Handedness of crossover connections in β sheets

(protein conformation/protein folding/structure prediction/β-α-β loops/topology of β structures)

JANE S. RICHARDSON

Department of Anatomy, Duke University, Durham, North Carolina 27710

ABSTRACT In a crossover connection, the polypeptide chain leaves one end of a β sheet, forms a loop of any length and any conformation, and reenters the same β sheet from the opposite end. Of the 85 examples of crossover connections which occur in the known protein structures, 83 are righthanded and only two are lefthanded. It is proposed that consistent handedness, even in long irregular loops, could be produced by the preferred twist direction of extended chain and the righthandedness of α-helices, provided certain conditions hold during the protein folding process.

β pleated sheets of parallel, antiparallel, or mixed types have proven to be one of the major structural features of globular proteins. Now that a rather large sample of different protein structures have been determined by x-ray crystallography, it has become possible to discern some patterns in the occurrence of features common to these structures. The current paper discusses the regularities that can be seen for one particular feature of β structure—the crossover connection.

Nomenclature

For the topological analysis of β sheet structures, the backbone connections between β strands can be classified into the two general types illustrated in Fig. 1. The first category is "hairpin," "plain," or "same-end" connections, where the chain reenters the sheet at the same end it left from (Fig. 1a). The second category is "crossover," "cross," or "opposite-end" connections, where the chain loops around to reenter the β sheet on the opposite end to that it left from (Fig. 1b). For either case the backbone loop between the two β strands may be of any length and take any conformation, except that it may not include another β strand that is part of the same β sheet. The two connected β strands may be nearest neighbors in the β sheet (as in Fig. 1) or one or more other strands may lie between them in the hydrogen-bonded β structure (as in Fig. 2). The name given each specific type of connection depends on the number (but not the direction) of intervening strands: "±1" is a hairpin and "±1x" a crossover connection between nearest-neighbor strands, "±2" is a hairpin and "±2x" a crossover connection with one intervening strand, etc. (1).

As illustrated in Fig. 2, crossover connections may be either righthanded or lefthanded.

Another type of handedness involved in this discussion is the characteristic twist of the strands in β sheet, which has been summarized and discussed by Chothia (2). This twist can be defined as the angle between successive peptide planes viewed along the direction of the chain (so that a flat, model β strand would have 0° of twist)*. The β strands occurring in the known protein structures all have a righthanded twist of between 5°

and 30° per residue; that is, they would make one full 360° twist in from 12 to 72 residues.

Type, length, and handedness of occurring crossover connections

Fig. 3 shows a selection of ±1x and ±2x crossover connections from five different proteins, all rotated into a standard orientation. Table 1 summarizes all crossover connections which occur in 20 different proteins whose structures are known. The preference for type ±1x connections is evident, as has previously been noted (1). The classic crossover connection (as illustrated, for instance, in refs. 4 and 5) consists of two parallel β strands joined by an antiparallel α-helical segment; in 57 of the 85 cases listed in Table 1, the connecting loop contains one or more helices, but there is great variation in conformation, as can be seen from the examples in Fig. 3.

Counting the length of a crossover loop between any pair of α-carbons that are directly opposite each other on the two β strands, the length falls in the range of 17 to 40 residues for 75% of the examples. The two shortest loops are nine residues for a ±2x in glyceraldehyde-phosphate dehydrogenase [D-glyceraldehyde-3-phosphate:NAD$^+$ oxidoreductase (phosphorylating), EC 1.2.1.12] (see Fig. 3e) and 12 residues for a ±1x in phosphoglycerate kinase (ATP:3-phospho-D-glycerate 1-phosphotransferase, EC 2.7.2.3); the two longest (excluding one case in phosphoglycerate kinase which includes an entire domain) are 86 residues for a ±1x in carboxypeptidase A (peptidyl-L-amino-acid hydrolase, EC 3.4.12.2) and 93 residues for a ±2x in phosphoglyceromutase (2,3-bisphospho-D-glycerate:2-phospho-D-glycerate phosphotransferase, EC 2.7.5.3).

The overwhelmingly consistent, and somewhat surprising, feature is that 83 cases out of the 85 are righthanded, even when the intermediate loop is long and convoluted. The two lefthanded examples are a ±1x in subtilisin (see Fig. 3d) and a ±2x in the more helical lobe of hexokinase. Only a few other features of protein structure are empirically observed at levels of reliability as high or higher than this: for instance, the righthandedness of α-helices, the solvent accessibility of charged groups, the glycine requirement in type II tight turns, and the characteristic twist of β sheets. All of these features can be derived from fairly straightforward theoretical considerations, and the first three of these examples were predicted before they were observed (6–8). The righthandedness of crossover connections, however, is an unanticipated type of long-range regularity.

Possible explanations for the righthandedness of crossover connections

What constraint could impose such a consistent long-range handedness on these loops in spite of the irregularity and variability of their conformations? One possibility is that a strongly preferred conformation at the ends of the β strands may prejudice the direction of the connecting loop. A count of the approximate direction the chain leaves (or enters) the end of

---

* Although this is probably the most useful convention for relatively extended chains, note that it differs from the convention used to describe helices (e.g., ref. 3) according to which a flat β strand would be described as having 180° of twist because its carbonyls alternate in direction.
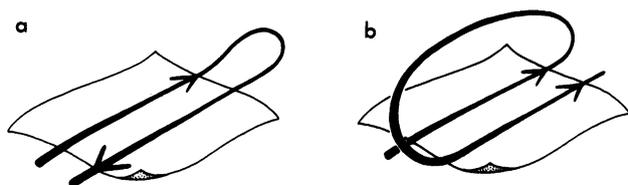
FIG. 1. Illustration of the two main classes of topological connection in β sheets: (a) "hairpin," "plain," or "same-end" connection; this specific example is type ±1, and (b) "crossover," "cross," or "opposite-end" connection; this specific example is type $\pm1x$.

FIG. 2. (a) A righthanded $\pm2x$ crossover connection. (b) A lefthanded $\pm2x$ crossover connection. Direction is not indicated for the skipped strand, since it may be either parallel or antiparallel to the others.

the β strand at each end of the crossover connections indicates that 57% of the time the direction is within the quadrant which would favor righthandedness. This constitutes a strong preference for the expected direction, but certainly not strong enough to explain a 97% rate of righthandedness. The directional preference of local conformation at the ends of the β strands is most pronounced for the shortest crossover connections and less so for the more wandering loops, while righthandedness of the loop is apparently invariant over the entire range of lengths; such a distribution pattern is more consistent
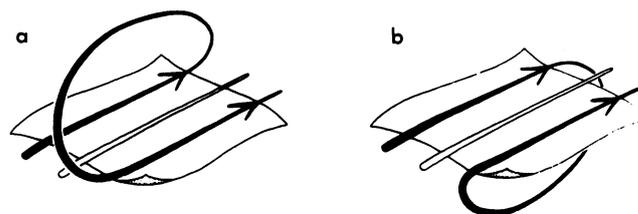
with the local conformational preference being an effect, rather than a cause, of overall loop handedness.

The type of explanation that seems to be required for this handedness phenomenon would involve a long-range constraint that could act during the protein folding process. A constraint that fits the requirements can be derived from the preferred twist direction of extended chain. As described in the section on nomenclature, β sheets are always observed to have a righthanded local twist, and this same twist preference can be deduced for isolated extended chains also, either from entropic

Table 1. Summary of the type and handedness of all crossover connections which occur in proteins of currently known three-dimensional structure

	±1x	±2x	±3x	±4x	±7x	±8x	±12x	Right-handed	Left-handed
Adenylate kinase (9) (EC 2.7.4.3)	2	2						4	
Bacteriochlorophyll protein (10)	1	2	1				1	5	
Carbonate dehydratase C (11) (EC 4.2.1.1)		1			1	1		3	
Carboxypeptidase A (12) (EC 3.4.12.2)	2	1						3	
Concanavalin A (13)		3		2				5	
Cytochrome b_5 (14)			2					2	
Erabutoxin b (31)		1						1	
Flavodoxin (15)	3	1						4	
Glyceraldehyde phosphate dehydrogenase (16) (EC 1.2.1.12)	4	1	3					8	
Hexokinase (17) (EC 2.7.1.1)	3	1	2					5	1 (±2x)
Lactate dehydrogenase (18) (EC 1.1.1.27)	4		1					5	
Liver alcohol dehydrogenase (19) (EC 1.1.1.1)	4	2	1					7	
Pancreatic trypsin inhibitor (20)		1						1	
Papain (21) (EC 3.4.22.2)		1		1				2	
Phosphoglycerate kinase (22, 23)* (EC 2.7.2.3)	6	1	1					8	
Phosphoglyceromutase (24) (EC 2.7.5.3)	1	2						3	
Thiosulfate sulfurtransferase (25) (rhodanese, EC 2.8.1.1)	2	1						3	
Subtilisin (26) (EC 3.4.21.14)	4	2						5	1 (±1x)
Thioredoxin (27)	1	1						2	
Triosephosphate isomerase (5) (EC 5.3.1.1)	7							7	
Totals	44	24	11	3	1	1	1	83	2

Essentially identical structures (such as the two domains of thiosulfate sulfurtransferase) are represented by only one example. Proteins are omitted entirely if they have no β structure or if their β sheets contain no crossover connections. Numbers in parentheses are references.

* Three somewhat different chain tracings have been reported for this enzyme; the above table entry is for the nucleotide-binding-domain version from ref. 22, but all three tracings have eight crossover connections, all righthanded. Handedness is unlikely to be wrong in a complete and consistent chain tracing because a minimum of four changes in connectivity are generally necessary to reverse the handedness of a crossover connection.
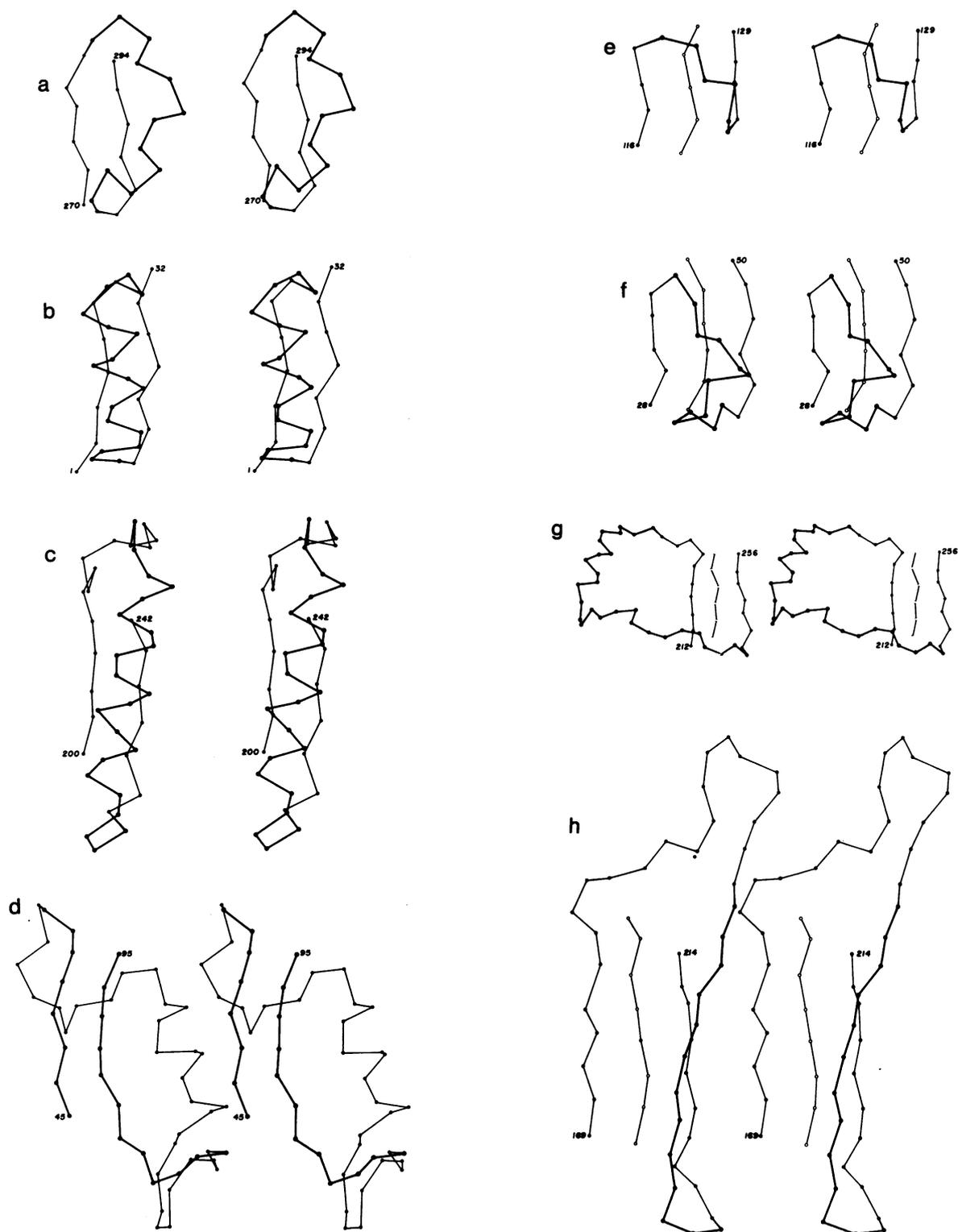
Biochemistry: Richardson

Proc. Natl. Acad. Sci. USA 73 (1976)   2621

FIG. 3. Stereo α-carbon drawings of an assortment of actual crossover connections, each rotated into a standard position with the β sheet in the plane of the paper and the β strands vertical. All are at the same scale except (g) which is reduced by 1/2. (a) A nonhelical ±1x from glyceraldehyde-phosphate dehydrogenase. (b) A helical ±1x from glyceraldehyde-phosphate dehydrogenase. (c) A helical ±1x from carboxypeptidase A. (d) A lefthanded ±1x from subtilisin. (e) A very short non-helical ±2x from glyceraldehyde-phosphate dehydrogenase with the central strand parallel. (f) A ±2x from subtilisin with the central strand parallel. (g) A helical ±2x from carbonate dehydratase C in which the loop makes a wide excursion to one side. (h) A rather long ±2x from concanavalin A with the central strand antiparallel and the connecting loop forming one strand in another β sheet.

(2) or from minimum energy (28, 29) considerations. Therefore, during the folding process any stretch of sequence in an approximately extended conformation would tend to resemble

Fig. 4a. It would not be a truly stiff ribbon, of course, but it has some degree of rigidity because of the preference for all residues to stay within the local energy well surrounding the twisted
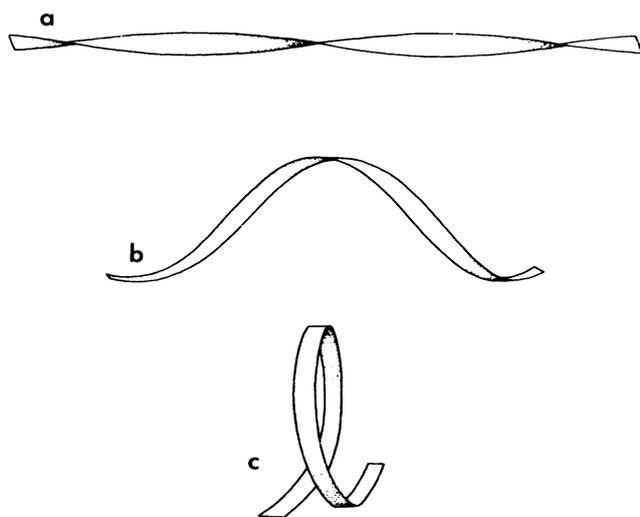
FIG. 4. A possible folding pathway which produces righthanded crossover loops from extended chain. In (a) the section of chain is extended, showing one full turn of the preferred righthanded twist for β strands. In (b) the two ends of this chain segment are moving toward one another, and the ribbon has started to buckle in a right-handed sense constrained by the chain twist. In (c) a complete righthanded loop is formed, with the two ends in position to form parallel β structure.

FIG. 5. A possible folding pathway which forms righthanded crossover loops from a righthanded α-helix with a β strand at each end of it.

β conformation. Since in general the stretch of chain in question will have large partly random sections of protein attached to both ends, it is unlikely that one end will wrap around in a circle to meet the other end. However, if any force now pushes the two ends of this chain segment toward one another, it can smoothly fold (as shown in Fig. 4) into a large righthanded loop of about the same size as the pitch of its former twist, with only very small changes in any of the ϕ,ψ angles. In contrast, forcing the chain segment into a lefthanded loop produces a very contorted structure. To prejudice the handedness of the resulting loop it is not necessary for the chain to remain smooth during the entire folding of the loop; it is only necessary to retain smoothness of the ribbon long enough to start the loop bulging in the right direction. The approximate loop size observed in crossover connections (9 to 93 residues) is about right to correspond with the pitch length of one full turn (12 to 72 residues) for the preferred range of chain twist. This type of constraint is implicitly built into the folding simulation used by Levitt and Warshel (29) to generate a model of pancreatic trypsin inhibitor which does in fact generate the righthanded ±2x crossover connection seen in native trypsin inhibitor.

For crossover connections with an α-helical connecting loop, the above mechanism of handedness constraint would not usually be applicable, because the presumption is that the helix would tend to form at least as early in the folding process as the β strands. A different, very simple folding pathway can be postulated for the case where a central α-helix has formed with a β strand extending from each end of it in approximately parallel directions. If the two β strands move toward one another, then as shown in Fig. 5 there can be a smooth transition to a righthanded loop with the helix unwinding slightly at the ends. In this case, the preference for righthandedness depends on the righthandedness of the α-helix, combined with the fact that it is much more likely for the backbone to leave both ends of the helix in a fairly smooth continuous direction than for both ends to reverse sharply at that point. This is undoubtedly not the only plausible folding pathway by which the handedness of an α-helix can be imagined to influence the handedness of
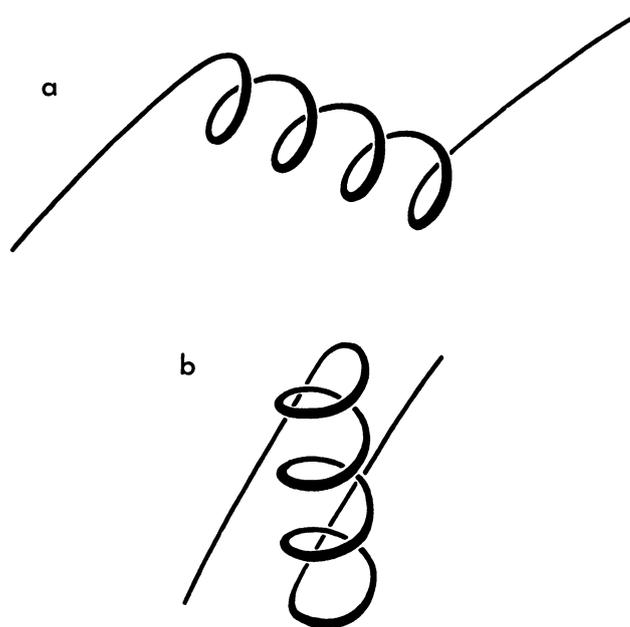
a resulting large loop. However, although it is not possible to enumerate all such pathways and determine their relative probabilities of occurrence, the observed empirical regularity of loop handedness requires that all commonly occurring loop folding pathways must in fact constrain righthandedness.

The two hypothetical folding pathways shown in Figs. 4 and 5 have in common the fact that they fold in a one-step process of bringing together the two β strands and throwing up the intervening loop. It is much less easy to see how long-range loop handedness could be strongly constrained by a two-step folding process in which first the intervening chain (whether α-helical or not) folds against one β strand, and then the second β strand folds down next to the first. An additional characteristic which enables the two proposed pathways to prejudice handedness is the condition that the ends of the folding section of chain are held approximately in place over very short time spans by the rest of the protein mass.

It may be that the nearly obligatory righthandedness of crossover-connection loops is a combined result of several contributing causes (both local conformational preferences and folding constraints, arising both from β strand twist handedness and from α-helix handedness), each of which happens to prejudice the result in the same direction.

Discussion

The fact that crossover connections are essentially always righthanded is useful information in several contexts. First of all, it can help in any attempt to predict how elements of secondary structure fold to form the tertiary structure of a protein. Secondly, this regularity must be taken into account in any analysis of topological similarity between β structures (e.g., refs. 1 and 30); for the case of n parallel strands it reduces the number of allowable possibilities by a factor of 2^{n-1}. Also, a presumption of righthandedness for crossover loops can sometimes aid the initial tracing of backbone chain from an electron density map.

The existence of long-range handedness constraints in the backbone structure of globular proteins is also significant because it forces us to consider the possibility that at least some

of the simple dominating processes in protein folding involve longer portions of the sequence than has usually been thought. According to the classic schema, the elements of secondary structure form first, and then come together as nearest-neighbor pairs or add on, one at a time, to nucleation sites that begin as pairs. It seems likely from the above discussion, however, that another frequent occurrence may be the one-step formation of a loop, bringing together two β strands which are not nearest-neighbor elements of secondary structure. The folding pathways suggested here may not be the correct ones or the only correct ones; but because crossover connections make up approximately 40% of the backbone length of the 37 distinct protein structures which are now known, and because their consistent handedness strongly suggests that they fold as a concerted unit, the formation of crossover connections is unquestionably an important element in understanding the folding of proteins.

Note Added in Proof. The righthandedness of β-α-β loops was noticed independently by K. Nagano, and an analysis of it is being published by M. J. E. Sternberg and J. M. Thornton (*J. Mol. Biol.*, in press).

1. Richardson, J. S., Richardson, D. C., Thomas, K. A., Silverton, E. W. & Davies, D. R. (1976) *J. Mol. Biol.* **102,** 221–235.
2. Chothia, C. (1973) *J. Mol. Biol.* **75,** 295–302.
3. Dickerson, R. E. & Geis, I. (1969) in *The Structure and Action of Proteins* (Harper & Row, New York), chap. 2.
4. Rao, S. T. & Rossmann, M. G. (1973) *J. Mol. Biol.* **76,** 241–256.
5. Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I. & Wilson, I. A. (1975) *Nature* **255,** 609–614.
6. Pauling, L., Corey, R. B. & Branson, H. R. (1951) *Proc. Natl. Acad. Sci. USA* **37,** 205–211.
7. Kauzmann, W. (1959) *Adv. Protein Chem.* **14,** 1–63.
8. Venkatachalam, C. M. (1968) *Biopolymers* **6,** 1425–1436.
9. Schulz, G. E., Elzinga, M., Marx, F. & Schirmer, R. H. (1974) *Nature* **250,** 120–123.
10. Fenna, R. E. & Matthews, B. M. (1975) *Nature* **258,** 573–577.
11. Liljas, A., Kannan, K. K., Bergstén, P.-C., Waara, I., Fridborg, K., Strandberg, B., Carlbom, U., Järup, L., Lövgren, S. & Petef, M. (1972) *Nature New Biol.* **235,** 131–137.
12. Quiocho, F. A. & Lipscomb, W. N. (1971) *Adv. Protein Chem.* **25,** 1–78.
13. Reeke, G. N., Jr., Becker, J. W. & Edelman, G. M. (1975) *J. Biol. Chem.* **250,** 1525–1547.
14. Mathews, F. S., Levine, M. & Argos, P. (1972) *J. Mol. Biol.* **64,** 449–464.
15. Anderson, R. D., Apgar, P. A., Burnett, R. M., Darling, G. D., Lequesne, M. E., Mayhew, S. G. & Ludwig, M. L. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 3189–3191.
16. Moras, D., Olsen, K. W., Sabesan, M. N., Buehner, M., Ford, G. C. & Rossmann, M. G. (1975) *J. Biol. Chem.* **250,** 9137–9162.
17. Fletterick, R. J., Bates, D. J. & Steitz, T. A. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 38–42.
18. Chandrasekhar, K., McPherson, A., Adams, M. J. & Rossmann, M. G. (1973) *J. Mol. Biol.* **76,** 503–528.
19. Brändén, C.-I., Eklund, H., Nordström, B., Boiwe, T., Söderlund, G., Zeppezauer, E., Ohlsson, I. & Åkeson, Å. (1973) *Proc. Natl. Acad. Sci. USA* **70,** 2439–2442.
20. Huber, R., Kukla, D., Rühlmann, A., Epp, O. & Formanek, H. (1970) *Naturwissenschaften* **57,** 389–392.
21. Drenth, J., Jansonius, J. N., Koekoek, R. & Wolthers, B. G. (1971) *Adv. Protein Chem.* **25,** 79–115.
22. Blake, C. C. F. & Evans, P. R. (1974) *J. Mol. Biol.* **84,** 585–601.
23. Bryant, T. N., Watson, H. C. & Wendell, P. L. (1974) *Nature* **247,** 14–17.
24. Campbell, J. W., Watson, H. C. & Hodgson, G. I. (1974) *Nature* **250,** 301–303.
25. Bergsma, J., Hol, W. J. G., Jansonius, J. N., Kalk, K. H., Ploegman, J. H. & Smit, J. D. G. (1975) *J. Mol. Biol.* **98,** 637–643.
26. Alden, R. A., Birktoft, J. J., Kraut, J., Robertus, J. D. & Wright, C. S. (1971) *Biochem. Biophys. Res. Commun.* **45,** 337–344.
27. Holmgren, A., Söderberg, B.-O., Eklund, H. & Brändén, C.-I. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 2305–2309.
28. Ramachandran, G. N. & Sasisekharan, V. (1968) *Adv. Prot. Chem.* **23,** 284–437.
29. Levitt, M. & Warshel, A. (1975) *Nature* **253,** 694–698.
30. Schulz, G. E. & Schirmer, R. H. (1974) *Nature* **250,** 142–144.
31. Low, B. W., Preston, H. S., Sato, A., Rosen, L. S., Searl, J. E., Rudko, A. D. & Richardson, J. S. (1976) *Proc. Natl. Acad. Sci. USA*, in press.