Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structural analysis of *Bordetella pertussis* BugE solute receptor in a bound conformation

Isabelle Huvent,[a,b]‡ Hassan Belrhali,[c]‡ Rudy Antoine,[b,d,e] Coralie Bompard,[a,b] Camille Locht,[b,d,e] Françoise Jacob-Dubuisson[b,d,e]* and Vincent Villeret[a,b]*

[a]UMR8161 CNRS, Institut de Biologie de Lille, Institute Pasteur de Lille, 1 rue du Professeur Calmette, BP245, 59021 Lille CEDEX, France, [b]IFR142, Institute Pasteur de Lille, 1 rue du Professeur Calmette, 59021 Lille CEDEX, France, [c]EMBL-Grenoble Outstation, 6 rue Jules Horowitz, BP181, 38042 Grenoble CEDEX 9, France, [d]INSERM-U629, Lille, France, and [e]Institute Pasteur de Lille, 1 rue du Professeur Calmette, BP245, 59021 Lille CEDEX, France

‡ These authors made equal contributions.

Correspondence e-mail: francoise.jacob@ibl.fr, vincent.villeret@ibl.fr

Received 27 April 2006
Accepted 16 August 2006

PDB Reference: BugE, 2dvz, r2dvzsf.

The Bug proteins form a large family of periplasmic solute-binding receptors present in a number of bacterial species. Here, the crystal structure of *Bordetella pertussis* BugE, a member of the Bug family coded by the gene BP0250, is reported. It adopts the Venus flytrap architecture of periplasmic binding proteins, with two domains separated by a deep cleft. BugE has a bound ligand, identified as a glutamate. The structure of *B. pertussis* BugD, which is an aspartic acid transporter, has recently been reported. These structures reveal high conservation of the Bug architecture, despite limited sequence identity. They share a common carboxylate-binding motif defined by two strand–β-turn–α-helix motifs, also involving two water molecules to bridge the carboxylate O atoms to the protein. The two water molecules are hydrogen bonded to a common main-chain carbonyl group. Although the features of the carboxylate-binding motif are totally conserved, the ligand in BugE is bound by its side-chain carboxylate group rather than by its α-carboxylate as in BugD. This specific ligand-binding motif is highly conserved in Bug proteins and the BugE structure suggests that the cavity of the Bug proteins might also accommodate carboxylated solutes other than amino acids. The vast expansion of the Bug family in several bacterial genera is likely to be explained by the possible diversity of ligands. No charged residues are involved in glutamate binding by BugE, unlike what has been described for all glutamate receptors reported so far.

1. Introduction

Bacterial genome sequences contain large families of paralogous genes that encode proteins of related, although not identical, function (Saier & Paulsen, 1999). In *Bordetella pertussis*, the whooping cough agent, a large family of periplasmic receptors has been identified: the Bugs (*Bordetella* uptake genes; Antoine *et al.*, 2003). *In silico* analyses of the available bacterial genomic sequences have indicated that several hundred *bug* homologues exist in the prokaryotic world, although very few have been characterized so far (Antoine *et al.*, 2003; unpublished data).

Periplasmic receptors in Gram-negative bacteria specifically bind solutes such as sugars, organic and inorganic ions, iron-chelator complexes and amino acids and deliver them to membrane-embedded uptake transporters (Tam & Saier, 1993). Thus, some Bug proteins are part of tripartite tricarboxylate-transport (TTT) systems in various bacterial species (Antoine *et al.*, 2003, 2005; Winnen *et al.*, 2003). TTT systems consist of a Bug-like receptor and two membrane proteins, one well conserved with 12 putative transmembrane α-helical spanners and the other poorly conserved with four putative

TMSs. Solute transport by TTT systems is most likely to be powered by free energy stored in electrochemical gradients across the cytoplasmic membrane. In several bacteria, including *B. pertussis*, *B. parapertussis*, *B. bronchiseptica* and *Ralstonia metallidurans*, the *bug* genes largely exceed the genes coding for predicted membrane components of TTT transporters and thus the vast majority of Bug proteins are 'orphans' without identified membrane partners (Antoine *et al.*, 2003). However, at least in *B. pertussis*, several orphan Bugs are very abundant periplasmic proteins, indicating that they are likely to perform important functions.

The first structure of a Bug protein, that of BugD, has recently been reported (Huvent *et al.*, 2006). The structure showed BugD to be an aspartate receptor. Here, we present the crystal structure of a second member of the Bug family, coded by BP0250. The protein from *B. pertussis* was crystallized with a bound fortuitous ligand that we have identified as a glutamate from the electron density and hydrogen-bonding pattern. Thus, the protein has been renamed BugE in order to refer to its ligand using the one-letter code. BugE, one of the most abundant periplasmic proteins of *B. pertussis*, is likely to participates in the uptake of glutamate, an important carbon source for this organism (Thalen *et al.*, 1999).

2. Materials and methods

2.1. Cloning, protein expression and purification

The BP0250 gene of *B. pertussis* was amplified by the polymerase chain reaction using genomic DNA from the Tohama I derivative BPSM as a template. The oligonucleotide BP0250-UP (upper) 5′-ATAT**GGATCC**GCCGATGCCTATCCCAGCAA-3′ with a *Bam*HI site (in bold) was used as the 5′-primer and the oligonucleotide BP0250-LO (lower) 5′-TATA**AAGCTT**CGGTGTCAGGCAGCCATCAG-3′ with a *Hin*dIII site (in bold) was used as the 3′-primer. The primers were designed to amplify the sequence encoding the mature part of the protein, *i.e.* without its signal peptide. The amplicon was directly inserted into the plasmid pCR II-TOPO (Invitrogen) and sequenced. The amplified gene was then introduced as a *Bam*HI–*Hin*dIII fragment into the corresponding sites of pQE-30 (Qiagen), yielding pIH01. This plasmid codes for BP0250 with an N-terminal six-histidine tag. Expression was conducted in *Escherichia coli* SG13009 (pREP4, pIH01), with pREP4 encoding the *lac* repressor *in trans* (Qiagen), thus ensuring tightly regulated expression. An overnight culture of freshly transformed cells in 20 ml LB medium supplemented with 100 µg ml^{-1} ampicillin and 25 µg ml^{-1} kanamycin was used to inoculate 2 l of the same medium. The culture was incubated at 310 K under rotatory shaking (200 rev min^{-1}). Overexpression of BP0250 was induced at an optical density at 600 nm (OD_{600}) of 0.6 by adding 1 m*M* IPTG. After 3 h of incubation, the cells were harvested by centrifugation at 6000*g* for 10 min at 277 K. The selenomethionyl (SeMet) protein was produced according to the method described by Doublié (1997).

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the outer resolution shell.

Crystal data	
Unit-cell parameters (Å)	
a	53.6
b	53.6
c	421.8
Space group	$P6_522$
Data collection	
Wavelength (Å)	0.9758
Resolution (Å)	15–2.30 (2.36–2.30)
Measured reflections	142647
Unique reflections	15977
Completeness (%)	96.6 (72.0)
$I/\sigma(I)$	6.0 (5.3)
R_{sym}† (%)	7.9 (12.5)
Redundancy	9.0 (8.7)
Refinement	
Resolution (Å)	15–2.30 (2.36–2.30)
R (%)	17.6 (19.4)
R_{free}‡ (%)	24.9 (36.4)
Protein atoms	2248
Water molecules	223
R.m.s. deviations	
Bonds (Å)	0.012
Angles (°)	1.3
Average *B* factors (Å^2)	
Main-chain atoms	18.7
Side-chain atoms	20.4
Ligand atoms	13.4
Cd^{2+} ions	23.9
Water molecules	35.8
Ramachandran plot (%)	
Favoured	91.9
Allowed	7.7
Generous	0.4
Disallowed	0

† $R_{sym} = \sum|I_i - I_{avg}|/\sum I_i$, where I_i is the observed intensity and I_{avg} is the average intensity. ‡ R_{free} is calculated for 5% of randomly selected reflections excluded from refinement.

All purification steps were carried out at 277 K. Typically, the cell pellets from a 2 l culture were resuspended in 50 ml native binding buffer (50 m*M* Tris–HCl pH 8.0, 250 m*M* NaCl, 50 m*M* imidazole and a tablet of Complete EDTA-free protease-inhibitor cocktail; Roche) containing 5 µg ml^{-1} DNase I. Cells were lysed by two passages through a French Press and the cell debris was removed by centrifugation for 20 min at 31 000*g*. The supernatant was loaded onto a nickel-Sepharose affinity column (Chelating Sepharose Fast Flow, Amersham Biosciences) pre-equilibrated in native binding buffer. The bound protein was eluted using a linear gradient of 70–200 m*M* imidazole in 50 m*M* Tris–HCl pH 8.0, 250 m*M* NaCl. For the SeMet protein, 0.2 m*M* EDTA and 10 m*M* DTT were added to the collection vessel to prevent oxidation of the SeMet residues. The fractions containing the recombinant protein were pooled and dialysed against 50 m*M* Tris–HCl pH 8.0, 1.4 *M* $(NH_4)_2SO_4$, which resulted in the precipitation of several contaminating proteins. The sample was then clarified by ultracentrifugation at 100 000*g* for 1 h. Next, an extensive dialysis of the supernatant was performed against 10 m*M* Tris–HCl pH 8.0, 250 m*M* NaCl. In the case of the selenomethionyl protein, dialysis steps were performed with buffers containing 0.2 m*M* EDTA and 10 m*M* DTT. The protein was then

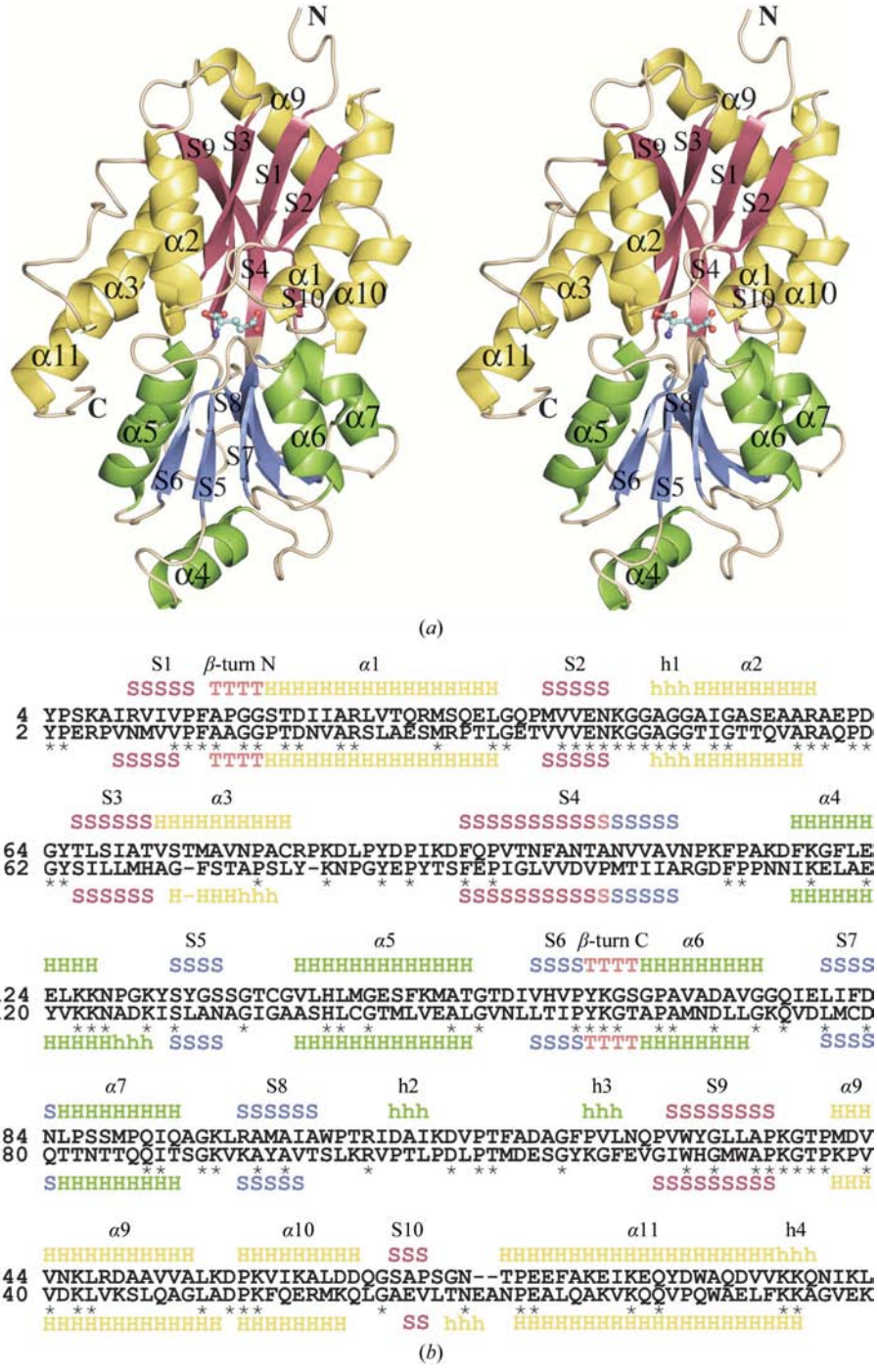

Figure 1
Structure and sequence of *B. pertussis* BugE. (*a*) Stereoview of the overall BugE structure. The protein is illustrated as a ribbon and the bound glutamate is shown in stick representation between the two domains. The α-helices of domains 1 and 2 are shown in yellow and green, respectively, and the β-strands of domains 1 and 2 are in red and blue, respectively. (*b*) Definition of the secondary structures of BugE (upper sequence). The aligned sequence of BugD is also shown (lower sequence). Asterisks show residues that are identical in the two sequences. The secondary-structure elements are indicated with the same colour code as in (*a*). Small hs indicate 3_{10}-helices. Secondary structures were assigned with *PROMOTIF* (Hutchinson & Thornton, 1996).

concentrated to 22 mg ml^{-1} using a centrifugal filter device (Centriplus YM-10, Millipore) and the protein concentration was determined by the Bradford method. Full incorporation of selenomethionine into the nine methionine (including the N-terminal methionine) sites was confirmed by electrospray mass spectrometry.

2.2. Crystallization and data collection

Crystals of the native and selenomethionyl protein were obtained at 277 K using the hanging-drop vapour-diffusion method. A 2 µl aliquot of protein solution (22 mg ml^{-1} protein) was mixed with an equal volume of a solution containing 0.1 *M* $CdCl_2$, 0.1 *M* sodium acetate pH 4.3, 30–32%(*v*/*v*) PEG 400 or 34–40%(*v*/*v*) PEG 200. For data collection, these crystals were flash-frozen in a 100 K dry nitrogen stream. With more than 30% PEG 200 or PEG 400, the crystallization solution functioned as a cryoprotectant and was used without modification.

Native crystals were initially characterized on a rotating-anode generator equipped with a MAR 345 imaging-plate detector. These crystals diffracted to at least 2.7 Å resolution, but presented a very long *c* axis that could not be resolved on the MAR 345. All subsequent data collections were thus pursued on a synchrotron beamline.

A single SeMet-labelled crystal was used to collect a diffraction data set at beamline BM14 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. The wavelength of 0.9758 Å was selected on the basis of an X-ray fluorescence scan of the frozen crystal in order to maximize the signal from the anomalous Se scatterers. A single data set was collected at the given wavelength to a maximum resolution of 2.0 Å using a MAR Mosaic 225 mm CCD detector. The diffraction intensities were indexed and integrated with *MOSFLM* (Leslie, 1992), scaled with *SCALA* and converted into amplitudes with the program *TRUNCATE* from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). The crystal belongs to space group $P6_522$, with unit-cell parameters $a = b = 53.6$, $c = 421.8$ Å, $\alpha = \beta = 90$, $\gamma = 120^\circ$. The crystal contains one monomer in the asymmetric unit, which corresponds to a solvent content of 52% and specific volume V_M of 2.57 Å^3 Da^{-1}. The effective resolution is 2.3 Å as esti-

mated with the program *DATAMAN* (Kleywegt & Jones, 1996). Data-collection statistics are summarized in Table 1.

2.3. Structure determination and refinement

The structure was solved by the SAD method. The position of the eight expected Se atoms of the monomer, not including the N-terminal methionine, were determined with the program *SHELXD* (Schneider & Sheldrick, 2002) using standard parameters and the input files generated by the program *XPREP* (Bruker AXS). Refinement of the selenium sites, phase calculation and density modification were carried out with the program *SHARP* (de La Fortelle & Bricogne, 1997). Model building was performed with the program *ARP/wARP* (Perrakis *et al.*, 1997, 1999) and was inspected using the program *Coot* (Emsley & Cowtan, 2004). The quality of the electron-density map allowed a nearly complete automatic attribution of the electron density. There is a *cis*-proline in position 5 of the BugE sequence. Residues 3–5 were not built automatically with *ARP/wARP* and were thus included manually after building the *cis*-proline. The last 11 residues (from Gln291) are less defined and thus only the backbone has been traced. Visual inspection of the electron-density map also revealed that BugE had a bound ligand, subsequently identified as a glutamate residue, which fits unambiguously into the electron-density map. Five cadmium ions were also identified in the electron density. The refinement and addition of solvent molecules were performed with *REFMAC*5 using standard parameters. A random set of 5% reflections was excluded from refinement for cross-validation of the refinement strategies, the insertion of solvent water and as a basis for a maximum-likelihood refinement with *REFMAC*5 (Murshudov *et al.*, 1999) and to monitor $R_{\rm free}$. Water molecules were assigned automatically for peaks >3σ in $F_{\rm o} - F_{\rm c}$ difference maps by cycling the *REFMAC*5 refinement with *ARP/wARP* and were retained if they obeyed hydrogen bonding. The ligand present in BugE was identified from $F_{\rm o} - F_{\rm c}$ and $2F_{\rm o} - F_{\rm c}$ electron-density maps. As the ligand is an amino acid, parameters defining it were already present in *REFMAC*5 dictionaries and were used without modification. The final $R_{\rm work}$ is 17.6% with an $R_{\rm free}$ of 24.9% using all data to 2.3 Å resolution. The final model has an excellent stereochemistry as indicated by *PROCHECK* (Laskowski *et al.*, 1993) (Table 1).

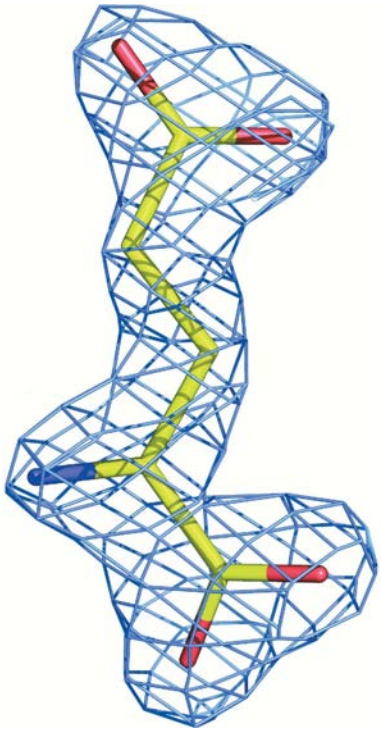

Figure 2
Electron-density map of the bound glutamate in the crystal structure. The $2F_{\rm o} - F_{\rm c}$ map was calculated by omitting the ligand coordinates and is contoured at the 1.5σ level.

The final model consists of one BugE molecule, a bound glutamate, five cadmium ions and 223 water molecules. Refinement statistics are summarized in Table 1.

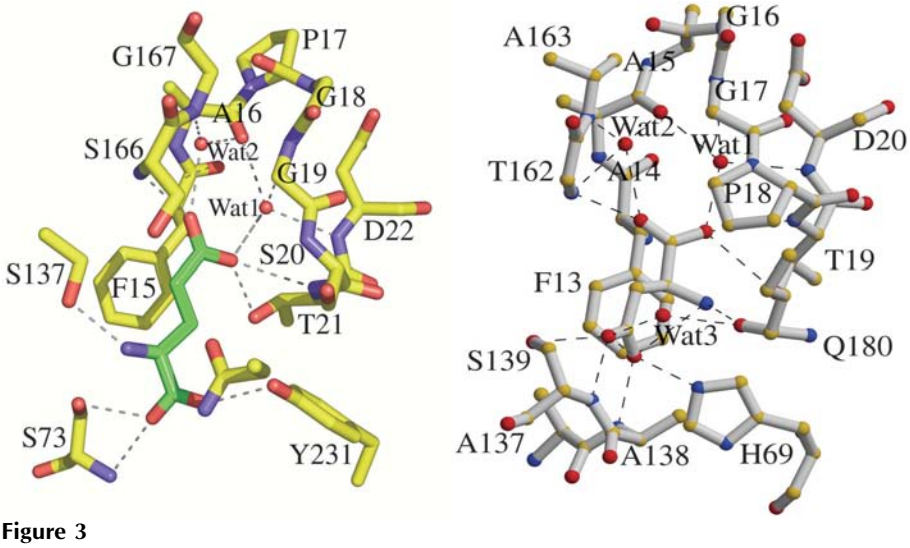

Figure 3
Comparison of the ligand-binding sites in BugE and BugD. (*a*) View of the ligand-binding site in BugE, showing interactions between the ligand and the protein. For clarity, the C atoms of the glutamate are shown in green. (*b*) View of the ligand-binding site in BugD, as described in Huvent *et al.* (2006).

3. Results and discussion

3.1. Structural analysis

The BugE structure was solved by the SAD method using a selenomethionyl-substituted recombinant protein crystal. The presence of cadmium ions in the crystallization solution was essential in order to obtain crystals. Five Cd^{2+} ions are clearly identified in the structure and are coordinated by aspartate or glutamate residues and water molecules. There is one monomer in the asymmetric unit and the coordination of three Cd^{2+} ions involves acidic residues

that belong to two different monomers, thus bridging different monomers in the crystal packing and explaining why cadmium ions are required for crystallization. The role of cadmium ions in protein crystallization has already been documented in numerous studies and BugE is another example where cadmium ions bridge neighbouring symmetry-related molecules within the crystal.

BugE is a monomeric protein of 302 amino-acid residues. Residues 3–302 have been included in the 2.0 Å structure. Only residue 1 (and the histidine tag) is not visible in the electron density. The structure is illustrated in Fig. 1. Overall, BugE adopts the Venus flytrap architecture of periplasmic binding proteins, with two domains separated by a deep cleft, the floor of which is formed by a pair of β-strands, S4 and S9. Domain 1 is composed of two separate polypeptide segments comprising residues 1–103 and 229–302 from the NH_2- and COOH-termini, respectively. The intermediate region defines domain 2 and domain junction is accomplished by β-strands S4 and S8–S9. There is one disulfide bridge in the structure, between Cys81 and Cys140.

The two domains consist each of a five-stranded β-sheet, surrounded by six and four α-helices in domains 1 and 2, respectively. The topological disposition of the various strands is as follows for domains 1 and 2, respectively: S2–S1–S3–S9–S4 and S6–S5–S7–S4–S8. The structure contains four short 3_{10}-helices consisting of residues 48–50 (h1), 209–211 (h2), 223–225 (h3) and 296–298 (h4) (Fig. 1).

3.2. BugE is liganded

The initial electron-density maps revealed the presence of a ligand in the structure. BugE adopts a closed conformation resulting from the fortuitous capture of its ligand, most likely found in the culture medium, and identified as a glutamine or glutamate residue (Fig. 2). Based on the hydrogen-bonding pattern, the ligand is a glutamate. It is completely buried in a pocket formed between the two domains and makes contacts with both of them.

Ligand binding involves residues Phe15, Ala16, Thr21, Asp22, Ser73, Ser137, Ser166, Gly167 and Tyr231. Many interactions involve backbone atoms and are as follows (Figs. 3 and 4). OT1 is hydrogen bonded to the hydroxyl of Ser73 and also to its backbone amide nitrogen, while OT2 is hydrogen bonded to the hydroxyl of Tyr231 and also to a buried water molecule. The C^{α} amine group of the ligand is hydrogen bonded to the hydroxyl group of Ser137. Asn184 lies in the vicinity of the α-amino and α-carboxylate groups, with its side-chain atoms ND2 and OD1 3.5 and 3.6 Å away from OT2 and the α-amino group, respectively.

Binding of the ligand side-chain carboxylate involves two water molecules, Wat1 and Wat2 (Figs. 3 and 4). OE1 is hydrogen bonded to Wat1, to the backbone amide N atom and to the hydroxyl group of Thr21. OE2 is hydrogen bonded to Wat2 and to the backbone amide N atom of Ser166. The structure reveals the presence within the bound ligand of a carboxylate group at that position, because each putative ligand O atom is hydrogen bonded to an amide main-chain group. This is consistent only with the presence of a glutamate and not with that of a glutamine.

Wat1 and Wat2 are stabilized by additional contacts with protein atoms. Wat1 interacts with the carbonyl of Ala16 and with the backbone amides of Gly19 and Asp22. Wat2 interacts with the backbone amide of Gly167 and also with the carbonyl of Ala16. Thus, Wat1 and Wat2 are both hydrogen bonded to the carbonyl O atom of Ala16 and form additional hydrogen bonds with the backbone amide groups of residues either from

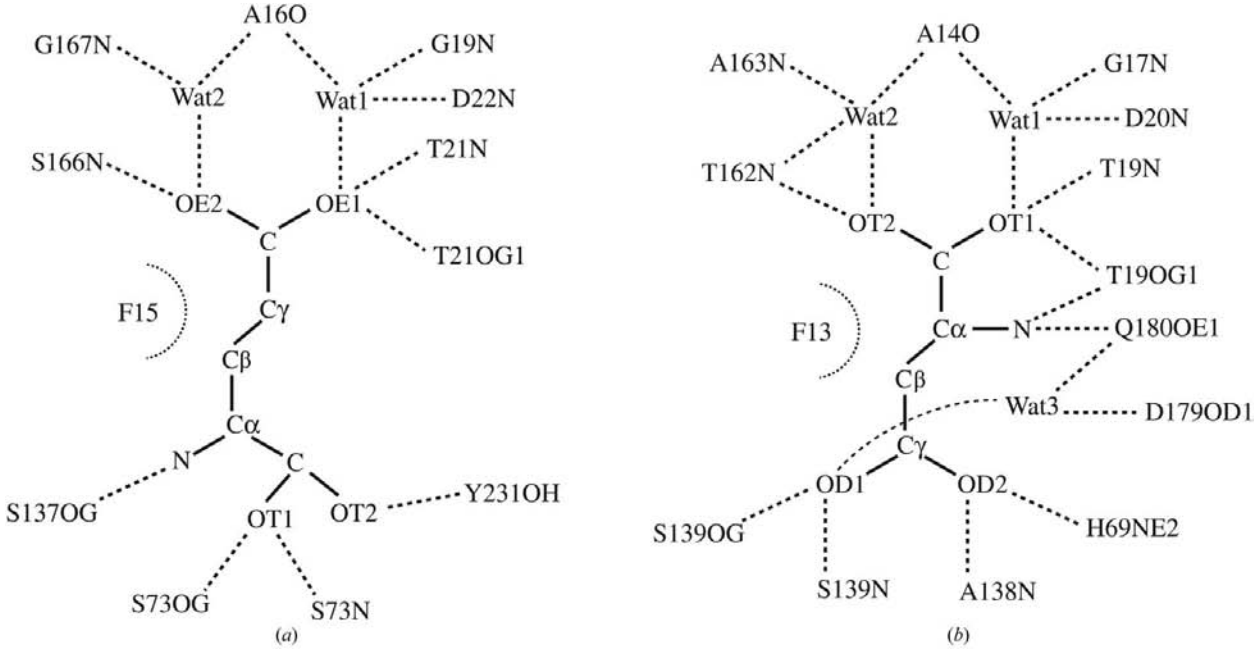

Figure 4
Illustration of the hydrogen-bonding pattern between (*a*) BugE and the glutamate and (*b*) BugD and the aspartate.

domain 1 (Wat1, Gly19, Asp22) or from domain 2 (Wat2, Gly167). The two water molecules are 3.5 Å apart.

Finally, Phe15 lies along the glutamate, which is in an extended conformation and is nearly parallel to the $C^{\alpha}-C^{\delta}$ segment of the ligand (Fig. 3).

3.3. Structural comparison with BugD

The first structure of a member of the *B. pertussis* Bug family has been reported recently: that of BugD, which is an aspartate transporter (Huvent *et al.*, 2006). The sequence identity between BugE and BugD is ∼33%. The two structures have a well conserved architecture, with an overall r.m.s. deviation of 1.4 Å for all C^{α} atoms (Fig. 5), as calculated with *LSQKAB* (Kabsch, 1976).

A conserved region between the two structures encompasses the secondary structures from S1 to helix α2. Another well conserved region is found between strand S6 and helix α6. These segments define the side-chain carboxylate-binding site of BugE. Indeed, BugD and BugE share a common carboxylate-binding motif defined by these stretches of sequence, involving two water molecules to bridge the carboxylate O atoms to the protein. Residues Ala16 and Gly19 (or Ala14 and Gly17 in BugD) are required for water binding and participate in a specific type II β-turn which is located between the first β-strand and α-helix. The second part of the carboxylate-binding motif is also made up from a strand (Ser6)–β-turn–helix (α6) structural motif.

The conservation among the Bug sequences of such a carboxylate-binding motif had been predicted from sequence analyses based on the BugD structure (Huvent *et al.*, 2006). However, there is a fundamental difference between BugD and BugE, because in BugD this structural motif is implicated in the binding of the C^{α} carboxylate group of the amino acid, while in BugE it is involved in the binding of the side-chain carboxylate group, despite the presence in both structures of Thr21 which, in BugD, binds the C^{α} amine group in addition to one carboxylate O atom. Thus, the BugE structure confirms the hypothesis that this site may be a conserved carboxylate-binding site common to a vast majority of Bugs, but clearly illustrates that it is not devoted solely to the binding of α-carboxylate groups.

The most divergent regions between the BugD and BugE structures are the helices α3 and α5 that are involved in binding either the side-chain carboxylate of BugD or the α-carboxylate and α-amino group in BugE. In BugE these secondary structures move in such a way as to enlarge the ligand-binding site, which allows space for the glutamate. This conformation may be stabilized by the presence of a disulfide bridge between Cys81 (α3) and Cys140 (helix α5). There is a limited number of sequence insertions/deletions between the two structures. Two deletions are observed close to helix α3 in BugD when compared with BugE. One is located just before the beginning of the helix, while the other is located in the loop linking helix α3 to sheet S4. These differences might also contribute to the positioning of helix α3 in BugE such that it can accommodate its ligand. The other area showing two insertions in BugD is located before the C-terminal helix α11, further away from the ligand-binding site.

3.4. Comparison with other glutamate receptors

Structures of ligand-bound glutamate receptors related to periplasmic binding proteins have been reported from both prokaryotic and eukaryotic organisms. These receptors are GluR0, a prokaryotic glutamate receptor ion channel (Mayer *et al.*, 2001), GluR2, a eukaryotic AMPA-sensitive glutamate receptor ion channel (Armstrong & Gouaux, 2000), and the GluR5 and GluR6 kainate receptors from mammals (Mayer, 2005). A receptor from *Thermus thermophilus*, TtGluBP (Takahashi *et al.*, 2004), has also been reported as a glutamate-binding protein, although its hydrogen-bonding pattern does not allow discrimination between a glutamine and a glutamate.

In glutamate-receptor ion channels, the peptide sequences that make up the two domains are interrupted by insertion of the first and second transmembrane ion-channel segments. The various glutamate receptors show overall structural homology with BugE and can be superimposed onto it with C^{α} r.m.s. deviations of 2.6 Å and ∼3.2–3.5 Å for GluR0 and all the other receptors, respectively. The ligand-binding mode of BugE is not found in any of those glutamate receptors. In the GluR0, GluR2, GluR5 and GluR6 complexes the binding of the ligand α-carboxylate and α-amino groups is quite similar. In particular, Arg residues which form hydrogen bonds and salt links with the ligand α-carboxylate O atoms are highly conserved, as well as an acidic residue (Asp or Glu) which interacts with the α-amino group. There is no charged residue in the ligand-binding site of BugE; thus, only the hydrogen-bonding pattern dissipates the charges on the carboxylate groups. In TtGluBP, there is only one glutamate which interacts with the α-amino group, while hydrogen bonds stabilize the α-carboxylate group, without any arginine. Here, also, the ligand binding is different from that of BugE.

In the TtGluBP and GluR0 complexes with glutamate, the ligands adopt comparably extended conformations. When the two structures are superimposed with *SSM* (Krissinel & Henrick, 2004), the C^{α} atoms of

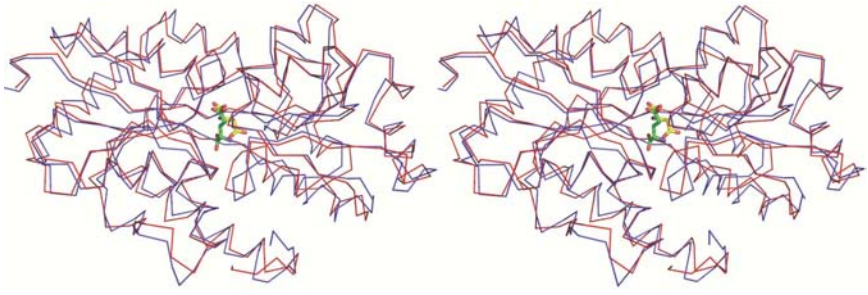

Figure 5
Stereoview showing the superimposition of the C^{α} traces of BugE (in blue) and BugD (in red). The glutamate and aspartate residues bound to BugE and BugD are shown with C atoms in green and yellow, respectively.

both glutamate ligands are only separated by 0.4 Å, which indicates comparable positions of the two domains in the receptors. An extended conformation is also observed in BugE, however, with inverted positions for the ligand α-amino and α-carboxylate groups and a slight displacement of the ligand towards domain 1, illustrated by the distance of 1.7 Å between the C^{α} atoms of the BugE and GluR0 ligands.

In contrast, in the GluR2, GluR5 and GluR6 complexes the ligand side chain has undergone a rotation such that its γ-carboxylic group projects towards and interacts with the base of the helix in domain 2 corresponding to α5 in BugE.

In all structures, binding of the side-chain carboxylate O atoms occurs without basic residues and involves hydrogen bonds with side-chain hydroxyl groups from Ser or Thr, with main-chain amide or carbonyl groups and with water-mediated hydrogen bonds. The binding motif observed in BugE as well as in BugD is unique to the Bug proteins, with two water molecules hydrogen bonded to the same carbonyl group of an alanine.

4. Conclusion

The Bug family of genes encoding extracytoplasmic solute receptors is strongly represented in several β-proteobacteria, including *B. pertussis*, and is also present in more limited numbers in many other bacterial species. The first structure of a Bug protein has been reported recently and allowed the description of the overall architecture of the Bug family and the identification of a ligand (Huvent *et al.*, 2006). The structure enabled us to characterize a structural motif for carboxylate binding, which was predicted to be conserved in the vast majority of Bug proteins. We have determined in this work the structure of a second member of the Bug family, BugE. Similarly to BugD, the protein has captured a ligand during the course of its production/purification, identified as a glutamate. This structure illustrates the high conservation of the Bug architecture, despite the limited overall sequence identity. All the β-strands and α-helices are well conserved, as well as the β-turns that have been shown to be important for carboxylate binding. The most divergent regions between BugD and BugE, helices α3 and α5, are located at the other end of the ligand-binding pocket and allow the accommodation of ligands of different sizes.

In BugD, the carboxylate-binding motif binds the α-carboxylate and α-amino groups of the amino acid, while in BugE it binds the side-chain carboxylate. This observation leads to the hypothesis that the conserved structural motif may enable the Bug family to bind and transport a variety of carboxylated ligands in addition to amino acids.

In *B. pertussis*, BugE is a very abundant periplasmic protein that is likely to participate in the uptake of glutamate, an important carbon source for this organism. *B. pertussis* cannot perform glycolysis (Parkhill *et al.*, 2003) and therefore amino acids, in particular glutamate, are its main carbon and nitrogen sources (Imaizumi *et al.*, 1983; Thalen *et al.*, 1999). It can also metabolize other carboxylated substrates such as lactate, succinate, fumarate and malate (Thalen *et al.*, 1999). The metabolism of *Bordetella* might thus be a reason for the expansion of the Bug family in this genus. The availability of two Bug structures provides us with the framework to perform further analyses of this large family.

We thank H. Drobecq for the mass-spectrometry measurements. FJD and VV are at the Centre National de la Recherche Scientifique (CNRS), France. We thank the MRC-BM14 Staff and the EMBL Staff for making the UK CRG BM14 beamline available to us. This work was supported in part by the Centre National de la Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Médicale (INSERM), the Institut Pasteur de Lille and the Région Nord-Pas de Calais through the CPER and FEDER. VV is a recipient of an Action Thématique et Incitative sur Programme Jeunes Chercheurs Grant from the CNRS.

References

Antoine, R., Huvent, I., Chemlal, K., Deray, I., Raze, D., Locht, C. & Jacob-Dubuissson, F. (2005). *J. Mol. Biol.* **351**, 799–809.
Antoine, R., Jacob-Dubuisson, F., Drobecq, H., Willery, E., Lesjean, S. & Locht, C. (2003). *J. Bacteriol.* **185**, 1470–1474.
Armstrong, N. & Gouaux, E. (2000). *Neuron*, **28**, 165–181.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Hutchinson, E. G. & Thornton, J. M. (1996). *Protein Sci.* **5**, 212–220.
Huvent, I., Belrhali, H., Antoine, R., Bompard, C., Locht, C., Jacob-Dubuisson, F. & Villeret, V. (2006). *J. Mol. Biol.* **356**, 1014–1026.
Imaizumi, A., Suzuki, Y., Ono, S., Sato, H. & Sato, Y. (1983). *Infect Immun.* **41**, 1138–1143.
Kabsch, W. (1976). *Acta Cryst.* A**32**, 922–923.
Kleywegt, G. J. & Jones, T. A. (1996). *Acta Cryst.* D**52**, 826–828.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Mayer, M. L. (2005). *Neuron*, **45**, 539–552.
Mayer, M. L., Olson, R. & Gouaux, E. (2001). *J. Mol. Biol.* **311**, 815–836.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Parkhill, J. *et al.* (2003). *Nature Genet.* **35**, 32–40.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct Biol.* **6**, 458–463.
Perrakis, A., Sixma, T. A., Wilson, K. S. & Lamzin, V. S. (1997). *Acta Cryst.* D**53**, 448–455.
Saier, M. H. Jr & Paulsen, I. T. (1999). *Res. Microbiol.* **150**, 689–699.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Takahashi, H., Inagaki, E., Kuroishi, C. & Tahirov, T. H. (2004). *Acta Cryst.* D**60**, 1846–1854.
Tam, R. & Saier, M. H. Jr (1993). *Microbiol. Rev.* **57**, 320–346.
Thalen, M., van den Ijssel, J., Jiskoot, W., Zomer, B., Roholl, O., de Gooijer, C., Beuvery, C. & Tramper, J. (1999). *J. Biotechnol.* **75**, 147–159.
Winnen, B., Hvorup, R. N. & Saier, M. H. Jr (2003). *Res. Microbiol.* **154**, 457–465.