Crystallographic structure and substrate-binding interactions of the molybdate-binding protein of the phytopathogen

*Xanthomonas axonopodis* pv. *citri*

Andrea Balan\(^a,b\), Carolina Santacruz-Pérez\(^a,b\), Alexandre Moutran\(^a\), Luís Carlos Souza Ferreira\(^a\),
Goran Neshich\(^c\), João Alexandre Ribeiro Gonçalves Barbosa\(^b,*\)

\(^a\) Departamento de Microbiologia, Instituto de Ciências Biomédicas II, Universidade de São Paulo, Cidade Universitária, São Paulo, SP, 05008-900, Brazil
\(^b\) Center for Structural Molecular Biology (CeBiME), Brazilian Synchrotron Light Laboratory (LNLS), CP 6192, Campinas, SP, 13084-971, Brazil
\(^c\) Laboratório de Bioinformática, EMBRAPA/CNPTIA, Campinas, SP, Brazil

Received 27 August 2007; received in revised form 8 November 2007; accepted 13 November 2007
Available online 26 December 2007

Abstract

In *Xanthomonas axonopodis* pv. *citri* (Xac or *X. citri*), the *modA* gene codes for a periplasmic protein (ModA) that is capable of binding molybdate and tungstate as part of the ABC-type transporter required for the uptake of micronutrients. In this study, we report the crystallographic structure of the Xac ModA protein with bound molybdate. The Xac ModA structure is similar to orthologs with known three-dimensional structures and consists of two nearly symmetrical domains separated by a hinge region where the oxyanion-binding site lies. Phylogenetic analysis of different ModA orthologs based on sequence alignments revealed three groups of molybdate-binding proteins: bacterial phytopathogens, enterobacteria and soil bacteria. Even though the ModA orthologs are segregated into different groups, the ligand-binding hydrogen bonds are mostly conserved, except for *Archaeglobus fulgidus* ModA. A detailed discussion of hydrophobic interactions in the active site is presented and two new residues, Ala\(^{38}\) and Ser\(^{151}\), are shown to be part of the ligand-binding pocket.

© 2007 Elsevier B.V. All rights reserved.

Keywords: ModA; Molybdate-binding protein; *Xanthomonas axonopodis* pv. *citri*; ABC transporters; X-ray crystal structure

1. Introduction

Molybdenum and tungsten are normally available in trace amounts in the environment and must be incorporated into the cells in an efficient way [1,2]. Many proteins bind molybdenum and tungsten and for molybdoenzymes such as nitrate reductase, formate dehydrogenase, dimethyl-sulfoxide reductase, trimethylamine-N-oxide reductase, and biotin-sulfoxide reductase, molybdenum is required as a cofactor [3,4]. These enzymes are responsible for transduction of energy, uptake of nutrients and gene regulation and can be found in the cytoplasm and associated with the cytoplasmic membrane [5–8].

The molybdate-binding protein (ModA) belongs to a distinct class of periplasmic proteins pertaining to the group of ABC (ATP-Binding Cassette) transporters specifically involved in the uptake of nutrients including metals such as molybdenum [9]. Among the periplasmic nutrient-binding proteins, ModA shows the highest degree of both specificity and affinity in comparison with other soluble components of ATP-dependent uptake systems [10]. The enhanced affinity of ModA to tetrahedral oxyanions, such as molybdate and tungstate, with regard to other potential ligands including sulfate, phosphate and vanadate, was attributed either to the higher coordination number or the larger anion size [11], as deduced from the structural studies.

The structures of the *Escherichia coli* (Eco ModA), *Azotobacter vinelandii* (Avi ModA) and more recently, *Archaeglobus fulgidus* (Afu ModA) ModAs were solved and shown to share high structural similarities in spite of the low amino acid sequence identity [12–13]. The proteins show an ellipsoidal shape with two globular domains connected by a hinge region and delimiting a cleft where the anion is bound and remains
completely entrapped and inaccessible to the solvent. The molybdate is kept linked to the protein by hydrogen bonds to the oxygen atoms of the molybdate.

Molybdate transporters have been recently reported in more than 25 different bacterial species including *Rhodobacter capsulatus*, *Haemophilus influenzae* and the archaeon *Pyrococcus furiosus* [14–17]. Phylogenetic analyses of molybdate and sulphate-binding proteins have shown that ModA orthologs of *E. coli* and *A. vinelandii* belong to different sub-families, due to differences in the amino acid composition of the ligand-binding pocket, particularly the presence of the tyrosine residue at position 170, which, in the *E. coli* ortholog, forms hydrogen bonds with the molybdate anion [18]. Even bigger differences were noted for the anion-binding mode in *Afu* ModA [13]. The presence of a ModA ortholog in a plant pathogen was first reported following the sequencing of the complete genomes of *Xanthomonas axonopodis* pv. *citri* and *X. campestris* pv. *campestris* [19]. The successful expression and purification of a recombinant form of the *Xac* ModA was recently reported by our group [20]. The protein has a molecular weight of 25 kDa, binds molybdate and tungstate with high affinity and shows enhanced thermal stability following anion binding. Sequence alignment of the *Xac* ModA with different bacterial orthologs revealed a close similarity level with the *Eco* ModA. Moreover, crystals of the recombinant *Xac* ModA, obtained in the presence of molybdate, diffracted to a resolution of 1.7 Å [21]. In this work, we report the definition of the *Xac* ModA structure solved by molecular replacement based on the *Eco* ModA crystallographic structure.

2. Material and methods

2.1. Computational analysis

The nucleotide and corresponding amino acid sequences of the *X. citri modA* gene (gene ID: 1157429) were made available by the *Xanthomonas* genome project supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (http://genoma4.iq.usp.br/xanthomonas/) (Genbank accession number AE011982). Search of ModA ortholog sequences were first carried out using the KEGG2 program of the Bioinformatics Center Institute for Chemical Research Kyoto University (www.genome.jp). The prediction of the signal peptide was made with the SignalP program (http://www.cbs.dtu.dk/services/SignalP/). The alignments of ModA amino acid sequences were carried out using the BLASTp and PSI-BLAST, available at the National Center of Biotechnology Information server (http://www.ncbi.nlm.nih.gov/BLAST) and ClustalX [22]. Genodex was used to edit the ClustalX results (http://www.psc.edu/biomed/genodex/). The tree produced from the alignments was obtained using the Phylip tool (http://evolution.genetics.washington.edu/phylip.html), and was visualized with Treeview (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html).

2.2. Crystallization, data collection and processing

*Xac* ModA was expressed in *Eco* BL21(DE3) and purified in high amounts from the cytoplasm by immobilized metal affinity chromatography, as previously described [20]. Purified protein was submitted to crystallization trials at 18 °C using the sitting-drop vapour-diffusion technique. Crystallographic data from a single cryocooled crystal of *ModA* in the presence of sodium molybdate were collected at the protein crystallography beamline D03B-MX1 of the Brazilian Synchrotron Light Laboratory (LNLS) in Campinas, Brazil. The data set was processed using HKL2000 [23] and the CCP4 package (Collaborative Computational Project, Number 4, 1994). The crystal belongs to the C2221 space group with unit cell dimensions *a*=68.15, *b*=171.14, *c*=112.04 Å and presents one trimer in the asymmetric unit [21]. Data processing statistics are shown in Table 1.

2.3. Structure solution, refinement and analysis

The structure of ModA bound to molybdate was solved by molecular replacement using the MOLREP program [24] and taking *Eco* ModA (PDB code 1AMF) as the search model. The resulting model was refined using REFMAC [25] and its stereochemistry validated with PROCHECK [26]. The program COOT [27] was used for model building. The final refinement statistics for the structure are listed in Table 1. The refined coordinates have been deposited in the PDB and received the entry code 2HSY. Comparisons between *Xac* ModA and the orthologs were performed using the COOT program [27] and the Java Protein Dossier (JPD) of the Diamond Sting Suite (DMS 3.0) [28]. The cutoff distances used for the definition of hydrogen bonds, salt bridges, and aromatic interactions were 2.0–3.2 Å, 2.0–6.0 Å, and 4.0 Å, respectively. The secondary-structure matching (SSM) algorithm [29] implemented within the COOT program was used for the superposition of ModA structures and r.m.s.d. calculations. All pictures were generated with the PYMOL software [30].

3. Results and discussion

3.1. Overall structure of Xac ModA bound to molybdate

The structure of the molybdate-bound form of *Xac* ModA was obtained by molecular replacement using the *Eco* ModA coordinates as the searching template. The model was refined

| Data collection, processing and refinement statistics of the *Xac* ModA structure |
|----------------------------------|------------------|
| **Space group**                  | C222,            |
| **Unit cell parameters (Å)**     | *a*=68.15, *b*=171.14, *c*=112.04 |
| **Mosaicity (°)**                | 0.5              |
| **Resolution limits (Å)**        | 40.00–1.70 (1.76–1.70) |
| **Mean I/σ(I) after merging**    | 26.6 (3.3)       |
| **Completeness (%)**             | 99.8 (99.3)      |
| **Multiplicity**                 | 5.7 (5.7)        |
| **Rmerge**                       | 0.062 (0.415)    |
| **Number of reflections**        | 417669          |
| **Number of unique reflections** | 72786 (7166)     |
| **Rfree**                        | 0.168           |
| **Rmerge**                       | 0.202           |
| **r.m.s.d. bond lengths (Å)**    | 0.012           |
| **r.m.s.d. bond angles (°)**     | 1.752           |
| **Total number of atoms/waters** | 6036/841        |
| **Average B factor (Å²)**        | 25.88           |
| **Ramachandran plot statistics by region (%)** | 95.6 |
| **Most favored**                 | 4.4             |
| **Allowed**                      |                 |

Values in parenthesis correspond to data regarding the last resolution shell of 1.76–1.70 Å. Data collection was performed at 110 K with wavelength radiation of 1.421 Å and using a MARCCD detector to record the rotation data with Δφ=1.0°.  

* Rfactor = Σ|Fo(h)–Fc(h)|/ΣFo(h), where Fo(h) and Fc(h) are observed and calculated amplitudes for reflection h. Rfree is calculated by the same equation using 5% of the data, chosen randomly and omitted from the refinement.

b There were no residues in the generously allowed or disallowed regions of the plot.
against data to a maximum resolution of 1.7 Å, yielding an $R_{\text{factor}}$ of 0.168 and $R_{\text{free}}$ of 0.202 (Table 1). The quality of Xac ModA model was validated with the PROCHECK program [26]. The Ramachandran plots revealed that 95.6% of the residues were in the most favored regions and no residue was placed in disallowed regions. The crystalline structure has three molecules of ModA in the asymmetric unit, all of which have a molybdate anion bound. Analysis of the monomer–monomer contacts within the unit cell using the PISA program [31] indicates that the trimer in the asymmetric unit and other complexation possibilities are formed due to crystallographic packing and should not be present in solution (data not shown). The r.m.s.d. between the individual monomers of the asymmetric unit is 0.3 Å. Similar to orthologs with solved three-dimensional structure, the Xac ModA has an ellipsoidal structure with dimensions of approximately 32 Å × 55 Å with a venus flytrap shape typical of ABC transporters’ periplasmic protein. The protein is formed by two similar domains (α/β sandwich), named N- and C-domains, separated by the substrate-binding cleft (Fig. 1A). The N-domain is formed by residues 1–82 and 196–234, encompassing a region defined by one β-sheet with 5 strands surrounded by 5 α-helices, while the C-domain, spanning from residue 83 to 195, is formed by one 5-stranded β-sheet surrounded by 7 α-helices. As shown in Fig. 1B, the Cα alignment of the four ModA orthologs revealed that the similarity with the Xac ModA increases in the following order: A. fulgidus, A. vinelandii and E. coli with sequence identities of 22%, 26% and 52% and r.m.s.d. of 1.8 Å (214 residues aligned), 1.6 Å (211 residues aligned) and 1.1 Å (226 residues aligned) for Cα atoms, respectively. The higher r.m.s.d against the recently reported archaeal Afu ModA [13], reflects the presence of several structural differences including the presence of two big insertions that form an additional 4-stranded β-sheet (Fig. 1B). The Xac ModA presents inter domain connections characteristic of the group II hinge region of periplasmatic nutrient-binding proteins (amino acids 193–195 and amino acids 82–84) [32].

3.2. Ligand-binding pocket of Xac ModA

The molybdate-binding pocket mediates key protein–substrate interactions affecting both specificity and affinity of the bacterial ModA proteins. Among these interactions, the most important ones are seven hydrogen bonds between the protein and the molybdate anion oxygens (O1 to O4): Ser$_{12}$/O$^\gamma$H–O4, Ser$_{12}$/NH–O4, Ser$_{39}$/NH–O2, Ser$_{39}$/O$^\gamma$H–O2, Ala$_{125}$/NH–O3, Val$_{152}$/NH–O1 and Tyr$_{170}$/O$^\eta$H–O3 (Fig. 2). Four of these hydrogen bonds involve main chain amino groups located at the N-terminal region of α-helices, where the helical positive dipoles contribute for the stabilization of the molybdate negative charges (Fig. 2B). One exception concerning the helix dipole stabilization might be the Ser$_{12}$ amino group interaction, located one residue prior to the helix and thus not aligned with the helix axial dipole. These electrostatic interactions are conserved between Xac and Eco ModA [12], but differ from Avi ModA in residues Ser$_{12}$, Ala$_{125}$ and Tyr$_{170}$, which are changed to Asn, Tyr and Ala, respectively (Fig. 2C). The Ser$_{12}$Asn and
 Ala125Tyr changes do not affect the ligand hydrogen bonding pattern, but Tyr170Ala prevents the O\(\eta\)H–O3 hydrogen bond. This loss is compensated by the Ala11Thr change in which the Thr residue forms a new O\(\gamma\)H–O1 hydrogen bond found only in Avi ModA [11]. This network of hydrogen bonds is very specific for coordinating the tetrahedral disposition of the molybdate oxygens in bacteria. In contrast, the archaeal Afu ModA presents octahedral oxyanion coordination sustained by similar interactions and the additional ones provided by the changes Val123Asp and Val152Glu. These acid residues provide one side-chain’s carboxylate oxygen each to form the octahedral geometry [13].

The selectivity for the anion has been explained by the size and polarity of the binding pocket [12,11]. Theoretical calculation using quantum mechanical/continuum dielectric methods have indicated that this selectivity is mainly due to the pocket size in the case of bacterial sulfate-binding proteins (SBP) [33], while with ModA, the most important effect would be anion desolvation [34]. Desolvating the larger molybdate anion is more energetically favored than the smaller sulfate. The polarity of the binding site affects the desolvation process by changing the environment dielectric constant. It has been shown that the ModA binding site has an apolar surface when compared

Fig. 2. The ModA binding site. (A) Amino acid structural alignment of the ModA orthologs from Xac, Eco, Avi and Afu produced with ClustalX [22] and corrected by visual inspection of the superposed structures. The numbering scheme follows the Xac ModA and the regions in boxes are structurally dissimilar from the Xac ModA structure. The asterisks underneath the sequences denote conservation, while the cylinders (helices) and arrows (β-strands) indicate the secondary structures present in Xac ModA. Amino acid residues involved in molybdate interactions are shown in yellow, when forming hydrogen bonds, or in green, when possessing carbon atoms within a 3.7 Å shell. (B) Ligand-binding pocket stereo view of Xac ModA evidencing the residues that interact with molybdate. Red cylinders indicate α-helix’s axis and blue dashed lines indicate hydrogen bonds. The α-helix close to residue Ser12 is outside the limits of the figure. (C) Superposed ligand-binding site residues of ModA orthologs showing the equivalent residues in Xac (purple), Eco (green), Avi (blue) and Afu (yellow) proteins, respectively, that coordinate the ligand via hydrogen bonds. Residues in equivalent positions of the alignment that do not form hydrogen bonds are represented by small traces.
to the negative potential of the SBP cavity [11]. This difference arises from several residues that have carbon atoms in the binding pocket. In Xac ModA, residues Ala10, Ala11, Ala38, Ala58, Val123, Pro124 and Ser151 have carbons located at less than 3.7 Å from the anion (marked in green in Fig. 2A). Residues Ala38 and Ser151 have not been previously cited as forming part of the binding pocket of bacterial ModAs, nevertheless they are certainly enhancing the molybdate selectivity by contributing to the formation of an apolar environment. In the case of these residues, the differences between Xac and Eco are: Ala38—Ser and Ser151—Asp; while the differences to the Avi structure are: Ala11—Thr, Ala38—Ser and Val123—Ala. All these changes do not affect the positioning of the carbon atoms that form the binding pockets in the Eco or Avi orthologs. In the case of the Afu ModA, the carbon atoms forming the binding pocket are from residues occupying the same position in the structure (Fig. 2A), with the noticeable exception of the aspartate that replaces the Val123 and participates in the octahedral coordination.

The Xac ModA binds molybdate with a dissociation constant ($K_d$) of 0.29 μM [20]. Thus, molybdate affinity of Xac ModA is comparable to those reported for other bacterial molybdate-binding proteins such as the Eco K12 ModE ($K_d=0.8 \mu M$) [35], the H. influenzae Mop protein ($K_d<1 \mu M$) and the two A. vinelandii ModG binding sites ($K_d$ of 10 nM and 10 μM) [5]. Two different $K_d$ values for the molybdate-binding affinity of Eco ModA have been reported (20 nM and 3 μM), probably reflecting differences in the methodological approaches employed by the two research groups [36,37]. Since the amino acid residues directly interacting with molybdate were the same in Xac and Eco ModA, $K_d$ differences may be reasoned by other residues indirectly affecting the attraction, binding and release of the anion to the oxyanion-binding pockets of the two proteins.

3.3. Sequence homology and phylogenetic analysis

The Xac modA gene encodes a 257 residue protein, including a putative 24 amino acid-long signal peptide. Amino acid sequence alignment of Xac ModA and twelve bacterial and one archaeal ModA orthologs revealed two highly conserved regions including all residues directly interacting with molybdate. An unrooted phylogenetic tree generated from the amino acid sequence of bacterial ModA and the recently reported sequence of the archaeal Afu ModA protein [13], revealed the presence of three main branches or families (Fig. 3). The first branch (Family 1) included the ModA orthologs expressed by enterobacteria including E. coli, Salmonella typhimurium, Shigella flexneri and Yersinia pestis. The second group (Family 2) encompassed ModA orthologs found in the genomes of plant-interacting bacterial species, such as the X. citri, X. campestris, Mezorhizobium loti, Sinorhizobium meliloti and Agrobacterium tumefaciens orthologs. Finally, the last and third phylogenetic group (Family 3) was characterized by a group of the more distantly related species.
including soil bacteria such as *A. vinelandii*, *Geobacter sulfurreducens*, *Streptomyces coelicolor*, *Ralstonia solanacearum* and the archaeon *A. fulgidus*. Analysis of each group showed that enterobacteria and plant-interacting bacteria share high sequence identity values ranging from 44% to 91% (data not shown). Differently, the sequence comparison for soil bacteria *ModA* excluding the sequence from *Afu ModA*, showed a larger diversity as demonstrated by the low sequence identity values (maximum of 15%). Phylogenetic analysis of the *ModA* orthologs including the four proteins with solved structure showed that the oxyanion-binding pocket residues are not solely responsible for differences in the phylogenetic classification such as was suggested before [11,18]. Fig. 3 reveals that *Eco* and *Xac* *ModA* belong to different families in spite of the conserved residues involved in the hydrogen bond network of the oxyanion-binding pockets. On the other hand, *Avi* and *Afu* *ModA* proteins were classified in the same sequence similarity family despite the differences in their binding pocket amino acid residues.

The recent description of the complete structure of the molybdate transporter of *A. fulgidus* [13] together with the present elucidation of the *Xac ModA* structure, will certainly contribute for the understanding of the molybdate transport in *X. citri*, including the definition of *ModA* amino acid residues interacting with the membrane-bound *ModB* component.

**Acknowledgments**

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, SMolBNet grants 01/07540-3 and 00/10266-8), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Associação Brasileira de Tecnologia de Luz Síncrotron (ABTLuS). C.S.-P. also thanks the Academy of Sciences for the Developing World (TWAS) for the fellowship. We also thankfully acknowledge the support from Dr. Rogério Meneghini.

**References**


