Distorted octahedral coordination of tungstate in a subfamily of specific binding proteins

Kaspar Hollenstein[†], Mireia Comellas-Bigler[†], Loes E. Bevers, Martin C. Feiters, Wolfram Meyer-Klaucke, Peter-Leon Hagedoorn, Kaspar P. Locher

[†] These authors contributed equally to this work.

K. Hollenstein, M. Comellas-Bigler, K.P. Locher (✉)
Institute of Molecular Biology and Biophysics
ETH Zurich
Schafmattstrasse 20
CH-8093 Zurich, Switzerland
E-mail: kaspar.locher@mol.biol.ethz.ch
Fax: +41-44-633-1182

L.E. Bevers, P.L. Hagedoorn
Department of Biotechnology
Delft University of Technology
Julianalaan 67
2628 BC Delft, The Netherlands

M.C. Feiters
Department of Organic Chemistry
Institute for Molecules and Materials
Faculty of Science
Radboud University Nijmegen
Heyendaalseweg 135
6525 AJ Nijmegen, The Netherlands

W. Meyer-Klaucke
European Molecular Biology Laboratory
Hamburg Unit
Deutsches Elektronen Synchrotron
Notkestrasse 85
D-22607 Hamburg, Germany

Abstract

Bacteria and archaea import molybdenum (Mo) and tungsten (W) from the environment in the form of the oxyanions molybdate (MoO_4^{2-}) and tungstate (WO_4^{2-}). These substrates are captured by an external, high-affinity binding protein, and delivered to ATP binding cassette (ABC) transporters, which move them across the cell membrane. We have recently reported a crystal structure of the molybdate/tungstate binding protein ModA/WtpA from *Archaeoglobus fulgidus*, which revealed an octahedrally coordinated central metal atom. By contrast, the previously determined structures of three bacterial homologs showed tetracoordinate Mo and W atoms in their binding pockets. Until then, coordination numbers above four had only been found for Mo/W in metalloenzymes where these metal atoms are part of the catalytic cofactors and coordinated by mostly non-oxygen ligands. We now report a high resolution structure of *A. fulgidus* ModA/WtpA, as well as crystal structures of four additional homologs, all bound to tungstate. These crystal structures match X-ray absorption spectroscopy (XAS) measurements from soluble, tungstate-bound protein, and reveal the details of the distorted octahedral coordination. Our results demonstrate that the distorted octahedral geometry is not an exclusive feature of the *A. fulgidus* protein, and suggest distinct binding modes of the binding proteins from archaea and bacteria.

Keywords: Tungsten, Binding proteins, Octahedral coordination, Crystal structure, Extended X-ray absorption fine structure

Abbreviations

ABC	ATP binding cassette
ATP	Adenosine triphosphate
DDT	1,4-Dithio-DL-threitol
EDTA	Ethylenediaminetetraacetic acid
EXAFS	Extended X-ray absorption fine structure
MES	2-Morpholinoethanesulphonic acid
PEG	Polyethyleneglycol
SAD	Single-wavelength anomalous dispersion
Tris	Tris(hydroxymethyl)aminomethane
XAS	X-ray absorption spectroscopy

Introduction

Molybdenum (Mo) is an essential element throughout biology. Certain archaea and bacteria, however, are able to use tungsten (W) and some hyperthermophiles even appear to be obligately dependent on tungsten and incapable of utilizing molybdenum [1, 2]. The two trace metals are taken up in the form of the oxyanion complexes molybdate (MoO_4^{2-}) and tungstate (WO_4^{2-}). Like other essential nutrients, these are accumulated from the environment by active transport across the bacterial or archaeal cell membrane. Uptake is catalyzed by ATP binding cassette (ABC) transporters that consist of a membrane-integral transport complex, composed of two transmembrane and two nucleotide-binding domains, and an external substrate binding protein [3]. The transporter provides a gated passageway across the membrane and powers the energetically 'uphill' translocation through ATP hydrolysis. The specificity is determined by the binding protein that serves as high-affinity receptor of the substrate with dissociation constants (K_D) in the micromolar, occasionally even in the nanomolar range [4, 5].

Significant insight into the structure and function of bacterial and archaeal binding proteins has been accumulated over the past years. They fall into three classes based on their folds, and all feature a bi-lobed architecture with a substrate binding site buried between the two halves [4, 5]. We have recently determined the crystal structure of a molybdate/tungstate-binding protein from the hyperthermophile *Archaeoglobus fulgidus* in its molybdate- and tungstate-bound forms, free and in complex with the cognate ABC transporter ModBC [6]. Due to the evident similarity of its fold and sequence to bacterial homologs from *Escherichia coli* and *Azotobacter vinelandii*, we had termed the protein ModA, and the corresponding transporter ModBC. At present it remains unknown if *A. fulgidus* accumulates MoO_4^{2-}, WO_4^{2-}, or both ions through this Mod system. Independently, however, it has been proposed that based on biochemical data and sequence comparisons this transporter family would preferentially facilitate the uptake of WO_4^{2-},

and an alternative name (WtpA) was introduced for the binding protein [7]. In the following, we will refer to these binding proteins as ModA/WtpA wherever possible.

The structure of *A. fulgidus* ModA/WtpA has revealed a binding site for molybdate and tungstate where the central metal atom is in a hexacoordinate configuration. This octahedral geometry was rather unexpected and in contrast to the previously solved structures of the two ModA homologs from *E. coli* and *A. vinelandii,* where the oxygen atoms are tetrahedrally disposed around the metal center [8, 9]. In line with these two structures the recently published structure of a homolog from the plant pathogen *Xanthomonas axonopodis* pv. *citri* showed a tetracoordinate Mo atom in the substrate binding site [10].

To determine whether the octahedral geometry was a unique feature of the *A. fulgidus* protein, and to study the binding mode in more detail, we have determined crystal structures of four additional archaeal homologs of the ModA/WtpA protein, in addition to a much higher resolution structure of *A. fulgidus* ModA/WtpA. We find that all five homologs indeed bind tungstate with an octahedral geometry, and by performing X-ray absorption spectroscopy (XAS) [11] studies, we demonstrate that this binding mode is also present in solution.

Materials and methods

Expression and purification of ModA/WtpA proteins. The relevant genes were amplified by PCR from the genomic DNA of *Archaeoglobus fulgidus* (ATCC 49558D)*, Methanosarcina acetivorans* (ATCC 35395), *Methanocaldococcus jannaschii* (ATCC 43067D), *Pyrococcus horikoshii* (ATCC 7860), and *Pyrococcus furiosus* (ATCC 43587). The precise starting and end points of the constructs are indicated in Fig. 1. The resulting fragments were inserted into a modified pET-19b (Novagen), attaching an N-terminal decahistidine affinity tag and a Tobacco Etch Virus (TEV) protease cleavage site. All plasmids were verified by DNA sequencing, which indicated that AfModA was expressed with a nonahistidine tag only.

AfModA, MaModA, MjModA, PhModA, and PfModA (the first two letters indicating the source organism) were over-expressed in the cytoplasm of *Escherichia coli* BL21-CodonPlus(DE3)-RIPL (Stratagene). Cells were grown in Terrific Broth supplemented with 1 % (w/v) glycerol at 37 °C. At an optical density (A_{600}) of 0.5-1.0 the temperature was lowered to 25°C and cells were grown into stationary phase where expression was auto-induced. All subsequent purification steps were carried out at room temperature unless stated differently. Cells were harvested by centrifugation and disrupted using a M-110L microfluidizer (Microfluidics) at 15.000 psi external pressure. Insoluble material was removed by centrifugation (40.000 x *g* for 45 min at 4°C) and the supernatant was loaded onto a NiNTA superflow affinity column (Qiagen). The column was washed with 25 mM and 60 mM imidazole (for AfModA, 80 mM imidazole for all others), and the proteins were eluted with 300 mM imidazole. The buffer was exchanged to 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.5 mM EDTA, and 2 mM 1,4-dithio-DL-threitol (DTT) by desalting and the protein was incubated with TEV protease for 16-18 hours at 20°C. The buffer was exchanged to 25 mM Tris-HCl pH 8.0, 250 mM NaCl, and 5 mM imidazole prior to removal of cleavage fragments and uncleaved material by a re-run on a NiNTA superflow column. The proteins were desalted into 5 mM Tris-HCl pH 7.5, 50 mM NaCl

(AfModA, MaModA, MjModA, PfModA) or 10 mM Tris-HCl pH 7.5, 100 NaCl (PhModA). For MaModA, MjModA, PfModA, and PhModA, the desalt buffer also contained 1.5 mM Na_2WO_4. The proteins were concentrated using Amicon Ultra-15 concentrator units (Millipore) with molecular cutoffs of 10 kDa. Concentrations were 25-30 mg/ml (AfModA), 10-15 mg/ml (MaModA), 10-15 mg/ml (MjModA), 5-10 mg/ml (PhModA), 5-10 mg/ml (PfModA). For AfModA, 1.5 mM Na_2WO_4 was added to the protein before crystallization.

Crystallization and structure determination. AfModA, MaModA, MjModA, PhModA, and PfModA (or WtpA) were crystallized by vapor diffusion in sitting drops at 20 °C against reservoir solutions as follows: AfModA (30 mM Tris-HCl pH 8.4, 160 mM Na_2SO_4, 5 % glycerol, 28 % PEG 4000); MaModA (0.5 mM MES-NaOH pH 6.5, 210 mM Mg-acetate, 18 % PEG 8000); MjModA (250 mM NH_4Cl, 16 % PEG 3350); PhModA (100 mM Bicine-NaOH pH 9.0, 100 mM NaCl, 25 % PEG 600); PfModA (200 mM K-formate, 20 % PEG 3350). The protein to reservoir volume ratio was 1:1. Crystals were cryo-protected as follows prior to flash cooling in liquid nitrogen: AfModA (15 % glycerol); MaModA (30 % glycerol); MjModA (30 % ethylene glycol); PhModA (30 % PEG 600); PfModA (30 % glycerol). Diffraction data was collected at the protein crystallography beamline X06SA at the Swiss Light Source (SLS) and processed using programs from the HKL2000 package [12] or XDS [13]. SAD phases exploiting bound tungstate were used for structure solution throughout, collecting data either at the L_3 absorption edge of tungsten or at a high remote wavelength. The wavelengths used were 0.90018 Å for AfModA, 1.21440 Å for MaModA, 1.2134 Å for PfModA, 1.21420 Å for MjModA, and 1.21475 Å for PhModA. Data statistics are presented in Table 1. Initial phases were obtained using the program SHELXD [14] or CNS [15] and were refined using SHARP [16] in combination with the programs DM [17] and Solomon [18]. Model building was performed using the program O [19]. The atomic model of AfModA was first refined isotropically to convergence using CNS [15] against structure factor amplitudes resulting in R_{work} and R_{free} of

0.1894 and 0.1986, respectively. Further refinement (using anisotropic B-factors) was carried out against the measured intensities using the program SHELXL [14] in the conjugate gradient (CGLS) mode. All residues were in the most favored or additional allowed regions of the Ramachandran plot except for Ala111 in chain A which exhibited a distinct backbone conformation from the equivalent residue in chain B. Refinement statistics are summarized in Table 1. For the other ModA/WtpA proteins, refinement and automatic water building were performed using CNS [15]. Individual B-factors were refined throughout and all residues were in the most favored or additional allowed regions of the Ramachandran plot, except for one MaModA residue, Ser215, which falls into a generously allowed region.

XAS measurements. For the X-ray absorption measurements full-length *A. fulgidus* ModA/WtpA (residues 32-342) was used instead of the C-terminally truncated version used for the crystallographic experiments. Concentrated protein (4.5 mM) in 20 mM Tris-HCl pH 8.0, 150 mM NaCl was incubated with a sub-stoichiometric amount of tungstate (final concentration 3 mM) and compared to 3 mM tungstate ($Na_2WO_4 \cdot 2\ H_2O$) in the same buffer. The L_3-edge tungsten X-ray absorption spectra were recorded at beamline D2 of the EMBL Outstation Hamburg at DESY, Germany. The DORIS storage ring operated at 4.5 GeV with the positron beam current ranging from 145 to 80 mA. A Si(*111*) double-crystal monochromator scanned X-ray energies around the tungsten L_3-edge (9.8–11.0 keV). Harmonic rejection was achieved by a focusing mirror (cut-off energy at 20.5 keV) and a monochromator detuning to 50 % of its peak intensity. The sample cells were mounted in a two-stage Displex cryostat and kept at about 20 K. The X-ray absorption spectra were recorded as tungsten L_3 fluorescence spectra with a Canberra 13-element germanium solid-state detector. While processing a pulse, the detector is frozen (dead time). We ensured that no more than 20 % of the counts occurred in this period and corrected each data point for this effect. For each sample, 15-20 scans were collected and averaged to ensure comparable statistics. X-ray energy calibration was achieved by recording

Bragg reflections from a static Si(*220*) crystal in back-reflection geometry during each scan [20]. Data reduction, such as background removal, normalization, and extraction of the fine structure, was performed with KEMP [21], assuming a threshold energy E_0 for W of 10.210 eV. Sample integrity during exposure to synchrotron radiation was checked by monitoring the position and shape of the absorption edge on sequential scans. No change in redox state or metal environment was detectable.

Extended X-ray absorption fine structure (EXAFS) analysis. The extracted tungsten L_3-edge (25–800 eV) EXAFS data were converted to photoelectron wave vector *k*-space and weighted by k^3. The spectra were analyzed with EXCURVE [22, 23] version 9.272. The program calculated the theoretical EXAFS for defined structural models based on the curved wave theory. Parameters of each structural model, namely the atomic distances (R), the Debye-Waller factors ($2\sigma^2$), and a residual shift of the threshold energy (EF), were refined, minimizing the fit index (FI) [22-24]. An amplitude reduction factor of 1.0 was used throughout the data analysis. For comparison with the crystallographic coordinates of AfModA, the coordinates of chain A were read into the EXAFS simulation program with W as the central atom. Before refinement, the W-O distances that were not expected to be resolved in the EXAFS measurements were lumped together and averaged: the distances W-O_1, W-O_2, and W-O_3 were averaged to 1.71 Å, the distances W-$O^{\varepsilon 2}$ (Glu218) and W-$O^{\delta 1}$ (Asp153) were averaged to 2.23 Å, while the W-O_4 distance of 2.00 Å was refined independently. The spatial arrangement of the atoms was taken from the crystallographic coordinates, and the multiple scattering was calculated on the basis of 3 units, each including the central atom with a pair of *trans* ligands. The number of free parameters was always inspected to be less than the number of independent data points [25]. The reduced χ^2 test was used to verify the significance of an additional ligand contribution [22-24].

Results

We cloned, expressed, purified and crystallized five homologous ModA/WtpA proteins from the archaea *A. fulgidus* (abbreviated as AfModA), *M. acetivorans* (MaModA), *M. jannaschii* (MjModA), *P. horikoshii* (PhModA), and *P. furiosus* (PfModA). Truncations were applied to these proteins as indicated in Fig. 1. These improved the stability of the proteins in solution and also facilitated better lattice contacts, thus providing better crystals that diffracted to higher resolution. The resolutions of the crystallographic studies were all better than 2 Å, with the lowest being 1.8 Å for the *P. horikoshii* homolog and the highest being 1.07 Å for the *A. fulgidus* protein (Table 1). The fold of the five ModA/WtpA proteins is highly similar to those of the previously studied molybdate-binding proteins ModA from *E. coli* and *A. vinelandii*, and to that of the plant pathogen *X. citri* [8-10]. Comparison of the eight structures reveals only one significant alteration of the fold: the five homologs described in this study contain an additional, four-stranded beta-sheet for which no equivalent is found in the proteins from *E. coli, A. vinelandii* and *X. citri* (Fig. 2). However, these inserted secondary structure elements are located on the surface of one of the halves of the protein and, thus, do not affect the architecture of the rest of the molecule nor do they change the arrangement of the oxyanion binding site (Fig. 2a).

Whereas all other molybdate- or tungstate-binding proteins described in the literature feature tetrahedral coordination of the metal, the tungsten atoms in the ModA/WtpA proteins described here all have six oxygen ligands in an octahedral coordination. Even though this geometry was observed at lower resolution for the AfModA protein earlier [6], the high resolution and excellent quality of the data allow, for the first time, a precise and detailed description of the unusual tungstate binding site at atomic resolution. Based on the crystallographic data a precision of 0.036 Å for the position of an atom with average B-factor (19.8 $Å^2$) can be estimated [26, 27]. However, the positional errors for the atoms in the oxyanion are clearly lower (average B-factor of the tungsten atom and its six oxygen ligands is 9.8 $Å^2$), and therefore, the

precision of the crystal structure and the EXAFS measurements discussed below are estimated to be similar.

In the *A. fulgidus* ModA/WtpA protein, the coordination sphere of the tungsten atom consists of three terminal (O_{1-3}), one protonated (O_4) and two carboxyl oxygens from the protein (Figs. 3 and 4). The carboxyl oxygen ligands are in *cis* disposition and are provided by the side-chains of Asp153 and Glu218, both located in the same lobe of the protein. These two acidic residues are replaced by hydrophobic side-chains of either Val or Ala in the binding pockets of the other ModA proteins (Fig. 2b) and, hence, the central metal atom is bound with a coordination number of four in the bacterial homologs (Fig. 3b). Inspection of the high-resolution structure of AfModA reveals that the tungsten atom is offset from the center of the octahedron and displaced away from the carboxyl oxygens, giving rise to a distorted octahedral geometry. The W-O distances vary significantly depending on the nature of the ligand and some of the bond angles differ strongly from the regular octahedron values (Fig. 3a and Table 2). The two W-O(carboxylate) bonds (2.18-2.24 Å) are much longer than the W-O distances involving terminal oxygen atoms (1.69-1.74 Å). This is in good agreement with the distance observed between the carboxyl oxygen of *R*-homocitrate and molybdenum in the FeMo-cofactor of *A. vinelandii* nitrogenase MoFe-protein and with W/Mo-O(carboxyl) bond lengths determined in the crystal structures of mandelato, glycolato, and *S*-lactato molybdates and tungstates [28-30]. The short distances between the metal and the three terminal oxygen atoms are indicative of a partial double bond character. As discussed by Zhou *et al*. [30] the stronger repulsion between such oxygen atoms results in increased bond angles between them and, in turn, compressed angles for bonds involving other atoms (Table 2). The increased length of the W-O_4 bond (1.92-2.00 Å) is probably caused by a postulated proton that we expect to be shared by oxygen O_4 and the second carboxylate oxygen $O^{\varepsilon 1}$ of Glu218. This intermediate distance complies well with the length of the bonds between the tungsten atom and the α-alkoxyl oxygens in mandelato and *S*-lactato tungstates and molybdates reported by Zhou *et al.* [29, 30], Cambridge Crystallographic Data

Centre (CCDC) reference codes 227929–227931, 203217–203219, 237819 and 237820. This supports the presence of a shared hydrogen atom that, at the available resolution, cannot be observed in the crystal structure. While the W-O distances involving the three terminal as well as the two carboxyl oxygens match well in the two non-crystallographically related AfModA monomers, the length of the W-O_4 bond does not and appears to be more flexible (Fig. 3a). Furthermore, the octahedral geometry is even more distorted than that of the mandelate complexes which feature two instead of one W-O distance in the 2.0 Å range. Like in the ModA proteins of *E. coli*, *A. vinelandii* and *X. citri* the oxygen ligands $O_{1\text{-}4}$ are bound to AfModA via several hydrogen bonds. Residues from both lobes of the protein contribute to ligand binding with a total of eight hydrogen bonds (Figs. 3a and 4). The hydrogen donors are the four backbone NH groups of Gly41, Ser42, Ser70, and Cys155 and the four side-chain hydroxyl groups of Ser42, Ser70, Glu218, and Tyr236. In the *E. coli* and *X. citri* homologs only seven such interactions are present with no equivalent for the O_4 to HN-Gly41 hydrogen bond (Fig. 3b). Although the structures of the other four ModA/WtpA homologs were determined at lower resolution, we nevertheless conclude that they have a very similar arrangement of their binding sites with distorted octahedral coordination of the metal centers.

Even though the crystallographic evidence for an octahedral tungsten coordination in ModA/WtpA proteins was convincing, we sought to obtain independent evidence for this geometry by studying the AfModA protein in solution rather than in a crystal lattice. To this end we recorded extended X-ray absorption fine structure (EXAFS) spectra of the tungstate-bound protein and, for comparison, of buffer-dissolved monomeric tungstate ($Na_2WO_4 \cdot 2\ H_2O$). The EXAFS traces (Fig. 5, upper panel) show interferences between different contributions, which is in part visualized by the peaks in the corresponding Fourier transforms (Fig. 5, lower panel). These are dominated by a peak at 1.7 Å representing the shell of oxygens closely bound to tungsten. For the protein spectrum (trace A), additional contributions at 2.2 and 3.7-4.0 Å are resolved which are absent in the spectrum of dissolved WO_4^{2-} (trace B). Both features are

characteristic of the distorted octahedral geometry found in the crystal structures of AfModA and the other four ModA/WtpA homologs. The 2.2 Å shell is evidence for the presence of longer W-O distances in that range. The 3.7-4.0 Å feature is due to the geometry-dependent multiple scattering (4th order) paths W-O_a-W-O_b-W, where O_a and O_b occupy opposite (*trans*) positions with respect to W. While such contributions are enhanced for O_a-W-O_b angles close to 180° as in an octahedral geometry, they are absent in complexes with tetrahedral disposition where all O-W-O angles are 109°. This is the case for monomeric tungstate in buffer (Fig. 5, traces B), which was satisfactorily simulated with four oxygen atoms placed at identical distances from the tungsten center without any assumptions regarding their spatial arrangement. In contrast to the AfModA EXAFS spectra, the previously reported EXAFS spectra of *A. vinelandii* ModA and of three cytoplasmic molybdate-binding proteins all lack the contribution at 2.2 Å as well as the features at 3.7-4.0 Å, indicating tetrahedral coordination geometry in these proteins [31]. Multiple scattering pathways going back and forth through the central atom have been observed before for both biological [32] and non-biological [33] square-planar nickel complexes. In the present case, however, their contribution is very dominant (Fig. S1), in particular in the high-frequency oscillations in the *k*-region 3-5 $Å^{-1}$ (Fig. 5, upper panel, trace A). The final simulation (black lines in Fig. 5, traces A) includes W-O contributions of 1.79 (3 O), 2.06 (1 O), and 2.24 Å (2 O). The contribution of the 2.06 Å W-O shell is not resolved in the Fourier transform, yet its inclusion is justified as the quality of the fit to the EXAFS data increases significantly (χ^2 value drops from $1.615 \cdot 10^{-6}$ to $1.0592 \cdot 10^{-6}$). Moreover, the position of the feature at 3.7-4.0 Å in the Fourier transform, ascribed to multiple scattering contributions, points to combinations of distances W-O_a and W-O_b of 1.8 + 2.2 Å and 1.8 + 2.0 Å, but not 1.8 + 1.8 Å or 2.2 + 2.2 Å. This is in line with the steric disposition of the ligands found in the crystal structure where the two carboxylate oxygens of Asp153 and Glu218 and the protonated oxygen (O_4) are in facial (*fac*) configuration with the three ligands each occupying coordination sites on the corners of a face of the coordination octahedron (Fig. 3).

1

Discussion

The recent discovery of octahedral molybdate and tungstate in the binding site of the periplasmic binding protein ModA/WtpA from the archaeon *A. fulgidus* came as a surprise [6]. Until then, coordination numbers above four had only been found in molybdo- and tungstoenzymes, where the Mo and W atoms are components of the catalytic cofactors and are coordinated by several, mostly non-oxygen ligands [1, 2]. Furthermore, in bacterial extracellular as well as intracellular proteins that bind MoO_4^{2-}, molybdenum is bound with a coordination number of four [31]. We solved the structures of five archaeal homologs of the bacterial periplasmic molybdate binding proteins and invariably found hexacoordinate tungsten in their binding pockets. This suggests distinct binding modes of the ModA/WtpA proteins from archaea and bacteria. While the previously determined structures of AfModA with bound tungstate and molybdate unambiguously showed that the two oxyanions were both bound with the same octahedral disposition, the resolution (1.60 Å and 1.55 Å, respectively) [6] was not sufficient for an in-depth analysis of the binding geometry. The 1.07 Å crystal structure presented here allows a detailed description of the distorted octahedral geometry including bond lengths and angles, which agree well with those determined from small molecule crystal structures of hexacoordinate monomeric tungstates and molybdates. In addition, our simulations of the X-ray absorption data based on the coordinates of the previously solved structure of tungstate-bound AfModA (PDB ID 2ONS) only poorly matched the EXAFS spectra (not shown), but drastically improved when the more accurate coordinates from the high-resolution structure became available. This indicates the absence of significant photo-reduction during the X-ray diffraction experiment of the protein crystals.

Enhancement of the coordination number and the associated change in geometry is not a requirement for high-affinity binding, as the bacterial homologs from *E. coli* and *X. citri* exhibit affinities with dissociation constants (K_D) in the low micromolar range [34, 35]. Yet, imposing a

change in coordination chemistry is an effective way to distinguish molybdate and tungstate from highly abundant sulfate (SO_4^{2-}), since, under physiological conditions, sulfur is unable to undergo this transition [36]. From the structures of *A. vinelandii* and *E. coli* ModA, where molybdate and tungstate have only four oxo ligands, it was concluded that these proteins discriminate between molybdate/tungstate and sulfate mainly by the sizes of these oxyanions [8, 9].

With the presence of two additional oxygen ligands, donated by acidic amino acid side-chains, the affinity for the substrate drastically increases. For *P. furiosus* ModA/WtpA, K_D values of 17 pM (tungstate) and 11 nM (molybdate) have recently been reported [7]. However, whether the rather large difference in affinity for these two oxyanions means that this transport system preferentially takes up tungstate remains to be elucidated. Furthermore, although high affinity is essential for effective capturing of substrates of low abundance such as trace metals, efficient transport may only occur if binding proteins readily release their cargo into the translocation pathway of their cognate ABC transporters. Further biochemical investigation of the transport process is needed to clarify this point.

References

1. Hille R (2002) Trends Biochem Sci 27:360-367
2. Johnson MK, Rees DC, Adams MW (1996) Chem Rev 96:2817-2840
3. Holland IB, Cole SPC, Kuchler K, Higgins CF (eds) (2003) ABC Proteins : From Bacteria to Man. Academic Press, London
4. Quiocho FA, Ledvina PS (1996) Mol Microbiol 20:17-25
5. Wilkinson AJ, Verschueren KHG (2003) In: Holland IB, Cole SPC, Kuchler K, Higgins CF (eds) ABC Proteins : From Bacteria to Man. Academic Press, London, pp. 187-207
6. Hollenstein K, Frei DC, Locher KP (2007) Nature 446:213-216
7. Bevers LE, Hagedoorn PL, Krijger GC, Hagen WR (2006) J Bacteriol 188:6498-6505
8. Hu Y, Rech S, Gunsalus RP, Rees DC (1997) Nat Struct Biol 4:703-707
9. Lawson DM, Williams CE, Mitchenall LA, Pau RN (1998) Structure 6:1529-1539
10. Balan A, Santacruz-Perez C, Moutran A, Ferreira LC, Neshich G, Goncalves Barbosa JA (2008) Biochim Biophys Acta 1784:393-399
11. Strange RW, Feiters MC (2008) Curr Opin Struct Biol 18:609-616
12. Otwinowski Z, Minor W, Charles W. Carter, Jr. (1997) Methods Enzymol 276:307-326
13. Kabsch W (1993) J Appl Crystallogr 26:795-800
14. Sheldrick GM (2008) Acta Crystallogr A 64:112-122
15. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Acta Crystallogr D Biol Crystallogr 54:905-921
16. de La Fortelle E, Bricogne G, Charles W. Carter, Jr. (1997) Methods Enzymol 276:472-494
17. Cowtan KD, Main P (1996) Acta Crystallogr D Biol Crystallogr 52:43-48
18. Abrahams JP, Leslie AGW (1996) Acta Crystallogr D Biol Crystallogr 52:30-42
19. Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Acta Crystallogr A 47:110-119

20. Pettifer RF, Hermes C (1985) J Appl Crystallogr 18:404-412
21. Korbas M, Marsa DF, Meyer-Klaucke W (2006) Rev Sci Instrum 77 DOI 10.1063/1.2209954
22. Gurman SJ, Binsted N, Ross I (1984) J Phys C Solid State Phys 17:143-151
23. Gurman SJ, Binsted N, Ross I (1986) J Phys C Solid State Phys 19:1845-1861
24. Binsted N, Strange RW, Hasnain SS (1992) Biochemistry 31:12117-12125
25. Stern EA (1993) Phys Rev B Condens Matter 48:9825-9827
26. Cruickshank DW (1999) Acta Crystallogr D Biol Crystallogr 55:583-601
27. Blow DM (2002) Acta Crystallogr D Biol Crystallogr 58:792-797
28. Einsle O, Tezcan FA, Andrade SL, Schmid B, Yoshida M, Howard JB, Rees DC (2002) Science 297:1696-1700
29. Zhou ZH, Hou SY, Cao ZX, Wan HL, Ng SW (2004) J Inorg Biochem 98:1037-1044
30. Zhou ZH, Zhao H, Tsai KR (2004) J Inorg Biochem 98:1787-1794
31. Duhme AK, Meyer-Klaucke W, White DJ, Delarbre L, Mitchenall LA, Pau RN (1999) J Biol Inorg Chem 4:588-592
32. Ha SW, Korbas M, Klepsch M, Meyer-Klaucke W, Meyer O, Svetlitchnyi V (2007) J Biol Chem 282:10639-10646
33. Metselaar GA, Schwartz E, de Gelder R, Feiters MC, Nikitenko S, Smolentsev G, Yalovega GE, Soldatov AV, Cornelissen JJ, Rowan AE, Nolte RJ (2007) Chemphyschem 8:1850-1856
34. Rech S, Wolin C, Gunsalus RP (1996) J Biol Chem 271:2557-2562
35. Balan A, Santacruz CP, Moutran A, Ferreira RC, Medrano FJ, Perez CA, Ramos CH, Ferreira LC (2006) Protein Expr Purif 50:215-222
36. Fraústo da Silva JJR, Williams RJP (2005) The biological chemistry of the elements: The inorganic chemistry of life. Oxford University Press, Oxford
37. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) J Mol Biol 340:783-795

38. Holm L, Park J (2000) Bioinformatics 16:566-567

Acknowledgments. We thank the beamline staff at the Swiss Light Source (SLS) for assistance with diffraction data collection, and D. C. Rees for discussions. This work was supported by the Roche Research Fund, the National Center for Competence in Research (NCCR) Structural Biology Zurich, the Swiss National Science Foundation, and the European Commission, Research Infrastructure Action under FP6 "Structuring the European Research Area Specific Programme", Contract # RII3-CT-2004-506008 for supporting access to the EMBL.

Coordinates. The coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under the PDB IDs 3CIJ (AfModA), 3CFX (MaModA), 3CG1 (PfModA), 3CFZ (MjModA), and 3CG3 (PhModA).

Table 1: Diffraction data collection and refinement statistics

	AfModA[a]	MaModA[a]	PfModA[a]	MjModA[a]	PhModA[a]
Data Collection					
Space group	$P2_12_12_1$	$C2$	$P2_1$	$P2_12_12_1$	$P2_12_12_1$
Molecules per AU	2	2	2	1	1
Cell dimensions					
a, b, c (Å)	73.901	115.500	46.460	36.954	37.880
	75.396	51.860	109.792	59.180	89.589
	116.996	124.048	55.430	150.759	97.868
α, β, γ (°)	90.000	90.000	90.000	90.000	90.000
	90.000	116.933	96.221	90.000	90.000
	90.000	90.000	90.000	90.000	90.000
Resolution (Å)[b]	30.0-1.07	30.0-1.6	30.0-1.6	30.0-1.7	30.0-1.8
	(1.09-1.07)	(1.7-1.6)	(1.7-1.6)	(1.75-1.7)	(1.85-1.8)
R_{meas} or R_{sym} (%)[b]	11.8 (39.1)	4.9 (27.6)	8.7 (19.6)	12.8 (21.6)	15.1 (36.1)
$I/\sigma I$[b]	16.8 (5.1)	15.3 (5.0)	9.5 (5.8)	15.0 (11.5)	16.3 (7.7)
Completeness (%)[b]	99.8 (99.3)	97.1 (95.3)	96.9 (98.0)	99.5 (98.4)	100.0 (100.0)
Redundancy[b]	9.1 (6.6)	3.3 (3.3)	2.6 (2.7)	11.8 (11.9)	13.5 (13.4)
Refinement					
Refinement program(s) used	CNS, SHELXL-97	CNS	CNS	CNS	CNS
Resolution (Å)	10.0-1.07	30.0-1.6	30.0-1.6	30.0-1.7	30.0-1.8
No. of reflections	270746/14081	156891/8211	126782/13648	64186/4812	54468/4037
R_{work}/ R_{free} (%)	12.64[c] /14.93	21.03/23.53	18.63/20.46	18.52/20.04	19.90/22.75
No. of atoms					
Protein	4827	4566	4740	2343	2532
WO_4^{2-}	10	10	10	5	5
Water	824	424	525	316	217
Mean *B*-factors (Å^2)					
Protein	17.3	26.0	18.7	11.5	24.3
WO_4^{2-}	9.9	20.1	10.9	12.2	16.2
Water	34.6	37.3	39.7	28.0	35.9
R.m.s deviations					
Bond lengths (Å)	0.007	0.0104	0.0097	0.0119	0.0112
Bond angles (°)	0.021[d]	1.7	1.7	1.6	1.8

[a] The first two letters of the protein name denote the source organism, see Methods.
[b] Highest resolution shell is shown in parenthesis.
[c] This *R*-factor was calculated including all data and is therefore called *R* instead of R_{work}.
[d] R.m.s. deviation for angle distances in Å calculated by SHELXL-97.

Table 2: Oxyanion geometry in AfModA

	Chain A (°)	Chain B (°)	Average (°)	Δ (°)
O_1–W–O_2	98.7	103.4	101.1	4.7
O_1–W–O_3	100.2	100.3	100.3	0.1
O_1–W–O_4	157.6	156.0	156.8	1.6
O_1–W–$O^{\delta1}Asp^{153}$	81.5	82.0	81.8	0.5
O_1–W–$O^{\varepsilon2}Glu^{218}$	85.0	83.9	84.5	1.1
O_2–W–O_3	102.3	102.5	102.4	0.2
O_2–W–O_4	92.2	90.0	91.1	2.2
O_2–W–$O^{\delta1}Asp^{153}$	86.6	84.5	85.6	2.1
O_2–W–$O^{\varepsilon2}Glu^{218}$	170.4	167.6	169.0	2.8
O_3–W–O_4	96.6	96.0	96.3	0.6
O_3–W–$O^{\delta1}Asp^{153}$	170.5	171.7	171.1	1.2
O_3–W–$O^{\varepsilon2}Glu^{218}$	85.6	85.8	85.7	0.2
O_4–W–$O^{\delta1}Asp^{153}$	79.7	79.5	79.6	0.2
O_4–W–$O^{\varepsilon2}Glu^{218}$	81.5	79.9	80.7	1.6
$Asp^{153}O^{\delta1}$–W–$O^{\varepsilon2}Glu^{218}$	85.2	86.5	85.9	1.3

Figure legends

Fig. 1 - Schematic of the expression constructs used for the crystallographic studies. The grey surfaces depict the parts of the proteins expressed and crystallized. Numbers above the bars indicate the amino acid positions. Signal sequences were predicted using the program SignalP [37].

Fig. 2 - Structural comparison and sequence alignment. **a**, *A. fulgidus* ModA/WtpA (*left panel*) is compared to ModA from *E. coli* (PDB ID 1WOD). The main structural differences are the distorted octahedral coordination of the oxyanion and an additional beta sheet (colored red) in AfModA. **b**, Structure-based alignment of the various ModA/WtpA sequences determined to date. The alignment was performed using DALI-lite [38] followed by minor manual adjustments. Secondary structure motifs are indicated above the sequences, and conserved residues involved in binding of substrate are highlighted in light blue for ModA/WtpA proteins discussed in this study (octahedral coordination) and in light green for ModA proteins with tetrahedral coordination.

Fig. 3 - Schematic of distinct oxyanion binding sites in related binding proteins ModA/WtpA from *A. fulgidus* (**a**) and ModA from *E. coli* (**b**, adapted from [8]). The schematic highlights the distinct tungsten coordination. Distances are given in angstroms for the two non-crystallographically related AfModA molecules (upper line, chain A; lower line, chain B). For a compilation of the bond angles see Table 2.

Fig. 4 - Stereo figure of the experimental electron density of the tungstate binding site of AfModA at 1.07 Å resolution, contoured at 3.5 σ (blue mesh) and at 60 σ(red mesh). For clarity only electron density around amino acid residues directly involved in the binding of WO_4^{2-} is

shown. The color code for the atoms is: black=carbon, red=oxygen, blue=nitrogen, yellow=sulfur, green=tungsten. Labels indicate amino acid residues in single letter code.

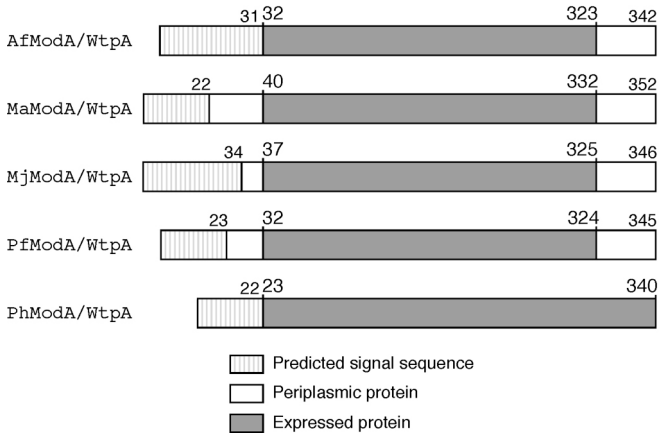
Fig. 5 - k^3-weighted EXAFS spectra (*upper panel*) and corresponding phase-corrected (assuming O in main shell) Fourier transforms (*lower panel*) at the W L_3 edge of tungstate in *A. fulgidus* ModA/WtpA (trace A) and buffer (trace B). Red line, experimental data; black line, simulation with the following parameters (distances in Å, Debye-Waller-type factors as $2\sigma^2$ in $Å^2$ in parentheses, Fit Index with k^3-weighting): trace A, EF -14.619 V, 3 O at 1.786 (0.008), 1 O at 2.060 (0.002), 2 O at 2.241 (0.003), FI $0.2998 \cdot 10^{-3}$; trace B, EF -13.54 V, 3.7 O at 1.770 (0.005), FI $0.1406 \cdot 10^{-3}$.

1 **Figures**

2

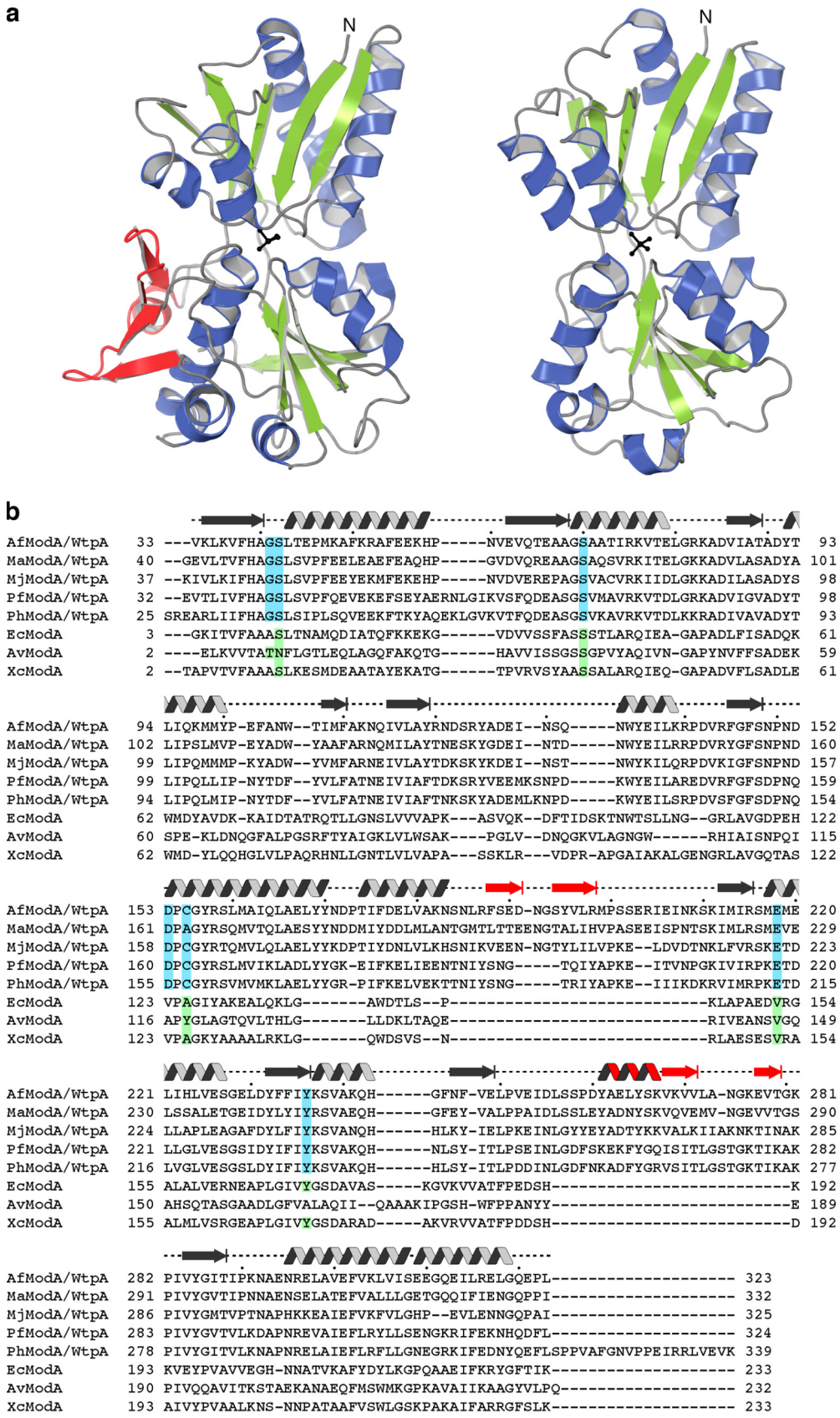
3 Fig. 1

4



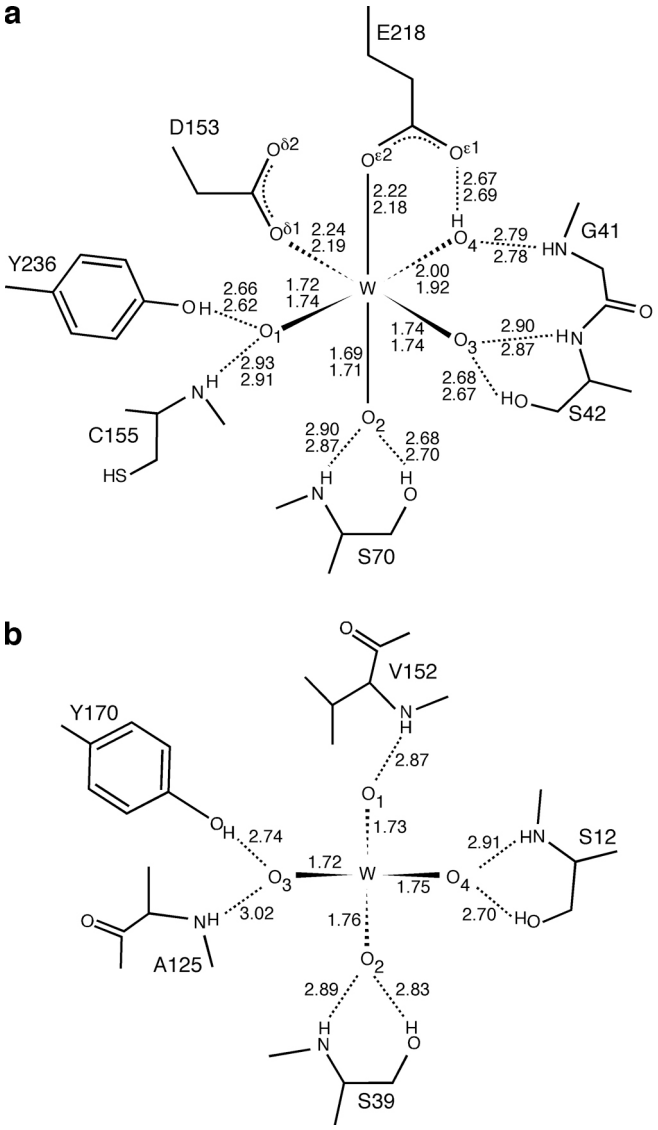
5
6

Fig. 2


a
N
N
b
AfModA/WtpA 33 ---VKLKVFHAGSLTEPMKAFKRAFEEKHP-----NVEVQTEAAGSAATIRKVTELGRKADVIATADYT 93
MaModA/WtpA 40 --GEVLTVFHAGSLSVPFEELEAEFEAQHP-----GVDVQREAAGSAQSVRKITELGKKADVLASADYA 101
MjModA/WtpA 37 --KIVLKIFHAGSLSVPFEEYEKMFEKEHP-----NVDVEREPAGSVACVRKIIDLGKKADILASADYS 98
PfModA/WtpA 32 --EVTLIVFHAGSLSVPFQEVEKEFSEYAERNLGIKVSFQDEASGSVMAVRKVTDLGRKADVIGVADYT 98
PhModA/WtpA 25 SREARLIIFHAGSLSIPLSQVEEKFTKYAQEKLGVKVTFQDEASGSVKAVRKVTDLKKRADIVAVADYT 93
EcModA 3 ---GKITVFAAASLTNAMQDIATQFKKEKG------VDVVSSFASSSTLARQIEA-GAPADLFISADQK 61
AvModA 2 ----ELKVVTATNFLGTLEQLAGQFAKQTG------HAVVISSGSSGPVYAQIVN-GAPYNVFFSADEK 59
XcModA 2 --TAPVTVFAAASLKESMDEAATAYEKATG------TPVRVSYAASSALARQIEQ-GAPADVFLSADLE 61
AfModA/WtpA 94 LIQKMMYP-EFANW--TIMFAKNQIVLAYRNDSRYADEI--NSQ-----NWYEILKRPDVRFGFSNPND 152
MaModA/WtpA 102 LIPSLMVP-EYADW--YAAFARNQMILAYTNESKYGDEI--NTD-----NWYEILRRPDVRYGFSNPND 160
MjModA/WtpA 99 LIPQMMMP-KYADW--YVMFARNEIVLAYTDKSKYKDEI--NST-----NWYKILQRPDVKIGFSNPND 157
PfModA/WtpA 99 LIPQLLIP-NYTDF--YVLFATNEIVIAFTDKSRYVEEMKSNPD-----KWYEILAREDVRFGFSDPNQ 159
PhModA/WtpA 94 LIPQLMIP-NYTDF--YVLFATNEIVIAFTNKSKYADEMLKNPD-----KWYEILSRPDVSFGFSDPNQ 154
EcModA 62 WMDYAVDK-KAIDTATRQTLLGNSLVVVAPK---ASVQK--DFTIDSKTNWTSLLNG--GRLAVGDPEH 122
AvModA 60 SPE-KLDNQGFALPGSRFTYAIGKLVLWSAK----PGLV--DNQGKVLAGNGW------RHIAISNPQI 115
XcModA 62 WMD-YLQQHGLVLPAQRHNLLGNTLVLVAPA---SSKLR---VDPR-APGAIAKALGENGRLAVGQTAS 122
AfModA/WtpA 153 DPCGYRSLMAIQLAELYYNDPTIFDELVAKNSNLRFSED-NGSYVLRMPSSERIEINKSKIMIRSMEME 220
MaModA/WtpA 161 DPAGYRSQMVTQLAESYYNDDMIYDDLMLANTGMTLTTEENGTALIHVPASEEISPNTSKIMLRSMEVE 229
MjModA/WtpA 158 DPCGYRTQMVLQLAELYYKDPTIYDNLVLKHSNIKVEEN-NGTYLILVPKE--LDVDTNKLFVRSKETD 223
PfModA/WtpA 160 DPCGYRSLMVIKLADLYYGK-EIFKELIEENTNIYSNG-----TQIYAPKE--ITVNPGKIVIRPKETD 220
PhModA/WtpA 155 DPCGYRSVMVMKLAELYYGR-PIFKELVEKTTNIYSNG-----TRIYAPKE--IIIKDKRVIMRPKETD 215
EcModA 123 VPAGIYAKEALQKLG-------AWDTLS--P----------KLAPAEDVRG 154
AvModA 116 APYGLAGTQVLTHLG-------LLDKLTAQE----------RIVEANSVGQ 149
XcModA 123 VPAGKYAAAALRKLG-------QWDSVS--N----------RLAESESVRA 154
AfModA/WtpA 221 LIHLVESGELDYFFIYKSVAKQH------GFNF-VELPVEIDLSSPDYAELYSKVKVVLA-NGKEVTGK 281
MaModA/WtpA 230 LSSALETGEIDYLYIYRSVAEQH------GFEY-VALPPAIDLSSLEYADNYSKVQVEMV-NGEVVTGS 290
MjModA/WtpA 224 LLAPLEAGAFDYLFIYKSVANQH------HLKY-IELPKEINLGYYEYADTYKKVALKIIAKNKTINAK 285
PfModA/WtpA 221 LLGLVESGSIDYIFIYKSVAKQH------NLSY-ITLPSEINLGDFSKEKFYGQISITLGSTGKTIKAK 282
PhModA/WtpA 216 LVGLVESGSLDYIFIYKSVAKQH------HLSY-ITLPDDINLGDFNKADFYGRVSITLGSTGKTIKAK 277
EcModA 155 ALALVERNEAPLGIVYGSDAVAS-----KGVKVVATFPEDSH----------K 192
AvModA 150 AHSQTASGAADLGFVALAQII--QAAAKIPGSH-WFPPANYY----------E 189
XcModA 155 ALMLVSRGEAPLGIVYGSDARAD-----AKVRVVATFPDDSH----------D 192
AfModA/WtpA 282 PIVYGITIPKNAENRELAVEFVKLVISEEGQEILRELGQEPL---------- 323
MaModA/WtpA 291 PIVYGVTIPNNAENSELATEFVALLLGETGQQIFIENGQPPI---------- 332
MjModA/WtpA 286 PIVYGMTVPTNAPHKKEAIEFVKFVLGHP--EVLENNGQPAI---------- 325
PfModA/WtpA 283 PIVYGVTVLKDAPNREVAIEFLRYLLSENGKRIFEKNHQDFL---------- 324
PhModA/WtpA 278 PIVYGITVLKNAPNRELAIEFLRFLLGNEGRKIFEDNYQEFLSPPVAFGNVPPEIRRLVEVK 339
EcModA 193 KVEYPVAVVEGH-NNATVKAFYDYLKGPQAAEIFKRYGFTIK---------- 233
AvModA 190 PIVQQAVITKSTAEKANAEQFMSWMKGPKAVAIIKAAGYVLPQ---------- 232
XcModA 193 AIVYPVAALKNS-NNPATAAFVSWLGSKPAKAIFARRGFSLK---------- 233

1 Fig. 3

2

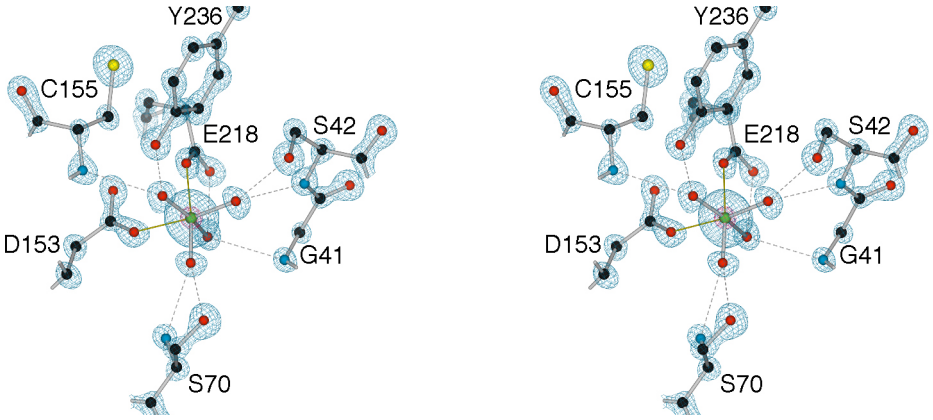

a
E218
D153
Oδ2
Oε2
Oε1
2.67
2.69
2.22
2.18
Oδ1
2.24
2.19
H
O4
2.79
2.78
HN
G41
Y236
2.66
2.62
1.72
1.74
W
2.00
1.92
OH
O1
1.74
1.74
O3
2.90
2.87
HN
O
1.69
1.71
2.93
2.91
NH
2.68
2.67
HO
S42
N
C155
HS
O2
2.90
2.87
2.68
2.70
H
H
N
O
S70
b
O
V152
N
H
2.87
Y170
O1
1.73
OH
2.74
2.91
HN
S12
1.72
W
1.75
O4
O3
O
NH
3.02
1.76
2.70
HO
A125
O2
2.89
2.83
H
H
N
O
S39

3

4

1 Fig. 4

2

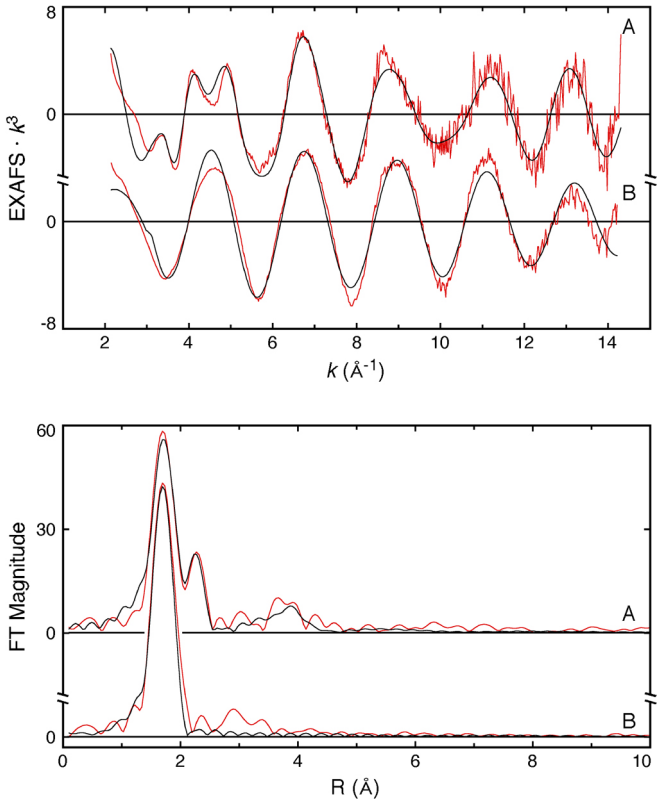

Y236
C155
E218
S42
D153
G41
S70
Y236
C155
E218
S42
D153
G41
S70

3

4

1 Fig. 5

2



3

4