

Reprint Series
24 December 1993, Volume 262, pp. 2027-2030

SCIENCE

Crystal Structure of the Repetitive Segments of Spectrin

Y. Yan, E. Winograd,* A. Viel, T. Cronin, S. C. Harrison, and D. Branton

Crystal Structure of the Repetitive Segments of Spectrin

Y. Yan, E. Winograd,* A. Viel, T. Cronin,
S. C. Harrison, D. Branton

The elongated proteins of the spectrin family (dystrophin, α -actinin, and spectrin) contain tandemly repeated segments and form resilient cellular meshworks by cross-linking actin filaments. The structure of one of the repetitive segments of α -spectrin was determined at a 1.8 angstrom resolution. A segment consists of a three-helix bundle. A model of the interface between two tandem segments suggests that hydrophobic interactions between segments may constrain intersegment flexibility. The helix side chain interactions explain how mutations that are known to produce hemolytic anemias disrupt spectrin associations that sustain the integrity of the erythrocyte membrane.

Spectrin, α -actinin, and dystrophin are actin cross-linking proteins. Their associations with actin generate meshworks that support the plasma membrane and sustain interactions between cellular structures responsible for cell motility and shape (1, 2). Although all spectrins have a similar domain organization, important functional distinctions among these proteins appear to arise from the differing number of tandemly repeated segments that each contains. Because the repeated segments account for most of the length of these proteins, the structure of these segments must in large part determine the flexibility, and hence the functional and mechanochemical properties, of the meshworks they form (3).

Sequence analysis of spectrin and measurements of contour length have suggested that the repeated segments each consist of 100 to 120 amino acid motifs that fold into a three-helix bundle (4, 5). Our recent demonstration that single repeats of spectrin can fold into stable conformations similar to those in the native protein (6) presented a fresh opportunity to examine segment structure. Polypeptides representing single repeats exist in solution as a mixture of monomers and homodimers (7). We have been able to crystallize homodimers of the 14th segment of *Drosophila* α -spectrin (Fig. 1) and determine its structure.

The details of the structure determination are presented in Table 1. A 3 Å resolution, multiple isomorphous replacement (MIR) map based on two heavy-atom derivatives and anomalous dispersion showed that the α 14 fragment had crystallized as a dimer, with the two polypeptide

partners related by a noncrystallographic twofold (Fig. 2A). Most of the polypeptide chain could be built into the map (Fig. 3), and clear side chain density made the sequence alignment unambiguous. Helices A (8) and B form a hairpin, as expected, but helix C is continuous with helix B and thus does not fold back on helices A and B. In place of helix C, helix C' (residues from the second polypeptide chain in the dimer) pack against helices A and B to create the anticipated three-helix bundle (Fig. 2B). Because the two polypeptide partners are of identical amino acid sequence, the "invading" C' helix can make precisely the same contacts with the AB hairpin as would the C helix if it were folded back on the AB hairpin. Because the intact molecule has an extended structure whose NH₂- and COOH-termini lie at opposite ends (9),

the NH₂- and COOH-termini of a segment must extend in opposite directions. Therefore, the C helix of each segment must fold back against helices A and B. Thus, although the residues between helices B and C maintain a continuous α -helical conformation in the crystalline α 14 dimer (Fig. 2A), we believe that in the intact spectrin chain these residues form a small loop (Fig. 2B) (10).

In the description here, we concentrate on one three-helix bundle. The last two residues at the COOH-terminus are poorly ordered in the fragment, and eight residues in the AB loop also lie in weak density. Otherwise, the helices are well defined along their entire lengths. The mean main chain dihedral angles are -63° for ϕ and -42° for ψ , as in globular proteins (11). The three helices interact with each other over eight turns, forming a left-handed three-helix bundle about 50 Å long and 20 Å in diameter (Fig. 2B). The buried area is 4400 Å², which is 38% of the total accessible surface area of three helices. Helix B is longer by two turns than helix A. Helix C is bent, helix B is slightly bent, and helix A is straight. The respective mean angles between the axes of the helices are 20°, 10°, and 15° for A to B, B to C, and A to C, and the mean interaxial distances are 12.3 ± 2 , 10.6 ± 1 , and 10.9 ± 1 Å, respectively. Each of the three helices exhibits the helical pattern found in extended α -helical structures, with positions conventionally labeled by the letters a to g (12) (Fig. 4A). Residues a and d are generally hydrophobic and lie on the inward-facing surface of the

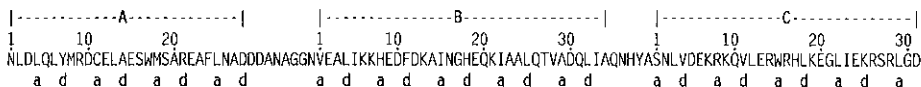


Fig. 1. Sequence of the 106 amino acids in segment 14 of *Drosophila* α -spectrin (23). The brackets above the sequence designate the regions that form α helices A, B, and C (8) as predicted (5). Segment phasing (starting with Asn 1392 and ending with Asp 1497) is derived from Winograd and co-workers (6) and has been illustrated for β -spectrin (24).

Table 1. Data collection, structure determination, and refinement (25, 26).

Parameter	Native III	Native II	Native I*	CH ₃ HgCl	K ₃ OsCl ₆
Resolution (Å)	1.8	2.0	3	3	3
Measured reflections	88201	46546	19853	9998	16417
Unique reflections	20541	13437	4928	4459	4779
Number of crystals	1	3	4	4	3
Completeness (%)	95	86	94	85	90
R_{merge}^\dagger	0.046	0.069	0.066	0.053	0.060
R_{iso}^\ddagger				0.29	0.13
Number of sites				2(4)	1
R_{cullis}^\S				0.48	0.63
Root mean square F_H/E (3 to 10 Å)				2.29	1.75

*Mean figure of merit for Native I = 0.50. $\dagger R_{\text{merge}} = \sum || - \langle I \rangle \Sigma I$, where I is intensity. $\ddagger R_{\text{iso}} = \sum |F_{\text{PH}} - F_{\text{P}}|/\Sigma F_{\text{P}}$, where F_{PH} and F_{P} are the derivative and native structure factor amplitudes, respectively. $\S R_{\text{cullis}} = \sum |F_{\text{H}}(\text{obs}) - F_{\text{H}}(\text{calc})|/\Sigma F_{\text{H}}(\text{obs})$, where F_{H} is the heavy-atom structure factor amplitude. ||Root mean square $F_H/E = \{(\Sigma f_H^2)/\Sigma [F_{\text{PH}}(\text{obs}) - F_{\text{PH}}(\text{calc})]^2\}^{1/2}$.

Y. Yan, E. Winograd, A. Viel, T. Cronin, D. Branton, Departments of Biochemistry and Molecular Biology and Cellular and Developmental Biology, Harvard University, Cambridge, MA 02138.
S. C. Harrison, Departments of Biochemistry and Molecular Biology and Cellular and Developmental Biology and Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138.

*Current address: Departamento de Quimica, Universidad Nacional de Colombia, Bogota, Colombia.

helix. The remaining residues tend to be more polar.

The antiparallel A and B helices in spectrin are in a register such that positions *d* of one helix are at the same level as positions *a* of the other (Fig. 4B). Thus, the *d* side chains project into a cavity on the surface of the opposite helix, packing laterally against *a* side chains and bounded axially by *d* side chains of successive heptads. This ridge-to-ridge packing arrangement is similar to that found in coiled-coil proteins (13), except that helix A does not coil around helix B. In contrast, helix C is axially displaced relative to helices B and A so that each turn of helix C runs roughly midway between two adjacent turns in helices B and A (Fig. 4B), and the side chain packing between helices C and B and between C and A conforms to the ridges-into-grooves arrangement common in globular proteins (14) (Fig. 5). The ridges-into-grooves packing of hydrophobic residues imposes different constraints on the angle between helices than does the ridge-to-ridge packing seen between helices A and B. Consequently, the B and C helices and the A and C helices pack together at angles of 10° and 15°, respectively, instead of the usual coiled-coil angle of 20° (13, 15).

In addition to the hydrophobic packing at the interior of the three-helix bundle, electrostatic interactions also appear to stabilize the spectrin repeat structure (Fig. 4A). Between helices A and C, interchain

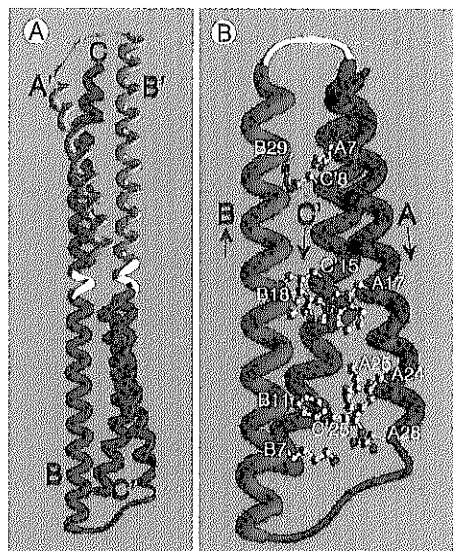


Fig. 2. Structure of α -spectrin segment 14. (A) Side view of the dimer. One polypeptide in the dimer is shown in red hues, the other in green hues. (B) Side view of one segment. The BC loop (white) has been inserted by model building. Some of the side chains that pack to maintain the spacing between the α helices are shown with carbon yellow, nitrogen blue, and oxygen red.

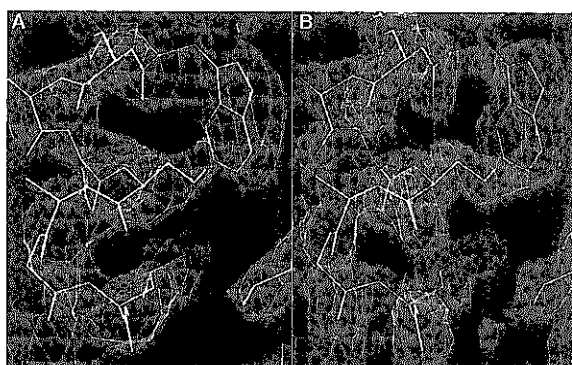


Fig. 3. Part of the electron density maps. (A) MIR map (3 Å). (B) $2F_o - F_c$ map (1.8 Å). The model shown is the final, refined model, including a portion of helix C, residues C'14 to C'26. Both maps are contoured at 1σ .

salt bridges are often formed by charged residues at positions *e* and *g*, as in conventional coiled-coil structures, and additional residues at position *c* in helix C are also involved in some ionic interactions. Salt bridges between the antiparallel B and C helices also involve charged residues at positions *e*.

Comparisons of the sequence of segment $\alpha 14$ with those of the other repetitive segments of α - and β -spectrin chains show that the key interactions present in our structure are broadly conserved. We there-

fore believe that all spectrin repeats are likely to have the same three-helix packing. In particular there is conservation of hydrophobic residues at those positions in all three helices that form the critical hydrophobic interface at the inside of the helix bundle. Moreover, the conserved charged residues often form interchain salt bridges. Among the most conserved of all residues are those that are important for maintaining contacts between the straight helix A and the coiled helices B and C. In particular, the residue at A7 (usually Tyr or Phe)

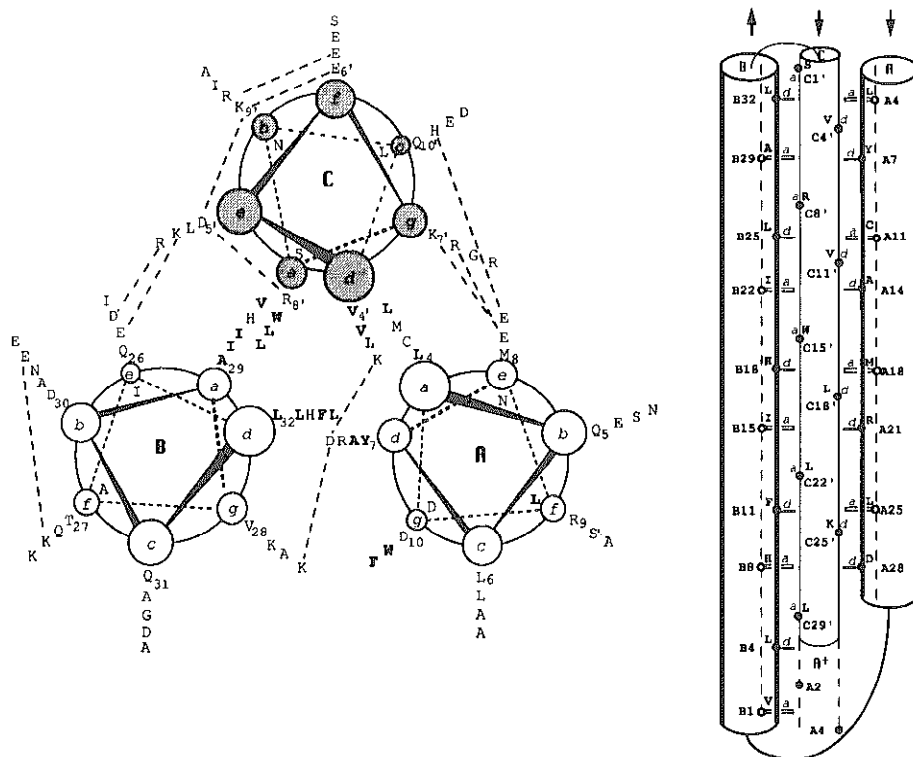


Fig. 4. Helical projection and side-view cartoons. (Left) Axial helical projection looking down the helix bundle from the NH_2 -terminus of the A helix. The heptad positions are shown by circles from which the actual residues at these positions radiate; the numerical position within the helix is indicated for the amino acids of the first heptad repeat closest to the viewer whose residues interact in the bundle. Residues closer to the viewer than the first interacting heptad repeat are shown inside each helix. Hydrophobic residues are bold. (Right) Cartoon side view of the three helices. The axial displacement of helix C with respect to A and B places the *a* position residues of C between the axially underlying or overlying *d* and *a* residues in helices A and B. Dashed lines extending from the bottom of helix C indicate the proposed position of the next succeeding segment's A helix.

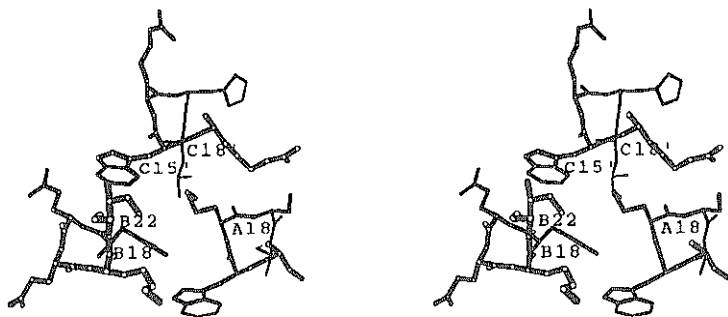


Fig. 5. Stereo view of α -spectrin segment 14 sectioned normal to the long axis of the helices at roughly the middle of the helix bundle. Only **a**- and **d**-position residues are labeled. The Met A18 (**a** position) and His B18 (**d** position) residues pack with their $C\alpha$ atoms at roughly the same depth, whereas the axial displacement of helix C' places the $C\alpha$ atoms of Met A18 above Leu C'18 (**d** position) but below Trp C'15 (**a** position) and the $C\alpha$ atoms of Trp C'15 above His B18 and below Ile B22 (**a** position).

helps space helix A away from B and C at one end, whereas at the other end the conserved, bulky residue at B11 (often Phe) packs against A25 and A24, the latter exceptionally conserved as a hydrophobic residue at a **g** position. Between these two, near the middle of the helix bundle, the invariant Trp A17 is involved in the way helix A crosses over from B to C, as is the nearly invariant Trp at C15. Indeed, these two tryptophans lie opposite each other, with His B18 between them. Finally, near the AB loop, the interaxial distance be-

tween helices A and B is unusually large (14 Å), and rather open salt bridges between the **d** position Asp A28, **g** position Lys B7, and **d** position Lys C'25 replace the usual hydrophobic contacts.

The integrity of the erythrocyte membrane and its skeleton depends on the association of spectrin heterodimers into tetramers (3). Tetramers form by associations between the partial repetitive motifs found near the NH_2 -terminus of the spectrin α chain and the $COOH$ -terminus of the β chain (16, 17). In all spectrins that have been sequenced, the partial motif on the α chain contains only those residues typical of a C helix, whereas the partial motif on the β chain contains residues typical of just the A and B helices. If the mode of association in tetramer formation corresponds to the packing seen in the crystallized fragment (helix C of one chain with helices A and B of the second chain) (Fig. 2A), the observed hydrophobic and ionic interactions should also be those that stabilize the corresponding association between α and β chains in the tetramer. Subgroups of hereditary elliptocytosis are human hemolytic disorders associated with the defective association of erythroid spectrin heterodimers into tetramers (18). Different point mutations have been found to be responsible for the defect. One of these (16), Ala to Pro at position B10 (19) in the incomplete terminal segment of the β chain, would be expected to disrupt the B helix and destabilize the three-helix bundle. Another (20), Arg to Leu, Ser, Cys, or His at position C'8 (21) in the partial NH_2 -terminal segment of the α chains, could disrupt critical interactions with conserved residues at A7 (usually Phe or Tyr) and B29 (usually Ala or Gly) (Fig. 2B). This disruption could occur because the manner in which the charged end of Arg C'8 sticks out of the helix bundle to interact with the neighboring **e** position residue

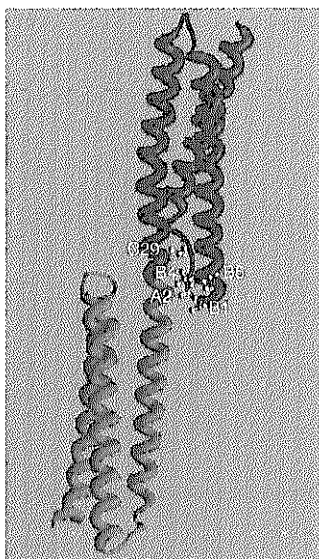


Fig. 6. Proposed model of two succeeding segments. Because the last two $COOH$ -terminal residues are poorly ordered in the fragment, the continuity between helix C of the first segment (red) and helix A of the second segment (pink) is hypothetical. Hydrophobic residues that stabilize the connection between the two segments (Val B1; Leu B4; Ile B5; the invariant Leu at C29; and the often hydrophobic residue located at an **f** position at A2, Leu in *Drosophila* α -14) are modeled in yellow.

(Fig. 4A) leaves the base of Arg C'8 as a small hydrophobic surface against which the side chains of A7 and B29 pack.

In general, the repetitive segments of spectrin preserve a fixed register between the C helix of one and the A helix of the next, with no prolines or glycines at the segment boundary. It has therefore been proposed that the "two" helices are in fact continuous (Fig. 4B), with the C helix of the first segment and the A helix of the next segment forming a single helix (5, 22). Because the residue positions are well defined along nearly the entire length of the helices, we were able to build a model of two successive repeats (Fig. 6). This model shows that interactions at the segment boundary explain several conserved, but otherwise puzzling, features of the spectrin amino acid sequence. First, the relatively open ends of the three-helix bundle allow the connection between segments to be made without any rearrangement of the AB or BC loops. Second, the α -helical turns that form the CA boundary (the last turns of helix C and the first of A) are involved in multiple interactions with the two turns of the helix at the NH_2 -terminus of B. Third, because the model predicts that each successive segment is axially rotated 60° in a right-handed sense with respect to the previous segment, it explains why the **f** position at A2 is frequently occupied by a hydrophobic, rather than a hydrophilic, residue. The 60° rotation places this first **f** position in the A helix of the next successive segment between the **a** and **d** (B1 and B4) residues of the preceding segment's B helix (Fig. 6). If this model is correct, the packing of this hydrophobic residue from A2 of one segment against the hydrophobic residues at B1, B4, and B5 of the previous segment should confine the segment-to-segment flexibility within a limited number of planes.

REFERENCES AND NOTES

1. R. R. Dubreuil, *BioEssays* 13, 219 (1991).
2. P. Matsuudaira, *Trends Biochem. Sci.* 16, 87 (1991).
3. A. Elgsaeter, B. T. Stokke, A. Mikkelsen, D. Branton, *Science* 234, 1217 (1986).
4. D. W. Speicher and V. T. Marchesi, *Nature* 311, 177 (1984); M. Koenig, A. P. Monaco, L. M. Kunkel, *Cell* 53, 219 (1988); R. A. Cross, M. Stewart, J. Kendrick-Jones, *FEBS Lett.* 262, 87 (1990).
5. D. A. D. Parry and C. Cohen, *A.I.P. Conf. Proc.* 226, 367 (1991); D. A. D. Parry, T. W. Dixon, C. Cohen, *Biophys. J.* 61, 858 (1992).
6. E. Winograd, D. Hume, D. Branton, *Proc. Natl. Acad. Sci. U.S.A.* 88, 10788 (1991).
7. Unpublished observations based on chromatography, cross-linking, and equilibrium sedimentation. Interconversion of the monomer and dimer was slow, probably because the helix-helix contacts must break to convert in either direction.
8. The first (NH_2 end) α helix in a three-helix bundle is labeled A, and the succeeding helices are labeled B and C (this differs from the notation in previous work, in which the "A" referred to our

- helix B.) The first residue in each helix is numbered 1; thus, A1 refers to the first residue in helix A. Amino acids are represented by their single-letter, uppercase abbreviations, and McLachlan (12) positions are in lowercase bold. When the crystalline dimer is referred to, the helices from one monomer are labeled with plain letters and those from the other are labeled with a prime.
9. D. M. Shotton, B. Burke, D. Branton, *J. Mol. Biol.* **131**, 303 (1979); D. W. Speicher, J. S. Morrow, W. J. Knowles, V. T. Marchesi, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5673 (1980).
 10. Unlike a number of the other spectrin repetitive motifs, segment 14 of the *Drosophila* α chain contains no Pro in the region between helix B and C, so that a continuous BC helix is possible. Furthermore, we note that the BC loop is variable in length from segment to segment; in α -spectrin segment 9, it contains an entire Src homology 3 domain.
 11. T. Blundell, D. Barlow, N. S. Borkalsoti, J. Thornton, *Nature* **306**, 281 (1983).
 12. A. D. McLachlan and M. Stewart, *J. Mol. Biol.* **98**, 293 (1975).
 13. D. W. Banner, M. Kokkinidis, D. Tsemoglou, *ibid.* **196**, 657 (1987); E. K. O'Shea, J. D. Klemm, P. S. Kim, T. Alber, *Science* **254**, 539 (1991); B. Lovejoy *et al.*, *ibid.* **259**, 1288 (1993).
 14. C. Chothia, M. Levitt, D. Richardson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4130 (1977).
 15. F. H. C. Crick, *Acta Crystallogr.* **6**, 89 (1953).
 16. W. T. Tse *et al.*, *J. Clin. Invest.* **86**, 909 (1990).
 17. D. W. Speicher *et al.*, *J. Biol. Chem.* **268**, 4227 (1993); L. Kotula, T. M. DeSilva, D. W. Speicher, P. J. Curtis, *ibid.*, p. 14788.
 18. J. Palek, in *Hematology*, W. Williams, E. Beutler, A. Erslev, M. Lichtman, Eds. (McGraw-Hill, New York, ed. 4, 1990), pp. 569-581.
 19. Corresponding to Asp B10 in *Drosophila* α 14.
 20. T. L. Coetzer *et al.*, *J. Clin. Invest.* **88**, 743 (1991).
 21. Corresponding to Arg C8 in *Drosophila* α 14.
 22. R. R. Dubreuil *et al.*, *J. Cell Biol.* **109**, 2197 (1989).
 23. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
 24. T. J. Byers, E. Brandin, R. A. Lue, E. Winograd, D. Branton, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6187 (1992).
 25. Crystals (in space group $P2_12_12$) were grown at room temperature with vapor diffusion against a well solution containing 10% polyethylene glycol (PEG) 400, 1 M ammonium acetate, 50 mM MES (4-morpholineethane sulfonic acid), pH 6.0. The starting mixture contained a 1:1 mixture of protein (20 mg/ml) and well solution. The polypeptide used for crystallization (originally called B14) was cleaved from a fusion protein generated in *Escherichia coli* as described (6). The mercury derivative was obtained by soaking crystals for 2 days in stabilization buffer (14% PEG 400, 1.2 M ammonium acetate, pH 6.0) saturated in CH_3HgCl , followed by a 1-hour backsoak. To prepare the K_3OsCl_6 derivative, the ammonium acetate in the stabilization buffer was exchanged for 1.2 M sodium acetate, and the crystals were soaked for 1 day in stabilization buffer containing 1 mM K_3OsCl_6 . Data were collected with a Siemens area detector on an Elliot GX-13 rotating anode generator (Avionics, Borehamwood, United Kingdom). $\text{CuK}\alpha$ radiation was provided by a Frank's double-focusing mirror assembly [S. C. Harrison, *J. Appl. Crystallogr.* **1**, 84 (1968)]. Data collection was controlled as described by M. Blum, P. Metcalf, S. C. Harrison, and D. C. Wiley [*J. Appl. Crystallogr.* **20**, 235 (1987)], and the data were processed by the XDS package of W. Kabsch [*ibid.* **21**, 916 (1988)]. Anomalous dispersion measurements were included in calculations of the phasing for the CH_3HgCl derivative. The Hg atom positions for the CH_3HgCl derivative were determined from a difference Patterson function. There were two sites in the asymmetric unit, and each site splits into two subsites, 3 Å apart. The K_3OsCl_6 derivative was solved from a difference Fourier synthesis phased on the CH_3HgCl derivative. There is one site per asymmetric unit. Heavy-atom parameters were refined, and initial phases were calculated with the program HEAVY [T. C. Terwilliger and D. Eisenberg, *Acta Crystallogr. Sect. A* **39**, 813 (1983)].
 26. Low-resolution data (3 Å) were collected at room temperature with unit cell $a = 47.91$ Å, $b = 48.70$ Å, $c = 105.29$ Å for native I and two derivatives. High-resolution data (2 and 1.8 Å) were collected at cryogenic temperature (-160°C) (native II and III; before data collection crystals were soaked in 25% PEG 400 and 1.2 M ammonium acetate, pH 6.0, for one day) with unit cell $a = 46.80$ Å, $b = 47.22$ Å, and $c = 104.43$ Å. The model built into the map was partially refined, with XPLOR [A. T. Brunger, *XPLOR Manual, Version 3.1* (Yale University, New Haven, CT, 1992); *Acta Crystallogr. Sect. A* **46**, 585 (1990)] against the native I (room temperature) data set. We then used the model to determine phases in the native II data set by refining six rigid segments in the resolution range of 12 to 3 Å ($R = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}} = 0.36$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes, respectively). A randomly selected 10% subset of the data was set aside for use in "free R factor" calculations. [A. T. Brunger, *Nature* **355**, 472 (1992)]. The results of the rigid-body search were manually adjusted in $2F_o - F_c$ and $F_o - F_c$ maps, and the resolution was extended to 2 Å in several stages. Cycles of simulated annealing refinement against 6 to 2 Å resolution data, alternated with manual adjustment and the addition of ordered water molecules, gave an R factor of 0.21 ($R_{\text{free}} = 0.34$). A still better low-temperature data set (Native III), extending to 1.8 Å resolution, then became available. Simulated annealing was used to refine against these data, first at 2 Å and then at 1.8 Å resolution ($R = 0.23$, $R_{\text{free}} = 0.32$); further refinement with TNT [D. E. Tronrud, L. F. TenEyck, B. W. Matthews, *Acta Crystallogr. Sect. A* **43**, 489 (1987)] yielded an R factor of 0.203 ($R_{\text{free}} = 0.304$, 6 to 1.8 Å, $F > 2\sigma$, 91% complete), with excellent geometry ($\sigma_{\text{bond}} = 0.017$ Å; $\sigma_{\text{angle}} = 2.2^\circ$). The final model had 107 ordered residues in each monomer (including one from the NH_2 -terminal cloning artifact) and a total of 156 ordered water molecules in the asymmetric unit. The coordinates of the refined model have been deposited in Brookhaven Protein Data Bank.
 27. We thank W. Stafford for equilibrium sedimentation analyses. Supported by National Institutes of Health grants CA 13202 (S.C.H.) and HL 17411 (D.B.).

30 July 1993; accepted 19 October 1993

