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 ($\text{MoO}_4^{2-}$) and tungstate ($\text{WO}_4^{2-}$). 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 \((\text{MoO}_4^{2-})\) and tungstate \((\text{WO}_4^{2-})\). 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 \(\text{Archaeoglobus fulgidus}\) 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 \(\text{Escherichia coli}\) and \(\text{Azotobacter vinelandii}\), we had termed the protein ModA, and the corresponding transporter ModBC. At present it remains unknown if \(\text{A. fulgidus}\) accumulates \(\text{MoO}_4^{2-}\), \(\text{WO}_4^{2-}\), 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 \(\text{WO}_4^{2-}\).
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 (A600) 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$_2$WO$_4$. 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$_2$WO$_4$ 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$_2$SO$_4$, 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$_4$Cl, 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 (Na₂WO₄·2 H₂O) in the same buffer. The L₃-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 L₃-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 L₃ 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 L₃-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-O₁, W-O₂, and W-O₃ were averaged to 1.71 Å, the distances W-O^{G2} (Glu218) and W-O^{S1} (Asp153) were averaged to 2.23 Å, while the W-O₄ 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 $\chi^2$ 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 Å²) 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 Å²), 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 (O1-3), one protonated (O4) 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-O4 bond (1.92-2.00 Å) is
probably caused by a postulated proton that we expect to be shared by oxygen O4 and the second
carboxylate oxygen Oe1 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-O₄ 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 O₁−₄ 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 O₄ 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 (Na₂WO₄·2 H₂O). 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 WO₄²⁻ (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-Oa-W-Ob-W, where Oa and Ob occupy opposite (trans) positions with respect to W. While such contributions are enhanced for Oa-W-Ob 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 (χ² value drops from 1.615·10⁻⁶ to 1.0592·10⁻⁶). 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-Oa and W-Ob 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 (O₄) 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).
Hollenstein, Comellas-Bigler et al. Tungstate binding geometry in ModA/WtpA proteins
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 tungstenzymes, 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 $\text{MoO}_4^{2-}$, 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

Hollenstein, Comellas-Bigler et al.  Tungstate binding geometry in ModA/WtpA proteins

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

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>AfModA&lt;sup&gt;a&lt;/sup&gt;</th>
<th>MaModA&lt;sup&gt;a&lt;/sup&gt;</th>
<th>PtModA&lt;sup&gt;a&lt;/sup&gt;</th>
<th>MjModA&lt;sup&gt;a&lt;/sup&gt;</th>
<th>PhModA&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2&lt;sub&gt;1&lt;/sub&gt;,P2&lt;sub&gt;1&lt;/sub&gt;,P2&lt;sub&gt;1&lt;/sub&gt;</td>
<td>C2</td>
<td>P2&lt;sub&gt;1&lt;/sub&gt;</td>
<td>P2&lt;sub&gt;1&lt;/sub&gt;,P2&lt;sub&gt;1&lt;/sub&gt;,P2&lt;sub&gt;1&lt;/sub&gt;</td>
<td>P2&lt;sub&gt;1&lt;/sub&gt;,P2&lt;sub&gt;1&lt;/sub&gt;,P2&lt;sub&gt;1&lt;/sub&gt;</td>
</tr>
<tr>
<td>Molecules per AU</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>73.901</td>
<td>115.500</td>
<td>46.460</td>
<td>36.954</td>
<td>37.880</td>
</tr>
<tr>
<td></td>
<td>75.396</td>
<td>51.860</td>
<td>109.792</td>
<td>59.180</td>
<td>89.589</td>
</tr>
<tr>
<td></td>
<td>116.996</td>
<td>124.048</td>
<td>55.430</td>
<td>150.759</td>
<td>97.868</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90.000</td>
<td>90.000</td>
<td>90.000</td>
<td>90.000</td>
<td>90.000</td>
</tr>
<tr>
<td></td>
<td>90.000</td>
<td>116.933</td>
<td>96.221</td>
<td>90.000</td>
<td>90.000</td>
</tr>
<tr>
<td>Resolution (Å)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>30.0-1.07</td>
<td>30.0-1.6</td>
<td>30.0-1.6</td>
<td>30.0-1.7</td>
<td>30.0-1.8</td>
</tr>
<tr>
<td></td>
<td>(1.09-1.07)</td>
<td>(1.7-1.6)</td>
<td>(1.7-1.6)</td>
<td>(1.75-1.7)</td>
<td>(1.85-1.8)</td>
</tr>
<tr>
<td>R&lt;sub&gt;meas&lt;/sub&gt; or R&lt;sub&gt;sym&lt;/sub&gt; (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>11.8 (39.1)</td>
<td>4.9 (27.6)</td>
<td>8.7 (19.6)</td>
<td>12.8 (21.6)</td>
<td>15.1 (36.1)</td>
</tr>
<tr>
<td>R&lt;sub&gt;iso&lt;/sub&gt;/R&lt;sub&gt;free&lt;/sub&gt;&lt;sup&gt;c&lt;/sup&gt; (%)</td>
<td>16.8 (5.1)</td>
<td>15.3 (5.0)</td>
<td>9.5 (5.8)</td>
<td>15.0 (11.5)</td>
<td>16.3 (7.7)</td>
</tr>
<tr>
<td>Completeness (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>99.8 (99.3)</td>
<td>97.1 (95.3)</td>
<td>96.9 (98.0)</td>
<td>99.5 (98.4)</td>
<td>100.0 (100.0)</td>
</tr>
<tr>
<td>Redundancy&lt;sup&gt;b&lt;/sup&gt;</td>
<td>9.1 (6.6)</td>
<td>3.3 (3.3)</td>
<td>2.6 (2.7)</td>
<td>11.8 (11.9)</td>
<td>13.5 (13.4)</td>
</tr>
</tbody>
</table>

Refinement

<table>
<thead>
<tr>
<th>Refinement program(s) used</th>
<th>CNS, SHELXL-97</th>
<th>CNS</th>
<th>CNS</th>
<th>CNS</th>
<th>CNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>10.0-1.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of reflections</td>
<td>270746/14081</td>
<td>156891/8211</td>
<td>126782/13648</td>
<td>64186/4812</td>
<td>54468/4037</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt;/R&lt;sub&gt;free&lt;/sub&gt; (%)</td>
<td>12.64&lt;sup&gt;d&lt;/sup&gt; /14.93</td>
<td>21.03/23.53</td>
<td>18.63/20.46</td>
<td>18.52/20.04</td>
<td>19.90/22.75</td>
</tr>
<tr>
<td>No. of atoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>4827</td>
<td>4566</td>
<td>4740</td>
<td>2343</td>
<td>2532</td>
</tr>
<tr>
<td>WO&lt;sub&gt;4&lt;/sub&gt;&lt;sup&gt;2-&lt;/sup&gt;</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Water</td>
<td>824</td>
<td>424</td>
<td>525</td>
<td>316</td>
<td>217</td>
</tr>
<tr>
<td>Mean B-factors (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>17.3</td>
<td>26.0</td>
<td>18.7</td>
<td>11.5</td>
<td>24.3</td>
</tr>
<tr>
<td>Protein</td>
<td>9.9</td>
<td>20.1</td>
<td>10.9</td>
<td>12.2</td>
<td>16.2</td>
</tr>
<tr>
<td>WO&lt;sub&gt;4&lt;/sub&gt;&lt;sup&gt;2-&lt;/sup&gt;</td>
<td>34.6</td>
<td>37.3</td>
<td>39.7</td>
<td>28.0</td>
<td>35.9</td>
</tr>
<tr>
<td>Water</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R.m.s deviations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.007</td>
<td>0.0104</td>
<td>0.0097</td>
<td>0.0119</td>
<td>0.0112</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>0.021&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1.7</td>
<td>1.7</td>
<td>1.6</td>
<td>1.8</td>
</tr>
</tbody>
</table>

<sup>a</sup>The first two letters of the protein name denote the source organism, see Methods.
<sup>b</sup>Highest resolution shell is shown in parenthesis.
<sup>c</sup>This R-factor was calculated including all data and is therefore called R instead of R<sub>work</sub>.
<sup>d</sup>R.m.s. deviation for angle distances in Å calculated by SHELXL-97.
Table 2: Oxyanion geometry in AfModA

<table>
<thead>
<tr>
<th></th>
<th>Chain A (°)</th>
<th>Chain B (°)</th>
<th>Average (°)</th>
<th>Δ (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O₁–W–O₂</td>
<td>98.7</td>
<td>103.4</td>
<td>101.1</td>
<td>4.7</td>
</tr>
<tr>
<td>O₁–W–O₃</td>
<td>100.2</td>
<td>100.3</td>
<td>100.3</td>
<td>0.1</td>
</tr>
<tr>
<td>O₁–W–O₄</td>
<td>157.6</td>
<td>156.0</td>
<td>156.8</td>
<td>1.6</td>
</tr>
<tr>
<td>O₁–W–O^{51}Asp^{153}</td>
<td>81.5</td>
<td>82.0</td>
<td>81.8</td>
<td>0.5</td>
</tr>
<tr>
<td>O₁–W–O^{52}Glu^{218}</td>
<td>85.0</td>
<td>83.9</td>
<td>84.5</td>
<td>1.1</td>
</tr>
<tr>
<td>O₂–W–O₁</td>
<td>102.3</td>
<td>102.5</td>
<td>102.4</td>
<td>0.2</td>
</tr>
<tr>
<td>O₂–W–O₄</td>
<td>92.2</td>
<td>90.0</td>
<td>91.1</td>
<td>2.2</td>
</tr>
<tr>
<td>O₂–W–O^{51}Asp^{153}</td>
<td>86.6</td>
<td>84.5</td>
<td>85.6</td>
<td>2.1</td>
</tr>
<tr>
<td>O₂–W–O^{52}Glu^{218}</td>
<td>170.4</td>
<td>167.6</td>
<td>169.0</td>
<td>2.8</td>
</tr>
<tr>
<td>Bond</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Value 4</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>$O_3$–$W$–$O_4$</td>
<td>96.6</td>
<td>96.0</td>
<td>96.3</td>
<td>0.6</td>
</tr>
<tr>
<td>$O_3$–$W$–$O^{\delta 1}$ Asp$^{153}$</td>
<td>170.5</td>
<td>171.7</td>
<td>171.1</td>
<td>1.2</td>
</tr>
<tr>
<td>$O_4$–$W$–$O^{\epsilon 2}$ Glu$^{218}$</td>
<td>85.6</td>
<td>85.8</td>
<td>85.7</td>
<td>0.2</td>
</tr>
<tr>
<td>$O_4$–$W$–$O^{\delta 1}$ Asp$^{153}$</td>
<td>79.7</td>
<td>79.5</td>
<td>79.6</td>
<td>0.2</td>
</tr>
<tr>
<td>$O_4$–$W$–$O^{\epsilon 2}$ Glu$^{218}$</td>
<td>81.5</td>
<td>79.9</td>
<td>80.7</td>
<td>1.6</td>
</tr>
<tr>
<td>Asp$^{153}$–$O^{\delta 1}$–$W$–$O^{\epsilon 2}$ Glu$^{218}$</td>
<td>85.2</td>
<td>86.5</td>
<td>85.9</td>
<td>1.3</td>
</tr>
</tbody>
</table>
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 1WUD). 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 0.2998·10$^{-3}$; trace B, EF -13.54 V, 3.7 O at 1.770 (0.005), FI 0.1406·10$^{-3}$. 

Hollenstein, Comellas-Bigler *et al.* Tungstate binding geometry in ModA/WtpA proteins
Figures

Fig. 1

AfModA/WtpA
MaModA/WtpA
MjModA/WtpA
PfModA/WtpA
PbModA/WtpA

Predicted signal sequence
Periplasmic protein
Expressed protein
Fig. 3

![Diagram of Tungstate binding geometry in ModA/WtpA proteins]

- Fig. 3
Hollenstein, Comellas-Bigler et al.  Tungstate binding geometry in ModA/WtpA proteins

Fig. 4
Fig. 5
Fig. S1 - $k^3$-weighted EXAFS spectrum (upper panel) and corresponding phase-corrected (assuming O in main shell) Fourier transform (lower panel) at the W L3 edge of tungstate in *A. fulgidus* ModA/WtpA. Red line, experimental data; black line, simulation including multiple scattering, as in main text, 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): 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$\cdot 10^{-3}$; Blue line, simulation with the same parameters but just single scattering; Green line, difference between the simulations, representing the multiple scattering contribution.