Crystal Structures of ModA from *Escherichia coli* and Formaldehyde Ferredoxin Oxidoreductase from *Pyrococcus furiosus*

Thesis by

Yonglin Hu

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1999

(Submitted December 21, 1998)

Acknowledgements

This project is only possible with the help from many people. I would like to thank Prof. Douglas C. Rees, my advisor, for his help, encouragement, and guidance throughout this project. I also would like to thank Prof. Jesse L. Beauchamp, Prof. Barbara Imperiali, and Prof. Harry B. Gray, the members of my thesis committee.

Many former and current Rees group members have given me a lot of help. I would like to thank Dr. Jamie L. Schlessman for helping me with anaerobic techniques, Drs. Michael H. B. Stowell and John W. Peters for their assistance with data collection, model building, and structural refinement. Salem Faham kindly provided me with the initial crystallization conditions of *Pyrococcus furiosus* formaldehyde ferredoxin oxidoreductase. I also would like to thank Drs. Hermann Schindelin and Caroline Kisker for their assistance with crystallographic programs. I thank Andrew P. Yeh and William A. Wehbi for proofreading the manuscripts.

At last but not at least, I offer my special thanks to my wife, Hong Liao, for her love and encouragement, and our son, Allen, who has brought us so much joy. I also would like to thank our parents for their invaluable help and support.

Abstract

The crystal structures of two proteins, ModA from *Escherichia coli* and formaldehyde ferredoxin oxidoreductase from *Pyrococcus furiosus*, are reported in this paper.

ModA is a periplasmic molybdate binding protein. Crystals of this protein complexed with tungstate or molybdate belong to space groups $P3_221$, with cell dimensions of a=b=82.6Å[1], c=81.5Å, α=β=90°, and γ=120°. The structure of ModA was solved by the Single Isomorphous Replacement and Anomalous Scattering method and refined to 1.75Å resolution for both molybdate- and tungstate-bound crystal forms. The R and free R factors are 16.2% and 20.3%, respectively, for the molybdate-bound model, and 16.3% and 18.6%, respectively, for the tungstate-bound model. Based on the structural comparisons with other periplasmic binding proteins, such as sulfate and phosphate binding proteins and ModA from *Azotobactor vinelandii*, the structural bases of the high specificity of ModA for molybdate were identified.

P. furiosus formaldehyde ferredoxin oxidoreductase (FOR) was crystallized in space group $P2_12_12_1$, with cell dimensions a=99.03Å, b=171.10Å, c=179.86Å, and α=β=γ=90°. Its crystal structure was solved by the molecular replacement method, and refined to 1.85Å resolution to an R factor of 17.4%, and free R factor of 22.0%. Complexes of FOR with glutarate, an inhibitor, and *P. furiosus* ferredoxin, its physiological electron acceptor, were solved and refined to 2.4Å and 2.15Å resolution, respectively. A structural comparison revealed that FOR may have an enzymatic mechanism similar to that of *Desulfovibrio gigas* Mop, an unrelated molybdenum-containing enzyme. Residues related to the substrate specificity of FOR were identified based on the FOR-glutarate interactions. From the arrangement of the redox centers in the FOR-ferredoxin complex, an electron transfer pathway between these two partners was proposed.

[1] 1Å is 10^{-10}m.

Contents

List of Figures

List of Tables

Chapter 1 Crystal Structure of ModA from *Escherichia coli*

1.1 Introduction to ModA

Molybdenum is an essential trace element for virtually all kinds of life. Its biological significance was first established in the 1930s when it was discovered that this metal was essential for the growth of the nitrogen-fixing organism Azotobacter (Bortels, 1930). Molybdenum-containing enzymes participate in the metabolisms of nitrogen, sulfur and carbon in microorganisms, plants, and animals (Stiefel, 1993). Molybdenum is also an essential element for human beings and is incorporated into various enzymes, such as sulfite oxidase (Cohen *et al.*, 1971) and xanthine oxidase (Wright *et al.*, 1993). Defects in molybdenum enzymes lead to various diseases.

Molybdenum is also known to play an antagonistic role with copper, another essential element, in ruminant animals. Excess molybdenum in soil can lead to copper deficiency, while excess copper can lead to molybdenum deficiency.

In nature, molybdenum usually exists in the form of soluble anionic molybdate ion. At physiological pH, the dominant form of molybdate is the mononuclear MoO_4^{2-}. To survive under conditions of molybdenum deficiency, many organisms have developed high-affinity active transport systems for the acquisition of molybdenum, in the form of molybdate. After the anion is transported into the cell, it is subsequently reduced and the molybdenum is incorporated into various enzymes. With the exception of nitrogenase MoFe protein, in which the molybdenum atom is part of a Mo–Fe–S cluster (Kim & Rees, 1992), molybdenum is incorporated into enzymes as part of Moco

(molybdenum-molybdopterin cofactor) (Rajagopalan & Johnson, 1992), in which the mononuclear molybdenum is coordinated by the sulfur atoms of molybdopterins.

In *Escherichia coli*, molybdenum is found in several enzymes, including nitrate reductase, formate dehydrogenase, dimethyl sulfoxide reductase, and biotin sulfoxide reductase. Molybdenum is accumulated in cells of *E. coli* by a high-affinity molybdate transport protein system encoded by the *modABCD* operon (Johann & Hinton, 1987, Maupin-Furlow *et al.*, 1995, Rech *et al.*, 1995, Walkenhorst *et al.*, 1995, Grunden & Shanmugam, 1997), formerly known as *chlD* locus (Shanmugam *et al.*, 1992) . The *mod* genes are required for the maximum expression of molybdenum-containing enzymes in *E. coli* (Pascal & Chippaux, 1982). Mutations at the *mod* genes result in the inability of *E. coli* to uptake molybdate from the environment, preventing the synthesis of active molybdoenzymes. The phenotype of the *mod* mutants can be suppressed by the addition of high concentrations of molybdate to the growth medium (Glaser & DeMoss, 1971). Molybdate is then transported into cells either by the sulfate transport system or by a nonspecific anion transport system (Lee *et al.*, 1990). High concentrations of molybdate in the growth medium will suppress the expression of *modABCD* operon (Miller *et al.*, 1987).

The *E. coli* molybdate transport system is a member of the osmotic-sensitive transport system family. More than 50 transport systems from *E. coli* and *Salmonella typhimurium* and other Gram-negative bacteria have been extensively studied in the past three decades (for reviews, see references (Ames, 1986, Furlong, 1987, Boos & Lucht, 1996)). These systems are responsible for uptaking nutrients, such as amino acids, oligopeptides, sugars, oxyanions, and vitamins, from the environment. Usually these transport systems are classified into osmotic-shock-sensitive and osmotic-shock-resistant transport systems, according to their response to osmotic shock (Neu & Heppel, 1965). The osmotic-shock-sensitive transport systems are also referred to as periplasmic transport systems since they are composed of a periplas-

mic binding protein, and usually three additional membrane-associated components. Since the periplasmic binding protein is required for transport through this type of system, loss of the protein by osmotic shock results in the inactivation of this system. Studies on the periplasmic transport systems show that although they have very different substrates, they are structurally, mechanistically, and probably evolutionarily related, and form a distinct class. The periplasmic transport systems belong to a superfamily of ABC (ATP-binding cassette) transporters. Members of this family are powered by the hydrolysis of ATP (Higgins, 1992).

By correlating the temperature sensitivity of a mutant histidine-binding protein to temperature sensitivity of histidine transport in *Salmonella typhimurium*, Ames & Lever (1972) demonstrated that the periplasmic binding proteins are obligatory components of periplasmic transport systems . These periplasmic binding proteins have high affinities for their substrates and serve as the primary receptors for transport. They are the initial determinants of the specificity of the transport systems (Quiocho & Ledvina, 1996). These binding proteins are the best characterized components of the periplasmic transport systems for several reasons: they are water soluble, abundant in the bacterial cells, and can be easily assayed due to their high affinity and specificity for their substrates.

That the periplasmic transport systems have membrane-associated components was first known in 1978 (Ames & Nikaido, 1978), but contrary to the periplasmic binding proteins, they are poorly understood. Based on biochemical and mostly genetic studies on these proteins, models have been proposed for the transport mechanism and the functions of these proteins (Ames, 1986, Higgins *et al.*, 1990). It is generally believed that there are two highly hydrophobic trans-membrane proteins, and another protein with a more hydrophilic amino acid sequence. The first two proteins are homologous to each other, and in some species they fuse into one larger protein. The last one is thought to be membrane-associated, but it has been debated as to

whether it is a trans-membrane protein (Ames, 1986) or not (Higgins *et al.*, 1990). A nucleotide binding site has been identified in this protein. It is generally believed that the transport by the periplasmic transport systems is powered by ATP hydrolysis catalyzed by this protein (Berger, 1973, Berger & Heppel, 1974). The transport process is powered by a proton-gradient for the shock-resistant transport systems.

The periplasmic binding proteins from bacteria usually consist of single peptides chain and function as a monomers. They have molecular weights between 23kD and 52kD, with most around 33kD (Furlong, 1987). All of these proteins have high substrate affinities and specificity, with binding constants in the micromolar range (Furlong, 1987), usually between 0.1μM and 1μM for sugar substrates and between 0.01μM and 0.1μM for amino acids. Most of these proteins are resistant to heat and proteases, and have broad pH optima. From the X-ray crystallographic studies on some of these proteins (Quiocho, 1990, Quiocho, 1991), all these proteins have two domains that are similarly folded, and a deep cleft separates these two domains. Although the substrates for these proteins are very different, they were all found to be bound in the clefts of their respective binding proteins.

The *E. coli modABCD* operon shares strong structural and sequence similarities with genes encoding other periplasmic binding protein-dependent transport systems (Johann & Hinton, 1987, Maupin-Furlow *et al.*, 1995, Walkenhorst *et al.*, 1995). Based on the comparison with the active transport systems for other solutes, such as maltose (Gilson *et al.*, 1982) and histidine (Higgins *et al.*, 1982), it was suggested that the *modABCD* operon of *E. coli* encoded a periplasmic transport system. The ModA protein is predicted to be a periplasmic binding protein (Rech *et al.*, 1995). It binds and transfers molybdate to ModB, a trans-membrane protein (Johann & Hinton, 1987), at the outer surface of the cytoplasmic membrane. ModB, together with ModC, then transports the anion across the cytoplasmic membrane to the cell cytoplasm in an ATP-dependent process. ModC

is proposed to contain an ATP hydrolase activity that provides energy for the transport process (Rech *et al.*, 1995). The function of the *modD* gene product is unknown. Two more genes, *modE* and *modF*, are also involved in molybdate transport. The *modE* gene encodes a molybdate and DNA binding protein of 28kD that affects molybdenum-dependent repression at the *modA* promoter (Grunden *et al.*, 1996, Anderson *et al.*, 1997, McNicholas *et al.*, 1997). The exact role of *modF* is unknown (Grunden *et al.*, 1996).

Based on cloning and DNA sequence studies, genes encoding molybdate transport systems similar to that of *E. coli* have been found in other organisms, such as *Aquifex aeolicus* (Deckert *et al.*, 1998), *Arthrobacter nicotinovorans* (Menendez *et al.*, 1997), *Azotobacter vinelandii* (Luque *et al.*, 1993) , *Bacillus subtilis* (Kunst *et al.*, 1997), *Haemophilus influenzae Rd* (Fleischmann *et al.*, 1995), *Helicobacter pylori* (Tomb *et al.*, 1997), *Methanobacterium thermoautotrophicum* (Smith *et al.*, 1997), *Mycobacterium tuberculosis* (Cole *et al.*, 1998) , *Rhodobacter capsulatus* (Wang *et al.*, 1993), and *Synechocystis sp* (Kaneko *et al.*, 1995).

Based on the presence of a leader-like sequence at its N-terminus, *E. coli* ModA was predicted to be a periplasmic protein (Rech *et al.*, 1995). The *E. coli modA* gene has been overexpressed, and its product purified and studied (Rech *et al.*, 1996). The studies confirmed that ModA is located in the periplasmic space of the cell, and is released following a gentle osmotic shock. Amino acid sequencing of its N-terminus showed that a leader region of 24 amino acids was removed upon export from the cell.

Biochemical studies (Rech *et al.*, 1996, Imperial *et al.*, 1998) on *E. coli* ModA showed that this protein bound molybdate and tungstate with high affinity. ModA does not bind other anions, such as sulfate, chromate, selenate, phosphate, and chlorate, even at concentrations as high as 2mM. Its high affinity for molybdate and tungstate is unchanged with pH in the range 3–9. Several changes were observed on ModA upon the binding of molybdate or tungstate, including its increase of mobility

in native polyacrylamide gel experiment, decrease of the pI from 7.0 to 5.6, changes in the UV absorption and fluorescence emission spectra, and the increased resistance to limited proteolysis by chymotrypsin. These changes indicate that ModA undergoes significant conformational changes when it binds its substrates.

By monitoring the changes in gel mobility and UV-visible spectrum of ModA upon the binding of molybdate and tungstate, the dissociation constants K_d were measured as 3μM and 7μM, for molybdate and tungstate, respectively, by Rech *et al.* (1996). But in a separate study by using an isotopic binding method with $^{99}MoO_4^{2-}$, a K_d of 20nM was observed for molybdate both *in vitro* and *in vivo*. This number was approximately the same for tungstate and is very close to the K_d's of other anion binding proteins (Imperial *et al.*, 1998).

Given the similarities in terms of size, geometry, and charge between molybdate and other tetrahedral oxyanions such as sulfate and phosphate (Williams, 1994), the mechanisms by which ModA can specifically bind molybdate and discriminate against closely related oxyanions are of general interest. In this chapter, we will describe the crystal structure of *E. coli* ModA determined and refined at 1.75Å resolution by the SIRAS (Single Isomorphous Replacement and Anomalous Scattering) method, and discuss the structural basis of the high specificity of this protein toward molybdate and tungstate in contrast to other oxyanions, such as sulfate. The results of this study have been previously published (Hu *et al.*, 1997).

ModA has an α/β structure. It consists of two domains that are structurally very similar. Two hinge areas connect the two domains. A large cleft separating these two domains provides the binding site for the substrate. The molybdate and tungstate anions are completely desolvated and buried in this cleft, held by hydrogen bonds with peptide groups from both domains. No charged groups are observed within 6Å of the bound anion.

Structural comparisons with other periplasmic binding proteins (Quiocho, 1990,

Quiocho, 1991) show striking similarities in their structures, although they don't have significant amino acid homology. In this chapter, a detailed comparison with the structure of sulfate binding protein (SBP) (Pflugrath & Quiocho, 1988) from *Salmonella typhimurium* will be presented. Since sulfate, the substrate of SBP, is very similar to molybdate in terms of size, shape, and charge, a comparison of the structures of ModA and SBP should lead to some insight into the specificity of ModA for molybdate over other anions such as sulfate. It will be shown that the anion binding pockets of SBP and ModA in general are similar to each other, reflecting the similarities between their substrate. For example, both proteins use hydrogen bonds exclusively for the binding of the anions. The high specificity of ModA is achieved through some variations of its binding pocket, such as the increase of the volume and decrease of the polarity.

The crystal structure of a homologous periplasmic molybdate binding protein from *Azotobacter vinelandii* has been reported recently (Lawson *et al.*, 1997, Lawson *et al.*, 1998). Although the structure of the anion binding pocket of this protein is not completely conserved compared to that of *E. coli* ModA, it agrees very well with the structural features of *E. coli* ModA identified to be possibly important for the specificity of ModA molecules. Amino acid sequence comparisons with ModA from other organisms suggests that this might also be true for other ModA's.

1.2 Crystallization and Data Collection

1.2.1 Crystallization

ModA from *E. coli* was expressed and purified as molybdate- and tungstate-free preparations as described (Rech *et al.*, 1996). ModA crystals were obtained with the hanging drop method at room temperature. Five microliters of protein solution (20mg/ml protein in 50mM of Tris, pH 7.6) were mixed with 5μl of precipitating solution (2mM MoO_4^{2-} or WO_4^{2-}, 27.5% PEG 8000(w/v), and 0.1M sodium acetate, pH 4.7) and equilibrated against 1ml of precipitating solution. Irregularly shaped crystals appeared in about one week and grew to an approximate dimension of $0.5\times0.5\times0.5mm^3$ in about two to three weeks. High protein concentration (20mg/ml or higher) was essential to the success of crystallization. While crystals of high quality was readily obtained with 20mg/ml protein solution, protein solution with lower concentrations did not yield crystals. Since the protein solutions were prepared at a concentration of 10mg/ml, they were concentrated to 20mg ModA per milliliter using an Amicon concentrator (molecular-weight cutoff: 10 kDa).

Attempts at crystallizing ModA in the absence of tungstate and molybdate, or with other anions such as sulfate and selenate, were unsuccessful.

The tungstate- and molybdate-bound ModA crystals belong to the space groups $P3_121$ or $P3_221$. They have virtually identical cell dimensions, a=b=82.6Å, c=81.5Å, $\alpha=\beta=90°$, and $\gamma=120°$. Assuming one ModA molecule per asymmetric unit, the Matthews coefficient (Matthews, 1968) is calculated to be $3.2Å^3$/Dalton.

1.2.2 Data Collection

Data collection was carried out at room temperature on an RAXIS IIc imaging plate system mounted on a Rigaku RU-200 rotating anode X-ray generator producing CuKα radiation (λ=1.5418Å). The radiation was monochromatized with a graphite

monochromator. The ModA crystals were mounted in quartz capillaries during data collection. A distance of 80.0mm between the crystals and the detector and an oscillation angle of 1°/frame were used in data collection.

Data sets collected from crystals with bound molybdate and tungstate were designated data set I and II, respectively. Both data sets were collected from single crystals. The data were processed using the programs DENZO and SCALEPACK (Otwinowski & Minor, 1997). The intensities were then scaled with the program ROTAVATA and merged with the program AGROVATA, and reduced to amplitudes with the program TRUNCATE (CCP4, 1994).

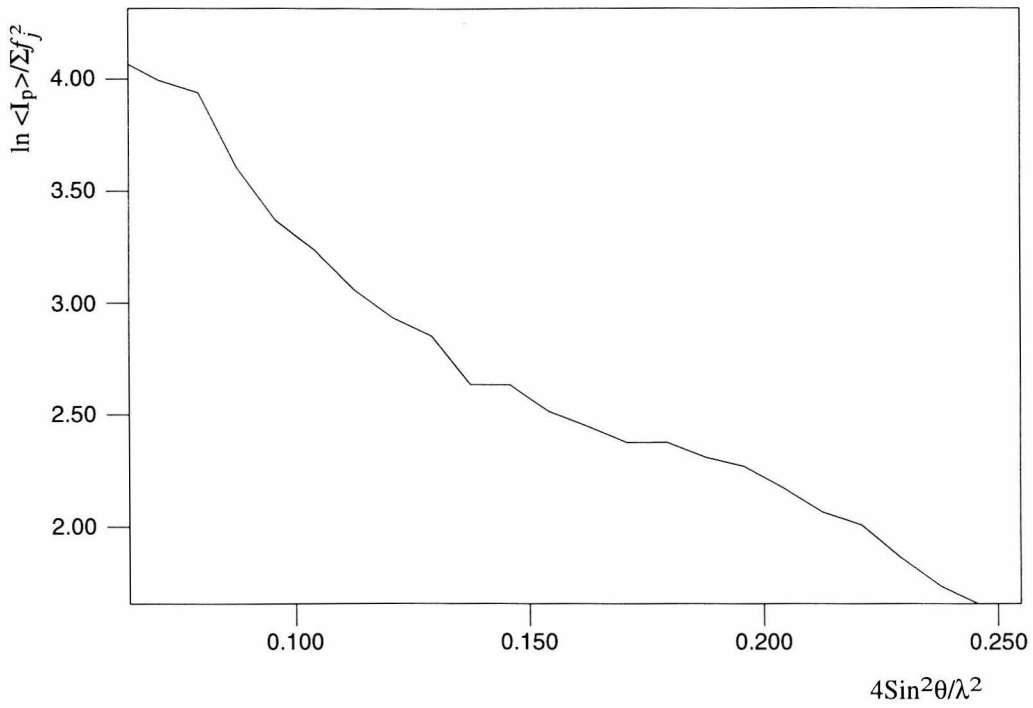


Figure 1.1: Wilson plot for data set I.

Both data sets were collected to 1.75Å resolution. The statistics of data collection are listed in Table 1.1. The mosaicities were refined to 0.28° and 0.23° for data sets I and II, respectively. The overall B-factors of both data sets were estimated to be 25Å^2 from the Wilson plot (Wilson, 1942) (Figures 1.1 and 1.2).

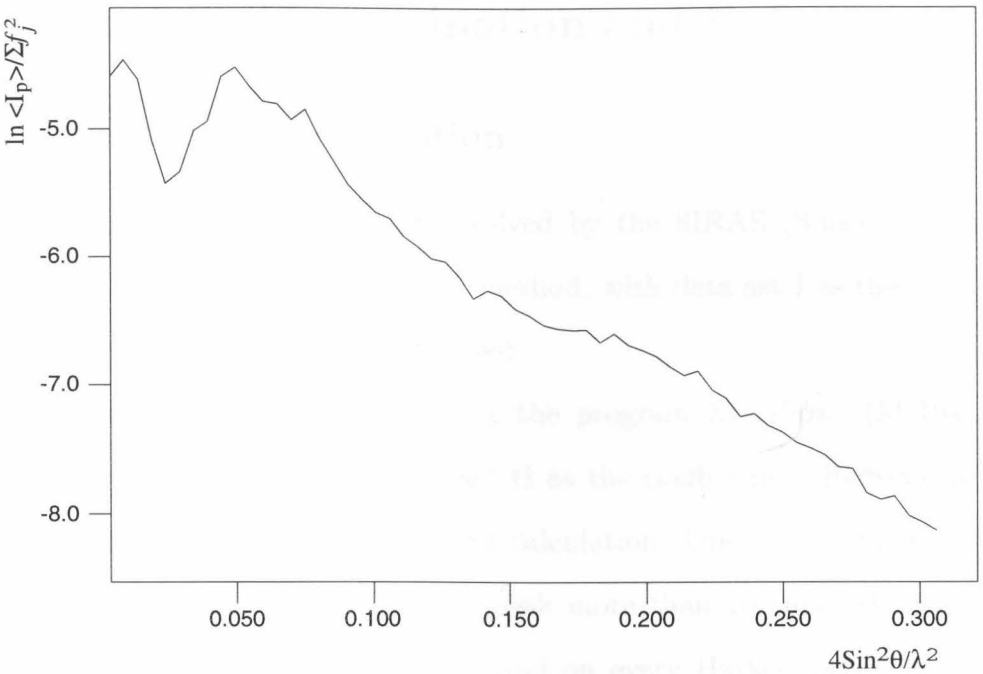


Figure 1.2: Wilson plot for data set II.

Data Sets	**I (MoO_4^{2-})**	**II (WO_4^{2-})**
Space Group	$P3_221$	$P3_221$
Unit Cell Dimensions	a=b=82.57Å, c=81.51Å	a=b=82.59Å, c=81.52Å
Resolution Range	30.0–1.75Å	40.0–1.75Å
Last Shell	1.81–1.75Å	1.81–1.75Å
$\langle I \rangle / \sigma(I)$ (last shell)	21.4(1.9)	21.0(2.0)
R_{sym} (last shell)	5.5% (24.4%)	6.5% (35.3%)
Completeness (last shell)	93.2% (60.8%)	93.8% (61.2%)
Observed Reflections	112116	124197
Unique Reflections	30575	30630

Table 1.1: Data collection statistics.

ModA crystals from both molybdate and tungstate solutions diffracted only slightly better (to about 1.60Å resolution) with synchrotron radiation X-ray sources than with rotating anode X-ray sources. Since they did not diffract at liquid nitrogen temperatures and suffered serious radiation damage at room temperature, attempts were not made to collect diffraction data with synchrotron X-ray sources.

1.3 Structure Determination and Refinement

1.3.1 Structure Determination

The structure of *E. coli* ModA was solved by the SIRAS (Single Isomorphous Replacement and Anomalous Scattering) method, with data set I as the native data set and data set II as the derivative data set.

A Patterson map was calculated using the program XtalView (McRee, 1992) with the anomalous differences of data set II as the coefficients. Reflections in the resolution range 8.0–2.5Å were used in the calculation. One of the Harker sections, z=1/3, is shown in Figure 1.3. A single peak more than 20 times the root-mean-square (rms) deviation of the map was found on every Harker section, confirming that there is only one molecule per asymmetric unit. Because the tungsten has an f'' of 6.0 electrons at 1.5418Å and is the only atom in the crystal that may have significant anomalous scattering at this wavelength, these peaks were assumed to be from the anomalous scattering of the bound tungstate anion. From these peaks, the position of the tungsten atom was determined to be x=0.83, y=0.50, z=0.028, for space group $P3_121$. When a native Patterson map was calculated with the magnitudes of the structure factors as coefficients, peaks were found for data set II at the same positions as those found in the anomalous Patterson map, although weaker (data not shown).

The molybdenum atom has 42 electrons, whereas tungsten has 74. A Patterson map calculated with the differences between the magnitudes of the structure factors of data sets I and II as coefficients has strong peaks at the same position as those in the anomalous Patterson map of data set II (Figure 1.4). This shows that the difference of 32 electrons between molybdenum and tungsten atoms makes a significant difference between data sets I and II and this difference can be used in phasing.

The initial experimental phases were calculated with the program MLPHARE

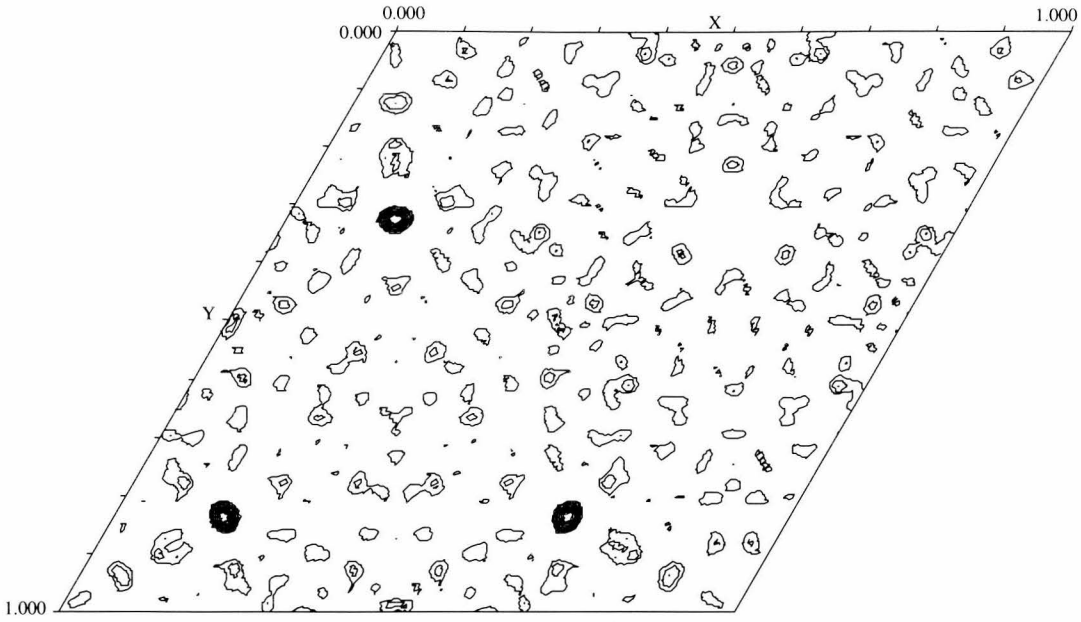


Figure 1.3: Harker section z=1/3 of the native anomalous Patterson map calculated with data set II.

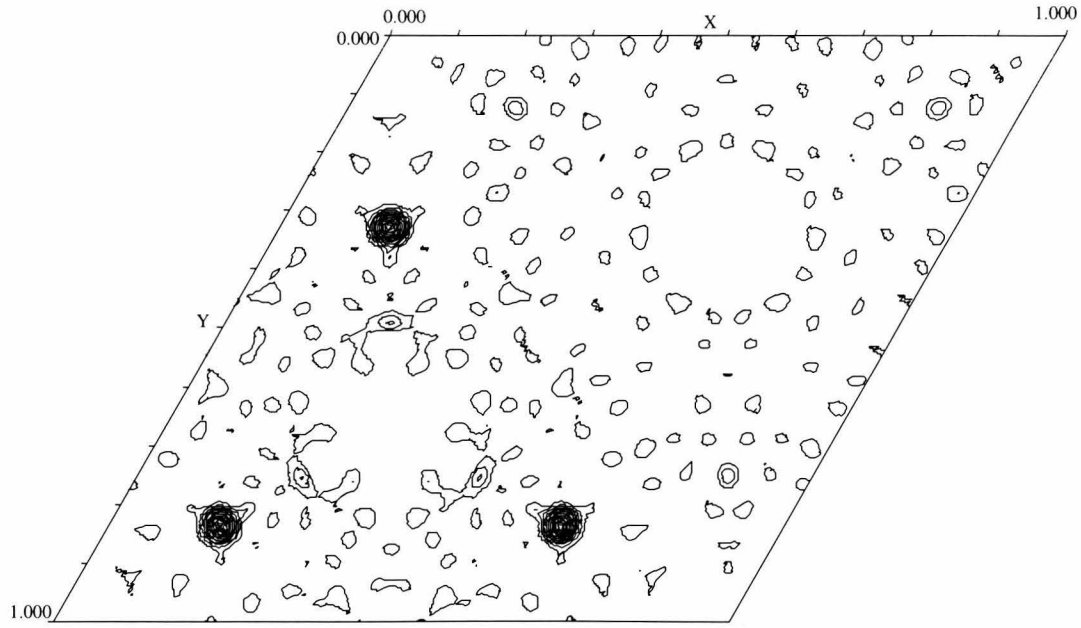


Figure 1.4: Harker section z=1/3 of the difference Patterson map calculated with data sets I and II.

(CCP4, 1994), with phase information from two sources: the anomalous scattering of data set II and the differences between the magnitudes of the structure factors of data sets I and II. The position of the tungsten atom was first refined for 15 cycles

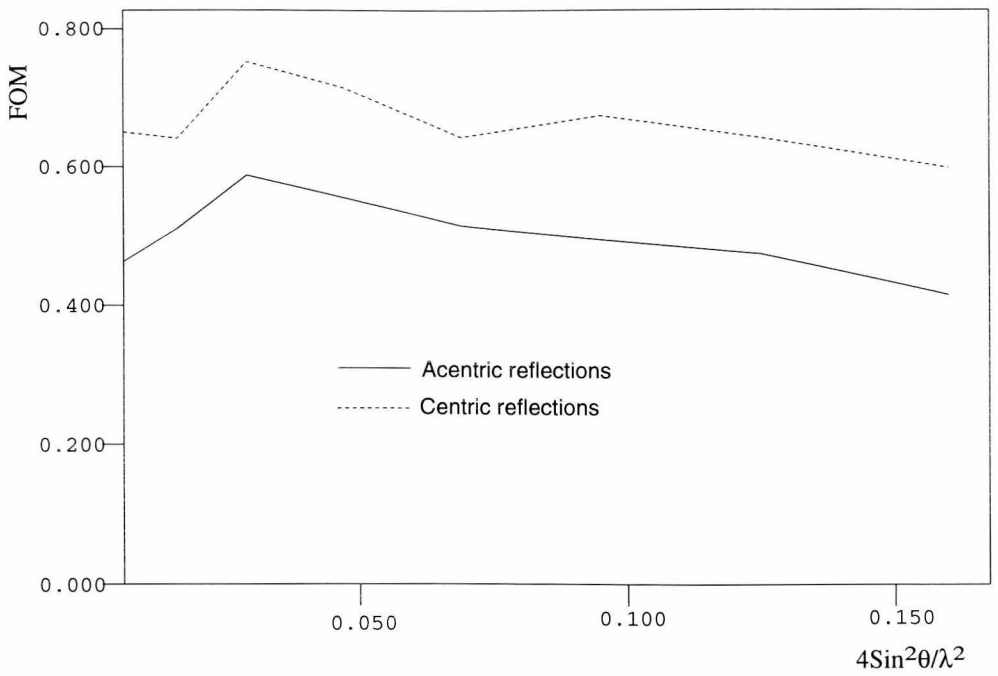


Figure 1.5: Figure of merit as a function of resolution.

with the reflections in the 30.0–2.5Å resolution range; phases were then calculated in the same resolution range. For all the reflections in this resolution range, the figure of merit[1] is 0.50, and the phasing power[2] is 2.3 (Figures 1.5 and 1.6).

The initial phases were then refined with 10 cycles of solvent flattening using the program SOLOMON (Abrahams & Leslie, 1996). The reflections in the 30.0–2.5Å resolution range of data set II were used in the refinement. The figure of merit for these reflections increased to 0.96 after the refinement.

All the calculations described above were performed with both possible space groups, $P3_121$ and $P3_221$. The results, such as phasing power and figure of merit, were identical for these two space groups, but the electron density map calculated with the refined phases and space group $P3_121$ was unrecognizable, while the one calculated

[1]Figure of merit is defined as the weighted mean of the cosine of the deviation of the phase angle from α_{best}.

[2]Phasing power is the ratio of the rms value of the calculated heavy atom scattering factor amplitude to the rms lack of closure error: $[\sum_n |F_H|^2 / \sum_n (|F_{PH}|_{obs} - |F_{PH}|_{calc})^2]^{1/2}$.

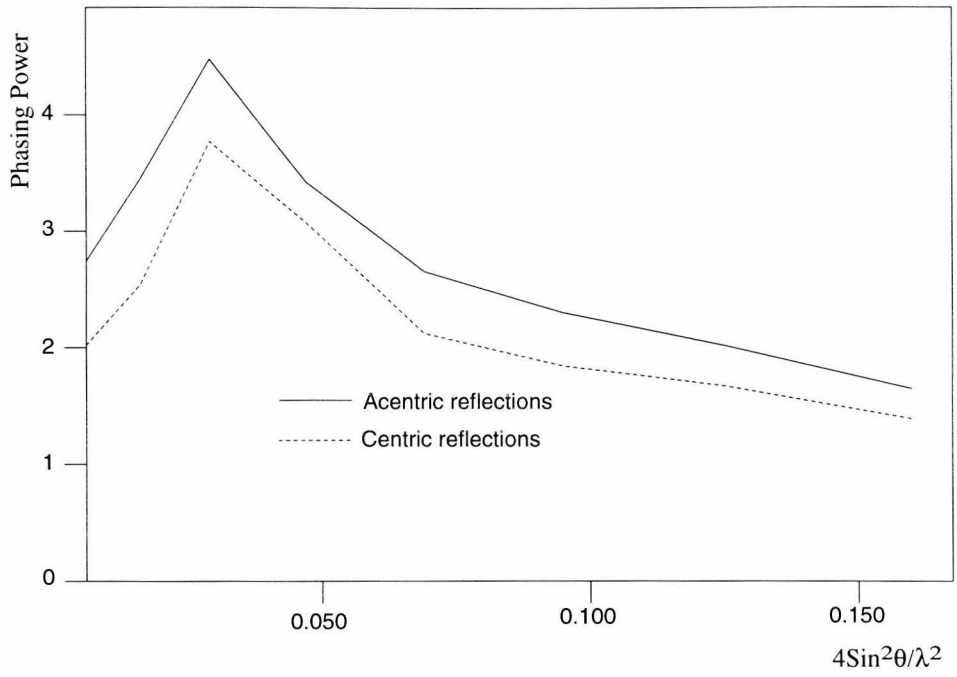


Figure 1.6: Phasing power as a function of resolution.

with the space group $P3_221$ was of excellent quality, indicating that $P3_221$ is the correct space group for *E. coli* ModA crystals. A segment of the experimental density map is shown in Figure 1.7. Figure 1.7, and all the other molecular model figures, are produced with the program MOLSCRIPT (Kraulis, 1991). All the electron density contours are prepared with the program TOM/FRODO (Jones, 1985).

The electron density of the peptide main chain in the experimental map was continuous from the fourth residue to the last one at 1σ level, and most side chains were clearly identifiable. The density at the tungsten binding site was above 20σ. This map was used in model building without further refinement.

Two factors may contribute to the high quality of the map. First, the strong anomalous scattering signals from the data set II. Phase information from anomalous scattering is inherently more accurate than phase information from isomorphous differences (Blundell & Johnson, 1976, pages 371–373). Second, the crystals used in the data collection were almost perfectly isomorphous to each other. The cell dimensions

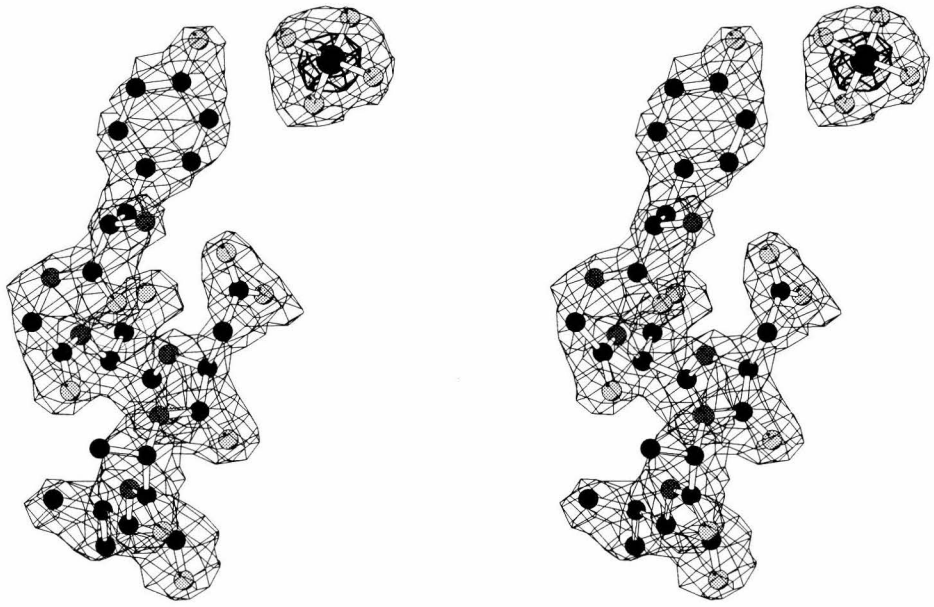


Figure 1.7: Stereoview of a segment of the experimentally phased SIRAS electron density map after solvent flattening, with the refined model superimposed. The residues shown here are Tyr 170, Gly 171, Ser 172, Asp 173, Ala 174, and Val 175, plus the bound tungstate. The map was calculated at 2.5Å resolution and contoured at 1σ and 15σ (bold lines).

are virtually identical for the tungstate and molybdate bound ModA crystal, consistent with the structure of the ModA molecule remaining unchanged when bound to either molybdate or tungstate.

1.3.2 Model Building

The amino acid sequence of *E. coli* ModA was built into the electron density map calculated with data set II. The program O (Jones *et al.*, 1991) version 5 running on an SGI workstation was used for model building. The command *baton* in O was used to choose the positions of the Cα atoms from the skeletons calculated from the electron density map with the program xdlMAPMAN (Kleywegt & Jones, 1996), and the *lego* commands were then used to build in the main chain and side chain atoms. The first three residues were missing from the experimental map. Residues 4–233 were built into the map, as well as a mononuclear, tetrahedral tungstate ion

(WO_4^{2-}), based on the density at the tungstate site.

1.3.3 Structure Refinement

Structure refinement was performed using the program X-PLOR (Brünger *et al.*, 1987) version 3.1, with the parameters of Engh & Huber (1991). The initial model contained residues 4–233 and the tungstate anion (WO_4^{2-}). The structure factors used in X-PLOR refinement were converted from CCP4 mtz files using the program MTZ2VARIOUS (CCP4, 1994). The refinement was first carried out against data set II. Five percent of the reflections were set aside for free R factor calculation and were not used in the refinement (Brünger, 1992).

For the initial model, the R factor was 43.5%, and the free R factor was 44.6%, for the reflections in the 8.0–2.8Å resolution range. After 100 cycles of positional refinement, the R factor dropped to 29.7%, free R 37.6%. A simulated annealing to 3000k further dropped the R factor to 25.6%, free R 34.4%.

The model was then subjected to alternate cycles of manual rebuilding with the program O, addition of solvent molecules, positional and B factor refinement, and resolution extension. During the manual rebuilding, the model was changed if needed according to difference Fourier electron density maps (Fo–Fc and 2Fo–Fc maps). Water molecules were added if their Fo–Fc density was above 3σ and 2Fo–Fc density above 1σ, and if they could form hydrogen bond interactions with protein or other water molecules. The current model contains residues 3–233, one tungstate anion, and 69 water molecules. All the reflections in data set II were used in the refinement, without applying any σ- or resolution-cutoff. A bulk solvent correction was performed (Jiang & Brünger, 1994). The R factor is 16.3%, free R factor 18.6%.

A molybdate-bound ModA model was then built based on the refined tungstate-bound model. The water molecules were removed from the tungstate-bound ModA model, and the tungstate was replaced with a molybdate anion. The model was

then subjected to a rigid body refinement and a simulated annealing to 3000k, followed by alternate cycles of manual rebuilding with the program O, addition of solvent molecules, and positional and B factor refinement. The current model contains residues 3–233, a molybdate anion, and 72 water molecules. All the reflections in data set I were used in the refinement, without applying any σ- or resolution-cutoff. A bulk solvent correction was performed (Jiang & Brünger, 1994). The R factor is 16.2%, free R factor 20.3%. Five percent of the reflections in data set I were set aside for free R factor calculation and were not used in the refinement.

ModA structures complexed with molybdate and tungstate have been deposited with Protein Data Bank, access code 1AMF and 1WOD, respectively.

The following geometry and energy parameters for the tungstate and molybdate anions were used in the refinement. The O–X–O (X=W, Mo) bond angles are 109.48° with an energy of 100kcal/mol·rad^2. The X–O bond lengths are 1.78Å (Koster *et al.*, 1969) and 1.77Å (Gatehouse & Leverett, 1969), for X=W and X=Mo, respectively, with an energy of 1000kcal/mol·Å^2.

1.4 Structure of ModA

1.4.1 Quality of the Models

Both molybdate- and tungstate-bound models were refined to good geometry. For the molybdate-bound model, the rms deviation of bond length from ideality is 0.010Å, 1.56° for bond angle, and 24.0° for dihedral angle. These values are 0.010Å, 1.57°, and 24.1°, respectively, for the tungstate-bound model.

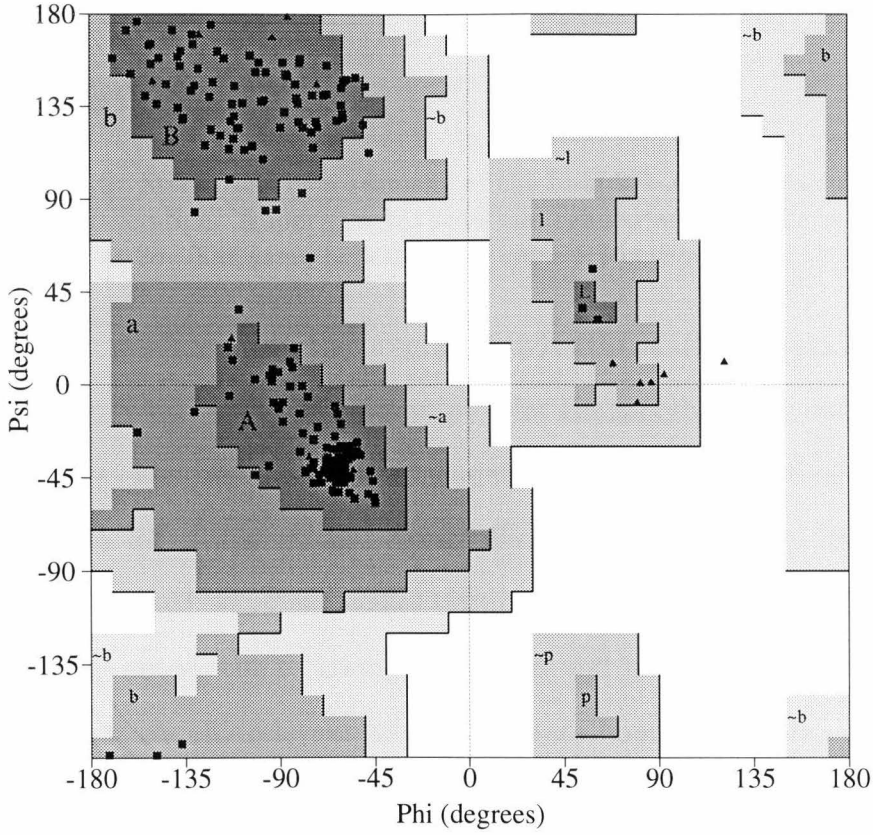


Figure 1.8: The Ramachandran plot for the molybdate-bound ModA model. Triangles represent glycine residues. The shadings, from darkest grey to white, designate the most favored regions, additonal allowed regions, generously allowed regions, and disallowed regions, respectively. The letters A, B, and L indicate the most favored regions for α-helix, β-strand, and left-handed helix, respectively.

For both models, 93% of the residues are in the most favored regions of the Ramachandran plot (Ramachandran & Sasisekharan, 1968) and none are in the disal-

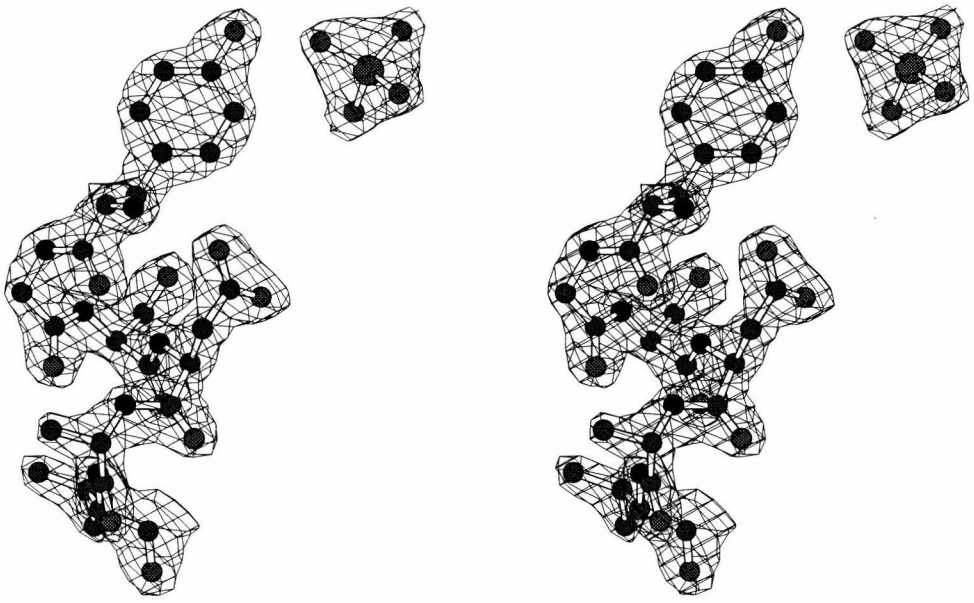


Figure 1.9: Stereoview of a segment of the final 2Fo-Fc map calculated at 1.75Å resolution. Superimposed with the refined model. The map is contoured at 2σ. The same region as in Figure 1.7 is shown.

lowed regions, as calculated with the program PROCHECK (Laskowski *et al.*, 1993). Shown in Figure 1.8 is the Ramachandran plot for the molybdate-bound model.

The molybdate- and tungstate-bound models are essentially identical with an rms deviation of 0.06Å for all Cα atoms. No significant differences were found between them. Most well defined water molecules occupy identical positions in both models. Unless specifically noted, the following discussion will be based on the molybdate-bound model but is equally relevant to the tungstate-bound molecule.

Most residues in the model have well defined electron density (Figure 1.9), with the exceptions of residues Lys 26, Lys 96, Ser 102, Lys 103, and Lys 178, which have part or all of their side chains disordered. The 2Fo–Fc difference Fourier electron density map is continuous at 1σ from the first residue (Gly 3) to the last one, even though residues 92–116 are flexible and refined to high temperature factor (Figure 1.10). The backbone of residue Asn 111 has two possible conformations, but only the dominant one was refined.

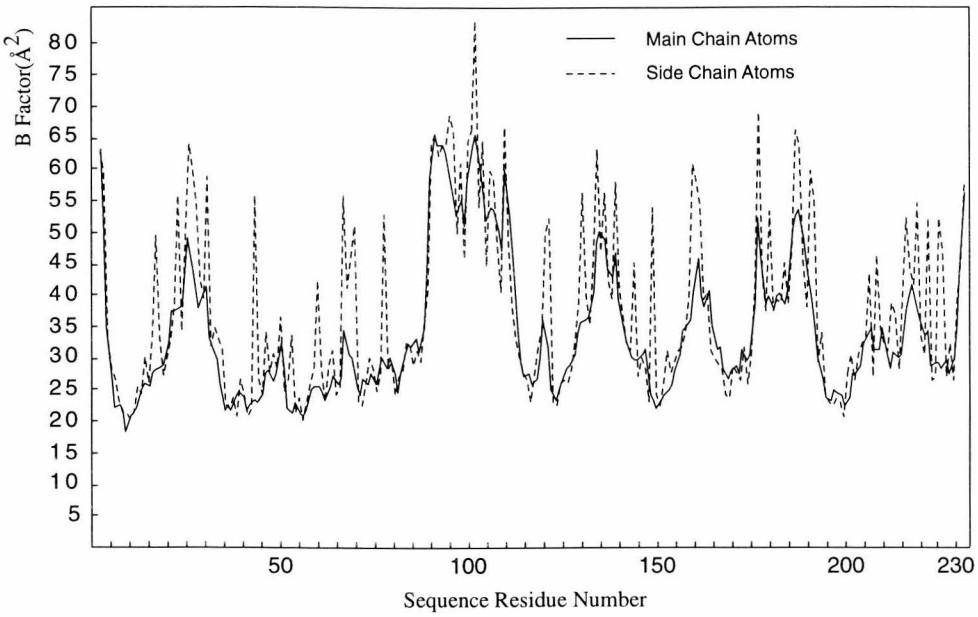


Figure 1.10: Average temperature factors for backbone and side chain atoms of the tungstate-bound model.

1.4.2 ModA Structure

Overall folding

The overall folding of ModA (Figure 1.11) is very similar to other periplasmic binding proteins with known structures (Quiocho, 1990, Quiocho, 1991). ModA molecule has the shape of an ellipsoid, with dimensions of 55Å×33Å×20Å. The overall structure of ModA is of the α/β type, with 41% of the residues in 14 helices, and 25% of them in 10 β-strands, as calculated by the program PROMOTIF (Hutchinson & Thornton, 1996). Helices are designated h1–h14 and are formed by residues 11–13 (3_{10}), 14–28 (α), 39–48 (α), 60–68 (α), 74–76 (3_{10}), 106–110 (α), 124–135 (α), 139–142 (α), 143–145 (3_{10}), 152–160 (α), 171–176 (α), 188–190 (3_{10}), 207–216 (α), and 219–227(α), respectively. Four of them (h1, h5, h9, and h12) are short 3_{10} helices with 3 residues. β-strands are designated b1–b10 and are formed by residues 4–10, 31–37, 54–56, 78–83, 85–90, 115–118, 146–149, 166–170, 180–185, and 194–200, respectively.

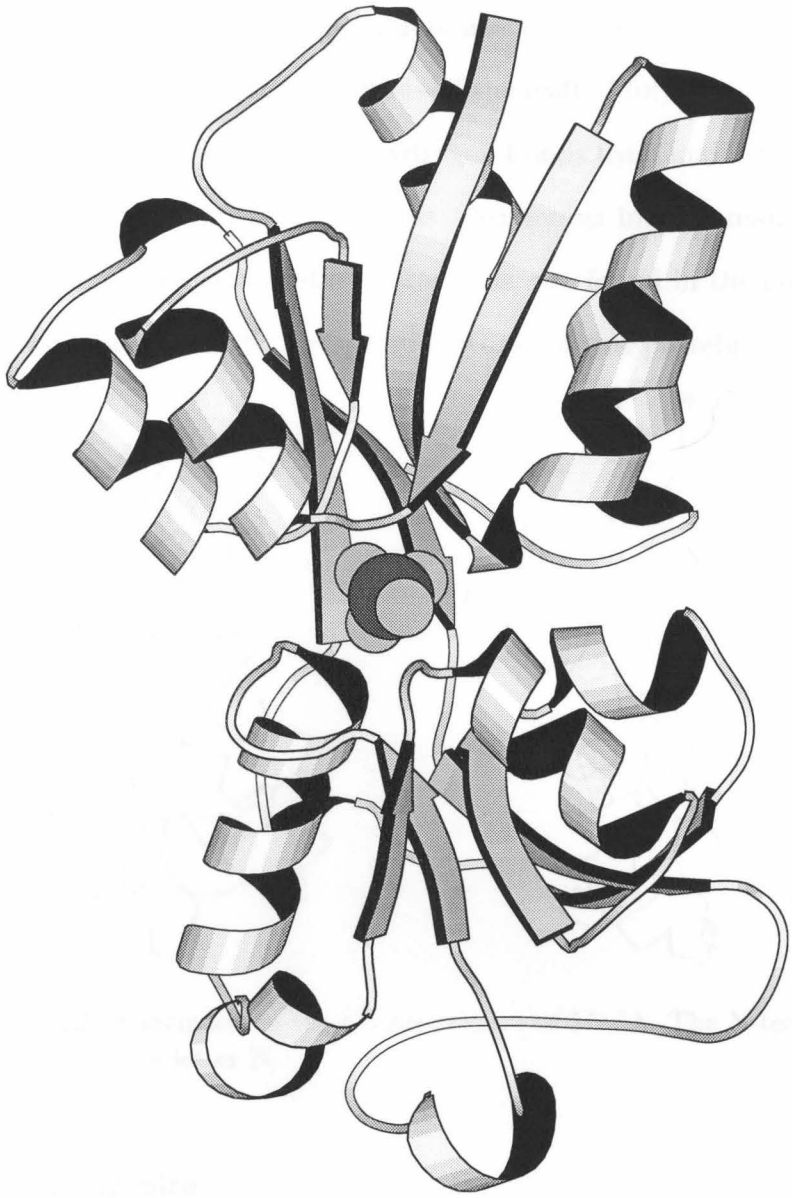


Figure 1.11: The overall folding of ModA. The molybdate anion, which is held between the two domains, is shown as a space filling model.

ModA consists of two domains, designated N-domain and C-domain, respectively (Figures 1.11 and 1.12). Please note that the same scheme is used in designating the domains of ModA as in other periplasmic binding proteins, although in ModA both N- and C-termini are in the N-domain. The two domains are similarly folded, both have a 5-stranded β-sheet at their centers, and 2 or 3 helices flank both sides of the sheets. The two domains are connected by two hinge regions. Between the

two domains is a deep cleft where the molybdate anion resides. The hinge regions connecting N- and C-domains form the basis of the cleft. Molybdate anion is held in the protein structure by accepting seven hydrogen bonds from main chain NH groups or side chain OH groups from both domains. Numerous inter-domain interactions, both hydrogen bonds and electrostatic interactions, are found in the cleft. Following is a detailed discussion about the structural features of this protein.

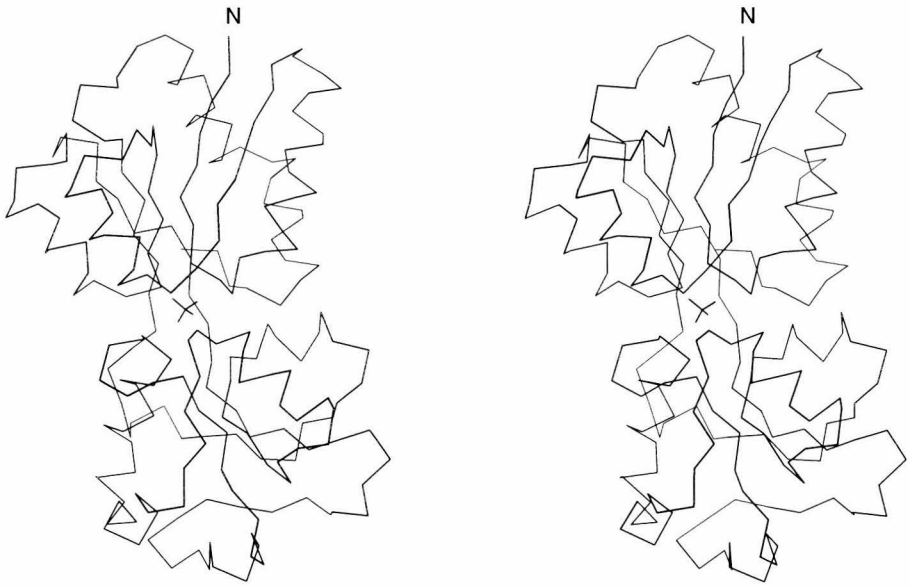


Figure 1.12: Stereoview of the Cα atom trace of ModA. The N-terminus is labeled with the letter N.

Molybdate Binding Site

The molybdate anion is sequestered at the interface between the two domains. It is approximately 6Å beneath the surface of the protein and is inaccessible to the solvent. Since octahedral molybdate and tungstate complexes are known, we studied the possibility of molybdate being octahedral in the ModA structure. To do this, we removed all the geometric restraints on the molybdate in our refinement. The result clearly showed that the anions bound to ModA are of tetrahedral coordination.

After having refined with restraints imposed (Chapter 1.3.3, page 16), the molybdate and tungstate in ModA models have geometries very close to that of a tetrahe-

dron. The X–O bond distances were refined to 1.69–1.72Å and 1.72-1.76Å, for X=Mo and X=W, respectively, and the O–X–O bond angles are 103.7°–113.9° and 104.2°–111.5°, for X=Mo and X=W, respectively. The bond angles and bond distances of the molybdate and tungstate are listed in Table 1.2.

MoO_4^{2-}					WO_4^{2-}				
O–Mo–O Bond Angles					O–W–O Bond Angles				
	O1	O2	O3	O4		O1	O2	O3	O4
O1	–	109.5°	113.9°	108.1°	O1	–	111.6°	111.1°	107.3°
O2		–	103.7°	109.7°	O2		–	104.2°	111.2°
O3			–	111.8°	O3			–	111.5°
O4				–	O4				–
Mo–O Bond Distances					W–O Bond Distances				
	1.69Å	1.72Å	1.70Å	1.72Å		1.73Å	1.76Å	1.72Å	1.75Å

Table 1.2: Bond angles and bond distances of the bound molybdate and tungstate.

Molybdate has a pK_a (K_a is acid constant) of 3.8 (Tytko & Trobisch, 1987). The dominant form of molybdate in the crystallization condition (pH 4.7) is the deprotonated MoO_4^{2-}. The molybdate oxygen atoms act as proton acceptors in all the seven hydrogen bonds between the molybdate and ModA. This confirms that the bound anion is fully deprotonated MoO_4^{2-}.

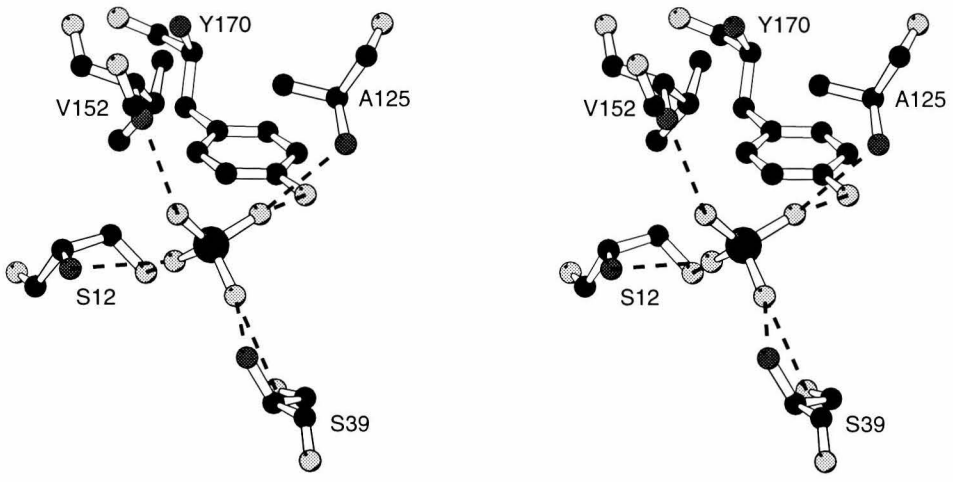


Figure 1.13: Stereoview of the interactions between the bound molybdate and protein groups. The hydrogen bonds are represented by dashed lines.

The molybdate is held in the structure by accepting seven hydrogen bonds, from

three side chain OH groups and four main chain amide groups (Figures 1.13 and 1.14). The lengths of the hydrogen bonds are shown in Figure 1.14. Residues Ser 12, Ser 39, Ala 125, Val 152, and Tyr 170 are involved in binding the anion. The oxygen atoms O2, O3, and O4 of the molybdate each accepts two hydrogen bonds, one from a main chain NH group and a second one from a side chain OH group. Atom O1 forms only one hydrogen bond, with the main chain NH group of Val 152. Residues Ser 12 and Ser 39 use both their main chain NH groups and side chain OH groups to form two hydrogen bonds with one oxygen atom, O4 and O2, respectively, of the molybdate in a chelate-type interaction (Figures 1.13 and 1.14). A similar mode of interaction has been previously observed in the binding of phosphate by phosphate binding protein (PBP) (Luecke & Quiocho, 1990).

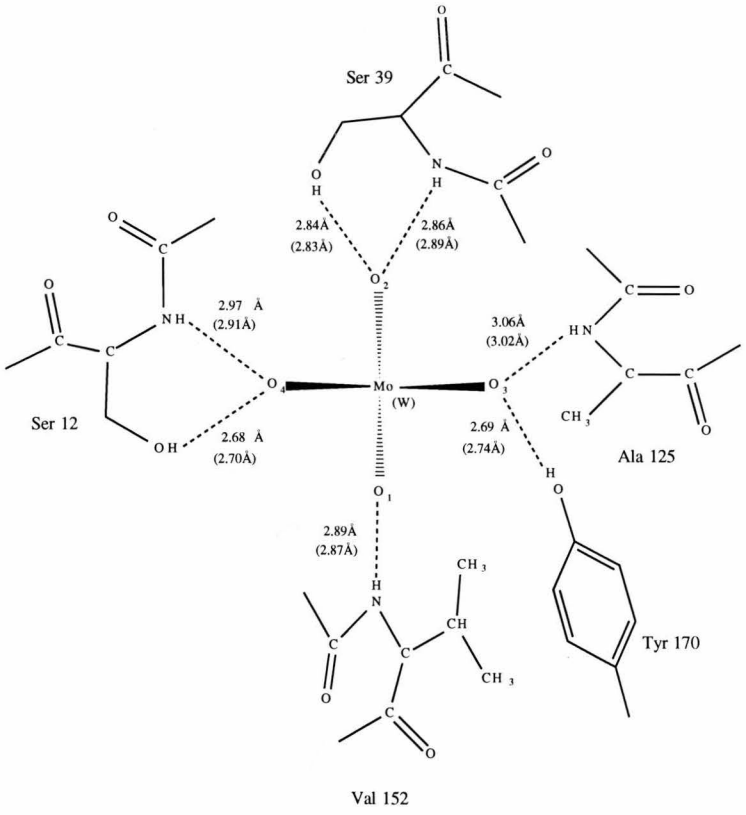


Figure 1.14: Hydrogen bonds between the molybdate/tungstate anion and ModA. The hydrogen bond lengths for the molybdate-bound model are indicated, with those for the tungstate-bound model in parentheses.

Several structural features of the molybdate binding site have been identified.

1. As with both SBP (Pflugrath & Quiocho, 1988) and PBP (Luecke & Quiocho, 1990) , no positively charged groups or water molecules are found in the vicinity of the ModA anion binding site. Furthermore, ModA, like SBP, also lacks carboxylate residues in the anion binding pocket. This observation is significant, since the ability of PBP to bind protonated oxyanions (phosphate) was achieved through the use of a single residue, Asp 56, that could serve as a hydrogen bond acceptor (Luecke & Quiocho, 1990). In ModA, the closest water molecule to the bound molybdate is on the protein surface and is more than 7Å away from the molybdenum, while the closest charged group is the side chain of Asp 151, which is also on the surface of the protein and is directed away from the molybdate. The lack of charged groups in the binding site might be one of the reasons why the binding of molybdate by ModA is insensitive to pH. This structural feature leaves the –2 charge on the anion uncompensated.

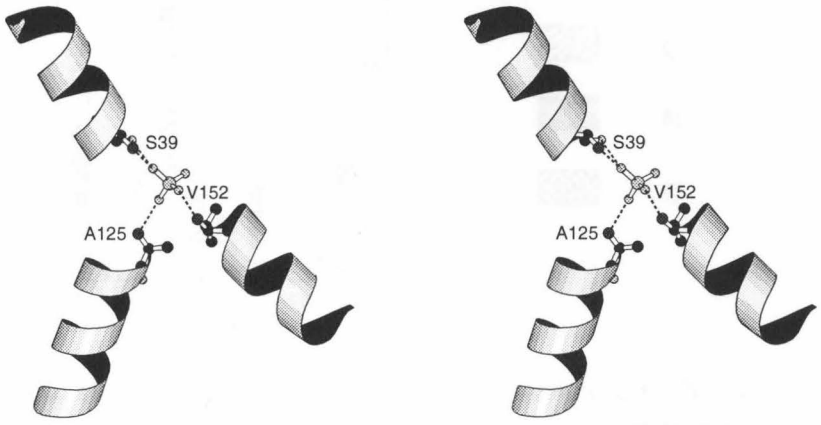


Figure 1.15: The molybdate is located near the N-termini of several α-helices, indicating the macrodipoles of the helices may be important in stabilizing the bound anion.

2. With the exception of Ser 12 and Tyr 170, residues involved in the binding of molybdate are found in the first turns of three α-helices. Ser 39 is the first residue of h3, Val 152 is the first residue of h10, and Ala 125 is the second residue

of h7 (Figure 1.15). Ser 12 is located in the 3_{10} helix h1. But the anion is not at the axis of any of these α-helices. This structural feature indicates that the macrodipoles of the helices may play a role in stabilizing the uncompensated charge on the anion, although it has been suggested that in SBP the local dipoles are more important in the stabilization of the bound sulfate than the macrodipoles of helices (He & Quiocho, 1993).

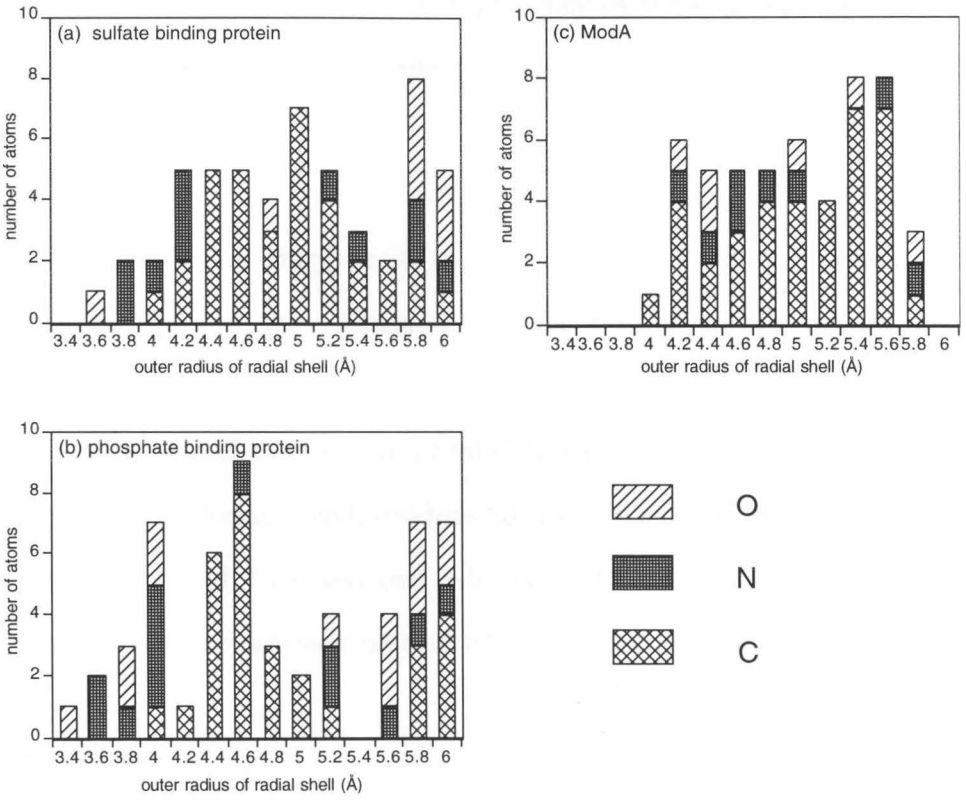


Figure 1.16: Histogram of the radial distribution of C, N and O atoms about the S, P, and Mo central atom of the bound anion in SBP, PBP and ModA respectively. Atoms were accumulated in shells of 0.2Å thickness.

3. Besides the protein groups that donate hydrogen bonds to the molybdate, a large number of hydrophobic groups, including the side chains of residues Ala 10, Ala 11, Ala 58, Val 123, Pro 124, Ala 125, and Val 152, the side chain Cϵ1 atom of Tyr 170, and the side chain Cβ atom of Asp 151, are found near the

molybdate, making van der Waals contacts with the oxygen atoms. All of them are within 3.8Å of at least one of the molybdate oxygen atoms. Figure 1.16 shows the radial distribution of C, N and O atoms about the S, P, and Mo central atom of the bound anion in SBP, PBP and ModA respectively. A much higher proportion of carbon atoms is observed in the anion binding site of ModA than those of SBP and PBP. We studied, but excluded, the possibility of C–H$\cdots$O hydrogen bonds between the molybdate and these groups, based on inappropriate bond angles. The presence of these groups means ModA has a rather hydrophobic binding site.

N-domain

The folding of ModA starts with β-strand b1. If the short 3_{10}-helices and the loop regions are ignored, the folding of residues 3–83 is an alternation of α-helices and β-strands (b1-h2-b2-h3-b3-h4-b4) and it contributes 4 of the 5 β-strands of the N-domain. These four β-strands are parallel to each other. Then the peptide chains goes on to fold the C-domain with residues 83–194. Residues 195–233 complete the N-domain with a β-strand, b10, and two α-helices, h13 and h14. Strand b10 is inserted into the central β-sheet between b3 and b4 and is anti-parallel to other strands.

C-domain

This domain is formed by residues 84–194. It has a topology virtually identical to that of the N-domain if we compare the secondary structures between residues 113 and 185, the first residue of b6 and the last residue of b10, respectively, with those between residues 3–83 of N-domain (Figure 1.17).

The first and the second β-strands and the α-helix between them are arranged right-handedly, and the two β-strands are parallel to each other. The α-helix following the second β-strand is on the opposite side of the β-sheet as the helix between the first

two β-strands, and the third β-strand is placed parallel to the first two and sandwiches the first strand with the second one. The α-helix following the third β-strand is on the same side of the β-sheet as the α-helix that precedes the third β-strand. The fourth β-strand is also parallel to the first three. Between the third and the fourth β-strands is a β-strand that is anti-parallel with other β-strands, and formed by residues far separated in the primary sequence of the protein. For N-domain, this fifth β-strand is b10, formed by residues 194–200, and is flanked by β-strands b3 and b4, which are formed by residues 54–56 and residues 78–83, respectively. For C-domain, the fifth β-strand is b5, formed by residues 85–90, and is flanked by b8 and b9, which are formed by residues 166–170 and residues 180–185, respectively.

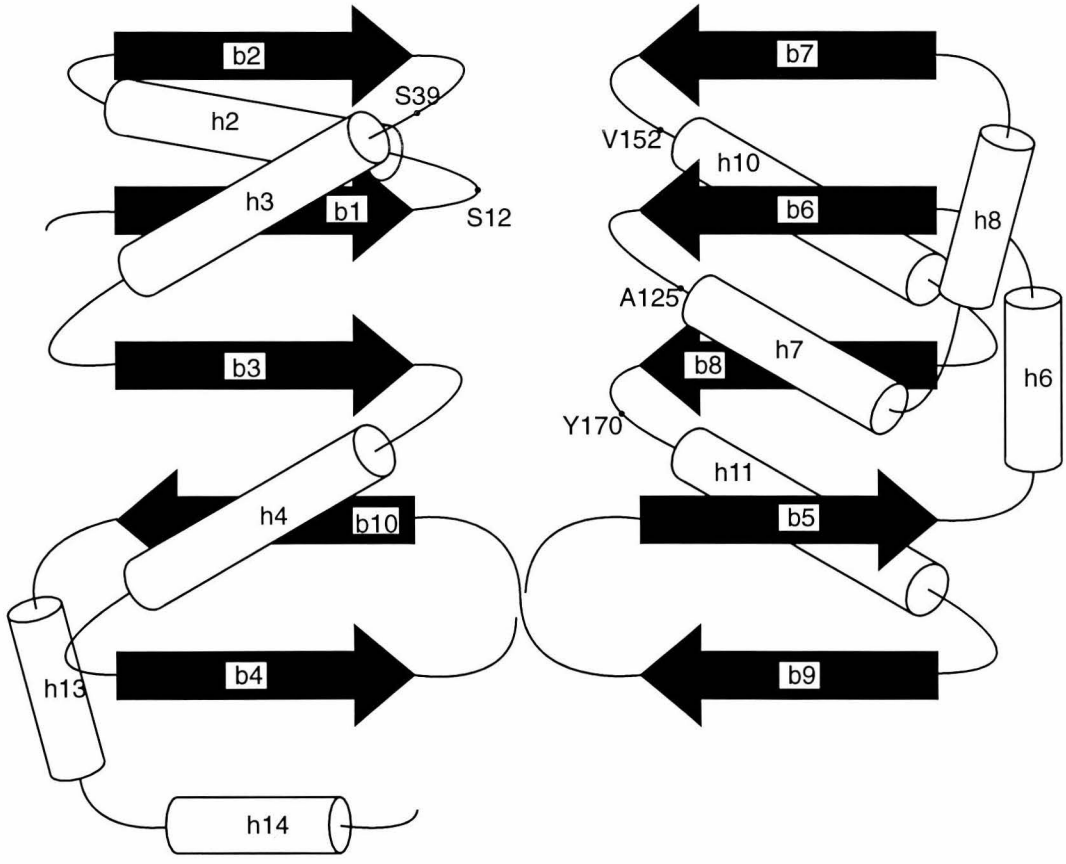


Figure 1.17: A topological drawing of the folding of ModA. 3_{10}-helices are omitted for simplicity. Residues donating hydrogen bonds to the bound anion are labeled. The dimensions of the secondary structures are not to scale.

Inter-domain Interactions

In addition to the interactions mediated by the bound molybdate, some direct inter-domain hydrogen bonds are found in the substrate-binding cleft. The carbonyl oxygen of Ala 58 (N-domain) accepts a hydrogen bond from the side chain OH group of Tyr 128 (C-domain); the side chain Oδ1 atom of Asn 15 (N-domain) accepts two hydrogen bonds from the side chain guanidine group of Arg 153 (C-domain); the side chain Nϵ1 atom of Trp 62 (N-domain) donates a hydrogen bond to the main chain carbonyl oxygen atom of His 122 (C-domain). A group of three water molecules held inside the cleft also mediates inter-domain hydrogen bond interactions which involve the main chain NH groups of Asn 15 and Ala 16 and carbonyl oxygen atoms of Ser 12 and Tyr 228 of N-domain, and the side chains of Ser 172 and Asp 173 from the C-domain.

Hinge Regions

Biochemical studies on ModA indicated that the binding of substrates induce significant conformational changes (Rech *et al.*, 1996). It has been suggested that unligated periplasmic binding proteins exist as an "open" form in which the substrate binding clefts are wide open. This assumption has been validated by crystallographic studies on unligated leucine- and Leu/Ile/Val binding proteins (Sack *et al.*, 1989*a*, Sack *et al.*, 1989*b*). Crystallographic studies on both liganded and unliganded structures of *E. coli* maltodextrin binding protein unequivocally demonstrated that the binding of the substrate induced a major conformational change in this protein. Upon the binding of the substrate, the two domains rotate by an angle of 35° around the hinge region and are twisted against each other by 8° during closing and opening of the binding site (Sharff *et al.*, 1992). These changes are mediated solely by the conformational changes in the hinge areas, whereas the structures of the two domains virtually remain unchanged.

The two hinge regions of ModA are formed by two segments of residues, residues Leu 81 to Leu 85 and residues Lys 192 to Tyr 196, respectively. The first sequence is part of the fouth β-strand of N-domain, while the second one contains residues from a β-turn and the last β-strand of N-domain. ModA is similar to the other binding proteins and many other single polypeptide, two-domain proteins, such as hexokinase and phosphofructokinase, in that the hinge is composed of two or three segments. Very few proteins, such as lysozyme, have a single segment domain linkage, indicating that too much flexibility might be disadvantagous for the functions of these proteins.

These two regions are far apart in the primary amino acid sequence of ModA, but are very close to each other in the tertiary structure. They are on the surface of the protein and form the base of the substrate binding cleft. Several hydrogen bonds are formed between these two regions. The carbonyl oxygen atom of Leu 81 accepts a hydrogen bond from the main chain NH group of Tyr 196; the carbonyl oxygen atom of Val 194 accepts a hydrogen bond from the main chain NH group of Asn 183. The carbonyl oxygen of Asn 83 and the main chain NH group of Val 194 form hydrogen bonds with a common water molecule.

The structures of the hinge regions in ModA are well suited for their functions. In the anion-bound state, they have a stable conformation, so they provide a base for the anion-binding cleft and help hold the two domains together, thus stabilizing the overall structure of ModA. But the hydrogen bond interactions stabilizing the hinge regions can be easily disrupted, so when the protein is in unliganded form, the hinge regions can adopt flexible conformations, enabling the two domains to move relatively freely during the ligand binding processes.

1.5 Structural Basis of the Specificity of ModA

1.5.1 Structural Comparison with SBP

Because molybdate and tungstate (pK_a's 3.8 and 3.7, respectively (Tytko & Trobisch, 1987)) likely bind to ModA as the deprotonated anions, a detailed comparison was made between ModA and the sulfate binding protein from *Salmonella typhimurium* (SBP) (Pflugrath & Quiocho, 1988), which has been established to bind sulfate (pK_a=1.9 (Tytko & Trobisch, 1987)) in the deprotonated form.

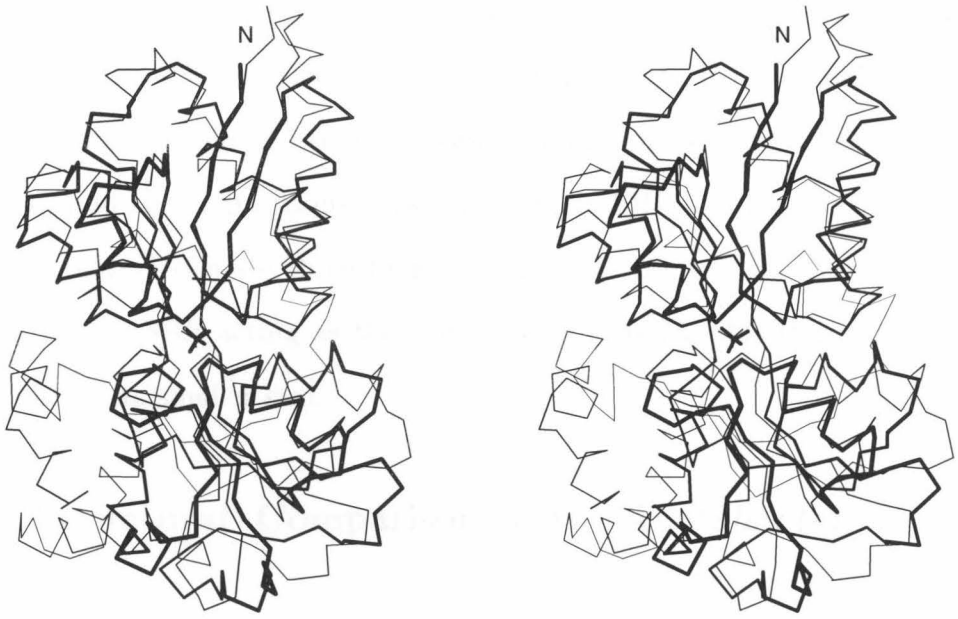


Figure 1.18: Stereoview of the superposition of ModA and SBP. The bound anions are present in the inter-domain cleft. The least-square superposition was based on residues 4–28, 30–72, 116–135, 147–160, 168–177, 194–201, and 207–231 of ModA, and residues 4–28, 36–78, 123–142, 168–181, 190–199, 219–226, and 237–261 of SBP. The ModA is in bold line.

Although no significant homology has been detected between the primary amino acid sequences of ModA and SBP, these two proteins exhibit similar polypeptide folds, including the region around the anion binding sites. Based on 145 equivalent Cα atoms from seven segments of both molecules, ModA and SBP are superimposable with an rms deviation of 1.4Å (Figure 1.18). When the two structures are so

superimposed, most of the main chain atoms at their anion binding sites are less than 1Å from their counterparts. It is also noticed that in general the residues at the binding sites are not conserved between ModA and SBP, nor are the positions of the hydrogen bonds between bound anions and the proteins. The sulfate anion is bound to SBP by accepting seven hydrogen bonds from peptide groups. The hydrogen bond donors include five main chain amide groups (Asp 11, Ser 45, Gly 131, Gly 132, and Ala 173), the side chain Nϵ1 atom of Trp 192, and the side chain OH group of Ser 132.

The most significant structural difference between ModA and SBP occurs at the C-terminus. With 233 residues, ModA is smaller than SBP (310 residues) and most other periplasmic binding proteins. Missing from ModA are the α-helices X, XI, and XII of SBP and the loop regions between them. In SBP and most other periplasmic binding proteins, there are three hinge regions that connect the two domains, with the α-helix X in SBP acting as the third one. In ModA, this third hinge region is missing due to its smaller size.

1.5.2 Structural Comparison with *Azotobacter vinelandii* ModA

Two distinct periplasmic molybdate binding proteins were found in *Azotobacter vinelandii*. The gene of one of them has been sequenced (Luque *et al.*, 1993), but its structure has not been solved. The crystal structure of the second *A. vinelandii* ModA has been solved and refined to 1.2Å resolution (Lawson *et al.*, 1997, Lawson *et al.*, 1998) and sequenced based on the crystal structure, although its exact sequence is still unknown. This protein will be referred to as AvModA2.

The three-dimensional structure of AvModA2 is very similar to that of *E. coli* ModA, as evidenced by the fact that the two proteins are superimposable with an

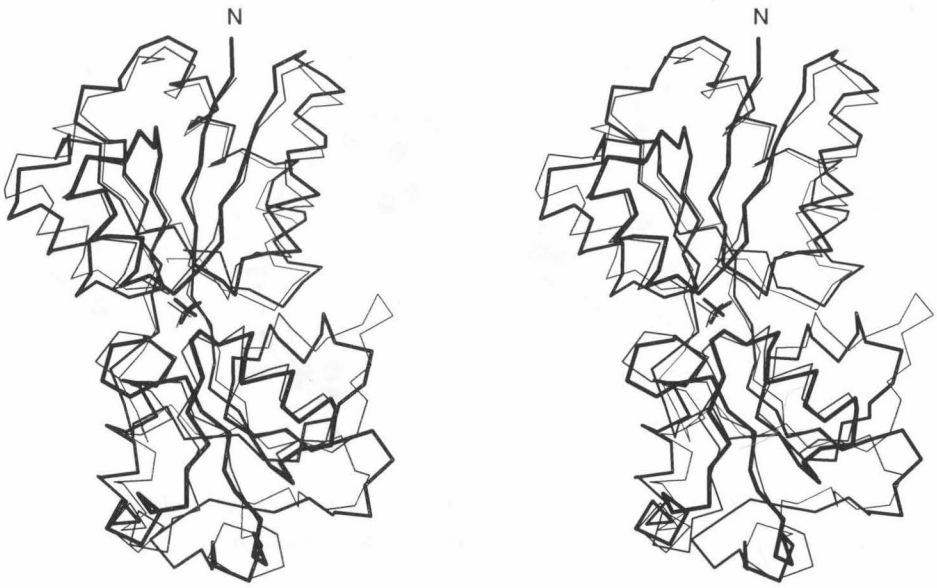


Figure 1.19: Stereoview of the superposition of the Cα atom traces of *E. coli* ModA (bold lines) and *A. vinelandii* ModA2. The least-square superposition was based on residues 4–90, 96–99, 113–117, 119–143, 144–157, 158–175, 183–204, 206–214, and 215–233 of *E. coli* ModA, and residues 2–88, 94–91, 106–110, 112–136, 139–153, 153–180, 180–201, 204–213, and 213–231 of AvModA2.

overall rms deviation of 1.4Å for 203 Cα atoms from both structures. The Cα atoms of the residues near the anion binding sites are within 1Å of each other between these two proteins (Figure 1.19). With 232 amino acid residues, AvModA2 has a size very similar to that of ModA. Consequently, it also lacks the α-helices X, XI, and XII found in the structure of SBP.

Compared to SBP, the anion binding pocket of AvModA2 is much more similar to that of *E. coli* ModA (Figure 1.20). Residues Ser 39 and Val 152 of *E. coli* ModA and their hydrogen bonds to molybdate are conserved in AvModA2 (as Ser 37 and Val 147, respectively). Ser 12 is replaced by Asn 10 in AvModA2, but the chelating-type interaction with molybdate is conserved, as the Asn 10 of AvModA2 donates two hydrogen bonds to the anion with its main chain amide group and side chain Nδ2 atom. Ala 125 of *E. coli* ModA is replaced by Tyr 118 in AvModA2, but the hydrogen bond donated by the main chain amide group is conserved. But the residue

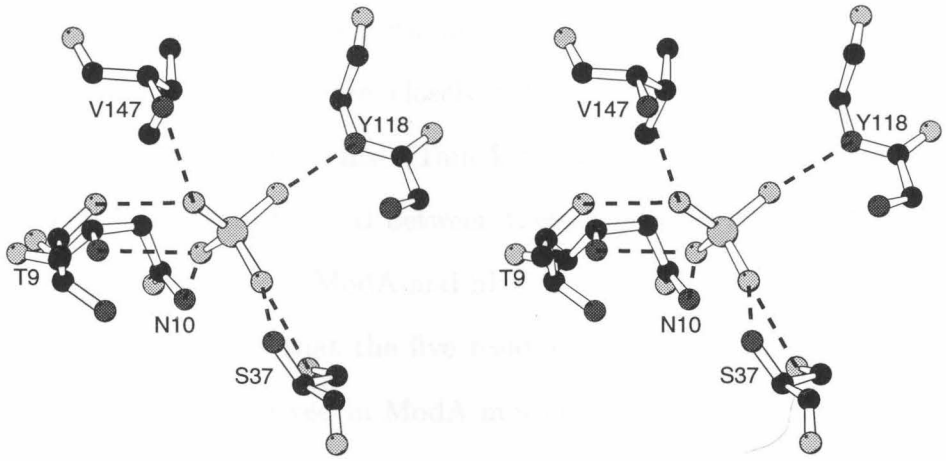


Figure 1.20: Stereoview of the anion binding site of AvModA2. The orientation of the molecule is the same as ModA in Figure 1.13.

corresponding to Tyr 170 in *E. coli* ModA is Ala 165 in AvModA2 and this residue does not form hydrogen bond with the bound molybdate. Instead, Thr 9, which is Ala 10 in *E. coli* ModA, donates a hydrogen bond to molybdate with its main chain amide group.

1.5.3 Sequence Comparisons with ModA from Other Organisms

In addition to *E. coli* and *A. vinelandii* (Luque *et al.*, 1993) ModA's, genes encoding putative periplasmic molybdate-binding proteins homologous to *E. coli* ModA can be found in the following organisms in the GenBank and SWISS-PROT data bases using the BLAST search algorithm (Altschul *et al.*, 1997): *Aquifex aeolicus* (Deckert *et al.*, 1998), *Arthrobacter nicotinovorans* (Menendez *et al.*, 1997), *Bacillus subtilis* (Kunst *et al.*, 1997), *Haemophilus influenzae Rd* (Fleischmann *et al.*, 1995), *Helicobacter pylori* (Tomb *et al.*, 1997), *Methanobacterium thermoautotrophicum* (Smith *et al.*, 1997), *Mycobacterium tuberculosis* (Cole *et al.*, 1998), *Rhodobacter capsulatus* (Wang *et al.*, 1993), and *Synechocystis sp* (Kaneko *et al.*, 1995). These sequences are aligned in Figure 1.21 using the program CLUSTALW

(Thompson *et al.*, 1994). Phylogenetic analysis based on amino acid sequences has suggested that ModA's are more closely related to each other than with other periplasmic anion binding proteins (Tam & Saier, 1993). Indeed, sequence homology to different degrees is observed between these ModAs (Figure 1.21). In contrast, the homology between *E. coli* ModA and SBP is not obvious.

It is interesting to note that the five residues involved in molybdate binding in *E. coli* ModA are not conserved in ModA molecules from other organisms. Of these residues, Ser 39 is the most conserved, since it is only replaced by a similar Thr in *A. vinelandii* ModA1 and the chelating-type hydrogen bonds are likely to be conserved. Ser 12 is replaced by Asn in some of the molecules, but the two hydrogen bonds donated by this residue are expected to be conserved, like in the structure of AvModA2. However the ModA from *M. thermoautotrophicum* has a Gly residue at this position. Val 152 is replaced by a similarly hydrophobic Ile in some of the molecules. Residues Ala 125 and Tyrosine 170 are not conserved nor are the hydrogen bonds donated by these residues, as shown in the structure of AvModA2.

Figure 1.21: Sequence alignment of ModAs from various organisms. Residues that are completely conserved or are replaced in fewer than four proteins by similar residues are marked by boxes. The secondary structures of *E. coli* ModA are shown. Residues in *E. coli* ModA that provide hydrogen bonds to the molybdate are labeled with a dot. Abbreviations: E_co: *Escherichia coli*; A_ae: *Aquifex aeolicus*; A_ni: *Arthrobacter nicotinovorans*; A_vi: *Azotobacter vinelandii*; B_su: *Bacillus subtilis*; H_in: *Haemophilus influenzae Rd*; H_py: *Helicobacter pylori*; M_th: *Methanobacterium thermoautotrophicum*; M_tu: *Mycobacterium tuberculosis*; R_ca: *Rhodobacter capsulatus*; S_sp: *Synechocystis sp.*

1.5.4 Possible Determinants of the Specificity of ModA

Based on the comparisons with SBP and AvModA2 structures, several possible determinants of ModA specificity have been identified.

Size of the binding pocket

X-ray crystallography studies on salt crystals have shown that molybdate and tungstate anions are significantly larger than sulfate. The average metal-oxygen distance observed in the crystal structures of molybdate and tungstate salts are 1.77Å (Gatehouse & Leverett, 1969) and 1.78Å (Koster *et al.*, 1969), respectively, compared to 1.47Å for the sulfate anion (Baur, 1964). Consistent with these observations, ModA has a larger anion binding pocket than SBP. Calculations using the program VOIDOO (Kleywegt & Jones, 1994) show that the volumes of the binding sites are 64Å^3 and 77Å^3 for SBP and ModA, respectively. For these calculations, the bound anions were removed from the structures and a probe radius of 1.2Å (Hubbard & Argos, 1995) was used. A similar calculation with the structure of AvModA2 showed that the corresponding volume of this protein is 74Å^3 (Lawson *et al.*, 1998). Clearly, the anion binding pockets of ModA and AvModA2 are very similar in size and are significantly larger than that of SBP.

The radial distributions of the anion ligands around the central atom also reflect the larger size of the anion binding pocket in ModA as compared to SBP (Figure 1.16). The distances between the central molybdenum/tungsten or sulfur atom and the protein atoms donating hydrogen bonds to the anions average 3.9Å and 4.4Å for SBP and ModA, respectively. These distances agree very well with the corresponding values observed in aqueous solutions, which have been determined as 3.8Å and 4.1Å, for sulfate and molybdate/tungstate, respectively (Caminiti *et al.*, 1979, Johansson & Caminiti, 1986).

A role for steric factors has been previously discussed in the binding of ligand

to SBP (He & Quiocho, 1991). In that study, Ser 130 of SBP, which provides one of the hydrogen bonds to the bound sulfate, was replaced by Cys, Ala, and Gly residues. It was found that the replacement with Cys, which has a larger side chain than that of Ser, caused a 3200-fold decrease in the sulfate-binding activity relative to the wild-type activity. In contrast, replacement with Ala and Gly resulted in only 100- and 15-fold decreases, respectively. It was suggested that the effect of the Cys substitution was attributed largely to steric effect, whereas the Gly substitution more nearly reflected the loss of one hydrogen bond to the bound sulfate with a strength of only 1.6kcal/mole.

The expansion of the anion binding pocket in ModA relative to SBP does not occur isotropically, however, as is best demonstrated by superimposing the sulfate anion of SBP on the molybdate/tungstate anion of ModA. While most of the main chain atoms in the anion binding site of the N-domains of the two proteins superimpose to within 0.5Å, those in the C-domain of SBP are shifted by 1.2Å to 1.8Å toward the bound anions (Figure 1.22). If ModA were to bind a smaller anion like sulfate, it

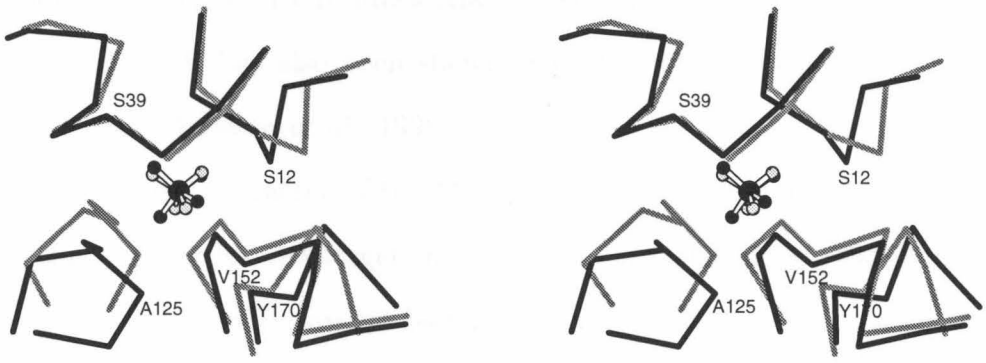


Figure 1.22: The superposition of the anion binding sites of ModA and SBP, showing the isotropical expansion of the anion binding pocket of ModA relative to SBP. The N domains of the two proteins are above the bound anions, while the C domains are below them. ModA is in black, SBP in grey. ModA residues donating hydrogen bonds to the bound anions are labeled. The two structures were first superimposed by least square method as described in Figure 1.18; the structure of SBP was then translated so that the sulfur atom of the bound sulfate coincided with the molybdenum atom of the molybdate in ModA.

would need to move the two domains closer together. Because four main chain NH groups at the termini of α-helices are involved in the binding of molybdate/tungstate anion, the binding site of ModA would appear relatively inflexible, since small, local conformational changes of side-chains can not be used to compensate for the difference in the sizes of the anions. Readjustments may cause steric clashes in the inter-domain interface and in other parts of the protein, or result in unfavorable hydrogen bond angles.

Polarity of binding pocket

Examination of the radial distribution of atom types around the central atom of the bound anions illustrates not only the differences in cavity size, but also differences in the polarity of the protein environment surrounding the anions. In particular, the protein surface adjacent to the molybdate/tungstate of ModA has appreciable apolar character relative to SBP and PBP, as indicated by the relatively high proportion of carbon atoms intermixed with ligand atoms (Figure 1.16). These atoms are provided by the side chains of ModA residues Ala 10, Ala 11, Ala 58, Val 123, Pro 124, Val 152 and Tyr 170. It has also been shown that the binding pocket of AvModA2 is virtually neutral (Lawson *et al.*, 1998).

The more apolar character of the ModA binding pocket may reinforce the specificity of this protein for the larger molybdate/tungstate anions relative to sulfate; simple electrostatic considerations based on Born charging energies suggest that it is energetically more unfavorable to transfer smaller ions to an apolar environment than larger ions.

Type of ligands

The differences in pK_a's between molybdate and sulfate should be reflected in the strength of hydrogen bonds to the surrounding ligands. Using a series of organic

compounds, Shan *et al.* (1996) have shown that there was a linear correlation between the increase in hydrogen bond energy and the decrease in ΔpK_a between the donor and acceptor of the hydrogen bond, as expected from simple electrostatic effects. Applying this to the hydrogen bonds between sulfate or molybdate and their binding proteins, it can be expected that the more basic molybdate would form stronger hydrogen bonds.

Furthermore, although both ModA and SBP donate seven hydrogen bonds to the bound anion, there are differences in the chemical nature of the hydrogen bond donors. More specifically, the number of NH and OH hydrogen bond donors to the anion are four and three, respectively, for the *E. coli* ModA, as compared to six and one, respectively, for SBP. The increased number of OH ligands in ModA may be mechanistically significant for the selective binding of molybdate and tungstate. For example, molybdate and tungstate bind with high affinity to poly-hydroxyl compounds such as sugars (Bourne *et al.*, 1961); this forms the basis of a molybdate and tungstate specific anion-exchange technique (Schilde & Uhlemann, 1993). The importance of this structural feature of *E. coli* ModA should not be overestimated, however, since in AvModA2, only two OH groups are involved in the binding of molybdate (Lawson *et al.*, 1998).

1.6 Conclusion

The general features identified in the binding of sulfate to SBP are consistent with the present observations of the binding of molybdate/tungstate to ModA; both proteins bind desolvated and deprotonated oxyanions through hydrogen bonds donated by protein residues. As the comparison between ModA and SBP demonstrates, however, discrimination between different oxyanions can be achieved through variations in the details of these binding interactions. In particular, the presence of a larger binding site and variation in the type of ligands are used by ModA to achieve specificity for molybdate binding, relative to sulfate.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallographica,* **D52**, 30–42.

Altschul, S. F., Modden, T. L., Schäffer, A. A., Zhan, J. H., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* **25** (17), 3389–3402.

Ames, G. F.-L. (1986). Bacterial periplasmic transport systems - structure, mechanism, and evolution. *Annual Review of Biochemistry,* **55**, 397–425.

Ames, G. F.-L. & Lever, J. E. (1972). The histidine-binding protein J is a component of histidine transport. *Journal of Biological Chemistry,* **247**, 4309–4316.

Ames, G. F.-L. & Nikaido, K. (1978). Identification of a membrane protein as a histidine transport component in *Salmonella typhimurium. Proceedings of the National Academy of Sciences of the United States of America,* **75**, 5447–5451.

Anderson, L. A., Palmer, T., Price, N. C., Bornemann, S., Boxer, D. H. & Pau, R. N. (1997). Characterization of the molybdenum-responsive ModE regulatory protein and its binding to the promoter region of the *modABCD* (molybdenum transport) operon of *Escherichia coli. European Journal of Biochemistry,* **246** (1), 119–126.

Baur, W. (1964). On the crystal chemistry of salt hydrates. IV. The refinement of the crystal structure of $MgSO_4 \cdot 7H_2O$ (Epsomite). *Acta Crystallographica,* **17**, 1361–1369.

Berger, E. A. (1973). Different mechanisms of energy coupling for the active transport of proline and glutamine in *Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America,* **70**, 1514–1518.

Berger, E. A. & Heppel, L. A. (1974). Different mechanisms of energy coupling for the shock-sensitive and shock-resistant amino acid permeases of *Escherichia coli. Journal of Biological Chemistry,* **249**, 7747–7755.

Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography.* Molecular Biology: An International Series of Monographs and Textbooks, Academic Press Inc., New York, New York.

Boos, W. & Lucht, J. M. (1996). Periplasmic binding protein-dependent ABC transporters. In Escherichia coli *and* Salmonella typhimurium *Cellular and Molecular Biology*, (Neidhardt, F. C., ed.), vol. 1, pp. 1175–1209. ASM Press Washington, D.C.

Bortels, H. (1930). Molybdan als katalysator bei biologischen Strickstoffbindung. *Archives of Microbiology,* **1**, 333–342.

Bourne, E. J., Hutson, D. H. & Weigel, H. (1961). Complex between molybdate and acyclic polyhydroxy compounds. *Journal of Chemical Society,* 35–38.

Brünger, A. T. (1992). Free R-value - a novel statistical quantity for assessing the accuracy of crystal structures. *Nature,* **355** (6359), 472–475.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R-factor refinement by molecular dynamics. *Science,* **235** (4787), 458–460.

Caminiti, R., Paschina, G., Pinna, G. & Magini, M. (1979). Experimental evidence of interactions SO_4^{2-}–H_2O in an aqueous solution. *Chemical Physics Letters,* **64** (2), 391–395.

CCP4 (1994). The CCP4 suite - programs for protein crystallography. *Acta Crystallographica,* **D50** (Part 5), 760–763.

Cohen, H. J., Fridovich, E. & Rajagopalan, K. V. (1971). Hepatic sulfite oxidase. A functional role for molybdenum. *Journal of Biological Chemistry,* **246**, 374–382.

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S. & Barrell, B. G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature,* **393** (6685), 537–544.

Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Aujay, M., Huber, R., Feldman, R. A., Short, J. M., Olsen, G. J. & Swanson, R. V. (1998). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus. Nature,* **392** (6674), 353–358.

Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallographica,* **A47** (JUL), 392–400.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., Fitzhugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L. I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C.,

Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O. & Venter, J. C. (1995). Whole genome random sequencing and assembly of *Haemophilus influenzae Rd. Science,* **269** (5223), 496–512.

Furlong, C. E. (1987). Osmotic-shock-sensitive transport systems. In Escherichia coli *and* Salmonella typhimurium *Cellular and Molecular Biology*, (Neidhardt, F. C., ed.), vol. 1, pp. 768–796. American Society for Microbiology Washington, D.C.

Gatehouse, B. M. & Leverett, P. (1969). Crystal structure of potassium molybdate, K_2MoO_4. *Journal of Chemical Society,* 849–854.

Gilson, E., Nikaido, H. & Hofnung, M. (1982). Sequence of the *malK* gene in *Escherichia coli* k12. *Nucleic Acids Research,* **10** (22), 7449–7458.

Glaser, J. H. & DeMoss, J. A. (1971). Phenotypic restoration by molybdate of nitrate reductase activity in *chlD* mutants of *Escherichia coli. Journal of Bacteriology,* **108**, 854–860.

Grunden, A. M., Ray, R. M., Rosentel, J. K., Healy, F. G. & Shanmugam, K. T. (1996). Repression of the *Escherichia coli modABCD* (molybdate transport) operon by *modE. Journal of Bacteriology,* **178** (3), 735–744.

Grunden, A. M. & Shanmugam, K. T. (1997). Molybdate transport and regulation in bacteria. *Archives of Microbiology,* **168**, 345–354.

He, J. J. & Quiocho, F. A. (1991). A non-conservative serine to cysteine mutation in the sulfate-binding protein, a transport receptor. *Science,* **251** (5000), 1479–1481.

He, J. J. & Quiocho, F. A. (1993). Dominant role of local dipoles in stabilizing uncompensated charges on a sulfate sequestered in a periplasmic active transport protein. *Protein Science,* **2** (10), 1643–1647.

Higgins, C. F. (1992). ABC transporters - from microorganisms to man. *Annual Review of Cell Biology,* **8**, 67–113.

Higgins, C. F., Gallagher, M. P., Hyde, S. C., Mimmack, M. L. & Pearce, S. R. (1990). Periplasmic binding protein-dependent transport systems - the membrane-associated components. *Philosophical Transactions of the Royal Society of London,* **B326**, 353–365.

Higgins, C. F., Haag, P. D., Nikaido, K., Ardeshir, F., Garcia, G. & Ames, G. F.-L. (1982). Complete nucleotide sequence and identification of membrane components of the histidine transport operon of *Salmonella typhimurium. Nature,* **298** (5876), 723–727.

Hu, Y. L., Rech, S., Gunsalus, R. P. & Rees, D. C. (1997). Crystal structure of the molybdate binding protein ModA. *Nature Structural Biology,* **4** (9), 703–707.

Hubbard, S. J. & Argos, P. (1995). Detection of internal cavities in globular proteins. *Protein Engineering,* **8** (10), 1011–1015.

Hutchinson, E. G. & Thornton, J. M. (1996). PROMOTIF - a program to identify and analyze structural motifs in proteins. *Protein Science,* **5**, 212–220.

Imperial, J., Hadi, M. & Amy, N. K. (1998). Molybdate binding by ModA, the periplasmic component of the *Escherichia coli mod* molybdate transport system. *Biochimica Et Biophysica Acta-Biomembranes,* **1370** (2), 337–346.

Jiang, J. S. & Brünger, A. T. (1994). Protein hydration observed by X-ray diffraction - solvation properties of penicillopepsin and neuraminidase crystal structures. *Journal of Molecular Biology,* **243** (1), 100–115.

Johann, S. & Hinton, S. M. (1987). Cloning and nucleotide sequence of the *chlD* locus. *Journal of Bacteriology,* **169** (5), 1911–1916.

Johansson, G. & Caminiti, R. (1986). The hydration of tungstate and molybdate ions in aqueous solution. *Zeitschrift Fur Naturforschung,* **A41**, 1325–1329.

Jones, T. A. (1985). Interactive computer-graphics - FRODO. *Methods in Enzymology,* **115**, 157–171.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallographica,* **A47**, 110–119.

Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugiura, M. & Tabata, S. (1995). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis sp.* strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. *DNA Research,* **2** (4), 153–166.

Kim, J. S. & Rees, D. C. (1992). Structural models for the metal centers in the nitrogenase molybdenum-iron protein. *Science,* **257** (5077), 1677–1682.

Kleywegt, G. J. & Jones, T. A. (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallographica,* **D50**, 178–185.

Kleywegt, G. J. & Jones, T. A. (1996). xdlMAPMAN and xdlDATAMAN - programs for reformatting, analysis and manipulation of biomacromolecular electron density maps and reflection data sets. *Acta Crystallographica,* **D52**, 826–828.

Koster, A. S., Kools, F. X. N. M. & Rieck, G. D. (1969). The crystal structure of potassium tungstate, K_2WO_4. *Acta Crystallographica,* **B25**, 1704–1708.

Kraulis, P. J. (1991). MOLSCRIPT - a program to produce both detailed and schematic plots of protein structures. *Journal of Applied Crystallography,* **24** (OCT), 946–950.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V. *et al.* (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis. Nature,* **390** (6657), 249–256.

Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK - a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography,* **26**, 283–291.

Lawson, D. M., Williams, C. E., White, D. J., Choay, A. P., Mitchenall, L. A. & Pau, R. N. (1997). Protein ligands for molybdate - specificity and charge stabilization at the anion-binding sites of periplasmic and intracellular molybdate-binding proteins of *Azotobacter vinelandii. Journal of the Chemical Society Dalton Transactions,* 3981–3984.

Lawson, D. M., Williams, C. E. M., Mitchenall, L. A. & Pay, R. N. (1998). An alternative binding site for molybdate: the 1.2Å resolution crystal structure of a periplasmic molybdate-binding protein from *Azotobacter vinelandii. Structure,* **submitted**.

Lee, J. H., Wendt, J. C. & Shanmugam, K. T. (1990). Identification of a new gene, *molR*, essential for utilization of molybdate by *Escherichia coli*. *Journal of Bacteriology,* **172** (4), 2079–2087.

Luecke, H. & Quiocho, F. A. (1990). High specificity of a phosphate-transport protein determined by hydrogen bonds. *Nature,* **347** (6291), 402–406.

Luque, F., Mitchenall, L. A., Chapman, M., Christine, R. & Pau, R. N. (1993). Characterization of genes involved in molybdenum transport in *Azotobacter vinelandii*. *Molecular Microbiology,* **7** (3), 447–459.

Matthews, M. W. (1968). Solvent content of protein crystals. *Journal of Molecular Biology,* **33**, 491–497.

Maupin-Furlow, J. A., Rosentel, J. K., Lee, J. H., Deppenmeier, U., Gunsalus, R. P. & Shanmugam, K. T. (1995). Genetic analysis of the *modABCD* (molybdate transport) operon of *Escherichia coli*. *Journal of Bacteriology,* **177** (17), 4851–4856.

McNicholas, P. M., Rech, S. A. & Gunsalus, R. P. (1997). Characterization of the ModE DNA-binding sites in the control regions of *modABCD* and *moaABCDE* of *Escherichia coli*. *Molecular Microbiology,* **23** (3), 515–524.

McRee, D. E. (1992). A visual protein crystallographic software system for X11/XView. *Journal of Molecular Graphics,* **10**, 44–46.

Menendez, C., Otto, A., Igloi, G., Nick, P., Brandsch, R., Schubach, B. & Bottcher, B. (1997). Molybdate-uptake genes and molybdopterin-biosynthesis genes on a bacterial plasmid - characterization of MoeA as a filament-forming protein with adenosinetriphosphatase activity. *European Journal of Biochemistry,* **250** (2), 524–531.

Miller, J. B., Scott, D. J. & Amy, N. K. (1987). Molybdenum sensitive transcriptional regulation of the *chlD* locus of *Escherichia coli*. *Journal of Bacteriology,* **169** (5), 1853–1860.

Neu, H. C. & Heppel, L. A. (1965). The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts. *Journal of Biological Chemistry,* **240**, 3685–3692.

Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods In Enzymology,* **276**, 307–326.

Pascal, M. C. & Chippaux, M. (1982). Involvement of a gene of the *chlE* locus in the regulation of the nitrate reductase operon. *Molecular & General Genetics,* **185** (2), 334–338.

Pflugrath, J. W. & Quiocho, F. A. (1988). The 2Å resolution structure of the sulfate-binding protein involved in active-transport in *Salmonella typhimurium*. *Journal of Molecular Biology,* **200** (1), 163–180.

Quiocho, F. A. (1990). Atomic structures of periplasmic binding proteins and the high-affinity active transport systems in bacteria. *Philosophical Transactions of the Royal Society of London,* **B326** (1236), 341–352.

Quiocho, F. A. (1991). Atomic structures and functions of periplasmic receptors for active transport and chemotaxis. *Current Opinion in Structure Biology,* **1**, 922–933.

Quiocho, F. A. & Ledvina, P. S. (1996). Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Molecular Microbiology,* **20** (1), 17–25.

Rajagopalan, K. V. & Johnson, J. L. (1992). The pterin molybdenum cofactors. *Journal of Biological Chemistry,* **267** (15), 10199–10202.

Ramachandran, G. N. & Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Advances in Protein Chemistry,* **23**, 283–437.

Rech, S., Deppenmeier, U. & Gunsalus, R. P. (1995). Regulation of the molybdate transport operon, *modABCD*, of *Escherichia coli* in response to molybdate availability. *Journal of Bacteriology,* **177** (4), 1023–1029.

Rech, S., Wolin, C. & Gunsalus, R. P. (1996). Properties of the periplasmic ModA molybdate-binding protein of *Escherichia coli*. *Journal of Biological Chemistry,* **271** (5), 2557–2562.

Sack, J. S., Saper, M. A. & Quiocho, F. A. (1989*a*). Periplasmic binding protein structure and function - refined X-ray structures of the leucine/isoleucine/valine-binding protein and its complex with leucine. *Journal of Molecular Biology,* **206** (1), 171–191.

Sack, J. S., Trakhanov, S. D., Tsigannik, I. H. & Quiocho, F. A. (1989*b*). Structure of the L-leucine-binding protein refined at 2.4Å resolution and comparison with the Leu/Ile/Val-binding protein structure. *Journal of Molecular Biology,* **206** (1), 193–207.

Schilde, U. & Uhlemann, E. (1993). Separation of several oxoanions with a special chelating resin containing methylamino-glucitol groups. *Reactive Polymers,* **20**, 181–188.

Shan, S. O., Loh, S. & Herschlag, D. (1996). The energetics of hydrogen-bonds in model systems: implications for enzymatic catalysis. *Science,* **272** (5258), 97–101.

Shanmugam, K. T., Stewart, V., Gunsalus, R. P., Boxer, D. H., Cole, J. A., Chippaux, M., Demoss, J. A., Giordano, G., Lin, E. C. C. & Rajagopalan, K. V. (1992). Proposed nomenclature for the genes involved in molybdenum metabolism in *Escherichia coli* and *Salmonella typhimurium*. *Molecular Microbiology,* **6** (22), 3452–3454.

Sharff, A. J., Rodseth, L. E., Spurlino, J. C. & Quiocho, F. A. (1992). Crystallographic evidence of a large ligand-induced hinge-twist motion between the 2 domains of the maltodextrin binding-protein involved in active transport and chemotaxis. *Biochemistry,* **31** (44), 10657–10663.

Smith, D. R., DoucetteStamm, L. A., Deloughery, C., Lee, H. M., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D. Y., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Prabhakar, S., McDougall, S., Shimer, G., Goyal, A., Pietrokovski, S., Church, G. M., Daniels, C. J., Mao, J. I., Rice, P., Nolling, J. & Reeve, J. N. (1997). Complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH: functional analysis and comparative genomics. *Journal of Bacteriology,* **179** (22), 7135–7155.

Stiefel, E. I. (1993). Molybdenum enzymes, cofactors, and chemistry: an introductory survey. In *Molybdenum enzymes, cofactors, and model systems*, (Stiefel, E. I., Coucouvanis, D. & Newton, W. E., eds), vol. 535, of *ACS Symposium Series* pp. 1–19. American Chemical Society.

Tam, R. & Saier, M. H. (1993). Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiological Reviews,* **57** (2), 320–346.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research,* **22** (22), 4673–4680.

Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L. X., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzegerald, L. M., Lee, N., Adams, M. D., Hickey, E. K., Berg, D. E., Gocayne, J. D., Utterback, T. R., Peterson, J. D., Kelley, J. M., Cotton, M. D., Weldman, J. M., Fujii, C., Bowman, C., Watthey, L., Wallin, E., Hayes, W. S., Weidman, J. M., Borodovsky, M., Karp, P. D., Smith, H. O., Fraser, C. M. & Venter, J. C. (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori. Nature,* **388** (6642), 539–547.

Tytko, K.-H. & Trobisch, U. (1987). Molybdenum: Molybdenum oxide hydrates, oxomolybdenum species in aqueous solutions. In *Gmelin Handbook of Inorganic Chemistry*, (Katscher, H. & Schröder, F., eds), vol. Supplement Volume B 3a, pp. 316–318. Springer-Verlag Berlin 8th edition.

Walkenhorst, H. M., Hemschemier, S. K. & Eichenlaub, R. (1995). Molecular analysis of the molybdate uptake operon, *modABCD*, of *Escherichia coli* and *modR*, a regulatory gene. *Microbiological Research,* **150**, 347–361.

Wang, G., Angermuller, S. & Klipp, W. (1993). Characterization of *Rhodobacter capsulatus* genes encoding a molybdenum transport system and putative molybdenum-pterin-binding proteins. *Journal of Bacteriology,* **175** (10), 3031–3042.

Williams, R. J. P. (1994). The biochemistry of molybdenum. In *Molybdenum: an outline of its chemistry and uses*, (Braithwaite, E. R. & Haber, J., eds), pp. 419–451. Elsevier Amsterdam.

Wilson, A. J. C. (1942). Determination of absolute from relative X-ray intensity data. *Nature,* **150**, 151–152.

Wright, R. M., Vaitaitis, G. M., Wilson, C. M., Repine, T. B., Terada, L. S. & Repine, J. E. (1993). cDNA cloning, characterization, and tissue-specific expression of human xanthine dehydrogenase xanthine oxidase. *Proceedings of the National Academy of Sciences of the United States of America,* **90** (22), 10690–10694.

Chapter 2 Crystal Structure of Formaldehyde Ferredoxin Oxidoreductase from *Pyrococcus furiosus*

2.1 Introduction

In this chapter, we report the 1.85Å resolution crystal structure of formaldehyde ferredoxin oxidoreductase (FOR) from *Pyrococcus furiosus* and an enzymatic mechanism proposed for FOR and homologous proteins based on structural studies. *Pf*FOR (Roy *et al.*, 1998) is a tungsten-containing protein. It is the third tungsten-containing protein, after aldehyde ferredoxin oxidoreductase (Mukund & Adams, 1991) and glyceraldehyde-3-phosphate ferredoxin oxidoreductase (Mukund & Adams, 1995), to be purified from the hyperthermophilic archaeon *Pyrococcus furiosus*.

2.1.1 Molybdenum- and Tungsten-containing Proteins

Molybdenum-containing enzymes (molybdoenzyme(s) for abbreviation) are ubiquitous in nature. The essential roles played by the molybdoenzymes in the metabolism of carbon, nitrogen, and sulfur have been known for more than six decades (Stiefel, 1993). However, the possibility of a biological function for the element tungsten (W, atomic number 74), which is chemically analogous to Mo, was only established in the early 1970s (Andreesen & Ljungdahl, 1973). This was not surprising, considering that besides W, the heaviest element with biological significance is iodine, which has an atomic number of 53. Instead, W has traditionally been regarded

as an antagonist of the biological functions of Mo. The first tungsten-containing protein (tungstoenzyme(s) for abbreviation), a formate dehydrogenase, was purified from *Clostridium thermoaceticum* in 1983 (Yamamoto *et al.*, 1983). Following the discovery that tungstate stimulated the growth of some hyperthermophilic archaea (Bryant & Adams, 1989), a series of tungstoenzymes have been identified.

The chemical properties of Mo and W are very similar. Not surprisingly, similarities exist between the structures and enzymatic mechanisms of W- and Mo-containing proteins. With the sole exception of nitrogenase MoFe protein, in which the molybdenum atom is part of a polymetallic Mo–Fe–S cluster (Kim & Rees, 1992), all the known Mo- and W-containing enzymes, including FOR, have mononuclear Mo or W sites that are coordinated by the sulfur atoms of molybdopterins.

Structures of Moco

The structure of the organic cofactor coordinating the Mo and W atoms in Mo- and W-containing enzymes is shown in Figure 2.1. This structure was first proposed by Rajagopalan and coworkers based on chemical analysis on modified or inactivated forms of the cofactor extracted from purified enzymes (Rajagopalan & Johnson, 1992) (Figure 2.1(a)), and was termed molybdopterin. The originally proposed structure was a bicyclic pterin derivative, with the pterin ring substituted at position 6 by a phosphorylated dihydroxybutyl side chain containing a *cis*-dithiolene group. X-ray crystallographic studies on *Pf* aldehyde ferredoxin oxidoreductase (AOR) established the general validity of this model (Chan *et al.*, 1995), but with the unanticipated observation that the molybdopterin is tricyclic. The third ring, a pyran, is formed by the attack of the 3′ hydroxyl group from the dihydroxybutyl side chain on the C7 atom of the pterin (Figure 2.1(b)).

The term molybdopterin has been widely used, but it causes confusion in some cases. First, it only refers to the metal-free organic cofactor despite its molybdo-

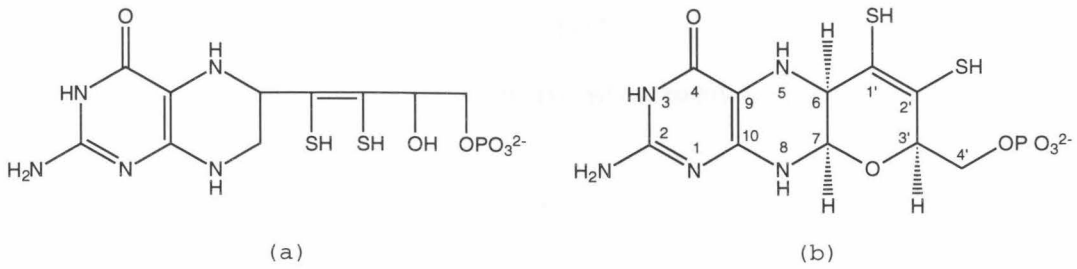


Figure 2.1: Structure of molybdopterin (a) proposed by Rajagopalan & Johnson (1992) and (b) observed in the crystal structure of *P. furiosus* AOR. In some enzymes the phosphate group is replaced by a dinucleotide. The atom numbering scheme is indicated.

prefix. Second, it is also found in tungstoenzymes. For this reason, the term Moco (molybdenum cofactor) has been suggested and will be used in this chapter. When Moco is used, it refers to the mononuclear Mo atom or W atom, the organic components that coordinate the metals, and any inorganic ligands of the Mo or W atoms (most frequently terminal oxo and sulfido groups). The term molybdopterin refers to the organic component of Moco, as shown in Figure 2.1(b). For convenience, terms like "molybdoenzyme" and "tungstoenzyme" will be used in addition to "Moco-containing proteins" in this chapter, but all of them refer to enzymes containing mononuclear Mo or W sites coordinated by molybdopterins. The nitrogenase MoFe protein will not be included in the discussions.

The molybdopterin was proposed to bind Mo and W atoms with its S atoms, with a stoichiometry of one or two molybdopterins per metal atom. The proposed coordination by the S atoms was verified by X-ray crystallographic studies on *Pf* AOR (Chan *et al.*, 1995) and all the subsequent studies on other Moco-containing enzymes. In enzymes from eukaryotic and archaeal systems, the molybdopterin possesses the structure shown in Figure 2.1, but in enzymes from prokaryotic systems it can be found as the dinucleotide of guanine, adenine, inosine, or cytidine-5′-monophosphate. The biosynthesis of molybdopterins has been reviewed (Hille, 1996). Enemark & Garner (1997) discussed the possible roles played by the Moco cofactors during the

enzymatic processes. With the exception of DMSO reductase, all the Moco-containing enzymes also possess additional redox-active sites, most frequently Fe_4S_4 clusters.

Families of Moco-Containing Proteins

Based on the types of reactions they catalyze, Moco-containing proteins may be divided into those that catalyze the transfer of oxygen atoms to or from electron lone pairs of substrates, and those that catalyze oxidative hydroxylation reactions of aldehydes (Hille, 1996, Kisker *et al.*, 1997*b*, Rees *et al.*, 1997). The former type includes two distinct families that are exemplified by DMSO reductase and sulfite oxidase, respectively. The latter type can also be subdivided into two families, a xanthine oxidase family which is exemplified by *Desulfovibrio gigas* aldehyde oxidoreductase (Mop), and a tungstoenzyme family which is exemplified by the aldehyde ferredoxin oxidoreductase (AOR) from *P. furiosus*. But exceptions have been observed. For example, *Escherichia coli* formate dehydrogenase H is classified in the DMSO family based on sequence similarities, but the substrate of this enzyme (formate) does not have electron lone pairs, and the enzymatic mechanism of this protein does not involve transfer of oxygen atoms; and instead proceeds through hydride (H^-) transfer from substrate formate to molybdenum center (Khangulov *et al.*, 1998). Some Moco-containing proteins can not be classified to any of the four families (Hille, 1996).

The reactions catalyzed by some of the enzymes are as follows:

DMSO reductase:

$$(CH_3)_2S{=}O + 2H^+ + 2e^- \longleftrightarrow (CH_3)_2S + H_2O$$

Sulfite oxidase:

$$SO_3^{2-} + H_2O \longleftrightarrow SO_4^{2-} + 2H^+ + 2e^-$$

Aldehyde oxidoreductase:

$$\text{R-CHO} + H_2O \longleftrightarrow \text{R-COOH} + 2H^+ + 2e^-$$

Members inside each family share substantial amino acid sequence homology

(Johnson *et al.*, 1996, Wootton *et al.*, 1991). For example, enzymes of the xanthine oxidase family typically have approximately 25% sequence identity and 60–70% sequence similarity with each other. Some structural features have also been observed that are common to the members within a family. For example, the molybdenum atoms in proteins from DMSO reductase and sulfite oxidase families are coordinated by peptide groups in addition to the molybdopterin ligands. For DMSO reductases and *E. coli* formate dehydrogenase H, the peptide ligands are serine and selenocysteine residues, respectively. For sulfite oxidases, it is a cysteine.

Crystal structures are available for at least one enzyme from each family: for sulfite oxidase family, it is the sulfite oxidase from chicken liver (Kisker *et al.*, 1997*a*); for DMSO reductase family, they are the DMSO reductases from *Rhodobacter sphaeroides* (Schindelin *et al.*, 1996) and *Rhodobacter capsulatus* (Schneider *et al.*, 1996, McAlpine *et al.*, 1998) and formate dehydrogenase H from *Escherichia coli* (Boyington *et al.*, 1997); for xanthine oxidase family, it is the aldehyde oxidoreductase (Mop) from *D. gigas* (Romão *et al.*, 1995); for the aldehyde oxidoreductase family, it is *P. furiosus* aldehyde ferredoxin oxidoreductase (AOR) (Chan *et al.*, 1995), which was the first Moco-containing protein to be characterized by X-ray crystallography. The structures of these proteins and their mechanistic implications have been discussed by Kisker *et al.* (1997*b*) .

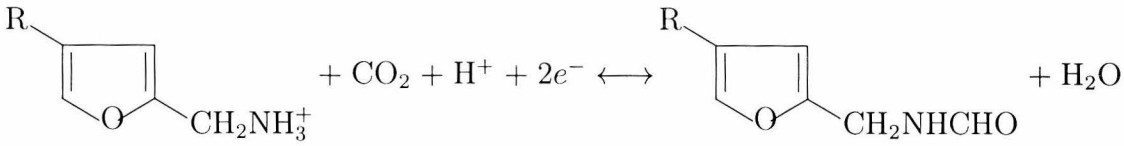
In this chapter, we report the 1.85Å resolution crystal structure of FOR, a member of the AOR-family. An enzymatic mechanism has been proposed for FOR and related tungstoenzymes including AOR based on the structural studies and comparisons with other Moco-containing proteins that is similar to that of Mop, an unrelated molybdoenzyme.

2.1.2 Tungsten-Containing Enzymes

At present, fourteen tungsten-containing enzymes have been purified and biochemically characterized, in contrast to more than 60 molybdoenzymes that have been studied (Hille, 1996). There have been great progresses in this field in the past few years and all but three of these tungstoenzymes were purified after 1990.

Three of the tungstoenzymes, including the formate dehydrogenase from *Clostridium thermoaceticum*, the first tungstoenzyme to be purified (Yamamoto *et al.*, 1983), are formate or formylmethanofuran dehydrogenases that utilize CO_2 as substrate. They catalyze the reversible oxidation of formate (formate dehydrogenase, or FDH) or reversible formation of N-formylmethanofuran (formylmethanofuran dehydrogenase, or FMDH), the first step in CO_2 utilization:

$$CO_2 + 2H^+ + 2e^- \longleftrightarrow HCOO^-$$



The FDHs from aerobic microorganisms usually do not contain metals or other cofactors, and use NAD(H) as electron carrier. In most cases, they function to oxidize formate. In contrast, FDHs from anaerobes usually contain metal centers, in most cases molybdenum, and catalyze the formation of formate. Growth studies have indicated that some Gram-positive bacteria and methanogens have W-containing FDHs, but only the W-containing FDH from *Clostridium thermoaceticum* (Yamamoto *et al.*, 1983) has been purified.

The FDH from *C. themoaceticum* does not have sequence similarities with AOR-family. Instead, it is homologous to molybdenum-containing FDHs, such as the FDH from *Methanobacterium formicicum* (Shuber *et al.*, 1986), a member of the the DMSO family.

FMDHs are only found in methanogens and usually Mo-containing enzymes, but two W-containing FMDHs have been purified recently, from *Methanobacterium thermoautotrophicum* (Bertram *et al.*, 1994) and *Methanobacterium wolfei* (Schmitz *et al.*, 1992*b*). Remarkably, both *Mb. wolfei* and *Mb. thermoautotrophicum* contain two FMDHs: one has a molybdenum while the other has a tungsten at the active site, and both proteins are enzymatically active (Bertram *et al.*, 1994, Schmitz *et al.*, 1992*b*). These are the only examples where molybdenum and tungsten can be interchanged to yield active enzymes. Both tungsten-containing FMDHs are multisubunit proteins and a sequence identity of 95% has been observed between the largest subunits of these two proteins, which are also the Moco-containing subunits. Both enzymes are also homologous to the molybdenum-containing FMDH from *Methanosarcina barkeri* (Vorholt *et al.*, 1996), which is also a multisubunit protein and a member of the DMSO family (Hille, 1996). It is therefore likely that all the three W-containing F(M)DHs belong to the DMSO family.

Acetylene hydratase (AH) from anaerobe *Pelobacter acetylenicus* is the most recently discovered and the least characterized (Rosner & Schink, 1995). The presence of a CX_2CX_3C sequence suggests that this protein also contains an Fe_4S_4 cluster. AH catalyzes the non-redox reaction of acetylene hydration:

$$HC{\equiv}CH + H_2O \longrightarrow CH_3CHO$$

but it is only catalytically active in the presence of strong reducing agents. It has been proposed that the mechanism of this protein may proceed through the initial reduction of acetylene, followed by hydration and oxidation (Rosner & Schink, 1995). This mechanism involves the potentially redox active tungsten atom and Fe_4S_4 cluster. This protein does not have sequence similarities with any other Moco-containing proteins and may belong to a new family.

The remaining ten tungstoenzymes belong to the AOR family. They catalyze the reversible oxidation of aldehydes to carboxylic acids:

$$\text{R-CHO} + \text{H}_2\text{O} \longleftrightarrow \text{R-COOH} + 2\text{H}^+ + 2e^-$$

Members of this family are the best characterized tungstoenzymes.

Six of the AOR family proteins have been purified from hyperthermophilic archaea by Adams and coworkers (Johnson *et al.*, 1996). Of these, *Pf* AOR was the first to be purified (Mukund & Adams, 1991) and is the best characterized. It oxidizes a broad range of aldehydes, both aliphatic and aromatic, but its highest catalytical efficiencies are with aldehyde derivatives of the common amino acids, such as acetaldehyde (from alanine), isovalerylaldehyde (from valine), and phenylacetaldehyde (from phenylalanine) (Mukund & Adams, 1991). It was also the first Moco-containing enzyme to be characterized by X-ray crystallography (Chan *et al.*, 1995). AOR is a homodimer of 67kDa, where each monomer contains a tungsten atom and an Fe_4S_4 cluster. The tungsten atom is coordinated by four sulfur atoms from two molybdopterins, which is held inside protein by hydrogen bond and salt bridge interactions. The Fe_4S_4 cluster is located 10Å from the tungsten atom. A homologous AOR has also been purified from *Thermococcus* sp. ES-1 (Heider *et al.*, 1995).

Two FORs have been purified, from *P. furiosus* (Roy *et al.*, 1998) and *T. litoralis* (Mukund & Adams, 1993), respectively. With an amino acid sequence identity of 87% and similarity of 92% (Kletzin *et al.*, 1995, Roy *et al.*, 1998), these two proteins are virtually identical. They are also homologous with *Pf* AOR (~50% amino acid sequence similarities). Both FORs are homotetramers with each subunit containing one tungsten atom and one Fe_4S_4 cluster. One calcium/subunit has also been detected in *Pf* FOR (Roy *et al.*, 1998).

Pf Glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR) was the third type of tungstoenzyme to be purified from hyperthermophiles and is the least characterized (Mukund & Adams, 1995). It is a monomeric protein of 63kDa. In addition

to the tungsten atom and an Fe_4S_4 cluster, this protein also contains 2 zinc ions per molecule. In contrast to AOR, GAPOR uses only glyceraldehyde-3-phosphate as a substrate. The amino acid sequence of this protein is homologous to those of AOR and FOR.

A feature common to all the tungstoenzymes purified from hyperthermophiles is that they all catalyze the oxidation of aldehydes and use the redox protein ferredoxin as their physiological electron carrier.

In addition, the *Pf* genome contains two additional genes that encode AOR-like enzymes of unknown function (Roy *et al.*, 1998). Moreover, the genomes of two related hyperthermophilic archaea, the fermentative *Pyrococcus horikoshii* (Kawarabayasi *et al.*, 1998) and the sulfate-reducing *Archaeoglobus fulgidus* (Klenk *et al.*, 1997), contain six and four AOR-like genes, respectively, some (but not all) of which are clearly of the AOR type.

Tungstoenzymes have also been identified in some acetogenic clostrodia. Three such enzymes have been purified, two from *Clostridium thermoaceticum*, and one from *Clostridium formicoaceticum*. Due to their abilities to reduce carboxylic acids, these proteins were named carboxylic acid reductase (CAR), but they also catalyze the reverse reaction, the oxidation of aldehydes to carboxylic acids. Unlike the tungstoenzymes from hyperthermophiles, the two CARs from *C. thermoaceticum* are heterooliogomers, whereas the *Cf* CAR is a homodimer. But the W-containing subunits of *Ct* CARs, which are also the largest subunits in the molecules, as well as the subunits of *Cf* CAR, are similar to the subunits of the tungstoenzymes from hyperthermophiles in term of size (all are in the range 63~69kDa). They also have homologous N-terminal amino acid sequences with those of hyperthermophilic tungstoenzymes. Based on their similarities with the hyperthermophilic tungstoenzymes, these three CARs are classified into the AOR family.

In the presence of molybdate, *C. formicoaceticum* also produces a molybdenum-

containing aldehyde oxidoreductase equivalent to *Cf* CAR, although their sequences are not related (White *et al.*, 1993). Similar Mo-containing aldehyde oxidoreductase may also be present in *Ct* (White & Simon, 1992).

The tenth member of AOR family is the aldehyde dehydrogenase (ADH) from *Desulfovibrio gigas*. This protein has a quaternary structure, subunit size, and W and Fe contents very similar to that of *Pf* AOR. Its N-terminal sequence is also homologous to other members of the AOR family.

Figure 2.2 shows the amino acid sequence alignment of the AOR-family tungstoenzymes with known sequences, as well as several AOR-like molecules identified by genome sequencing. The program CLUSTALW (Thompson *et al.*, 1994) was used to align the sequences. Sequence similarities are obvious for these molecules, especially for residues near tungsten sites and those involved in Moco and Fe_4S_4 cluster binding.

Almost every tungstoenzyme purified from mesophiles has an analogous molybdoenzyme that is present within the same organism or in a very closely related species, and the organisms producing these tungstoenzymes can survive under molybdenum-only condition (Kletzin & Adams, 1996). But the growth of hyperthermophilic archaea is obligately dependent on W, and Mo can not substitute for W in tungstoenzymes (Kletzin & Adams, 1996, Mukund & Adams, 1996).

The molecular properties of all the purified tungstoenzymes are listed in Table 2.1. Abbreviations used in the table are: *Pf* AOR: aldehyde ferredoxin oxidoreductase from *Pyrococcus furiosus* (Kletzin *et al.*, 1995, Mukund & Adams, 1991); ES-1 AOR: aldehyde ferredoxin oxidoreductase from *Thermococcus* sp. ES-1 (Heider *et al.*, 1995); *Pf* FOR: formaldehyde ferredoxin oxidoreductase from *Pyrococcus furiosus* (Heider *et al.*, 1995); *Tl* FOR: formaldehyde ferredoxin oxidoreductase from *Thermococcus litoralis* (Heider *et al.*, 1995, Mukund & Adams, 1993); *Pf* GAPOR: glyceraldehyde-3-phosphate ferredoxin oxidoreductase from *Pyrococcus furiosus* (Mukund & Adams, 1995); *Ct* CAR I: carboxylic acid reductase I from

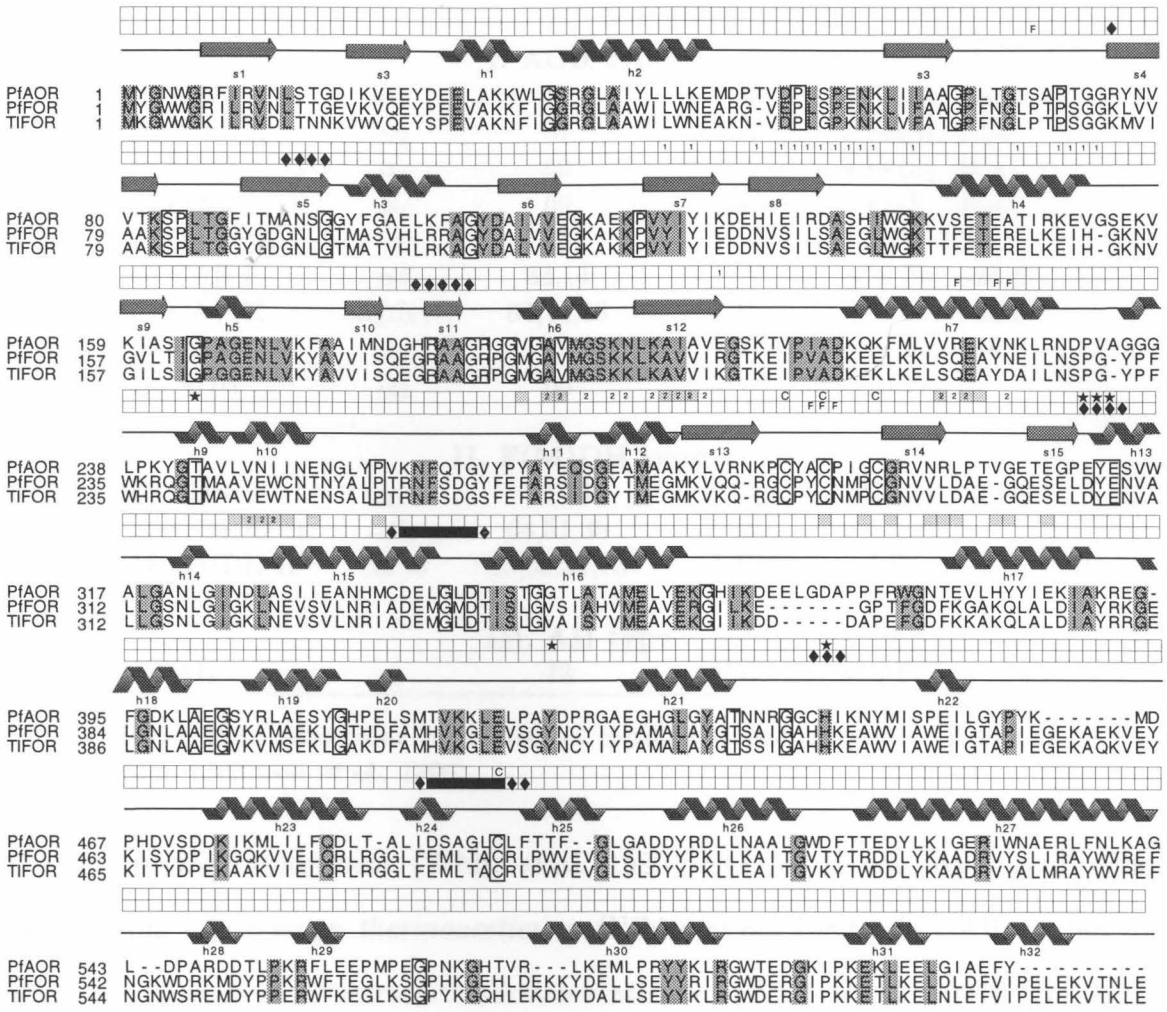


Figure 2.2: Amino acid sequence alignment for AOR-family tungstoenzymes. The sequences of *Pf* FOR, AOR, GAPOR, *Tl* FOR, and four AOR-like molecules identified in the genome of *A. fulgidus* were used in the alignment, but only the sequences of *Pf* FOR, *Pf* AOR, and *Tl* FOR are shown for simplicity. Residues conserved in all proteins are in boxes, while the highly conserved ones (only replaced by one or two similar residues) are in grey shadow. Secondary structures of FOR are marked by helices (for α- and 3_{10}-helices) and arrows (for β-strands). Residues involved in Moco binding are labeled by black squares. Residues in contact with Moco are labeled with $\diamond$; those closest to the tungsten site are labeled with $\star$. Cysteine ligands to the Fe_4S_4 cluster are designated by the letter C. Residues marked by 1 or 2 are those involved in the first and second inter-subunit interface of FOR, respectively, and those involved in Fd binding are marked by the letter F. Residues involved in the intermolecular interface of AOR are marked grey squares.

Enzymes	Subunits	Subunit Mass (kDa)	W Content	Other Cofactors
		I. AOR type		
Pf AOR	α_2	67	2 W	2 Fe_4S_4, 1Fe
ES-1 AOR	α_2	67	2 W	2 Fe_4S_4, 1 Fe?
Pf FOR	α_4	69	4 W	4 Fe_4S_4, 4Ca
Tl FOR	α_4	69	4 W	4 Fe_4S_4, 4Ca?
Pf GAPOR	α	63	1 W	~6 Fe, 2 Zn
Ct CAR I	$\alpha\beta$	64,14	1 W	~29 Fe, ~25 S
Ct CAR II	$\alpha_3\beta_3\gamma$	64,14,43	3 W	~82 Fe, ~54 S
Cf CAR	α_2	67	2 W	~11 Fe, ~16 S
Dg ADH	α_2	65	2 W	~10 Fe
		II. F(M)DH type		
Ct FDH	$\alpha_2\beta_2$	96,76	2 W	2 Se, 20–40 Fe
Mw FMDH II	$\alpha\beta\gamma$	64,51,35	1 W	2–5 Fe
Mt FMDH II	$\alpha\beta\gamma\delta$	65,53,31,15	1 W	~8 Fe
		III. AH type		
Pa AH	α	73	1 W	4–5 Fe

Table 2.1: Molecular properties of purified tungstoenzymes. Adopted from (Johnson *et al.*, 1996).

Clostridium thermoaceticum (White *et al.*, 1989); *Ct* CAR II: carboxylic acid reductase II from *Clostridium thermoaceticum* (White *et al.*, 1989); *Cf* CAR: carboxylic acid reductase from *Clostridium formicoaceticum* (White *et al.*, 1991); *Dg* ADH: aldehyde dehydrogenase from *Desulfovibrio gigas* (Trautwein *et al.*, 1994); *Ct* FDH: formate dehydrogenase from *Clostridium thermoaceticum* (Yamamoto *et al.*, 1983); *Mw* FMDH II: formylmethanofuran dehydrogenase II from *Methanobacterium wolfei* (Schmitz *et al.*, 1992*a*, Schmitz *et al.*, 1992*b*); *Mt* FMDH II: formylmethanofuran dehydrogenase II from *Methanobacterium thermoautotrophicum* (Bertram *et al.*, 1994, Hochheimer *et al.*, 1995); *Pa* AH: acetylene hydratase from *Pelobacter acetylenicus* (Rosner & Schink, 1995).

2.1.3 Physiological Functions and Enzymatic Mechanisms of FOR

AOR, FOR, and GAPOR have been purified from various species (see Table 2.1). All three enzymes catalyze the oxidation of various aldehydes using ferredoxin (Fd) as the electron acceptor:

$$\text{R-CHO} + \text{H}_2\text{O} + 2\text{Fd}_{ox} \longrightarrow \text{R-COOH} + 2\text{H}^+ + 2\text{Fd}_{red}$$

but they have very different substrate specificities: AOR oxidizes a broad range of aliphatic and aromatic aldehydes, FOR only oxidizes $C_1 \sim C_3$ aldehydes with activities much lower than that of AOR, and GAPOR only utilizes glyceraldehyde-3-phosphate. Based on their substrate specificities, AOR and GAPOR have been proposed to participate in the metabolism of peptides and carbohydrates, respectively. The physiological function of FOR is unknown.

Properties of AOR and GAPOR

Kinetic studies on AOR from *Thermococcus* strain ES–1 showed that the substrates with the highest k_{cat}/K_m values were acetaldehyde, isovalerylaldehyde, indoleacetaldehyde, and phenylacetaldehyde, all of which exhibit apparent $K_m \leq 100\mu M$ (Heider *et al.*, 1995, Adams & Kletzin, 1996). Similar data have been obtained for *Pf* AOR. Acetaldehyde, isovalerylaldehyde, indoleacetaldehyde, and phenylacetaldehyde are the aldehyde derivatives of alanine, leucine, tryptophan, and phenylalanine, respectively. It was thus proposed that AOR is part of

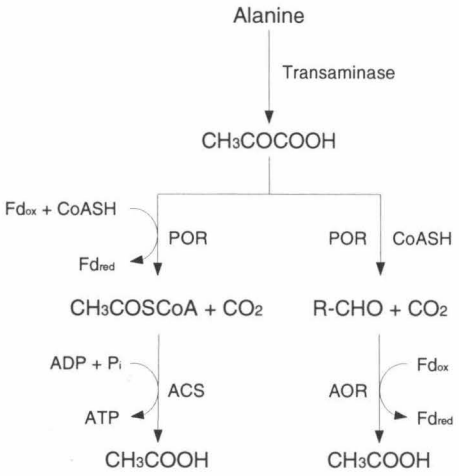


Figure 2.3: Proposed pathway for the metabolism of amino acids in heterotrophic hyperthermophilic archaea, using alanine as a example. Adopted from (Ma *et al.*, 1997).

the amino acid metabolism pathway of the hyperthermophilic archaea and functions to oxidize aldehydes generated during amino acid catabolism (Heider *et al.*, 1995).

Experiments by Ma *et al.* (1997) showed that the aldehyde substrates of AOR are generated from 2-keto acids by 2-keto acid ferredoxin oxidoreductases functioning as decarboxylases in CoA-dependent reactions . Corresponding 2-keto acid ferredoxin oxidoreductases, such as pyruvate ferredoxin oxidoreductases (POR), indolepyruvate ferredoxin oxidoreductases (IOR), and 2-keto isovelarate ferredoxin oxidoreductases (VOR), are abundant in *Pyrococcus furiosus* and have been purified and characterized. The 2-keto acids, in turn, are the products of transamination reactions of amino acids. The role played by AOR in amino acid metabolism is illustrated in Figure 2.3.

GAPOR, on the other hand, only oxidizes glyceraldehyde-3-phosphate (K_m 28μM). *P. furiosus* has been found to ferment sugars by an unusual Embden-Meyerhof pathway that involves ADP dependent kinases (Kengen *et al.*, 1994). But its cell extracts contain very low activities of two key glycolytic enzymes, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and phosphoglycerate kinase (PGK). Instead, these cell extracts have an abundance of GAPOR. Based on the high activity and strict specificity of GAPOR toward glyceraldehyde-3-phosphate, it is proposed that this protein functions in place of GAPDH and possibly PGK in the Embden-Meyerhof glycolytic pathway of *Pyrococcus furiosus* (Figure 2.4) (Mukund & Adams, 1995).

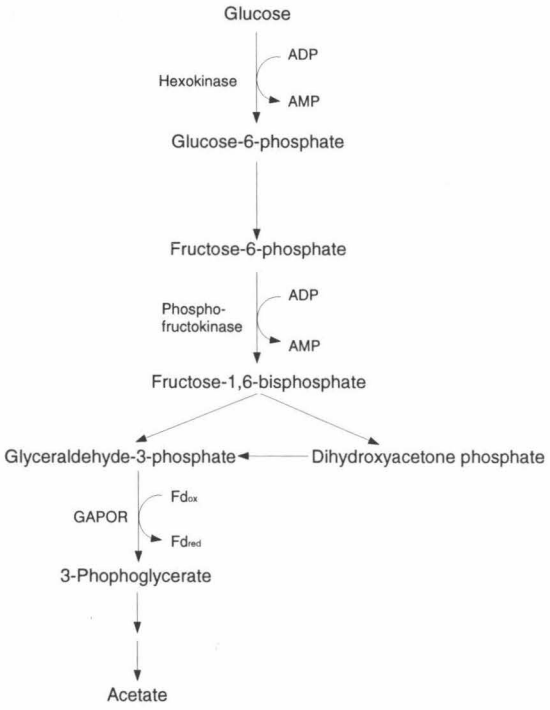


Figure 2.4: Proposed role of GAPOR in the conversion of glucose to acetate in *P. furiosus*. Adopted from (Mukund & Adams, 1995).

Both AOR and GAPOR use ferredoxin (Fd) as their physiological electron carrier. They have high affinities for this protein, with the K_m values less than 10μM (Heider *et al.*, 1995, Mukund & Adams, 1995). In *P. furiosus*, the oxidation of the reduced Fd is coupled to the reduction of H^+ to H_2, or, if S^0 is present, to the production of H_2S, via the enzymes NADP ferredoxin oxidoreductase and hydrogenase (Ma *et al.*, 1993).

Properties of FOR

Formaldehyde is not considered the real substrate of FOR, although this protein is named formaldehyde ferredoxin oxidoreductase and was purified by its ability to oxidize formaldehyde. The real substrate of this protein is still unknown. *P. furiosus* FOR does oxidize C2–C4 aliphatic aldehydes, but they are unlikely to be of physiological significance, as the enzyme has a very low affinity for such compounds ($K_m \geq$ 25 mM) (Table 2.2). FOR also exhibits weak activity with butyraldehyde (C4) and phenylacetaldehyde, but it does not oxidize isovaleraldehyde (C5) (Roy *et al.*, 1998).

Substrate	Apparent K_m (mM)	k_{cat} (s^{-1})
Formaldehyde	25	7.1×10^4
Acetaldehyde	60	4.4×10^4
Propionaldehyde	62	1.1×10^4
Phenylpropionaldehyde	15	2.5×10^4
Indole-3-acetaldehyde	25	2.3×10^3
Succinic semialdehyde	8	5.7×10^3
Glutaric dialdehyde	0.8	4.2×10^4
Adipic acid semialdehyde	33	3.0×10^4

Table 2.2: Kinetic properties of *Pf* FOR. Adopted from (Roy *et al.*, 1998).

It has been suggested that C5- or C6-type aldehydes might be physiologically relevant for *Pf* FOR (Roy *et al.*, 1998). The only chemical that exhibits a submillimolar K_m is glutaric dialdehyde (C5), but this compound is not in any known biochemical pathway. Like AOR and GAPOR, FOR utilizes ferredoxin as its physiological elec-

tron carrier, but with an apparent K_m of 100μM, the interaction between FOR and Fd is much weaker than the interactions between AOR or GAPOR and Fd.

The enzymatic mechanism of FOR is unknown. Nor do we know the enzymatic mechanisms of any other members of AOR family. In this chapter, we propose an enzymatic mechanism for FOR that is very similar to that of *Dg* Mop, an unrelated Mo-containing xanthine oxidase, based on crystal structure studies of *Pf* FOR and structural comparisons with Mop. This mechanism may also apply to other members of the AOR family since the residues close to the active sites are highly conserved for all members of AOR family with known sequences (Figure 2.2). We have also identified the residues that might be responsible for the specificity of FOR based on the interactions between FOR and glutarate, the oxidation product of glutaric dialdehyde. We also propose an electron transfer pathway between FOR and Fd, based on the structure of the FOR-Fd complex.

2.2 Crystallization and Data Collection

2.2.1 Crystallization

*Pf*FOR was purified (Roy *et al.*, 1998) and crystallized at room temperature under an argon-hydrogen atmosphere using a modification of the melting-point capillary method (Georgiadis *et al.*, 1992); ten microliters of precipitant solution were introduced into melting-point capillary, and separated by 2–3 mm from 10μl of 50:50 mixture of protein and precipitant solutions. The capillary was flipped several times to make the mixture of protein-precipitant solution uniform, and then left undisturbed for several weeks, with the precipitant solution on the top and the mixture of protein-precipitant solution at the bottom (Figure 2.5).

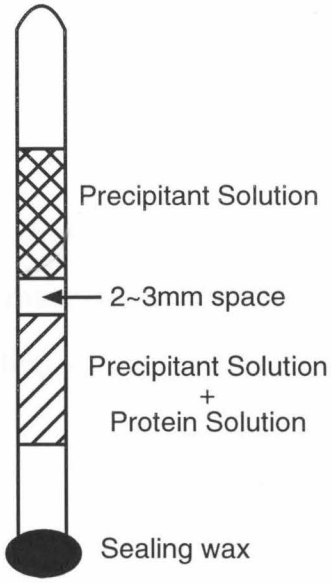


Figure 2.5: Setup of the capillaries.

The crystallization of FOR was much more reproducible with this modified method and the crystals were larger in size.

The protein solution contained 55–65 mg/ml of FOR, 50mM Tris at pH 8.0, 0.5mM dithionite, 2mM dithiothreitol, and 0.2M KCl. The precipitant solution contained 30% glycerol (v/v), 20% PEG 4000 (w/v), 0.1M sodium citrate at pH 5.6, and 0.2M NaCl. Sometimes, 10% glycerol was included in the protein solution, and the glycerol concentration in the precipitant solution would be lowered to 20% accordingly. Brown crystals generally appeared in 2 weeks and achieved maximum dimensions of 1.5mm×0.3mm×0.3mm after 4–6 weeks. A high concentration of protein was essential for the success of the crystallization experiment: FOR could only be crystallized when the protein concentration was 50mg/ml or higher. Glycerol was needed for the FOR crystals to be stable and of high diffracting quality. This is in agreement with the observation that the presence of 10% glycerol in solution stabilized FOR dur-

ing purification (Roy *et al.*, 1998). FOR crystals could be obtained in the absence of glycerol, but they were extremely fragile and difficult to handle, and only diffracted to low resolution. Including glycerol in precipitant solutions significantly enhanced the quality and stability of FOR crystals. Glycerol also increased the solubility of FOR, so high concentrations of PEG were needed in the precipitant solution to crystallize FOR.

The FOR-glutarate co-crystals were obtained under conditions similar to those of the native FOR crystals, except 50mM of sodium glutarate was included in the crystallization condition, and the concentration of the buffer, sodium citrate, was reduced to 30mM.

The FOR-Fd co-crystals were also obtained under similar conditions, with the presence of an excess amount of *Pf* ferredoxin (Fd) (molar ratio of 2:1 to 4:1). The concentration of sodium citrate in the precipitant solution was reduced to 10–30mM, and the concentration of NaCl was reduced to 0–50mM. The Fd solution used in the crystallization contained 60mg/ml of protein in 50mM Tris pH 8.0, 0.2M NaCl, and 2mM dithionite, and 0.5μl of this solution was added to 10μl of the mixture of FOR-precipitant solution. The FOR-Fd co-crystals have slightly different shapes compared to the native FOR crystals and are darker in color. Low salt concentrations were used to facilitate the formation of a specific FOR-Fd complex, but at very low salt concentration (10mM sodium citrate, no NaCl), the proteins did not crystallize; but instead tended to precipitate.

The native FOR crystals, FOR-glutarate co-crystals, and FOR-Fd co-crystals are isomorphous and belong to the space group $P2_12_12_1$, with cell dimensions a=100Å, b=170Å, c=180Å, and $\alpha=\beta=\gamma=90°$.

2.2.2 Data Collection

X-ray diffraction data were collected at liquid nitrogen temperature. The crystals were flash cooled to -180°C in liquid nitrogen on rayon loops prior to data collection, and were kept in a continuous nitrogen stream at approximately -180°C during the experiments. Two data sets, designated data set I and data set II, were collected to 2.55Å and 1.85Å resolution on FOR crystals on beam-line 7-1 and beam-line 9-1 at the Stanford Synchrotron Radiation Laboratory, respectively. For data set I, a total of 123 frames were collected with an oscillation angle of 1°/frame at a distance of 330mm. The exposure time was 3 minutes per frame. For data set II, an exposure time of 5 minutes and an oscillation angle of 1° per frame were used, and a total of 130 frames were collected at a distance of 250mm. The FOR crystals diffracted beyond 1.85Å resolution with SSRL beam-line 9-1 radiation, but it was only possible to collect data to this resolution due to the geometry limitation of the detector.

Crystals	I(Native FOR)	II(Native FOR)	FOR-glutarate Complex	FOR-Fd Complex
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell Dimensions	a=99.03Å b=171.10Å c=179.86Å	a=100.34Å b=170.85Å c=180.64Å	a=100.55Å b=170.98Å c=181.09Å	a=99.78Å b=170.66Å c=180.17Å
X-ray Source	SSRL BL7-1	SSRL BL9-1	Rigaku	SSRL BL7-1
Wavelength	1.08Å	0.98Å	1.54Å	1.08Å
Resolution	2.55Å	1.85Å	2.40Å	2.15Å
Last Shell	2.59–2.55Å	1.88–1.85Å	2.49–2.40Å	2.19–2.15Å
R_{sym} (last shell)	8.1%(23.4%)	7.4%(22.3%)	5.9%(15.8%)	8.3%(33.4%)
$<I>/\sigma(I)$ (last shell)	20.3(4.2)	13.9(2.6)	9.9(2.2)	13.1(2.1)
Completeness (last shell)	97.1%(73.4%)	91.4%(83.6%)	77.6%(40.3%)	90.6%(66.1%)
Total Reflections	403964	928734	193320	403899
Unique Reflections	97499	240510	93750	150319

Table 2.3: Statistics of the data collection.

Data collection for the FOR-glutarate co-crystal was performed on an RAXIS IIc image plate detector with CuKα (λ=1.54Å) radiation from a Rigaku rotating anode X-ray generator. The radiation was monochromatized with a mirror system. A data set to 2.4Å resolution was collected.

A data set to 2.15Å resolution was collected on an FOR-Fd co-crystal on SSRL beam line 7-1.

The data were processed using the programs DENZO and SCALEPACK (Otwinowski & Minor, 1997). The intensities were then scaled with the program ROTAVATA and merged with the program AGROVATA, and reduced to amplitudes with the program TRUNCATE (CCP4, 1994). The statistics of the data sets are summarized in Table 2.3. The mosaicities are 0.43° and 0.36° for data sets I and II, respectively, as calculated with the program SCALEPACK. The average temperature factors are 32Å^2 and 20Å^2 for data sets I and II, respectively, as estimated by the Wilson plots (Wilson, 1942) (Figure 2.6).

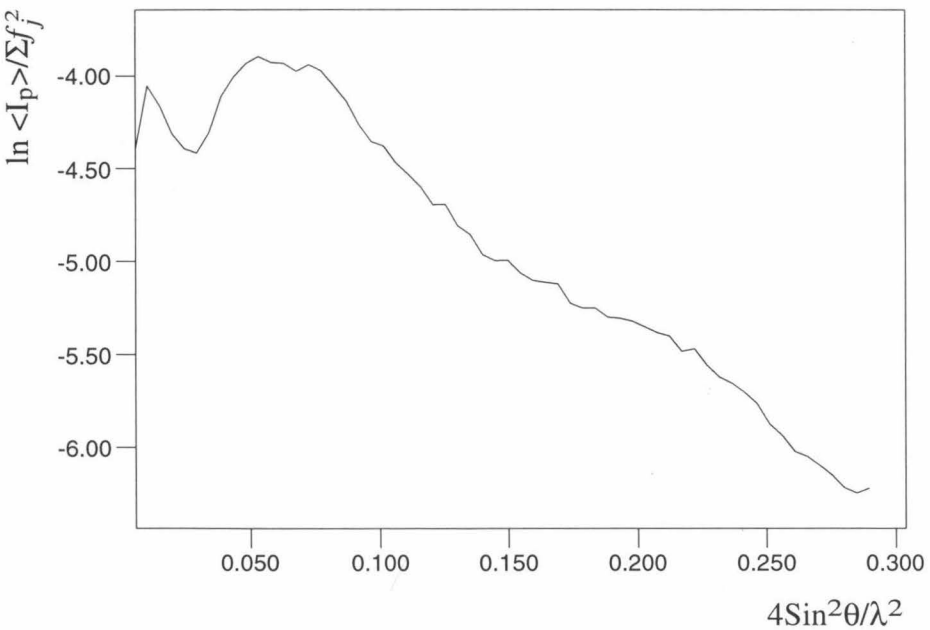


Figure 2.6: The Wilson plot of the native data set II.

2.3 Structure Determination and Refinement

2.3.1 Structure Determination of FOR

The crystal structure of FOR was solved by the molecular replacement method, with the structure of *Pf* AOR (Chan *et al.*, 1995, PDB access code 1AOR) as the initial model. Phase information from the anomalous scattering of the data set II was also used in the structural determination.

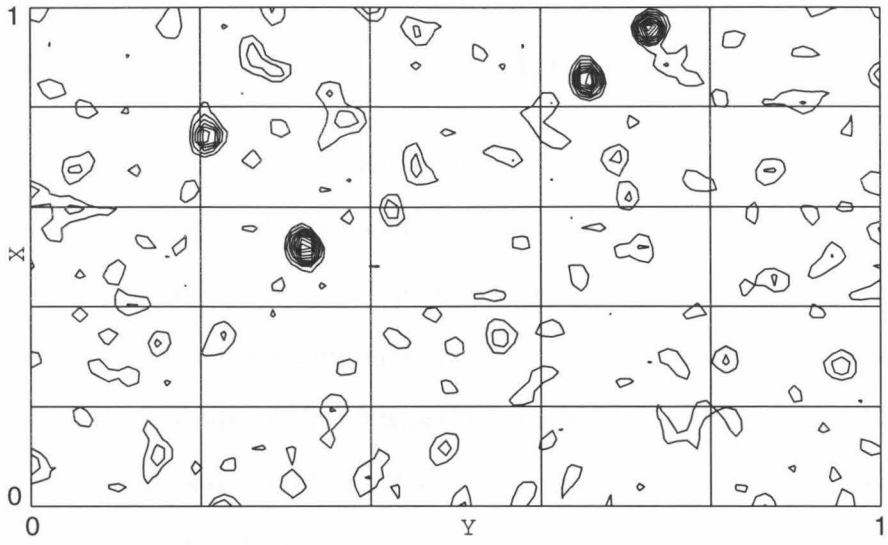


Figure 2.7: Harker section z=1/2 of the native anomalous Patterson map calculated with the data set II. Reflections in the 8–2.5Å resolution range were used.

Anomalous Patterson maps were calculated with reflections from native data set II in the resolution range of 8.0 to 2.5Å, using the program FFT (CCP4, 1994). From these maps (Figure 2.7), the positions of the four tungsten atoms in the asymmetric unit were derived. Assuming four monomers in an asymmetric unit, the FOR crystals have a Matthews coefficient (Matthews, 1968) of 2.9Å^3/Dalton.

The molecular replacement solution was obtained with the program AMoRe (Navaza & Saludjian, 1997). Various conditions, such as different data sets, resolution ranges, and search models, were tested, but the molecular replacement search

was successful only when the reflections in the resolution range 15.0–4.5Å from data set I were used. A polyalanine AOR monomer with Moco and Fe_4S_4 cluster served as the search model. Four solutions were obtained as follow:

1. α=118.8°, β=114.6°, γ=273.4°, t_x=–46.4Å, t_y=28.0Å, t_z=94.7Å;
2. α=174.7°, β=55.6°, γ=197.6°, t_x=63.0Å, t_y=80.9Å, t_z=–8.6Å;
3. α=151.9°, β=101.1°, γ=289.7°, t_x=11.4Å, t_y=51.3Å, t_z=119.3Å;
4. α=122.0°, β=147.6°, γ=276.6°, t_x=–26.2Å, t_y=11.4Å, t_z=114.8Å.

Here (t_x, t_y, t_z) is the translational vector of the solution, and α, β, and γ are the Eular angles of the rotational components of the solution.

Solutions 1, 2, and 3 were found to be correct, judged by the positions of the tungsten atoms calculated from anomalous Patterson maps. Based on the distances between the tungsten atoms, a dimer was then made from solutions 2 and 3 and used as the model for a second round of AMoRe search, which gave the correct solution for the second dimer (α=115.1°, β=85.9°, γ=245.3°, t_x=77.0Å, t_y =–97.2Å, and t_z =124.9Å).

Twenty cycles of rigid body refinement were then carried out in the resolution range 15.0–4.5Å to refine the positions of the four subunits, using the program X-PLOR version 3.1 (Brünger *et al.*, 1987). The refinement with X-PLOR also showed that there were no severe steric clashes between the subunits.

Initial phases were calculated from the molecular replacement solution and the anomalous scattering of data set II, to 4.5Å and 3.0Å resolution, using the programs SFALL and MLPHARE, respectively. The phase calculation with MLPHARE assumed that the four tungsten atoms were the only anomalous scatters, and their positions were first refined with the program VECREF for 20 cycles. The phases

calculated with SFALL and MLPHARE were then combined using the program SIGMAA (CCP4, 1994). The overall figure of merit was 0.49.

The phases were then refined with the program SOLOMON (Abrahams & Leslie, 1996) , using data set II. The refinement started with phases to 4.5Å resolution and were then extended to 3.0Å resolution in 50 steps. Solvent flipping and four-fold non-crystallographic symmetry (NCS) averaging were carried out during the refinement. The masks used in the NCS averaging were calculated from the molecular replacement solution using the program MAMA. The NCS operations were calculated from the X-PLOR refined positions of the subunits. The SOLOMON refinement increased the figure of merit to 0.95.

The electron density map calculated at 3.0Å resolution with the refined phases was of excellent quality (Figures 2.8 and 2.9) and was used for model building without further manipulations. There was only one discontinuity, between residues Gly 454 and Glu 461, in the main chain density of every monomer when contoured at 1σ. The Moco and Fe_4S_4 cofactors were clearly identifiable. Although a polyananine AOR model was used in the calculation of inital phases, most side chains were identifiable from the electron density map and corresponded very well with the amino acid sequence of FOR, indicating that the model bias was minimum after phase refinement. Experiments with different starting phases showed that if only the phase information from either molecular replacement or anomalous scattering were used, the calculated electron density map was of much inferior quality.

One monomer was first built into the electron density map using the program O (Jones *et al.*, 1991) running on an SGI workstation, and the other three monomers were then generated using the NCS operations. The program xdlMAPMAN (Kleywegt & Jones, 1996) was used to convert the electron density map calculated with SOLOMON to O format and to skeletonize of the map. The command *baton* in O was used to choose the positions of the Cα atoms from the skeletons,

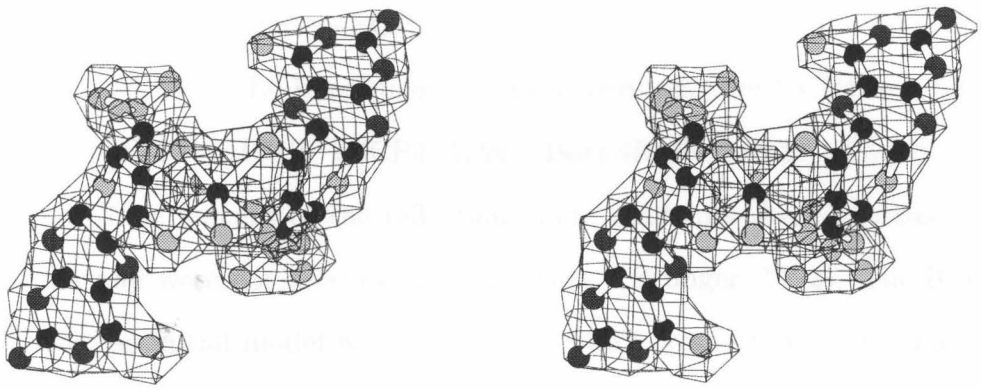


Figure 2.8: Stereoview of the electron density map around one of the Moco cofactors. Superimposed with the refined model. Calculated with the SOLOMON refined phases and contoured at 1σ.

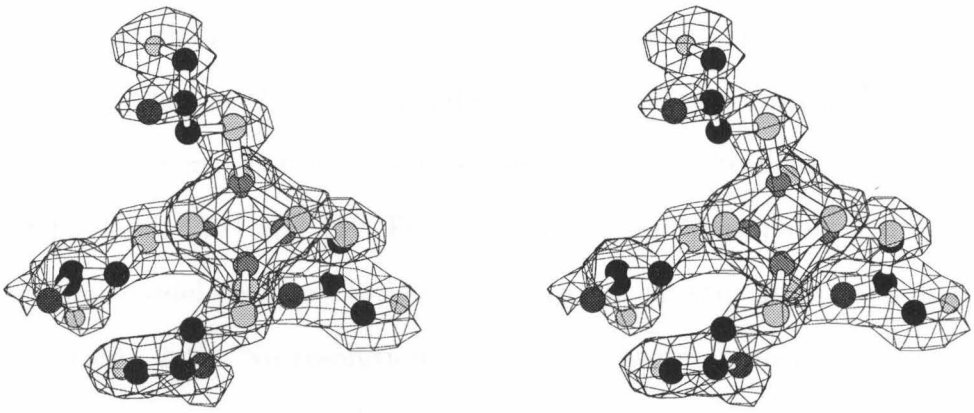


Figure 2.9: Stereoview of the electron density map around one of the Fe_4S_4 clusters and its cysteine ligands. Calculated with the SOLOMON refined phases and contoured at 1σ.

and *lego* commands were then used to build in the main chain and side chain atoms. Because the sequence of *Pf* FOR was not available when the model building was performed, the sequence of *Tl* FOR (Kletzin *et al.*, 1995) was used, and it was changed to the correct sequence of *Pf* FOR when it became available.

2.3.2 Structure Refinement of Native FOR Model

The structural refinement was only carried out for data set II, using the program X-PLOR version 3.1 and the parameters of Engh & Huber (1991).

The refinement of the native FOR structure started at 2.5Å resolution. The structure factors used in X-PLOR refinement were converted from CCP4 mtz files using the program MTZ2VARIOUS (CCP4, 1994). Both the F^+ and F^- of every reflection were output. Five percent of the reflections were randomly chosen to calculate the free R factor and were not used in the refinement (Brünger, 1992). The R and free R factors for the initial model were 41.4% and 44.6%, respectively. A simulated annealing refinement at 3000K dropped these values to 29.7% and 37.6%, respectively. Alternative cycles of positional refinement, B factor refinement, resolution extension, addition of water molecules, and manual rebuilding of the model were then carried out, and reduced the R and free R factors to 17.4% and 22.0%, respectively, for all the reflections in data set II. The manual rebuilding was performed independently for all the four monomers. Water molecules were built into the model only when they had good Fo–Fc ($> 3\sigma$) and 2Fo–Fc ($> 1\sigma$) maps and reasonable hydrogen bonding. The final model contains 19,132 non-hydrogen protein atoms and 1349 water molecules (Table 2.4). No resolution- or σ-cutoffs were applied in the refinement, and a bulk solvent correction (Jiang & Brünger, 1994) was performed. Anomalous

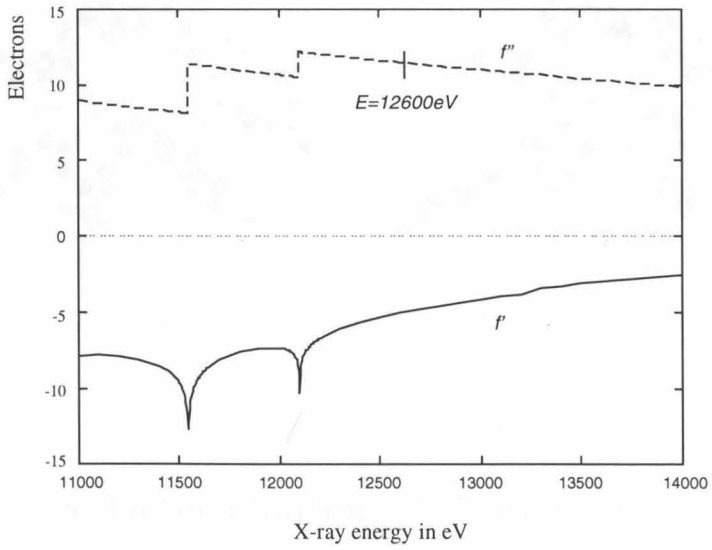


Figure 2.10: Anomalous scattering factors of tungsten at wavelengths near 0.98Å (energy=12600eV).

scattering was used in the refinement, using an f'' of 11.5 electrons (Figure 2.10) for the tungsten atoms. The NCS restraints were not used in the refinement, but small rms deviations of 0.126Å to 0.143Å are observed between monomer Cα atoms.

2.3.3 Structure Determination of FOR-glutarate Complex

The glutarate binding site was located in the FOR-glutarate complex by difference Fourier techniques, using the structure of the native FOR tetramer without water molecules as the initial model. Following rigid body refinement and simulated annealing at 3000K, positional refinement and individual B factor refinement were performed, and densities for glutarate molecules clearly showed up in the active site cavity of every FOR monomer (Figure 2.11). The glutarate molecules were then built into the model. The R and free R factors for the present model are 17.7% and 24.4%, respectively. The same set of reflections was used to calculate the free R factor as for the native FOR structure. In the refinement process, four-fold NCS restraints were applied to both protein (weight=75) and the bound glutarate (weight=250).

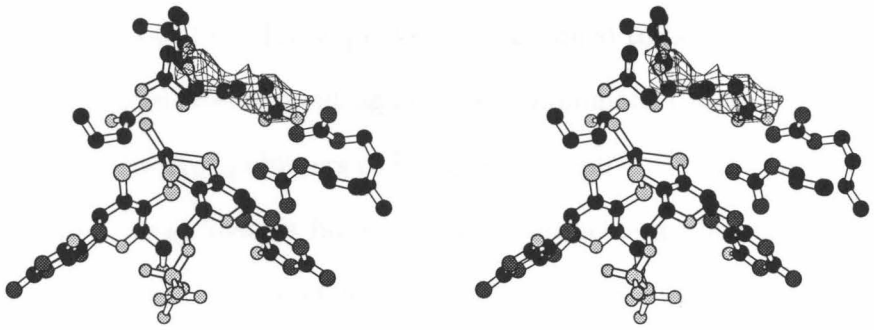


Figure 2.11: Fo–Fc omit map contoured at 2.5σ around one of the glutarates.

2.3.4 Structure Determination of FOR-Fd Complex

For the FOR-Fd complex structure, anomalous Fourier maps calculated with phases from the FOR model show two strong peaks near the Fe_4S_4 clusters of two

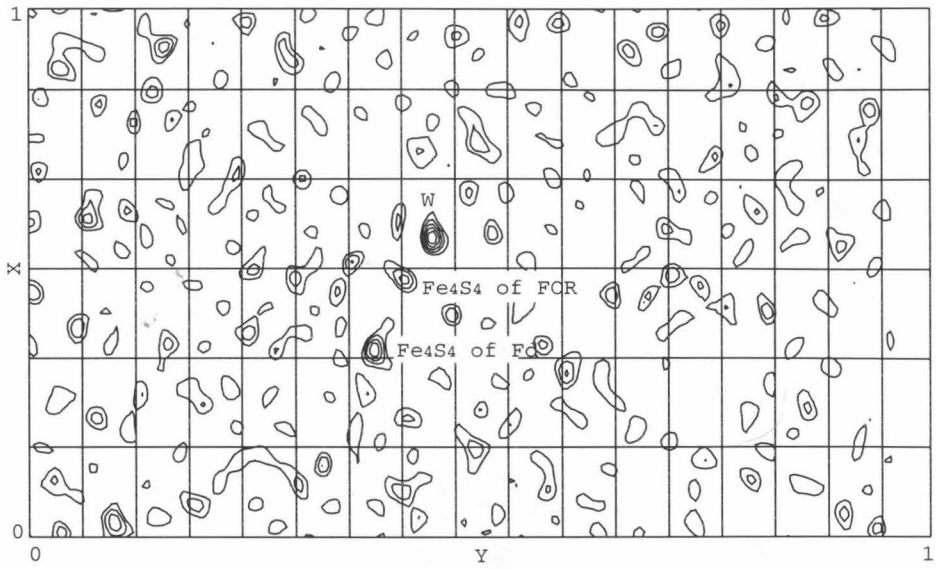


Figure 2.12: Anomalous Fourier map calculated at 5Å resolution with FOR-Fd data set, contoured at 1σ. Two of the peaks were generated by tungsten atom and Fe_4S_4 cluster, respectively, of FOR, and the third one was only present in the map calculated from FOR-Fd data set, but not in the map calculated with native FOR data sets, and is assumed to be from the Fe_4S_4 cluster of the bound Fd.

of the four FOR monomers which are not present with native FOR data sets (Figure 2.12). These peaks are about 9Å from the surface of FOR, and 15Å from center of the Fe_4S_4 clusters of FOR. These peaks were assumed to be from the Fe_4S_4 clusters of bound Fd molecules. Refinement against the anomalous differences indicated that the occupancies of the Fe_4S_4 clusters of Fd were 50%–60% of that of the Fe_4S_4 clusters of FOR, assuming both clusters have the same temperature factors. Corresponding positions of other two FOR monomers are blocked by FOR molecules from other tetramers.

Two Fd molecules were initially modeled using the structure of the homologous Fd from *Desulfovibrio africanus* (Sery *et al.*, 1994, PDB access cod 1FXR) by superimposing their α-helices in the appropriate electron density surrounding the cluster positions. While residues involved in the interface with FOR are reasonably well defined, other parts of the Fd molecules are of poorer quality. In general, the more

distant Fd residues are from FOR, the poorer the quality of the model, and many of these residues could not be definitively modeled. This suggests the Fd molecule "wobbles" around the site of attachment to FOR. Because of this behavior, the refinement of the FOR-Fd complex did not progress as well as for the other FOR structures.

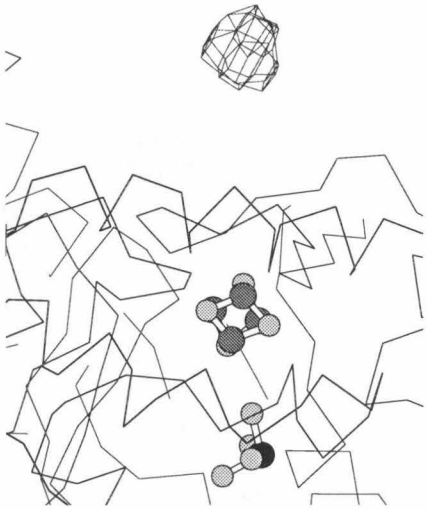


Figure 2.13: Fo–Fc maps contoured at 3σ clearly show strong peaks at identical positions as the peaks in Figure 2.12 for the Fe_4S_4 clusterof ferredoxin.

The refined 2Fo–Fc and Fo–Fc maps around the iron-sulfur cluster clearly showed that it is of Fe_4S_4 type, although it can be quantitatively converted to the 3Fe-form by ferricyanide treatment (Conover *et al.*, 1990). Residues Cys 21 and Cys 48 are close to each other in the model, but the density for possible disulfide bond is not evident. This is in agreement with the experimental observation that only after prolonged exposure to air do the free thiol groups of Cys 21 and Cys 48 form disulfide bond (Gorst *et al.*, 1995) (the Fd sample used in crystallization was in reduced form). As it has been shown by H–1 NMR studies (Calzolai *et al.*, 1995), residue Asp 14 coordinates one of the iron atoms of the Fe_4S_4 cluster with its side chain carboxylate group, which is likely a mono-dentate ligand based on its refined electron density.

Statistics of the refinement of the structures of FOR, FOR-glutarate complex, and FOR-Fd complex are listed in Table 2.4. The coordinations of the structures

of FOR and FOR-glutarate complex have deposited with Protein Data Bank, access code 1B25 and 1B4N, respectively.

Model	Native FOR	FOR-glutarate Complex	FOR-Fd Complex
Resolution	1.85Å	2.40Å	2.15Å
R Factor	17.4%	17.7%	25.3%
Free R	22.0%	24.4%	30.5%
No. Reflections	240450	93750	150222
No. of Residues	2404	2404	2432
Non-H Atoms	19132	19168	20054
Water Molecules	1349	168	0
RMS Deviations			
Bond Distance	0.008Å	0.008Å	0.013Å
Bond Angle	1.37°	1.29°	1.61°
Dihedral	22.9°	25.3°	24.9°
Average B (main)	20.5Å^2	22.8Å^2	27.1Å^2
Average B (side)	23.5Å^2	24.0Å^2	27.3Å^2

Table 2.4: Statistics of structure refinement.

2.4 Structure of FOR

The four subunits of FOR tetramer, which will be referred to as subunits A-D, have very similar structures with each other. The rms deviations between them are 0.126Å to 0.143Å for all the Cα atoms. Some side chains on protein surfaces, most of which are from lysine and arginine residues, have different conformations in different subunits, but no significant differences have been observed in other parts of the structures. Discussions in this chapter will be based on the structure of subunit A and in general apply to all the other molecules.

2.4.1 Quality of the Model

Residues Gly 454 to Glu 461 are disordered and missing from the FOR model. All the other residues, the Fe_4S_4 clusters (Figure 2.14), and the Moco cofactors (Figure 2.15), are in good density. Figure 2.16 shows a 2Fo–Fc map representative of the final model.

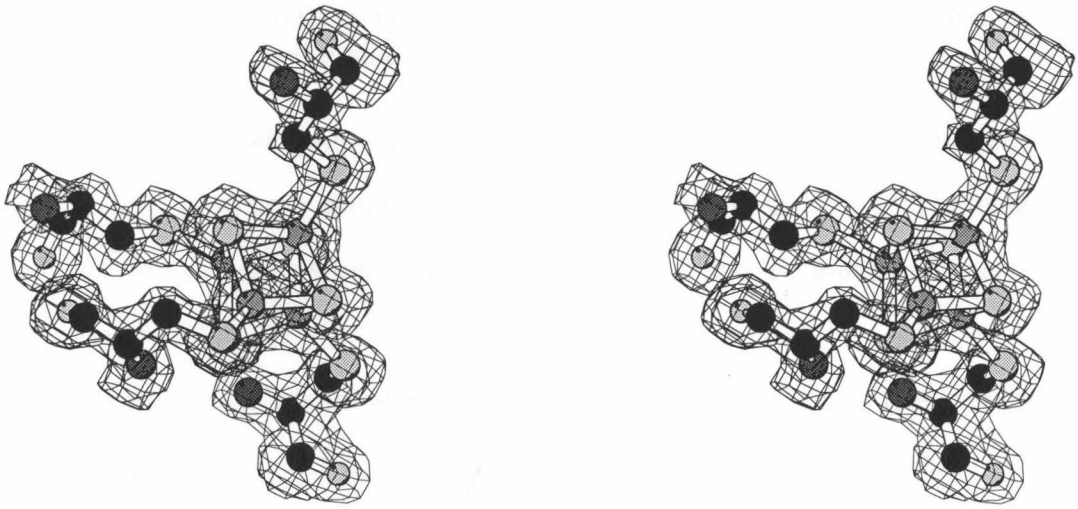


Figure 2.14: The 2Fo–Fc map around one of the Fe_4S_4 clusters and its four cysteine ligands. Contoured at 1.5σ.

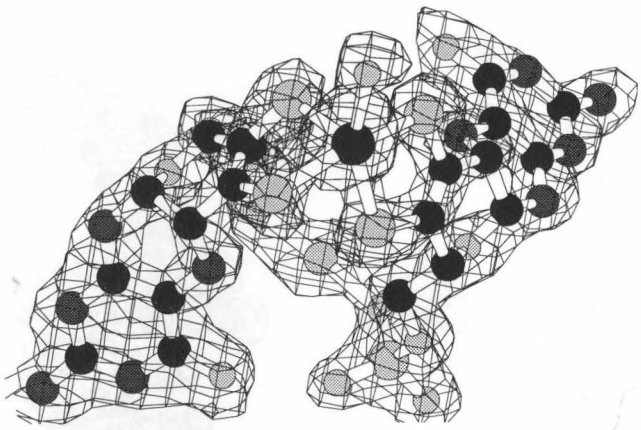


Figure 2.15: The 2Fo–Fc map around one of the Moco's. Contoured at 1.5σ.

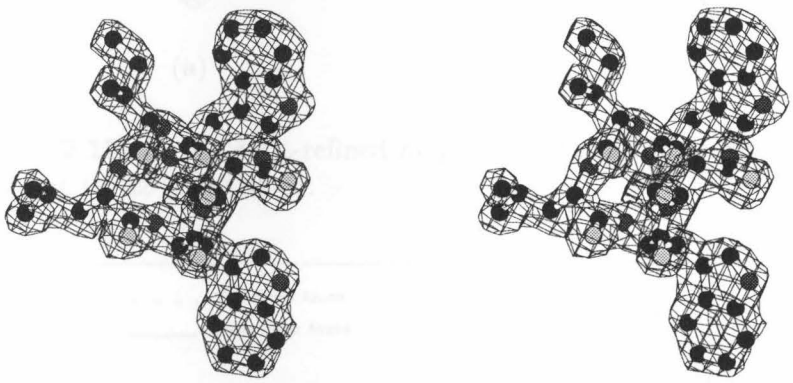
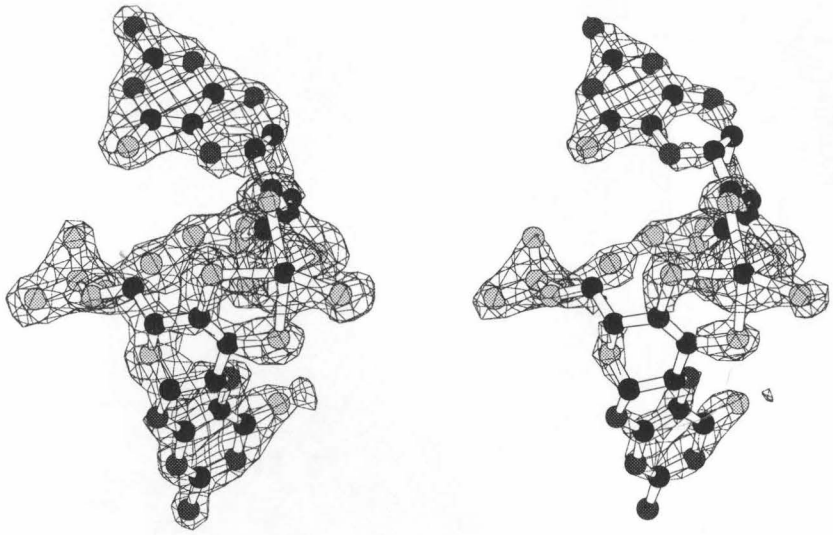


Figure 2.16: Stereoview of the 2Fo–Fc map around residues Trp 40, Ile 41, Leu 42, Trp 43, and Asn 44. Contoured at 1.5σ. This map is representative of the entire refined model.

To validate the model, one of the Moco cofactors was removed and the remaining model was subjected to a simulated annealing refinement at 3000k. The density of the removed Moco clearly showed up in the Fo–Fc and 2Fo–Fc maps and was of good quality (Figure 2.17).

The overall average temperature factor for the model is 22.8Å^2. The average temperature factors are 20.4Å^2 and 25.2Å^2 for the main chain and side chain atoms of the polypeptide chains, respectively, and the average temperature factors for most residues are refined to 15$\sim$25Å^2 (Figure 2.18). The average temperature factor is 12.3Å^2 for atoms in Moco's and Fe_4S_4 clusters. Temperature factors for water molecules range from 10.1Å^2 to 77.0Å^2 and the average value is 36.6Å^2.



(a) (b)

Figure 2.17: Fo–Fc omit-refined maps around one of the Moco's. Contoured at (a) 2σ and (b) 4σ.

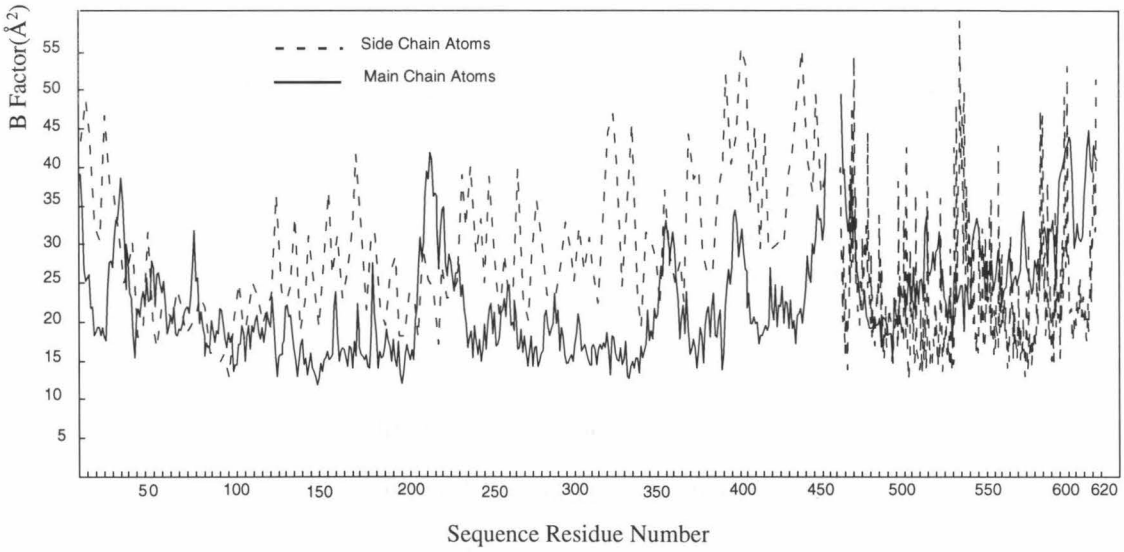


Figure 2.18: Distribution of average B factor as a function of residue number.

Ninety-one percent of the residues are in the most favored regions in the Ramachandran plot (Figure 2.19), as calculated with PROCHECK (Laskowski *et al.*, 1993) . Two residues, Ile 32 and Asp 125, occupy the disallowed region. The former is located in a short loop (residues Phe 31 to Gly 34) between two α-helices, while

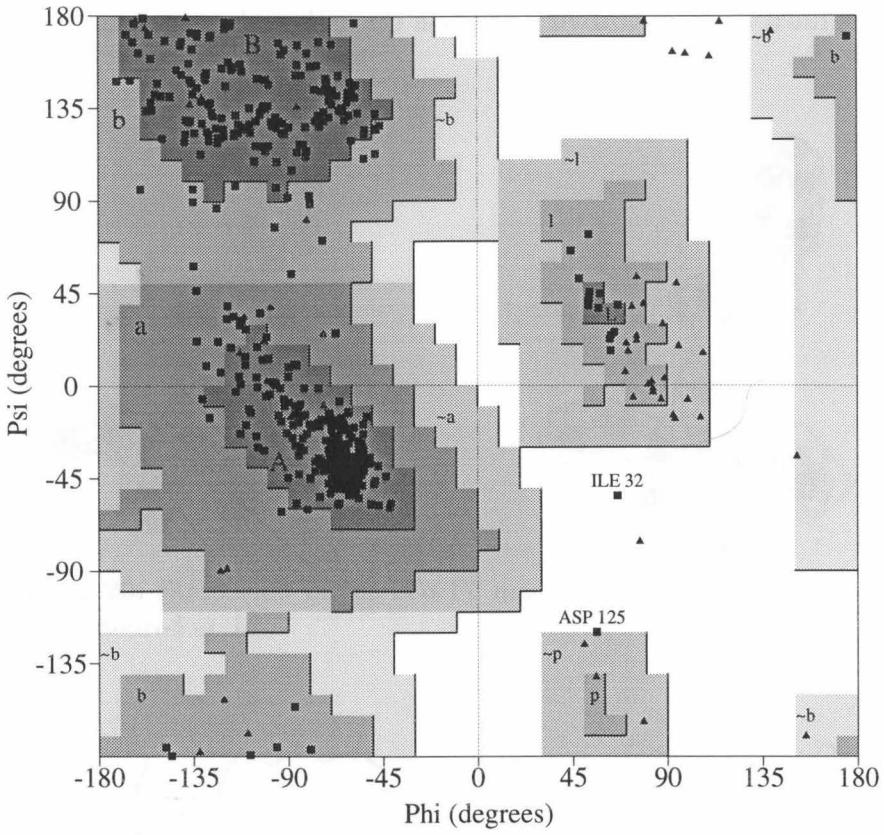


Figure 2.19: Ramachandran plot for subunit A. Triangles represent glycine residues. The shadings, from darkest grey to white, designate the most favored regions, additonal allowed regions, generously allowed regions, and disallowed regions, respectively. A, B, and L are the most favored regions for α-helix, β-strand and left-handed helix, respectively. Residues Ile 32 and Asp 125 are in disallowed regions and are labeled.

the latter is in a short loop (residues Glu 124 to Asn 127) between two β-strands, and both have very well defined electron density (Figure 2.20). The conformations of these two residues are stabilized by multiple hydrogen bonds with nearby peptide groups. The side chain carboxylate group of Asp 125 accepts two hydrogen bonds from Thr 204, one from the side chain OH group and the second one from main chain amide group; the main chain amide group of Asp 125 donates a hydrogen bond to the main chain carbonyl group of Arg 202, while the main chain carbonyl group of Asp 125 accepts a hydrogen bond from side chain OH group of Tyr 106 (Figure 2.21). For residue Ile 32, its main chain amide group donates a hydrogen bond to the carbonyl

group of Ala 28, and its carbonyl group accepts a hydrogen bond from side chain Nϵ of Trp 4.

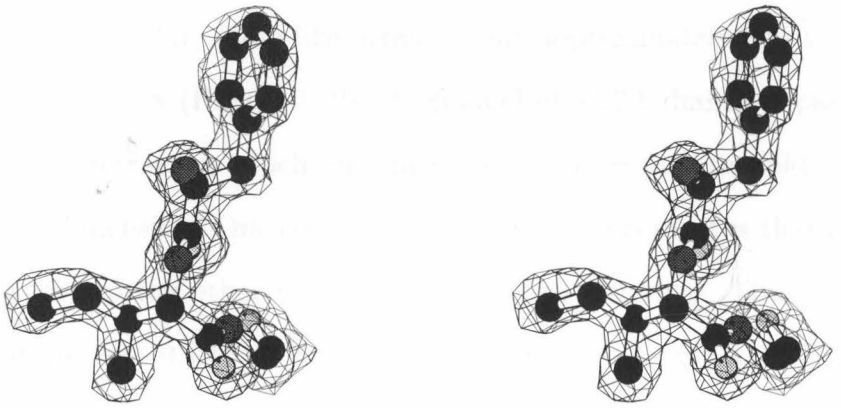


Figure 2.20: Stereoview of the 2Fo–Fc map around residues Phe 31 and Ile 32, contoured at 1σ.

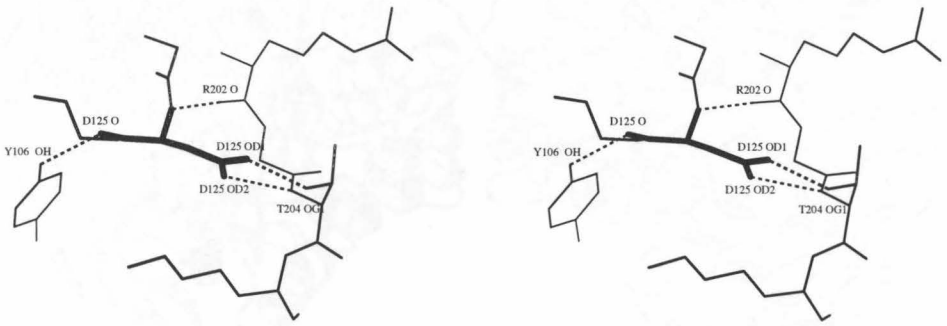


Figure 2.21: The conformation of Asp 125 is stablilized by hydrogen bond interactions with nearby peptide groups.

2.4.2 Overall Structure

The subunits in the FOR tetramer are related by 222 molecular symmetry that generates a relatively flat, plate-like arrangement approximately 115Å on each side with a thickness of 50Å (Figure 2.22). A channel of ~27Å diameter passes through the center of the tetramer, which encompasses the molecular twofold axis oriented along the short dimension. One consequence of this arrangement is that each subunit contacts only two of the other three subunits in the tetramer. Thus, the distances between W atoms are 40Å (between the closest subunits, A and B, or C and D), 63Å

Figure 2.22: FOR tetramer in an asymmetric unit. Viewed approximately along one of the NCS twofold axes. Subunits A, B, C, and D are arranged counter-clockwise starting from the lower left corner.

with the second neighbor (subunits A and D, or B and C), and 74Å (between two subunits across the corner, A and C, or B and D).

2.4.3 Polypeptide Fold

The FOR monomer is approximately spherically shaped with a diameter of 60Å. The overall secondary structure of FOR is predominantly helical, with 43% of the residues in α- or 3_{10}-helices, while β-strands only account for 13% of the residues. In the following discussions, the secondary structures of FOR are prefixed with h, stands for α- and 3_{10}-helices, or s, for β-strands, respectively. There are 32 helices and 15 β-strands in FOR, and are formed by: h1: residues 25–31; h2: 34–45; h3: 96–103; h4: 141–152; h5: 163–166; h6: 187–193; h7: 212–228; h8: 232–238; h9: 239–242 (3_{10}); h10: 243–249; h11: 266–269 (3_{10}); h12: 271–277; h13: 307–313; h14: 315–317 (3_{10}); h15: 322–335; h16: 339–354; h17: 367–378; h18: 382–388; h19: 392–399; h20: 402–404 (3_{10}); h21: 422–430; h22: 443–447; h23: 469–481; h24: 484–487; h25: 493–498; h26: 504–513; h27: 519–540; h28: 547–550 (3_{10}); h29: 555–558; h30: 573–587; h31: 597–602; h32: 608–613; s1: 7–12; s2: 18–22; s3: 58–61; s4: 75–81; s5: 88–94; s6:

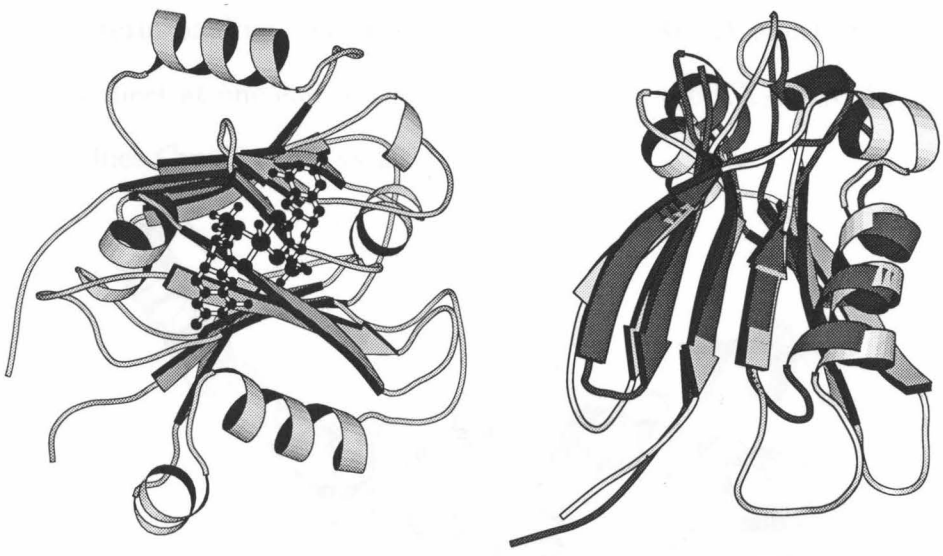
Figure 2.23: Stereoview of FOR molecule. Viewed approximately along the twofold axis that relates the first and second molybdopterins and the first and second halves of domain 1. Domain 1 is in black, domain 2 dark grey, domain 3 light grey. The Moco is represented by ball–and–stick model.

108–112; s7: 119–124; s8: 127–132; s9: 157–160; s10: 174–176; s11: 180–182; s12: 196–202; s13: 278–282; s14: 292–296; s15: 302–305.

As was observed for AOR (Chan *et al.*, 1995), FOR polypeptide chain consists of three domains: the ellipsoid-shaped domain 1 (residues 1–208) that forms the base of the protein, while domains 2 (residues 209–406) and 3 (residues 407–619) are on top of the first domain. The Moco and the Fe_4S_4 cluster are located at the interface of the domains (Figure 2.23).

Domain 1

Domain 1 (Figure 2.24a) consists of two similarly folded halves, formed by residues 1–115 and 116–208, respectively. Both halves are folded into a 6-stranded β-sheet and 3 α-helices. The two β-sheets form the center of domain 1, with α-helices h2 and h4, the two longest in this domain, on either side of the sheets and parallel with the



(a) (b)

Figure 2.24: (a). Domain 1 of FOR and the Moco. Viewed approximately along the twofold axis relates the two halves. (b). The superposition of the first and second (darker) halves of domain 1.

β-strands. The first and second halves are superimposable with an rms deviation of 1.9Å, based on 72 Cα atoms from either half (Figure 2.24b). Domain 1 is also distinctive in that it contains a relatively large portion of β-strands, since 12 (s1-s12) of the 15 β-strands in FOR are found in this domain.

The Moco is located right on top of domain 1. The phosphate groups of Moco form multiple hydrogen bonds and salt bridges with the side chains of residues Lys 75 and Arg 138 (Figure 2.29), and numerous hydrogen bond interactions with the main chain amide groups from domain 1. But this domain is not involved in the binding of the ring systems of molybdopterin moieties. Instead, only domains 2 and 3 provide the peptide groups for the interactions with the ring systems (see below).

Domain 2

Domain 2 contains 14 helices including the long α-helix (h7) that connects domain 1 and domain 2. The loop area between helices h15 and h16 contains the first molybdopterin-binding sequence (residues 333–338). A short, three stranded anti-parallel β-sheet at one end of the domain contains the CX_2CX_3C Fe_4S_4-binding sequence (residues Cys 284 to Cys 291) (Figure 2.25).

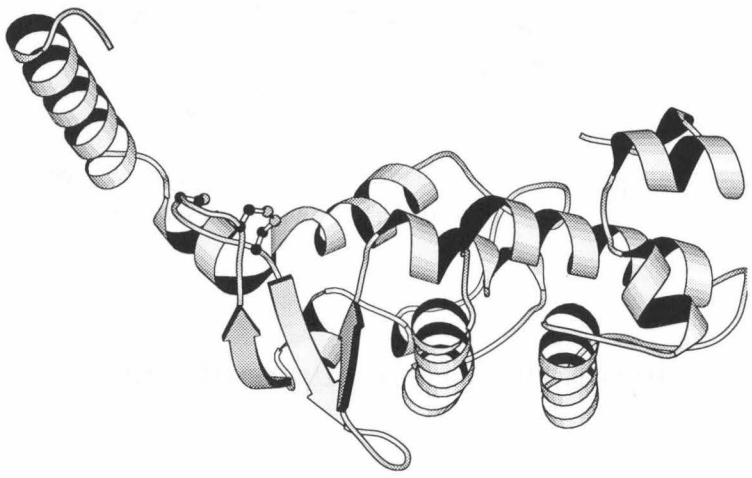


Figure 2.25: Domain 2 of FOR. The cysteine residues coordinating Fe_4S_4 cluster are shown as ball-and-stick models.

Domain 3

Domain 3 is exclusively helical and consists of eleven α-helices (h21 to h27 and h29 to h32) and one 3_{10}-helix (h28) (Figure 2.26). At the center of this domain are the α-helices h23 and h24, followed by the second molybdopterin-binding sequence (residues 486–491). Residue Cys 491 also provides the fourth ligand of Fe_4S_4 cluster. The longest α-helix in FOR, h27, formed by residues 519–540, is positioned perpendicular to h23 and h24.

Domains 2 and 3 exhibit similarities in the polypeptide fold. Helices h12, h13, h15, h16 and h17 in domain 2 correspond to helices h21, h22, h24, h25 and h26 in domain 3, respectively; the relative positions and topological connections of these helices are conserved between the two domains, although the structural similarity is sufficiently weak that this correspondence is not picked up by automated searches.

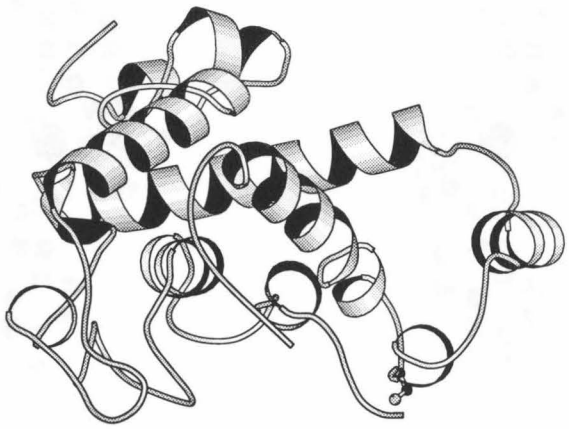


Figure 2.26: Domain 3 of FOR. Residue Cys 491, which coordinates Fe_4S_4 cluster, is shown as ball-and-stick model.

2.4.4 Structure of Moco and Its Environment

Structure of Moco

The Moco in FOR consists of two molybdopterin molecules and one tungsten atom (Figure 2.27). The two molybdopterins are designated as the first and the

second molybdopterin, respectively, according to the order their binding motifs occur in FOR sequence. Like those found in other Moco-containing protein structures (Rees *et al.*, 1997), the molybdopterins in FOR are tricyclic, formed from the fusion of pterin and pyran ring systems. The tungsten atom is coordinated by all four dithiolene sulfur atoms present in the two molybdopterins, with a range of W–S distances of 2.41–2.53Å (average 2.49Å) (Table 2.5). As in AOR, no protein side chain coordinates the tungsten atom. The two molybdopterins are further linked to each other by a magnesium ion, which bridges the phosphate groups of each molybdopterin. The two molybdopterins are approximately related by a two-fold rotation about an axis that passes through both the tungsten and the magnesium, and which is approximately coincident with the twofold axis relating the two halves of domain 1.

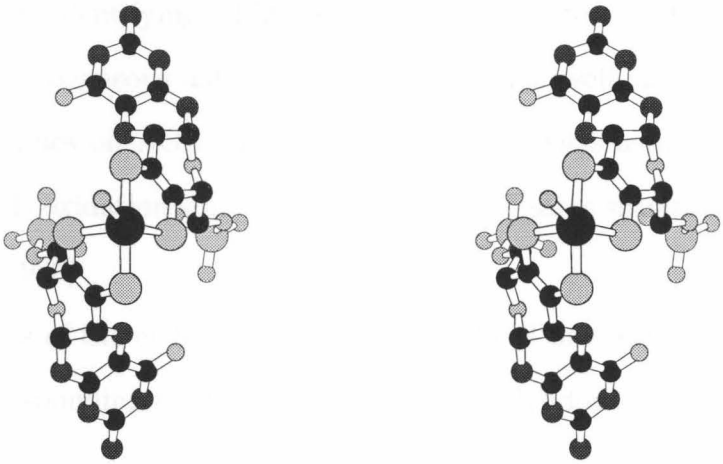


Figure 2.27: Stereoview of the Moco of FOR.

Molybdopterin	S atom	Subunit A	Subunit B	Subunit C	Subunit D
1	S1	2.53Å	2.44Å	2.51Å	2.46Å
1	S2	2.52Å	2.49Å	2.50Å	2.47Å
2	S1	2.49Å	2.50Å	2.53Å	2.50Å
2	S2	2.41Å	2.51Å	2.52Å	2.44Å

Table 2.5: W–S distances in Moco's.

In addition to the four sulfur atoms, one other ligand was found to coordinate

the W in *Pf* FOR. Based on its electron density, this ligand was assumed to be an oxygen atom. This group refines to a distance of 2.10Å from the W (range from 2.09 to 2.10Å) and occupies equivalent positions in all the four monomers. In the refinement, an equilibrium bond distance of 2.10Å and a bond strength of 500kcal/mol·Å^2 were used for the W–O bond (see Appendix B). The equilibrium W–O distance was determined by refining the structure with no restraints on this bond. Caution must be taken in interpreting the W–O distance, however, since the "ripples" around a W atom caused by series termination effects are predicted to occur at this distance (Schindelin *et al.*, 1997). For comparison, an EXAFS study on *Pf* AOR indicated that an oxo group coordinates the tungsten atom, with a W–O distance of 1.7Å, and an additional O or N atom is possibly present at 2.1Å (George *et al.*, 1992). A potential difficulty in identifying additional coordination sites on the tungsten may be caused by the heterogeneous nature of the tungsten site itself, as was found for AOR. Spectroscopic studies on AOR show that the W site exists as a mixture of W(IV), W(V), and W(VI) oxidation states, and it is likely the same situation occurs for FOR (Koehler *et al.*, 1996).

The structures of the eight molybdopterins in FOR are very similar and fall into two classes, corresponding to the first and second molybdopterins, respectively (Fig-

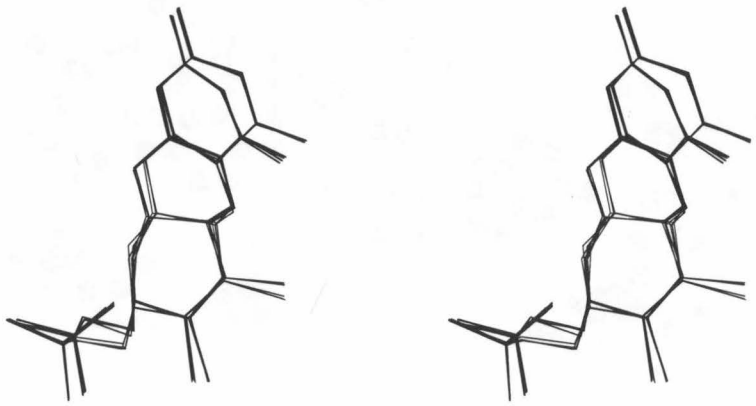


Figure 2.28: Stereoview of the superposition of all the eight molybdopterins in FOR.

ure 2.28). The rms deviation is 0.1Å between the non-hydrogen atoms of any two molybdopterins from the same class, and about 0.5Å between any two molybdopterins from different classes.

Moco-Protein Interactions

The two molybdopterins in each subunit are sandwiched between domain 1 on one side and domains 2 and 3 on the other. The two phosphate groups, one from each molybdopterin, form salt bridges with the side chains of Lys 75, Arg 180, and Lys 438, and hydrogen bonds with numerous main chain NH groups. With the exception of the Lys 438 side chain, all the protein groups involved in the binding of the phosphate groups are from domain 1, while only domains 2 and 3 are involved in the interactions with the ring systems of the two molybdopterins. The ring systems of the two molybdopterins each interact with a binding motif corresponding to residues Asp 333 to Asp 338 and Glu 486 to Cys 491, from domains 2 and 3, respectively.

The pattern of hydrogen bonds between the ring systems of the two molybdopterins and their respective binding sequences are virtually identical (Figure 2.29).

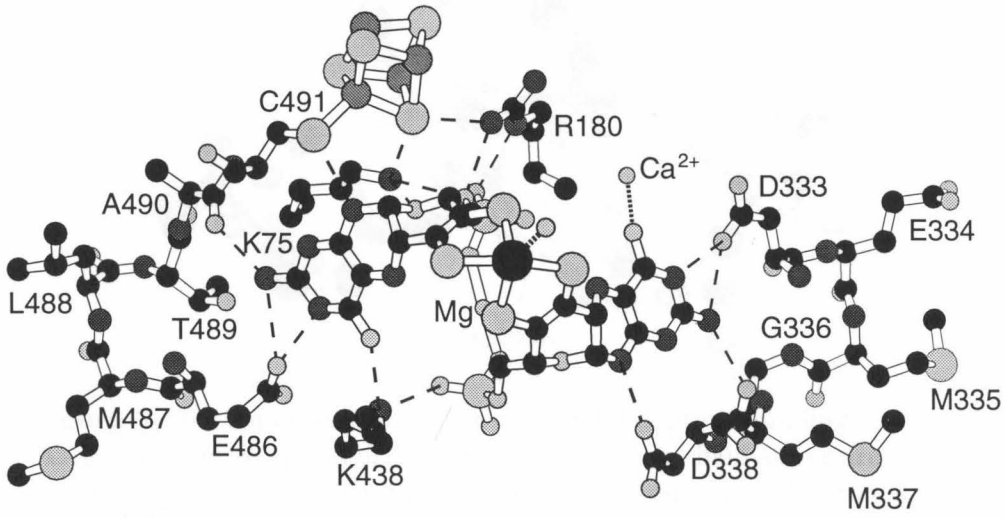


Figure 2.29: Moco and Fe_4S_4 cluster and surrounding residues. The first and second molybdopterin binding motifs are at the right and left, respectively. Hydrogen bond interactions are represented as dashed lines.

The molybdopterin N3 atom donates a hydrogen bond to the side chain carboxylate group of the first residue of the molybdopterin binding motif (Asp 333 or Glu 486); the amino group of molybdopterin donates two hydrogen bonds, one to the carbonyl oxygen of the fifth residue and one to the side chain carboxylate group of the first residue; and atom N8 donates a hydrogen bond to the side chain of the last residue (Oδ1 of Asp 338 or Sγ of Cys 491). In addition, the carbonyl oxygen atom of the second molybdopterin accepts a hydrogen bond from the side chain of Lys 438, while the pyran ring oxygen of the second molybdopterin is hydrogen-bonded to the side chain of Lys 75. Please refer to Figure 2.1 on page 57 for the numbering of the molybdopterin atoms. Information about these interactions is listed in Table 2.6. The calcium ion identified in elemental analyses was found to provide additional interactions between the first molybdopterin and the protein (see Chapter 2.4.7).

In addition to the similar interaction patterns with their respective molybdopterins, the two Moco binding motifs in FOR have very similar conformations. Their Cα atoms can be superimposed with an rms deviation of 0.47Å (Figure 2.30).

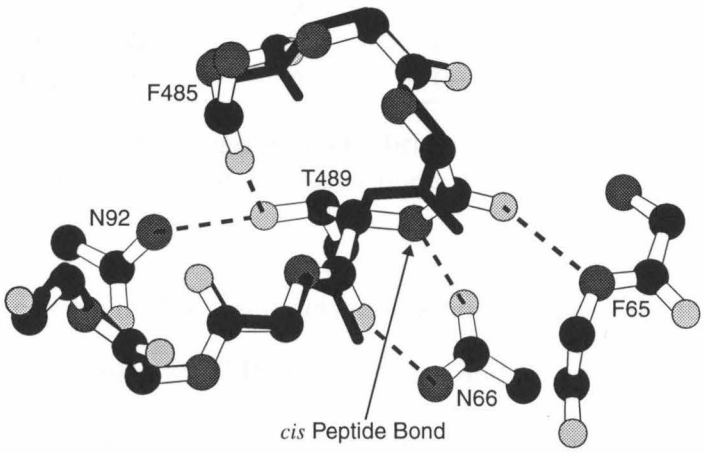


Figure 2.30: Superposition of the two molybdopterin-binding motifs in FOR. With the exception of Thr 489, only the main chain atoms are shown. The second molybdopterin-binding motif and some peptide groups near the *cis* peptide bond are shown in a ball-and-stick model, while the first molybdopterin binding motif with a corresponding *trans* peptide bond is in black. Hydrogen bonds stabilizing the *cis* peptide bond are indicated by dashed lines.

Interactions Involving Moco Ring Systems					
From		To		Properties	
Residue	Atom	Molybdopterin	Atom	Type	Distance
Asp 333	$O\delta 2$	1	N3	D	2.73Å
Asp 333	$O\delta 2$	1	NH_2	D	2.86Å
Met 337	O	1	NH_2	D	2.96Å
Asp 338	$O\delta 1$	1	N8	D	2.74Å
Glu 486	$O\epsilon 1$	2	N3	D	2.66Å
Glu 486	$O\epsilon 1$	2	NH_2	D	3.24Å
Ala 490	O	2	NH_2	D	2.72Å
Cys 491	$S\gamma$	2	N8	D	3.15Å
Lys 75	$N\zeta$	2	O	A	3.04Å
Lys 438	$N\zeta$	2	O4	A	2.94Å

Interactions Involving Phosphate Groups					
From		To		Properties	
Residue	Atom	Molybdopterin	Atom	Type	Distance
Asn 92	N	1	$O\gamma 3$	A	2.77Å
Gly 183	N	1	$O\gamma 2$	A	2.80Å
Arg 184	N	1	$O\gamma 2$	A	3.09Å
His 436	$N\epsilon 2$	1	$O\gamma 4$	A	3.07Å
Lys 438	$N\zeta$	1	$O\gamma 3$	A	2.75Å
Lys 75	$N\zeta$	2	$O\gamma 4$	A	3.12Å
Gly 94	N	2	$O\gamma 2$	A	2.75Å
Arg 180	$N\eta 2$	2	$O\gamma 4$	A	2.98Å
Arg 180	$N\epsilon$	2	$O\gamma 2$	A	2.90Å
Ala 181	N	2	$O\gamma 3$	A	3.07Å

Table 2.6: Hydrogen bond and salt bridge interactions between Moco and peptide groups. D for hydrogen bond donor, A for hydrogen bond acceptor.

The symmetric protein-Moco interaction was not observed in the structures of *Rh. spaeroides* and *Rh. capsulatus* DMSO reductases and *E. coli* formate dehydrogenase H which also contain two molybdopterins. In these proteins, as well as in chicken liver sulfite oxidase and *D. gigas* Mop, the Moco's are bound by residues scattered in their amino acid sequences.

In AOR, both molybdopterin binding motifs have the sequence profile DXXGL(C/D)X, where X is any amino acid. While the first motif is conserved in FOR (starting with Asp 333), the second one is replaced by the sequence EML-

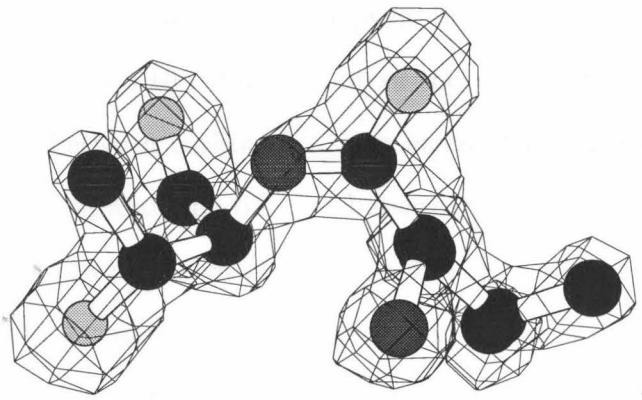


Figure 2.31: The *cis* peptide bond between Leu 488 and Thr 489. Superimposed with the refined 2Fo–Fc map contoured at 1σ.

TAC, starting with Glu 486. The most dramatic change is the substitution of the Gly (which adopts a left-handed α-helical polypeptide conformation) by Thr 489. As a consequence of this substitution, the peptide bond between Leu 488 and Thr 489 adopts an unusual *cis* peptide bond configuration. This *cis* peptide bond conformation is stabilized by multiple hydrogen bond interactions with nearby protein groups (Figure 2.30). To refine this *cis* peptide bond, a new amino acid type for X-PLOR was made in which the C–N–Cα–C dihedral is 0°. This *cis* peptide bond has very well defined density (Figure 2.31).

Theoretical calculations showed that there is an energy difference of about 2.8kcal/mol between *cis* and *trans* conformations of non-proline peptide bonds (Stewart *et al.*, 1990). A frequency of about 0.1% was predicted for the occurrence of *cis* peptide bond in denatured polypeptide, but a survey of the Brookhaven Protein Data Bank found that only 0.05% of 31005 amide bonds are *cis* (Stewart *et al.*, 1990). Interestingly, non-proline *cis* peptide bonds are often found near active sites of proteins (Rees *et al.*, 1981). The structural basis of the proximity of *cis* peptide bonds to protein active sites is unknown, but it was suggested that these *cis* peptide bonds might be required for the precise positioning of the protein groups around the active sites, even though they introduce strain (Herzberg & Moult, 1991).

In the case of FOR structure, the *cis* peptide bond conformation is stabilized by the hydrogen bonds between the amide group of the peptide bond and the side chain Oδ1 atom of Asn 66, between carbonyl oxygen atom of Leu 488 and the main chain nitrogen atoms of Pro 64 and Phe 65, and between the side chain OH group of Thr 489 and the carbonyl oxygen atom of Phe 485 and the side chain Nδ2 atom of Asn 92 (Figure 2.30). The formation of these hydrogen bonds is only possible when the peptide bond between Leu 488 and Thr 489 adopts the *cis* conformation. Corresponding hydrogen bonds do not exist for the first Moco-binding domain. Furthermore, if the peptide bond between Leu 488 and Thr 489 adopted a *trans* conformation, the side chain of Thr 489 would be in steric clash with the side chain of Asn 66. The DXXGLC/DX motif was not found in Moco-containing proteins from other families.

2.4.5 Fe_4S_4 Cluster and Its Environment

The Fe_4S_4 cluster is located approximately 10Å from the W atom (Figure 2.29), and is buried approximately 6Å below the protein surface. As anticipated, it is coordinated by the Sγ atoms of residues Cys 284, Cys 287, Cys 291, and Cys 491. The Fe_4S_4 cluster is surrounded by hydrophobic side chains from residues Trp 235, Met 289, Pro 290, Leu 493, and Pro 494. The cluster is buried inside the protein structure, but a water molecule is found in the environment of the cluster and is hydrogen-bonded to one of the inorganic sulfurs and the carbonyl oxygen of the residue Pro 285 (Figure 2.32). The cluster is linked to the second molybdopterin through several hydrogen bond interactions (Figure 2.29 and Table 2.6): the Sγ atom of Cys 491 accepts a hydrogen bond from N8 of the molybdopterin; one of the inorganic sulfurs accepts two hydrogen bonds, the first is from the side chain of Lys 75, which also forms a hydrogen bond to the pyran ring oxygen and a salt bridge to the phosphate group; the second is from the side chain of Arg 180, which also forms two salt bridges with the phosphate group. The proximity of this pterin to the Fe_4S_4

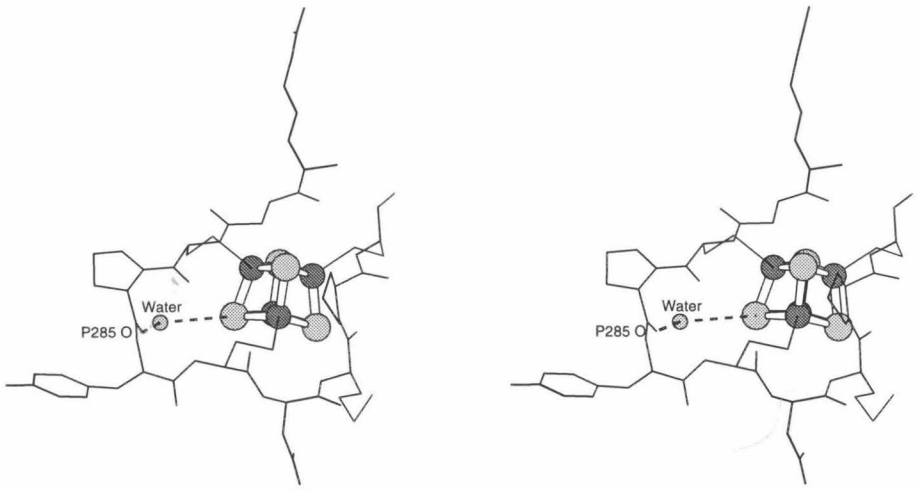


Figure 2.32: The Fe_4S_4 cluster and some of the surrounding residues (Arg 282, Gly 283, Cys 284, Pro 285, Tyr 286, Cys 287, Asn 288, Met 289, Pro 290, and Cys 291.) A water molecule is found in its environment.

cluster and the intimate interaction between them suggest that the molybdopterin may play an important role in electron transfer.

The refined structures of Fe_4S_4 clusters are very similar to those of the Fe_4S_4 clusters found in ferredoxins and other proteins. All the Fe–S distances are in the 2.25–2.33Å range with an average of 2.29Å (Table 2.7).

		Subunit A	Subunit B	Subunit C	Subunit D
Fe1	S1	2.25Å	2.29Å	2.28Å	2.27Å
	S2	2.28Å	2.28Å	2.28Å	2.29Å
	S3	2.30Å	2.30Å	2.31Å	2.31Å
Fe2	S1	2.30Å	2.31Å	2.28Å	2.30Å
	S2	2.24Å	2.29Å	2.25Å	2.26Å
	S4	2.30Å	2.27Å	2.31Å	2.31Å
Fe3	S1	2.29Å	2.31Å	2.34Å	2.33Å
	S3	2.29Å	2.30Å	2.29Å	2.32Å
	S4	2.29Å	2.32Å	2.29Å	2.30Å
Fe4	S2	2.28Å	2.25Å	2.30Å	2.27Å
	S3	2.31Å	2.32Å	2.30Å	2.30Å
	S4	2.26Å	2.31Å	2.26Å	2.28Å

Table 2.7: Fe–S distances in Fe_4S_4 clusters.

2.4.6 Active Site Cavity

A channel connecting the tungsten site and the protein surface was observed at the interface of domains 2 and 3 in the AOR structure that was porposed to permit substrates and products to enter and leave the active site, respectively (Chan *et al.*, 1995). In FOR, a cavity, rather than an open channel, is found at the same position. The volume of this cavity is calculated as 1500Å^3 by the program VOIDOO (Kleywegt & Jones, 1994), using a 1.2Å probe radius (Hubbard & Argos, 1995). This cavity consists of two distinctive parts: a large chamber at the bottom and a much narrower channel leading toward the protein surface (Figure 2.33). The bottom chamber is lined by the Moco and the side chains of Thr 240, Tyr 307, Glu 308, Tyr 416, His 437, Trp 441, Arg 481, Arg 492, Leu 493, and Val 496. The channel leading to the protein surface is lined primarily by hydrophobic residues, including Phe 234, Trp 235, Ala 242, Ala 243, Pro 451, Ile 452, and Glu 497. Residues Pro 451 and Ile 452, which are followed by a disordered region of the polypeptide chain (residues Gly 454 to Glu 461), along with the Phe 234 side chain

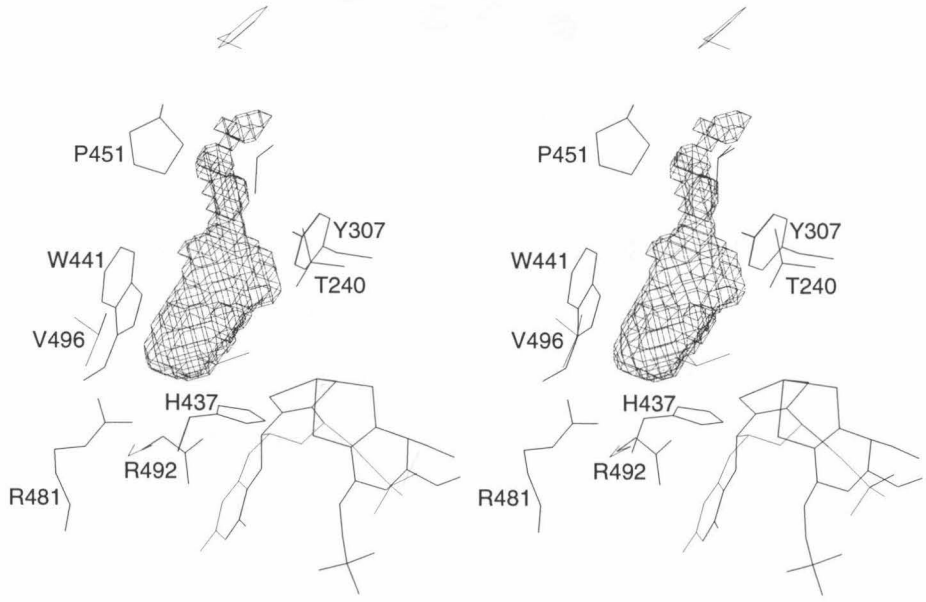


Figure 2.33: Stereoview of the active site cavity, defined by VOIDOO, and some of the surrounding residues.

seal the cavity from the environment.

2.4.7 Calcium Site

In the process of refinement, a relatively electron dense feature was found in all four subunits near the O4 of one of the molybdopterins that refined to very low B factors for a water molecule. The density level and the geometry suggest that it is a cation, most likely calcium or potassium, with the former consistent with the elemental analysis of FOR (Roy *et al.*, 1998). Calcium ions were modeled into each site, and they all refined to reasonable B factors ranging from 21.5Å^2 to 25.6Å^2.

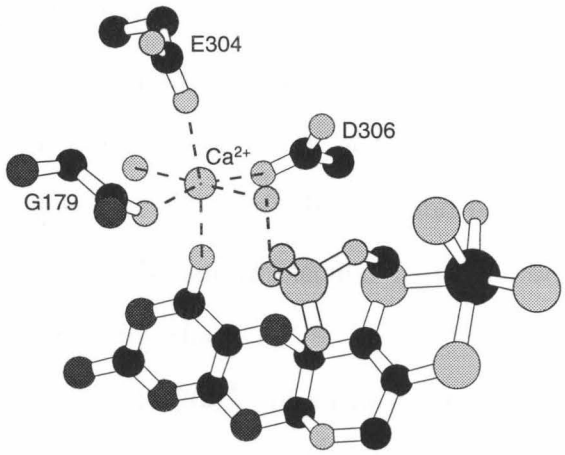


Figure 2.34: The calcium site found near the W atom, and the ligands.

The calcium site is 7.8Å from tungsten atom (Figure 2.29) and is octahedrally coordinated by the carbonyl oxygen atoms of the first molybdopterin and residues Gly 179, the side chain carboxylate groups of Glu 304 and Asp 306 (both monodentate), and two water molecules (Figure 2.34). One of the water molecules is also hydrogen-bonded to the phosphate group of the second molybdopterin. With loose geometry restraints (100 or 200kcal/mol·Å^2, see Appendix B), the Ca–O distance refined to 2.1Å for the molybdopterin oxygen atom, and 2.3–2.5Å for water and peptide oxygens. This site is inaccessible to substrate and is not in the putative

electron transfer pathway, so its function is likely to be structural despite its proximity to the active site.

2.4.8 Inter-Subunit Interfaces

Two distinct types of subunit-subunit interfaces are present in FOR which differ significantly in the nature of the complementing interactions.

Interface I: This subunit-subunit interface buries approximately 710Å^2 of accessible surface area between domains 1 of monomers A and D and monomers B and C. Van der Waals contacts between nonpolar residues and hydrophobic interactions between the following side chains from both monomers dominate the inter-subunit contacts: Tyr 120, Tyr 122, Leu 131, Leu 148, Ile 151, and His 152. Two intermolecular hydrogen bonds are found at the edge of the interface between the carbonyl oxygen of Ile 151 and the amide group of Gly 135, and between the carbonyl oxygen of Ser 132 and Nδ1 of His 152. No intermolecular hydrogen bonds or salt bridges are observed in the center of this interface.

Interface II: Approximately 700Å^2 of surface area are buried at this interface between the second domains of monomers A and B and monomers C and D. At the center of this interface, an extensive hydrogen bond network mediated by four buried water molecules exists that involves the carbonyl groups of Leu 296 and Met 277 and the side chain Nδ2 atom of Asn 323 from both monomers. Near the surface is another hydrogen bond network mediated by three water molecules and the side chains of Arg 267, Arg 270, and Thr 273 from both monomers. A water molecule on protein surface forms hydrogen bonds with the side chain Oγ atom of Ser 268 of one monomer and the side chain NH2 atom of Arg 237 of another monomer. Several direct intermolecular protein-protein hydrogen bonds are also observed in this interface. The carbonyl oxygen atom of Lys 278 from one monomer accepts two hydrogen bonds from another monomer, one from the Nζ atom of Lys 321, and the second one from the Nδ2

atom of Asn 323. The carbonyl oxygen atom of Gly 276 from one monomer accepts a hydrogen bond from the amide nitrogen atom of Asn 323 from another monomer. The side chain Oγ atom of Ser 268 donates a hydrogen bond to the carbonyl oxygen atom of Tyr 272 of another monomer.

The distinctive natures of the types of contacts at the two types of subunit-subunit interfaces in FOR are rather typical of oligomeric proteins; a recent survey of these interfaces demonstrated that only one-third of the interfaces have predominantly apolar character, with the remainder often exhibiting significant polar character, including the participation of bridging water molecules (Larsen *et al.*, 1998).

The area buried in the two intermolecular interfaces accounts for 7.4% of the total accessible surface of an FOR monomer, which is about 19000Å^2. This percentage is typical for specific protein-protein interactions (Janin & Chothia, 1990), but is at the low end for intersubunit contacts in oligomeric proteins. A survey of 23 oligomeric proteins showed that the areas buried in intersubunit interfaces account for between 9% (for superoxide dismutase) and 40% (for catalse) (Janin *et al.*, 1988). The relatively small intersubunit interfaces in FOR might indicate that the formation of the FOR tetramer is not essential for the stability of this protein (Janin *et al.*, 1988).

2.5 Structural Comparison with *Pf* AOR

FOR and AOR monomers have very similar folds and can be superimposed with an rms deviation of 1.5Å, based on 576 Cα atoms from 15 segments of both molecules (Figure 2.35). This rms deviation is consistent with the estimated value (1.2Å) expected from the degree of sequence identity (40%) between these two proteins (Chothia & Lesk, 1986). In agreement with the observation that domain 1 is most highly conserved between AOR and FOR (Kletzin *et al.*, 1995), the Cα atoms of the 208 residues which comprise this domain in FOR can be superimposed on the corresponding Cα atoms in AOR with an rms deviation of 1.1Å. The most significant structural difference between AOR and FOR occurs in the region between residues Ile 447 and Ile 464 of FOR. In both structures, this sequence is located between two α-helices (h22 and h23 in FOR), and is longer in FOR due to an insertion of five residues relative to AOR. In FOR, this sequence covers the opening of the active site cavity, with residues Pro 451 and Ile 452 sealing the opening of the cavity. Residues Gly 454 to Glu 461 appear mobile and are not visible in the electron density map, although the corresponding residues are well defined in AOR. This region may act

Figure 2.35: Superposition of an FOR monomer (black line) with an AOR monomer.

as a "lid" for the cavity that could modulate accessibility of the FOR active site to substrates.

Despite the strong similarities in amino acid sequences, AOR and FOR have different quaternary structures. While FOR is a tetramer, AOR exists as a dimer both in solution and in the crystal form. A mononuclear iron site was found in AOR at the center of the dimerization interface, coordinated by Glu 332 (bi-dentate) and His 383 from both monomers. Extensive hydrogen bond interactions were also found at the inter-subunit interface (Chan *et al.*, 1995). In contrast, the iron site is not present in FOR, and the Glu and His residues are replaced with Val 327 and Lys 371, respectively.

In addition to these specific substitutions between AOR and FOR, in general, different parts of the protein surfaces of AOR and FOR are involved in inter-subunit interactions. There are some overlaps between the second inter-subunit interface of FOR and the inter-subunit interface of AOR, but none of the residues in the first inter-subunit interface of FOR are involved in the subunit-subunit interactions in AOR (Figure 2.2).

Differences are also found in the active site cavities of FOR and AOR. In FOR, the Moco is located at one end of the bottom chamber of the active site cavity (see Chapter 2.4.6). Residues Thr 240, Asp 306, Tyr 307, Glu 308, Tyr 416, and His 437 are closest to the tungsten atom and, together with Moco, form half of the chamber. These residues have very similar conformations in AOR and FOR, and are highly conserved in AOR-like enzymes with known sequences. In fact, Thr 240, Tyr 307, and Glu 308 are conserved in all the AOR-like molecules with known sequences; Residue Tyr 416 has its side chain pointed to the tungsten site and very well positioned to donate a hydrogen bond to a bound substrate. This residue is also conserved in most of these molecules, but is replaced by Arg in *Pf* GAPOR and by His in two AOR-like molecules from *Archaeoglobus fulgidus*. Both substitutions are by similar

amino acids capable of donating hydrogen bonds. His 437 is conserved in all these molecules but *Pf*GAPOR, in which it is replaced by a similar Asn; Asp 306 is replaced by a similar Glu residue in some of these molecules (Figure 2.2). The proximity of these residues to the tungsten site and the fact that they are very conserved suggest that the immediate environment of the tungsten atoms in FOR, AOR, and other members in this family are highly conserved and that these enzymes may have very similar enzymatic mechanisms, and that the residues near the tungsten atom may be important to the functions of AOR-family tungstoenzymes.

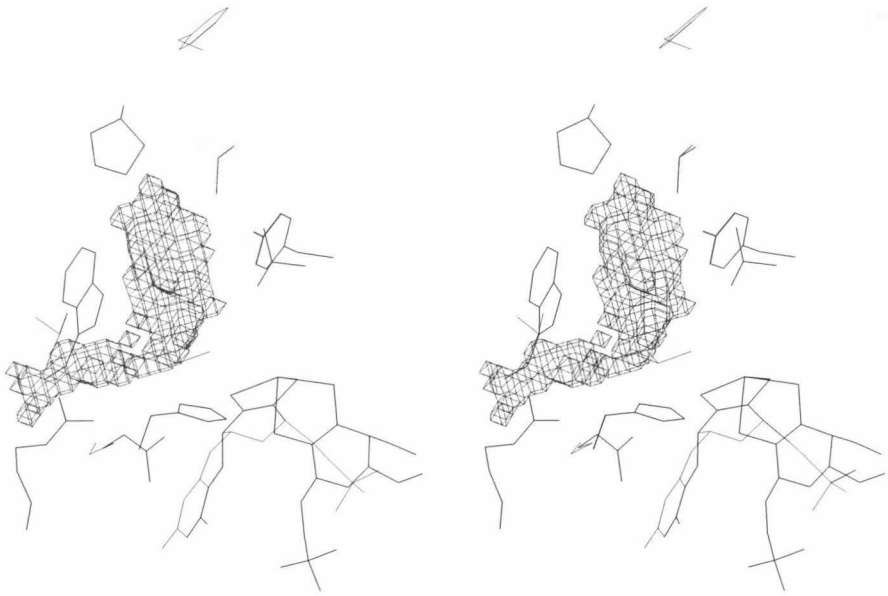


Figure 2.36: Stereoview of the active site cavity of AOR superimposed with **FOR** model. An extra pocket is obvious compared with the cavity of FOR (Figure 2.33) that is filled by Arg 481, Arg 492 and other residues in FOR structure.

Residues lining the opposite end of Moco across the bottom chamber of the active site cavity of FOR generally are not conserved. Structures at this end of the cavities of FOR and AOR are very different. Calculations with the program VOIDOO (Kleywegt & Jones, 1994) suggested that part of the AOR cavity, a hydrophobic pocket formed by the side chains of Ile 449, Tyr 452, Ser 455, Tyr 461, Phe 481, Leu 484, Thr 485, Ile 488, Leu 495, Thr 498, and Phe 499, is filled by the side chains

of Arg 481 and Arg 492 in FOR (Figure 2.36), which replace smaller residues Thr 485 and Leu 495 in AOR, respectively. An insertion of one residue between Ile 480 and Gly 483 in FOR also helps reducing the volume of FOR cavity in this area by shifting the main chain atoms toward the center of the cavity. The substitutions of Thr 485 and Leu 495 in AOR by Arg 481 and Arg 492 also increase the polarity of the FOR cavity. The implications for the specificity of FOR by these differences will be discussed later in Chapter 2.6.

Residues lining the channel leading to protein surfaces are also different in FOR and AOR and are not conserved in other AOR-family proteins. This results in a wider channel in AOR that is open to the environment, while the channel in FOR is narrower and sealed from the environment.

2.6 Interactions Between FOR and Glutarate

In native FOR structure, the active site cavity is occupied by uninterpretable density, with the highest peak near the side chains of Arg 481 and Arg 492. The density was approximately the size of a citrate molecule, which is present in the crystallization solution. Since glutaric dialdehyde has the lowest K_m for any characterized substrate, FOR was crystallized in the presence of the oxidation product, glutarate. In the crystal structure of the FOR-glutarate complex, a glutarate molecule clearly shows up in the cavities of each monomer in both the 2Fo–Fc and Fo–Fc maps (Figure 2.37), and refines to good geometry and reasonable B factors (average 26Å^2). The structure of FOR molecules in the FOR-glutarate complex is very similar to the native FOR crystal structure. No significant conformational changes were observed.

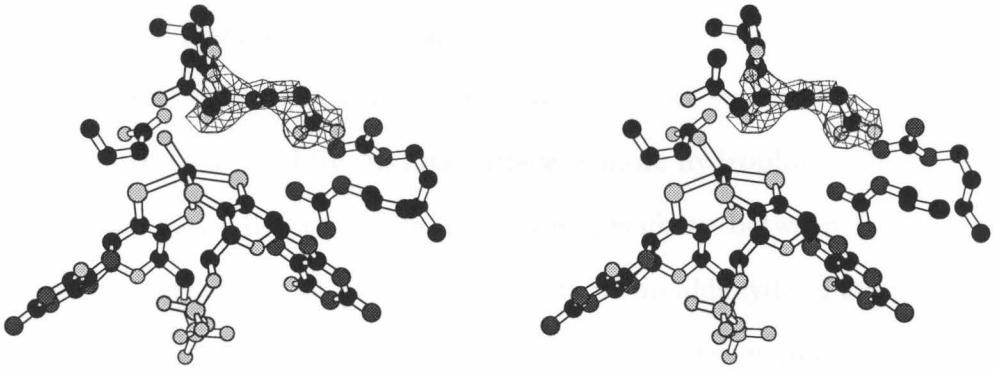


Figure 2.37: Stereoview of the 2Fo–Fc map around one of the glutarate molecules. Contoured at 1σ.

One of carboxylate groups of the glutarate is located near the tungsten site, within hydrogen bonding distance of the side chains of Glu 308, Tyr 416, and His 437, and forms a hydrogen bond with the side chain carboxylate group of Glu 308. The electron density of this carboxylate group of glutarate is not very well defined. The oxygen ligand coordinated to the tungsten in the native FOR structure is not evident in the glutarate complex. The second carboxylate group of glutarate is located at the opposite end of Moco across the active site cavity of FOR. It is anchored to the

protein through electrostatic interactions with the side chains of Arg 481 and Arg 492 (Figure 2.38).

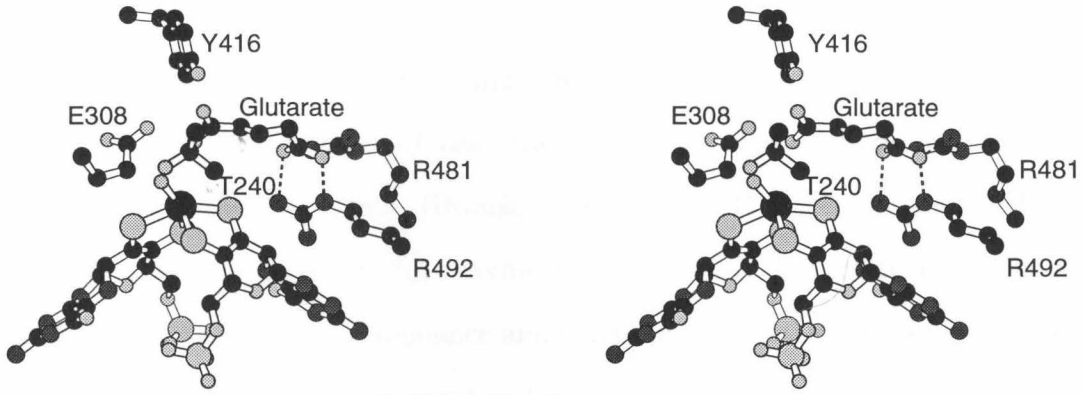


Figure 2.38: Stereoview of the bound glutarate at the active site. Dashed lines indicate hydrogen bond interactions between FOR and glutarate.

As described in Chapter 2.5, The structures of FOR and AOR are very different in the area around residues Arg 481 and Arg 492 of FOR, and the residues generally are not conserved. More specifically, the cavity in FOR structure is smaller in this area compared to that of AOR, and its surface is more hydrophilic. These variations may contribute to the differences in substrate specificity between FOR and AOR. The physiological substrate of FOR is thought to be an aldehyde of the size similar to that of glutaric dialdehyde, which can obviously fit into the small cavity of FOR. The true substrate for this enzyme may have a functional group that can be specifically recognized by Arg 481 and/or Arg 492 by accepting hydrogen bonds or salt bridges from their side chains, as shown in the structure of FOR-glutarate complex. The hydrophobic pocket in AOR, which is filled by the side chains of Arg 481 and Arg 492 (see Chapter 2.5 and Figure 2.36), may serve to accommodate the apolar side chains of the aldehyde substrates, e.g., the phenyl group of phenylacetaldehyde.

2.7 Structural Comparison with Mop and Implications for the Catalytic Mechanism

The proximity of Glu 308 to the tungsten site of FOR is reminiscent of a glutamic acid residue, Glu 869, positioned near the Mo site of the aldehyde oxidoreductase (Mop) from *Desulfovibrio gigas* (Romão *et al.*, 1995, PDB access code 1ALO). Although Mop is a Moco-containing enzyme that, like FOR and AOR, catalyzes aldehyde oxidation, its amino acid sequence and three-dimensional structure are unrelated to those of FOR and AOR. Structural and mechanistic studies of Mop have identified Glu 869 as playing an essential catalytic role. Interestingly, when Mop and FOR are superimposed based on the single molybdopterin group of Mop and the first molybdopterin of FOR, the side chain of Glu 308 of FOR corresponds to Glu 869 of Mop. In addition, the side chain of Tyr 416 of FOR is close to that of Tyr 622 of Mop, the oxygen ligand of the FOR tungsten atom occupies a position similar to that of water molecule 910 of Mop, and the sulfido group coordinated to the Mo of Mop corresponds to one of the dithiolene sulfurs of the second molybdopterin of FOR (Figure 2.39).

The mechanisms of Moco-containing enzymes have been extensively discussed (Hille, 1996) and they can generally be described either in terms of a coupled proton-electron transfer mechanism or a stoichiometrically equivalent oxygen-atom transfer mechanism (Stiefel, 1997). Based on the active site comparison between FOR and Mop, the sequence comparisons within the AOR-family of enzymes, and the glutarate-protein interactions observed in the FOR-glutarate complex, a catalytic mechanism based on coupled electron-proton transfer similar to that of Mop (Huber *et al.*, 1996, Romão *et al.*, 1997) can be proposed for FOR. In this analogy, Glu 308 of FOR plays a role similar to that of Glu 869 of Mop in the activation of a water molecule to attack the carbonyl group of the bound aldehyde substrate. In concert with this nucleophilic attack, the hydrogen atom from the –CHO group of the substrates is

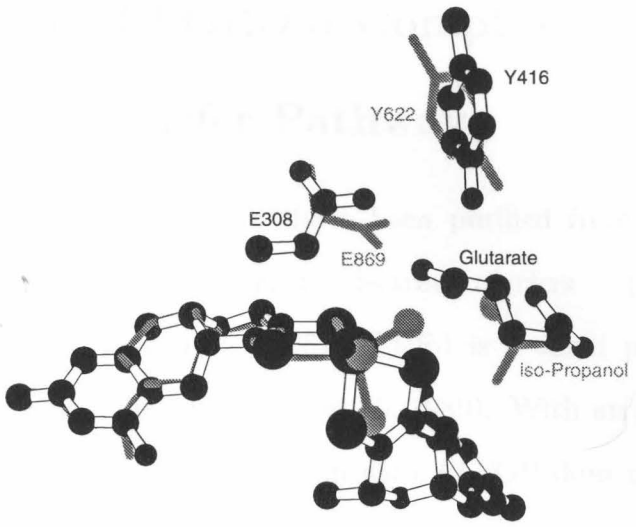


Figure 2.39: Superposition of the active sites of Mop (grey) and FOR (ball-and-stick model), illustrating the correspondence between active sites and bound ligands (iso-propanol of Mop and glutarate of FOR). The oxygen atoms of the bound water molecules and iso-propanol of Mop are shown as balls.

formally transferred to the active site of FOR as a hydride. The acceptor of this hydride could be one of the non-protein groups coordinating the tungsten atom (either an O or an S group (Roy *et al.*, 1998), or one of the dithiolene sulfur atoms). Tyr 416 is proposed to participate in the enzymatic mechanism by donating a hydrogen bond to the carbonyl oxygen of the substrate, thus serving to activate the aldehyde group for nucleophilic attack.

Residue Thr 240 might participate in proton transfer at the active site coupled to electron transfer. The side chain OH group of this residue is hydrogen bonded to a group of three well-defined water molecules held inside FOR. Rotation about the Cα–Cβ bond allows this OH group to approach the oxygen ligand of the tungsten atom to within hydrogen bonding distance. It is tempting to propose that this residue shuttles protons between the active site and solvent molecules by rotating the Cα–Cβ bond.

2.8 Structure of FOR-Fd Complex and the Electron Transfer Pathway

All the three tungstoenzymes that have been purified from hyperthermophilic archaea use Fd as their physiological electron carriers (Heider *et al.*, 1995, Mukund & Adams, 1995, Roy *et al.*, 1998). *Pf* Fd is a small protein (7.5kDa) of 66 residues and one Fe_4S_4 cluster (Aono *et al.*, 1989). With an apparent K_m value of 100μM (Roy *et al.*, 1998), its interaction with *Pf* FOR does not appear to be as strong as it is with AOR and GAPOR, whose K_m values are less than 10μM.

The docking regions on both molecules are clearly identified in the FOR-Fd complex (Figure 2.40), although the Fd structure becomes increasingly less well defined away from the interface, likely due to wobbling of the Fd about the binding site (see Chapter 2.3.4 on page 80). The Fd binding site on FOR is created by a shallow depression centered around residues Tyr 286 and Cys 287 and consists of the side chains of Pro 69, Tyr 223, Tyr 286, and Asn 288, and the main chain atoms of Tyr 286 and Cys 287, one of the Fe_4S_4 cluster ligands. This area is rather hydrophobic, does not contain any charged groups, and, somewhat surprisingly, the residues are not highly conserved in any other of the AOR-like enzymes (Figure 2.2, page 65). The majority of the binding region in Fd is provided by residues Asp 14, Ala 15, and Ile 16. These residues are part of the Fe_4S_4 cluster binding motif, and Asp 14 coordinates the Fe_4S_4 cluster with its side chain carboxylate group replacing the expected cysteine residue often found in ferredoxins from other organisms. The relatively small surface area of 350Å^2 buried at the FOR-Fd inter-molecular interface may contribute to the apparent weak interaction between these two proteins.

The Fe_4S_4 cluster of Fd is positioned near the line passing through the tungsten atom and the Fe_4S_4 cluster of FOR. The 15Å distance between Fe_4S_4 clusters of Fd and FOR is about the shortest possible distance between these centers, considering

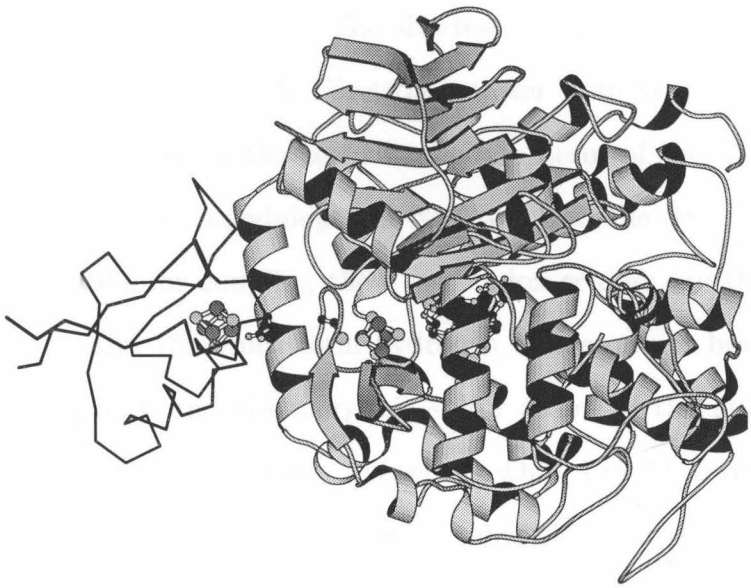


Figure 2.40: The interactions between Fd and FOR. Fd is illustrated as the Cα trace, while FOR is depicted as ribbons. Asp 14 of Fd and Cys 287 of FOR are shown as ball-and-stick models.

the shapes of the two proteins and the locations of Fe_4S_4 clusters with respect to the surfaces.

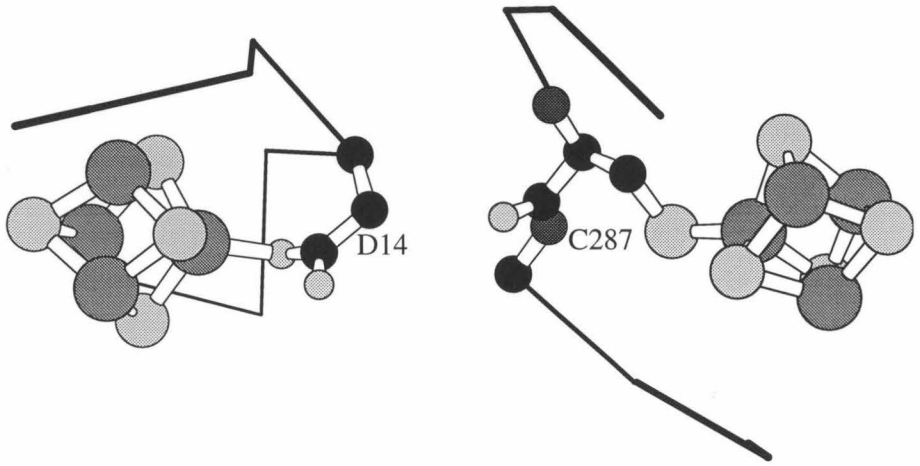


Figure 2.41: The arrangement of the FOR (right) and Fd (left) Fe_4S_4 clusters and surrounding residues in the structure of FOR-Fd complex. The Asp 14 of Fd is in van der Waals contact with Cys 287 of FOR.

A possible electron transfer pathway can be identified that leads from the tungsten atom of FOR to the Fe_4S_4 cluster of Fd. Following substrate oxidation, electrons from the reduced tungsten site are proposed to be transferred to the Fe_4S_4 cluster of FOR,

and then to the Fe_4S_4 cluster of Fd. The shortest atom-to-atom distance between the two Fe_4S_4 clusters is about 13Å. Electron transfer from W to the Fe_4S_4 cluster of FOR could be facilitated by the second molybdopterin of the Moco, through a hydrogen bond between the molybdopterin N8 atom and the Sγ atom of Cys 491. Other possible pathways for coupling these two redox centers are through the side chains of Arg 180 or Lys 75, which form hydrogen bonds or salt bridges with both the second molybdopterin and the inorganic sulfur atoms of the Fe_4S_4 cluster (see Chapter 2.4.5). Arg 180 is completely conserved while Lys 75 is only replaced by arginine in AOR and some other AOR-family tungstoenzymes, in which cases the hydrogen bonds are expected to be conserved. Interestingly, two ligands to the Fe_4S_4 clusters for FOR and Fd, Cys 287 of FOR and Asp 14 of Fd, are in van der Waals contact (Figure 2.41) and should serve to couple these two centers for electron transfer. This observation emphasizes the important role of position 14 in the electron transfer functions of Fd, as suggested by recent mutagenesis studies (Brereton *et al.*, 1998).

2.9 Concluding Remarks

Based on crystal structure analyses of FOR and its complex with glutarate, we propose that FOR and Mop, which are unrelated in their amino acid sequences and three-dimensional structures but which catalyze similar oxidation reactions, have similar enzymatic mechanisms. This proposal is supported by structural similarities between these two proteins at their active sites. Furthermore, this mechanism should also be relevant to other proteins in the AOR-family, given the conservation of residues in the immediate environment of the tungsten site. However, other residues surrounding the active site cavity are less well conserved, and may contribute to the difference in substrate specificity evident between different members of the AOR family. From the structural analysis of the FOR-Fd complex, it appears likely that the Fe_4S_4 cluster plays an important role in the electron transfer pathway between these two redox partners.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallographica,* **D52**, 30–42.

Adams, M. W. W. & Kletzin, A. (1996). Oxidoreductase-type enzymes and redox proteins involved in fermentative metabolisms of hyperthermophilic archaea. *Advances in Protein Chemistry,* **48**, 101–180.

Andreesen, J. R. & Ljungdahl, L. G. (1973). Formate dehydrogenase of *Clostridium thermoaceticum*: incorporation of selenium-75, and the effects of selenite, molybdate, and tungstate on the enzyme. *Journal of Bacteriology,* **116**, 867–873.

Aono, S., Bryant, F. O. & Adams, M. W. W. (1989). A novel and remarkably thermostable ferredoxin from the hyperthermophilic archaebacterium *Pyrococcus furiosus*. *Journal of Bacteriology,* **171** (6), 3433–3439.

Bertram, P. A., Schmitz, R. A., Linder, D. & Thauer, R. K. (1994). Tungstate can substitute for molybdate in sustaining growth of *Methanobacterium thermoautotrophicum* - identification and characterization of a tungsten isoenzyme of formylmethanofuran dehydrogenase. *Archives of Microbiology,* **161** (3), 220–228.

Boyington, J. C., Gladyshev, V. N., Khangulov, S. V., Stadtman, T. C. & Sun, P. D. (1997). Crystal structure of formate dehydrogenase H - catalysis involving Mo, molybdopterin, selenocysteine, and an Fe_4S_4 cluster. *Science,* **275** (5304), 1305–1308.

Brereton, P. S., Verhagen, M. F. J. M., Zhou, Z. H. & Adams, M. W. W. (1998). Effect of iron-sulfur cluster environment in modulating the thermodynamic properties and biological function of ferredoxin from *Pyrococcus furiosus*. *Biochemistry,* **37** (20), 7351–7362.

Brünger, A. T. (1992). Free R-value - a novel statistical quantity for assessing the accuracy of crystal structures. *Nature,* **355** (6359), 472–475.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R-factor refinement by molecular dynamics. *Science,* **235** (4787), 458–460.

Bryant, F. O. & Adams, M. W. W. (1989). Characterization of hydrogenase from the hyperthermophilic archaebacterium, *Pyrococcus furiosus*. *Journal of Biological Chemistry,* **264** (9), 5070–5079.

Calzolai, L., Gorst, C. M., Zhao, Z. H., Teng, Q., Adams, M. W. W. & Lamar, G. N. (1995). H-1 NMR investigation of the electronic and molecular structure of the 4-iron cluster ferredoxin from the hyperthermophile *Pyrococcus furiosus* - identification of Asp-14 as a cluster ligand in each of the 4 redox states. *Biochemistry,* **34** (36), 11373–11384.

CCP4 (1994). The CCP4 suite - programs for protein crystallography. *Acta Crystallographica,* **D50** (Part 5), 760–763.

Chan, M. K., Mukund, S., Kletzin, A., Adams, M. W. W. & Rees, D. C. (1995). Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science,* **267** (5203), 1463–1469.

Chothia, C. & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *The EMBO Journal,* **5** (4), 823–826.

Conover, R. C., Kowal, A. T., Fu, W., Park, J. B., Aono, S., Adams, M. W. W. & Johnson, M. K. (1990). Spectroscopic characterization of the novel iron-sulfur cluster in *Pyrococcus furiosus* ferredoxin. *Journal of Biological Chemistry,* **265** (15), 8533–8541.

Enemark, J. H. & Garner, C. D. (1997). The coordination chemistry and function of the molybdenum centres of the oxomolybdoenzymes. *Journal of Biological Inorganic Chemistry,* **2** (6), 817–822.

Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallographica,* **A47** (JUL), 392–400.

George, G. N., Prince, R. C., Mukund, S. & Adams, M. W. W. (1992). Aldehyde ferredoxin oxidoreductase from the hyperthermophilic archaebacterium *Pyrococcus furiosus* contains a tungsten oxo-thiolate center. *Journal of the American Chemical Society,* **114** (9), 3521–3523.

Georgiadis, M. M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J. J. & Rees, D. C. (1992). Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii. Science,* **257** (5077), 1653–1659.

Gorst, C. M., Zhou, Z. H., Ma, K. S., Teng, Q., Howard, J. B., Adams, M. W. W. & Lamar, G. N. (1995). Participation of the disulfide bridge in the redox cycle of the ferredoxin from the hyperthermophile *Pyrococcus furiosus* - H-1 nuclear-magnetic-resonance time resolution of the 4 redox states at ambient temperature. *Biochemistry,* **34** (27), 8788–8795.

Heider, J., Ma, K. & Adams, M. W. W. (1995). Purification, characterization, and metabolic function of tungsten-containing aldehyde ferredoxin oxidoreductase from the hyperthermophilic and proteolytic archaeon *thermococcus* strain ES-1. *Journal of Bacteriology,* **177** (16), 4757–4764.

Herzberg, O. & Moult, J. (1991). Analysis of the steric strain in the polypeptide backbone of protein molecules. *Proteins Structure Function And Genetics,* **11** (3), 223–229.

Hille, R. (1996). The mononuclear molybdenum enzymes. *Chemical Reviews,* **96**, 2757–2816.

Hochheimer, A., Schmitz, R. A., Thauer, R. K. & Hedderich, R. (1995). The tungsten formylmethanofuran dehydrogenase from *Methanobacterium thermoautotrophicum* contains sequence motifs characteristic for enzymes containing molybdopterin dinucleotide. *European Journal of Biochemistry,* **234** (3), 910–920.

Hubbard, S. J. & Argos, P. (1995). Detection of internal cavities in globular proteins. *Protein Engineering,* **8** (10), 1011–1015.

Huber, R., Hof, P., Duarte, R. O., Moura, J. J. G., Moura, I., Liu, M. Y., Legall, J., Hille, R., Archer, M. & Romão, M. J. (1996). A structure-based catalytic mechanism for the xanthine oxidase family of molybdenum enzymes. *Proceedings of the National Academy of Sciences of the United States of America,* **93** (17), 8846–8851.

Janin, J. & Chothia, C. (1990). The structure of protein-protein recognition sites. *Journal of Biological Chemistry,* **265** (27), 16027–16030.

Janin, J., Miller, S. & Chothia, C. (1988). Surface, subunit interfaces and interior of oligomeric proteins. *Journal of Molecular Biology,* **204** (1), 155–164.

Jiang, J. S. & Brünger, A. T. (1994). Protein hydration observed by X-ray diffraction - solvation properties of penicillopepsin and neuraminidase crystal structures. *Journal of Molecular Biology,* **243** (1), 100–115.

Johnson, M. K., Rees, D. C. & Adams, M. W. W. (1996). Tungstoenzymes. *Chemical Reviews,* **96** (7), 2817–2839.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallographica,* **A47**, 110–119.

Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y. *et al.* (1998). Complete sequence and gene organization of the genome of a hyperthermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Research,* **5** (2), 55–76.

Kengen, S. W. M., de Bok, F. A. M., van Loo, N. D., Dijkema, C., Stams, A. J. M. & de Vos, W. M. (1994). Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by *Pyrococcus furiosus. Journal of Biological Chemistry,* **269** (26), 17537–17541.

Khangulov, S. V., Gladyshev, V. N., Dismukes, G. C. & Stadtman, T. C. (1998). Selenium-containing formate dehydrogenase H from *Escherichia coli*: A molybdopterin enzyme that catalyzes formate oxidation without oxygen transfer. *Biochemistry,* **37** (10), 3518–3528.

Kim, J. S. & Rees, D. C. (1992). Structural models for the metal centers in the nitrogenase molybdenum-iron protein. *Science,* **257** (5077), 1677–1682.

Kisker, C., Schindelin, H., Pacheco, A., Wehbi, W. A., Garrett, R. M., Rajagopalan, K. V., Enemark, J. H. & Rees, D. C. (1997*a*). Molecular basis of sulfite oxidase deficiency from the structure of sulfite oxidase. *Cell,* **91**, 973–983.

Kisker, C., Schindelin, H. & Rees, D. C. (1997*b*). Molybdenum-cofactor-containing enzymes - structure and mechanism. *Annual Review of Biochemistry,* **66**, 233–267.

Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., Richardson, D. L., Kerlavage, A. R., Graham, D. E., Kyrpides, N. C., Fleischmann, R. D., Quackenbush, J., Lee, N. H., Sutton, G. G., Gill, S., Kirkness, E. F., Dougherty, B. A., Mckenney, K., Adams, M. D., Loftus, B., Peterson, S., Reich, C. I., Mcneil, L. K., Badger, J. H., Glodek, A., Zhou, L. X., Overbeek, R., Gocayne, J. D., Weidman, J. F., Mcdonald, L., Utterback, T., Cotton, M. D., Spriggs, T., Artiach, P., Kaine, B. P., Sykes, S. M., Sadow, P. W., Dandrea, K. P., Bowman, C., Fujii, C., Garland, S. A., Mason, T. M., Olsen, G. J., Fraser, C. M., Smith, H. O., Woese, C. R. & Venter, J. C. (1997). The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus. Nature,* **390** (6658), 364–370.

Kletzin, A. & Adams, M. W. W. (1996). Tungsten in biological systems. *FEMS Microbiology Reviews,* **18** (1), 5–63.

Kletzin, A., Mukund, S., Kelley-Crouse, T. L., Chan, M. K., Rees, D. C. & Adams, M. W. W. (1995). Molecular characterization of the genes encoding the tungsten-containing aldehyde ferredoxin oxidoreductase from *Pyrococcus furiosus* and formaldehyde ferredoxin oxidoreductase from *Thermococcus litoralis. Journal of Bacteriology,* **177** (16), 4817–4819.

Kleywegt, G. J. & Jones, T. A. (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallographica,* **D50**, 178–185.

Kleywegt, G. J. & Jones, T. A. (1996). xdlMAPMAN and xdlDATAMAN - programs for reformatting, analysis and manipulation of biomacromolecular electron density maps and reflection data sets. *Acta Crystallographica,* **D52**, 826–828.

Koehler, B. P., Mukund, S., Conover, R. C., Dhawan, I. K., Roy, R., Adams, M. W. W. & Johnson, M. K. (1996). Spectroscopic characterization of the tungsten and iron centers in aldehyde ferredoxin oxidoreductases from 2 hyperthermophilic archaea. *Journal of the American Chemical Society,* **118** (49), 12391–12405.

Larsen, T. A., Olson, A. J. & Goodsell, D. S. (1998). Morphology of protein-protein interfaces. *Structure,* **6** (4), 421–427.

Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK - a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography,* **26**, 283–291.

Ma, K., Hutchins, A., Sung, S. J. S. & Adams, M. W. W. (1997). Pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon, Pyrococcus furiosus, functions as a CoA-dependent pyruvate decarboxylase. *Proceedings of the National Academy of Sciences of the United States of America,* **94**, 9608–9613.

Ma, K., Schicho, R. N., Kelly, R. M. & Adams, M. W. W. (1993). Hydrogenase of the hyperthermophile Pyrococcus furiosus is an elemental sulfur reductase or sulfhydrogenase - evidence for a sulfur-reducing hydrogenase ancestor. *Proceedings of the National Academy of Sciences of the United States of America,* **90** (11), 5341–5344.

Matthews, M. W. (1968). Solvent content of protein crystals. *Journal of Molecular Biology,* **33**, 491–497.

McAlpine, A. S., McEwan, A. G. & Bailey, S. (1998). The high resolution crystal structure of DMSO reductase in complex with DMSO. *Journal of Molecular Biology,* **275** (4), 613–623.

Mukund, S. & Adams, M. W. W. (1991). The novel tungsten-iron-sulfur protein of the hyperthermophilic archaebacterium, *Pyrococcus furiosus*, is an aldehyde ferredoxin oxidoreductase - evidence for its participation in a unique glycolytic pathway. *Journal of Biological Chemistry,* **266** (22), 14208–14216.

Mukund, S. & Adams, M. W. W. (1993). Characterization of a novel tungsten-containing formaldehyde ferredoxin oxidoreductase from the hyperthermophilic archaeon, *Thermococcus litoralis* - a role for tungsten in peptide catabolism. *Journal of Biological Chemistry,* **268** (18), 13592–13600.

Mukund, S. & Adams, M. W. W. (1995). Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon *Pyrococcus furiosus. Journal of Biological Chemistry,* **270** (15), 8389–8392.

Mukund, S. & Adams, M. W. W. (1996). Molybdenum and vanadium do not replace tungsten in the catalytically active forms of the 3 tungstoenzymes in the hyperthermophilic archaeon *Pyrococcus furiosus. Journal of Bacteriology,* **178** (1), 163–167.

Navaza, J. & Saludjian, P. (1997). AMoRe - an automated molecular replacement program package. *Methods in Enzymology,* **276**, 581–594.

Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods In Enzymology,* **276**, 307–326.

Rajagopalan, K. V. & Johnson, J. L. (1992). The pterin molybdenum cofactors. *Journal of Biological Chemistry,* **267** (15), 10199–10202.

Rees, D. C., Hu, Y. L., Kisker, C. & Schindelin, H. (1997). A crystallographic view of the molybdenum cofactor. *Journal of the Chemical Society Dalton Transactions,* 3909–3914.

Rees, D. C., Lewis, M., Honzatko, R. B., Lipscomb, W. N. & Hardman, K. D. (1981). Zinc environment and *cis* peptide bonds in carboxypeptidase A at 1.75Å resolution. *Proceedings of the National Academy of Sciences of the United States of America,* **78** (6), 3408–3412.

Romão, M. J., Archer, M., Moura, I., Moura, J. J. G., Legall, J., Engh, R., Schneider, M., Hof, P. & Huber, R. (1995). Crystal structure of the xanthine oxidase-related aldehyde oxidoreductase from *D. gigas. Science,* **270** (5239), 1170–1176.

Romão, M. J., Rösch, N. & Huber, R. (1997). The molybdenum site in the xanthine oxidase-related aldehyde oxidoreductase from Desulfovibrio gigas and a catalytic mechanism for this class of enzymes. *Journal of Biological Inorganic Chemistry,* **2** (6), 782–785.

Rosner, B. M. & Schink, B. (1995). Purification and characterization of acetylene hydratase of *Pelobacter acetylenicus*, a tungsten iron-sulfur protein. *Journal of Bacteriology,* **177** (20), 5767–5772.

Roy, R., Mukund, S., Schut, G., Dunn, D. M., Weiss, R. & Adams, M. W. W. (1998). Purification and molecular characterization of tungsten-containing formaldehyde ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*: The third of a putative five member tungstoenzyme family. *Journal of Bacteriology,* **submitted**.

Schindelin, H., Kisker, C., Hilton, J., Rajagopalan, K. V. & Rees, D. C. (1996). Crystal structure of DMSO reductase - redox-linked changes in molybdopterin coordination. *Science,* **272** (5268), 1615–1621.

Schindelin, H., Kisker, C. & Rees, D. C. (1997). The molybdenum cofactor - a crystallographic perspective. *Journal of Biological Inorganic Chemistry,* **2** (6), 773–781.

Schmitz, R. A., Albracht, S. P. J. & Thauer, R. K. (1992*a*). A molybdenum and a tungsten isoenzyme of formylmethanofuran dehydrogenase in the thermophilic archaeon *Methanobacterium wolfei. European Journal of Biochemistry,* **209** (3), 1013–1018.

Schmitz, R. A., Richter, M., Linder, D. & Thauer, R. K. (1992*b*). A tungsten-containing active formylmethanofuran dehydrogenase in the thermophilic archaeon *Methanobacterium wolfei. European Journal of Biochemistry,* **207** (2), 559–565.

Schneider, F., Löwe, J., Huber, R., Schindelin, H., Kisker, C. & Knäblein, J. (1996). Crystal structure of dimethyl sulfoxide reductase from *Rhodobacter capsulatus* at 1.88Å resolution. *Journal of Molecular Biology,* **263** (1), 53–69.

Sery, A., Housset, D., Serre, L., Bonicel, J., Hatchikian, C., Frey, M. & Roth, M. (1994). Crystal structure of the ferredoxin-I from *Desulfovibrio africanus* at 2.3 Angstrom resolution. *Biochemistry,* **33** (51), 15408–15417.

Shuber, A. P., Orr, E. C., Recny, M. A., Schendel, P. F., May, H. D., Schauer, N. L. & Ferry, J. G. (1986). Cloning, expression, and nucleotide sequence of the formate dehydrogenase genes from *Methanobacterium formicicum. Journal of Biological Chemistry,* **261** (28), 2942–2947.

Stewart, D. E., Sarkar, A. & Wampler, J. E. (1990). Occurrence and role of *cis* peptide bonds in protein structures. *Journal of Molecular Biology,* **214** (1), 253–260.

Stiefel, E. I. (1993). Molybdenum enzymes, cofactors, and chemistry: an introductory survey. In *Molybdenum enzymes, cofactors, and model systems*, (Stiefel, E. I.,

Coucouvanis, D. & Newton, W. E., eds), vol. 535, of *ACS Symposium Series* pp. 1–19. American Chemical Society.

Stiefel, E. I. (1997). Chemical keys to molybdenum enzymes. *Journal of the Chemical Society Dalton Transactions,* 3915–3923.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research,* **22** (22), 4673–4680.

Trautwein, T., Krauss, F., Lottspeich, F. & Simon, H. (1994). The (2R)-hydroxycarboxylate-viologen-oxidoreductase from *Proteus vulgaris* is a molybdenum-containing iron-sulfur protein. *European Journal of Biochemistry,* **222** (3), 1025–1032.

Vorholt, J. A., Vaupel, M. & Thauer, R. K. (1996). A polyferredoxin with eight [4Fe-4S] clusters as a subunit of molybdenum formylmethanofuran dehydrogenase from *Methanosarcina barkeri. European Journal of Biochemistry,* **236** (1), 309–317.

White, H., Feicht, R., Huber, C., Lottspeich, F. & Simon, H. (1991). Purification and some properties of the tungsten-containing carboxylic acid reductase from *Clostridium formicoaceticum. Biological Chemistry Hoppe-Seyler,* **372** (11), 999–1005.

White, H., Huber, C., Feicht, R. & Simon, H. (1993). On a reversible molybdenum-containing aldehyde oxidoreductase from clostridium-formicoaceticum. *Archives of Microbiology,* **159** (3), 244–249.

White, H. & Simon, H. (1992). The role of tungstate and or molybdate in the formation of aldehyde oxidoreductase in clostridium-thermoaceticum and other

acetogens - immunological distances of such enzymes. *Archives of Microbiology,* **158** (2), 81–84.

White, H., Strobl, G., Feicht, R. & Simon, H. (1989). Carboxylic acid reductase - a new tungsten enzyme catalyzes the reduction of non-activated carboxylic acids to aldehydes. *European Journal of Biochemistry,* **184** (1), 89–96.

Wilson, A. J. C. (1942). Determination of absolute from relative X-ray intensity data. *Nature,* **150**, 151–152.

Wootton, J. C., Nicolson, R. E., Cock, J. M., Walters, D. E., Burke, J. F., Doyle, W. A. & Bray, R. C. (1991). Enzymes depending on the pterin molybdenum cofactor - sequence families, spectroscopic properties of molybdenum and possible cofactor-binding domains. *Biochimica Et Biophysica Acta,* **1057** (2), 157–185.

Yamamoto, I., Saiki, T., Liu, S. M. & Ljungdahl, L. G. (1983). Purification and properties of NADP-dependent formate dehydrogenase from *Clostridium thermoaceticum*, a tungsten-selenium-iron protein. *Journal of Biological Chemistry,* **258** (3), 1826–1832.

Appendix A. Topology File for Moco

```
remarks topology file for pterin

    autogenerate angles=true dihedrals=false end

    MASS  AC    40.08
    MASS  MG    24.30500
    MASS  OG1   15.99940
    MASS  OG2   15.99940
    MASS  OG3   15.99940
    MASS  OG4   15.99940
    MASS  HOG    1.00794

    MASS  PP    30.97400
    MASS  OPH   15.99940
    MASS  OPP   15.99940
    MASS  OSP   15.99940
    MASS  HOP   1.007940

    MASS  CTP   12.01100
    MASS  C1P   13.01900
    MASS  C2P   14.02700
    MASS  C3P   15.03500
    MASS  CRP   12.01100
    MASS  CR1P  13.01900
    MASS  NHP   14.00670
    MASS  NRP   14.00670
    MASS  OP    15.99940
    MASS  HNP   1.007940

    MASS  SP    32.06600
    MASS  W     183.8500
    MASS  OXO   15.99940

    RESIdue PTE

     GROUP
       ATOM MG1   TYPE=MG    CHARges= 2.00   END
       ATOM AC1   TYPE=AC    CHARges= 2.00   END
       ATOM OG1   TYPE=OG1   CHARges=-0.834  END
       ATOM HA1   TYPE=HOG   CHARGes= 0.417  END
       ATOM HB1   TYPE=HOG   CHARGES= 0.417  END
```

```
      ATOM OG2   TYPE=OG2   CHARges=-0.834  END
      ATOM HA2   TYPE=HOG   CHARGes= 0.417  END
      ATOM HB2   TYPE=HOG   CHARGES= 0.417  END
      ATOM OG3   TYPE=OG3   CHARges=-0.834  END
      ATOM HA3   TYPE=HOG   CHARGes= 0.417  END
      ATOM HB3   TYPE=HOG   CHARGES= 0.417  END
      ATOM OG4   TYPE=OG4   CHARges=-0.834  END
      ATOM HA4   TYPE=HOG   CHARGes= 0.417  END
      ATOM HB4   TYPE=HOG   CHARGES= 0.417  END

{first pterin}
    GROUP
      ATOM P1   TYPE=PP     CHARge= 1.50    END
      ATOM OP1  TYPE=OPH    CHARge=-1.00    END
      ATOM OP2  TYPE=OPH    CHARge=-0.75    END
      ATOM OP3  TYPE=OPP    CHARge=-0.50    END
      ATOM OP4  TYPE=OSP    CHARge=-0.75    END
      ATOM HP2  TYPE=HOP    CHARge= 0.25    END
    GROUP
      ATOM C1   TYPE=C2P    CHARge= 0.25    END
    GROUP
      ATOM C2   TYPE=C1P    CHARge= 0.25    END
      ATOM O2   TYPE=OSP    CHARge=-0.50    END
    GROUP
      ATOM C3   TYPE=CTP    CHARge= 0.19    END
      ATOM S3   TYPE=SP     CHARge=-1.19    END
    GROUP
      ATOM C4   TYPE=CTP    CHARge= 0.19    END
      ATOM S4   TYPE=SP     CHARge=-1.19    END
    GROUP
      ATOM C5   TYPE=C1P    CHARge= 0.00    END
    GROUP
      ATOM N6   TYPE=NHP    CHARge=-0.30    END
      ATOM H6   TYPE=HNP    CHARge= 0.30    END
    GROUP
      ATOM C7   TYPE=CRP    CHARge= 0.00    END
    GROUP
      ATOM C8   TYPE=CRP    CHARge= 0.55    END
      ATOM O8   TYPE=OP     CHARge=-0.55    END
    GROUP
      ATOM N9   TYPE=NHP    CHARge=-0.30    END
      ATOM H9  TYPE=HNP    CHARge= 0.30    END
    GROUP
      ATOM C10  TYPE=CRP    CHARge= 0.12    END
```

```
      ATOM N10  TYPE=NHP    CHARge=-0.80    END
      ATOM H10 TYPE=HNP    CHARge= 0.46    END
      ATOM H1A TYPE=HNP    CHARge= 0.46    END
    GROUP
      ATOM N11  TYPE=NRP    CHARge= 0.00    END
    GROUP
      ATOM C12  TYPE=CRP    CHARge= 0.00    END
    GROUP
      ATOM N13  TYPE=NHP    CHARge=-0.30    END
      ATOM H13  TYPE=HNP    CHARge= 0.30    END
    GROUP
      ATOM C14  TYPE=C1P    CHARge= 0.25    END

{second pterin}
    GROUP
      ATOM P2    TYPE=PP     CHARge= 1.50    END
      ATOM OP5   TYPE=OPH    CHARge=-1.00    END
      ATOM OP6   TYPE=OPH    CHARge=-0.75    END
      ATOM OP7   TYPE=OPP    CHARge=-0.50    END
      ATOM OP8   TYPE=OSP    CHARge=-0.75    END
      ATOM HP6   TYPE=HOP    CHARge= 0.25    END
    GROUP
      ATOM C21   TYPE=C2P    CHARge= 0.25    END
    GROUP
      ATOM C22   TYPE=C1P    CHARge= 0.25    END
      ATOM O22   TYPE=OSP    CHARge=-0.50    END
    GROUP
      ATOM C23   TYPE=CTP    CHARge= 0.19    END
      ATOM S23   TYPE=SP     CHARge=-1.19    END
    GROUP
      ATOM C24   TYPE=CTP    CHARge= 0.19    END
      ATOM S24   TYPE=SP     CHARge=-1.19    END
    GROUP
      ATOM C25   TYPE=C1P    CHARge= 0.00    END
    GROUP
      ATOM N26   TYPE=NHP    CHARge=-0.30    END
      ATOM H26  TYPE=HNP    CHARge= 0.30    END
    GROUP
      ATOM C27   TYPE=CRP    CHARge= 0.00    END
    GROUP
      ATOM C28   TYPE=CRP    CHARge= 0.55    END
      ATOM O28   TYPE=OP     CHARge=-0.55    END
    GROUP
      ATOM N29   TYPE=NHP    CHARge=-0.30    END
```

```
      ATOM H29  TYPE=HNP    CHARge= 0.30    END
    GROUP
      ATOM C30  TYPE=CRP    CHARge= 0.12    END
      ATOM N30  TYPE=NHP    CHARge=-0.80    END
      ATOM H30 TYPE=HNP    CHARge= 0.46    END
      ATOM H3A TYPE=HNP    CHARge= 0.46    END
    GROUP
      ATOM N31  TYPE=NRP    CHARge= 0.00    END
    GROUP
      ATOM C32  TYPE=CRP    CHARge= 0.00    END
    GROUP
      ATOM N33  TYPE=NHP    CHARge=-0.30    END
      ATOM H33  TYPE=HNP    CHARge= 0.30    END
    GROUP
      ATOM C34  TYPE=C1P    CHARge= 0.25    END

    GROUP
      ATOM W1   TYPE=W      CHARge= 6.00    END
      ATOM OX1  TYPE=OXO    CHARge=-2.00    END

    BOND MG1 OG1     BOND MG1 OG2
    BOND OG1 HA1     BOND OG1 HB1
    BOND OG2 HA2     BOND OG2 HB2

    BOND MG1 OP1      BOND MG1 OP5
    BOND AC1 OG3     BOND AC1 OG4
    BOND OG3 HA3     BOND OG3 HB3
    BOND OG4 HA4     BOND OG4 HB4
{end}

    BOND P1  OP1      BOND P1 OP2
    BOND P1 OP3       BOND P1 OP4
    BOND OP4 C1       BOND OP2 HP2

    BOND C1 C2
    BOND C2 C3        BOND C2 O2
    BOND C3 C4        BOND C3 S3
    BOND C4 C5        BOND C4 S4

    BOND C5 N6        BOND C5 C14
    BOND N6 C7        BOND N6 H6
    BOND C7 C8        BOND C7 C12
    BOND C8 N9        BOND C8 O8
    BOND N9 C10       BOND N9 H9
```

```
BOND C10 N11      BOND C10 N10
BOND N10 H10      BOND N10 H1A
BOND N11 C12
BOND C12 N13
BOND N13 C14      BOND N13 H13
BOND C14 O2

BOND P2 OP5       BOND P2 OP6
BOND P2 OP7       BOND P2 OP8
BOND OP8 C21      BOND OP6 HP6

BOND C21 C22
BOND C22 C23        BOND C22 O22
BOND C23 C24        BOND C23 S23
BOND C24 C25        BOND C24 S24

BOND C25 N26        BOND C25 C34
BOND N26 C27        BOND N26 H26
BOND C27 C28        BOND C27 C32
BOND C28 N29        BOND C28 O28
BOND N29 C30        BOND N29 H29
BOND C30 N31        BOND C30 N30
BOND N30 H30       BOND N30 H3A
BOND N31 C32
BOND C32 N33
BOND N33 C34        BOND N33 H33
BOND C34 O22

BOND W1 S3       BOND W1 S4
BOND W1 S23      BOND W1 S24
BOND W1 OX1

DIHE P1 OP4 C1 C2       DIHE OP1 P1 OP4 C1
DIHE OP4 C1 C2 C3

IMPR C2 C3 C4 C5

IMPR N11 C12 C7 C8

IMPR N6 C12 C8 C7      IMPR O8 C7 N9 C8
IMPR N10 N9 N11 C10
IMPR N13 N11 C7 C12

DIHE P2 OP8 C21 C22   DIHE OP5 P2 OP8 C21
```

```
 DIHE OP8  C21 C22 C23

 IMPR C22 C23 C24 C25

 IMPR N31 C32 C27 C28

 IMPR N26 C32 C28 C27   IMPR O28 C27 N29 C28
 IMPR N30 N29 N31 C30   IMPR N33 N31 C27 C32

 END {PTE}
PREsidue PPTE
  GROUp
    MODIfy ATOM 1O           CHARge=-0.55 END
    MODIfy ATOM 1C           CHARge= 0.55 END
  GROUp
    MODIfy ATOM 2O           CHARge=-0.55 END
    MODIfy ATOM 2C           CHARge= 0.55 END

  ADD BOND 3MG1   1O
  ADD BOND 3MG1   2O

  ADD ANGLe  1O 3MG1 2O

  ADD ANGLe  1O 3MG1 3OP1    ADD ANGLe  1O 3MG1 3OP5
  ADD ANGLe  1O 3MG1 3OG1    ADD ANGLe  1O 3MG1 3OG2

  ADD ANGLe  2O 3MG1 3OP1    ADD ANGLe  2O 3MG1 3OP5
  ADD ANGLe  2O 3MG1 3OG1    ADD ANGLe  2O 3MG1 3OG2


  END {PPTE}

-------------------------------------------------
PREsidue PCA              {*PATCh command.*}
  GROUp
    MODIfy ATOM 1O           CHARge=-0.55 END
    MODIfy ATOM 1C           CHARge= 0.55 END
  GROUp
    MODIfy ATOM 2OD2 TYPE=OC CHARge=-1.00 END
  GROUp
    MODIfy ATOM 3OE2 TYPE=OC CHARge=-1.00 END

  ADD BOND 4AC1   1O
  ADD BOND 4AC1   2OD2
```

```
ADD BOND 4AC1   3OE2
ADD BOND 4AC1   4O8

ADD ANGLe  1O 4AC1 4O8
ADD ANGLe  1O 4AC1 3OE2
ADD ANGLe  1O 4AC1 4OG3
ADD ANGLe  1O 4AC1 4OG4

ADD ANGLE 4O8 4AC1 2OD2
ADD ANGLE 4O8 4AC1 4OG3
ADD ANGLE 4O8 4AC1 4OG4

ADD ANGLE 2OD2 4AC1 3OE2
ADD ANGLE 2OD2 4AC1 4OG3
ADD ANGLE 2OD2 4AC1 4OG4

ADD ANGLE 3OE2 4AC1 4OG3
ADD ANGLE 3OE2 4AC1 4OG4


END
```

Appendix B. Parameter File for Moco

```
remarks parameters for pterin file

   bonds AC OG3  200. 2.42
   bonds AC OG4  200. 2.42
   bonds AC OP   200. 2.09
   bonds AC O    200. 2.37
   bonds AC OC   100. 2.45
   BONDS OG3 HOG  450.0  0.9572
   BONDS OG4 HOG  450.0  0.9572
   BONDS HOG HOG   0.0  1.5139

angles OG3 AC OP  30.  90.
angles OG3 AC OG4 30. 180.
angles OG4 AC OP  30.  90.
angles OG3 AC O   30.  90.
angles OG4 AC O   30.  90.
angles OG3 AC OC  30.  90.
angles OG4 AC OC  30.  90.
angles OP  AC OC  30.  90.
angles O   AC OP  30.  90.
angles O   AC OC  30.  90.
angles OC  AC OC  30.  90.

angles AC OG3 HOG 30. 110.5
angles HOG OG3 HOG 30.0 104.52
angles AC OG4 HOG 30. 110.5
angles HOG OG4 HOG 30.0 104.52
{end of calcium}

   bonds MG OG1  500. 2.10
   bonds MG OG2  500. 2.10
   bonds MG OPH  500. 2.03
   bonds MG O    500. 2.05
   BONDS OG1 HOG  450.0  0.9572
   BONDS OG2 HOG  450.0  0.9572
   BONDS HOG HOG   0.0   1.5139

   angles OG1 MG OG2 30. 90.
   angles OG1 MG OPH 30. 90.
   angles OG2 MG OPH 30. 90.
   angles OG1 MG O 0. 180.
   angles OG2 MG O 0. 180.
```

```
angles OPH MG OPH 30. 180.
angles OPH MG O 30. 90.
angles O MG O 30. 90.

angles MG OG1 HOG   30.0     110.5
ANGLE HOG OG1 HOG   30.0     104.52
angles MG OG2 HOG   30.0     110.5
ANGLE HOG OG2 HOG   30.0     104.52

bond PP OPH   500.0  1.61
bond PP OPP   528.0  1.48
bond PP OSP   237.0  1.61
bond OSP C2P  292.0  1.43
bond OPH HOP  450.0  0.96

angles HOP OPH PP    47.0   107.3
angles OPH PP OPH    48.1   102.0
angles OPH PP OPP    100.0  108.0
angles OPH PP OSP    48.1   102.0
angles OPP PP OSP    100.0  108.0
angles PP OSP C2P    47.0   120.0

dihedral  X PP OPH X      0.5  3 0.0
dihedral  X PP OSP X      0.5  3 0.0
dihedral  X PP OPP X      0.5  3 0.0
dihedral  X OSP C2P X     0.5  3 0.0
dihedral  X C1P C2P X     0.5  3 0.0

IMPR C1P CTP CTP C1P 45.0 0 0.0
IMPR NRP CRP CRP CRP 45.0 0 0.0
IMPR NHP NRP CRP CRP 45.0 0 0.0
IMPR NHP NHP CRP CRP 45.0 0 0.0
IMPR NHP NHP NRP CRP 45.0 0 0.0
IMPR NHP CRP CRP CRP 45.0 0 0.0
IMPR OP  CRP CRP CRP 45.0 0 0.0
IMPR OP  CRP NHP CRP 45.0 0 0.0

bond C2P C1P  500.  1.54

bond C1P OSP  500.  1.38

bond C1P CTP  500.  1.54
bond CTP CTP  200.  1.34
bond CTP SP   500.  1.74
```

```
BOND C1P NHP  500.  1.46
BOND C1P C1P  500.  1.54
BOND NHP CRP  500.  1.34
BOND NRP CRP  500.  1.34
BOND CRP CRP  500.  1.42
BOND CRP OP   500.  1.39

BOND NHP HNP  500.  0.97

angle OSP C2P C1P 50. 110.5

angle C2P C1P CTP 0. 110.5
angle C2P C1P OSP 0. 110.5
angle OSP C1P CTP 0. 110.5
angle C1P OSP C1P 0. 111.7
angle OSP C1P NHP 0. 110.5
angle OSP C1P C1P 0. 110.5

angle C1P CTP CTP 50. 120.
angle SP CTP CTP  50. 120.
angle SP CTP C1P  50. 120.

angle W SP CTP  50. 106.5

angle CTP C1P NHP 0. 110.5
angle NHP C1P C2P 0. 110.5

angle C1P NHP CRP 50. 120.0
angle C1P NHP HNP 50. 120.0
angle HNP NHP CRP 50. 120.0
angle HNP NHP HNP 50. 120.0

angle NHP CRP CRP 50. 120.0
angle NRP CRP CRP 50. 120.0

angle CRP CRP OP  50. 120.0
angle CRP CRP CRP 50. 120.0
angle NHP CRP OP  50. 120.0

angle CRP NHP CRP 50. 120.0

angle NHP CRP NHP 50. 120.0
angle NHP CRP NRP 50. 120.0
```

```
 angle CRP NRP CRP 50. 120.0

 angle CRP NHP C2P 50. 120.0
 angle HNP NHP C2P 50. 120.0

 angle NHP C1P C1P 0. 110.5

 angle C1P C1P CTP 0. 110.5

 bond W SP 500. 2.47
 bond W OXO 500. 2.10

 angle SP W SP 0. 90.
 angle SP W OXO 0. 90.

dihedral X SP W  X  0. 2 180.

 angle MG OPH PP 10. 137.5

             {*   Lennard-Jones parameters              *}
             {*                          ------1-4------- *}
             {*  epsilon   sigma        epsilon   sigma *}
             {* (kcal/mol)   (A)       (kcal/mol) (A)   *}

 nonbonded MG        0.1000   1.3200     0.1000   1.3200
 nonbonded AC        0.1000   1.3200     0.1000   1.3200
 nonbonded OG1       0.1591   2.8509     0.1591   2.8509
 nonbonded OG2       0.1591   2.8509     0.1591   2.8509
 nonbonded OG3       0.1591   2.8509     0.1591   2.8509
 nonbonded OG4       0.1591   2.8509     0.1591   2.8509
 nonbonded HOG       0.0498   1.4254     0.0498   1.4254

 nonbonded PP        0.5849   3.3854     0.5849   3.3854
 nonbonded OPH       0.1591   2.8509     0.1591   2.8509
 nonbonded OPP       0.1591   2.8509     0.1591   2.8509
 nonbonded OSP       0.1591   2.8509     0.1591   2.8509

 nonbonded CTP       0.1200   3.7418     0.1      3.3854
 nonbonded C1P       0.0486   4.2140     0.1      3.3854
 nonbonded C2P       0.1142   3.9823     0.1      3.3854
 nonbonded C3P       0.1811   3.8576     0.1      3.3854
 nonbonded CRP       0.1200   3.7418     0.1      3.3854
 nonbonded CR1P      0.0486   4.2140     0.1      3.3854
```

```
nonbonded NHP       0.2384   2.8509     0.2384   2.8509
nonbonded NRP       0.2384   2.8509     0.2384   2.8509
nonbonded OP        0.1591   2.8509     0.1591   2.8509
nonbonded HNP       0.0498   1.4254     0.0498   1.4254
nonbonded HOP       0.0498   1.4254     0.0498   1.4254

nonbonded SP        0.0430   3.3676     0.0430   3.3676
nonbonded W         0.1000   1.4000     0.1000   1.4000
nonbonded OXO       0.1591   2.8509     0.1591   2.8509
```