites too near one another to be simultaneously occupied. Crambin and erabutoxin additionally have alternative networks of solvent molecules; in crambin one of these must represent three-way disorder. A large heterogeneous water network from the erabutoxin crystal is shown in Figure 5. The hydrogen-bonding versatility of water molecules is shown by the ability of the two distinct solvent networks, each using about the same number of hydrogen bonds, to link two protein molecules in a crystalline lattice contact.

DISCUSSION

Distribution of Heterogeneity in Protein Models. Several

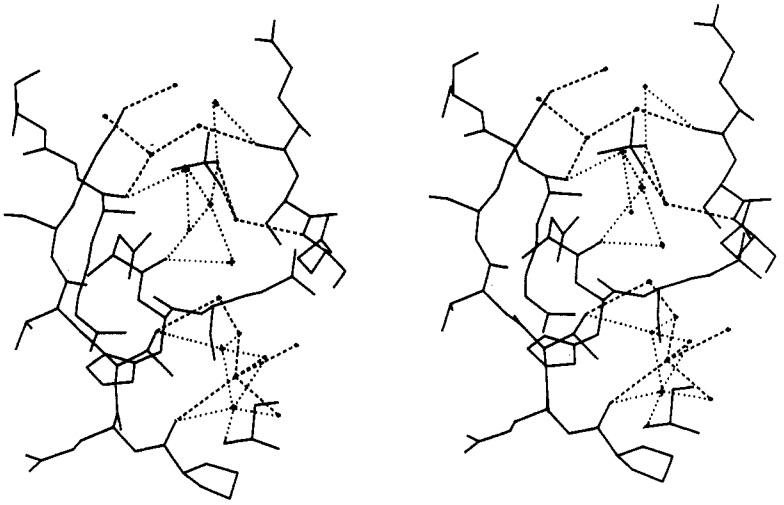

FIGURE 5: Stereo drawing of alternative water networks in erabutoxin. The two sets of hydrogen bonds are shown in dashed and dotted lines. These two networks were determined to be mutually exclusive on the basis of solvent sites too near one another to be simultaneously occupied. There are six water molecules in each partially occupied network.

trends emerge from our observations of a total of 38 structurally heterogeneous amino acid side chains in these four proteins. First, the vast majority (36) of these amino acids are found on the protein surface where flexibility is least likely to perturb the overall stability of the protein molecule. Of these, about half (19) are charged groups, predominantly long lysine, glutamate, and arginine side chains. Many of these charged groups form alternative salt bridges to other parts of the structure, usually nonheterogeneous. In a few cases, however, heterogeneity of two charged groups is correlated so that they salt bridge either to each other or to other charged groups. Other, more subtle, correlations may also exist, but only when groups approach too closely can it be said with certainty that their heterogeneities are correlated. Of the seventeen uncharged, heterogeneous, surface side chains, nine are polar (Ser, Tyr, Thr, Asn), seven are aliphatic (Ile, Leu, Met, Pro, Val), and one is a sequence isomer (Ser/Pro). Two heteroeneous residues are buried in the protein interior: Phe-13 of crambin and Leu-38 of lamprey hemoglobin.

The conformational substates we have modeled generally involve only atoms in amino acid side chains. However, rearrangements of a side chain also can cause subtle changes in the backbone structure. In a few instances it was necessary to extend heterogeneity into the main chain in order to maintain acceptable geometry in the alternative side chain during refinement. This was the case with Thr-2 of crambin, with Ser-9 of erabutoxin, and with Thr/Asx-29 of lamprey hemoglobin. The peptide Thr-45-Val-46-Lys-47-Pro-48-Gly-49-Ile-50 in erabutoxin is an example of extensive heterogeneity in the main chain. The two conformers that have been modeled are not grossly different—the mean separation of equivalent atoms is 0.9 Å with a maximum of 3.5 Å. Inasmuch as so many atoms are observed to be heterogeneous, it is unlikely that only two substates exist in this region. However, the limits of the diffraction analysis at 1.4 Å prevent more elaborate modeling of this region.

Most intermolecular contacts between protein molecules in crystals are mediated by a shared hydration shell. These waters provide facile adaptation between intrinsically noncomplementary surfaces. Here we observe considerable heterogeneity. Several flexible side chains alternate in sites with water molecules, and there are many water molecules with multiple sites for favorable hydrogen-bonding interactions. We identify as heterogeneous those solvent sites that are too close to one another to be simultaneously occupied. In crambin and erabutoxin, networks of mutually exclusive solvent sites are observed. Less and less order is observed in passing from solvent at the well-packed protein interfaces to the continuum. It is likely that the unrestrained thermal parameters refined for solvent molecules are sometimes falsely high and mask further discrete disorder in solvent regions of the crystal.

A significant proportion of direct protein–protein lattice contacts are affected by structural heterogeneity in the four crystals. Crambin has a rather hydrophobic surface and 18 of the 84 interprotein lattice contacts shorter than 4.0 Å include atoms in partially occupied sites. While none of the four direct hydrogen bonds between crambin molecules involve heterogeneous groups, this is not the case for the other proteins. There are 14 intermolecular hydrogen bonds and salt bridges in the erabutoxin crystal. Ten of these involve at least one heterogeneous group. Similarly, three of eight such interactions in the myohemerythrin crystal and six of seventeen in lamprey hemoglobin have one group partially occupied. While each protein–protein interaction surface also includes contacts unaffected by observed heterogeneity, these results are somewhat surprising and indicate that the crystal lattices can accommodate a significant amount of heterogeneity at molecular interfaces. In contrast, other observations indicate that structural heterogeneity in some lattice contacts inhibits crystal growth (Bott et al., 1982), and reduced atomic mobilities at lattice contacts have also been demonstrated (Sheriff et al., 1985).

Heterogeneity and Resolution Limits of Diffraction Data. The ability to detect structural heterogeneity is strongly influenced by the resolution limits imposed by the diffraction data and by the thermal parameters for the protein molecule in its particular crystal lattice. In crambin, at 0.945 Å, heterogeneity could be modeled without the aid of interactive graphics and conformers whose atomic positions differed by less than 1 Å could be placed accurately. A few cases of three-way heterogeneity were also observed in crambin. In erabutoxin, at 1.4 Å, it was generally straightforward to fit heterogeneous side chains but quite difficult to fit two subtly different main-chain conformers. Solvent molecules alternating with protein atoms in closely spaced sites made a few areas difficult to interpret. We observed less discrete disorder in myohemerythrin, at 1.7/1.3 Å, than in the other protein crystals. This may be due to the large and highly anisotropic thermal parameters arising from the rather loose packing of myohemerythrin in the crystal lattice. Also, alternative solvent

networks were not observed in the myohemerythrin crystal as they had been in crambin and erabutoxin. Heterogeneous side-chain conformers in lamprey hemoglobin, at 2.0 Å, often were assigned on the basis of persistent $|F_o| - |F_c|$ electron density, which seldom appeared as well-formed side chains. Confidence in such assignments came from the appearance of density for a conformer that had been omitted from the model being refined, as well as from results on the other structures at higher resolution.

The resolution limits of a structure analysis affect both the minimum population for a structural substate to be observable and the spatial separation for distinct substates to be resolved. Three-way heterogeneity is observed in crambin (0.945 Å) with occupancies as low as 20% in Thr-2, but only two-way heterogeneity is seen in the other structures (1.4–2.0 Å). This suggests that until resolution is significantly better than 1.4 Å, occupancies must be significantly greater than 33% for interpretable electron density to appear. The electron density features representing heterogeneity in lamprey hemoglobin (2.0 Å) were generally of poor quality. Many of these conformers are not widely separated, and the quality of the electron density features correlates well with conformer separation, suggesting that the conformers in lamprey hemoglobin are simply not as well resolved as those in the other structures. In general, conformer separation in each of the four protein crystals is correlated with the resolution of the structure analysis. Conformer separation is measured here as the distance between those atoms that represent the extremes of separation of the conformers. The ranges of conformer separation are 0.72–2.70 Å (average 1.75 Å) for crambin, 0.77–4.92 Å (average 2.74 Å) for erabutoxin, 3.11–9.04 Å (average 4.56 Å) for myohemerythrin, and 1.24–5.53 Å (average 2.68 Å) for lamprey hemoglobin. Conformers that are not widely separated relative to the resolution of the structure analysis are prevented from collapsing to their mean position by the stereochemical restraints employed during refinement.

Heterogeneity and Atomic Thermal Parameters. Some of the differences in degree of observed heterogeneity in our four protein crystals seem to be correlated with the thermal parameters of each of the atomic models. Crambin and myohemerythrin are the extremes in this regard. Crambin has exceptionally low thermal parameters and the several instances of structural heterogeneity that we observed were very clear in electron density maps. Myohemerythrin has very high thermal parameters and the lowest proportion of heterogeneous side chains. Although the density features interpreted as heterogeneity in myohemerythrin were more recognizable as side chains than were those in the lamprey hemoglobin analysis at lower resolution, less heterogeneity was observed in the myohemerythrin crystal. Indeed, the diffraction data for lamprey hemoglobin were unnaturally truncated to 2.0-Å spacings as is reflected in the relatively low value of the average refined atomic thermal parameter (16 Å^2). The results suggest that there is no less heterogeneity in the myohemerythrin crystal but rather that the high thermal parameters serve to make less of the existing heterogeneity observable. This is supported by the fact that the heterogeneous side chains that are observed in myohemerythrin have a significantly larger separation of conformers than those found in the other proteins. Local heterogeneity in the protein molecule may be hidden by the loose packing of the molecule in the crystal lattice.

Discrete Substates vs. Continuous Disorder. These crystallographic observations of discrete heterogeneity are possible because the flexible groups spend virtually all the time in a few localized positions. Continuous heterogeneity, as evidenced by lack of electron density or very large thermal parameters, is observed much less frequently. There are two such side chains in lamprey hemoglobin, seven in myohemerythrin, two in erabutoxin, and none in crambin. This preponderance of discrete over continuous heterogeneity indicates that, for displacements of the magnitude seen in the cases presented here, protein dynamics are best veiwed as a population of conformational substates. These observations allow a lower bound to be placed on the number of substates available to each protein molecule in solution. Taking interdependence of heterogeneity into account, we infer 192 lattice-independent substates for crambin, 512 for erabutoxin, 64 for myohemerythrin, and 512 for lamprey hemoglobin. In solution, further conformational perturbations might occur that are prevented in the crystal by lattice contacts. The flexible hexapeptide in erabutoxin, which forms no hydrogen bonds or salt bridges in the crystal, may be an example of the more extensive perturbations that could occur in solution.

The very closely spaced conformers of Phe-13 in crambin (Figure 1) were an unexpected result. The χ_1 torsional rotations are trans for both conformers, differing by only 11°, and the χ_2 torsions are virtually identical. This suggests that even very small conformational perturbations may be modeled best as discrete substates. Such substates would not be resolved in the three other protein crystals discussed here, just as the Phe-13 conformers were not resolved until data to at least 1.0-Å spacings were included in the analysis. Unresolved discrete heterogeneity would be expected to increase the thermal parameters of affected atoms. Evidence for this is found in the thermal parameters for terminal atoms of long side chains relative to those for their respective C_β atoms. Extreme thermal "explosion" along side chains is prevented by the stereochemical restraints that are applied to thermal parameters during refinement (Konnert & Hendrickson, 1980). The ratio of thermal parameters for the terminal atom(s) to those of C_β in the long side chains Lys, Arg, Glu, and Gln never exceeds 2.50, with an average value of 1.51 for 95 such residues in our four protein models. Unresolved discrete disorder would be expected to increase this thermal-parameter ratio for side chains where heterogeneity has not been modeled. This is indeed the case. The average terminal atom to C_β thermal-parameter ratio is 1.64 for the 56 nonheterogeneous Lys, Arg, Glu, and Gln side chains and is 1.23 for the 38 heterogeneous Lys, Arg, and Glu side chains.

Dynamic or Static Heterogeneity? The question naturally arises whether these substates exist dynamically in the crystal or are examples of lattice disorder where an average over all the static crystalline unit cells is observed. One crystallographic approach to this problem is to see which conformers can be moved to their alternatives in the crystal without prohibitively short contacts with nonheterogeneous parts of the model. In nearly all cases, if it is assumed that the ordered solvent is in equilibrium with the continuum, such conformational changes can be made in the crystals. Possible exceptions are Leu-38 in lamprey hemoglobin and the Arg-33, Asn-62 pair in erabutoxin. Leu-38 is a buried side chain and may not be able to move between conformers without bad contacts at the C_δ atoms. Arg-33 and the C-terminal Asn-62 form a salt bridge in one state of erabutoxin. This is a very tight lattice contact, perhaps frozen in some pairs of molecules during crystallization while others formed the alternative intramolecular salt bridges. Slight amounts of molecular "breathing" during short transition times would allow both of these conformational changes to occur.

The diffraction data for crystals of these four proteins were all measured at room temperature. Frauenfelder et al. (1979) and Hartmann et al. (1982) have shown that reducing the temperature of the crystallographic experiment reduces the thermal parameters of the refined model. The myohemerythrin result, by comparison with the other proteins, implies that large thermal parameters disguise heterogeneity, and we suggest that falsely large, "exploded" thermal parameters do the same. Data collection at low temperature should facilitat observation of heterogeneity by lowering thermal parameter and also may allow analysis at higher resolution. For flexibility that is dynamic, slow cooling of crystals may eliminate some conformers at higher energy as they become depopulated. Studies at various temperatures may pinpoint groups that are statically disordered.

Observability of Substates in Protein Crystals. Remarkably few cases of heterogeneity in refined protein models have been reported in the literature. Phillips (1980) in his refinement of oxymyoglobin at 1.6-Å resolution modeled each of four side chains in two alternative conformations and showed by calculating nonbonded energy that two potential wells exist for each of these side chains. The model of actinidin determined at 1.7-Å resolution (Baker, 1980) contains five discretely disordered side chains as does that for *Escherichia coli* dihydrofolate reductase at the same resolution (Bolin et al., 1982). Two examples are reported in trypsin inhibitor at 0.94-Å resolution (Wlodawer et al., 1984). It is not clear whether heterogeneous models were refined for all these proteins. In several cases side-chain flexibility has been noted in electron density maps but not modeled [for example, Almassy et al. (1983), Arnone et al. (1982) and Glover et al. (1983)]. The parts of protein molecules implicated in biological function generally receive close scrutiny during a structure analysis, and heterogeneity has been reported in catalytic side chains of enzymes (James et al., 1980; Borkakoti et al., 1982), in a bound ligand molecule (Quiocho & Vyas, 1984), in active site ions or water molecules (Rees et al., 1983; Bolognesi et al., 1982), and in a critical subunit interface (Shaanan, 1983). At least one case of heterogeneity in amino acid sequence has been reported also to have evidence in electron density maps (Steigemann & Weber, 1979).

The results on our four proteins suggest that conformational heterogeneity in protein crystals is widespread and should be detectable in structure analyses at 2-Å resolution or better. These four proteins each contain a single one- or two-layer folding domain. While their small size facilitated the task of sorting out the structural heterogeneity, we expect that observations like those reported here could also be made for proteins with larger domains. Most of the structural heterogeneity was observed in the course of systematic inspection of $|F_o| - |F_c|$ difference Fourier maps in an effort to locate ordered solvent and to correct the remaining errors in the protein models. These are not cases of missing electron density. For example, eight heterogeneous side chains in erabutoxin were built into density of height from 0.26 to 0.52 e/Å^3 (average value 0.33 e/Å^3). While none of these were modeled until more than 40 waters were included, all but one had density (average height of 0.33 e/Å^3) in the $|F_o| - |F_c|$ map calculated before any solvent was added to the model. In all the proteins, virtually all significant $|F_o| - |F_c|$ peaks that were too near the protein models to be solvent could be interpreted as stereochemically sensible heterogeneity. Fourier syntheses of the type $2|F_o| - |F_c|$ often did not suggest heterogeneity. The tight stereochemical restraints applied in the refinement of thermal parameters certainly affected the appearance of $|F_o| - |F_c|$ density and may have facilitated interpretation of heterogeneity. The interpretation was aided further by the convenience of a molecular graphics system in our laboratory that was connected via a fast link to the computer where refinements were carried out. Frequent inspection of electron density maps was very efficient, while the peak inspection programs focused our graphics attention on those areas of the models and maps tht needed interpretation. Finally, the differences in heterogeneity observed in these four protein cyrstals certainly reflect the fact that the observations were made, each structure essentially independently, by four investigators.

Acknowledgments

We thank Alex Rashin for providing a computer program to locate cavities in protein molecules and Michael Connolly for pointing out to us the cavity in crambin.

Registry No. Erabutoxin, 59536-69-5.

References

Almassy, R. J., Fontecilla-Camps, J. C., Suddath, F. L., & Bugg, C. E. (1983) *J. Mol. Biol. 170*, 497–527.

Arnone, A., Briley, P. D., Rogers, P. H., & Hendrickson, W. A. (1982) in *Hemoglobin and Oxygen Binding* (Ho, C., Ed.) pp 127–133, Elsevier/North-Holland, Amsterdam.

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.

Baker, E. N. (1980) *J. Mol. Biol. 141*, 441–484.

Barry, C. D., Molnar, C. E., & Rosenberger, F. U. (1976) *Technical Memorandum No. 229*, Computer Systems Laboratory, Washington University, St. Louis, MO.

Bennett, W. S., & Huber, R. (1984) *CRC Crit. Rev. Biochem. 15*, 291–384.

Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C., & Kraut, J. (1982) *J. Biol. Chem. 257*, 13650–13662.

Bolognesi, M., Gatti, G., Menegatti, E., Guarneri, M., Marquart, M., Papamokos, E., & Huber, R. (1982) *J. Mol. Biol. 162*, 839–868.

Borkakoti, N., Moss, D. S., & Palmer, R. A. (1982) *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. B 38*, 2210–2217.

Bott, R. R., Navia, M. A., & Smith, J. L. (1982) *J. Biol. Chem. 257*, 9883–9886.

Bourne, P. E., Sato, A., Corfield, P. W. R., Rosen, L. S., Birken, S., & Low, B. W. (1985) *Eur. J. Biochem. 153*, 521–527.

Englander, S. W., & Kallenbach, N. R. (1983) *Q. Rev. Biophys. 16*, 521–655.

Frauenfelder, H., Petsko, G. A., & Tsernoglou, D. (1979) *Nature (London) 280*, 558–563.

Glover, I., Haneef, I., Pitts, J., Wood, S., Moss, D., Tickle, I., & Blundell, T. (1983) *Biopolymers 22*, 293–304.

Hartmann, H., Parak, F., Steigemann, W., Petsko, G. A., Ponzi, D. R., & Frauenfelder, H. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4967–4971.

Hendrickson, W. A. (1981) in *Refinement of Protein Structures* (Machin, P. A., Campbell, J. W., & Elder, M., Eds.) pp 1–7, Science and Engineering Research Council, Daresbury, U.K.

Hendrickson, W. A. (1985) in *Molecular Dynamics and Protein Structure* (Hermans, J., Ed.) pp 75–76, University of North Carolina, Chapel Hill.

Hendrickson, W. A., & Konnert, J. H. (1980) in *Computing in Crystallography* (Diamond, R., Ramaseshan, S., &

Venkatesan, K., Eds.) pp 13.01–13.23, Indian Academy of Sciences, Bangalore.
Hendrickson, W. A., & Teeter, M. M. (1981) *Nature (London) 290*, 107–113.
Hendrickson, W. A., Klippenstein, G. L., & Ward, K. B. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2160–2164.
Hombrados, I., Rodewald, K., Neuzil, E., & Braunitzer, G. (1983) *Biochimie 65*, 247–257.
Honzatko, R. B., Hendrickson, W. A., & Love, W. E. (1985) *J. Mol. Biol. 184*, 147–164.
James, M. N. G., Sielecki, A. R., Brayer, G. D., Delbaere, L. T. J., & Bauer, C. A. (1980) *J. Mol. Biol. 144*, 43–88.
Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268–272.
Jones, T. A. (1982) in *Computational Crystallography* (Sayre, D., Ed.) pp 303–317, Clarendon Press, Oxford.
Karplus, M., & McCammon, J. A. (1981) *CRC Crit. Rev. Biochem. 9*, 293–349.
Karplus, M., & McCammon, J. A. (1983) *Annu. Rev. Biochem. 52*, 263–300.
Konnert, J. H. (1976) *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. A 32*, 614–617.
Konnert, J. H., & Hendrickson, W. A. (1980) *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr. A 36*, 344–350.
Levitt, M., Sander, C., & Stern, P. S. (1985) *J. Mol. Biol. 181*, 423–447.
Li, S. L., & Riggs, A. (1970) *J. Biol. Chem. 245*, 6149–6169.
Phillips, S. E. V. (1980) *J. Mol. Biol. 142*, 531–554.
Quiocho, F. A., & Vyas, N. K. (1984) *Nature (London) 310*, 381–386.
Rees, D. C., Lewis, M., & Lipscomb, W. N. (1983) *J. Mol. Biol. 168*, 367–387.
Shaanan, B. (1983) *J. Mol. Biol. 171*, 31–59.
Sheriff, S., & Hendrickson, W. A. (1986) *Acta Crystallogr.* (in press).
Sheriff, S., Hendrickson, W. A., Stenkamp, R. E., Sieker, L. C., & Jensen, L. H. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1104–1107.
Steigemann, W., & Weber, E. (1979) *J. Mol. Biol. 127*, 309–338.
Steitz, T. A., Harrison, R., Weber, I. T., & Leahy, M. (1983) in *Ciba Foundation Symposium 93: Mobility and Function in Proteins and Nucleic Acids*, pp 25–46, Pitman, London.
Teeter, M. M., & Hendrickson, W. A. (1979) *J. Mol. Biol. 127*, 219–223.
Teeter, M. M., Mazer, J. A., & L'Italien, J. J. (1981) *Biochemistry 20*, 5437–5443.
Wagner, G. (1983) *Q. Rev. Biophys. 16*, 1–57.
Weiss, M. A., Sauer, R. T., Patel, D. J., & Karplus, M. (1984) *Biochemistry 23*, 5090–5095.
Wlodawer, A., Walter, J., Huber, R., & Sjolin, L. (1984) *J. Mol. Biol. 180*, 301–329.