Acta Cryst. (1994). D**50**, 186–191

Ta_6Br_{14} is a Useful Cluster Compound for Isomorphous Replacement in Protein Crystallography

BY GUNTER SCHNEIDER AND YLVA LINDQVIST

Department of Molecular Biology, Swedish University of Agricultural Sciences, Uppsala Biomedical Center, Box 590, S-751 24 Uppsala, Sweden

(*Received* 10 *March* 1993; *accepted* 14 *September* 1993)

Abstract

The metal cluster Ta_6Br_{14} has been used to prepare heavy-metal derivatives of two large proteins, ribulose-1,5-bisphosphate carboxylase/oxygenase and transketolase. In both cases, this cluster compound produced a single-site derivative for which a difference Patterson map, calculated to 5.5 Å resolution, could be solved straightforwardly. Ta_6Br_{14} provided enough phase information to unambiguously locate the heavy-atom positions in other multiple-site derivatives. In transketolase, the heavy-metal complex binds at the surface of the protein in a dominantly hydrophobic pocket. In ribulose bisphosphate carboxylase/oxygenase, it binds between two molecules in the crystal lattice. There are negatively charged glutamic and/or aspartic acid residues in the vicinity of the bound clusters. Ta_6Br_{14} is useful over a wide range of pH. For large proteins and/or large unit cells, this compound should be included in the initial screening for heavy-metal derivatives.

Introduction

One of the bottlenecks in the structure determination of large proteins is the preparation of suitable heavy-metal derivatives. In many cases such derivatives can be prepared, but because of the size of the protein, heavy-metal compounds very often bind at several sites, resulting in non-trivial difference Patterson maps. The correct localization of the first metal site can, therefore, be a rate-limiting step in the structure solution. In the event of a single-site substitution, the problem is in the size of the heavy-atom contribution, *e.g.* binding of the heavy metal might not cause enough changes in the diffracted intensities to result in reliable measurements of these differences.

As a remedy to these problems in the application of isomorphous replacement methods to large proteins, heavy-metal cluster compounds were introduced several decades ago. The unique properties of these complexes are that they contain a high number of electrons and, because of their size, tend to bind only to few specific sites on the surface of proteins.

Despite an early report on the use of large heavy-metal complexes in protein crystallography, successful applications of this type of compound in multiple isomorphous replacement are scarce. Corey and co-workers (Corey, Stanford, Marsh, Leung & Kay, 1962; Stanford, Marsh & Corey, 1962) co-crystallized lysozyme with Ta_6Cl_{14} and Nb_6Cl_{14}, respectively, and could also show that these cluster compounds bind in a specific manner. The difference in the number of electrons between the two complexes was sufficient to locate the cluster from a difference Patterson map. However, because of non-isomorphism between these crystals and crystals of native lysozyme, the cluster compounds were not used in the subsequent structure determination of lysozyme. More recently, a smaller heavy-metal complex, tetrakis-(acetoxymercuri)methane (TAMM) was used successfully in the structure determination of several large proteins and protein complexes, *e.g.* glutathione transferase (Reinemer *et al.*, 1991), the photosynthetic reaction center from *Rhodopseudomonas viridis* (Deisenhofer, Epp, Miki, Huber & Michel, 1984), the nucleosomal core histone octamer (Richmond, Finch, Rushton, Rhodes & Klug, 1984; O'Halloran, Lippard, Richmond & Klug, 1987) and an idiotype–anti-idiotype complex (Bentley, Boulot, Riottot & Poljak, 1990). The tungsten cluster $\{[W_3O_2(O_2CCH_3)_6](H_2O)_3\}(CF_3SO_3)_2$ was used in the structure determination of heavy riboflavin synthase (Ladenstein *et al.*, 1988). It has been observed that the symmetry of tungsten cluster compounds correlates well with the symmetry of their binding sites in the protein, *e.g.* trigonal and pentagonal tungsten clusters bind at sites on the threefold and fivefold symmetry axes of the protein, respectively (Ladenstein, Bacher & Huber, 1987). Gold clusters are presently being used to prepare heavy-metal derivatives of large protein complexes such as ribosomes (Yonath, Saper, Frolow, Makowski & Wittman, 1986; Boeckh & Wittman, 1991).

In this report, we summarize our experience with the heavy-metal cluster Ta_6Br_{14} in the structure determination of two large proteins, ribulose-1,5-bisphosphate carboxylase (Rubisco, 2 × 500 residues

in the asymmetric unit) (Schneider, Lindqvist, Brändén & Lorimer, 1986) and transketolase (2 × 680 residues in the asymmetric unit) (Lindqvist, Schneider, Ermler & Sundström, 1992). The primary intention in using these cluster compounds was to prepare single-site derivatives which could provide initial phase information for the subsequent localization of the heavy-metal positions in other multiple-site derivatives.

Materials and methods

Heavy-metal cluster

$Ta_6Br_{14}.7H_2O$ (dodeca-μ-bromo-hexatantalum dibromide) is a stable compound and its synthesis has been described (Kuhn & McCarley, 1965; Brauer, 1981). In short, the synthesis of the cluster compound consists of the following steps. $TaBr_5$ (available from Alfa) is heated together with aluminium foil in a sealed, evacuated quartz tube in a temperature gradient ranging from 723 to 553 K for 2 d. $AlBr_3$ and excess $TaBr_5$ are then removed by sublimation from the product at 553 K. Subsequently, the mixture is heated to 843 K for 5 h. The remaining product consists of a deep green powder of Ta_6Br_{14}.

The molecular structures of the analogues Nb_6Cl_{14} and Ta_6Cl_{14} were elucidated by Pauling and coworkers (Vaughan, Sturdivant & Pauling, 1950). The $Ta_6Br_{12}^{2+}$ cation consists of an octahedron of tantalum ions, surrounded by 12 bromine ions. Binding of one $Ta_6Br_{12}^{2+}$ ion adds 856 electrons to a protein, which results in a considerable contribution to the scattering power even for large proteins.

Because of its symmetric nature, the cluster can, in principle, bind to a protein in a number of different, equivalent orientations. It has been shown that within 5 Å resolution limit, the combined scattering from six octahedrally arranged ions is practically independent of the orientation of the octahedron (Stanford, 1962).

Preparation of heavy-metal derivatives and data collection

Crystals of Rubisco (Schneider, Brändén & Lorimer, 1986) and transketolase (Schneider, Lindqvist & Sundström, 1989) were soaked in a solution of mother liquor containing 1.1 m*M* Ta_6Br_{14} for 4–5 d.

Diffraction data for native and derivatized Rubisco crystals were collected on a four-circle STOE diffractometer at 277 K. X-ray data for native and derivative crystals of transketolase were collected on a Xentronics area detector, mounted on a Rigaku rotating anode. Details of the data-collection statistics are given in Table 1. Data scaling, merging and all other crystallographic calculations were performed with the program *PROTEIN* (Steigemann, 1974).

Phase refinement

Refinement of the heavy-atom parameters was carried out with a program originally written by M. Rossmann and modified by L. Ten Eyck and S.

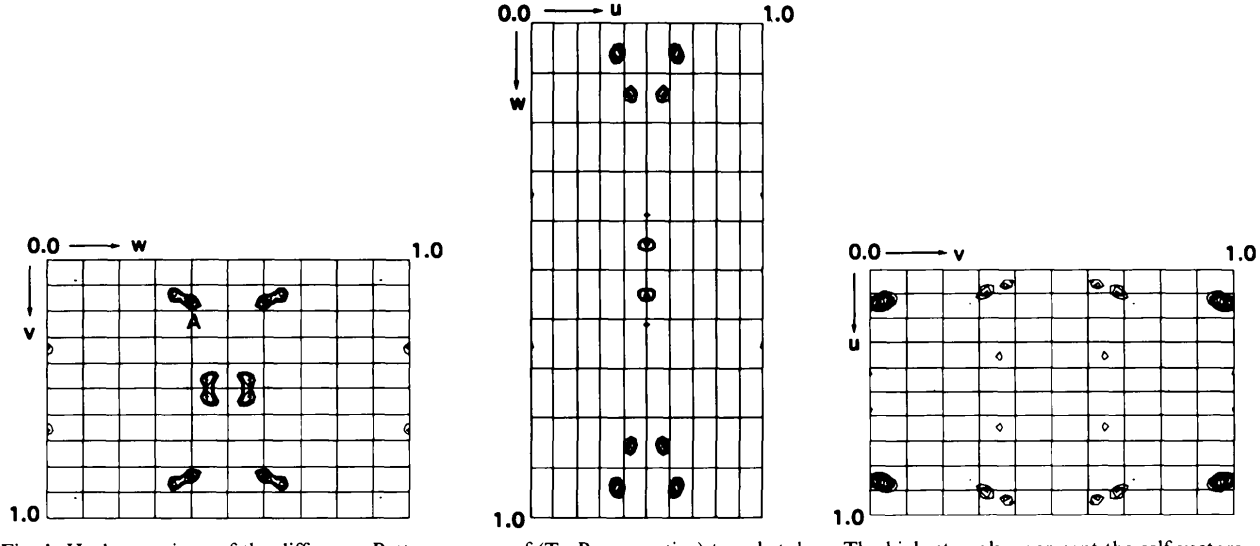

Fig. 1. Harker sections of the difference Patterson map of (Ta_6Br_{14} − native) transketolase. The highest peaks represent the self vectors for the two cluster positions. The lowest contour level is two times the standard deviation (σ) of the map, contoured in steps of 0.5σ. Peak *A* in the section $u = 1/2$ at $v = 0.161$ and $w = 0.397$ corresponds to a cross vector between the two cluster positions.

Table 1. *Data collection*

Data set	No. of crystals	Resolution (Å)	Measured reflections	Unique reflections (% of total)	R_{merge}*	R_{deriv}†
Transketolase						
NATI	2	2.9	71017	23796 (75)	0.078	
Ta_6Br_{14}	1	5.5	17192	4195 (71)	0.067	0.154
Rubisco						
NATI	2	5.0	4316	4227 (99)	0.036	
Ta_6Br_{14}	1	5.5	3203	3082 (95)	0.024	0.257

* $R_{merge} = \sum\sum_i |I_i - \langle I \rangle| / \sum |\langle I \rangle|$ where I_i are the intensity measurements for a reflection and $\langle I \rangle$ is the mean value for this reflection.
† $R_{deriv} = \sum |F_{PH} - F_P| / \sum |F_P|$ where F_{PH} is the structure-factor amplitude of the derivative crystal and F_P is that of the native.

Table 2. *Phasing statistics*

Protein	No. of sites	R_{Cullis}*	Phasing power†
Transketolase‡	2	0.61	1.65
Transketolase§	2	0.56	2.08
Rubisco‡	2	0.69	1.56
Rubisco§	2	0.63	1.81

* $R_{Cullis} = \sum ||F_{PH} \pm F_P| - |F_H(\text{calc})|| / \sum |F_{PH} - F_P|$, where F_{PH} and F_P are defined as in Table 1 and F_H(calc) is the calculated heavy-atom structure-factor amplitude summed over centric reflections only.

† Phasing power = $F(H)/E$, the r.m.s. heavy-atom structure-factor amplitudes divided by the lack of closure error.

‡ Cluster treated as as single point scatterer in phase refinement and phase calculation.

§ Cluster approximated as six point scatterers, arranged in octahedral geometry.

Remington. The program uses the least-squares procedure described by Dickerson, Kendrew & Strandberg (1961) and is incorporated in the *PROTEIN* program package. The $Ta_6Br_{12}^{2+}$ cluster was treated in two different ways. In one type of calculation, it was considered a single point scatterer and no attempt was made to model the structure of the complex. In all calculations, the temperature factor used for the Ta atoms was 30 Å^2. In one case (Rubisco), calculations were also performed using a temperature factor of 75 Å^2 to partially account for the larger volume of the cluster as compared to a single point scatterer. In the second type of calculation, phase refinement and phase calculations were performed using a Ta_6 octahedron, which is a more appropriate model of $Ta_6Br_{12}^{2+}$. In this approach, the octahedron was placed in random orientation with its center of gravity at the position of the single point scatterer.

Results

Transketolase from Saccharomyces cerevisiae

Transketolase from *S. cerevisiae*, a thiamin-dependent enzyme, consists of two polypeptide chains of 680 residues (Sundström, Lindqvist, Schneider, Hellman & Ronne, 1994). The enzyme crystallizes in space group $P2_12_12_1$ with cell dimensions $a = 76.3$, $b = 113.3$ and $c = 160.9$ Å. The crystal asymmetric unit contains the dimer. Soaking of transketolase crystals in a 1.1 m*M* Ta_6Br_{14} solution resulted in a green crystals and X-ray data for these were collected to 5.5 Å. A difference Patterson map calculated on data for the Ta_6Br_{14} derivative and the native crystal (Fig. 1) can be easily interpreted in terms of one binding site per subunit. The two $Ta_6Br_{12}^{2+}$ ions bind at sites that obey the non-crystallographic symmetry operator relating the two protein subunits. After refinement of the heavy-metal parameters (Table 2), SIR phases were calculated and difference Fourier maps for other derivatives were calculated. In all of these maps, the highest

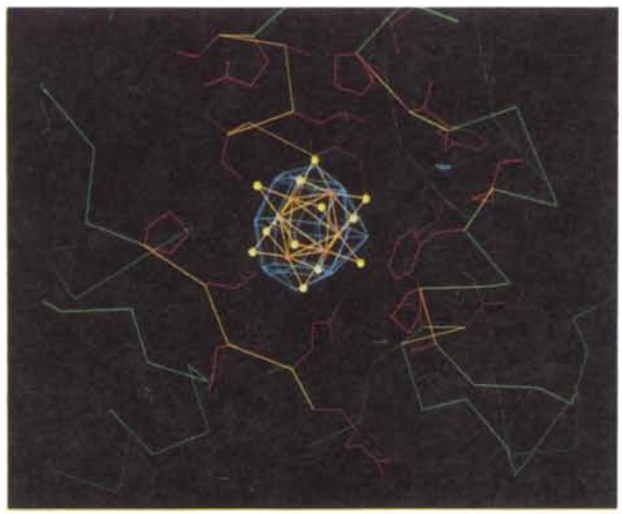

Fig. 2. View of the binding site for the $Ta_6Br_{12}^{2+}$ ion in transketolase. Tantalum ions are shown in red and bromine ions in yellow. The difference electron density in an $|F_o| - |F_c|$ electron-density map, calculated with Ta_6Br_{14} transketolase data and phases calculated from the model of holotransketolase is superimposed in blue, contoured at three times the standard deviation of the map.

Table 3. *Comparison of peaks in difference Fourier electron-density maps for a mercury derivative of transketolase, methoxyethylmercuric chloride*

Phases were SIR phases calculated from the refined tantalum positions. Case 1: cluster treated as point scatterer; case 2: cluster modeled as octahedron of six point scatterers.

	Case I		Case 2	
Heavy metal	Peak height*	Peak/σ	Peak height	Peak/σ
Hg(1)	4.5	10.9	4.7	11.3
Hg(2)	3.9	9.5	4.2	10.1
Ta(1)	3.6	8.7	3.4	8.2
Hg(3)	3.0	7.2	3.1	7.5
Ta(2)	2.4	5.8	2.2	5.3
Hg(4)	1.7	4.2	2.0	4.8

* Arbitrary units.

peaks corresponded to the positions of the heavy-metal ions. Even for a UO_2Cl_2 derivative with multiple sites and low occupancies, the uranium positions could be located as the highest peaks in the difference Fourier maps.

The two ways of treating the cluster in the phase-refinement procedure gave similar results. In both cases, the difference Fourier electron-density maps for the derivatives, based on SIR phases from the tantalum derivative, clearly had the unknown heavy-metal sites as the highest peaks, with the exception of two 'ghost' peaks for the tantalum positions (Table 3). However, as judged from both the refinement statistics and signal-to-noise ratio for the peaks in the difference Fourier maps, the phase calculation based on the more realistic model of the Ta_6 octahedron resulted in more accurate phases. Based on these phases, maxima in the difference Fourier maps, corresponding to the unknown heavy-metal positions, have a significantly higher signal-to-noise ratio. This is especially true for the weaker of the four mercury sites, Hg(4) (Table 3). Also, the spurious signal for the tantalum positions is lower in these maps. Clearly, even at the low resolution, treatment of the cluster compound as a single point scatterer is inadequate and the phases obtained are of better quality when the cluster is at least partially modeled.

A difference Fourier map was calculated to check the solution of the difference Patterson map, based on the three-dimensional model for transketolase (Lindqvist, Schneider, Ermler & Sundström, 1992). In the $|F_o| - |F_c|$ electron-density map, the two highest peaks (ten times and eight times the standard deviation of the map, respectively) are at the expected $Ta_6Br_{12}^{2+}$ binding sites and a model of the cluster compound fits well into the spherical density (Fig. 2). The cluster binds in a predominantly hydrophobic pocket at the surface of the protein through mostly non-specific interactions. The binding site is built up of residues Ala355, Leu531, Pro532, Leu568, Val570, Ala642, Pro643, Phe646, Pro653 and the more polar residues Gln529 and Asn530. The two positive charges of the $Ta_6Br_{12}^{2+}$ cluster ion are compensated for by the side chains of Glu565 and Glu571 which are in the proximity of the cluster (Fig. 2).

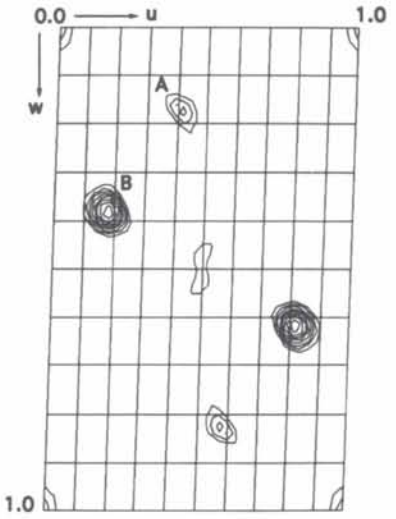

Fig. 3. Harker section $v = 1/2$ of the difference Patterson map of (Ta_6Br_{14} − native) Rubisco. The contour level is two times the standard deviation (σ) of the map, in steps, of 0.5σ. Peak *A* at position $u = 0.41$, $w = 0.157$ represents a self vector and the strong peak *B* at $u = 0.18$, $w = 0.38$ a self vector overlapping with a cross vector between the two cluster positions.

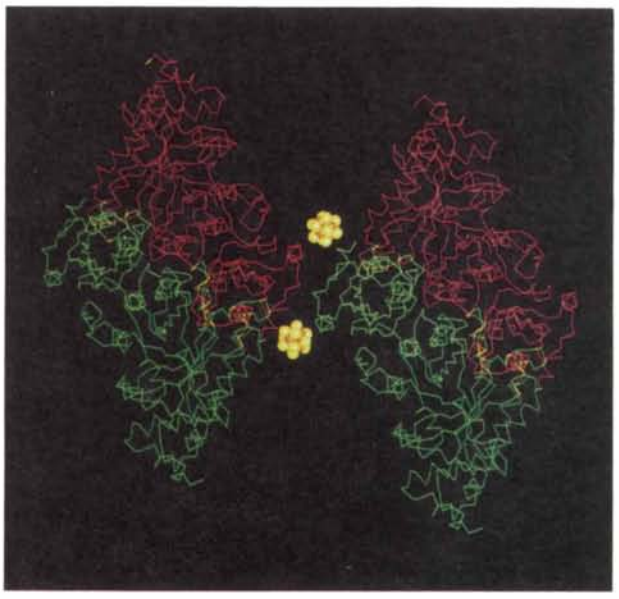

Fig. 4. Positions of the major binding sites of $Ta_6Br_{12}^{2+}$ in crystals of Rubisco from *Rhodospirillum rubrum*. The two subunits in the Rubisco dimer are color coded differently. The $Ta_6Br_{12}^{2+}$ ions are shown as CPK models.

Rubisco from Rhodospirillum rubrum

Rubisco from the photosynthetic bacterium *R. rubrum* is a dimer of identical subunits of molecular weight 54 kDa per subunit. The enzyme crystallizes in space group $P2_1$ with cell dimensions $a = 65.5$, b 70.6 and $c = 104.1$ Å, $\beta = 92.1°$ (Schneider, Brändén & Lorimer, 1986). Soaking of native crystals in a solution of 1.1 m*M* Ta_6Br_{14} results in green–black crystals. The soaking procedure did not change the cell dimensions of the crystals and an X-ray data set was collected on a diffractometer to 5.5 Å resolution.

Fig. 3 shows the Harker section for the (Ta_6Br_{14} – native Rubisco) difference Patterson map. The Harker section contains two peaks above 3σ. Peak *A* at 0.41, 0.157 corresponds to a self vector. The strong peak *B* at 0.18, 0.38 corresponds to a self vector for the second site, overlapping with a cross vector between the two cluster positions. The parameters for the two heavy-metal clusters were refined in the usual way (Table 2).

Difference Fourier maps, based on SIR phases from the Ta_6Br_{14} derivative, were used to locate the heavy-metal positions in the other derivatives. In all the cases tested, the highest peaks in the difference Fourier maps corresponded to a heavy-metal sites in the derivatives. Amongst these cases was a $KAu(CN)_2$ derivative, which contained multiple sites with low occupancies, making the localization of the sites from a difference Patterson map rather difficult. Assigning a larger temperature factor for the Ta atoms in the structure-factor calculations (75 instead of 30 Å^2) gave no significant improvement in the signal-to-noise ratio of these difference Fourier maps. However, as with transketolase, phase refinement and phase calculation using a tantalum octahedron as a model of the complex resulted in better refinement statistics (Table 2) and an improved signal-to-noise ratio in the difference Fourier maps, calculated with the SIR phases (data not shown). Even at low resolution, this approach seems to represent the cluster better than simple smearing out of the point scatterer by assigning a large temperature factor.

To further analyze the binding site, we calculated a difference Fourier map for the Ta_6Br_{14} derivative with phases calculated from the refined model of Rubisco (Schneider, Lindqvist & Lundqvist, 1990). In this difference electron-density map, the highest peaks (seven times the standard deviation of the map) are at the two sites determined from the analysis of the difference Patterson map. There are, however, five other peaks in the difference Fourier map above four times the standard deviation of the map. These positions do not have a corresponding self vector in the Harker section of the difference Patterson map above a 2σ threshold. Three of these peaks are found at the surface of the protein at suitable binding sites and probably represent minor binding sites of the $Ta_6Br_{12}^{2+}$ ion with low occupancies.

The two major sites in the Ta_6Br_{14} derivative of Rubisco obey the non-crystallographic symmetry relating the two subunits. However, in crystals of Rubisco, the two cluster ions bind at the interface between two molecules in the crystal lattice and are close to residues from both molecules (Fig. 4). Both these binding sites show less interaction between the $Ta_6Br_{12}^{2+}$ ion and the protein than is seen in transketolase. Contrary to the transketolase case, where a number of hydrophobic residues are in contact with the metal cluster, mostly polar residues – in particular the side chains of Gln3, Asp15 and Glu16 – are in the proximity of the $Ta_6Br_{12}^{2+}$ ion and probably participate in the binding of the complex.

Discussion

The present study confirms that large cluster compounds can be employed successfully to prepare heavy-atom derivatives for large proteins. Such derivatives generally contain only a few sites which allows the straightforward determination of initial phase information. In the two cases studied, the phase information has been of sufficient quality to determine the positions of heavy metals in derivatives with multiple sites and low occupancies. The work on Rubisco and transketolase has been carried out at pH values as different as 5.6 and 7.8, which illustrates the broad range of pH values in which the Ta_6Br_{14} cluster can be used.

So far, Ta_6Br_{14} has only been employed to obtain phase information at low resolution. The simplistic treatment of the cluster as a point scatterer will certainly be inadequate at higher resolution. Also, the approximation that the combined scattering of the tantalum ions is independent of the orientation of the cluster breaks down at high resolution. Thus for more precise structure-factor calculations, the orientation of the cluster must be determined. It has not yet been shown, however, whether the cluster does in fact bind in defined orientation to protein sites. Given the symmetrical arrangement of atoms in $Ta_6Br_{12}^{2+}$ and the observed lack of specific interactions between the cluster ion and protein, this might not be the case.

In conclusion, from the positive experience with the use of the compound Ta_6Br_{14} in isomorphous replacement with medium-sized or large proteins, we suggest that it should be included in a screen for derivatives, if a single-site derivative is desired for initial phase information.

We gratefully acknowledge the gift of a sample of Ta_6Br_{14} from Professor Robert Huber, München. This work was supported by a grant from the Swedish Natural Science Research Council.

References

BENTLEY, G. A., BOULOT, G., RIOTTOT, M. M. & POLJAK, R. J. (1990). *Nature (London)*, **348**, 254–257.

BOECKH, T. & WITTMAN, H.-G. (1991). *Biochim. Biophys. Acta*, **1075**, 50–55.

BRAUER, G. (1981). *Handbuch der Präparativen Anorganischen Chemie*, 3 ed., Vol. III, p. 1454. Stuttgart: F. Enke Verlag.

COREY, R. B., STANFORD, R. H., MARSH, R. E., LEUNG, Y. C. & KAY, L. M. (1962). *Acta Cryst.* **15**, 1157–1163.

DEISENHOFER, J., EPP, O., MIKI, K., HUBER, R. & MICHEL, H. (1984). *J. Mol. Biol.* **180**, 385–398.

DICKERSON, R. E., KENDREW, J. C. & STRANDBERG, B. E. (1961). *Acta Cryst.* **14**, 1188.

KUHN, P. H. & MCCARLEY, R. E. (1965). *Inorg. Chem.* **4**, 1482–1486.

LADENSTEIN, R., BACHER, A. & HUBER, R. (1987). *J. Mol. Biol.* **195**, 751–753.

LADENSTEIN, R., SCHNEIDER, M., HUBER, R., BARTUNIK, H.-D., WILSON, K., SCHOTT, K. & BACHER, A. (1988). *J. Mol. Biol.* **203**, 1045–1070.

LINDQVIST, Y., SCHNEIDER, G., ERMLER, U. & SUNDSTRÖM, M. (1992). *EMBO J.* **11**, 2373–2379.

O'HALLORAN, T. V., LIPPARD, S. J., RICHMOND, T. J. & KLUG, A. (1987). *J. Mol. Biol.* **194**, 705–712.

REINEMER, P., DIRR, H. W., LADENSTEIN, R., SCHÄFER, J., GALLAY, O. & HUBER, R. (1991). *EMBO J.* **10**, 1997–2005.

RICHMOND, T. J., FINCH, J. T., RUSHTON, B., RHODES, D. & KLUG, A. (1984). *Nature (London)*, **311**, 532–538.

SCHNEIDER, G., BRÄNDÉN, C.-I. & LORIMER, G. (1986). *J. Mol. Biol.* **187**, 141–143.

SCHNEIDER, G., LINDQVIST, Y., BRÄNDÉN, C.-I. & LORIMER, G. (1986). *EMBO J.* **5**, 3409–3415.

SCHNEIDER, G., LINDQVIST, Y. & LUNDQVIST, T. (1990). *J. Mol. Biol.* **211**, 989–1008.

SCHEIDER, G., LINDQVIST, Y. & SUNDSTRÖM, M. (1989). *J. Biol. Chem.* **264**, 21619–21620.

STANFORD, R. H. (1962). *Acta Cryst.* **15**, 805–806.

STANFORD, R. H., MARSH, R. E. & COREY, R. B. (1962). *Nature (London)*, **196**, 1176–1178.

STEIGEMANN, W. (1974). PhD thesis, Technische Univ. München, Germany.

SUNDSTRÖM, M., LINDQVIST, Y., SCHNEIDER, G., HELLMAN, U. & RONNE, H. (1994). *J. Biol. Chem.* In the press.

VAUGHAN, P. A., STURDIVANT, J. H. & PAULING, L. (1950). *J. Am. Chem. Soc.* **72**, 5477–5486.

YONATH, A., SAPER, M. A., FROLOW, F., MAKOWSKI, I. & WITTMAN, H. G. (1986). *J. Mol. Biol.* **192**, 161–162.