Common Structural Framework of the Two Ca²⁺/Mg²⁺ Binding Loops of Troponin C and Other Ca²⁺ Binding Proteins[†]

Osnat Herzberg and Michael N. G. James*

Medical Research Council of Canada Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

Received June 19, 1985

ABSTRACT: The refinement of the crystal structure of turkey skeletal muscle troponin C at 2.2-Å resolution reveals that the two calcium binding loops that are occupied by Ca^{2+} ions adopt conformations very similar to those of the two homologous loops of parvalbumin and to that of loop III-IV of the intestinal calcium binding protein. This specific fold assures suitable spatial positioning of the Ca^{2+} ligands. It consists of two reverse turns, one located at each end of the loop, and four Asx turns (a cyclic hydrogen-bonded structure involving an oxygen of the side chain of residue n and the main-chain amide nitrogen of residue n + 2) whenever such a side chain coordinates to the metal ion. The fifth Ca^{2+} coordination position in both loops of troponin C is occupied by a water molecule that is within hydrogen-bonding distance of an aspartic acid, thus mediating indirect interaction between the cation and the negatively charged carboxylate. The same loop framework is conserved in the two Ca^{2+} binding loops of parvalbumin and loop III-IV of the intestinal Ca^{2+} binding protein in spite of the variability in the nature of the side chains at equivalent positions. The disposition of the Ca^{2+} and of its coordinating water molecule relative to the protein main chain is conserved in all these cases.

The intracellular Ca²⁺ binding proteins exhibit a variety of affinities to Ca²⁺, with binding constants in the range 10⁵-10⁸ M⁻¹ (Levine & Williams, 1982). The molecular interactions responsible for this range of affinities are as yet unknown. The primary structural homology among these proteins was noted as soon as sequence information became available (Collins et al., 1973; Kretsinger & Barry, 1975), and the homology continues to be detected as more sequences accumulate (Gariépy & Hodges, 1983; Kobayashi et al., 1984). This sequence homology and the results of the crystal structure study of carp parvalbumin (Kretsinger & Nockolds, 1973) led to the prediction that there would be a common Ca²⁺ binding motif for TnC, 1 CaM, ICaBP, and other related Ca2+ binding proteins. The crystallographic structure determinations of ICaBP (Szebenyi et al., 1981), turkey and chicken TnC (Herzberg & James, 1985; Sundralingam et al., 1985), and CaM (Babu et al., 1985) have supported these proposals.

In the calcium-bound state, all the calcium binding units consist of a helix-loop-helix structure in which the helices are roughly perpendicular to each other and the Ca²⁺ ion binds to oxygen ligands from the loop. This motif has been designated the EF hand (Kretsinger & Nockolds, 1973). The Ca2+ binding loop consists of 12 sequential residues, of which those at positions 1, 3, 5, 7, 9, and 12 are involved in coordination to the cation (see Figure 1). The cation is octahedrally coordinated by oxygen atoms that are located at the axial positions x, y, z, -y, -x, -z in a Cartesian coordinate framework corresponding to the sequence of residues given above [see Kretsinger & Nockolds (1973) for a more complete description]. The majority of the oxygen atoms in the ligand field of the Ca²⁺ ion are from Asx or Glu side chains. However, the residue in position 7 contributes its main-chain carbonyl oxygen atom, and therefore this side chain is the least conserved. Position 9 of the loop also exhibits variability: some sequences have Asp residues at this position, Parv has Gly-98 in the EF loop and Glu-59 in the CD loop, and ICaBP has Ser-62 in the III-IV loop. Direct coordination to the Ca²⁺ ion by the carboxyl group of Glu-59 in Parv is observed, but for those loops with shorter side chains two possibilities exist: (a) a water molecule may occupy the fifth metal coordination site or (b) in cases involving an oxygen-containing side chain there may be an adjustment of the loop main chain to bring this side chain close enough for direct coordination. It is arrangement a and not arrangement b that occurs in the EF loop of Parv (Kretsinger & Barry, 1975) and in the loop III-IV of ICaBP (based on the coordinates deposited by Szebenyi and Moffat at the Brookhaven Protein Data Bank; Bernstein et al., 1977).

In TnC, position 9 of both high-affinity Ca^{2+}/Mg^{2+} binding loops is occupied by an aspartic acid residue. The multiple isomorphous replacement electron density map (Herzberg & James, 1985) indicated different conformations of these residues such that we interpreted Asp-114 in loop III so that it coordinated directly to the Ca^{2+} ion whereas in loop IV the side chain of Asp-150 was $\simeq 6.3$ Å from the calcium. We suggested that a water molecule might coordinate directly to the Ca^{2+} in the second case. Least-squares refinement of the turkey TnC structure at 2.2-Å resolution has defined the high-affinity Ca^{2+}/Mg^{2+} binding loops more sharply. This paper describes the picture that has emerged and relates the TnC binding loops to the other known structures.

MATERIALS AND METHODS

The crystal structure of turkey skeletal TnC was determined initially at 2.8-Å resolution by the multiple isomorphous replacement method (Herzberg & James, 1985). The molecular model obtained was then refined by the restrained-parameter

[†]Supported by the Medical Research Council of Canada grant to the Group in Protein Structure and Function and by an Alberta Heritage Foundation for Medical Research award to O.H.

¹ Abbreviations: TnC, troponin C; Parv, parvalbumin; ICaBP, vitamin D dependent intestinal calcium binding protein; CaM, calmodulin.

reciprocal space least-squares procedure (Hendrickson & Konnert, 1980) at 2.8-Å resolution for ten cycles to an R factor² of 0.30. Higher resolution data up to 2.2 Å have been added gradually and after an additional 55 cycles, the R factor has dropped to a value of 0.20 for the 8.0–2.2-Å data [6435 reflections with $I > 2\sigma(I)$]. The model at this stage includes most of the polypeptide chain (the first two residues and the last one are apparently too disordered to identify), two Ca²+ ions bound to the C-terminal domain, and 21 water molecules. Details of the results of the refinement will be published after completion of the process.

Coordinates for Parv and ICaBP were obtained from the Brookhaven Protein Data Bank (Bernstein et al., 1977). In the case of Parv, set 6A was used since it conforms more with commonly accepted protein geometry. Comparisons of the geometry of the Ca²⁺ binding loops of TnC, ICaBP, and Parv have been made by using a least-squares minimization procedure. In these comparisons, all of the atoms that comprise the main chain of the 12 residue loops were included. The loops were superposed in pairs. The resulting rotation matrix and translation vector were applied to the atoms of the whole fragment including the side chains, Ca²⁺ ions, and water molecules so that the results could be viewed in the MMS-X interactive graphics system using the program M3 written by Colin Broughton (Sielecki et al., 1982).

RESULTS

Restrained-parameter least-squares refinement of TnC at 2.2-Å resolution has improved the reliability of the molecular model relative to the initial 2.8-Å structure. We report here results that are unlikely to change in substance during further refinement. The reliability is indicated not just by the R value of 0.20 but also by the small root mean square (rms) deviation from accepted protein stereochemistry (0.027 Å for bond distances, 0.048 Å for bond angle distances, 0.009 Å for deviations from planarity of planar groups, and 2.0° for deviation from 180° for the ω angle of the peptide groups). Loops III and IV of TnC are well-defined at this stage; the main-chain atoms and Ca²⁺ coordinating ligands have some of the lowest temperature factors in the molecule ($B \simeq 6$ –17 Ų).

Figure 1 shows the amino acid sequence alignment for the Ca²⁺ binding loops of TnC, Parv, ICaBP, and CaM. The structural implication of amino acid substitutions at each position has been carefully discussed (Kretsinger, 1980). In general, the most conserved residues in these loops are those that have side-chain oxygen atoms coordinating directly to the Ca²⁺ ion (i.e., at positions 1, 3, 5, and 12). In addition, positions 6 and 8 are highly conserved because of structural constraints: (a) a glycine at position 6 is required to sample the ϕ, ψ space in a region not allowed for a C^{β} -containing amino acid $[(\phi,\psi) \simeq (90^{\circ},0^{\circ})]$, thus enabling both the side chain at position 5 and the carbonyl main chain at position 7 to coordinate to the Ca2+ ion, and (b) the residue in position 8 is involved in a β -sheet hydrogen-bonding interaction between adjacent calcium binding loops (see later). The highly conserved aliphatic hydrophobic side chain is needed for the construction of the hydrophobic core of the domain.

The sequence variability at positions 7 and 9 was already described in the introduction. The electron density map of TnC computed with the refined phases at cycle 65 clarifies the structural basis for the variability at position 9. It reveals a peak in the fifth coordination site of each of the Ca²⁺ ions

```
8
                                                 9
TnC
              30Asp Ala Asp Gly Gly Gly Asp Ile Ser Thr Lys Glu.
             .. Asp Glu Asp Gly Ser Gly Thr Ile Asp Phe Glu Glu,,
 Loop
      11
             , . . Asp Lys Asn Ala Asp Gly Phe Ile Asp Ile Glu Glu, , ,
 Loop III
             142Asp Lys Asn Asn Asp Gly Arg Ile Asp Phe Asp Glu153
 Loop
      ΙV
             siAsp Gln Asp Lys Ser Gly Phe Ile Glu Glu Asp Glusz
 Loop
      CD
 Loop
             .oAsp Ser Asp Gly Asp Gly Lys Ile Gly Val Asp Glu:o:
I CaBF
             , Ala Lys Glu Gly Asp Pro Gln Leu Ser Lys Glu Gluz,
 Loop I-II‡
 Loop III-IV saAsp Lys Asn Gly Asp Gly Glu Val Ser Phe Glu Gluss
             20Asp Lys Asp Gly Asn Gly Thr Ile Thr Thr Lys Glu31
 Loop
             seAsp Ala Asp Gly Asn Gly Thr Ile Asp Phe Pro Glus,
 Loop
             .: Asp Lys Asp Gly Asn Gly Tyr Ile Ser Ala Ala Glu.o.
 LOOD III
            , 2. Asn Ile Asp Gly Asp Gly Glu Val Asn Tyr Glu Glu, ac
```

FIGURE 1: Sequence alignment for Ca²⁺ binding loops of TnC (Wilkinson, 1976), Parv (Coffee & Bradshaw, 1973), ICaBP (Fullmer & Wasserman, 1981), and CaM (Watterson et al., 1980). An asterisk indicates a Ca²⁺ ligand that is either determined crystallographically or predicted on the basis of sequence homology. (‡) The coordinates in the Brookhaven Data Bank (Bernstein et al., 1977) are consistent with Ala-14 as a ligand. The publication (Szebenyi et al., 1981) reports Ala-15 as the ligand. This loop will not be used in the following comparisons, since its conformation is reported to be substantially different from that of the EF hand.

(see Figure 2). In light of their relative peak heights (the ratios of the Ca²⁺ peaks to these peaks are 2.2 and 2.6 for loops III and IV, respectively) and the Ca²⁺ to ligand distances of 2.4 and 2.7 Å for loops III and IV, respectively, we interpret these peaks as water molecules. Calcium ion to water molecule distances in small molecule crystal structures average 2.4 Å (Einspahr & Bugg, 1980), and this is also the average distance for Ca²⁺ to carboxylate oxygen atom for small molecules (Einspahr & Bugg, 1981). In the course of the refinement we have not applied distance restraints to any of the Ca²⁺ to ligand oxygen distances and therefore have a relatively large distribution of Ca²⁺—O distances of 2.0–2.7 Å averaging at 2.3 Å.

With a bound water molecule in the fifth ligand site of the Ca²⁺ ion it is not necessary to alter the main-chain conformation such that the shorter aspartic side chains may coordinate directly to the Ca²⁺ ion. The water molecules intervene between the two, with hydrogen-bonding interactions to Asp-114 and Asp-150 of 2.8 and 3.2 Å, respectively. Each of these two water molecules also has hydrogen-bonding interactions with other ligand side chains.

Ca²⁺ binding and structural integrity of the associated loops are assisted by intricate secondary and tertiary structure interactions (Figure 3). There are two reverse turns in each loop. The first, involving loop residues 1-4 (Figure 1), fits the criteria of a type I turn [main-chain torsion angles $[\phi_2, \psi_2]$; ϕ_3,ψ_3] around [-60°,-30°; -90°,0°] as defined by Venkatachalam (1968)], and the second, involving residues 9-12, is most similar to the type III turn (a helical turn with $[\phi_2, \psi_2]$; ϕ_3,ψ_3] around [-60°,-30°; -60°,-30°]). In addition, there are four tertiary interactions in each loop involving hydrogen bonding between an Asx oxygen of the residue at position nand the main-chain NH of the residue at position n + 2(Figures 3 and 4). Such conformations have been termed Asx turns (Rees et al., 1983). As pointed out by Baker & Hubbard (1984), this is too restrictive a term because similar n to n +2 hydrogen bonding occurs with serine, threonine, and cysteinyl residues. The cysteinyl interactions have been found in the

² The R factor is defined as $\sum ||F_0| - |F_c|| / \sum |F_0|$, where $|F_0|$ and $|F_c|$ are the measured and calculated structure factor amplitudes, respectively.

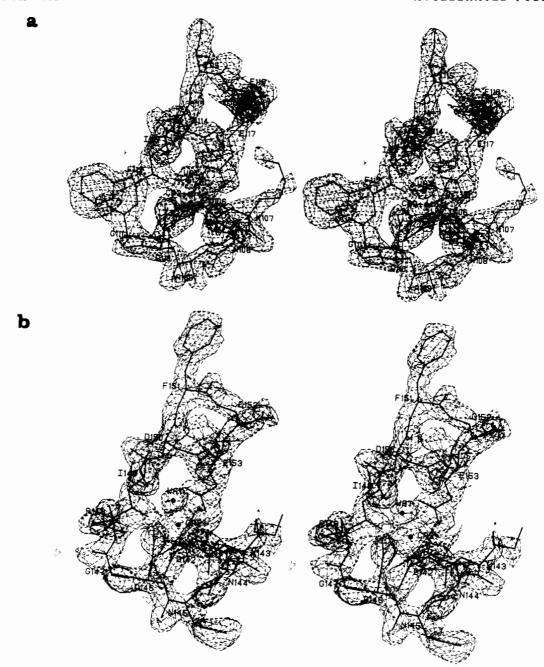


FIGURE 2: Stereoscopic representation of the high-affinity Ca^{2+} binding loops of TnC. Loop III is in (a) and loop IV is in (b). Both are oriented to present similar aspects to the viewer. Superimposed on the molecular model is the electron density map at 2.2-Å resolution. The maps were calculated with coefficients $2m|F_0| - D|F_0|$ and calculated phases (R. J. Read, unpublished results), where D is defined by Luzzati (1952). Contour surfaces are drawn at 0.45 e Å⁻³. Water molecules are labeled as WAT and calcium ions as CA2+. The reliability of these water molecules was confirmed by calculating a difference Fourier map that did not include them in the model.

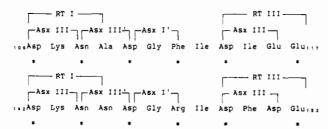


FIGURE 3: Reverse turns and Asx type turns found in the high-affinity Ca²⁺/Mg²⁺ binding loops of TnC. An asterisk indicates those residues that are Ca²⁺ ligands, acting directly or indirectly via a water molecule. RT refers to a reverse turn and Asx to an Asx-type turn. Roman numerals refer to the type of turn as explained in the text.

iron-sulfur proteins. Indeed, as shown in Figure 4a, in the Parv CD loop there is a serine in position 5, yet the n to n + 2 interaction is maintained.

The (ϕ,ψ) values of residue n+1 in an Asx turn are equivalent to (ϕ_3,ψ_3) in a reverse turn. Therefore, Asx turns were categorized by analogy to reverse turns (Rees et al., 1983). A type I Asx turn has (ϕ,ψ) angles of $(-90^{\circ},0^{\circ})$ and a type III $(-60^{\circ},-30^{\circ})$. In TnC we observe two Asx turns of type III and two Asx turns with sign reversals of the above angles. They are denoted I' and III' (Figure 3). Asx I' can also be regarded as an Asx II. Both examples of Asx I' in TnC involve glycine at position n+1 in a manner analogous to a reverse turn II, fulfilling the unique structural role of this residue as discussed earlier.

In addition to the n to n + 2 interactions, the Asx side chains are also stabilized by dipole interactions with one or two more main-chain nitrogen atoms. The disposition of the oxygen atoms in these Asx side chains is assisted by the multiple hydrogen-bonding scheme, thus assuring their correct orien-

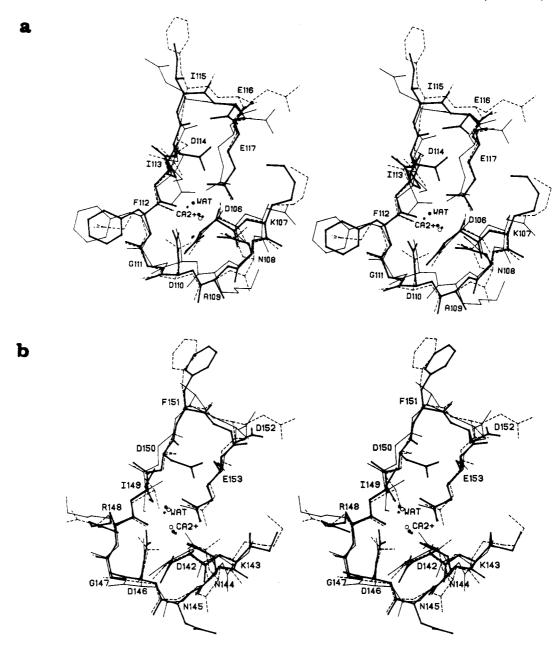


FIGURE 4: Stereoscopic representation of (a) loop III and (b) loop IV of TnC (solid thick lines). Superimposed on them are Parv loop CD (solid thin lines) and ICaBP loop III-IV (dashed lines) in (a) and Parv loop EF (solid thin lines) and ICaBP loop III-IV (dashed lines) in (b). The labels correspond to the sequence of TnC. (•) corresponds to water or Ca²⁺ in TnC, (*) ICaBP, and (O) Parv.

tation relative to the calcium ion. Similarly, the disposition of cysteinyl sulfurs in the coordination field of an iron metal is maintained in a related manner (Adman et al., 1975). In ferredoxin, rubredoxin, and the high potential iron protein, two sulfurs from conserved sequences of the type Cys-X-Y-Cys participate both in coordination to the iron and in n to n+2 interactions with main-chain nitrogens.

Residues 7–9 in each calcium binding loop adopt a conformation similar to that of a β -sheet, with antiparallel hydrogen bonding between the central residues only (in loop III, Ile-113 N forms a hydrogen bond to Ile-149 O of loop IV, 2.7 Å, and Ile-149 N forms a hydrogen bond to Ile-113 O, 3.2 Å). The disposition of the two strands has the commonly observed right-handed twist when viewed parallel to the strand direction. Similar hydrogen-bonded interactions exist in the Ca²⁺-free loops of TnC (Ile-37 N to Ile-73 O distance of 2.9 Å; Ile-73 N to Ile-37 O distance of 2.7 Å). The preservation of the hydrogen-bonded interaction in both the Ca²⁺-free and Ca²⁺-bound loops implies a concerted movement of the loops

Table I: Root Mean Square Differences in Atomic Positions after Superposition of Ca²⁺ Binding Loops^a

		rms discrepancy (Å)
TnC loop III	TnC loop IV	0.47
TnC loop III	Parv loop CD Parv loop EF ICaBP loop III-IV	0.61 0.67 0.55
TnC loop IV	Parv loop CD Parv loop EF ICaBP loop III-IV	0.76 0.55 0.51

 $^{^{}a}$ Forty-eight main-chain atoms of the 12 residues of each loop are included.

at the transition between the two states.

To what extent the conformation is preserved in other loops of known calcium binding protein crystal structures can be seen from Table I, where the root mean square discrepancy between main-chain coordinates of relevant loops superpositioned on the TnC high-affinity loops are given. Loop I-II of

ICaBP was not included because it adopts a very different conformation that has three main-chain carbonyl oxygen atoms coordinating to the calcium. Superpositions of each high-affinity binding site of TnC with the best fitting loop of Parv and with ICaBP loop III-IV are shown in Figure 4. The similarity in conformation between the different loops is striking. Moreover, the positions of both of the Ca²⁺ ions and the water molecules agree very closely. The largest discrepancy is 0.4 Å.

Two other features are apparent from these comparisons: (a) The water molecules in loop III of TnC and loop III-IV of ICaBP fall at the center of the position occupied by the carboxyl group of Glu-59 in the Parv CD loop. (b) The third ligand of the Parv CD loop (position 5 in Figure 1) is a serine side chain rather than Asx. In this case, direct coordination of Ser-55 O^{γ} to the Ca^{2+} is achieved by adjustment of the main-chain conformation near Ser-55 to bring it closer to the cation (Figure 4). The interaction of the serine O^{γ} with the NH group of residue n+2 is maintained as in the cases of Asx turns.

DISCUSSION

The sample of five Ca2+ binding loop structures, each of which contains 12 residues homologous to the archetypal sequence of Parv, strongly suggests that any other Ca²⁺ binding site of such a sequence will adopt a similar main-chain conformation. In cases where a side-chain ligand cannot coordinate directly to the cation without disruption of that conformation, it is likely that a water molecule will be used as a ligand instead. Only Ser-55 in Parv does not conform to this generalization. This prediction can be especially relevant to the position of the fifth ligand of Ca²⁺ (i.e., position 9), which was described in the previous section. For example, all four binding sites in CaM contain amino acids with side chains too short to be used directly as fifth ligands, and in a manner similar to TnC they should be associated with water molecules (Figure 1). Those are Thr-28, Asp-64, Ser-101 and Asn-137. The same may apply to the regulatory domain of TnC, in which the candidates for coordination in the equivalent position are Ser-38 and Asp-74. There is less evidence to support the same prediction concerning candidates for the third ligands in the regulatory domain of TnC, which are Gly-34 and Ser-70. However, the option of a water molecule coordinating to the Ca²⁺ in the vicinity of Gly-34 looks highly favored since otherwise the main-chain peptide might have to be reversed so that the Gly-34 main-chain carbonyl oxygen would be directed toward the calcium. Presumably this would create a larger distortion than that involving Ser-55 in Parv.

One of the goals in structural studies of Ca²⁺ binding proteins is understanding what determines low- or high-affinity binding of Ca²⁺. This is especially important for TnC, in which the low-affinity sites play the regulatory role in muscle contraction. Sites with approximately the same Ca²⁺ affinity exhibit a wide range of coordination at the fifth position, from direct interaction of a Glu in the CD loop of Parv, through binding via a water molecule to an Asp in the two high-affinity sites of TnC, to no acidic group at all as in Gly-98 in the EF hand of Parv. There are no two binding loops with identical sequences within this class of proteins (Gariepy & Hodges, 1983). Obviously, numerous factors must affect the final Ca²⁺ binding constants, factors involving not only direct interactions of the Ca²⁺ ligands but also longer range properties. These could include charge-charge and dipole interactions, as well as helix-helix van der Waals packing interactions between and

within EF hands. The latter could affect the preference of helices to pack tighter in an antiparallel fashion as observed for the apo state in TnC or to be more open as in the Ca²⁺-bound state.

ACKNOWLEDGMENTS

We thank John Moult and Ted Baker for valuable discussions.

Registry No. Ca, 7440-70-2.

REFERENCES

- Adman, E., Watenpaugh, K. D., & Jensen, L. H. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 4854-4858.
- Babu, Y. S., Sack, J. S., Greenhough, T. G., Bugg, C. E., Means, A. R., & Cook, W. J. (1985) Nature (London) 315, 37-40.
- Baker, E. N., & Hubbard, R. E. (1984) Prog. Biophys. Mol. Biol. 44, 97-179.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shinanouchi, T., & Tasumi, M. (1977) J. Mol. Biol. 112, 535-542.
- Coffee, C. J., & Bradshaw, R. A. (1973) J. Biol. Chem. 248, 3305-3312.
- Collins, J. H., Potter, J. D., Horn, M. J., Wilshire, G., & Jackman, N. (1973) FEBS Lett. 36, 268-272.
- Einspahr, H., & Bugg, C. E. (1980) Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. B36, 264-271.
- Einspahr, H., & Bugg, C. E. (1981) Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. B37, 1044-1052. Fullmer, C. S., & Wasserman, R. H. (1981) J. Biol. Chem. 256, 5669-5674.
- Gariépy, J., & Hodges, R. S. (1983) FEBS Lett. 160, 1-6. Hendrickson, W. A., & Konnert, J. H. (1980) in Biomolecular Structure, Function, Conformation and Evolution (Srinivasan, R., Ed.) Vol. I, pp 43-57, Pergamon Press, Oxford. Herzberg, O., & James, M. N. G. (1985) Nature (London) 313, 653-659.
- Kobayashi, T., Takasaki, Y., Takagi, T., & Konishi, K. (1984) Eur. J. Biochem. 144, 401-408.
- Kretsinger, R. H. (1980) CRC Crit. Rev. Biochem. 8, 119-174.Kretsinger, R. H., & Nockolds, C. E. (1973) J. Biol. Chem. 248, 3313-3326.
- Kretsinger, R. H., & Barry, C. D. (1975) *Biochim. Biophys.* Acta 405, 40-52.
- Levine, B. A., & Williams, R. J. P. (1982) in Calcium and Cell Function (Cheung, W. Y., Ed.) Vol. II, pp 1-38, Academic Press, New York.
- Luzzati, V. (1952) Acta Crystallogr. 5, 802-810.
- Rees, D. C., Lewis, M., & Lipscomb, W. N. (1983) J. Mol. Biol. 168, 367-387.
- Sielecki, A. R., James, M. N. G., & Broughton, C. G. (1982) in *Proceedings of an International Summer School on Crystallographic Computing* (Sayre, D., Ed.) pp 409-419, Clarendon Press, Oxford.
- Sundralingam, M., Bergstrom, R., Strasburg, G., Rao, S. T., Roychowdhury, P., Greaser, M., & Wang, B. C. (1985) Science (Washington, D.C.) 227, 945-948.
- Szebenyi, D. M. E., Obendorf, S. K., & Moffat, K. (1981) Nature (London) 294, 327-332.
- Venkatachalam, C. M. (1968) Biopolymers 6, 1425-1436.
 Watterson, D. M., Sharief, F. S., & Vanaman, T. C. (1980)
 J. Biol. Chem. 255, 962-971.
- Wilkinson, J. M. (1976) FEBS Lett. 70, 254-256.