downstream of the *tac* promoter in pRW76. Then the region between the Nco I and Sph I sites was replaced by an oligonucleotide encoding *GAL4* sequences from codon 2 to the Sph I site, to yield pLPK76-7. The transcriptional terminator was derived from the P22 phage *ant* gene. The β-gal-GAL4(147–881) fusion is formed at the Eco RI site 17 codons from the end of *E. coli lacZ*, and is expressed from the *lac* promoter.

28. Soluble extracts for filter-binding assays were prepared from cells from 250-ml cultures of SF1148/pLPK76-7 (GAL4$^+$) and SF1148/pRW76 (GAL4$^-$); the cells were resuspended in 2 ml of buffer A(200) [buffer A, as modified from Bram and Kornberg (*9*), is 25 m*M* Hepes, *p*H 7.5; 5 m*M* $MgCl_2$; 0.1 m*M* EDTA; 5 m*M* 2-mercaptoethanol; 10 percent glycerol; 1 m*M* phenylmethylsulfonyl fluoride; 2 μ*M* pepstatin; 0.6 μ*M* leupeptin with KCl at the concentration given in parentheses)] and lysed by sonication below 10°C. The protein content of the cleared lysates was determined by the protein-dye binding method of M. M. Bradford [*Anal. Biochem.* 72, 248 (1976)], with bovine serum albumin as standard, and was typically ~15 mg/ml. In filter-binding experiments with pUC18 and pUC18 + 17mer DNA's we included 5 m*M* IPTG (isopropyl-β-thio-D-galactopyranoside) to reduce binding of *lac* repressor present in the extract to the *lac* operator on these plasmids. We estimated that the affinity of the GAL4-β-gal fusions for GAL4 binding sites was not more than tenfold lower than that of intact GAL4.

29. DNase I protection experiments were carried out in buffer A(50) containing sheared salmon testis DNA (50 μg/ml). Conditions for the footprinting experiments were established as follows: Filter-binding curves of the type shown in Fig. 2B were carried out at 17mer fragment concentrations from 10^{-13} to 10^{-8} *M*. The GAL4(1–147)-β-gal fusion concentration required for half-maximal retention was independent of 17mer concentration when this was at or below 10^{-10} *M*, but shifted up threefold at 10^{-9} *M* 17mer. Therefore, at the 17mer concentration used in the footprinting experiments (10^{-10} *M*), the ratio of specifically bound to free 17mer is independent of the 17mer concentration [A. D. Johnson, C. O. Pabo, R. T. Sauer, *Methods Enzymol.* 65, 839 (1980)]. The specific-binding activity of different preparations is now routinely quantitated by measuring the retention of 17mer by an appropriate dilution of the preparation under conditions of stoichiometric complex formation, at 10^{-8} *M* 17mer. For GAL4 and GAL4 derivatives, the concentration required to half-maximally occupy the 17mer is lower than that required to half-maximally occupy any of the natural sites in UAS_G.

30. The plasmid has been adapted for propagation in histidine-selective media, by replacement of LEU2 in a construct described in Silver *et al.* (*8*), with a *HIS3* gene. The plasmid pGG17 encodes GAL4(1–98) expressed from the *ADH1* promoter. The parental yeast strain in these experiments was YM335 (*a Δ-gal4 ura3-52 ade2-101 lysl his3 met*) obtained from M. Johnston [M. Johnston and R. Davis, *Mol. Cell. Biol.* 4, 1440 (1984)]. The plasmid pRY171 (*4*), which bears a *GAL1-lacZ* fusion but no 2 μ*M* origin of replication, was linearized at its only Apa I site in *URA3* [T. Orr-Weaver, J. Szostak, R. Rothstein, *Proc. Natl. Acad. Sci. U.S.A.* 78, 6354 (1981)], and introduced into cells of the YM335 strain to yield YM335::171. The YM335::GG9 recombinant was constructed as follows: The intact *HIS4* UAS (*18*), extending from 109 to 198 bp upstream of the *HIS4* transcription start site, was inserted, in its normal orientation, 128 bp upstream of the GAL1 start site, (in pLR1Δ20) (*5*). Two copies of a 30-bp DNA fragment containing the 17mer were then installed between the *HIS4* and *GAL1* sequences. The construct was then chromosomally integrated in YM335 (legend to Table 1). Plasmids (Fig. 6) that produce *GAL4*-β-gal fusion proteins were then introduced into these strains by the lithium acetate procedure [H. Ito, Y. Fukuda, K. Murata, A. Kimura, *J. Bacteriol.* 53, 163 (1983)]. Individual transformants were grown to 2 × 10^7 cell/ml in minimal medium lacking histidine [F. Sherman, G. Fink, J. Hicks, Eds., *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, rev. ed., 1983)], containing, 2 percent galactose, 3 percent glycerol, and 2 percent lactate as sodium lactate, *p*H 6. The β-galactosidase assays were performed in triplicate on cultures grown from separate transformant colonies. The standard errors were <20 percent.

31. We thank R. Brent, A. Courey, E. Giniger, A. Hochschild, M. Lamphier, S. Plon, P. Silver, R. Wharton, and members of the Ptashne laboratory for helpful discussions and comments on the manuscript. The initial detection of GAL4 in *E. coli* was made possible by a gift of antiserum to GAL4 from P. Silver. We thank M. Lamphier, R. Wharton and P. Silver, and R. Yocum for plasmids, J. Douhan III for antisera to λ repressor and β-galactosidase, A. Cowie for advice on protein-DNA complex immunoprecipitations, M. Johnston for yeast strains, and R. Yocum and P. Messenguy for permission to cite unpublished data. Supported by NIH grant GM32308 (M.P.), by a Paul Mazur Fellowship from Harvard University (L.K.), and by NIH training grant GM07598 (G.G.).

7 November 1985; accepted 27 December 1985

Crystal Structure of Cd,Zn Metallothionein

W. F. Furey, A. H. Robbins, L. L. Clancy, D. R. Winge, B. C. Wang, C. D. Stout

The anomalous scattering data from five Cd in the native protein were used to determine the crystal structure of cadmium, zinc (Cd,Zn) metallothionein isoform II from rat liver. The structure of a 4-Cd cluster was solved by direct methods. A 2.3 Å resolution electron density map was calculated by iterative single-wavelength anomalous scattering. The structure is folded into two domains. The amino terminal domain (β) of residues 1 to 29 enfolds a three-metal cluster of one Cd and two Zn atoms coordinated by six terminal cysteine thiolate ligands and three bridging cysteine thiolates. The carboxyl terminal domain (α) of residues 30 to 61 enfolds a 4-Cd cluster coordinated by six terminal and five bridging cysteine thiolates. All seven metal sites have tetrahedral coordination geometry. The domains are roughly spherical, and the diameter is 15 to 20 Å; there is limited contact between domains. The folding of α and β is topologically similar but with opposite chirality. Redundant, short cysteine-containing sequences have similar roles in cluster formation in both α and β.

Metallothioneins are low molecular weight, cysteine-rich proteins that bind metal ions (*1–4*). Metallothioneins and their genes have several potential kinds of physiological activity in that (i) the genes are induced by metal ions and glucocorticoid hormones; (ii) transcription is modulated during embryonic development; (iii) the genes may be involved in control of cell differentiation and proliferation; (iv) the proteins may function to activate Zn requiring apo-enzymes and regulate cellular metabolism; and (v) the proteins may act as free-radical scavengers, and they are synthesized in response to ultraviolet (UV) radiation (*5*). The metallothionein primary sequences are highly conserved, especially at positions of cysteine residues (*1, 6, 7*).

Metallothionein isoform II (MT II) from rat liver contains 61 amino acids including 20 cysteines (*7*). This molecule can be cleaved into two domains: residues 1 to 29 (denoted β, and having 9 cysteines) and residues 30 to 61 (denoted α, with 11 cysteines) (*8*). Nuclear magnetic resonance (NMR) with ^{113}Cd spectra demonstrate that Cd is bound to the protein in clusters of four and three

W. F. Furey and B. C. Wang are in the Biocrystallography Laboratory, VA Medical Center, P.O. Box 12055, Pittsburgh, PA 15240; A. H. Robbins and C. D. Stout are in the Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, CA 92037; L. L. Clancy is in the Department of Crystallography, University of Pittsburgh, Pittsburgh, PA 15260; and D. R. Winge is at the University of Utah Medical Center, Salt Lake City 84132.

metals formed by bridging and terminal cysteine thiolate ligands (*3, 9*). All 20 cysteines participate in tetrahedral coordination at the seven metal sites (*1, 10*). The four- and three-metal clusters are coordinated by the α and β domains, respectively (*3, 11*). The α domain binds four Cd cooperatively (*12*); the β domain binds both Cd and Zn, but binds Cu preferentially as compared to Cd or Zn (*13, 14*).

Rabbit liver MT II has been studied by two-dimensional NMR methods (*15, 16*). Heteronuclear ^{113}Cd-^{1}H NMR spectra have been used to assign the cysteine resonances in rabbit liver MT II, and two-dimensional nuclear Overhauser enhancement (NOESY) and two-dimensional correlated (COSY) spectra have been interpreted to yield a structure for the protein in solution (*17, 18*). We now report the x-ray crystal structure of MT II based on a 2.3 Å resolution electron density map.

Data collection. MT II isolated from rat liver crystallizes from potassium phosphate and sodium formate as single, colorless, square bipyramids 0.3 by 0.3 by 1.0 mm in size (*19*). The space group is tetragonal, $P4_12_12$, with $a = b = 30.9$, and $c = 120.4$ Å, and there is one molecule in the asymmetric unit (*19*). Dissolved single crystals contain 5 moles of Cd and 2 moles of Zn per mole of protein and the same amino acid composition as the native protein as isolated (*19*). The dissolved crystals also exhibit the same UV spectrum as the protein as isolated.

Native diffraction data were collected with CuKα radiation by oscillation photography to 2.3 Å resolution for one crystal and by diffractometry to 3.0 Å resolution for six crystals (Table 1). Data collection on the oscillation camera was done at 4°C and required 10.5 hours, and the crystal showed essentially no decay. Data collection on the diffractometer was done at 18°C and required 1 hour per 50 reflections. Crystals were discarded when monitor reflections decayed more than 10 percent, which occurred within 12 hours at 3.0 Å and within 2 days at 6.0 Å resolution. The accelerated decay in these crystals, in spite of their intrinsic order, may be due to the presence of heavy atoms. For each diffractometer data set the complete shell of 6.0 Å resolution Bijvoet pairs was collected for scaling purposes. All reflections were collected consecutively at ±2θ as $hk\ell, \bar{h}\bar{k}\bar{\ell}$.

The data were reduced and merged to preserve the Bijvoet pair observations of each reflection. The polarity of each crystal mount was established from a subset of large Bijvoet differences. The data were scaled together in shells of $\sin\theta/\lambda$ with the use of anisotropic shape factors. The β_{11}, β_{22}, and β_{33} factors ranged from 0.97 to 1.03; the β_{12}, β_{13}, and β_{23} factors ranged from −0.10 to +0.10. Because all the crystals were mounted on c this variation probably arises from differences in path length through the square, bipyramidal crystals for reflections at different θ angles, and from variations in crystal volume (0.10 to 0.15 mm^3). The diffractometer data sets were corrected for absorption by the use of 00ℓ reflections and empirical ϕ scans. The oscillation camera data were not corrected for absorption, which may account for the larger R_{merge} compared to the diffractometer data (0.080), while the internal agreement within this data set was good (0.024) (Table 1) (*20*).

The combined data set showed a significant anomalous scattering signal (Table 1). For CuKα radiation, λ = 1.54 Å; the imaginary components of the anomalous scattering factors, f'', are: f'' (Cd) = 5.0, f'' (Zn) = 0.6, and f'' (S) = 0.6. The Bijvoet differences have been used to derive phase angles for the protein. At the same time, a survey of 80 compounds for heavy atom derivatives was unsuccessful. Soaked crystals were evaluated by precession photography. The survey included sulfhydryl-specific compounds, such as $[PtCl_4]^{2-}$, $Hg(CN)_2$, $CdCl_2$, which destroyed the diffraction pattern, and inert complexes and ions, such as $[Pt(CN)_4]^{2-}$, Sm^{3+}, and UO_2^{2+}, which had no effect, even at high concentration. One compound, $(NH_4)_2WS_4$, stained the crystals yellow and introduced significant isomorphous intensity differences. A complete 2.3 Å oscillation camera data set was collected and scaled to the native film data (R_{merge} = 0.12), but a consistent solution to the isomorphous difference Patterson map was not found. Assay of dissolved, soaked crystals showed 0.7 mol of tungsten per mole of protein. Presumably, $[WS_4]^{2-}$ has multiple, weak binding sites. Crystallization experiments with MT II reconstituted or substituted with other metals have not as yet yielded suitable isomorphous crystals. This survey has included samples of Zn_7 MT II, Cd_7 MT II, and Cd_5Pb_2 MT II: the latter yielded rectangular crystals too small to characterize.

Table 1. Cd, Zn MT II native data.

Item	3.0 Å diffractometer data*	2.3 Å film data†	Combined 2.3 Å data
Crystals (No.)	6	1	7
Reflections measured (No.)	7585	4816	12401
Independent reflections in *4/mmm*	1121	2373	2530‡
R merge§	0.035 to 0.065	0.024	0.080
R acentric (1773 Bijvoet pairs)			0.080
R centric (757 Friedel pairs)			0.044
Reflections ≥6.0 σ_F ∞ −2.3 Å			84%
Reflections ≥6.0 σ_F 2.5–2.3 Å			72%

*Enraf-Nonius CAD-4 diffractometer, extended counter arm, He-filled diffracted beam path, θ/2θ scans, intensities corrected for background, absorption, Lorentz and polarization effects. †Rigaku 12-kW rotating anode generator, Arndt-Wonnacott oscillation camera, 4.95° oscillations for 14 cassettes, crystal mounted on the c axis data processed by method of Rossmann (*20*), only whole reflections accepted. ‡Total $hk\ell$ to 2.3 Å is 2962. §$R = \Sigma||F|_i - |\bar{F}||/\Sigma|\bar{F}|$. *R* merge for equivalent reflections in 422. *R* centric for $0k\ell$, $hh\ell$, $hk0$ reflections, and *R* acentric for Bijvoet pair reflections in *4/mmm* after crystal to crystal scaling.

Direct methods. Initially, an image of the 4-Cd cluster was derived from the 2.3 Å film data set and by direct methods. The theory of Hauptman for anomalous scatterers in proteins (*21*) was encoded and applied to the normalized structure factors in both space groups $P4_12_12$ and $P4_32_12$. Out of 11 million three-phase structure invariants linking 2786 reflections (1393 Bijvoet pairs with $|E| > 0.30$), the 105,525 lowest variance estimates ($A > 1.00$) were selected for phasing. In each space group, 80 phase sets were generated by the multisolution method. The sets were evaluated by examining Bijvoet difference Fourier maps in descending order of figure of merit for each set. In the ninth set examined (absolute figure of merit, 0.668; space group $P4_12_12$), the four highest peaks formed a cluster with all six Cd-Cd distances in the range 3.6 to 4.4

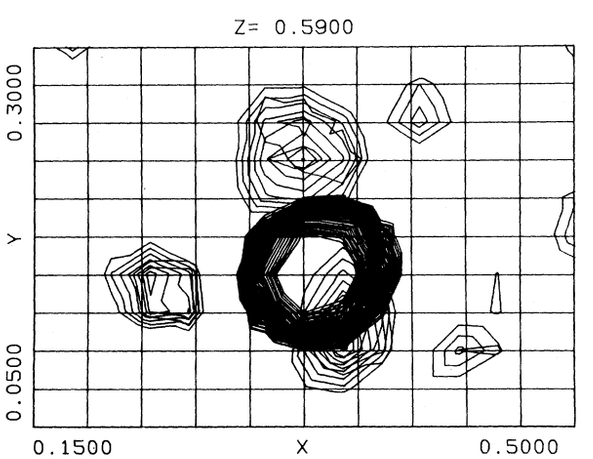

Fig. 1. Bijvoet difference Fourier map at 2.3 Å resolution based on direct methods phase set number 9 in $P4_12_12$ in which the four highest peaks are separated by distances in the range 3.6 to 4.4 Å.

Å (Fig. 1) (*22*). The four-atom model was consistent with a Patterson map based on the largest Bijvoet differences and is similar to a model proposed from ^{113}Cd NMR (*9*). In six of the eight higher ranked sets, the maps yielded only a single very high peak on a special position (twofold axis) and were deemed physically unrealistic. At this stage, there was no indication of a fifth Cd site in the direct methods map.

Subsequently, the 3 Å diffractometer data were collected. Although the short intracluster Cd-Cd vectors from the model were consistent with Bijvoet difference Patterson maps from both the film and diffractometer data sets, the indicated cluster location differed substantially in the two maps. The difference probably arises from the absence of blind region reflections along the c^* direction in the 2.3 Å film data. The required translation (~12 Å) was derived by an R value observed compared to calculated Bijvoet differences search against the 6.0 Å merged data on a 1 Å grid, which gave a minimum at $x = 0.59$, $y = 0.69$, and $z = 0.65$ for the center of gravity of the 4-Cd cluster. Although the translation map contained two R values only 1 percent higher than the minimum, the minimum position gave the best agreement with the observed Patterson map (Fig. 2a). After applying the translation, refinement (*23*) of the four sites as a rigid group against the 30 percent largest Bijvoet differences in the merged data set (Table 1) gave $R = 0.396$ at 3 Å resolution. Refinement as four individual sites gave $R = 0.371$, with shifts less than 0.85 Å for each atom. The location of a fifth Cd site was then determined from a single minimum in an R value search in which a Cd was translated relative to the fixed position of the 4-Cd cluster. Inclusion of the fifth Cd in refinement against the 30 percent largest Bijvoet differences reduced R to 0.332 at 3.0 Å resolution (*24*).

With the model structure at hand, it was possible to deconvolute the anomalous difference Patterson map, which previously had not been feasible because five Cd sites give rise to 100 equal-weighted

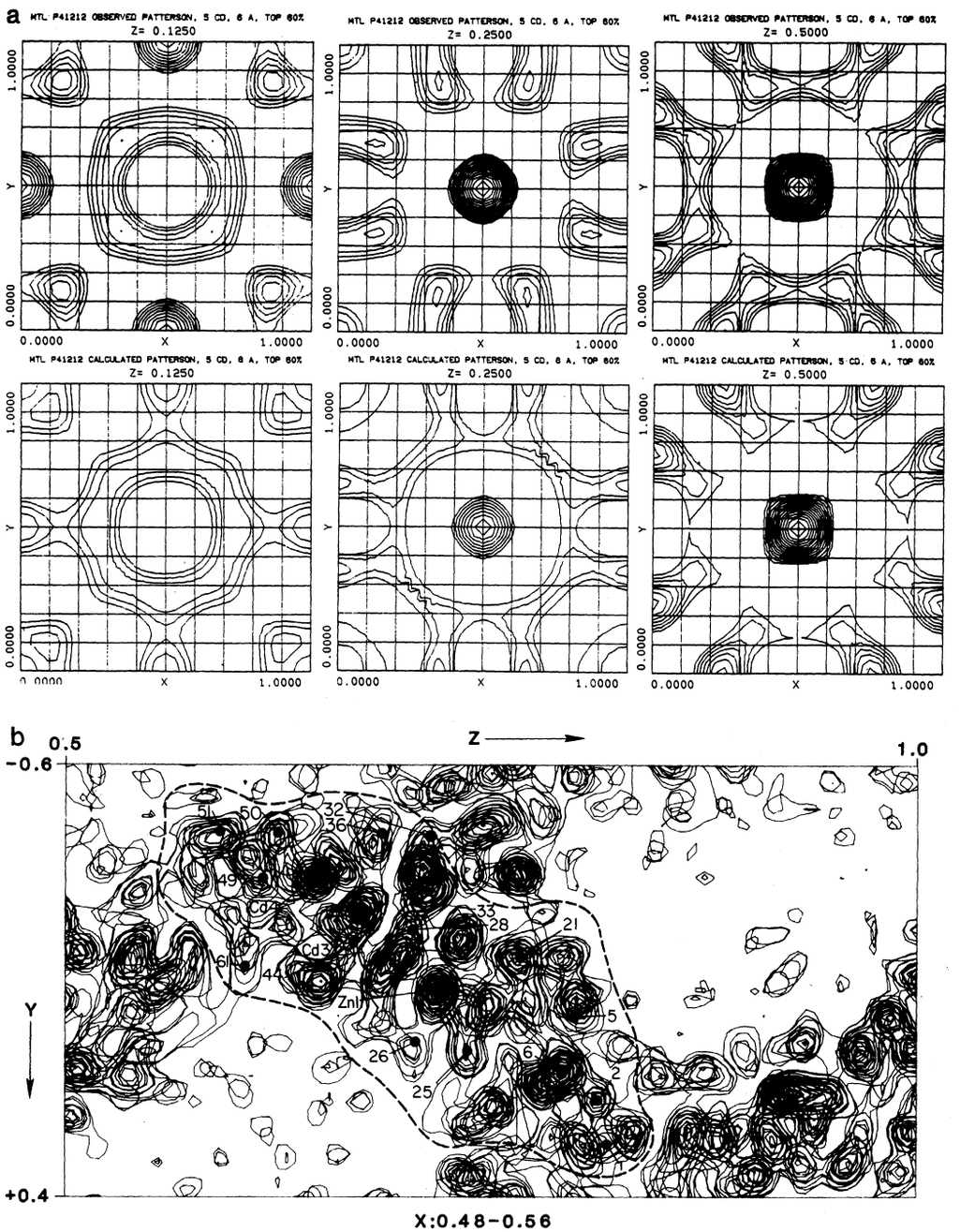

Fig. 2. (a) Three sections of the 6.0 Å resolution Bijvoet difference Patterson map calculated with the 60 percent largest anomalous differences. For each section, the upper map is calculated with the observed Bijvoet differences and the lower map is calculated with the same reflections and the five-site Cd model. The agreement of the two maps demonstrates that the Cd sites derived account for the experimentally observed anomalous scattering data. (b) Composite of five sections of the final 2.3 Å resolution ISAS electron density map. Intervals along the x axis are 0.6 Å, and each section is contoured with ten levels with respect to the highest peak in the map. The α-carbon positions and metal peaks are indicated, and the reference molecule is enclosed within the dashed line. The sections shown encompass 16 α-carbon positions, the Cd3, Cd4, and Zn1 sites of the reference molecule, and 32 percent of the volume of the asymmetric unit. Electron density outside the dashed line arises from symmetry related molecules in unit cell.

vectors per asymmetric unit. Sections of the observed and calculated 6.0 Å anomalous difference Patterson map are shown in Fig. 2a. The correlation coefficient between the two Patterson maps is 0.75. Because $z \simeq 5/8$ for the 4-Cd cluster, non-Harker vectors occur in the Harker sections at $w = 1/4$ and 1/2. Vectors from the fifth Cd to the 4-Cd cluster in $P4_12_12$ fit the 3.0 Å Patterson map, while those in $P4_32_12$ did not, confirming the space group assignment.

Electron density maps and refinement. The five refined Cd sites were used as a basis for iterative single wavelength anomalous scattering (ISAS) phase calculations (*25*) in both $P4_12_12$ (cluster at 0.59, 0.69, 0.65) and $P4_32_12$ (cluster at −0.59, −0.69, −0.65). Only 20 percent of the real part of the calculated heavy atom structure factors was used in computing the initial phase probability distributions for the 1773 acentric reflections. A solvent content of 50 percent was used in calculating the electron density filter. Initial estimates of phases for 757 centric reflections were obtained by phase extension following four cycles of map inversion and phase combination. The centric reflections and the initial ISAS acentric phases were then used to construct a new density filter, and four cycles of map inversion and phase combination on the acentric reflections were repeated. It was necessary to iterate through the entire process three times; that is, derive three density filters in order to reach convergence for the derived centric phase angles. Presumably, the starting density filter is degraded by the omission of the 757 centric reflections for which there is no experimental phase information. The figure of merit was consistently 1 to 2 percent higher in the $P4_12_12$ space group in agreement with the direct methods results.

Three 2.3 Å resolution electron density maps were calculated with the use of ISAS derived acentric and centric phases. The first map was based on the five-site Cd model and was used to interpret the positions of 11 sulfurs in the 4-Cd cluster. These appeared as tetrahedral connections of electron density from the protein to each Cd site. The presence of one long Cd-Cd distance (5.2 Å) indicated that only five of the Cd-Cd distances were linked by bridging sulfur ligands, necessitating six terminal ligands (*9*). Four ideal CdS_4 tetrahedra (Cd-S 2.5 Å, S-Cd-S 109.5°) were least-squares fit to the coordinates and refined as rigid bodies against the 30 percent largest 3.0 Å anomalous differences, resulting in decreases of 0.1 to 0.2 Å in the Cd-Cd distances. Only the translations of the CdS_4 tetrahedra were refined; refinement of the rotations failed to converge, presumably because of the large ratio of scattering factors, f'' (Cd)/f''(S) ~ 8/1. A calculated difference Patterson map at 3.0 Å with the 11 sulfurs included had very similar features to the map calculated on the basis of the Cd sites alone when either was compared to the observed 3.0 Å map. However, a second 2.3 Å electron density map with the four Cd and 11 sulfur sites from the rigid-body refinement, and the fifth Cd site, had a significantly improved connectivity within the protein region and better contrast compared with the solvent region. Apparently, at 2.3 Å the sulfurs are insufficiently resolved from Cd to be independently refined, but their presence affects the precise centers of the derived Cd sites. The 11 sulfur sites of the 4-Cd cluster were reinterpreted from the second map without major change in the model, and four sulfurs attaching to the fifth Cd site were interpreted. Although the other two metal sites of the three-metal cluster and their attaching sulfurs were apparent, they were not included in the model. The rigid body refinement was repeated with five CdS_4 tetrahedra, giving $R = 0.334$. A final 2.3 Å electron density map was calculated for 1773 acentric and 757 centric reflections with overall figure of merit of 0.67.

Structure factors were also calculated against the protein |F|'s, with the use of the five Cd and 15 sulfur sites as heavy atoms ($R = 0.53$ for observed protein structure factors, F_o, compared to calculated structure factors, F_c). Centric reflections with these calculated phases were weighted by the method of Sim (*26*) and spliced with acentric reflections having the ISAS phases and weights. The electron density map incorporating these heavy atom centric phases was virtually indistinguishable from the final ISAS map, confirming the validity of the ISAS derived centric phases.

The 2.3 Å electron density was interpreted with the use of a mini-map and a map on a scale of 1 cm/Å stacked in a Richard's box (*27*). The polypeptide chain was not traced upon inspection because the commonly recognizable features of α-helix and β-sheet were not present. However, the requirement for metal coordination placed stringent constraints on the interpretation. Consequently when model building into the electron density, the 28 cysteine sulfur to metal bonds and the sequence provided an unequivocally clear path through the density of the asymmetric unit (Fig. 2b). Coordinates of the model (407 protein atoms, 7 metals) were measured and refined (15 cycles) by restrained least squares (*28, 29*) in increasing steps of resolution (3.5, 3.0, 2.7, and 2.3 Å data) to $R = 0.398$ (1732 reflections > 6.0 σ_F, 5.0 to 2.3 Å, overall thermal factor $B = 15$ Å^2). Metal-sulfur bonds were constrained to 2.50 Å. The

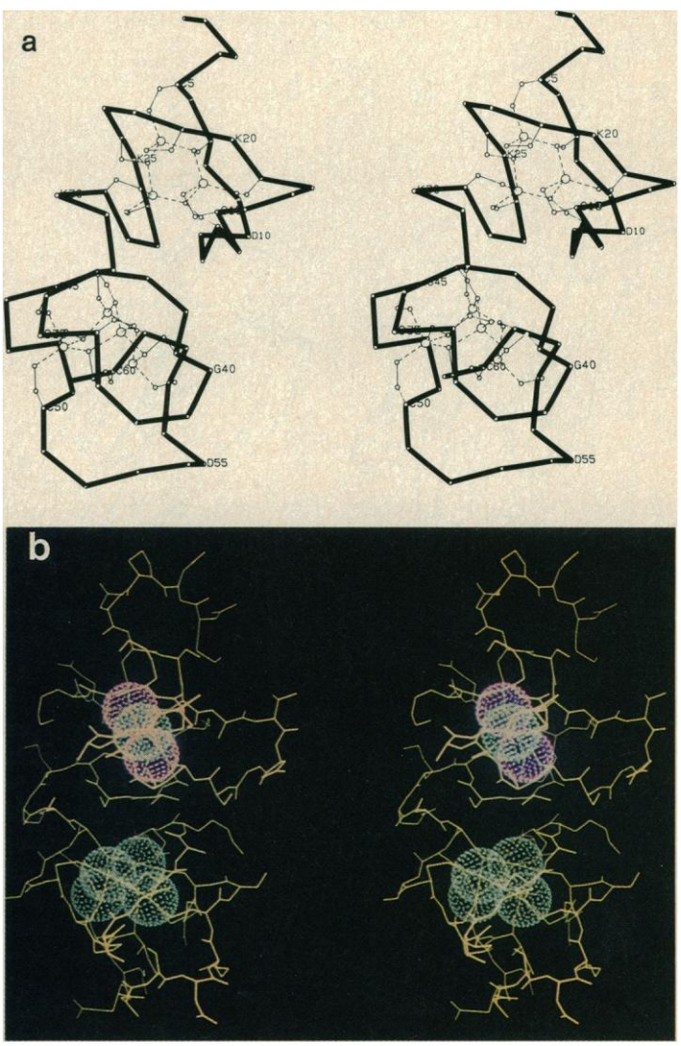

Fig. 3. (a) Stereofigure of the metallothionein structure showing the α-carbon, the β-carbon and the sulfur atoms of cysteine and Cd and Zn atoms. The α-carbon positions are labeled on every fifth residue. (b) Stereofigure of the structure showing all of the atoms. The Zn and Cd sites are highlighted with purple and green spheres, respectively. This view is roughly orthogonal to the view in (a). This figure was prepared by using graphics software written by M. L. Connolly and A. J. Olson (*42, 43*).

refined coordinates were used to calculate 27 residue-deleted, unbiased $2F_o - F_c$ Fourier maps: 19 structure factor calculations omitting residues 1 to 3, 4 to 6, and so forth; one omitting residues 58 to 61; and seven omitting each of seven metal(Cys)$_4$ centers. The entire structure was then refit to the 27 unbiased $2F_o - F_c$ maps with the use of interactive graphics. The $2F_o - F_c$ maps independently confirmed the sequence assignments of all 20 cysteines and the seven metal sites. The new coordinates were then refined to $R = 0.338$ (1732 reflections $> 6.0\ \sigma_F$, 5.0 to 2.3 Å, individual isotropic B factors). For this model, the root-mean-square (rms) deviations from ideality for bond, angle, and plane defining distances are 0.035, 0.094, and 0.034 Å, respectively, and B factors are 15.0 ± 2.1 Å ($\sigma_B = 2.0$ Å^2). Atomic coordinates have been deposited with the Protein Data Bank (*30*).

Description of structure. The α-carbon, cysteine side chain and metal cluster structure of MT II is shown in Fig. 3. A schematic representation of the protein folding is given in Fig. 4. The cysteine sulfur coordination of the metal clusters is depicted in Fig. 5 and listed in Table 2. The molecule folds into two domains of equal size; the NH_2-terminal domain of residues 1 to 29 (β) enfolds the three-metal cluster; the COOH-terminal domain of residues 30 to 61 (α) enfolds the 4-Cd cluster. Each domain is globular with a diameter of 15 to 20 Å; the domains are linked at residues 30 and 31 to form a molecule maximum linear dimension of 30 to 35 Å. The interface between α and β consists of contacts between C^{26}, T^{27}, and S^{28} with A^{42}, K^{43}, and C^{44} (see legend to Table 2) and contacts within the turn of residues 28 to 33 (Fig. 3).

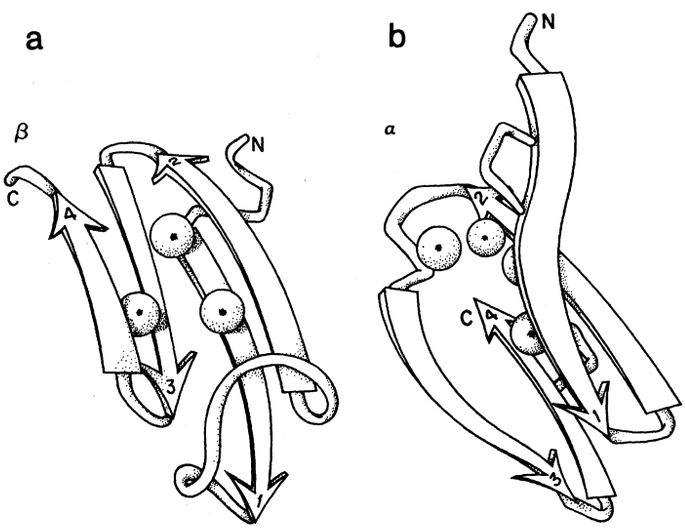

Fig. 4. Schematic drawing of the polypeptide chain folding of the (a) β and (b) α domains. Corresponding strands in the two domains are labeled 1 to 4. Strands 1 and 4 of the α domain are interrupted by "β bulges." This depiction of the structure depends on substituting cysteine-to-metal bonding in place of main chain–to–main chain hydrogen bonding.

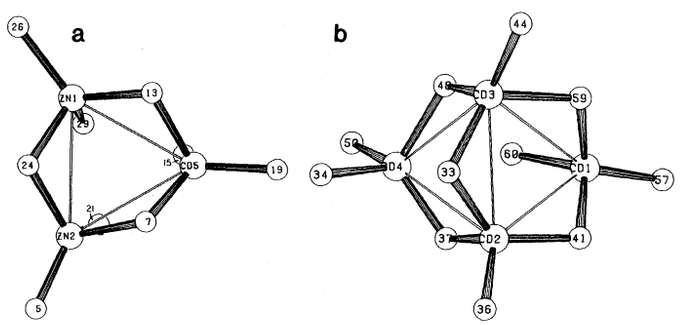

Fig. 5. Structures of the three-metal (a) and 4-Cd (b) clusters showing the Sγ and metal atoms only. The Sγ atoms are labeled with their cysteine residue numbers. Thin, open bonds represent metal-metal virtual bonds. This view of the three-metal cluster is inverted from that in Fig. 3a but is chosen to emphasize homology with the Cd1, Cd2, Cd3 portion of the 4-Cd cluster.

Because there is rather limited contact between α and β, it is reasonable to describe the folding in terms of the individual domains. If interaction from cysteine to the clusters is taken in place of normal main-chain hydrogen bonding, then each domain can be described as an antiparallel β sandwich (Fig. 4). In this view, the domains have homologous folds, such that strands 1 and 3 are parallel and strands 2 and 4 are parallel. There is no symmetry between the α and β domains, but simply a translational repeat of the domain fold along the long axis of the molecule. Another way to view the domain folding is as two turns of a single-stranded polypeptide spiral wrapping around a cluster (Fig. 3a). From this perspective, the spiral of the α domain is left-handed, while the spiral of the β domain is right-handed. Both domains contain a number of possible β- and γ-type reverse turns as is suggested from the α-carbon positions. Repeats of these turns, from G^{11} to A^{16}, and from S^{45} to C^{50}, form what appear to be short helical stretches, 2_7 and 3_{10}, respectively.

The structure of MT II is normal in that hydrophilic side chains are excluded from the interior of the molecule. The structure is unusual, however, in that it is essentially a monolayer of polypeptide wrapping around the metal-thiolate clusters. The clusters take the place of the hydrophobic core and hydrogen-bonded secondary structure of a larger protein. Ten of the 20 cysteines occur between C^{19} and C^{41} (Table 2). This segment of the sequence comprises a central region of the molecule bringing the two clusters into proximity at the α-β interface (minimum separation at Cd3–Zn1, 9.0 Å). The extreme economy of folding around the clusters is relaxed somewhat away from the α-β interface. In the α domain, a small hydrophobic pocket is formed below the 4-Cd cluster by the side chains of P^{38}, V^{39}, I^{49}, and A^{53}. The only residue, aside from the cysteines, whose side chain lies entirely within the interior of the structure is I^{49} (Fig. 3b). In the β domain, the side chains of K^{20} and K^{25} sandwich the NH_2-terminal chain at S^{6}C^{7}A^8, and K^{22} interacts with D^2 above the three-metal cluster. Although there is more lysine in β, there is no marked clustering of charged residues. There is, however, a definite tendency for peptide planes near the cluster to have the carbonyls oriented outward toward solvent, which directs the amides toward the γ-sulfur atoms of the cysteine. This kind of stereochemistry has been observed in rubredoxins and ferredoxins (*31*).

The metal cluster in the β domain consists of three bridging thiolate ligands, six terminal thiolates, and three metal sites (Figs. 3 and 5A and Table 2). From the anomalous scattering data, only one of these metal sites is occupied by Cd; from the crystal composition (*19*), the other two must therefore contain Zn. The three metal sites are tetrahedrally coordinated. The metal atoms form an equilateral triangle with distances 4.1 ± 0.1 Å. The six-ring atoms of the three-metal cluster, having tetrahedral valence angles, adopt a chair conformation. Conceptually, a chair conformation orients the terminal and bridging ligands toward the appropriate cysteines of the protein fold. The three axial cysteine ligands of the chair, C^{15}, C^{21}, and C^{29}, being more exposed by this folding of β, are protected to some extent by the α domain.

The metal cluster in the α domain consists of five bridging thiolate ligands, six terminal thiolates, and four Cd sites (Figs. 3 and 5b and Table 2). The metals are tetrahedrally coordinated. The metal atoms form a pair of triangles (Cd1, Cd2, Cd3 and Cd2, Cd3, Cd4), which

Table 2. Cysteine ligands in Cd, Zn MT II* amino acid sequence (*7*). The one-letter amino acid code is: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; G, glycine; I, isoleucine; K, lysine; M, methionine; N, asparagine; P, proline; Q, glutamine; S, serine; T, threonine; V, valine.

β domain																													
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
M	D	P	N	C	S	C	A	T	D	G	S	C	S	C	A	G	S	C	K	C	K	Q	C	K	C	T	S	C	K
				Zn2		Cd5 Zn2						Cd5 Zn1		Cd5				Cd5		Zn2			Zn1 Zn2		Zn1			Zn1	

α domain																														
31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61
K	S	C	C	S	C	C	P	V	G	C	A	K	C	S	Q	G	C	I	C	K	E	A	S	D	K	C	S	C	C	A
		Cd2 Cd3	Cd4		Cd2	Cd2 Cd4				Cd1 Cd2			Cd3				Cd3 Cd4		Cd4							Cd1		Cd1 Cd3	Cd1	

share an edge (Cd2-Cd3 virtual bond) and make a dihedral angle of 89°. The Cd–Cd distances at the bridging sulfurs, $S\gamma^{(59)}$, $S\gamma^{(33)}$, $S\gamma^{(41)}$, $S\gamma^{(37)}$, and $S\gamma^{(48)}$, are 3.9, 3.9, 4.1, 4.3, and 4.4 Å, respectively. The sixth Cd–Cd distance, Cd1–Cd4, is significantly longer at 5.2 Å and is not bridged by sulfur. The four Cd atoms form a distorted tetrahedron which could be described as a "butterfly." The five bridging sulfur atoms are arranged as a square pyramid with $S\gamma^{(33)}$ at the apex (Sγ–Sγ distances, 3.9 to 4.2 Å). The sulfur of C^{33}, the first cysteine in the sequence of α, is distinctive in the 4-Cd cluster as the bridging sulfur linking two triply-bridged Cd sites. Two six-atom rings are embedded in the four-metal cluster: a chair consisting of Cd1, Cd2, Cd3, $S\gamma^{(33)}$, $S\gamma^{(41)}$, and $S\gamma^{(59)}$; and a boat consisting of Cd2, Cd3, Cd4, $S\gamma^{(33)}$, $S\gamma^{(37)}$, and $S\gamma^{(48)}$.

The primary sequence of MT II contains redundant cysteine sequences, and it is of interest to consider these in relation to the clusters (Fig. 5, a and b). The sequence CXC occurs seven times (X, variable residue). Five of these sequences have one cysteine as a bridging thiolate and one cysteine as a terminal thiolate to the same metal site. Two others, $C^{19}XC^{21}$ and $C^{34}XC^{36}$, have both cysteines as terminal thiolate ligands to two metal centers. CXXC occurs at six positions (Table 2). Except for $C^{41}XXC^{44}$, these sequences have both cysteines coordinated to the same metal with at least one cysteine as a terminal ligand. The sequence CXXXC occurs four times; in each one both cysteines are coordinated to the same metal. In only four of 19 CC, CXC, CXXC, and CXXXC sequences are the cysteines not coordinated to a common metal. General stereochemical patterns for these cysteine sequences are not apparent at this resolution.

By contrast, the most widely spaced cysteines in the primary sequence span the longest metal-metal separation in either domain: C^{50} and C^{57} are terminal ligands of Cd1 and Cd4; this pair of Cd sites is the 5.2 Å unbridged edge of the 4-Cd cluster. The CXC sequences are often in an extended conformation, such that the first and third residues have their side chains oriented on the same side of the "β-sheet"—toward a cluster. However, CXXC is often a turn, returning the second cysteine to the same metal. The structure of $C^{19}KC^{21}KQC^{24}KC^{26}TSC^{29}$ illustrates these two patterns (Fig. 3a). The special peptide $C^{33}CSCC^{37}$ has $C^{33}XXC^{36}$ coordinating to Cd2 and $C^{34}XXC^{37}$ coordinating to Cd4. The homologous sequence $C^{57}SCC^{60}$ has $C^{57}XXC^{60}$ coordinated to Cd1.

The three largest noncysteinyl segments of the sequence participate in crystal packing contacts in the tetragonal unit cell. Residues of the two hydrophilic loops, $K^{51}EASDK^{56}$ and $A^{8}TDGS^{12}$, from adjacent molecules are in contact by 2_1-screw axis symmetry along *a(b)*. In particular, E^{52}, A^{53}, D^{10}, G^{11}, and S^{12} have distances between their α-carbons of less than 6 Å. $C^{36}CPVGC^{41}$ and $NacM^{1}DPN^{4}$ on a different pair of molecules come in contact about the diagonal twofold axis. Twofold related molecules are also in contact at K^{30}. The 4_1 screw axis related molecules have a head-to-tail interaction of NH_2- and COOH-termini. The crystal packing of MT II molecules, therefore, involves polar residues in loops and at the termini but there are few contacts involving the cysteines.

Correlation with data for protein in solution. The three-dimensional structure of MT II confirms the principal findings of biochemical and biophysical experiments on the molecule. The three- and four-metal clusters are linked by three- and five-bridging thiolate ligands, respectively, as deduced from ^{113}Cd homonuclear decoupling NMR data for rabbit liver MT II (*9*). The protein is folded into two domains, as demonstrated by subtilisin cleavage and sequence determination of rat liver MT II (*8*). All seven metal sites are tetrahedrally coordinated by cysteine sulfur, as shown on the basis of chemical exchange, and UV, visible, circular dichroic (CD), and electron paramagnetic resonance (EPR) spectroscopies (*1, 2, 10*). The α domain is the binding site of a 4-Cd cluster, as shown by ^{113}Cd NMR of the α domain (*11*) and cooperative binding of Cd to the α domain (*12*). The proposed connectivity of Cd and sulfur in the 4-Cd cluster (*9*) is in agreement with the crystal structure. Constraints derived from the same experiments also led to a model for the three-metal cluster (*9*). Again, the metal-sulfur connectivity of the cluster model and the crystal structure are the same. The discrepancy of the metal to cysteine assignments derived from ^{113}Cd-^{1}H NMR spectra (*18*) compared to the crystal structure suggests that reconstituted MT II may be able to adopt a different conformation than the protein as isolated.

Cleavage of MT II into domains by subtilisin occurs at $C^{29}K^{30}$ after incubation with EDTA for selective removal of Zn (*8*). In the structure this peptide bond lies in an exposed turn adjacent to the Zn1 site (Fig. 3a), so that we can visualize how removal of Zn might promote hydrolysis. The antigenic determinants of MT have been defined by radioimmunoassays to be the invariant sequence $NacM^{1}DPNC^{5}$ and residues 20 to 25 (*32*). In MT II, both sequences contain exposed turns; the side chains of K^{20}, K^{22}, and K^{25} may also be in proximity to the NH_2-terminus in solution. The overall dimensions of MT II in the crystal resemble a prolate ellipsoid, consistent with hydrodynamic experiments (*33*), although the axial ratio 2.5:1 is less than the 4:1 value calculated for high

ionic strength solutions (*34*). Although a ^{113}Cd NMR study of reconstituted rabbit liver MT II has indicated that multiple conformers exist in low ionic strength solutions (*35*), there is no evidence as yet for disorder in the crystal.

The metal binding specificity at the Cd5 site versus the two Zn sites in the β domain may be due to coordination by four consecutive cysteines in the sequence (Table 2), which includes $C^{15}XXXC^{19}$, the only five-residue loop in the β domain. In contrast, each of the Zn sites has a CXXC loop, analogous to the structural Zn site in liver alcohol dehydrogenase, which has the sequence $C^{97}GKC^{100}RVC^{103}$ and C^{111} (*36*). The rarity of CXC in Fe-S proteins (*37*) and the frequent occurrence of this sequence in MT II implies that CXC favors cysteine as a bridging ligand. The frequently occurring sequence CPV of Fe-S proteins occurs in MT II at C^{37}. Both the α and β domains of MT II also bind Cu(I) with a stoichiometry of six metals per domain; in the β domain this requires that all nine cysteines be bridging thiolates (*4, 14*). In view of the structure with one Cd, two Zn in the β and four Cd in the α domain, the data imply that the conformation of both domains must be different with Cu(I) bound (*4*).

The isolated α and β domains of MT II are capable of binding metals with the same stoichiometries as in the intact molecule (*38*), consistent with the limited contact between domains in the structure. Although the linking region from C^{19} to C^{41} contains a high proportion of conserved residues (*6, 7*), it is not clear from the structure whether this conservation reflects intradomain stereochemical requirements only, or whether it also involves specific interdomain interactions. The contacts between α and β are no greater than the packing contacts between symmetry-related domains in the lattice. The gene sequence of mouse MT I has exons corresponding to residues 1 to 9, 10 to 31, and 32 to 61 (*39*). Consistent with the hypothesis that exons code for intact protein domains (*40*), the third MT I exon correlates with the α domain structure. The separation of the β coding sequence into two exons is not obvious in the structure, except that strands 1 and 2 cross over differently in β and α (Fig. 4).

Amino acid sequences of mammalian MT isoforms are strongly conserved and the positions of all 20 cysteine residues are invariant (*1, 6, 7*). One would expect the structures of the homologous proteins to be similar to rat liver MT II, as indicated by the spectroscopic data for rabbit and equine MT isoforms (*1–3*). The sequence variations which do occur can be accommodated into the structure, and the most variable regions, residues 8 to 12 and 51 to 56 (*7*), correspond to exposed loops. Primary sequences of crab MT I and II contain 18 cysteines with changes at six cysteine positions compared to mammalian MT (*6*). The ^{113}Cd NMR for crab MT I establishes that both domains contain a three-metal, nine-cysteine cluster, in analogy with the β domain of MT II (*41*). While the changes at cysteine positions make it speculative to align the structure of MT II with crab MT I, the presence of a three-metal cluster in the α domain of the crab protein suggests an interesting comparison. As noted, the 4-Cd cluster has the chair conformation of a three-metal cluster embedded in it. The corresponding atoms are, respectively: Cd5, Zn2, Zn1 and the Sγ atoms of cysteines C^{24}, C^{7}, C^{13}, C^{5}, C^{26}, C^{19}, C^{21}, C^{29}, and C^{15} in β; and Cd1, Cd2, Cd3 and the Sγ atoms of C^{33}, C^{41}, C^{59}, C^{36}, C^{44}, C^{57}, C^{37}, C^{48} and C^{60} in α (Fig. 5). A least-squares fit of these 12 pairs of atoms gives an average discrepancy of 0.6 Å. This homology, and the fact that the ring of Cd2, Cd3, Cd4 has a boat conformation, suggest that Cd4 is the metal site deleted in the α domain cluster of crab MT I. In rat liver MT II, Cd4 is also the most exposed of the four sites in α (Fig. 3); in addition, this site has the 5.2 Å unbridged Cd–Cd distance with Cd1.

REFERENCES AND NOTES

1. J. H. R. Kägi and M. Nordberg, *Metallothionein* (Birkhauser, Basel, Switzerland, 1979).
2. M. Vašák and J. H. R. Kägi, in *Metal Ions in Biological Systems*, H. Siegel, Ed. (Dekker, New York, 1983), vol. 15, p. 213.
3. D. C. Dalgarno and I. M. Armitage, in *Advances in Inorganic Biochemistry*, G. L. Eichorn and L. Marzilli, Eds. (Elsevier, New York, 1984), vol. 6, p. 113.
4. K. B. Nielson, C. L. Atkin, D. R. Winge, *J. Biol. Chem.* **260**, 5342 (1985).
5. M. Karin, *Cell* **41**, 9 (1985); see also the references therein.
6. K. Lerch, D. Ammer, R. W. Olafson, *J. Biol. Chem.* **257**, 2420 (1982).
7. D. R. Winge, K. B. Nielson, R. D. Zeikus, W. R. Gray, *ibid.* **259**, 11419 (1984).
8. D. R. Winge and K. A. Miklossy, *ibid.* **257**, 3471 (1982).
9. J. D. Otvos and I. M. Armitage, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7094 (1980).
10. M. Vašák and J. H. R. Kägi, *ibid.* **78**, 6709 (1981).
11. Y. Boulanger, I. M. Armitage, K. A. Miklossy, D. R. Winge, *J. Biol. Chem.* **257**, 13717 (1982).
12. K. B. Nielson and D. R. Winge, *ibid.* **258**, 13063 (1983).
13. R. W. Briggs and I. M. Armitage, *ibid.* **257**, 1259 (1982).
14. K. B. Nielson and D. R. Winge, *ibid.* **259**, 4941 (1984).
15. J. D. Otvos, H. R. Engeseth, S. J. Wehrli, *J. Magn. Reson.* **61**, 579 (1985).
16. D. Live, I. M. Armitage, D. C. Dalgarno, D. J. Cowburn, *J. Am. Chem. Soc.* **107**, 1775 (1985).
17. D. Neuhaus, G. Wagner, M. Vašák, J. H. R. Kägi, K. Wüthrich, *Eur. J. Biochem.* **143**, 659 (1984); *ibid.* **151**, 257 (1985).
18. M. H. Frey *et al.*, *J. Am. Chem. Soc.* **107**, 6847 (1985).
19. K. A. Melis, D. C. Carter, C. D. Stout, D. R. Winge, *J. Biol. Chem.* **258**, 6255 (1983).
20. M. G. Rossmann, *J. Appl. Cryst.* **12**, 225 (1979).
21. H. Hauptman, *Acta Cryst.* **A38**, 632 (1982).
22. W. Furey, A. H. Robbins, C. D. Stout, *13th International Union of Crystallography Congress, Abstracts* (Hamburg, Germany, 9 to 18 August 1984), No. 01.2–5.
23. W. A. Hendrickson and M. M. Teeter, *Nature (London)* **290**, 107 (1981).
24. C. D. Stout, W. F. Furey, A. H. Robbins, L. L. Clancy, B. C. Wang, *International Chemical Congress of Pacific Basin Societies, Abstracts* (Honolulu, 16 to 21 December 1984), No. 04B12.
25. B. C. Wang, in *Diffraction Methods in Biological Macromolecules* (a volume of *Methods of Enzymology*), H. Wyckoff, Ed. (Academic Press, New York, 1985), p. 90.
26. G. A. Sim, *Acta Cryst.* **13**, 511 (1960).
27. F. M. Richards, *J. Mol. Biol.* **37**, 225 (1968).
28. W. A. Hendrickson and J. H. Konnert, in *Biomolecular Structure, Function, Conformation and Evolution*, R. Srinivasan, Ed. (Indian Academy of Sciences, Bangalore, 1980), vol. 1, p. 43.
29. W. Furey, B. C. Wang, M. Sax, *J. Appl. Cryst.* **15**, 160 (1982).
30. Coordinates deposited 12 June 1985 with the Protein Data Bank, Brookhaven National Laboratory, Upton, NY, 11973.
31. E. Adman, K. D. Watenpaugh, L. H. Jensen, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4854 (1975).
32. D. R. Winge and J. S. Garvey, *ibid.* **80**, 2472 (1983).
33. J. H. R. Kägi, S. R. Himmelhoch, P. D. Whanger, J. L. Bethune, B. L. Vallee, *J. Biol. Chem.* **249**, 3537 (1974).
34. M. Vašák, C. Berger, J. H. R. Kägi, *FEBS Lett.* **168**, 174 (1984).
35. M. Vašák, G. E. Hawkes, J. K. Nicholson, P. J. Sadler, *Biochemistry* **24**, 740 (1985).
36. H. Eklund *et al.*, *J. Mol. Biol.* **102**, 27 (1976).
37. C. D. Stout, in *Iron Sulfur Proteins* (a volume of *Metal Ions in Biology*), T. G. Spiro, Ed. (Wiley-Interscience, New York, 1982), vol. 4, p. 97.
38. K. B. Nielson and D. R. Winge, *J. Biol. Chem.* **260**, 8698 (1985).
39. N. Glanville, D. M. Durnam, R. D. Palmiter, *Nature (London)* **292**, 267 (1981).
40. W. Gilbert, *ibid.* **271**, 501 (1978).
41. J. D. Otvos, R. W. Olafson, I. M. Armitage, *J. Biol. Chem.* **257**, 2427 (1982).
42. M. L. Connolly and A. J. Olson, *Comput. Chem.* **9**, 1 (1985).
43. M.L. Connolly, *Science* **221**, 709 (1983).

44. We thank Drs. J. Rosenberg and C. Frederick for the use of a rotating anode generator and film scanner. Supported by NIH grant GM-36535. This is contribution number 4076-MB from the Department of Molecular Biology, Research Institute for Scripps Clinic, La Jolla, CA.

19 August 1985; accepted 30 December 1985