DALTON

Protein ligands for molybdate. Specificity and charge stabilisation at the anion-binding sites of periplasmic and intracellular molybdate-binding proteins of *Azotobacter vinelandii*‡

David M. Lawson, Clare E. Williams, Daniel J. White, Achille P. Choay, Lesley A. Mitchenall and Richard N. Pau*,†

Nitrogen Fixation Laboratory, John Innes Centre, Colney Lane, Norwich, UK NR4 7UH

Electrostatic interactions are important in the binding of anions to proteins. In Gram negative bacteria, molybdate binds specifically to a periplasmic binding protein and a number of cytoplasmic binding proteins. The molybdate-binding site in an *Azotobacter vinelandii* periplasmic binding protein has been determined at the atomic level from the crystal structure of the protein with bound tungstate at 1.2 Å resolution. The periplasmic molybdate-binding protein is very similar to the sulfate-binding protein of *Salmonella typhimurium*. In both, the anions are completely buried and bound by seven hydrogen bonds donated by main-chain and neutral residues at the ends of α-helices. The specificity of the two proteins for binding their respective anions may be related to small differences in the sizes of the anions and the lengths of the bonds formed. In the cytoplasm three distinct proteins have similar 7 kDa molybdate-binding domains. Secondary structure analysis indicates that the domains are all-β structures with anti-parallel β-strands. Analysis of molybdate binding by the cytoplasmic binding proteins suggests that, unlike the α/β periplasmic binding proteins, molybdate binding in these proteins involves electrostatic interactions with positively charged residues. These findings are important in understanding anion-binding in proteins of different structural classes.

Many proteins bind oxodianions. Analysis of the diverse binding sites for phosphate and sulfate shows that they tend to form one, two or three contacts with the protein, and that positively charged groups and polar residues are often involved.[1] By contrast, the highly specific binding of sulfate and phosphate to their cognate periplasmic binding proteins involves seven or twelve hydrogen bonds respectively to backbone NH and side chain residues. No ionic bonds are formed and the double negative charges on the anions are stabilised by multiple local dipoles.[2,3] Recognition of a proton in mono- or di-basic phosphate plays a key role in the very high specificity of the binding proteins. At physiological pH molybdenum and tungsten are present in the environment as the oxydianions molybdate, MoO_4^{2-}, and tungstate, WO_4^{2-}, respectively.[4] They resemble sulfate in charge, hydrogen-bonding properties and size.[5] The molecular basis for the specificity of molybdate-binding proteins is not known. Bacteria have two classes of molybdate-binding proteins: a periplasmic binding protein, and cytoplasmic proteins which bind molybdenum in a highly conserved 7 kDa domain.[6]

Herein we describe the atomic structure of the molybdate-binding site of a periplasmic molybdate-binding protein from *Azotobacter vinelandii* which has been determined to a resolution of 1.2 Å. This allows a detailed description of the binding site. Secondary structure predictions show that the cytoplasmic molybdenum-binding proteins or domains represent a distinct all-β class of metal-binding domain. Binding analysis and amino acid sequence comparisons allow inferences to be made about their binding-sites.

Results and Discussion

Periplasmic molybdate-binding protein of *Azotobacter vinelandii*

As in other periplasmic-binding proteins, the molybdate-binding protein consists of two similar domains (I and II) separated by a deep cleft.[7] The 232 amino acid polypeptide chain intertwines between the domains, which are each formed by a five-stranded β-sheet between two layers of α-helices (Fig. 1). The *A. vinelandii* molybdate-binding protein and the *Salmonella typhimurium* sulfate-binding protein have similar tertiary structures. A major difference is the absence in the molybdate-binding protein of a segment corresponding to 50 amino acids at the C-terminus of the sulfate-binding protein. These amino acids form one of three chains which link the two domains and form a hinge at the bottom of the cleft between the two domains of the sulfate-binding protein.

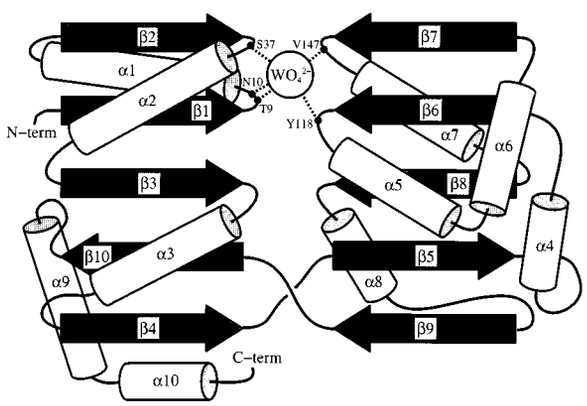

Fig. 1 Schematic diagram showing the topology of molybdate-binding protein. α-Helices are depicted as cylinders and β-strands are shown as arrows. Residues involved in anion binding are indicated

Molybdate binding site

The molybdate (and tungstate) oxydianions are completely dehydrated and bound by seven hydrogen bonds in the cleft of the molybdate-binding protein. The hydrogen bonds are all donated by uncharged polar groups on the protein. Four are from NH groups of main-chain peptide units [asparagine (Asn-10) and serine (Ser-37) in domain I and tyrosine (Tyr-118) and valine (Val-147) in domain II] and one each from the side

† E-Mail: pau@bbsrc.ac.uk

‡ Based on the presentation given at Dalton Discussion No. 2, 2nd–5th September 1997, University of East Anglia, UK.

Non-SI unit employed: Da = u ≈ 1.66×10^{-27} kg.

Table 1 Hydrogen bond lengths (Å) at the anion-binding site

Asn-10	$N \cdots WO_4^{2-}$	O(2)	2.93	Thr-9	$O\gamma1 \cdots WO_4^{2-}$	O(1)	2.73
	$N\delta2 \cdots WO_4^{2-}$	O(2)	2.79	Tyr-118	$N \cdots WO_4^{2-}$	O(4)	2.84
Ser-37	$N \cdots WO_4^{2-}$	O(3)	2.81	Val-147	$N \cdots WO_4^{2-}$	O(1)	2.84
	$O\gamma \cdots WO_4^{2-}$	O(3)	2.76				

Table 2 Geometry [bond lengths (Å) and angles (°)] of bound tungstate

O(1)–W	1.78	O(3)–W	1.77	O(1)–W–O(2)	111.08	O(2)–W–O(3)	107.67
O(2)–W	1.73	O(1)–W	1.77	O(1)–W–O(3)	111.03	O(2)–W–O(4)	118.24
				O(1)–W–O(2)	111.05	O(3)–W–O(4)	107.61

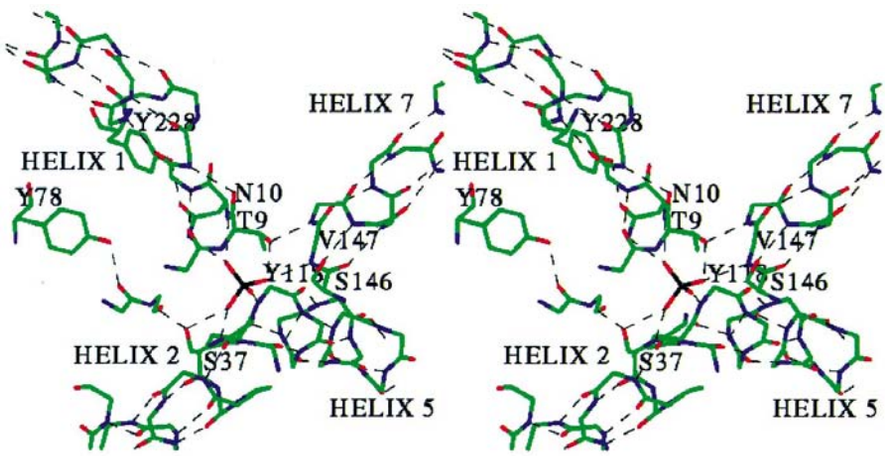

Fig. 2 Stereo figure produced by the program O showing hydrogen-bonding networks (dashed lines) extending away from the anion binding site in molybdate-binding protein.[8] For clarity only the side chains involved in these networks (with the exception of prolines) are displayed (key: green = carbon, blue = nitrogen, red = oxygen, black = tungsten). Important residues are labelled as close as possible to their α-carbons without compromising clarity

chains of Thr-9 (threonine), Asn-10 and Ser-37 (domain I). The hydrogen-bond donating groups are located at the amino-terminal ends of four α-helices (Fig. 2). The sulfate-binding protein similarly binds sulfate by donating seven hydrogen bonds to the oxo groups of the anion. Five are from main-chain NH groups, and two from side chains [Ser OH and tryptophan (Trp) NH].[3] The hydrogen-bond donor groups are all located at the amino-terminal ends of four helices that are in equivalent locations to those involved in binding molybdate in the molybdate-binding protein. Helix macrodipoles have been assumed to be involved in electrostatic interactions of proteins.[9] However, neutralisation of the positively charged residues that cap two of the binding helices of sulfate-binding protein, or that are coupled to peptide units close to the anion-binding helix termini, had little effect on sulfate binding.[10] This indicated that the charge on the anion is stabilised by highly localised charged dipoles at the helix termini rather than helix macrodipoles. In the molybdate-binding protein there are no positively charged residues coupled to the N-terminal ends of the binding helices, or capping the helices.

Determinants of specificity of the periplasmic molybdate-binding protein

The selectivity shown by the periplasmic-molybdate- and sulfate-binding proteins for their respective anions is intriguing. Since the anion is held entirely through interactions with its four oxygens, it is not immediately obvious which criteria are used by these proteins as the basis for selectivity. We believe that the size of the anion is crucial for successful complex formation since the protonation states at physiological pH and charges of the two anions are identical. The majority of the hydrogen bonds (five in sulfate-binding protein and four in molybdate-binding protein) involve main-chain nitrogens. Given that these all lie at the N-termini of α-helices, they will be quite inflexible. From high-resolution X-ray analysis of small molecule structures, the average distance between the central atom and the oxygens is 1.75 ± 0.04 Å and 1.76 ± 0.002 Å for molybdate and tungstate respectively, compared to 1.47 ± 0.02 Å for sulfate. The lengths of the hydrogen bonds and the geometry of bound tungstate are shown in Tables 1 and 2. When a sulfate anion is inserted into the binding pocket of the molybdate-binding protein model, or molybdate is inserted into the pocket of the sulfate-binding protein model, reoptimisation of the bond lengths could be obtained by slight tightening or slight loosening of the interface between domains I and II. However, reoptimisation of the bond angles would be less straightforward, requiring either a pivoting or a rolling of the α-helices involved. Therefore, we conclude that small changes in the size of the anion are sufficient to explain the differences between tight and weak binding, and therefore form the basis for selectivity in these proteins.

Intracellular molybdate-binding proteins

Three cytoplasmic proteins have been identified which bind molybdate after it has been transported across the bacterial inner cell membrane.[11] These proteins are similar to, or have a domain that is similar to, the tetrameric 28 kDa protein called Mop in *Clostridium pasteurianum*.[11] In *A. vinelandii* the homodimeric ModG has two tandem Mop-like sequences.[12] We term Mop and proteins that are similar to it molbindins. Two other distinct proteins have domains that are similar to Mop. They are: (1) the C-terminal domain of ModC, the peripheral membrane ATP-binding protein of the molybdate transporter and (2) the C-terminal half of molybdenum-responsive regulatory protein ModE, which has two tandem Mop-like repeats per dimer. Secondary structure analysis of 17 molbindins and Mop-like domains shows that they have an all-β secondary structure, composed of six (or possibly seven) antiparallel

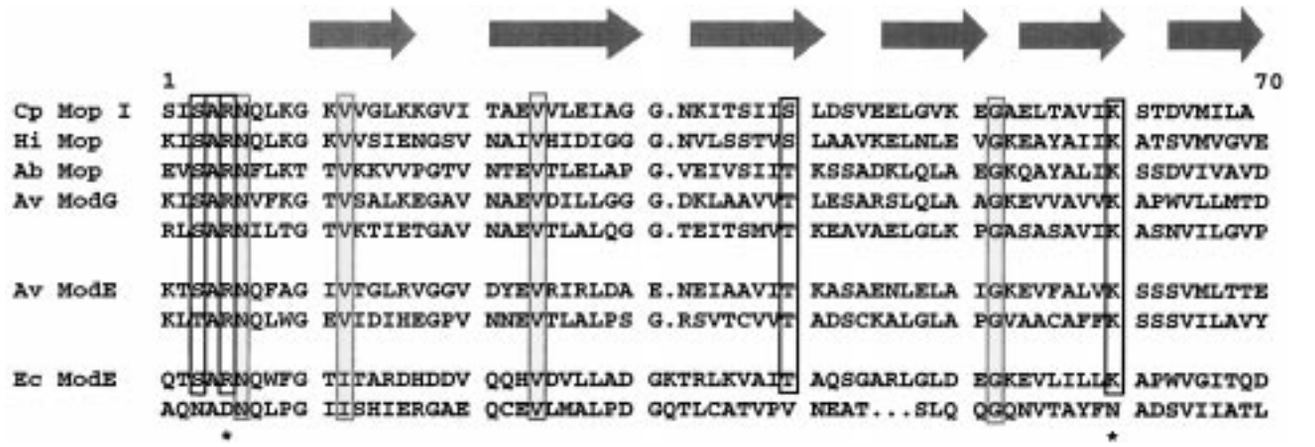

Fig. 3 Secondary structure prediction and amino acid sequence alignment of the Mop-like repeats of molbindins and the molybdenum-dependent regulatory ModE proteins from *A. vinelandii* and *E. coli*. The molbindins are: Cp Mop I, MopI from *Clostridium pasteurianum*; Hi Mop, Mop from *H. influenzae* and Av ModG, *A. vinelandii* ModG and ModE proteins. The ModG and the ModE proteins have two Mop-like repeats. The locations of predicted β-strands are indicated by arrows. Secondary structure was predicted from the alignment of 17 Mop-like sequences. The identical and similar residues conserved in all 17 protein sequences except the second repeat of *E. coli* ModE are shown in boxes with black lines. Other residues that are conserved in all Mop-like sequences are shown in boxes with grey lines. The positively charged residues, Arg and Lys, which are absent in the second repeat of *E. coli* ModE are labelled with asterisks

β-strands (Fig. 3). We have isolated the Mop protein from *Haemophilus influenzae*, ModG from *A. vinelandii*, and the C-terminal Mop-like domain from *A. vinelandii* ModE. Gel equilibrium chromatography and fluorescence titrations show that they bind four mol of Mo per four Mop-like domains.[13] In contrast to this, the four Mop-like domains of the dimeric *Escherichia coli* ModE only bind two mol of Mo.[14] This suggests that two of the four Mop-like domains of *E. coli* ModE are unable to bind Mo. Four similar or identical amino acids are conserved in all molbindins, in both Mop-like domains of *A. vinelandii* ModE and in the first Mop-like domain of *E. coli* ModE, but not in the second domain of *E. coli* and *H. influenzae* ModE. They are Ser/Thr, Arg (arginine), Ser/Thr and Lys (lysine) (Fig. 3). In the second domain of *E. coli* ModE they are changed: Ser to Asn, Arg to Asp (aspartic acid), Ser to Val, and Lys to Asn (Fig. 3). The amino acids which are changed in the second Mop-like domain of ModE include two positively charged residues, and are found in positions which are consistent with their being in a binding site at one end of an all-β protein (Fig. 3). They are located in an N-terminal conserved cluster of amino acids, and at the C-terminal ends of the predicted strands III and V. The N-terminal cluster also includes an invariant residue, Asn. This analysis indicates that a reasonable model for the structure of the molbindins is one in which molybdate is bound at a site in β-barrel protein that includes two positively charged residues. Analogous examples of all-β proteins with a bound metal are plastocyanin and azurin, in which the copper is bound within a β-barrel.[15]

Conclusion

The binding site in molybdate-binding protein provides further support for the important role of multiple localised dipoles at the end of helices in the specific binding in α/β periplasmic anion binding proteins. Secondary structure analysis shows that the cytoplasmic molybdate-binding proteins differ from the periplasmic binding proteins in being all-β proteins. Further, sequence and binding analysis suggest that, unlike molybdate-binding protein, positively charged residues are involved in binding molybdate at a site which may also include hydrogen bonds to hydroxy amino acids. This may be a consequence of the higher density of hydrogen bonds that are available at the ends of α-helices than in a β-sheet structure.

Experimental

X-Ray structure analysis

Molybdate-binding protein crystallises isomorphously with either molybdate or tungstate bound and the crystals diffract X-rays to high resolution. The space group is $P2_1$ with cell parameters $a = 32.90$, $b = 88.80$, $c = 41.75$ Å and $\beta = 93.55°$. Data were collected to 1.2 Å resolution at the European Molecular Biology Laboratory (EMBL) Deutsches Elektronen-Synchrotron (DESY) in Hamburg (beamline XII, $\lambda = 0.91$ Å) on the tungstate-bound form. Data to 2.0 Å resolution were collected 'in-house' ($\lambda = 1.54$ Å) on the molybdate-bound form. The data sets are extremely isomorphous and the tungstate data contained measurable anomalous differences. There is a single protein molecule per crystallographic asymmetric unit, which is confirmed by a clear single peak on the Harker section in both isomorphous and anomalous Patterson-difference maps. This has enabled us to solve the structure using the Single Isomorphous Replacement Anomalous Scattering (SIRAS) method at 2.0 Å resolution and refine the resultant model of the tungstate-bound form to 1.2 Å resolution. A more detailed description of the model and structure solution will be given elsewhere.

Purification of the molybdate-binding domain of *A. vinelandii* ModE

The DNA corresponding to the molybdate-binding domain of *A. vinelandii* ModE (amino acids 195–270) was amplified by polymerase chain reaction (PCR) using pMAC75 as a template.[12] The PCR products were inserted between the *Nde*I and *Bam*HI sites in the polylinker of the expression vector pET-15B (Novagen). The molybdate-binding domain was over-expressed in *E. coli* strain BL21(DE3).[16] Cells, grown at 37 °C in 3 dm^3 Luria–Bertani broth supplemented with 0.4% w/v glucose culture, were induced with 0.5 mM isopropyl β-D-thiogalactopyranoside, and harvested 3 h after induction. The cells were lysed by free-thawing in the presence of DNAase. The cell debris was removed by centrifugation at 18 000 g ($g \approx 9.806$ m s^{-2}) for 1 h. The supernatant was applied to a nickel-loaded metal chelate column (HiTrap: Pharmacia). The unbound proteins were washed from the column with 50 mM $(HOCH_2)_3C$-NH_2·HCl (TRIS·HCl), pH 7.6, 250 mM NaCl and the chelated protein eluted with a 0.5 mM imidazole gradient in the same buffer. The imidazole buffer was exchanged with TRIS·HCl pH 7.6 by gel filtration on Bio-Rad 6PG.

Estimation of molybdate binding sites

Fluoresence titrations were carried out by addition of small aliquots of sodium molybdate to the His_6-tagged Mo-binding proteins in 50 mM TRIS·HCl pH 7.6. The intrinsic fluoresence spectra were recorded with a Perkin-Elmer LS50B spectrofluor-

imeter thermostatted to 25 °C. Excitation was at 290 nm and emission at 353 nm. Gel equilibrium determinations of bound molybdate were carried out in a 110 × 10 mm diameter column of Bio-Gel P6 DG (Bio-Rad) equilibrated with 20 mM TRIS·HCl pH 7.6, 100 mM NaCl. Molybdenum was estimated by the method of Clark and Axley.[17]

Secondary structure analysis

Protein secondary prediction was carried out by the programs PHD,[18] TOPITS[19] and the PSA system.[20,21] The PHD prediction was based on a Multiple Sequence Format alignment produced by the GCG program PILEUP.[22] The following proteins or molybdate-binding domains were used for the alignment, SWISSPROT IDs: MOP1_CLOPA, MOP2_CLOPA, MOP3_CLOPA, MOP_HAEIN, MODG_AZOVI MODA_AZOVI, MODE_ECOLI, MODE_HAEIN, MOPA_RHOCA, MOPB_RHOCA and *Anabaena* Mop: EMBL database accession no. X95645.

Accession numbers

The coordinates and structure factor data for the *A. vinelandii* periplasmic molybdate-binding protein have been deposited with the Protein Data Bank with accession codes 1ATG and R1ATGSF, respectively.

Acknowledgements

This work has been supported by the BBSRC. We would like to acknowledge the use of the EPSRC funded Chemical Database Service at Daresbury. We thank the EU (HCMP LIP grant contract no. CHGE-CT93-0040) for supporting data collection at the EMBL Hamburg Outstation. We thank Dr. A. C. Pinder and Dr. R. G. Clark of the Institute of Food Research, Norwich for facilities for fluorimetry.

References

1 R. R. Copley and G. J. Barton, *J. Mol. Biol.*, 1994, **242**, 321.
2 H. Luecke and F. A. Quiocho, *Nature* (*London*), 1990, **347**, 402.
3 J. W. Pflugrath and F. A. Quiocho, *Nature* (*London*), 1985, **314**, 257.
4 M. T. Pope, E. R. Still and R. J. P. Williams, in *Molybdenum and molybdenum-containing enzymes*, ed. M. Coughlan, Pergamon Press, Oxford. 1980, p. 3.
5 J. J. R. Fraústo da Silva and R. J. P. Williams, Clarendon Press, Oxford, 1993, ch. 17, pp. 411–435.
6 R. N. Pau, W. Klipp and Leimkühler, in *Iron and related transition metals in microbial metabolism*, eds. G. Winkelmann and C. J. Carrano, Harwood Academic Publishers, Amsterdam, 1997, p. 217.
7 F. A. Quiocho and P. S. Ledvina, *Mol. Microbiol.*, 1996, **20**, 17.
8 T. A. Jones, J.-Y. Xou, S. W. Cowan and M. Kjeldgaard, *Acta Crystallogr.*, *Sect. A*, 1991, **47**, 110.
9 W. G. J. Hol, P. T. van Duijnen and H. J. C. Berendsen, *Nature* (*London*), 1985, **273**, 443.
10 J. J. He and F. A. Quiocho, *Protein Sci.*, 1993, **2**, 1643.
11 S. M. Hinton and B. Merritt, *J. Bacteriol.*, 1986, **168**, 688.
12 N. J. Mouncey, L. A. Mitchenall and R. N. Pau, *J. Bacteriol.*, 1995, **177**, 5294.
13 R. N. Pau, A. Choay, A. Mitchenall and D. J. White, unpublished work.
14 L. A. Anderson, T. Palmer, N. C. Price, S. Bornemann, D. H. Boxer and R. N. Pau, *Eur. J. Biochem.*, 1997, **246**, 119.
15 B. A. Fields, H. H. Bartsch, H. D. Bartunik, F. Cordes, M. Guss and H. C. Freeman, *Acta Crystallogr.*, *Sect. D*, 1994, **50**, 709.
16 F. W. Studier and B. A. Moffatt, *J. Mol. Biol.*, 1986, **189**, 113.
17 L. C. Clark and J. Axley, *Anal. Chem.*, 1955, **27**, 2000.
18 B. Rost and C. Sander, *Proteins: Struct. Funct. Genet*., 1994, **19**, 55.
19 B. Rost, TOPITS: Threading one-dimensional predictions into three-dimensional structures, AAAI Press, Menlo Park, CA, p. 314.
20 C. M. Stultz, J. V. White and T. F. Smith, *Protein Sci.*, 1993, **2**, 305.
21 J. V. White, C. M. Stultz and T. F. Smith, *Math. Biosci.*, 1994, **119**, 35.
22 J. F. Devereux, A. F. W. Coulson and A. Lyall, *Nucleic Acids Res.*, 1987, **12**, 387.

Received 9*th June* 1997; *Paper* 7/04006G