The EMBO Journal Vol.18 No.6 pp.1435–1446, 1999

The high-resolution crystal structure of the molybdate-dependent transcriptional regulator (ModE) from *Escherichia coli*: a novel combination of domain folds

David R.Hall, David G.Gourley, Gordon A.Leonard[1], Elizabeth M.H.Duke[2], Lisa A.Anderson[3], David H.Boxer[3] and William N.Hunter[4]

The Wellcome Trust Building, Department of Biochemistry, University of Dundee, Dundee DD1 5EH, [2]CCLRC-DL, Daresbury Laboratory, Warrington, Cheshire, WA4 4AD, [3]Department of Biochemistry, University of Dundee, Dundee DD1 5HN, UK and [1]Joint Structural Biology Group, ESRF, BP 220, F-38043 Grenoble cedex, France

[4]Corresponding author
e-mail: w.n.hunter@dundee.ac.uk

D.R.Hall and D.G.Gourley contributed equally to experimental work

The molybdate-dependent transcriptional regulator (ModE) from *Escherichia coli* functions as a sensor of molybdate concentration and a regulator for transcription of operons involved in the uptake and utilization of the essential element, molybdenum. We have determined the structure of ModE using multi-wavelength anomalous dispersion. Selenomethionyl and native ModE models are refined to 1.75 and 2.1 Å, respectively and describe the architecture and structural detail of a complete transcriptional regulator. ModE is a homodimer and each subunit comprises N- and C-terminal domains. The N-terminal domain carries a winged helix–turn–helix motif for binding to DNA and is primarily responsible for ModE dimerization. The C-terminal domain contains the molybdate-binding site and residues implicated in binding the oxyanion are identified. This domain is divided into sub-domains *a* and *b* which have similar folds, although the organization of secondary structure elements varies. The sub-domain fold is related to the oligomer binding-fold and similar to that of the subunits of several toxins which are involved in extensive protein–protein interactions. This suggests a role for the C-terminal domain in the formation of the ModE–protein–DNA complexes necessary to regulate transcription. Modelling of ModE interacting with DNA suggests that a large distortion of DNA is not necessary for complex formation.

Keywords: DNA binding/gene regulation/molybdate/OB-fold/winged helix–turn–helix

Introduction

Molybdenum is an essential trace element used to provide a redox active centre in the molybdopterin cofactor of a number of enzymes (Rajagopalan, 1996; Kisker *et al.*, 1997). This second row transition metal is bio-available as the water soluble molybdate, (MoO_4^{2-}), and specific systems have evolved to regulate its uptake and to utilize its complex chemical properties in biochemical reactions. The *modABCD* operon of *Escherichia coli* codes for a high-affinity molybdate transport system that is a member of the ABC transporter family (Higgins, 1992; Pau *et al.*, 1998), and in addition some molybdate may leach through the sulfate transporter. ModA encodes a periplasmic molybdate binding protein, the structure of which has recently been determined (Hu *et al.*, 1997; Lawson *et al.*, 1997). ModB and ModC are integral membrane proteins, the latter of which binds ATP. Little is known about ModD. Transcription of this operon is negatively controlled by the molybdate-responsive regulatory protein encoded by the *modE* gene. This gene is positioned immediately upstream and transcribed divergently from the *modABCD* operon (Grunden *et al.*, 1996). A schematic depiction of MoO_4^{2-} uptake and the regulation of the high-affinity transport system is given in Figure 1. ModE is a pleiotropic regulator of transcription since it appears to act as a positive regulator of the *moaABCDE* operon which encodes enzymes involved in molybdopterin biosynthesis (Walkenhorst *et al.*, 1995; McNicholas *et al.*, 1997, 1998a) and *dmsABC*, the structural operon for the molybdoenzyme DMSO reductase (McNicholas *et al.*, 1998b).

ModE has been cloned, overexpressed and biochemically characterized (Anderson *et al.*, 1997). The protein functions as a homodimer of subunits consisting of 262 amino acids. It displays a high affinity for MoO_4^{2-} and WO_4^{2-} (K_d = 0.8 μM), and the binding of the anion is accompanied by a large quench (50%) of the intrinsic protein fluorescence, producing a blue shift of the emission maximum. This suggests that a significant conformational change accompanies anion binding (Anderson *et al.*, 1997). With anion bound, ModE binds near to the transcription start of the *modABCD* promoter and functions as a repressor. The intracellular availability of molybdate determines the conformation of ModE, regulating transcription of the transport system operon and thereby controlling the acquisition and subsequent use of the element.

Analysis of the ModE sequence suggested two domains (Grunden *et al.*, 1996). The N-terminal domain sequence carries a putative helix–turn–helix (HTH) motif frequently observed in DNA binding proteins (Brennan, 1993). The C-terminal residues present significant homologies with domains found in a range of bacterial proteins, in particular with a small (~7 kDa molecular mass) molybdenum-containing protein from *Clostridium pasteurianum* termed Mop (Hinton and Freyer, 1986; Hinton and Merritt, 1986). Mop was originally thought to bind molybdopterin although this has not been confirmed. Studies with ModE strongly indicate that Mop and homologous domains share the ability to bind MoO_4^{2-}. Such domains are of wide significance in the biology of molybdenum (Grunden *et al.*, 1996; Anderson *et al.*, 1997). The C-terminal domain of *E.coli* ModE contains a tandem repeat of the Mop sequence and we refer to this as the Di-Mop domain. A model for this C-terminal molybdate binding domain has

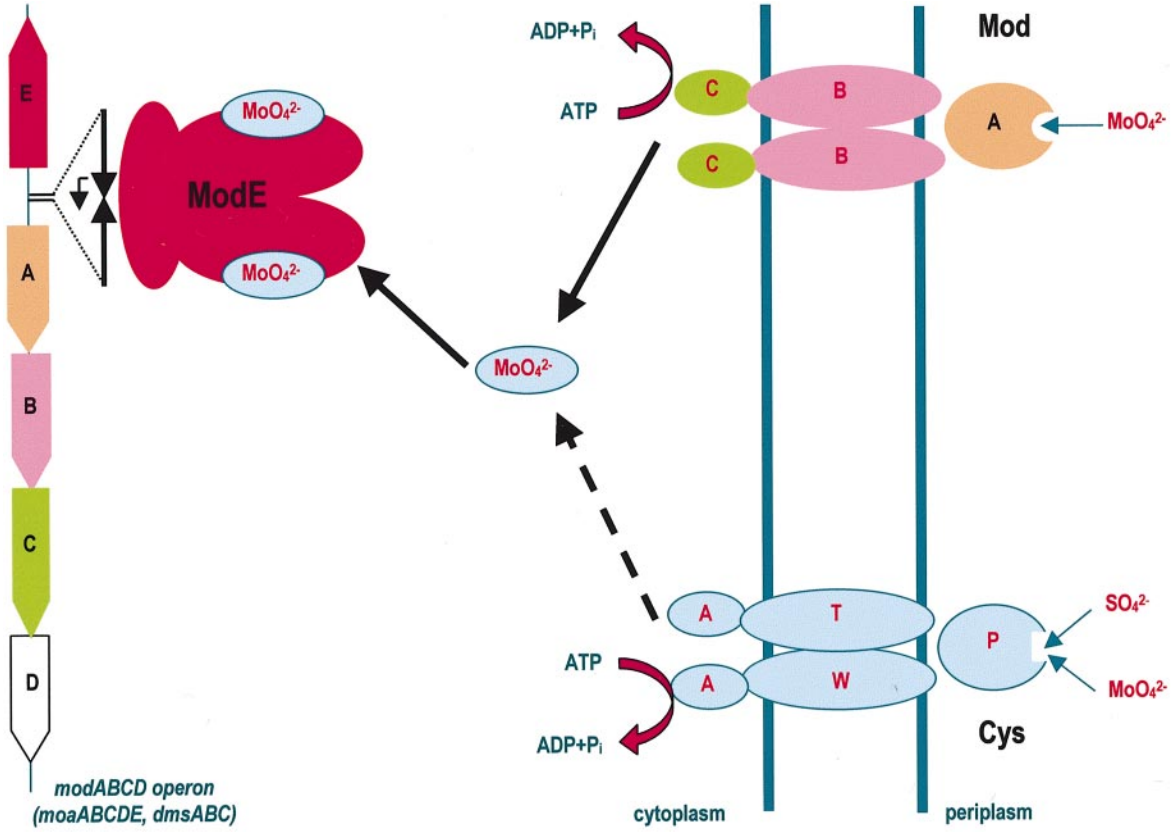


Fig. 1. A schematic depiction of molybdate-dependent gene regulation in *E.coli*. Molybdate enters the bacterium via a high-affinity transporter system which comprises three proteins (ModA, ModB and ModC), and also, less efficiently, through the sulfate transporter which consists of CysP, T, W and A. Once in the cytoplasm the anion binds to the dimeric molybdate-dependent transcriptional activator (ModE), which in turn binds to and represses the *modABCD* operon. ModE also regulates the *moaABCDE* and *dmsABC* operons. Modified from Pau *et al.* (1998).

been proposed based on secondary structure predictions (Lawson *et al.*, 1997).

It has been postulated that ModE is related to one of the most common families of prokaryotic transcription regulators, the LysR family (Anderson *et al.*, 1997). This group of proteins controls the divergent transcription of linked genes or unlinked regulons encoding a range of functions such as nitrogen fixation, amino acid biosynthesis and bacterial virulence (Schell, 1993). The family has a DNA binding N-terminal region of ~65 residues that includes a HTH motif and a co-inducer recognition domain. The co-inducers that stimulate LysR-type transcriptional regulators (LTTRs) are diverse molecular entities, including aromatic and aliphatic compounds, amino acid derivatives and ions. Binding of the co-inducer leads to LTTR–DNA association near the RNA polymerase binding site, DNA bending and transcription activation (Schell, 1993; Ansari *et al.*, 1995). The structure of a complete LTTR has not yet been published although crystal structures of the co-inducer domain of CysB (Tyrrell *et al.*, 1997) and the DNA binding domain of OmpR are available (Martinez-Hackert and Stock, 1997).

We embarked upon a structural analysis of *E.coli* ModE to define an accurate three-dimensional model of the complete transcriptional regulator, and now report the details of native and a selenomethionine-derived protein at resolutions of 2.1 and 1.75 Å, respectively. The model illustrates the structure and organization of the subunits in the dimer, the basis for the dimerization, the orientation of DNA binding domains with respect to each other and the architecture of the Mop and Di-Mop domains. The putative MoO_4^{2-} binding site and residues that might be important for interactions with DNA are also identified.

Results and discussion

The ModE subunit

The ModE monomer consists of two domains, the N-terminal domain (I) comprising all residues to 121, which interacts directly with DNA, and the C-terminal domain (II) which is the putative molybdate binding component. Domain II can be delineated into two sub-domains (*a* and *b*), corresponding to the two Mop sequence similarity regions. Sub-domain *a* consists of residues 122–183 and 256–262, whilst sub-domain *b* is formed from the contiguous stretch of residues 184–255. The folding of each domain and the assembly of each subunit of the dimer are illustrated in Figure 2. Secondary structure assignments have been made with a combination of automated methods (DSSP, Kabsch and Sander, 1983; PROMOTIF, Hutchinson and Thornton, 1996) and by visual inspection. Secondary structure elements are mapped onto the amino acid sequence in Figure 2D.

The N-terminal DNA binding domain (I) is composed mainly of α-helices (60% α-helix, 20% β-strand). There are five α-helices and four β-strands, with the DNA binding region being a winged HTH motif comprising the secondary structure elements α2, α3, β3 and β4. Helix α4 (residues 80–108) forms the backbone of this domain. The C-terminal domain (II or Di-Mop) consists mainly of

β-sheet (25% α-helix, 60% β-strand) and starts at strand β5. It consists of two β-barrels, the first of which is constructed from β6–β9 and β15. The second β-barrel is formed by β-strands 10–14 and this domain also contains a short segment of helix, α8. A turn of 3_{10} helix and a short α helix (α9) are positioned at the junction of the two barrels.

The ModE dimer

A homodimer constitutes the asymmetric unit but there are significant deviations from non-crystallographic symmetry in the arrangement of monomers with respect to each other. There are two distinct non-crystallographic 2-fold axes relating the N-terminal domains with each other and the C-terminal domains with each other. The relationship of these axes to each other is shown in Figure 3. The reorientation of the domains occurs through the linker region (residues 115–126) where few interactions occur between domain I and domain II. There is a considerable dimer interface involving interactions between domain I subunits that constrains the dimer in this region in a rigid structure. McNicholas *et al.* (1998b) explored the role of the N- and C-terminal domains of ModE with respect to dimerization and DNA binding. They concluded that dimerization of ModE was required for efficient DNA binding, and that the C-terminal domain was primarily responsible for dimer formation. The crystallographic results indicate that it is the N-terminal domain that plays the major role in dimerization although the C-terminal domain does make a contribution (see below). The relationship of domain II to domain I, which is partially responsible for asymmetry in the homodimer, is discussed later in terms of the dimer interface and the crystal lattice.

Similar domain folds occur in other proteins

The ModE structure presents a new combination of domains with previously observed folds. The model was submitted to the DALI server (http://www2.ebi.ac.uk/dali; Holm and Sander, 1993) to search for structural homologues. None were found for the ModE dimer or the intact monomer; however, fold similarities were observed separately for the C-terminal (Mop) sub-domains and the N-terminal DNA binding domain.

As expected structures containing the winged and non-winged HTH moiety were identified. The N-terminal monomeric DNA binding region has greatest structural similarity to the diphtheria toxin repressor (DTR) monomer (Ding *et al.*, 1996; White *et al.*, 1998), which also has a winged HTH fold for DNA recognition and binding. For all C_α atoms, an r.m.s. deviation of 3.5 Å was observed, whilst for 47 C_α atoms between residues 34 and 100 (ModE numbering), an r.m.s. fit of 1.7 Å was observed. This region corresponds to the winged HTH motif plus the long backbone helix α4. Figure 4A illustrates the overlay of the monomeric DNA binding domain of ModE with the intact DTR molecule.

The Di-Mop C-terminal domain has a novel assembly with no homologous structure being observed for this entire domain. Each of the two sub-domains aligned best with two chains of pertussis toxin (PT) and C_α traces of the superimposed structures are depicted in Figure 4B. These sub-domains display the oligonucleotide/oligosaccharide binding fold (OB-fold), which is based on a five-stranded Greek-key β-barrel capped by an α-helix located between the third and fourth strands. This fold has been observed in a number of proteins with little or no significant sequence homology (Murzin, 1993). Most of the binding site residues are from the loops protruding from the barrel axis and these loops differ in length, sequence and conformation. The OB-fold for sub-domain *a* of domain II of ModE is not formed in the same ordering of β-sheets/α-helix as that described for typical OB-folds (Murzin, 1993).

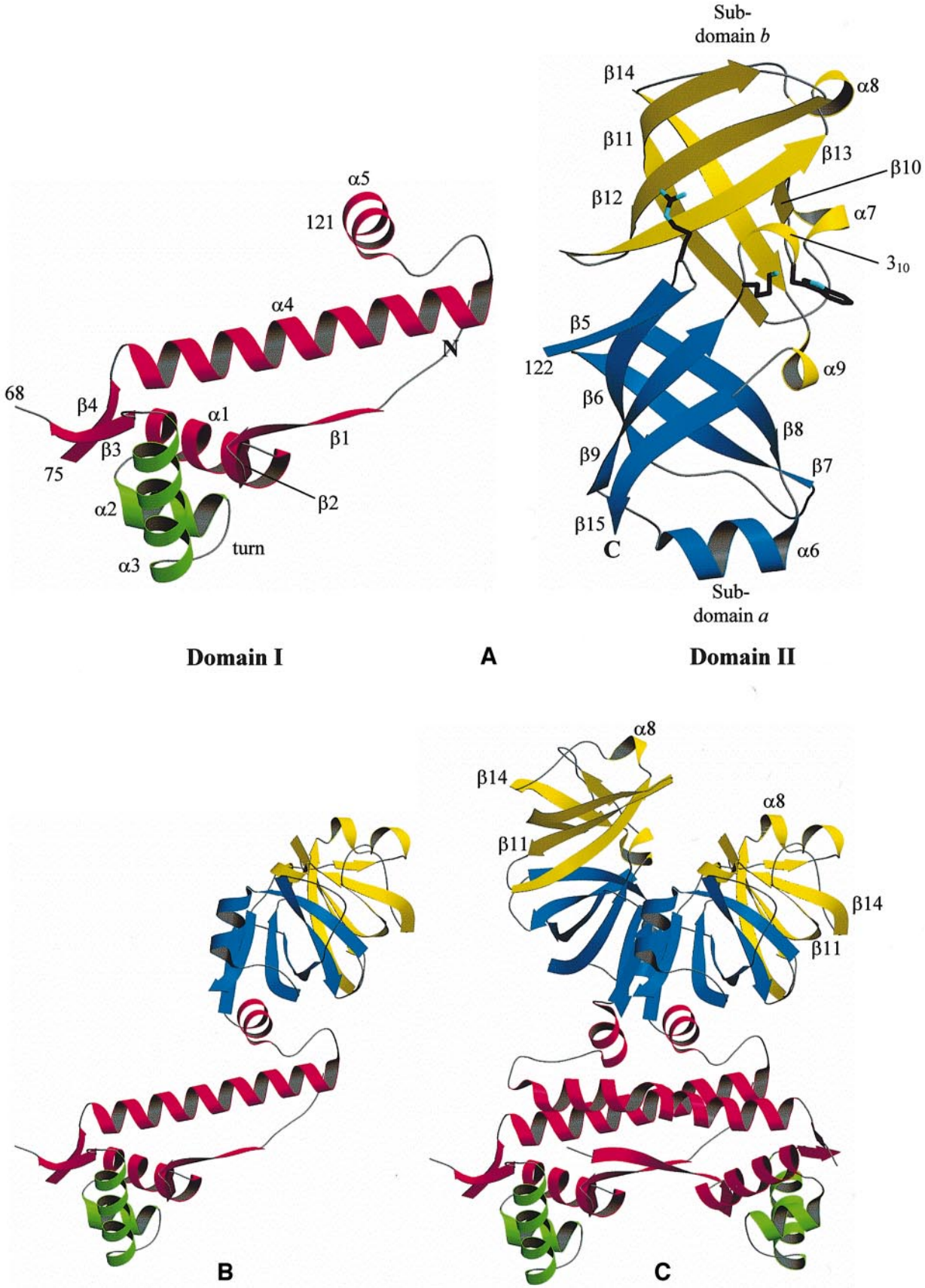
Subunit S4 of the PT–ATP complex (Hazes *et al.*, 1996) aligned best with sub-domain *a* (residues 127–182, 256–262) with all C_α atoms having an r.m.s. fit of 3.1 Å. Subunit S5 of PT (Stein *et al.*, 1994) aligned best with sub-domain *b* (residues 183–255), with all C_α atoms having an r.m.s. fit of 3.0 Å. These PT subunit structures are similar six-stranded anti-parallel β-sheets capped by an α-helix with an OB-fold. The strand arrangement of sub-domain *b* of ModE closely matches that of subunit S5 of PT, whilst sub-domain *a* has five of the six strands of subunit S4 of PT. The residues 136–144 of ModE form a loop instead of the related β-strand of subunit S4 of PT. Both sub-domains of the Di-Mop region of ModE have the capping helix although it is shorter than those observed in S4 and S5 of PT. The spatial relationship of S4 with S5 subunits in PT does not match that of the sub-domains in the Di-Mop region of ModE. However, that both S4 and S5 are involved in extensive protein–protein interactions within the PT structure suggests that the Di-Mop fold may interact with other proteins in order to carry out its function, a point that is discussed later.

The OB-fold was found for other structural matches using DALI. These homologues included the bacterial enterotoxins such as verotoxin-1 (Stein *et al.*, 1992), toxic shock syndrome toxin-1 (Papageorgiou *et al.*, 1996), heat-labile enterotoxin (van den Akker *et al.*, 1996) and the bacterial superantigen *Streptoccocus* pyrogenic exotoxin C (Roussel *et al.*, 1997). All these proteins have the OB-fold either as a subunit of the biological unit or as part of a larger monomeric chain.

The dimer interface

There are extensive contacts between the two subunits of the ModE dimer but these are mainly due to interactions between the N-terminal domains which account for almost 70% of the dimer interface. This suggests that the HTH motifs are held in a relatively fixed conformation in the ModE dimer.

The dimer presents a solvent accessible area of 23 795 $Å^2$ and the buried surface interactions between monomers occupies 6017 $Å^2$, with some 44 residues from subunit A and 46 residues from subunit B closer than 3.6 Å to their partner subunit. Of these, 38 residues are common to each subunit and form 41 inter-subunit hydrogen bonds using nitrogen and oxygen atoms from both main chains and side chains. The interactions are localized in two regions of the N-terminal domain I and at one region in the Di-Mop domain II (Figures 2C and 3). In domain I they are found between strand β1 of both subunits, resulting in an antiparallel two-stranded β-sheet, and helix α4 of both subunits, which has an aliphatic hydrophobic core. Most of the helix–helix contacts are


Sub-
domain *b*
β14
α8
β11
β13
β10
β12
α7
3_{10}
α5
121
α4
N
β5
122
α9
68
β4
α1
β6
β8
β1
β3
β9
β7
75
β2
α2
β15
C
α6
turn
α3
Sub-
domain *a*
Domain I
A
Domain II
α8
β14
α8
β11
β14
β11
B
C

formed at the opposing ends of these long helices. Differences in hydrogen bonding patterns between subunits are observed at these helix–helix contact regions. An inter-subunit salt bridge is observed between Asp106A (A and B are used to identify the subunits where a distinction is necessary) and Arg80B, and this contributes to a strained conformation for the acidic residue. Due to the asymmetry of the dimer, Asp106B has no such interaction.

Almost all of the inter-subunit contacts formed by the residues in the C-terminal Di-Mop domain are formed between strand β5 of one subunit and β14 of the partner subunit. This association occurs between two anti-parallel β-strands which, in conjunction with another from each monomer (strand β8), creates two three-stranded antiparallel β-sheets that form a β-barrel at the centre of the dimer interface.

Domain I–domain II interactions and the influence of the crystal lattice

There are only a few associations between domain I and domain II, and these consist of nine hydrogen bonds and one salt bridge, all of which occur between residues of the same subunit. Thus, there are no interactions between subunit A domain I and subunit B domain II or between subunit B domain I and subunit A domain II. The lack of contacts between domains permits a degree of flexibility in the organization of the dimer which may be necessary for the function of ModE.

The orientation of domain II with respect to domain I is influenced by crystal lattice contacts between symmetry-related molecules. Each dimer contacts 10 symmetry-related dimers using five different interfaces. There are 56 potential hydrogen bonds between symmetry-related molecules as well as solvent mediated interactions and van der Waals contacts. There are however, only a few interactions between domain I with its symmetry partners and these contacts mainly involve sub-domain *b* of domain II. Helix α3, strands β3 and β4 of chain A interact with strands β11 and β14 of a symmetry-related chain B. Helix α3 of chain B interacts with helices α7, α8 and α9 of a symmetry-related chain B, and helix α7 of chain A interacts with residues immediately following strand β1 in chain A of a symmetry-related molecule.

The average temperature factor per residue distribution for each subunit is similar in the DNA binding N-terminal domain, and is reasonably similar in the Di-Mop region although subunit B has slightly higher average temperature factors overall. This is borne out by the average temperature factors for domain I of subunits A and B (35 and 39 $Å^2$, respectively) and for domain II of subunits A and B (43 and 44 $Å^2$, respectively).

The Di-Mop domain

The OB-like fold of the sub-domain of this region has already been discussed and the arrangement of the sub-domains with respect to each other and their organization in the dimer are presented in Figure 2. Residues that are strictly conserved in the Di-Mop domain of ModE and the Mop domains of other proteins are Val147, Ile162, Ser166, Glu177, Lys183, Val187, Asn201, Glu218 and Gly244. The polar residues are worthy of further comment. Ser166A O_γ has no interactions but Ser166B O_γ forms a hydrogen bond (3.15 Å) with $N_{\eta 1}$ of Arg169B. In turn, Arg169B forms three hydrogen bonds to strand β5 of

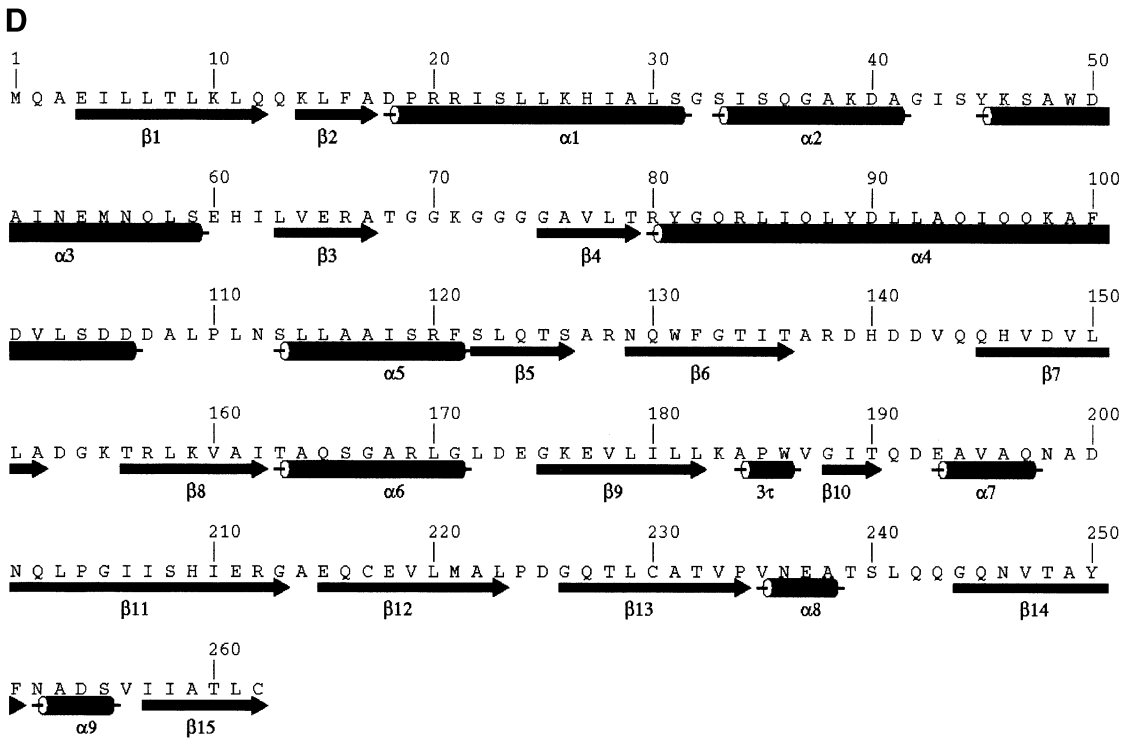


Fig. 2. Ribbon diagrams to depict the architecture of ModE. (**A**) Domains I and II. The helices of the HTH motif of domain I are coloured green. Sub-domains *a* and *b* of domain II are coloured blue and yellow. Residues Arg128, Lys183 and Trp186 are shown as sticks. (**B**) A ModE monomer. (**C**) The dimer with elements of secondary structure at the C-terminal domain assigned to aid orientation of this domain with respect to (A). Figures 2, 3, 4, 5 and 6 were generated with MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merritt and Murphy, 1994). (**D**) The amino acid sequence with elements of secondary structure that have been assigned.

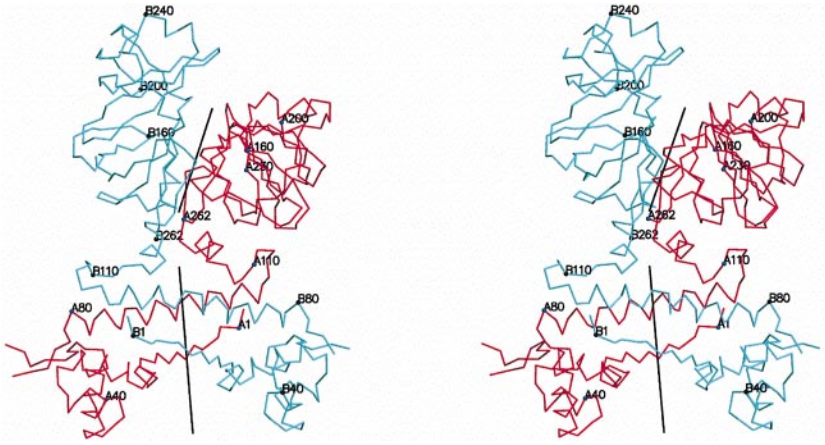


Fig. 3. Stereoview C_α trace of the ModE dimer. Monomer A is red, monomer B is cyan and the black lines represent the non-crystallographic 2-fold axes relating each type of domain.

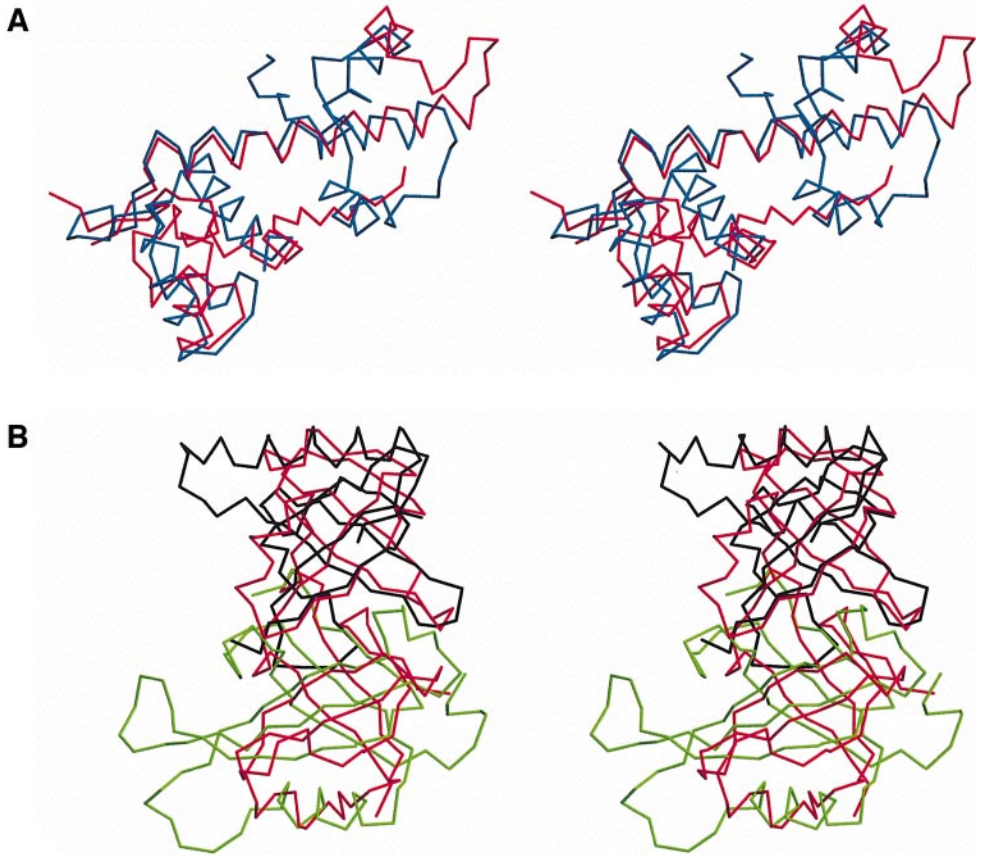


Fig. 4. Least-squares overlaps of components of the ModE structure with other protein fragments. (**A**) Domain I of ModE (red) with DTR (blue). (**B**) Domain II of ModE (red) with PT–ATP S4 (green) overlayed on sub-domain *a*, and PT S5 (black) overlayed on sub-domain *b*.

subunit A. Glu177 has differing interactions in each subunit and is involved in interaction between domains II and I. Glu177A $O_{\varepsilon1}$ hydrogen bonds to O_γ of Ser113A whilst Glu177B $O_{\varepsilon1}$ forms a salt bridge (2.57 Å) with the $N_{\eta1}$ of Arg120B and Glu177B $O_{\varepsilon2}$ forms a hydrogen bond (2.99 Å) with $N_{\eta2}$ of Arg120B. This is another example of the asymmetry observed between the domains of the two subunits. Lys183 is the only other conserved positively charged side chain found in the Di-Mop domain. The N_ζ atom of this residue interacts with the carbonyl of Asp254 of the opposing subunit and is therefore involved in dimerization. Asn201, in both subunits, forms hydrogen bonds with Lys159 and Phe251 of the same subunit. The interactions with Lys159 are between sub-domain *a* and sub-domain *b*. Asn201 is located in strand β11 and Phe251 in strand β14. Hydrogen bonds between the main-chain atoms of these residues form part of the β-barrel of sub-domain *b*.

As mentioned earlier, the two Mop-like sub-domains (*a* and *b*) are interdependent for their structure. This is a result of sub-domain *a* being formed from two stretches of polypeptide chain (residues 128–183 and 256–262).

This contrasts with the assignment of the Mop-like domains from sequence analysis (Walkenhorst *et al.*, 1995; Lawson *et al.*, 1997), whereby each sub-domain was assigned to continuous sequence. Sub-domain *b* is formed from continuous sequence (residues 184–255). The contacts between the sub-domains in each subunit are similar and include the residues Arg128 and Asn129 from the conserved sequence SARN. This tetrapeptide sequence (residues 126–129) has been implicated in dimer formation (McNicholas *et al.*, 1998b) and molybdate binding (Grunden *et al.*, 1996; Lawson *et al.*, 1997).

The putative molybdate binding site

The ModE dimer binds two molybdates (Anderson *et al.*, 1997), and by consideration of residues that are conserved in ModE and Mop proteins, their position in the structure and their chemistry, we can assign a molybdate binding site. The residues of interest are Arg128, which is contributed from the SARN segment, Ser166, Lys183 and Glu218. These residues are strictly conserved in ModE and Mop domain sequences. Arg169 is conserved between *E.coli* ModE and a ModE homologue identified in *Haemophilus influenzae* ORF7 (Fleischmann *et al.*, 1995), but in Mop sequences is changed to either a glutamic acid or asparagine. Thr232 is present in all ModEs and is homologous to a serine found in *C.pasteurianum* Mop. The positions of Arg128 and Lys183 are shown in Figure 2A to indicate their position at the junction of subunits *a* and *b*. The asymmetry of the ModE structure results in distinct arrangements of the two putative MoO_4^{2-} binding sites in the homodimer (Figure 5). The only positively charged and conserved residues are found near the interface of the sub-domains. In the periplasmic molybdate binding protein, ModA, the oxyanion binding site is created in a deep cleft at the junction of four helices oriented to use the helix macrodipole (Hol *et al.*, 1978). There are seven hydrogen bonds donated from uncharged donor groups to the oxyanion provided by four main-chain amides, hydroxyl groups from two serines and finally a tyrosine N_δ (Hu *et al.*, 1997; Lawson *et al.*, 1997). The binding site in ModA is different from that of ModE although the possible conservation of serine/threonine–MoO_4^{2-} interactions is noted. This is not unexpected since this residue type is typically found in binding sites for another oxyanion, namely phosphate (Copley and Barton, 1994).

Ser126 is positioned at the end of strand β5 and the amide forms a hydrogen bond with Ile258 O at the beginning of strand β15 of the opposing subunit. The sheet formed by these two strands is responsible for dimer interactions in domain II. Asn129 as found in both subunits forms hydrogen bonds with Leu182 and Cys230 of the same subunit. Glu218A forms a hydrogen bond (2.96 Å) to the symmetry-related Arg138A. This constrains Glu218A, permitting its $O_{\varepsilon 2}$ atom to form intramolecular salt bridges to $N_{\eta 1}$ (2.52 Å) and $N_{\eta 2}$ (2.46 Å) of Arg128A, a residue implicated in molybdate binding (Lawson *et al.*, 1997). The related site in the partner subunit has no such lattice contact. If molybdate binds to Arg128 and to Lys183 then this would disrupt the interactions between sub-domains. The movement of sub-domains with respect to each other would in turn influence the environment of Trp186 given its location at the interface of the sub-domains. The side chain of Trp186A is present in a double conformation (not shown). One side chain conformer has its $C_{\eta 2}$ atom 3.2 Å from the $C_{\eta 2}$ atom of Trp186B. Alterations in the environment of this particular residue following molybdate binding would account for the large fluorescence quenching observed on the addition of the oxyanion (Anderson *et al.*, 1997).

How does ModE bind to DNA?

DNA recognition and binding of ModE to the *modA* promoter involves recognition of a palindromic sequence (Anderson *et al.*, 1997). The dimerization of the protein provides the 2-fold related residues necessary for interaction with such a DNA motif. Ansari *et al*. (1995) surmised that the LTTR MerR, when in repressor mode or closed conformation was able to bend DNA towards itself in a similar fashion to the bacterial catabolite activator protein, CAP. The activated or open form induced unbending and provided the template for transcription. This was proposed as a paradigm for how the LTTR family operate.

A comparison of residues conserved, in terms of structural position and chemistry, on the HTH motifs of ModE, CAP and DTR gives some insight into components of ModE that may be involved in DNA binding. The structures of CAP (Schultz *et al.*, 1991; Parkinson *et al.*, 1996) and DTR (White *et al.*, 1998) have been solved in complex with oligonucleotides and indicate that DNA–protein interactions occur between helix α3 and the major groove, and a conserved serine of the turn binds to the phosphate backbone. From the CAP, DTR and ModE sequence alignments and by visual inspection of the ModE model (Hall, 1999) we can predict which ModE residues might interact with DNA. These are Ser35, Gln36, Lys39, Ser44 (conserved in all examples considered here), Tyr45 and Lys46 (CAP has interactions with DNA at the equivalent residues whilst DTR has glycine and proline which can have no side chain interactions), Ser47, Trp49 and Asp50.

When CAP binds to its target, the DNA is distorted significantly (reviewed by Kolb *et al.*, 1993). Such a large distortion is not observed when DTR is complexed to DNA (White *et al.*, 1998). A 30 bp B-DNA duplex model was generated with the program INSIGHT based on the promoter region of *modA* (Anderson *et al.*, 1997). A ModE–DNA model was assembled using knowledge of DTR– and CAP–DNA interactions (Figure 6), and suggested that the winged HTH motif present at either end of domain I of ModE will interact with both the major groove and adjacent minor groove of DNA. Helix α3 is positioned for major groove recognition. The wing, formed from residues 61–78, would interact with the adjacent minor groove. Residues at the tip of the wing are absent from the model, this region is glycine rich and predicted to be highly mobile. There is a lysine in the glycine-rich loop, the side chain of which could interact with the phosphate backbone of the DNA. Comparisons of ModE with the structures of DTR– and CAP–DNA complexes suggest that ModE is more similar to DTR than CAP since reasonable contacts can be formed by the protein interacting with standard B-form DNA (see Figure 6; Hall, 1999). There is no reason to invoke large-scale bending or distortion of DNA induced by ModE as playing an important role in its function. However, whilst studying the anaerobic regulation of the *dmsABC* operon by ModE,

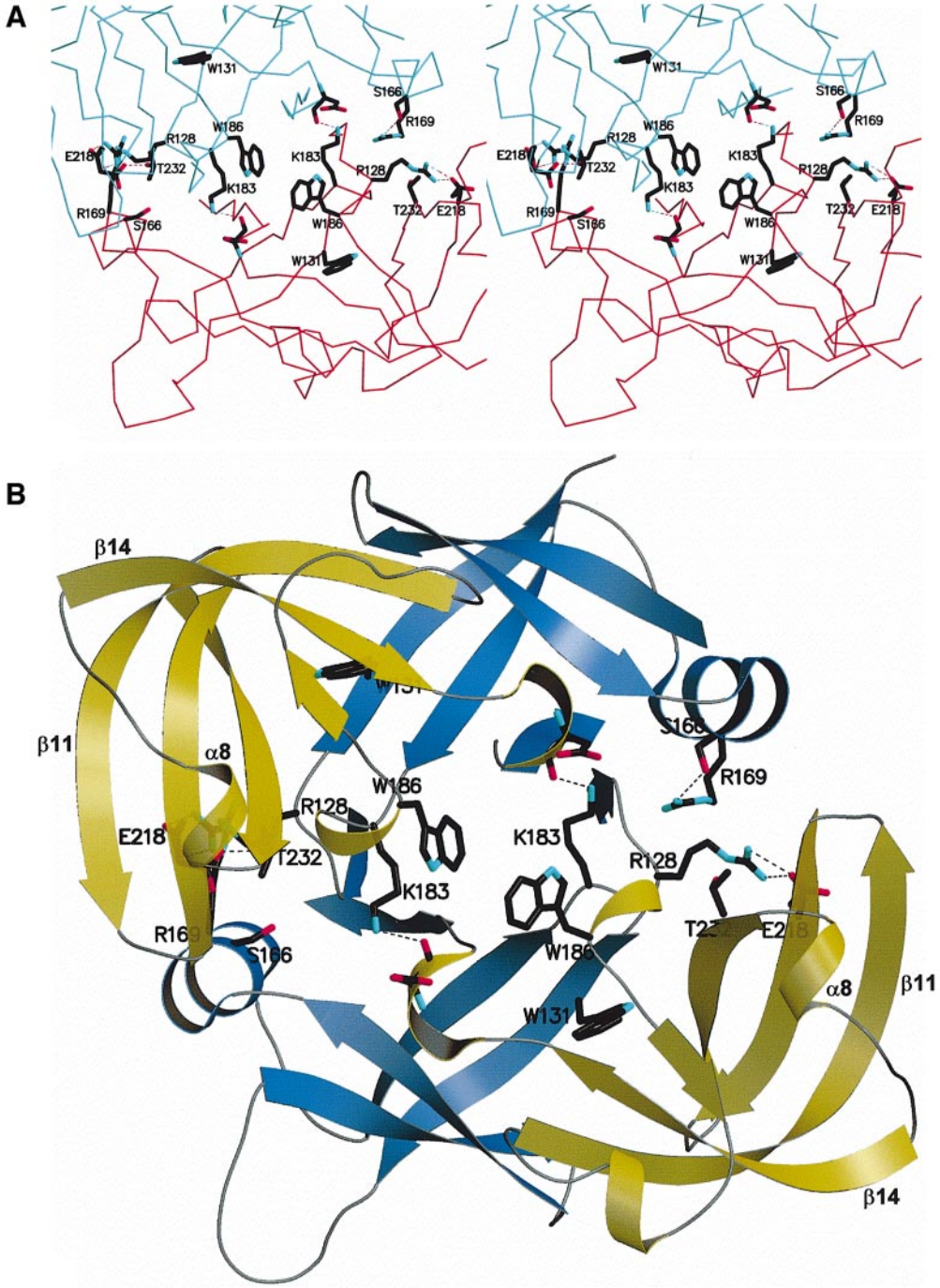


Fig. 5. (**A**) Stereoview to show the sub-domain II*a* interface with sub-domain II*b* and residues postulated to bind molybdate or to be influenced by oxyanion-induced conformational changes. Dashed lines represent hydrogen bonding interactions. (**B**) The same view as (A) but illustrating secondary structure with elements labelled as for Figure 2C.

McNicholas *et al.* (1998a) noted a binding site for Integration Host Factor (IHF) between the ModE-regulated promoter region and the operator site. They were able to show that IHF and ModE worked in concert to bring about the molybdate-dependent regulation. The structure of IHF complexed to DNA shows that this protein induces severe distortion of DNA (Rice *et al.*, 1996). We can speculate that protein–protein interactions between ModE and IHF, which are modulated by the Di-Mop domain, control transcription of this and perhaps other operons. The molybdate dependency would result from the conformation of the Di-Mop domains which in turn depends upon whether molybdate is bound or not.

Concluding remarks

This is the first structure of a ModE protein to be determined. It describes the fold and architecture of the dimer and will aid further biochemical and structural investigations designed to improve understanding of the function of this protein. The fold delineates the molecule into two domains, one for DNA binding (the N-terminus) and one for molybdate binding (the C-terminus). The

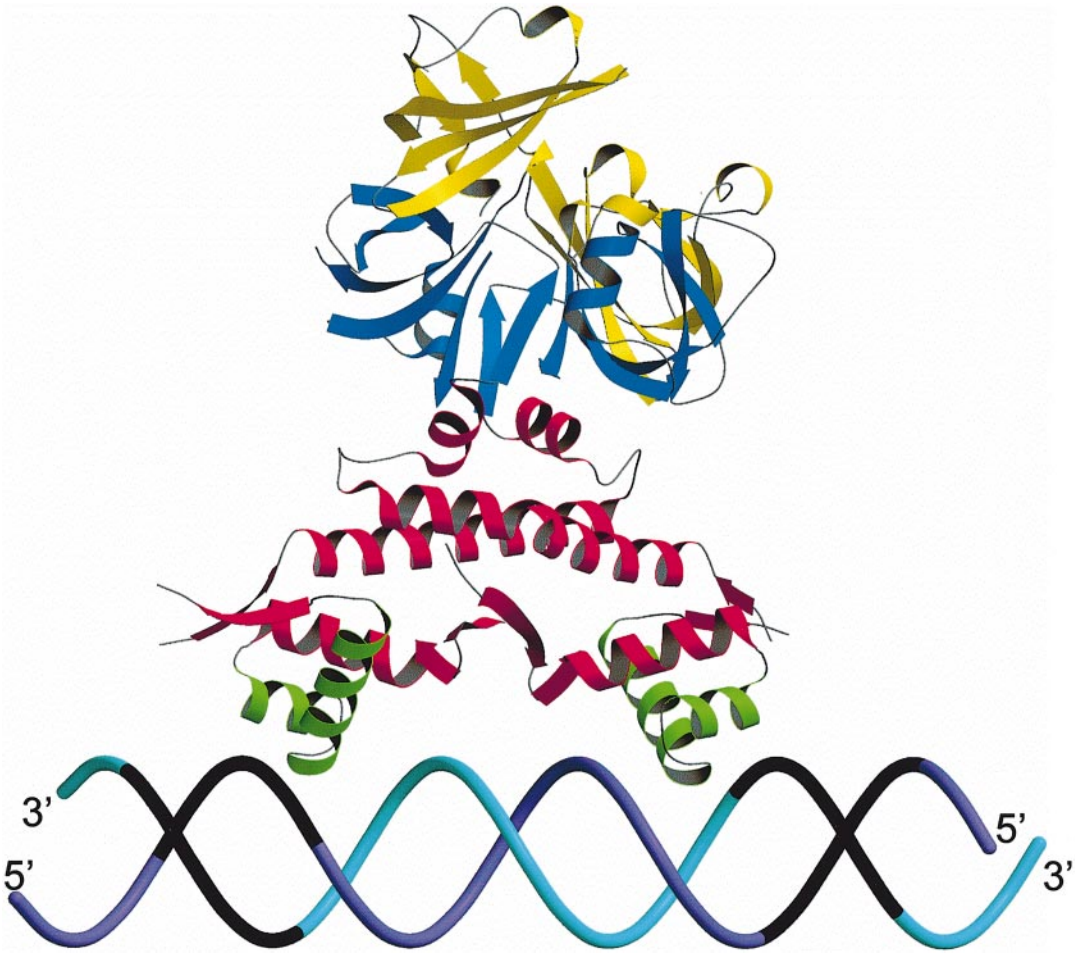


Fig. 6. A model of ModE interacting with double helix DNA. The DNA strands are purple and cyan, the palindromic DNA recognition site is black.

DNA binding domain has been shown to incorporate one winged HTH DNA binding motif per monomer and the residues implicated in DNA binding have been identified. This domain also provides most of the contacts that stabilize the formation of the ModE homodimer. The molybdate binding domain contains the Di-Mop fold and provides a model for the structure of Mop domains, which is an OB-like fold. It has been possible to identify the positions of residues used in molybdate binding from this structure. However, it has not been possible to suggest how the binding of molybdate to ModE effects regulation and it should be borne in mind that ModE will bind to the *modABCD* operon in the absence of molybdate (McNicholas *et al.*, 1998b). Asymmetry in the ModE dimer suggests a degree of conformational freedom in the domain–domain alignment that may allow the protein to adopt distinct structural states depending on the level of molybdate available. There may be a link between distinct structural states and the protein's ability to act as a repressor and activator. There is evidence to suggest that other proteins may be recruited for the regulation of transcription and the crystallographic model now provides a template for future studies to address these points.

ModE has tentatively been classed as a member of the LTTR family of transcriptional regulators (Anderson *et al.*, 1997). However, the structure of the ModE Di-Mop co-factor domain is quite different from the CysB co-factor binding domain structure which is more closely related to that of periplasmic substrate binding proteins (Tyrrell *et al.*, 1997). It is therefore questionable as to whether ModE and CysB should be classified in the same family of transcriptional regulators or if a new family or subset should be defined. Further studies are required to improve on the limited understanding of transcriptional regulators, in particular their classification and structure–function relationships.

Materials and methods

Sample preparation and data collection

The *E.coli* gene had previously been cloned into the T7-promoter based *E.coli* expression pET15b system creating the plasmid pLAA6 (Anderson *et al.*, 1997). The plasmid was transformed into *E.coli* BL21(DE3) and selected on Luria–Bertani (LB) agar plates containing 100 μg/ml ampicillin. A positive colony was grown in LB broth with ampicillin to mid-log phase at which point 0.4 mM (final concentration) isopropyl-β-D-thiogalactopyranoside was added, and cell growth continued with vigorous aeration for 4 h. Cells were harvested by centrifugation (2500 *g*) at 277 K then resuspended in 50 mM Tris–HCl pH 7.6, 250 mM NaCl, 5 mM benzamidine and stored in liquid nitrogen. Bacterial cells were broken by passage through a French press and the insoluble cell debris pelleted by centrifugation at 277 K (18 000 *g*) for 15 min. The cell extract was filtered and then applied to a 5 ml metal chelate affinity column (Hi-Trap, Pharmacia) charged with nickel. The unbound proteins were washed from the column with 50 mM Tris–HCl pH 7.6, 250 mM NaCl, 5 mM benzamidine. The His-tagged protein was eluted with a 0–500 mM imidazole gradient in the same buffer, dialysed overnight in 50 mM Tris–HCl pH 8.0, and then incubated with 15 U of thrombin (Pharmacia) and 2.5 mM $CaCl_2$ for 2 h at 301 K to remove the histidine

Table I. Data collection and processing statistics relevant to refinement of the native and selenomethionyl forms of ModE

	Native	Selenomethionyl form
Unit cell lengths (Å)	81.11, 127.36, 62.95	81.61, 127.24, 62.99
Detector	Mar345	Princeton and Mar CCDs
Wavelength (Å)	1.0000	0.8855 and 0.9310
Resolution range (Å)	29–2.1	24–1.75
No. of measurements	268 449	205 040
No. of unique reflections	37 763	53 298
Coverage overall (%)	99.0 (99.6)	80.2 (66.4)
$I/\sigma(I)$ all data	15.4 (2.0)	31.8 (2.9)
R_{sym} overall (%)	6.7 (42)	5.5 (37)

$R_{sym} = \Sigma|I–<I>|/\Sigma I$, where the summation is over all symmetry equivalent reflections.
Numbers in parentheses correspond to the highest resolution shell, a bin of ~0.1 Å.

Table II. Data collection and processing statistics relevant to the MAD experiment on the selenomethionyl form of ModE

	λ_1	λ_2	λ_3
Wavelength (Å)	0.9779	0.9782	0.8855
Resolution range (Å)	25–2.60	25–2.60	24–2.35
No. of measurements	81 505	71 863	97 163
No. unique reflections	19 797	19 782	26 615
Coverage overall (%)	95.9 (83.3)	95.8 (82.9)	95.7 (81.4)
$I/\sigma(I)$ all data	23.9 (5.1)	24.2 (5.4)	22.4 (3.2)
R_{sym} overall (%)	3.7 (11.5)	3.7 (10.9)	4.1 (15.9)
R_{anom} (%)	4.8	2.7	3.4
R_{iso} (%)	2.5		8.1

Numbers in parentheses correspond to the highest resolution shell, a bin of 0.14 Å.
$R_{anom} = \Sigma|I(+)–I(–)/\Sigma[I(+)–I(–)]$.
$R_{iso} = \Sigma|F_{PH}–F_P|/\Sigma F_P$ with respect to λ_2 as the reference structure factor F_P. F_{PH} is the structure factor of the derivative.

Table III. Phasing statistics for MAD data using an MIR approach

Data set	Phasing power		R_C		
	Acentric	Centric	Acentric	Centric	Anomalous
λ_1	0.82	0.59	0.98	0.98	0.71
λ_2					0.88
λ_3	1.57	1.40	0.76	0.62	0.82

Phasing power is the r.m.s. ($|F_H|/E$), where F_H is the calculated heavy-atom structure factor, i.e. from five Se positions, and E is the residual lack of closure.
$R_C = \Sigma|F_{PH} \pm F_P–F_H|/|F_{PH}–F_P|$.

tag. ModE was separated from thrombin, uncleaved fusion protein and N-terminal peptide by strong anion exchange chromatography using an HQ (Poros) column on a BioCAD 700E workstation. Pooled fractions were concentrated to 20 mg/ml in 50 mM Tris–HCl pH 7.6. Sample purity was assessed by SDS–PAGE and MALDI-TOF spectroscopy. The yield of purified protein is ~25 mg/l of bacterial culture. For selenomethionyl protein production the plasmid pLAA6 was heat shock transformed into the *E.coli* Met$^-$ strain, B834(DE3). The bacteria were cultured in M9 media supplemented with amino acids except that L-selenomethionine was substituted for methionine. Selenomethionyl ModE expression was induced, the sample purified and the N-terminal histidine tag removed using thrombin as described above. MALDI-TOF mass spectroscopy was used to check that the samples were pure and confirmed 100% incorporation of Se for S in the methionines. The use of the hexa-histidine affinity tag in the preparation of ModE results in a Gly-Ser-His extension to the protein. As such these residues have been numbered –3, –2 and –1, respectively where present in the molecular model and residue number 1 corresponds to the initiating methionine of ModE. A benefit for MAD phasing is the incorporation of this additional methionine.

Table IV. Refinement and model geometry statistics

	Native	Se-met form
Protein residues	503	505
Protein atoms	3822	3857
Waters/ions	309/1 Ni^{2+}	426/1 Ni^{2+}
R_{work} (%)/No. of observations	0.214/36 708	0.234/50 614
R_{free} (%)/No. of observations	0.286/1934	0.279/2663
Average isotropic thermal Parameters (Å^2)		
Subunit A overall	37.6	39.3
Subunit B overall	39.6	41.7
Main chain	37.2	39.1
Side chains	40.1	41.9
Solvents	45.9	53.9
Residues in dual conformations[a]	K14A, E174A, W186A, R84B, D173B, E218B, C230B	K14A, R20A, R84B, E218B
Residues truncated to alanine	S-2A, H-1A, Q13A, T68A, K155A, Q165A, E211A, E215A, N236A, T239A, R120B, Q145B, K155B, E215B	H-1A, H-1B, R66B, K155B, Q197B
R.m.s. bond lengths (Å)	0.016	0.016
R.m.s. bond angle associated distances (Å)	0.038	0.033
R.m.s. planarity (Å)	0.044	0.036
Overall *G* factor[b]	–0.34	–0.07
Ramachandran analysis (% and number of residues)		
favourable	90.7 (411)	93.0 (422)
additional	8.8 (40)	6.8 (31)
generous	0.2 (1)	0 (0)
disallowed	0.2 (1)	0.2 (1)

[a]The single letter amino acid code and the number in the sequence are given with A or B to indicate with which subunit the residue is associated.
[b]*G*-factor and Ramachandran analysis were derived using PROCHECK (Laskowski *et al.*, 1993).

Orthorhombic crystals in space group $P2_12_12$ were obtained under similar conditions for both the native (Hall *et al.*, 1999) and selenomethionyl proteins, mixing 3 µl of protein solution (incubated for 1 h with 20 µM Na_2MoO_4) with 1 µl of reservoir solution on a coverslip and then sealing over the reservoir using vacuum grease. The reservoirs used were 250 µl of 10–18% PEG 8000, 50 mM cacodylic acid pH 6.0, 100 mM magnesium acetate. Crystals grow at 277 K in 2 days to a size of 0.4×0.3×0.3 mm. Although molybdate was present in the crystallization conditions there is no evidence of ordered binding in the structure.

Crystals were characterized using station PX9.5 at Daresbury laboratory and data collections subsequently used BM14 and ID14-EH4 at ESRF, Grenoble. Crystals were maintained at 100 K in the presence of glycerol as cryo-protectant. A native dataset to 2.1 Å resolution was measured using a MAR image plate as the detector (Table I). A single crystal of selenomethionyl protein was used for MAD data collection. The choice of wavelengths for the peak (λ_1), inflection point (λ_2) and remote data (λ_3) were derived from a XANES scan on the sample. The three wavelengths were selected so as to obtain the largest value for the anomalous difference, f'' (λ_1) and the minimum value for f' (λ_2 or inflection point). Selecting a far remote wavelength (λ_3) maximized the dispersive difference, $\Delta f'$ ($\lambda_3 – \lambda_2$). The PREDICT program (Dr M.Noble, University of Oxford, UK) was used to determine the range for data collection and each wavelength dataset was collected in a single sweep of 1° oscillations through 120°. A further 44° of data, which extended to 1.75 Å resolution, were recorded from another cryo-cooled Se-met crystal at a later date using a short allocation of commissioning time on

beamline ID14-EH4. These measurements were combined with the data collected from the remote wavelength (λ_3) and subsequently used for the high resolution refinement. All data were processed, reduced and scaled using the HKL suite (Otwinoswki and Minor, 1997) then analysed using CCP4 software (CCP4, 1994). Details are given in Table II.

Location of Se positions, MAD phasing and model building

The homodimer that constitutes the asymmetric unit contains six selenomethionines. Anomalous difference Pattersons were analysed with RSPS (Knight, 1989) and in addition, FA values extracted using REVISE (Fan *et al.*, 1993) and input to the direct methods program RANTAN (Langs *et al.*, 1995). The independent methods yielded three common Se sites. Experimental phases were calculated from the MAD data based on these Se positions using an MIR approach (MLPHARE; Otwinowski, 1991). These phases were used in conjunction with $F(\lambda_3)$–$F(\lambda_2)$ dispersive differences to calculate Fouriers from which two further Se sites were identified. Phasing to 2.6 Å using five Se positions (overall figure of merit 0.53) followed by solvent flattening and histogram matching as implemented in DM (Cowtan, 1994) produced an overall figure of merit of 0.75. Statistics are presented in Table III. Large deviations from non-crystallographic symmetry were observed when the Se atom positions were overlapped and when the electron density map was inspected. Consequently, the map was interpreted using O (Jones *et al.*, 1991) and a model of the dimer constructed. The first model comprised residues 4–71, 77–265 of subunit A and 4–68, 79–140, 150–238 and 251–265 of subunit B.

Refinement and analysis

The native and selenomethionyl forms of ModE were independently refined from the first model onwards using the maximum-likelihood algorithm incorporated in REFMAC (Murshudov *et al.*, 1997). Rounds of refinement were interspersed with interactive computer graphics model building using O. For the selenomethionyl protein refinement with REFMAC was combined with ARPP (Lamzin and Wilson, 1993) for water placement. A number of residues displayed dual conformations and were modelled as such, several residues were poorly defined by the electron density maps and were truncated to alanines. The stereochemistry of both structures was assessed with PROCHECK (Laskowski *et al.*, 1993) and details/statistics for both models are presented in Table IV.

Comments on the quality of the models and identification of a Ni^{2+} ion

The refined native model comprises residues –2A to 68A, 75A to 262A, 1B to 65B, 76B to 137B and 145B to 262B. The selenomethionyl model comprises residues –1A to 68A, 74A to 262A, –1B to 66B, 76B to 137B and 145B to 262B.

In both models a divalent cation is bound to subunit A. The binding site is found at a loop formed by residues 138 to 148, which is not fully observed in subunit B due to disorder. In the Se-met form the ion displays a distorted octahedral environment with ligands provided by side chains of Asp139A, His140A, His146A, Asp148A (bidentate) and a water. Metal–ligand distances fall in the range 2.06–2.31 Å. The water ligand is not observed in the native structure. The binding site in the selenomethionyl model is well defined with the cation having a temperature factor of 38 $Å^2$ and is comparable with those of its environment. In the native model this cation has a temperature factor of 65 $Å^2$ which is ~20 $Å^2$ greater than those of its immediate environment, suggesting that this site is of reduced occupancy. A XANES scan identified the presence of Ni^{2+} in a crystal and a data set to 2.9 Å resolution measured on the high energy side of the absorption edge (λ = 1.483 Å; R_{sym} = 4.7%; R_{ano} = 2.7%; coverage = 99.1%) and an anomalous difference Fourier using phases derived from the model gave a well defined peak, 9σ in height, at the Ni^{2+} site. This site is assumed to have no functional role in ModE and is probably an artefact due to the use of a Ni^{2+}-affinity column during purification.

The r.m.s. deviations between the independently refined native and selenomethionyl model atom positions are 0.41 Å for main chain atoms and 0.77 Å between all protein atoms. In both native and selenomethionyl structures one residue (Asp106A) is in the disallowed region of a Ramachandran plot. This residue is well defined in the density modified experimental maps as well as the refined maps, and is constrained in this conformation by a salt bridge formed with Arg80B and by a hydrogen bond donated from the amide of Asp106A to the carbonyl of Val102A.

Structure factors corresponding to the three-wavelength MAD experiment, and data used in native and Se-met refinements together with two sets of coordinates have been deposited.

Acknowledgements

We thank Dr A.Mehlert for mass spectroscopy expertise, Dr R.Savva for a sample of *E.coli* B834, Dr R.Pau for discussions and contributions, Dr D.Norman for the DNA model, Dr S.McSweeney for assistance in high-resolution data collection, Dr D.Lawson and a referee for constructive comments. Funded by the BBSRC, the Wellcome Trust, EPSRC, the CCLRC-Daresbury Laboratory and the University of Dundee. We are grateful for support and access to the synchrotron sources at Daresbury, UK, and ESRF, France.

References

Anderson,L.A., Palmer,T., Price,N.C., Bornemann,S., Boxer,D.H. and Pau,R.N. (1997) Characterisation of the molybdenum-responsive ModE regulatory protein and its binding to the promotor region of the *modABCD* (molybdenum transport) operon of *Escherichia coli*. *Eur. J. Biochem.*, **246**, 119–126.

Ansari,A.Z., Bradner,J.E. and O'Halloran,T.V. (1995) DNA-bend modulation in a repressor-to-activator switching mechanism. *Nature*, **374**, 371–375.

Brennan,R.G. (1993) The winged-helix DNA-binding motif: another helix–turn–helix takeoff. *Cell*, **74**, 773–776.

Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

Copley,R.R. and Barton,G.J. (1994) A structural analysis of phosphate and sulphate binding sites in proteins. *J. Mol. Biol.*, **242**, 321–329.

Cowtan,K. (1994) DM program. *Joint CCP4 ESF-EACBM Newsletter*, **31**, 34–38.

Ding,X., Zeng,H., Schiering,N., Ringe,D. and Murphy,J.R. (1996) Identification of the primary metal ion-activation sites of the diphtheria tox repressor by X-ray crystallography and site-directed mutational analysis. *Nature Struct. Biol.*, **3**, 382–387.

Fan,H.-F., Woolfson,M.M. and Yao,J.-X. (1993) New techniques of applying multiwavelength anomalous scattering data. *Proc. R. Soc. Lond.*, **A442**, 13–32.

Fleischmann,R.D. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, **269**, 496–512.

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

Hall,D.R. (1999) Structure and Function of ModE, a Diferric Serum Transferrin and a Class II Aldolase. Thesis. University of Dundee, UK.

Hall,D.R., Gourley,D.G., Duke,E.M.H., Leonard,G.A., Anderson,L.A., Pau,R.N., Boxer,D.H. and Hunter,W.N. (1999) Two crystal forms of the molybdate dependent transcriptional regulator (ModE) from *Escherichia coli*. *Acta Crystallogr.*, **D55**, 542–543.

Hazes,B., Boodhoo,A., Cockle,S.A. and Read,R.J. (1996) Crystal structure of the pertussis toxin-ATP complex: A molecular sensor. *J. Mol. Biol.*, **258**, 661–671.

Higgins,C.F. (1992) ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol.*, **8**, 67–113.

Hinton,S.M. and Freyer,G. (1986) Cloning, expression and sequencing the molybdenum-pterin-binding protein (mop) gene of *Clostridium pasteurianum* in *Escherichia coli*. *Nucleic Acids Res.*, **14**, 9371–9380.

Hinton,S.M. and Merritt,B. (1986) Purification and characterisation of a molybdenum-pterin binding-protein (mop) in *Clostridium pasteurianum* W5. *J. Bacteriol.*, **168**, 688–693.

Hol,W.G.J., van Duijnen,P.T. and Berendsen,H.J.C. (1978) The α-helix dipole and the properties of proteins. *Nature*, **273**, 443–446.

Holm,L. and Sander,C. (1993) Protein-structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.

Hu,Y., Rech,S., Gunsalus,R.P. and Rees,D.C. (1997) Crystal structure of the molybdate binding protein ModA. *Nature Struct. Biol.*, **4**, 703–707.

Hutchinson,E.G. and Thornton,J.M. (1996) PROMOTIF—A program to identify and analyze structural motifs in proteins. *Protein Sci.*, **5**, 212–220.

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

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure–pattern-recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Kisker,C., Schindelin,H. and Rees,D.C. (1997) Molybdenum-cofactor-containing enzymes: structure and mechanism. *Annu. Rev. Biochem.*, **66**, 233–267.

Knight,S. (1989) Ribulose, 1,5-bisphosphate carboxylase/oxygenase—a structural study. Thesis. Swedish University of Agricultural Sciences, Uppsala, Sweden.

Kolb,A., Busby,S., Buc,H., Garges,S. and Adhya,S. (1993) Transcriptional regulation by cAMP and its receptor protein. *Annu. Rev. Biochem.*, **62**, 749–795.

Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Lamzin,V.S. and Wilson,K.S. (1993) Automated refinement of protein models. *Acta Crystallogr.*, **D49**, 129–147.

Langs,D.A., Miller,R., Hauptman,H.A. and Han,G.W. (1995) Use of the minimal function for partial structure development in direct-methods. *Acta Crystallogr.*, **A51**, 81–87.

Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Lawson,D.M., Williams,C.E., White,D.J., Choay,A.P., Mitchenall,L.A. and Pau,R.N. (1997) Protein ligands for molybdate. Specificity and charge stabilisation at the anion-binding sites of periplasmic and intracellular molybdate-binding proteins of *Azotobacter vinelandii*. *J. Chem. Soc. Dalton Trans.*, **21**, 3981–3984.

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

McNicholas,P.M., Chiang,R.C. and Gunsalus,R.P. (1998a) Anaerobic regulation of the *Escherichia coli dmsABC* operon requires the molybdate-responsive regulator ModE. *Mol. Microbiol.*, **27**, 197–208.

McNicholas,P.M., Mazzotta,M.M., Rech,S.A. and Gunsalus,R.P. (1998b) Functional dissection of the molybdate-responsive transcription regulator, ModE, from *Escherichia coli*. *J. Bacteriol.*, **180**, 4638–4643.

Martinez-Hackert,E. and Stock,A.M. (1997) The DNA-binding domain of OmpR: crystal structure of a winged helix transcription factor. *Structure*, **5**, 109–124.

Merritt,E.A. and Murphy,M.E.P (1994). Raster3d, a program for photorealistic molecular graphics. *Acta Crystallogr.*, **D50**, 869–873.

Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.*, **D53**, 240–255.

Murzin,A.G. (1993) OB (oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.*, **12**, 861–867.

Otwinowski,Z. (1991) Maximum likelihood refinement of heavy atom parameters. In Wolf,W., Evans,P.R. and Leslie,A.G.W. (eds), *Isomorphous Replacement and Anomalous Scattering, Proceedings of the CCP4 Study Weekend, January, 1991*. SERC Daresbury Laboratory, Warrington, UK, pp. 80–86.

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

Papageorgiou,A.C., Brehm,R.D., Leonidas,D.D., Tranter,H.S. and Acharya,K.R. (1996) The refined crystal structure of toxic shock syndrome toxin-1 at 2.07 Å resolution. *J. Mol. Biol.*, **260**, 553–569.

Parkinson,G., Wilson,C., Gunasekera,A., Ebright,Y.W., Ebright,R.E. and Berman,H.M. (1996) Structure of the CAP–DNA complex at 2.5 Å resolution: a complete picture of the protein-DNA interface. *J. Mol. Biol.*, **260**, 395–408.

Pau,R.N., Klipp,W. and Leimkuhler,S. (1998) Molybdenum transport, processing and gene regulation. In Winkelmann,G. and Carrano,C.J. (eds), *Transition Metals in Microbial Metabolism*. Harwood Academic Publishers, UK, pp. 217–234.

Rajagopalan,K.V. (1996) Biosynthesis of the molybdenum cofactor in *Escherichia coli* and *Salmonella typhimurium*. In Neidhardt,F.C. (ed.), *Echerichia coli and Salmonella: Cellular and Molecular Biology*. 2nd edn. American Society of Microbiology Press, Washington DC, pp. 674–679.

Rice,P.A., Yang,S.-W., Mizuuchi,K. and Nash,H.A. (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell*, **87**, 1295–1306.

Roussel,A., Anderson,B.F., Baker,H.M., Fraser,J.D. and Baker,E.N. (1997) Crystal structure of the streptoccocal superantigen SPE-C: dimerisation and zinc binding suggest a novel mode of interaction with MHC class II molecules. *Nature Struct. Biol.*, **4**, 635–643.

Schell,M.A. (1993) Molecular biology of the LysR family of transcriptional regulators. *Annu. Rev. Microbiol.*, **47**, 597–626.

Schultz,S.C., Shields,G.C. and Steitz,T.A. (1991) Crystal structure of a CAP–DNA complex: the DNA is bent by 90 degrees. *Science*, **253**, 1001–1007.

Stein,P.E., Boodhoo,A., Tyrrell,G.J., Brunton,J.L. and Read,R.J. (1992) Crystal structure of the cell-binding B oligomer of verotoxin-1 from *E. coli*. *Nature*, **355**, 748–750.

Stein,P.E., Boodhoo,A., Armstrong,G.D., Cockle,S.A., Klein,M.H. and Read,R.J. (1994) The crystal-structure of pertussis toxin. *Structure*, **2**, 45–57.

Tyrrell,R., Verschueren,K.H.G., Dodson,E.J., Murshudov,G.N., Addy,C. and Wilkinson,A.J. (1997) The structure of the cofactor-binding fragment of the LysR family member, CysB: a familiar fold with a surprising subunit arrangement. *Structure*, **5**, 1017–1032.

Van den Akker,F., Sarfaty,S., Twiddy,E.M., Connell,T.D., Holmes,R.K. and Hol,W.G. (1996) Crystal structure of a new heat-labile enterotoxin, LT-IIb. *Structure*, **4**, 665–678.

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

White,A., Ding,X.C., vanderSpek,J.C., Murphy,J.R. and Ringe,D. (1998) Structure of the metal-ion-activated diphtheria toxin repressor tox operator complex. *Nature*, **394**, 502–506.

Received December 21, 1998; revised and accepted January 26, 1999