Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the effector-binding domain of the LysR-type transcription factor RovM from *Yersinia pseudotuberculosis*

Nick Quade,[a] Marieke Dieckmann,[a] Matthias Haffke,[a] Ann Kathrin Heroven,[b] Petra Dersch[b] and Dirk W. Heinz[a]*

[a]Department of Molecular Structural Biology, Helmholtz Centre for Infection Research, D-38124 Braunschweig, Germany, and [b]Department of Microbiology, Helmholtz Centre of Infection Research, D-38124 Braunschweig, Germany

Correspondence e-mail: dirk.heinz@helmholtz-hzi.de

Received 2 September 2010
Accepted 28 November 2010

PDB Reference: effector-binding domain of RovM, 3onm.

In enteropathogenic *Yersinia*, the expression of several early-phase virulence factors such as invasin is tightly regulated in response to environmental cues. The responsible regulatory network is complex, involving several regulatory RNAs and proteins such as the LysR-type transcription regulator (LTTR) RovM. In this study, the crystal structure of the effector-binding domain (EBD) of RovM, the first LTTR protein described as being involved in virulence regulation, was determined at a resolution of 2.4 Å. Size-exclusion chromatography and comparison with structures of full-length LTTRs show that RovM is most likely to adopt a tetrameric arrangement with two distant DNA-binding domains (DBDs), causing the DNA to bend around it. Additionally, a cavity was detected in RovM which could bind small inducer molecules.

1. Introduction

Yersinia pseudotuberculosis and *Y. enterocolitica* are enteropathogenic food-borne bacteria that cause diarrhoea, enteritis and autoimmune disorders (Salyers & Whitt, 2002). Once in the intestine, the bacteria have to penetrate through the epithelial cells to reach the underlying lymphoid tissues (Heroven & Dersch, 2010). This critical step is mainly mediated by the outer membrane protein invasin, which binds to $\beta1$ integrins on the surface of M cells.

The expression of *invasin* (*inv*) is strictly dependent on the environment and the stage of infection and is controlled by a complex regulatory network that integrates various external signals, including temperature, nutrient availability and growth phase. A key player in this network is the MarR-type protein RovA, which directly activates the *inv* gene as well as its own expression (Heroven *et al.*, 2004). The temperature control of *inv* expression is directly governed by RovA (Herbst *et al.*, 2009). At moderate temperatures RovA is able to bind to DNA and activate transcription of *inv* and *rovA*. Upon a temperature shift from 298 to 310 K it undergoes a conformational change, resulting in a strong reduction in DNA-binding affinity and an increased susceptibility to proteolytic degradation by the Lon protease.

Another important element in this regulatory network is the LysR-type protein RovM. It represses RovA activity and thereby *inv* expression together with the nucleoid-associated H-NS protein, which is a repressor of *rovA* and *inv* (Heroven & Dersch, 2006). RovM has homologues in other bacteria such as HexA from *Erwinia carotovora* (65% sequence identity),

PecT from *E. chrysanthemi* (66% identity) and LrhA from *Escherichia coli* (65% identity), which are involved in different cellular processes such as biofilm formation and motility (Lehnen *et al.*, 2002; Mukherjee *et al.*, 2000; Surgey *et al.*, 1996).

LysR-type proteins consist of two domains: a winged helix–turn–helix (wHTH) DNA-binding domain (DBD) followed by an inducer-binding effector-binding domain (EBD). These domains are connected *via* a long flexible linker helix that is involved in the oligomerization of the proteins. LysR-type proteins have been shown to adopt a wide range of oligomerization states, such as dimers (Zhou *et al.*, 2010), tetramers (Muraoka *et al.*, 2003; Monferrer *et al.*, 2010) and even octamers (Sainsbury *et al.*, 2009). They regulate a variety of genes bearing functions such as antibiotic resistance, aromatic compound catabolism and amino-acid metabolism (Schell, 1993; Clark *et al.*, 2002; Maddocks & Oyston, 2008). However, RovM was the first LysR-type protein described as being involved in regulation of virulence factors (Heroven & Dersch, 2006). Since then, several LysR-type transcription regulators (LTTRs) have been described as being involved in virulence regulation, indicating the importance of these proteins for bacterial virulence (Kendall *et al.*, 2010; Burnside *et al.*, 2010; O'Grady *et al.*, 2010).

RovM has been shown to recognize a region about 50 bp upstream of the *rovA* promoter P1 containing two palindromic sequences which are similar to the binding sites of other LysR-type proteins (Heroven & Dersch, 2006). Bending of DNA upon LysR-type protein binding has been described (Deghmane *et al.*, 2004) and RovM also seems to bend its binding site, as hyper-reactive bases were detected in the middle of the RovM-binding sequence in the *rovA* promoter region in DNase I protection assays (Heroven & Dersch, 2006). LysR-type proteins often negatively regulate their own expression (Schell, 1993). RovM, on the other hand, does not bind directly to its own promoter, but can activate its own expression, probably *via* other transcriptional regulators (Heroven & Dersch, 2006).

One important role of RovM seems to be the control of *rovA* expression in response to the availability of nutrients and the growth phase (Heroven *et al.*, 2008). The carbon-storage regulator system (Csr) has been shown to be involved in this control mechanism. The RNA-binding protein CsrA activates the expression of *rovM via* an as yet unknown mechanism. On the other hand, small regulatory RNAs called CsrB and CsrC can bind and sequester multiple CsrA molecules and prevent them from activating *rovM* expression. CsrC synthesis is strongly dependent on growth phase and medium composition, whereas the environmental signal for CsrB production is still unclear. Additionally, RovM has been shown to activate flagellar motility by an as yet unknown mechanism (Heroven & Dersch, 2006; Heroven *et al.*, 2008).

Interestingly, RovA as well as RovM are identical in *Y. pseudotuberculosis* and *Y. pestis*, the causative agent of bubonic plague, even though the *inv* gene is nonfunctional in *Y. pestis*. Furthermore, RovA has been shown to be indispensable for *Y. pseudotuberculosis* and *Y. pestis* infection, as *rovA* mutants of both species were strongly attenuated in virulence (Ellison *et al.*, 2004; Heroven & Dersch, 2006). This shows that the RovA/RovM regulatory system plays a crucial role in *Yersinia* virulence in addition to the well studied invasin.

To elucidate the role of RovM in this complex regulatory network, we investigated the structural and biophysical properties of RovM from *Y. pseudotuberculosis*. Here, we describe the production, purification and analysis of the oligomeric state by size-exclusion chromatography of full-length RovM (36 kDa) and the effector-binding domain of RovM (RovM-EBD; residues 93–333; 24 kDa). Furthermore, we solved the three-dimensional structure of RovM-EBD. Comparison with the structure of a full-length LTTR indicates that full-length RovM probably also adopts a similar fold consisting of a dimer of dimers. Additionally, a cavity suitable for the binding of an as yet unknown inducer molecule was detected, pointing towards an even more complex regulation of the Rov/Inv system.

2. Materials and methods

2.1. Cloning and expression

The full-length RovM sequence from *Y. pseudotuberculosis* YPIII pIB1 cloned into a pET28 vector, named pAKH43 (Heroven & Dersch, 2006), was used for the production of full-length RovM. The sequence for RovM-EBD (RovM 92–310) was generated from this vector by polymerase chain reaction using the following primers: forward, 5′-GGGAATTC**CATATG**TACA-3′; reverse, 5′-AGGAGA**CTCGAG**GACTA-3′. The obtained insert was digested with *Nde*I and *Xho*I (cleaving at the sequences shown in bold) and cloned into a vector generating RovM-EBD in frame with an N-terminal thrombin cleavage site and a His_6 tag. The N-terminal amino acids preceding Met92 had the sequence MGSS**HHHHHH**SSG**LVPRGS**H (the His_6 tag and the thrombin cleavage site are shown in bold). Full-length RovM and RovM-EBD were both expressed in *E. coli* BL21 (DE3) cells. The cells were grown in LB medium containing 30 µg ml^{-1} kanamycin at 310 K to an optical density at 600 nm of ∼0.8. Following this, protein production was induced by addition of 2 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG). The cells were grown for an additional 2 h and harvested by centrifugation at 6000*g* at 277 K.

2.2. Purification

The two proteins were purified under similar conditions. The cell pellet was resuspended in cold lysis buffer consisting of 50 m*M* NaH_2PO_4 pH 8, 1 m*M* $MgCl_2$, 300 m*M* NaCl, 10 m*M* imidazole and a Complete EDTA-free protease-inhibitor tablet (Roche) and lysed by two passes through a homogenizer at 138 MPa. The lysate was centrifuged at 60 000*g* for 45 min at 277 K to remove cell debris. The supernatant was loaded onto a column containing 10 ml Ni–NTA Sepharose Superflow resin (Qiagen) pre-equilibrated with lysis buffer. The column was washed three times with five column volumes

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	SAD	Native
X-ray source	Rigaku MicroMax-007 HF [Cu $K\alpha$]	ESRF ID-29
Space group	$I4_122$	$I4_122$
Unit-cell parameters (Å)	$a = b = 69.03$, $c = 352.49$	$a = b = 69.47$, $c = 351.22$
Wavelength (Å)	1.54	0.97
Resolution (Å)	88.6–2.5	87.8–2.4 (2.53–2.4)
Measured reflections	192810 (23819)	182593 (22862)
Unique reflections	26240 (4108)	17557 (2485)
$\langle I/\sigma(I)\rangle$	28.7 (4.9)	16.4 (3.5)
Completeness (%)	99.4 (96.7)	99.9 (100)
Multiplicity	7.3 (5.8)	10.4 (9.2)
$R_{\rm merge}$ (%)	6.3 (41.6)	7.4 (54.0)
Wilson B factor (Å^2)	33.75	59.3
Solvent content (%)	43	43
$R_{\rm work}$ (%)		22.1 (23.4)
$R_{\rm free}$ (%)		29.5 (30.2)
Protein atoms		2142
Solvent atoms		142
R.m.s.d. from ideal geometry		
Bond lengths (Å)		0.014
Bond angles (°)		1.575
Average B factor (Å^2)		
Protein (chain A)		56.4
Protein (chain B)		62.7
Solvent		55.3
Ramachandran plot (%)		
Favoured region		97.8
Allowed region		2.2

of wash buffer consisting of 50 m*M* NaH_2PO_4 pH 8, 300 m*M* NaCl and 20 m*M* imidazole. The bound protein was eluted with five column volumes of elution buffer consisting of 50 m*M* NaH_2PO_4 pH 8, 300 m*M* NaCl and 250 m*M* imidazole.

For further purification using an anion-exchange column, the proteins were diluted in a 1:2 ratio with 20 m*M* Tris pH 8 and 5 m*M* DTT to reduce the salt concentration. The proteins were loaded onto a MonoQ 10/100 column (GE Healthcare) and eluted with a linear gradient from 100 m*M* to 1 *M* NaCl in 20 m*M* Tris pH 8 and 5 m*M* DTT.

The proteins were finally purified and their oligomerization states were estimated by size-exclusion chromatography using a calibrated Superdex 200 16/60 column (GE Healthcare) pre-equilibrated with 20 m*M* Tris pH 8, 100 m*M* NaCl and 5 m*M* DTT. The column was calibrated using gel-filtration calibration kits from GE Healthcare and the results were analysed according to the manual using *Microsoft Excel*. Sample fractions were analyzed for purity by SDS–PAGE and pure fractions were pooled and concentrated to 14 mg ml^{-1} for the full-length protein and 22 mg ml^{-1} for the effector-binding domain using Vivaspin 10 kDa molecular-weight cutoff concentrators (Sartorius).

2.3. Crystallization

Several commercial 96-well sitting-drop crystal screens (JCSG Core, PEGs and PEGs II Suites from Qiagen) for RovM-EBD were set up at 291 K. Initial crystals were reproduced and optimized using the hanging-drop vapour-diffusion method. The crystals used for data collection were grown using a mixture of 2 µl protein solution (22 mg ml^{-1} in 20 m*M* Tris pH 8, 100 m*M* NaCl and 5 m*M* DTT) and 2 µl reservoir solution consisting of 0.1 *M* HEPES pH 8.0 and 12.5% PEG 6000. The protein–reservoir mixture was equilibrated against 500 µl reservoir solution. Initial crystals grew at 291 K after one week and reached their final size within three weeks. Samarium-derivatized crystals were obtained by transferring crystals into a drop with reservoir solution containing 50 m*M* $SmCl_3$ and soaking for 10 min.

2.4. Data collection, processing and structure determination

Before measurement, the crystals were transferred into a drop consisting of 0.1 *M* HEPES pH 8.0, 25% PEG 6000 and 10% glycerol. The crystals were flash-cooled in liquid nitrogen and a native X-ray diffraction data set was collected to 2.4 Å resolution at 100 K on beamline ID-29 at the ESRF (Grenoble, France) equipped with an ADSC Q315R detector. A SAD data set was collected from the Sm-soaked crystals to 2.5 Å resolution using a MicroMax-007 HF rotating copper-anode X-ray generator (Rigaku) with a Saturn 944+ CCD detector (Rigaku).

The RovM-EBD crystals belonged to space group $I4_122$, with unit-cell parameters $a = b = 69.47$, $c = 351.22$ Å. The data sets were processed with *XDS* (Kabsch, 2010) and scaled with *SCALA* (Evans, 2006). For the calculation of the free *R* factor ($R_{\rm free}$) a randomly chosen 5% of the reflections were set aside. The calculated Matthews coefficient ($V_{\rm M}$; Matthews, 1968) of 2.12 Å^3 Da^{-1} indicated the presence of two monomers per asymmetric unit, corresponding to a solvent content of 43%. Molecular replacement using PDB entries 1al3 (CysB), 2fyi (Cbl), 2h98 (CatM) and 2h9b (BenM) as models failed to provide useful phases. Therefore, the samarium-derivatized crystal was used for phasing. Heavy-atom positions and initial phases were obtained using *SHELXC*/*D*/*E* (Sheldrick, 2008) as implemented in *CCP*4 (Collaborative Computational Project, Number 4, 1994). The resulting maps were of low quality and were not interpretable. Thus, the four best Sm positions were selected and were used for the refinement of heavy-atom positions, the detection of additional heavy-atom sites and phasing in *Phaser* (McCoy *et al.*, 2007). *Phaser* detected 14 Sm sites and the resulting phases were improved by density modification in *Parrot* (Cowtan, 2010). The improved density showed clear features such as helices, strands and even some side chains. *Buccaneer* (Cowtan, 2006) was successfully used to build most of the structure. The structure was completed in *Coot* (Emsley *et al.*, 2010) and was refined with *REFMAC*5 (Murshudov *et al.*, 1997). After several rounds of model building, the structure was used for phasing of the native data set by molecular replacement with *MOLREP* (Vagin & Teplyakov, 2010). A large part of the *B* chain that had poor density in the Sm data set now showed clear density, indicating some larger changes in the crystal that were induced by Sm soaking. The structure was further refined against the native data set and finally water molecules were added manually and using the water-find tool in *Coot*. Addi-

tional refinement, including a final TLS refinement (Painter & Merritt, 2006*a*) step using four TLS groups which were determined by the *TLS Motion Determination Server* (Painter & Merritt, 2006*b*), was carried out with *REFMAC*. The final *R* values were R_{work} = 22.0% and R_{free} = 29.5%.

Data-collection and refinement statistics are listed in Table 1. The final model was checked and validated using *MolProbity* (Chen *et al.*, 2010; Davis *et al.*, 2007), which indicated a good-quality model. All residues fall within allowed regions of the Ramachandran plot (Ramachandran & Sasisekharan, 1968). The *PISA* server was used to assess the relevance of the crystal contacts (Krissinel & Henrick, 2007).

The atomic coordinates and structure factors have been deposited in the Protein Data Bank (Bernstein *et al.*, 1977) under accession code 3onm.

3. Results

3.1. RovM-EBD structure

We have solved the crystal structure of RovM-EBD from *Y. pseudotuberculosis* at 2.4 Å resolution with two RovM-EBD monomers in the asymmetric unit. Both chains adopt the typical conformation of the effector-binding domain of LTTR EBDs (Fig. 1*a*; Ezezika *et al.*, 2007). Each chain is roughly divided into two $\alpha\beta$-domains, EBD-I and EBD-II, with two crossovers between them. EBD-I contains a five-stranded β-sheet with three α-helices around it. EBD-II also contains a five-stranded β-sheet, which is strongly twisted, and four α-helices (Fig. 1*b*). However, the two chains exhibit slight conformational differences arising from the crystal packing. For chain *A* residues 99–289 could be modelled, with the exception of residue 159, which was not visible in the density. Chain *B* has been modelled from residues 100 to 280, with the exception of residue 128. Probably as a consequence of missing crystal contacts, the last helix in the C-terminal α-helix of chain *B* is less well resolved than that in chain *A* and could not be fully modelled. A comparison of the two chains shows differences in several regions, mostly in EBD-I and particularly in the helices, whereas EBD-II remains mostly identical (Fig. 1*c*). The r.m.s.d. between chains *A* and *B* is 0.91 Å for common C^{α} atoms.

3.2. Oligomeric states of RovM-FL and RovM-EBD

The oligomeric states of full-length RovM and RovM-EBD were estimated using a calibrated size-exclusion chromatography column. The calculated oligomeric state of full-length RovM was 4.7, indicating the presence of tetrameric RovM in solution, as has been demonstrated for some other LTTRs such as CbnR (Muraoka *et al.*, 2003). In contrast, the EBDs of LTTRs have been shown to exist as dimers in solution (Stec *et al.*, 2006; Ezezika *et al.*, 2007). The oligomeric state of RovM-EBD was calculated to be 2.2, which indicates that RovM-EBD is also a dimer in solution. Indeed, the crystal structure shows two RovM-EBD molecules in the asymmetric unit. The chains mostly interact *via* helix 1, sheet 4 and sheet 9 as well as the loop preceeding sheet 9, covering an interface area of 845.7 $Å^2$ (Fig. 2*b*). However, a comparison with other solved EBDs of LTTRs showed that the dimer formation is different. Indeed, after inspecting the crystallographic symmetry mates it became clear that RovM-EBD exhibits several other relevant interactions in the crystal (Fig. 2). The *PISA* server was used to assess these interfaces, showing that the interface

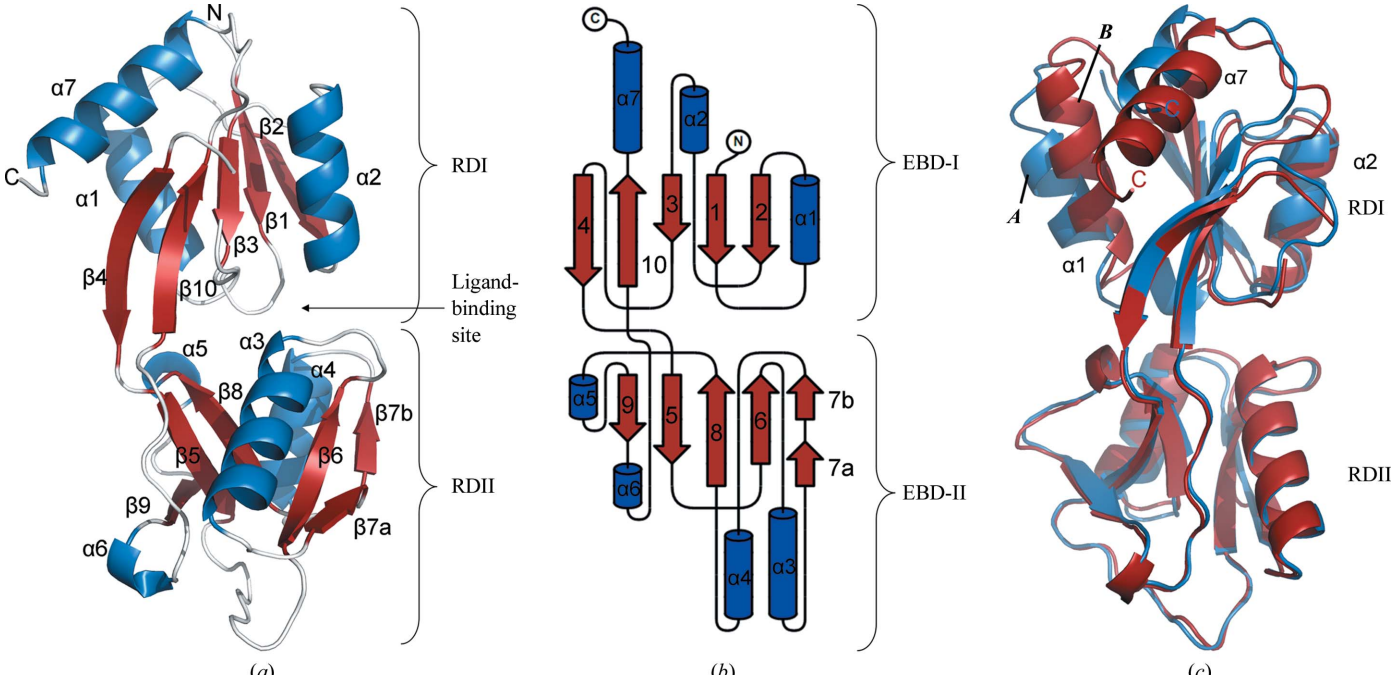

Figure 1
Structure of the RovM-EBD monomer. (*a*) Cartoon representation. (*b*) Topology plot generated by *TopDraw* (Bond, 2003). (*c*) Superposition of chains *A* and *B*. Chain *A* is shown in blue and chain *B* in red.

between chain *A* and chain *B* is roughly as extensive as the interface between chain *B* and its symmetry mate *B*′. The *BB*′ interface is the same as that shown for other LTTR EBDs (Fig. 2). Here, two molecules are arranged in a head-to-tail fashion, with the EBD-I of one molecule interacting *via* helix 1 with helix 4 of the EBD-II of the other molecule with an interface area of 883.8 Å^2. In contrast, the interaction between chain *A* and its symmetry mate *A*′ is different. Here, the interaction is still mediated by helices 1 and 4, but the two molecules are shifted in such a way that sheet 2 can interact with sheet 2 of the symmetry mate, covering an area of 945.6 Å^2 (Fig. 2). This interface was considered to be less relevant by the *PISA* server. Thus, in solution RovM-EBD probably exists as a dimer like *BB*′ and the comparably strong interaction between *A* and *B* in the crystal leads to a distortion in the geometry of the second dimer *AA*′ (Fig. 2). Inspection of the other symmetry mates revealed further interactions between two *A* or two *B* subunits (termed *A* and *A*″ or *B* and *B*″; Fig. 2), which are very similar for *AA*″ and *BB*″. This interaction is mediated by helices 2 and 3, covering interfaces of 681.8 and 642.7 Å^2 for *AA*″ and *BB*″, respectively.

3.3. Ligand-binding site

LTTR proteins are known to regulate gene expression in response to small ligands called inducers (Schell, 1993). These inducers are often related to the genes that they influence, such as benzoate in aromatic compound degradation (Ezezika *et al.*, 2007). These inducer molecules bind in a cavity between the two domains (EBD-I and EBD-II) of the EBD (Fig. 1*a*). Ligand binding is thought to induce a conformational change in the EBD, which leads to a repositioning of the DBDs and a relieved bending of the DNA (Schell, 1993). No inducer molecule for RovM is known to date. However, the protein structure of RovM-EBD shows a cavity (roughly 10.5 Å wide, 7.0 Å high and 6.5 Å deep) between EBD-I and EBD-II

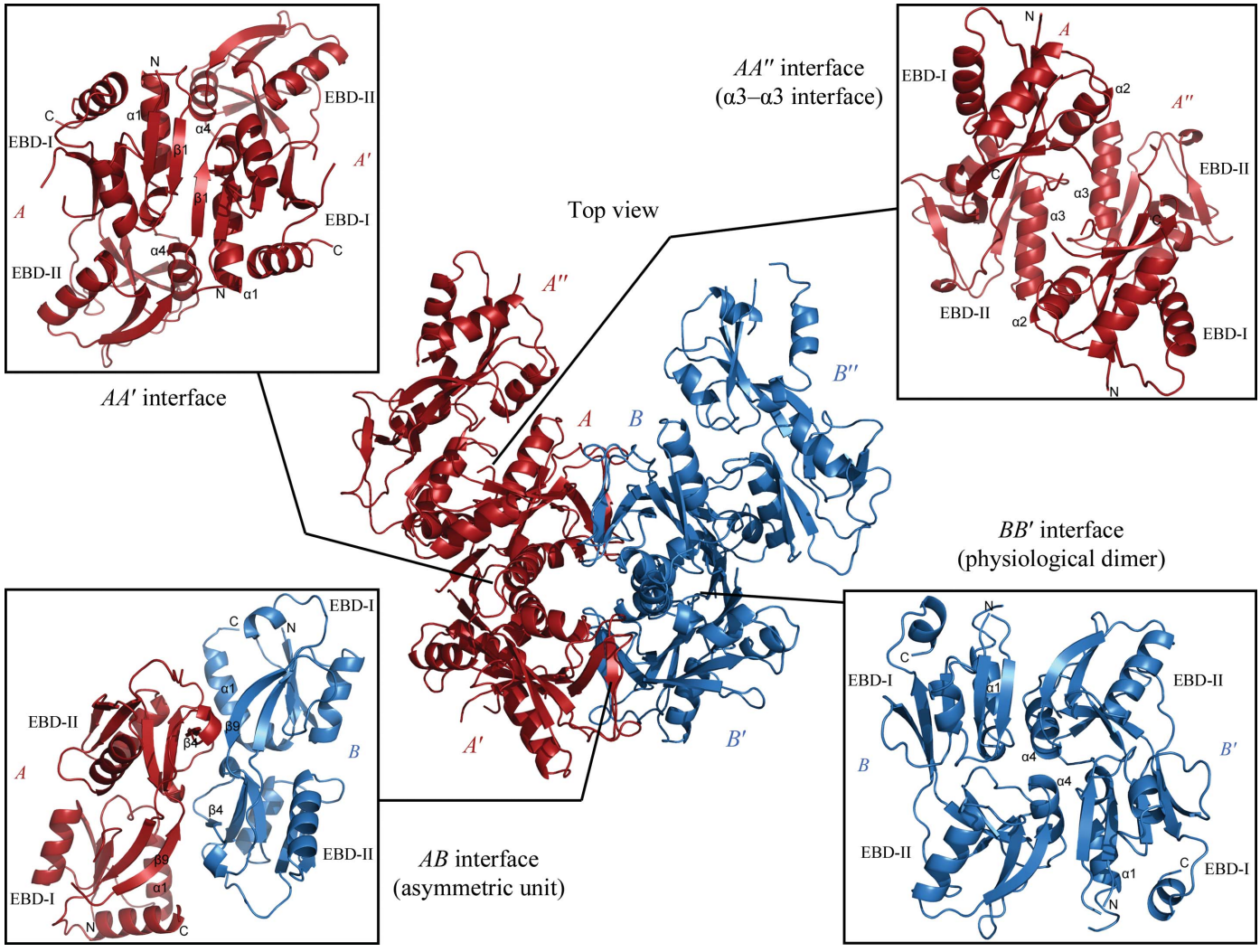

Figure 2
Structures of the interfaces between RovM-EBD monomers in the crystal. Chain *A* is shown in blue and chain *B* in red. The *AB* interface is seen in the asymmetric unit. The *BB*′ interface is seen in all structures of LTTRs and is thus termed the physiological dimer. The *AA*′ interface is similar to the *BB*′ interface but the two EBDs are shifted relative to each other. The *AA*″ interface (and the *BB*″ interface, which is not enlarged) is seen in some other LTTRs and seems to mediate tetramer formation.

(Fig. 6*b*) which could accommodate a small-molecule ligand. It is lined with two hydrophobic residues (Tyr198 and Leu222) and several hydrophilic residues (Asn106, Asn107, Ser137, Thr155 and Arg238), suggesting that a rather hydrophilic inducer molecule could bind in this cavity.

4. Discussion

4.1. Comparison with other LTTR structures

The structure of RovM-EBD was superposed with the structure of the EBD of CynR from *E. coli* (Knapp & Hu, 2009), with which it shares the highest sequence identity (21%) of all known LTTR EBD structures (Fig. 3*a*), with an r.m.s.d. of 2.15 Å for all common C^{α} atoms. Minor structural deviations are only found for exposed loop regions. CynR has been shown to exist as a dimer in solution. Thus, the structure of the CynR dimer was aligned with the *BB*′ dimer of RovM, which indicates that the *BB*′ interface is indeed the physiological interface, which is conserved among all LTTR EBDs crystallized to date (Fig. 3*b*).

An alignment of the sequences of RovM and its homologues CynR (*E. coli*; Knapp & Hu, 2009), CbnR (*Ralstonia eutropha*; Muraoka *et al.*, 2003) and BenM (*Acinetobacter baylyi*; Ezezika *et al.*, 2007) shows a high degree of conservation within the DBDs, indicating a conserved DNA-binding mechanism (Fig. 4). In contrast, the EBD has only a few conserved residues, which are mostly hydrophobic residues in the core of the protein. In particular, the C-terminal region is very dissimilar between the homologues. This distribution of conserved residues has also been observed in other classes of transcription factors such as the MarR family, which have a mostly conserved DBD with only a few mutations to adapt to new promoter sequences and a weakly conserved ligand-binding domain that allows binding of very different inducer molecules (Wu *et al.*, 2003).

Of the known LTTR structures, full-length RovM has the highest sequence identity to CbnR (20.4%; Fig. 4). A superposition of RovM with the structure of a monomer of full-length CbnR shows that the EBD-Is align quite well, whereas the EBD-IIs seem to be tilted with respect to the EBD-Is (Fig. 5*a*; Muraoka *et al.*, 2003). In full-length RovM the DBD would be present at the N-terminus of the protein, connected to the EBD *via* a long linker helix. The DBD and the linker helix could easily adopt a conformation very similar to that of CbnR as there are no clashes in this alignment.

Full-length CbnR is a tetramer, which was also shown in this study to be the case for full-length RovM. RovM-EBD, on the other hand, is a dimer in solution. Superposition of the RovM-EBD *BB*′ dimer with the CbnR tetramer shows that the RovM-EBD dimer is arranged in the same way as the dimers of the EBD of CbnR (Fig. 5*b*). The linker helices and DBDs could be oriented in the same fashion as in CbnR, allowing DNA binding at two distant sites and DNA bending, which has been shown for RovM (Heroven & Dersch, 2006).

In CbnR the interaction between the two dimers is mainly mediated by helix V of one subunit with the same helix of another subunit that is related by a twofold axis (Muraoka *et al.*, 2003). This kind of interaction, called a tetramer interface, has also been detected in other full-length LTTR structures, namely BenM (Ruangprasert *et al.*, 2010) and DntR (Smirnova *et al.*, 2004). A similar interaction is also present in the

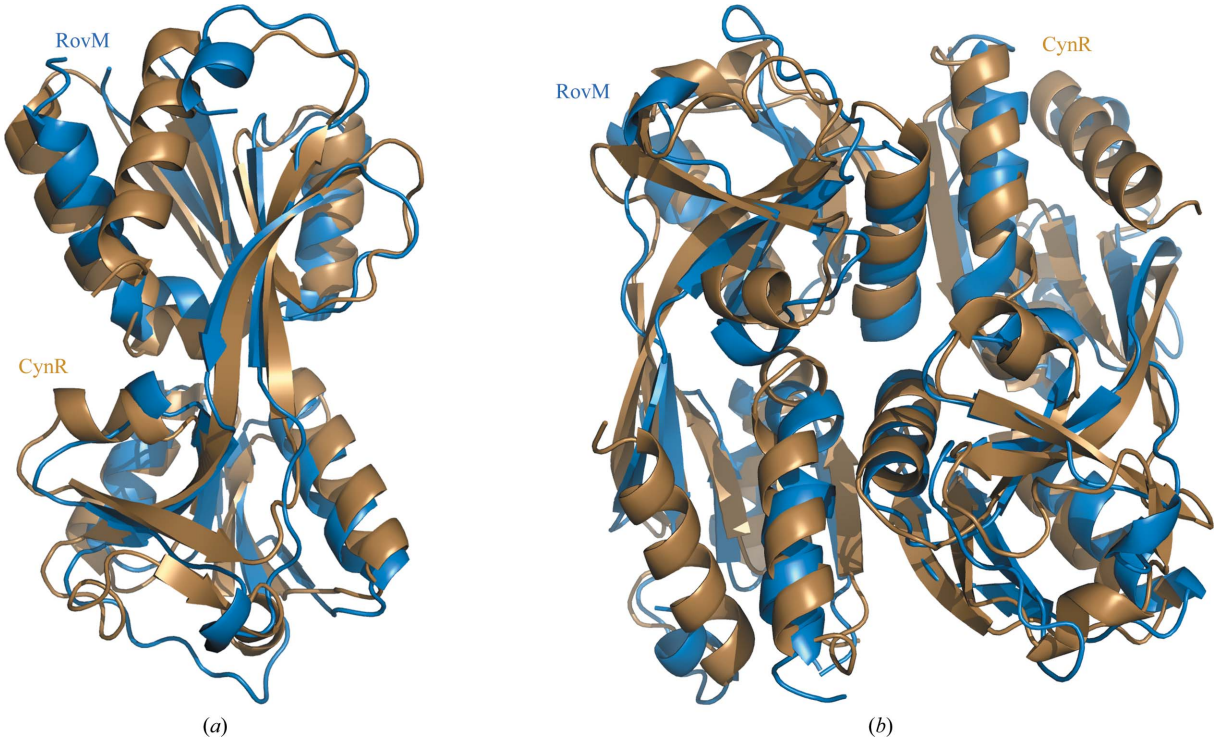

Figure 3
Superposition of RovM-EBD with CynR. RovM-EBD is shown in blue and CynR is shown in gold. (*a*) Monomers; (*b*) dimers.

RovM-EBD structure and is called the *AA*′′ interface. Here, the interactions are mostly mediated by helix 3 (which corresponds to helix V in CbnR) and helix 2. However, this interface varies considerably between the different structures. This is easily visualized by the angle between the two symmetry-related helices 3 and 3′. In the CbnR structure the two helices are approximately orthogonal to each other (86°), whereas the angle is 129° in DntR and 152° in BenM (Ruangprasert *et al.*, 2010). In contrast, in the RovM-EBD structure these two helices are nearly parallel, with an angle of 172°. Thus, the two EBDs related by the tetramer interface are roughly parallel to each other in CbnR, whereas they are nearly orthogonal in RovM-EBD (Fig. 5*c*). The structure in which the angle is closest to that in RovM-EBD is that of BenM, in which the EBDs are also nearly parallel. However, in comparison with RovM-EBD the EBDs in BenM are also shifted relative to each other (Fig. 5*d*). This shows that even though several LTTRs use the same interface, the resulting oligomers may look very different.

To determine whether it would be possible for RovM-EBD to form a tetramer similar to that of CbnR (two pairs of EBDs connected *via* the tetramer interface between one EBD from each pair), four *B* subunits were picked accordingly (Fig. 5*e*). In such an orientation the N-termini of two subunits were positioned very close to each other so that the linker helices in a full-length protein would probably clash. It thus seems unlikely that the full-length RovM tetramer would form in this manner. In full-length RovM the strong interaction between the linker helices would probably dominate over the weak tetramer interaction and there might be no interaction between the EBD pairs as seen in PDB entry 2esn (V. V. Lunin, C. Chang, T. Skarina, E. Gorodischenskaya, A.

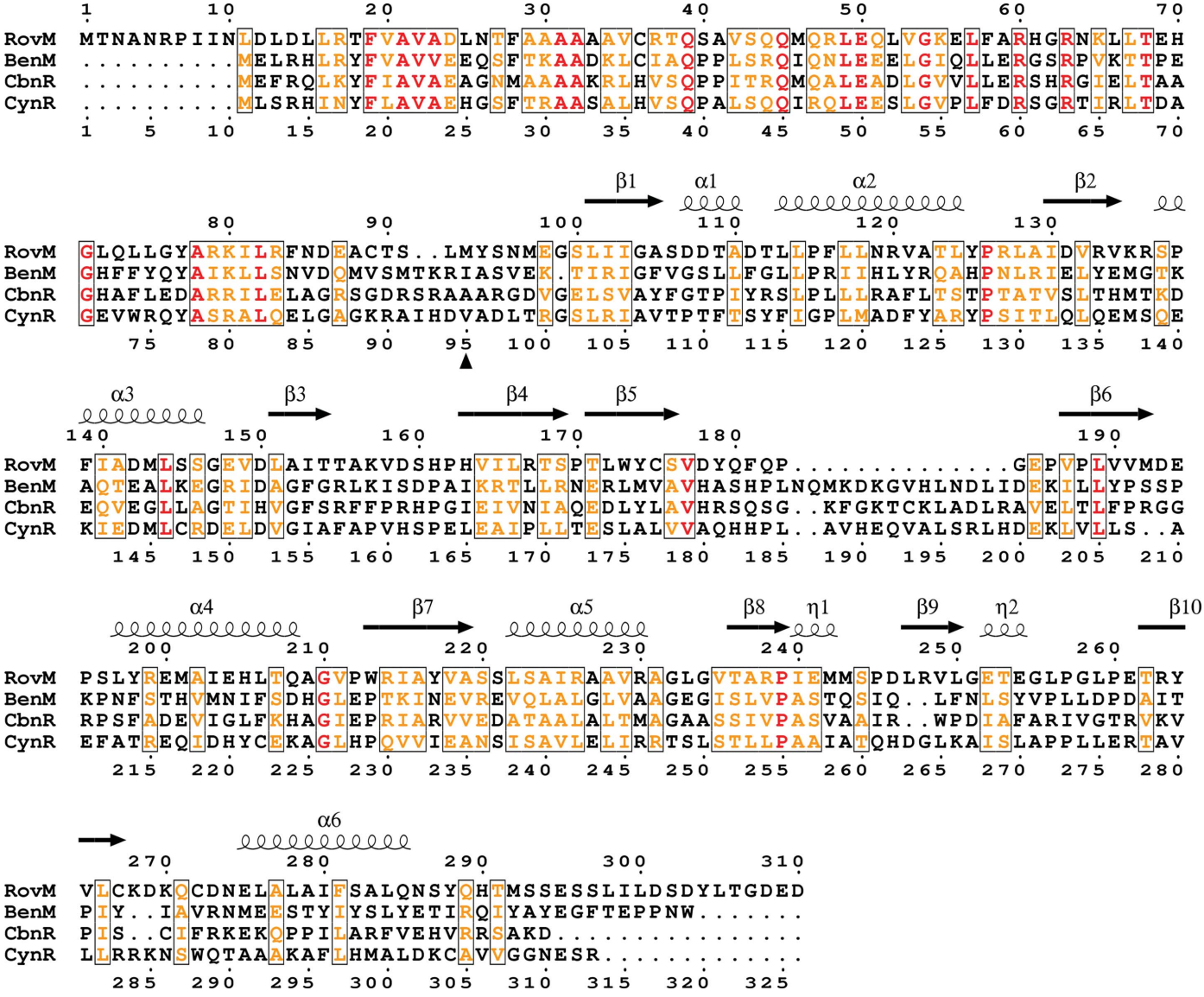

Figure 4
Structure-based sequence alignment of RovM with other LTTR proteins. The sequences of RovM (*Y. pseudotuberculosis*), BenM (*A. baylyi*; PDB code 2f7a); CbnR (*R. eutropha*; PDB code 1ixc) and CynR (*E. coli*; PDB code 3hfu) were aligned with *ClustalW* (Larkin *et al.*, 2007) and displayed with *ESPript* v.2.2 (Gouet *et al.*, 1999).

Joachimiak, A. Edwards & A. Savchenko, unpublished work). On the other hand, the strong EBD interface in the asymmetric unit of RovM-EBD might also play a role in forming the tetramer or in the cooperative binding of several RovM tetramers.

4.2. Analysis of inducer-binding sites

For several LTTR proteins inducers have been described that influence the gene regulation of the protein. For some of these proteins such as BenM it was possible to solve the structure of the complex with the inducer (Ezezika *et al.*, 2007). The inducer molecule (in this case *cis*,*cis*-muconate) was shown to bind in the cleft between EBD-I and EBD-II and binding resulted in a tilting of the two domains towards each other. This movement is thought to be relayed to the DBDs *via* the long linker helices and thus to influence DNA binding. BenM was also shown to contain a secondary ligand-binding site in EBD-I between helix 1 and sheet 4 in which a benzoate molecule was bound (Ezezika *et al.*, 2007). Fig. 6(*a*) shows the positions of both ligands after alignment with the RovM structure. Inspection of the primary binding site shows

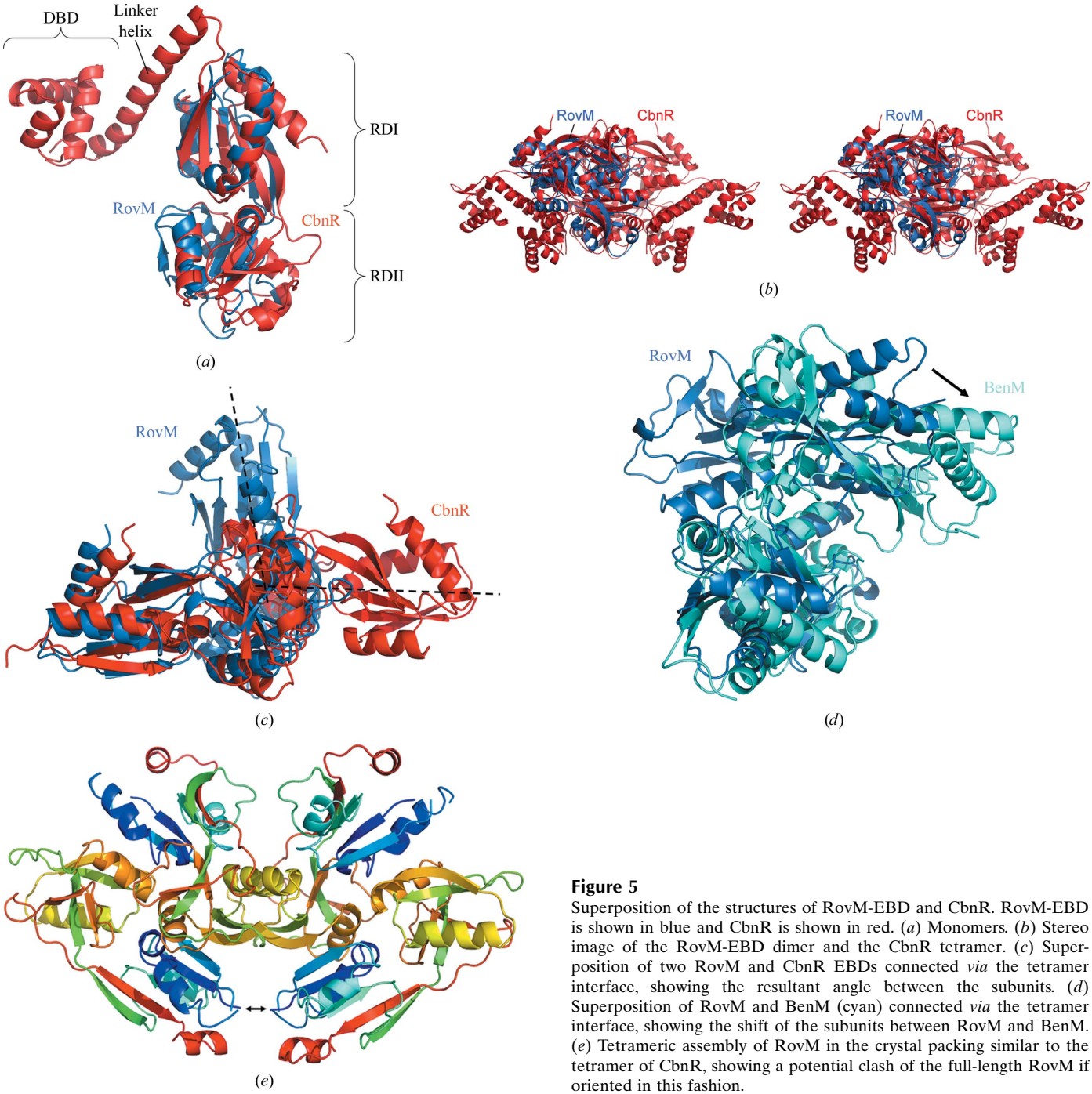

Figure 5
Superposition of the structures of RovM-EBD and CbnR. RovM-EBD is shown in blue and CbnR is shown in red. (*a*) Monomers. (*b*) Stereo image of the RovM-EBD dimer and the CbnR tetramer. (*c*) Superposition of two RovM and CbnR EBDs connected *via* the tetramer interface, showing the resultant angle between the subunits. (*d*) Superposition of RovM and BenM (cyan) connected *via* the tetramer interface, showing the shift of the subunits between RovM and BenM. (*e*) Tetrameric assembly of RovM in the crystal packing similar to the tetramer of CbnR, showing a potential clash of the full-length RovM if oriented in this fashion.

that there is a cavity in RovM which is lined with a few hydrophobic and several hydrophilic residues, some of which are similar in BenM (Fig. 6*b*). However, the cavity is much shallower than that in BenM, indicating that RovM might bind a smaller ligand than BenM.

A cavity for the secondary binding site is not visible for RovM. The ligand clashes with some side chains such as Leu113 and Leu166 and even the backbone of helix 1 (Fig. 6*c*). However, the residues that form the secondary binding site in BenM are very similar to those in RovM. Thus, after conformational changes in RovM a secondary binding site similar to that observed in BenM might become accessible. Future work will concentrate on finding inducer molecules for RovM and analysing their role in the regulation of invasin and other *Yersinia* virulence factors.

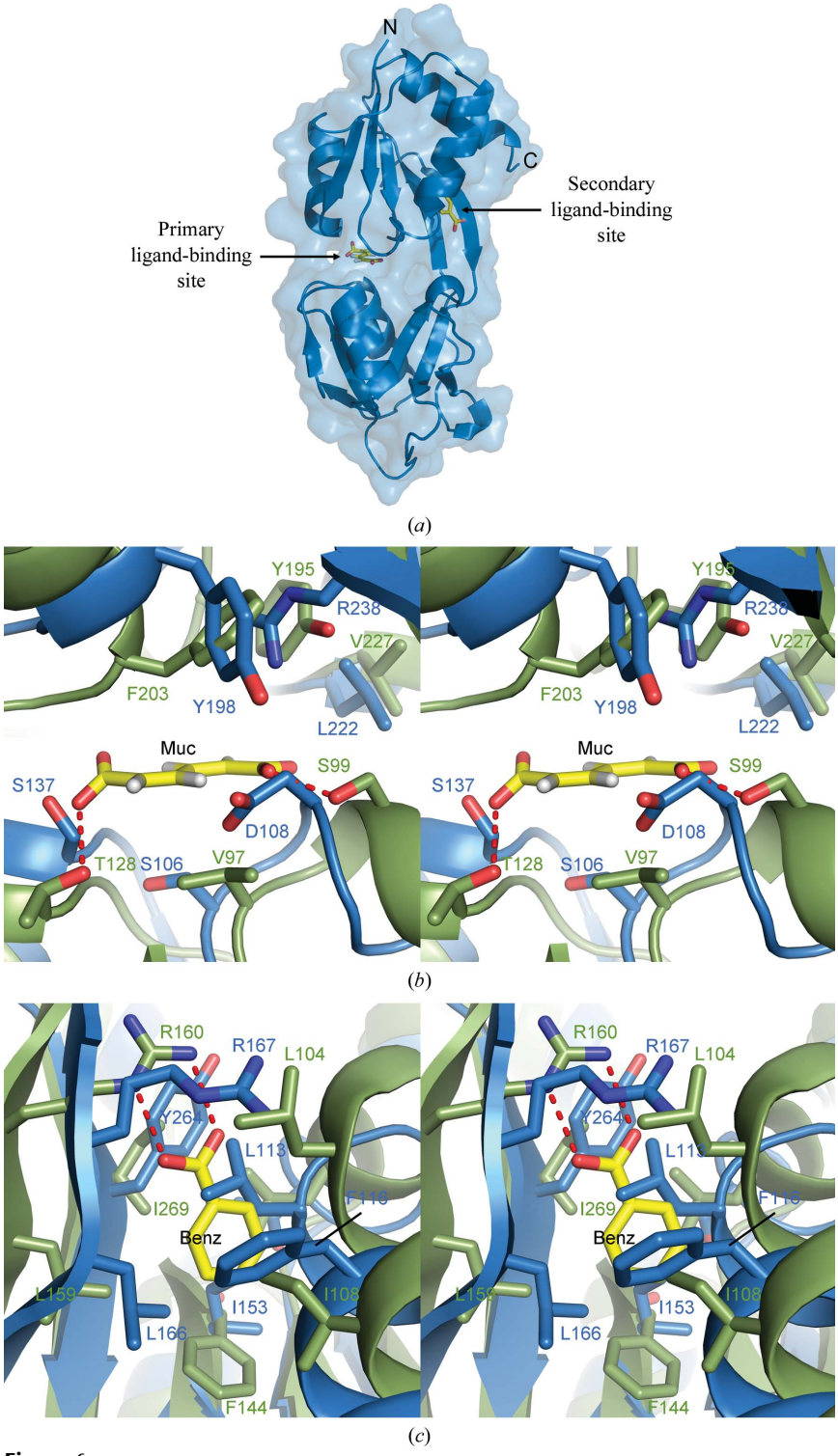

Figure 6
Comparing BenM (green) ligand-binding sites with the structure of RovM-Rd (blue). The BenM ligands *cis*,*cis*-muconate and benzoate are both shown as yellow sticks. (*a*) Cartoon and surface display of RovM-EBD. (*b*) Stereoview: comparison of primary binding-site residues between RovM-EBD and BenM. (*c*) Stereoview: comparison of secondary binding-site residues between RovM-EBD and BenM.

4.3. Comparison between RovM from pathogenic *Yersinia* species

Y. pseudotuberculosis and *Y. enterocolitica* are related pathogens which are both taken up *via* the faecal–oral route and cause similar diseases. They both use the same regulatory system involving RovA and RovM to regulate the expression of virulence factors such as invasin. Even *Y. pestis*, a vector-borne pathogen which enters its host *via* a completely different route and causes different diseases, relies on this system for virulence-factor regulation. The sequence of RovM is fully conserved between *Y. pseudotuberculosis* and *Y. pestis*, while there are some differences between *Y. pseudotuberculosis* and *Y. enterocolitica* (Supplementary Fig. 1[1]). Barely any mutations are found in the DBD, which shows a conserved DNA-binding mechanism, and the first 45 amino acids of the EBD. Most of the mutations can be found at the end of helix 7 and the following part, which is probably unstructured and was not detectable in the electron density owing to its flexibility. Mapping the changed residues onto the structure of RovM demonstrates that most of the mutations can be found on the surface of the protein (Supplementary Fig. 2[1]). However, two mutations (L197P and Y198F) are located within the potential inducer-binding site between EBD-I and EBD-II. So far, it is unclear what the inducer of RovM is or whether there is one at all. However, since RovM has been shown to mediate invasin expression in response to

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: BE5161). Services for accessing this material are described at the back of the journal.

nutrient availability, it is conceivable that some small molecule connected to metabolism may act as an inducer for RovM. Whether the mentioned mutations in the ligand-binding pocket affect the binding selectivity or affinity of the putative ligand is the subject of further investigation. RovA and RovM genes have also been found in virtually all other *Yersinia* species, with the most distant RovM from *Y. intermedia* sharing 88% sequence identity. However, these other species are nonpathogenic in humans and have no invasin gene. It is probable that the whole RovA/RovM regulatory system originally performed other functions and has evolved in pathogenic species to regulate virulence genes as well.

In conclusion, we present the structure of RovM-EBD from *Y. pseudotuberculosis* as well as an analysis of the oligomeric states of RovM-EBD and full-length RovM. RovM-EBD exists as a dimer, while full-length RovM probably assembles into a tetrameric dimer of dimers like its homologue CbnR. Additionally, a cavity was detected in RovM-EBD which could serve as a binding site for a small inducer molecule.

We thank Dr Björn Klink for help with the SAD data collection and analysis. We also thank Dr Jörn Krausse, Dr Anja Menzel and Sayantan Saha for proofreading the manuscript and for discussions and Dr Joachim Reichelt for help with bioinformatics.

References

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *Eur. J. Biochem.* **80**, 319–324.

Bond, C. S. (2003). *Bioinformatics*, **19**, 311–312.

Burnside, K., Lembo, A., de Los Reyes, M., Iliuk, A., BinhTran, N.-T., Connelly, J. E., Lin, W.-J., Schmidt, B. Z., Richardson, A. R., Fang, F. C., Tao, W. A. & Rajagopal, L. (2010). *PLoS ONE*, **5**, e11071.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Clark, T. J., Momany, C. & Neidle, E. L. (2002). *Microbiology*, **148**, 1213–1223.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.

Cowtan, K. (2010). *Acta Cryst.* D**66**, 470–478.

Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B. III, Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). *Nucleic Acids Res.* **35**, W375–W383.

Deghmane, A. E., Giorgini, D., Maigre, L. & Taha, M. K. (2004). *Mol. Microbiol.* **53**, 917–927.

Ellison, D. W., Lawrenz, M. B. & Miller, V. L. (2004). *Trends Microbiol.* **12**, 296–300.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

Ezezika, O. C., Haddad, S., Clark, T. J., Neidle, E. L. & Momany, C. (2007). *J. Mol. Biol.* **367**, 616–629.

Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.

Herbst, K., Bujara, M., Heroven, A. K., Opitz, W., Weichert, M., Zimmermann, A. & Dersch, P. (2009). *PLoS Pathog.* **5**, e1000435.

Heroven, A. K., Bohme, K., Rohde, M. & Dersch, P. (2008). *Mol. Microbiol.* **68**, 1179–1195.

Heroven, A. K. & Dersch, P. (2006). *Mol. Microbiol.* **62**, 1469–1483.

Heroven, A. K. & Dersch, P. (2010). *Res. Adv. Mol. Microbiol.* **1**, 1–17.

Heroven, A. K., Nagel, G., Tran, H. J., Parr, S. & Dersch, P. (2004). *Mol. Microbiol.* **53**, 871–888.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Kendall, M. M., Rasko, D. A. & Sperandio, V. (2010). *Mol. Microbiol.* **76**, 1306–1321.

Knapp, G. S. & Hu, J. C. (2009). *Protein Sci.* **18**, 2307–2315.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.

Lehnen, D., Blumer, C., Polen, T., Wackwitz, B., Wendisch, V. F. & Unden, G. (2002). *Mol. Microbiol.* **45**, 521–532.

Maddocks, S. E. & Oyston, P. C. (2008). *Microbiology*, **154**, 3609–3623.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 499–501.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Monferrer, D., Tralau, T., Kertesz, M. A., Dix, I., Sola, M. & Usón, I. (2010). *Mol. Microbiol.* **75**, 1199–1214.

Mukherjee, A., Cui, Y., Ma, W., Liu, Y. & Chatterjee, A. K. (2000). *Environ. Microbiol.* **2**, 203–215.

Muraoka, S., Okumura, R., Ogawa, N., Nonaka, T., Miyashita, K. & Senda, T. (2003). *J. Mol. Biol.* **328**, 555–566.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

O'Grady, E. P., Nguyen, D. T., Weisskopf, L., Eberl, L. & Sokol, P. A. (2010). *J. Bacteriol.*, doi:10.1128/JB.00852-10.

Painter, J. & Merritt, E. A. (2006*a*). *Acta Cryst.* D**62**, 439–450.

Painter, J. & Merritt, E. A. (2006*b*). *J. Appl. Cryst.* **39**, 109–111.

Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–438.

Ruangprasert, A., Craven, S. H., Neidle, E. L. & Momany, C. (2010). *J. Mol. Biol.* **404**, 568–586.

Sainsbury, S., Lane, L. A., Ren, J., Gilbert, R. J., Saunders, N. J., Robinson, C. V., Stuart, D. I. & Owens, R. J. (2009). *Nucleic Acids Res.* **37**, 4545–4558.

Salyers, A. A. & Whitt, D. D. (2002). *Bacterial Pathogenesis: A Molecular Approach*, 2nd ed., pp. 213–228. Washington DC: ASM Press.

Schell, M. A. (1993). *Annu. Rev. Microbiol.* **47**, 597–626.

Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Smirnova, I. A., Dian, C., Leonard, G. A., McSweeney, S., Birse, D. & Brzezinski, P. (2004). *J. Mol. Biol.* **340**, 405–418.

Stec, E., Witkowska-Zimny, M., Hryniewicz, M. M., Neumann, P., Wilkinson, A. J., Brzozowski, A. M., Verma, C. S., Zaim, J., Wysocki, S. & Bujacz, G. D. (2006). *J. Mol. Biol.* **364**, 309–322.

Surgey, N., Robert-Baudouy, J. & Condemine, G. (1996). *J. Bacteriol.* **178**, 1593–1599.

Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.

Wu, R.-Y., Zhang, R.-G., Zagnitko, O., Dementieva, I., Maltzev, N., Watson, J. D., Laskowski, R., Gornicki, P. & Joachimiak, A. (2003). *J. Biol. Chem.* **278**, 20240–20244.

Zhou, X., Lou, Z., Fu, S., Yang, A., Shen, H., Li, Z., Feng, Y., Bartlam, M., Wang, H. & Rao, Z. (2010). *J. Mol. Biol.* **396**, 1012–1024.