Crystal Structure of Ser-22/Ile-25 Form Crambin Confirms Solvent, Side Chain Substate Correlations*

(Received for publication, January 3, 1997, and in revised form, February 5, 1997)

Akihito Yamano, Nam-Ho Heo, and Martha M. Teeter

From the X-ray Research Laboratory, Rigaku Corporation, 3-9-12 Matsubara, Akishima, 196 Tokyo, Japan, the Department of Industrial Chemistry, Kyungpook National University, Taegu, 702-701 Korea, and the Merkert Chemistry Center, Department of Chemistry, Boston College, Chestnut Hill, Massachusetts 02167

It is not agreed that correlated positions of disordered protein side chains (substate correlations) can be deduced from diffraction data. The pure Ser-22/Ile-25 (SI form) crambin crystal structure confirms correlations deduced for the natural, mixed sequence form of crambin crystals. Physical separation of the mixed form into pure SI form and Pro-22/Leu-25 (PL form) crambin and the PL form crystal structure determination (Yamano, A., and Teeter, M. M. (1994) J. Biol. Chem. 269, 13956–13965) support the proposed (Teeter, M. M., Roe, S. M., and Heo, N. H. (1993) J. Mol. Biol. 230, 292–311) correlation model. Electron density of mixed form crambin crystals shows four possible pairs of side chain conformations for heterogeneous residue 22 and nearby Tyr-29 (2^2 = 4, two conformations for each of two side chains). One combination can be eliminated because of short van der Waals’ contacts. However, only two alternatives have been postulated to exist in mixed form crambin: Pro-22/Tyr-29A and Ser-22/Tyr-29B. In crystals of the PL form, Pro-22 and Tyr-29A are found to be in direct van der Waals’ contact (Yamano, A., and Teeter, M. M. (1994) J. Biol. Chem. 269, 13956–13965). Comparison of the SI form structure with the mixed form electron density confirms that the fourth combination of side chains does not occur and that side chain correlations are mediated by water networks.

Motion correlated over 5–8 Å (liquid-like movement) has been shown by the non-Bragg technique of x-ray diffuse scattering to be important in insulin and lysozyme crystals (1, 2). State of the art molecular dynamics methods cannot model such correlations (3), perhaps because of inadequate sampling of conformational substates (4). Multiple substates of nearly equal energy are also proposed for myoglobin based on spectroscopic evidence (5–7), but spectroscopy is not well suited to elucidate the nature of these substates. Neither is NMR, unless extremely tight distance restraints are used (8).

Diffraction from a crystal is averaged over many unit cells and over the time spent on data collection. It is generally believed that this averaging precludes extracting dynamic information, such as occurrence of multiple substate correlations from an x-ray structure. However, nonrandom correlations will contribute to Bragg reflections. Given diffraction data beyond 1.4 Å (9), the correlations can be modeled as substate disorder and provide insight into protein dynamics. If one could physically separate the substates and study each separately, one could prove such correlations exist and derive the rules for the correlation.

Crambin presents an excellent system for such an experiment. Crambin from the natural source contains two sequence isomers in a 3:2 ratio (10, 11), the so-called mixed form of crambin. The major isomer has Pro and Leu at positions 22 and 25, respectively (the PL form); the minor isomer has Ser and Ile at the same positions (the SI form). In the mixed form crystal structure, side chain electron densities for heterogeneous residues are superimposed (Pro and Ser at residue 22 and Leu and Ile at residue 25). The Tyr-29 side chain from a 2^1-screw axis-related molecule has close contacts with the Pro or Ser residue and adopts two conformations. A proposed correlation of the Tyr-29 conformation with the identity of the amino acid at residue 22 (12) has been supported by the PL form structure (13). Now the second or SI form of crambin has been purified by fast protein liquid chromatography and crystallized. It establishes the side chain correlations definitively and establishes associated solvent interactions.

In this paper, first the proposed mixed form protein networks are extended to water disorder using stereochemical “rules,” such as van der Waals’ contacts and hydrogen bonding. Second, these postulated networks are compared with the crystal structures of the physically separated pure forms of crambin: the PL form structure and the newly determined SI form structure. These results establish that the x-ray structure of the mixed form of crambin can elucidate substrate spatial correlations between side chains and solvent, as proven by the pure form structures.

Alternative conformations at disordered residues and water molecules are designated by attaching A and B to the residue number. Such disordered conformations are often correlated with neighboring residue disorder through space and may represent conformational substates of the protein. For example, Ser-22A, Tyr-29A, and Wat-132A represent one disordered substate correlated through space with the alternates Ser-22B, Tyr-29B, and Wat-132B.

Crambin was purified to a single sequence form (13), and crystals of the SI form were grown by vapor diffusion techniques (14). Conditions were similar to those previously used (15) but with an initial reservoir concentration of 50% ethanol. In contrast to other forms, seeding by methods such as the streak seeding technique (16) was essential to nucleate crystal growth. Here a submicroscopic mixed form crystal served as the seed crystal, and small crystals appeared along the streak.

* This work was supported by National Science Foundation Grants DMB 89-04337 and MCB-9219857 (to M. M. T.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 The abbreviations used are: PL, major isomer containing Pro-22 and Leu-25; SI, minor isomer containing Ser-22 and Ile-25; Wat, water; rms, root mean square.
Table I
Crystallographic data for crambin crystals

|                          | Mixed form (130 K) | PL form (150 K) | SI form (150 K) |
|--------------------------|-------------------|----------------|----------------|
| Space group              | P2₁              | P2₁            | P2₁            |
| a (Å)                    | 40.76            | 40.62          | 40.76          |
| b (Å)                    | 18.49            | 18.34          | 18.40          |
| c (Å)                    | 22.53            | 22.14          | 22.27          |
| β (°)                    | 90.61            | 91.07          | 90.7           |
| V (Å³)                   | 16833.4          | 16495.3        | 16706.2        |
| Resolution               | 0.83 Å           | 1.05 Å         | 0.89 Å         |
| Temperature              | 130 K            | 150 K          | 150 K          |
| Completeness (>2σ(Fo))  | 72.6%            | 85.1%          | 75.8%          |
| Luzzati error (Å)        | 0.08             | 0.06           | 0.06           |
| Standard R               | 10.5%            | 9.5%           | 14.7%          |

a Mixed form refers to crystals of the mixed sequence form of crambin, where residue 22 is either Pro or Ser and residue 25 is Ile or Leu.
b PL form refers to the Pro-22/Leu-25 pure sequence form of crambin.
c SI form is the Ser-22/Ile-25 pure sequence form of crambin.

Table II
Summary of the PROLSQ refinement residuals and weights for SI form crambin

| Stereological feature | rms deviation from ideality | Target variance |
|-----------------------|----------------------------|-----------------|
| Covalent bonds (Å)    | 0.014                      | 0.020           |
| Angle distance (Å)    | 0.033                      | 0.040           |
| Planar 1–4 distances (Å) | 0.055                    | 0.050           |
| H bond                | 0.010                      | 0.006           |
| Least squares planes (Å) | 0.010                    | 0.010           |
| Volume at chiral centers (Å³) | 0.063                | 0.050           |
| Single torsion contacts (DINC = −0.3°) | 187 at 0.158 A | 0.500           |
| Multiple torsion contacts (DINC = 0) | 258 at 0.160 A | 0.500           |
| Possible H bond contacts (DINC = −0.2°) | 73 at 0.158 A | 0.500           |
| Planar torsions (°)   | 5.1                        | 3.0             |
| Staggered torsions (°) | 10.2                       | 15.0            |
| Orthorotational torsions (°) | 25.6                  | 20.0            |
| Structure factor weighta | 9.3, −14.1                | 8.0, −10.0      |
| Agreement with diffraction data | Standard R factorb | 14.7            |
| Error R factorc | 7.0                        |                 |
| Error R-factord | 5.94                       |                 |

a DINC is the change in the minimum van der Waals’ contact distance.
b The weight for the structure factors in refinement (the “target” σ of |F_o|−|F_c|) was modeled by the function σ = (3σo) with σ = 8.0 + (11.0 − |F_o|/16). The fitted (|F_o|−|F_c|) 9.3 + (−14.1) × (sin(|F_o|/16)).
c Standard R factor = Σ| |F_o|−|F_c||/Σ |F_o|.
d Error R-factor = Σ σ_o |F_o|−2|F_c|/|F_o|.

Ser-22/Tyr-29B, 3) Pro-22/Tyr-29A, and 4) Pro-22/Tyr-29B. Because of a short Van der Waals’ contact, the fourth choice can be excluded immediately (the distance from O₆ of Tyr-29B to C₂ of Pro-22 is only 2.41 Å).

Tyr-29A forms a slightly short hydrogen bond to Wat-182, and Wat-182A makes a hydrogen bond to Wat-82. But Tyr-29B O₆ forms hydrogen bonds to either Wat-132A or Wat-132B as well as to Wat-47 in ring A of the pentagon water ring cluster (23). Wat-132A or Wat-132B hydrogen bonds to the backbone N of residue 22. However, this is only possible with the Ser side chain or site Tyr-29B because of the Pro C₃N covalent bond. Short contacts with these waters would result with Pro-22 or Tyr-29A (Wat-132A-C₂, 1.47 Å, Wat-132B-C₂, 1.52 Å, see Fig. 2).

The three remaining networks can be extended to include hydrogen-bonded waters (branches are indicated in parenthesis): 1) Ser-22/Tyr-29A/Wat-182A/Wat-82, 2) Ser-22/Wat-132/ (Wat-182B/Tyr-29B (the red network in Fig. 2); and 3) Pro-22/Tyr-29A/Wat-182A/Wat-82 (the green network in Fig. 2.)

Based on this analysis, one would predict for the pure form structures that waters associated with the missing form would have altered occupancies. Key would be the weak density sites Wat-123A/B. They should be considerably stronger in the SI structure, the water-protein conformations perfectly match the green network in Fig. 2. The rms deviation from the mixed form structure is 0.065 Å over Tyr-29A, Pro-22, Wat-47, Wat-82, and Wat-182A.

Fig. 4 shows the electron density and the atomic model for the pure SI form structure at the same region that is shown in Fig. 2. The electron density is consistent with the elimination of the Ser and Ile side chains. The density at residue 22 matches Pro. The phenolic ring of residue 29 takes the A conformation and Tyr-29A O₆ makes allowed van der Waals’ contacts with Pro-22 C₂, and O₆. Water sites Wat-132A and Wat-132B are absent in this structure. The water-protein conformations perfectly match the green network in Fig. 2. The rms deviation from the mixed form structure is 0.065 Å over Tyr-29A, Pro-22, Wat-47, Wat-82, and Wat-182A.

Considering only protein atoms and a single Ser conformation, four possible combinations exist: 1) Ser-22/Tyr-29A, 2)
Tyr-29 takes the Tyr-29B conformation, and waters 132A/132B are enhanced as predicted. This structure is nearly identical to the red network in Fig. 2, except for Wat-182B. The rms deviation between this and the mixed form structure is 0.267 Å for Tyr-29B, Ser-22A/B, Wat-47, Wat-132A/B, and Wat-182B. An additional water site (Wat-182C) could be modeled in Fig. 2 (elongated density on 182A), because an additional water site is visible from the Ser/Ile structure (Wat-182A).

From the above comparisons, one can conclude that interpenetrating disorder networks can be separated by optimizing van der Waals’ contacts and hydrogen bonds. Because networks 2 and 3 account for all the electron density in the mixed form crystal, these are the only networks needed to account for the mixed form disorder.

Why is the network 1 not present in nature? Stereochemical requirements alone cannot exclude this possibility, because it neither violates van der Waals’ contact limits nor has inappropriate hydrogen bonds. However, if the phenol ring of residue 29 took the Tyr-29A conformation and the side chain of residue 22 were Ser, there would be a large vacancy around Wat-132A, Wat-132B, and Pro-22 C. The potential empty space is eliminated by the spatial correlation among side chains and water molecules. In other words, space or vacuum is not allowed at a protein surface, probably because it is energetically unfavorable.

Further, in the SI form structure, this space filling can be seen from the alternate water conformations identified in electron density maps. Wat-182A is shifted downward (Fig. 4) to fill the empty space created by the absence of the Tyr-29A conformation. Another water site (182B) alternates with this site 2.31 Å away and hydrogen bonds to Wat-132. Wat-132A/B disorder appears for similar reasons. Both alternate pairs fill the available space and optimize packing and hydrogen bonding.

From these disordered water molecules, the importance of solvent for protein flexibility is evident. The full rationalization of the correlations derived from Fig. 2 must involve solvent-mediated interactions.

Proposed disorder networks in the mixed form crambin are extended to solvent and confirmed by the pure Pro-22/Leu-25 (13) and Ser-22/Ile-25 forms of crambin. Here the two disordered forms resulting from sequence differences were physi-
cally separated by fast protein liquid chromatography, and each was crystallized. The spatial correlations implied from the mixed form structure were proven by examining both protein and water from the two pure form structures. Water was critical for this confirmation.

Derived rules for correlations provide insight into the structure and dynamics of proteins in general. In this paper, we have proven that correlated conformations obey simple stereochemical rules and have alternates that fill space. The same logic used here should apply to assigning multiple conformational substates where sequence differences are not involved (13, 24). These results demonstrate that dynamic correlation does occur and can be deduced from an x-ray structure at 1 Å resolution using fundamental principles. Such elucidation is important for understanding the mechanisms of such important proteins as lysozyme and myoglobin.

Acknowledgments—The crystal structure determinations of the pure forms of crambin were initiated with a sample of the mixed form of crambin, which was extracted and purified by Hucheng Bei, whose work is gratefully acknowledged. Thanks is due to Ofer Markman, who assisted with the figures. We thank Jack Dunitz and Boguslaw Stec for helpful discussions.

REFERENCES
1. Caspar, D. L. D., Clarage, J. B., Salunke, D. M., and Clarage, M. S. (1988) Nature 332, 659–662
2. Clarage, J. B., Clarage, M. S., Phillips, W. C., Sweet, R. M., and Caspar, D. L. D. (1992) Proteins 12, 145–157
3. Clarage, J. B., Romo, T., Andrews, B. K., Pettit, B. M., and Phillips, G. N., Jr. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 3288–3292
4. Caspar, D. L. (1995) Structure 3, 327–329
5. Frauenfelder, H., Sligar, S. G., and Wolynes, P. G. (1991) Science 254, 1598–1603
6. Nienhaus, G. U., Mourant, J. R., and Frauenfelder, H. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 2902–2906
7. Nienhaus, G. U., Mourant, J. R., Chu, K., and Frauenfelder, H. (1994) Biochemistry 33, 13413–13430
8. Bonvin, A. M., and Brunger, A. T. (1996) J. Biomol. NMR 7, 72–76
9. Smith, J. L., Hendrickson, W. A., Honzatko, R. B., and Sheriff, S. (1986) Biochemistry 25, 5018–5027
10. Hendrickson, W. A., and Teeter, M. M. (1981) Nature 290, 109–113
11. Teeter, M. M., Mazer, J. A., and L'Italien, J. J. (1981) Biochemistry 20, 5437–5443
12. Teeter, M. M., Roe, S. M., and Heo, N. H. (1993) J. Mol. Biol. 230, 292–311
13. Yamano, A., and Teeter, M. M. (1994) J. Biol. Chem. 269, 13956–13965
14. McPherson, A. (1982) in Preparation and Analysis of Protein Crystals, pp. 94–96, John Wiley & Sons, New York
15. Teeter, M. M., and Hendrickson, W. A. (1979) J. Mol. Biol. 127, 219–223
16. Stura, E. A., and Wilson, I. A. (1992) in Crystallization of Nucleic Acids and Proteins (Ducruix, A., and Giege, R., eds) pp. 112–113, Oxford University Press, New York
17. Hope, H. (1988) Acta Crystallogr. Sect. B Struct. Sci. 44, 22–26
18. Hendrickson, W. A., and Konnert, J. H. (1980) in Computing in Crystallography, (Diamond, R. S., and Venkatesan, K., eds) pp. 13.01–13.23, Indian Academy of Sciences, Bangalore, India
19. Jones, T. A. (1985) in Methods Enzymol. 115, 157–171
20. Hendrickson, W. A., and Konnert, J. H. (1980) Acta Crystal. Sec. A 36, 344–350
21. Luzzati, V. (1952) Acta Crystallogr. 5, 802–810
22. Stec, B., Zhou, R., and Teeter, M. M. (1995) Acta Crystallogr. Sec. D 51, 663–681
23. Teeter, M. M. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 6014–6018
24. Teeter, M. M. (1991) in Annual Review of Biophysics and Biophysical Chemistry, pp. 577–600, Annual Reviews Inc., Palo Alto, CA
Crystal Structure of Ser-22/Ile-25 Form Crambin Confirms Solvent, Side Chain Substate Correlations
Akihito Yamano, Nam-Ho Heo and Martha M. Teeter

J. Biol. Chem. 1997, 272:9597-9600.
doi: 10.1074/jbc.272.15.9597

Access the most updated version of this article at http://www.jbc.org/content/272/15/9597

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 20 references, 5 of which can be accessed free at http://www.jbc.org/content/272/15/9597.full.html#ref-list-1