Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# Structure of *m*-carboxyphenyl-α-D-galactopyranoside complexed to heat-labile enterotoxin at 1.3 Å resolution: surprising variations in ligand-binding modes

**Wendy E. Minke,[a] Jason Pickens,[a] Ethan A. Merritt,[a] Erkang Fan,[a] Christophe L. M. J. Verlinde[a] and Wim G. J. Hol[a,b]***

[a]Department of Biological Structure and Biomolecular Structure Center, University of Washington, Box 357742, Seattle, Washington 98195, USA, and [b]Howard Hughes Medical Institute, University of Washington, Box 357742, Seattle, Washington 98195, USA

Correspondence e-mail: hol@gouda.bmsc.washington.edu

Received 20 December 1999
Accepted 4 April 2000

**PDB Reference:** PLT $B_5$–MCPG, 1djr.

In the quest to develop drugs against traveller's diarrhoea and cholera, the structure of the *B* pentamer of heat-labile enterotoxin (LT) complexed with a new receptor-binding antagonist, *m*-carboxyphenyl-α-D-galactopyranoside, has been determined. The high resolution obtained for this structure allowed anisotropic refinement of the model. It was also now possible to confirm at a near-atomic resolution the structural similarity between the *B* subunits of LT and the closely related cholera toxin (CT), including the similarity in deviations of planarity of the same peptide unit in LT and CT. The structure of the LT complex clearly revealed different conformations for the *m*-carboxyphenyl moiety of the ligand in the five *B* subunits of LT, while the binding modes of the well defined galactopyranoside moieties were identical. In two binding sites the *m*-carboxyphenyl moiety displayed no significant electron density, demonstrating significant flexibility of this moiety. In a third binding site the *m*-carboxyphenyl moiety could be modelled unambiguously into the density. The two remaining binding sites were involved in crystal packing contacts and the density for the ligands in these two binding sites clearly revealed different binding modes, of which one conformation was identical to and one completely different from the conformation of *m*-carboxyphenyl-galactopyranoside in the third subunit. The multiple binding modes observed in the crystal may represent the ensemble of conformations of *m*-carboxyphenyl-α-D-galactopyranoside complexed to LT in solution.

## 1. Introduction

Heat-labile enterotoxin (LT) and cholera toxin (CT) are two closely related proteins. Heat-labile enterotoxin is produced by *Escherichia coli* and causes traveller's diarrhoea. Cholera toxin is produced by *Vibrio cholerae*, which is the causative agent of cholera. The two toxins not only share 80% sequence identity, but also have very similar structures and modes of action. The proteins are comprised of one *A* subunit and five identical *B* subunits. The *B* subunits form a regular pentamer which has five receptor-binding sites through which LT and CT interact with the receptor GM1 [Galβ1–3GalNAcβ1–4-(Neu5Acα2–3)Galβ1–4Glc-ceramide] on the surface of human epithelial cells. Subsequently, the *A* subunit is translocated across the membrane and in the cytosol it ADP-ribosylates $G_{s\alpha}$. Upon ADP-ribosylation, $G_{s\alpha}$ remains in its active form and continuously stimulates adenylate cyclase. The resulting increased cAMP levels activate ion channels, leading to a greatly enhanced efflux of ions and water from the cells, which are the major symptoms of the diseases caused by LT and CT.

One strategy to prevent the toxic activity of LT and CT is to block their entry into host cells. Therefore, ligands were designed that are able to inhibit receptor binding by LT and CT. So far, all the receptor-binding antagonists tested and synthesized in our laboratory contain a galactose 'anchor', utilizing the fact that the terminal galactose of the natural receptor is important for binding despite the low $IC_{50}$ of 58 m*M* for galactose alone. The best monomeric galactose derivative obtained inhibits LT binding to the receptor 1500 times more efficiently than galactose itself. However, while the most potent inhibitor has an $IC_{50}$ of 40 µ*M* in the LT ELISA assay (Minke, Hong *et al.*, 1999), the oligosaccharide portion of GM1 (GM1 OS) is a much better inhibitor, with an $IC_{50}$ of 10 n*M* in the same assay (Minke, Roach *et al.*, 1999). Clearly, there is a need for considerably more powerful receptor-binding antagonists.

Design of high-affinity receptor-binding antagonists is greatly facilitated by structural information on the binding modes of known inhibitors. Fortunately, several ligands have already been co-crystallized with the toxins and their structures subsequently solved (Sixma *et al.*, 1992; Merritt *et al.*, 1994, 1997, 1998; van den Akker *et al.*, 1996). For other ligands, it has been possible to predict the binding mode using molecular-modelling techniques (Minke, Diller *et al.*, 1999). Recently, a new ligand was designed, synthesized and tested in our laboratory: *m*-carboxyphenyl-α-D-galactopyranoside (Pickens *et al.*, 2000). The $IC_{50}$ of *m*-carboxyphenyl-α-D-galactopyranoside (MCPG) in the LT ELISA assay is ten times better than the $IC_{50}$ of the parent compound galactose (Minke, Roach *et al.*, 1999; Pickens *et al.*, 2000). Since it might be beneficial to include structural features of MCPG's binding mode into future ligand-design efforts and since molecular modelling studies did not provide an unambiguous answer for the binding mode of MCPG to LT (Minke, Hol & Verlinde, unpublished), we decided to crystallize the ligand in complex with the porcine LT *B* pentamer (pLT $B_5$) and solve its structure.

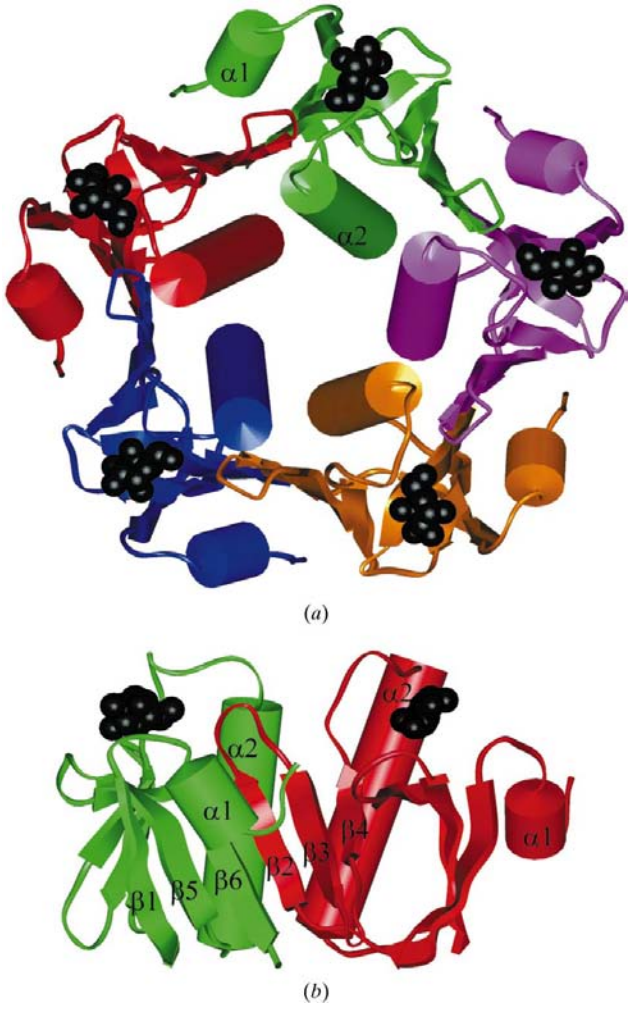


**Figure 1**
Schematic diagram of (*a*) the *B* pentamer of LT, with each of the *B* subunits depicted in a different colour and the galactoses in each subunit depicted in black, and (*b*) two of the *B* subunits forming the six-stranded antiparallel β sheet. The secondary-structure elements are α1, 5–9; β1, 15–22; β2, 26–30; β3, 37–41; β4, 47–50; α2, 59–78; β5, 82–88; β6, 94–102.

For a thorough general description of the three-dimensional structure of LT, the reader is referred to the initial structural LT papers by Sixma *et al.* (1991, 1993). However, in order to facilitate the description of the structure presented in this paper, the main features of the LT *B* subunits and their receptor-binding sites are reiterated briefly here. The five LT *B* subunits pack together to form a 'donut', with the monomers arranged symmetrically around a fivefold axis, leaving a pore in the middle of the pentamer (Fig. 1). Each of the *B* monomers has two three-stranded antiparallel β-sheets, a small N-terminal helix (α1) and a larger helix lining the pore (α2). Sheet I (β2, β3 and β4) of each *B* subunit forms a six-stranded antiparallel β-sheet with sheet II (β1, β5 and β6) of a neighbouring subunit. The LT *B* pentamer has five binding sites for its receptor GM1. Each receptor-binding site lies primarily within a single *B* monomer, in a cleft formed by the 10–14, 51–58 and 89–93 loops. At one end of this cleft, the 31–36 loop from an adjacent monomer forms the remainder of the binding site. The crystal structures that have been solved of LT and LT $B_5$ in complex with several galactose-containing ligands (Sixma *et al.*, 1992; Merritt *et al.*, 1994, 1997; Van den Akker *et al.*, 1996) have a number of features in common: (i) for each ligand, the crystallographically observed binding mode of the galactose substituent, which is a common building block in all ligands in these complexes, is similar in all five subunits of the LT pentamer, (ii) the galactose moiety invariably has the same binding mode in all LT structures and (iii) all crystallographically observed binding sites have conserved water molecules, of which the most relevant for ligand binding are classified as water molecules 1 to 5 (Merritt *et al.*, 1994).

A pleasant surprise of the pLT $B_5$–MCPG complex crystal was the high resolution to which it diffracted. Currently, the highest resolution LT or CT structure in the Protein Data Bank is CT $B_5$ complexed to the oligosaccharide portion of its natural receptor GM1, at 1.25 Å resolution (Merritt *et al.*, 1998). For LT, the highest resolution crystal structure in the Protein Data Bank was until now the porcine LT $B_5$–thiodigalactose complex, at 1.7 Å resolution (Merritt *et al.*, 1997).

**Table 1**
Crystallographic data for the pLT *B*5–MCPG complex.

| | |
|---|---|
| Space group | $P2_1$ |
| Unit-cell parameters | |
| *a* (Å) | 42.36 |
| *b* (Å) | 95.04 |
| *c* (Å) | 67.51 |
| $\beta$ (°) | 100.7 |
| Resolution limits (Å) | 20–1.3 |
| Observed reflections | 837038 |
| Unique reflections | |
| Working set | 121261 |
| Test set | 6479 |
| Mosaicity (°) | 0.63 |
| $R_{merge}$ (%) | |
| Overall | 7.2 |
| Highest shell (1.35–1.30 Å) | 46.9 |
| Completeness (%) | |
| Overall | 99.4 |
| Highest shell (1.35–1.30 Å) | 95.5 |
| Refined model | |
| $R_{work}$ (%) | 13.4 |
| $R_{free}$ (%) | 18.9 |
| Protein atoms | 4106 ($B_{avg}$ = 10.8 Å$^2$) |
| MCPG atoms | 87 ($B_{avg}$ = 17.9 Å$^2$) |
| Glycerol atoms | 18 ($B_{avg}$ = 18.4 Å$^2$) |
| Water molecules | 807 ($B_{avg}$ = 27.4 Å$^2$) |
| Protein stereochemistry | |
| R.m.s. non-ideality, bond length† (Å) | 0.021 |
| R.m.s. non-ideality, bond angles† (°) | 2.81 |
| R.m.s. $\Delta B_{iso}$ between bonded atoms (Å$^2$) | 2.25 |
| Overall average *G* factor‡ | −0.20 |
| Backbone quality index§ | 0.41 |

† Target stereochemistry from Engh & Huber (1991). ‡ *G* factor as reported by *PROCHECK* (Laskowski *et al.*, 1993). § Backbone *Z* score reported by *WHATIF* (Hooft *et al.*, 1996).

With the high resolution we obtained for the new pLT $B_5$–MCPG crystal structure presented here, we were not only able to determine the ligand-binding properties for MCPG, but the 1.3 Å resolution structure also enabled us to refine the model anisotropically and compare LT and CT *B* pentamers at near-atomic resolution.

## 2. Materials and methods

### 2.1. Sample preparation and crystallization

Porcine LT $B_5$ (pLT $B_5$) was obtained by thermo-induced expression of plasmid pROFIT-LTB in *E. coli* strain MC1061 (Feil *et al.*, 1996; Minke, Roach *et al.*, 1999). Purification was performed following the protocol developed by Uesaka *et al.* (1994), except that the clarified cell lysate was subjected to a 30% ammonium sulfate precipitation in 20 m*M* Tris–HCl buffer (pH 7.5) prior to affinity chromatography on immobilized D-galactose.

Synthesis of MCPG is described elsewhere (Pickens *et al.*, 2000). Crystals of MCPG complexed to LT $B_5$ formed after two weeks at room temperature. The crystals were grown in capillaries from an initially three-layered solution because this method had proven successful in the co-crystallization of an isosteric compound, *m*-nitrophenyl-$\alpha$-galactopyranoside (Merritt *et al.*, 1997). The bottom layer consisted of 4 µl 42% PEG 6000 in 50 m*M* NaCl and 100 m*M* Tris–HCl at pH 7.5, the middle layer of 4 µl 200 m*M* MCPG in 100 m*M* Tris–HCl at pH 7.5 and the top layer of 5 µl protein solution (pLT $B_5$ in 200 m*M* NaCl, 1 m*M* EDTA, 3 m*M* azide and 50 m*M* Tris–HCl pH 7.4). The protein solution had an OD of 4.4 at 280 nm, which corresponds to a pLT $B_5$ concentration of 3.6 mg ml$^{-1}$ (Minke, Roach *et al.*, 1999). The crystal chosen for data collection was first transferred from the capillary to an artificial mother liquor consisting of 15% PEG 6000, 92 m*M* NaCl and 80 m*M* Tris–HCl pH 7.5. Subsequently, the crystal was transferred to a cryosolution consisting of artificial mother liquor with 20% glycerol. After residing less than a minute in the cryosolution, the crystal was flash-frozen in liquid nitrogen. The crystal belongs to the monoclinic space group $P2_1$, with unit-cell parameters $a$ = 42.4, $b$ = 95.0, $c$ = 67.5 Å, $\beta$ = 100.7°.

### 2.2. Data collection and processing

Data were collected from a single crystal at SSRL beamline 9-1 with a MAR Research imaging-plate scanner and synchrotron radiation with a wavelength of 0.98 Å. Reflections were collected to 1.3 Å in 157 images with an oscillation angle of 1.0° per image. Data were processed with *DENZO* and were merged and scaled with *SCALEPACK* (Otwinowski & Minor, 1997). The merged data set was obtained from 837 038 observations and contained 127 740 unique reflections with an overall $R_{merge}$ of 7.2% and a completeness of 99.4%. 68.4% of all reflections were greater than three standard deviations. Statistics for the data set used in the refinement are given in Table 1.

### 2.3. Structure solution

An initial model for the pLT $B_5$–MCPG structure was obtained *via* molecular replacement using the program *AMoRe* (Navaza, 1994). The atomic coordinates of the pLT B5 pentamer as seen in the pLT $B_5$–thiodigalactoside complex (accession code 1lt5 in the Protein Data Bank; Merritt *et al.*, 1997) were used as a search model. The cross-rotation function for data in the resolution range 8.0–3.0 Å resulted in five clear solutions related by a fivefold rotation. The correlation coefficients of these peaks ranged from 0.38 to 0.40, while the correlation coefficient for the highest background peak was 0.15. Subsequently, the translation search resulted in solutions with correlation coefficients between 0.63 and 0.67 and *R* factors between 35.4 and 37.5%. In contrast, the highest background peak from the rotation search had a correlation coefficient of 0.14 and an *R* factor of 54.7% after the translation search. After rigid-body fitting of the whole pentamer with *AMoRe* using the data in the resolution range 8.0–3.0 Å, the *R* factor was 35.0%.

### 2.4. Refinement

Refinement of the atomic coordinates and displacement parameters was carried out using the program *SHELXL*97 (Sheldrick & Schneider, 1997). The auxiliary program *SHELXPRO* was used for preparation of the input parameters for the refinement, for map calculation and for model

**Table 2**
Summary for the different stages of the refinement of the pLT $B_5$–MCPG complex.

| Refinement step | Resolution limits (Å) | Water molecules | Ligand atoms | Disordered protein atoms | Number of parameters | $R$ (%) | $R_{free}$ (%) |
|---|---|---|---|---|---|---|---|
| 1. Rigid-body monomers | 20–2 | — | — | — | 33 | 43.3 | 43.9 |
| 2. Coordinates & isotropic $B$ | 20–1.5 | 148 | — | — | 17075 | 23.6 | 27.9 |
| 3. Waters | 20–1.3 | 446 | — | — | 18267 | 20.6 | 24.2 |
| 4. Anisotropic protein | 20–1.3 | 446 | — | — | 38867 | 18.6 | 23.9 |
| 5. MCPG | 20–1.3 | 825 | 96 | — | 41202 | 15.7 | 21.0 |
| 6. Glycerol | 20–1.3 | 796 | 114 | — | 41138 | 15.6 | 20.9 |
| 7. Disorder | 20–1.3 | 795 | 105 | 81 | 40583 | 15.4 | 20.7 |
| 8. Anisotropic water | 20–1.3 | 807 | 111 | 81 | 45970 | 14.2 | 20.0 |
| 9. Riding H atoms | 20–1.3 | 807 | 111 | 81 | 45970 | 13.4 | 18.9 |
| 10. Final model | 20–1.3 | 807 | 105 | 81 | 45792 | 13.4 | 18.9 |

analysis. Map fitting and real-space refinement were carried out using the graphics program *XTALVIEW* (McRee, 1992). In addition, the programs *PROCHECK* (Laskowski *et al.*, 1993), *WHATIF* (Vriend & Sander, 1993) and *PARVATI* (Merritt, 1999) were used for structure evaluation. The final results of the refinement are listed in Table 1 and a summary of the different stages of the refinement is given in Table 2.

The structure was refined against intensities rather than amplitudes and all parameters were refined simultaneously. During the refinement, 5% of the reflections were reserved for free $R$-factor calculation (Brünger, 1992). A full-matrix least-squares algorithm was used for rigid-body refinement, while a conjugate-gradient least-squares algorithm was used for all other isotropic and anisotropic refinement steps. Each round of refinement was alternated with a round of manual refitting and model building, using difference omit electron-density maps.

Distance, planarity, chiral volume and antibumping restraints were applied for the protein and its ligands during the refinement. The target values for the 1–2 and 1–3 distances in the protein were based on a study by Engh & Huber (1991). The target values for galactose were identical to those used in the refinement of the 1.25 Å CT $B_5$–GM1 OS structure (Merritt *et al.*, 1998), while those for MCPG and glycerol were based on small-molecule structures in the Cambridge Structural Database (Allen & Kennard, 1993*a*,*b*). The standard deviations of the bond lengths were initially restrained to 0.02 Å and those of the 1–3 distances to 0.04 Å, while later in the refinement they were relaxed to 0.03 and 0.06 Å, respectively. At the same time, the target planarity and chiral volume standard deviations were both relaxed from 0.1 to 0.15 Å$^3$.

Regions of diffuse solvent were modelled using Babinet's principle (Moews & Kretsinger, 1975). For modelling of defined water molecules, the positions as proposed by *SHELXL*97 after each refinement round were used. Water molecules were removed from the model if the $B$ factors increased above 60 Å$^2$.

Anisotropic displacement parameters ($U_{ij}$) were restrained during refinement by requiring that atoms closer than 1.7 Å have similar $U_{ij}$ components, with a standard deviation of 0.08 Å$^2$ for terminal atoms and 0.04 Å$^2$ for all other atoms. In addition, for pairs of chemically bonded atoms the components of the anisotropic displacement in the direction of the 1–2 or 1–3 bond were restrained to be equal, with a standard deviation of 0.01 Å$^2$. Anisotropic displacement parameters of water molecules were restrained to approximately isotropic behaviour, with a standard deviation of 0.1 Å$^2$.

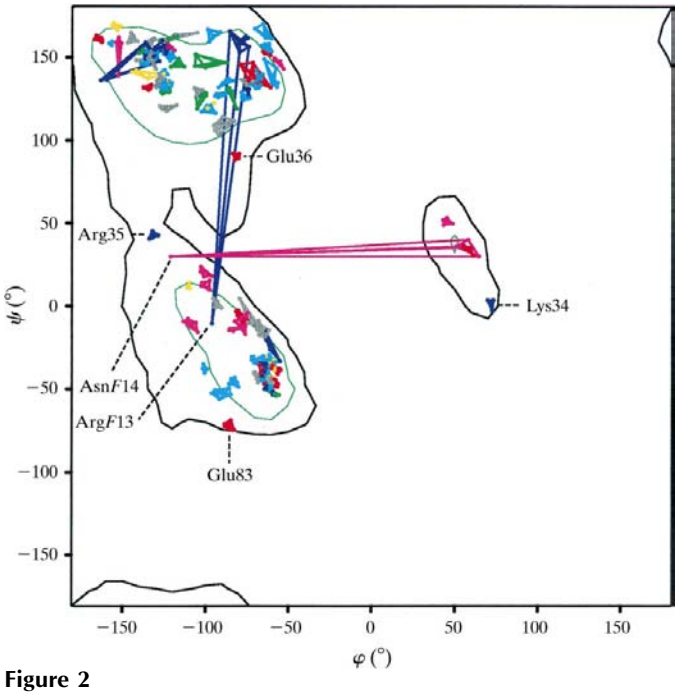


**Figure 2**
Ramachandran ($\varphi/\psi$) plot for 490 protein residues excluding glycines. Lines connect the fivefold-related residues to one another. Residues with $\varphi/\psi$ values outside of the most favourable regions are labelled, as well as residues that have deviating angles from their fivefold-related partners. Colour coding is as follows: yellow, Cys, Met; green, Phe, Tyr, Trp, His; cyan, Ala, Leu, Ile, Val, Pro; red, Glu, Asp; blue, Arg, Lys; purple, Gln, Asn; grey, Ser, Thr. 99.0% of the residues have $\varphi/\psi$ values in the core region and 85.1% are in the inner core region, according to Kleywegt & Jones (1996).

## 3. Results

### 3.1. Accuracy of the model

The final $R$ factor is 13.4% for the 121 261 unique reflections in the resolution range 20–1.3 Å and 10.6% for the 93 100 reflections with $F_o > 4\sigma(F_o)$; the restrained goodness of fit is 1.85 as calculated by *SHELX*. The final difference Fourier map was reasonably flat, with a standard deviation of 0.06 e Å$^{-3}$. Significant positive and negative peaks in the final ($F_o - F_c$) map are located in the vicinity of the disordered Met101 residues in the five subunits, as well as near Asn*H*94 (where *H* refers to one of the *B* subunits, which are labelled *D*

to *H*), where a glycerol was initially modelled and later removed owing to disorder.

The accuracy of the model is reflected by the overall coordinate estimated standard uncertainty of 0.47 Å based on *R* value (Cruickshank, 1999). In addition, the program *PROCHECK* (Laskowski *et al.*, 1993) classified the model as being of excellent quality and consistent with other structures of similar resolution.

The final model of the pLT $B_5$–MCPG complex also shows good stereochemistry, as was shown by analysis of the main-chain ($\varphi/\psi$) and side-chain ($\chi_1$) torsion angles and of the peptide planarity ($\omega$). In the Ramachandran plot, 99% of the $\varphi/\psi$ angles are in the core region (Fig. 2). The residues with $\varphi/\psi$ angles outside the most favourable regions are identical in each pLT *B* monomer and the fivefold-related $\varphi/\psi$ angles of these residues are clustered tightly. In addition, the same four unusual backbone angles of residues Lys34, Arg35, Glu36 and Glu83 have been observed in other CT and LT complex structures (Merritt *et al.*, 1997, 1998), indicating that these unusual dihedral angles are characteristic features of LT and CT *B* pentamers. The only two residues with $\varphi/\psi$ angles deviating from their fivefold-related partners, Arg13 and Asn14 from monomer *F*, are intimately involved in crystal packing contacts. Since these subunit *F* deviations have not been seen in any other structure, these dihedral angles in subunit *F* are specific for the current crystal form.

For the $\chi_1$ angles, the program *WHATIF* reports unusual rotamers based on a database of solved protein structures. In the pLT $B_5$–MCPG structure only six residues have a *WHATIF* rotamer value below 0.4: Ser95 in subunits *D*, *E* and *H* and Ile99 in subunits *D*, *F* and *H*. The occurrence of these unusual rotamers in multiple subunits suggests that the refined $\chi_1$ angles correctly describe the actual structure. For the $\omega$ angles the mean value is 178.2° with a standard deviation of 6.3°, which agrees well with the distribution of $\omega$ values found in the Cambridge Structural Database by MacArthur & Thornton (1996), who reported a mean value of 179.7° and a standard deviation of 5.9°. The most prominent outlier is the peptide group between Gln49 and Val50, with a mean $\omega$ value for the *B* pentamer of 153.5° and a standard deviation of 2.1° among the five *B* subunits. The non-planarity of this peptide has previously been observed in the CT $B_5$–GM1 OS structure, with a mean $\omega$ value for the *B* pentamer of 153.7° and a standard deviation of 0.7°. It was suggested that the non-planarity might be a result of backbone strain induced by ligand binding, since this peptide immediately precedes a loop which is part of the receptor-binding site (Merritt *et al.*, 1998).

The anisotropy of the LT $B_5$–MCPG model was analysed using the program *PARVATI* (Merritt, 1999). Anisotropy is defined here as the ratio of the minimum and maximum principal axes of the atomic ellipsoids. The distribution of the anisotropy by atom group is depicted in Fig. 3, which shows that none of the atoms is perfectly isotropic (anisotropy = 1.00), while only five atoms have very high anisotropy (<0.10). The mean anisotropy values for the protein, solvent and MCPG are 0.45, 0.40 and 0.48, respectively, with standard deviations of 0.13, 0.14 and 0.10, respectively, which corresponds to a typical distribution of anisotropy for high-resolution structures deposited with the Protein Data Bank (which have a mean of 0.45 and a standard deviation of 0.15; Merritt, 1999).

### 3.2. Overall structure

Some features of the overall structure are influenced by crystal packing, which will therefore be discussed briefly. Under the described crystallization conditions, the pLT $B_5$–MCPG complex packs in a unit cell (Table 1) which has not been observed in any LT or CT crystal form obtained previously and which is depicted in Fig. 4. The Matthews coefficient of this crystal is 2.2 $Å^3\ Da^{-1}$ (Matthews, 1968), corresponding to a solvent content of 46%. In this unit cell there are two similar contacts between symmetry-related pentamers: (i) $\beta$-strand 1*E* interacts with the loop between $\beta$-strands 3*D* and 4*D* of a symmetry-related pentamer, (ii) $\beta$-strand 1*F* interacts with the loop between $\beta$-strands 3*H* and 4*H* of a different symmetry-related pentamer and in this site the backbone carbonyl group of Tyr12*F* forms a hydrogen bond with the NZ of Lys44*H*. However, the major interaction between symmetry-related pentamers is between the ligand-binding site of subunit *F* and the ligand-binding site of subunit *H* of a symmetry-related pentamer.

The conformation of the protein in the 1.3 Å pLT $B_5$–MCPG structure is essentially the same as in previously reported structures. The average r.m.s. difference in CA positions with the search model (the pLT $B_5$–thiodigalactose complex) is 0.45 Å based on the five possible superpositions of the pentamers. The main differences are in the loops, especially in the 10–14 loop and in the 51–60 loop of subunit *F* of the pLT $B_5$–MCPG structure. The average r.m.s. difference in CA positions with the high-resolution CT-GM1 OS structure is 0.51 Å based on the five possible superpositions of the pentamers. Again, the main differences are in the loops: besides the significantly deviating 10–14 loop of subunit *F* of the pLT $B_5$–MCPG structure, in this comparison the 42–46 loop of subunits *E* and *G* also show relatively large deviations

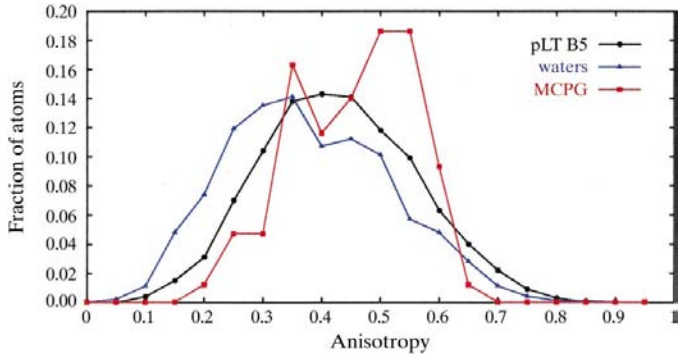


**Figure 3**
*PARVATI* (Merritt, 1999) plot describing the distribution of atomic anisotropy in the high-resolution model of the pLT $B_5$–MCPG complex. The anisotropy is the ratio of minimum and maximum eigenvalues of the $U_{ij}$ matrix. A perfectly isotropic atom has an anisotropy of 1.0. For the three atom groups the anisotropy corresponds to the distribution of high-resolution structures deposited in the Protein Data Bank (Merritt, 1999).

of up to 2.0 Å. Superposition of the *B* subunits of the pLT $B_5$–MCPG structure onto each other leads to an average r.m.s. difference of 0.27 Å for the 103 CA atoms. Again, the main differences are in the 10–14 loop, in the 42–46 loop and in the 51–60 loop. In addition, there are differences of up to 1 Å in the 31–36 loop. The differences in conformations for the four loops of the pLT $B_5$–MCPG crystal structure mentioned above can be explained as follows. The deviations in the 10–14 loop of subunit *F* result from ligand binding in combination with crystal packing. The differences in the conformation of the 51–60 loop have also been observed in the unliganded LT and CT structures, while the 51–60 loop becomes more ordered upon ligand binding. The 31–36 loop is also part of the ligand-binding site and therefore the differences in the conformation are probably also a consequence of ligand binding. The differences between the CT and LT structures in the 42–46 loop presumably arise from the sequence differences between CT and pLT; residues 44 and 46 are a serine and a glutamine in pLT *B* and an asparagine and an alanine in CT *B*, respectively.

The packing of the unit cell of the pLT $B_5$–MCPG crystal is slightly tighter than observed for most *B*-pentamer crystals, which in combination with the high resolution may explain the relatively high number of observed water molecules (807). The waters were analysed using the program *HBPLUS* (McDonald & Thornton, 1994), using a hydrogen-bond cutoff of 2.5 Å between the H and acceptor atoms, and with angular cutoff values of 120 and 90° (as minimum acceptable values) at the H and O atoms, respectively. It appeared that out of the 807 water molecules, 597 occur in the first hydration shell. The average $B_{eq}$ [equivalent isotropic displacement parameter; $B_{eq} = 8\pi^2 \times (U_{11} + U_{22} + U_{33})/3$] of these 597 first-shell water molecules is 22.3 Å$^2$, while the average $B_{eq}$ of the remaining 216 waters is 41.8 Å$^2$. There are 102 water molecules that make one hydrogen bond to either the protein or another crystallographically observed water, 192 water molecules that make two hydrogen bonds, 230 water molecules that make three hydrogen bonds and 283 water molecules that take full advantage of their hydrogen-bonding ability.

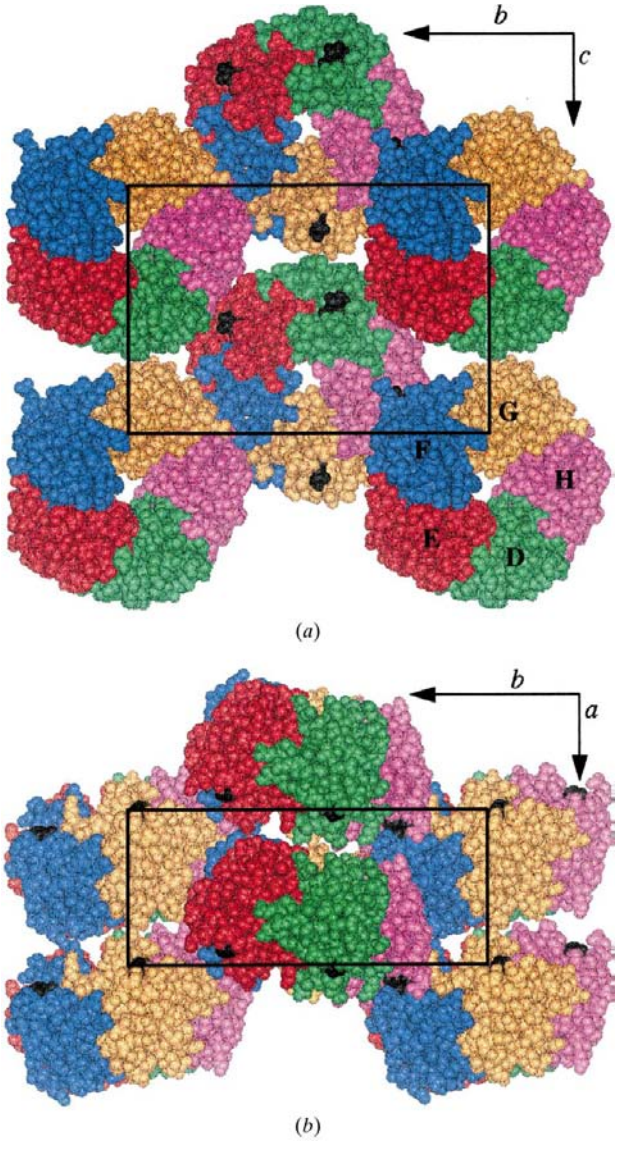


**Figure 4**
Lattice packing of the pLT $B_5$–MCPG crystal. Projection along the *a* axis (*a*) and along the *c* axis (*b*). The ligands are depicted in black, while the *B* subunits are color coded as follows: green, *D*; red, *E*; blue, *F*; orange, *G*; purple, *H*. The most extensive crystal packing contacts are between subunit *F* of one pentamer and subunit *H* of another pentamer.

### 3.3. Ligands, disordered residues and other specific structural characteristics

The final model of the pLT $B_5$–MCPG structure is complete with the exception of 13 atoms for which there is a lack of density. Three residues, Lys63*D*, Arg13*E* and Arg13*G*, have no density for some side-chain atoms and therefore these atoms were omitted from the model. In addition, the C-terminal residue of subunit *E* has very poor density and the carboxylate group of this residue was also omitted from the model.

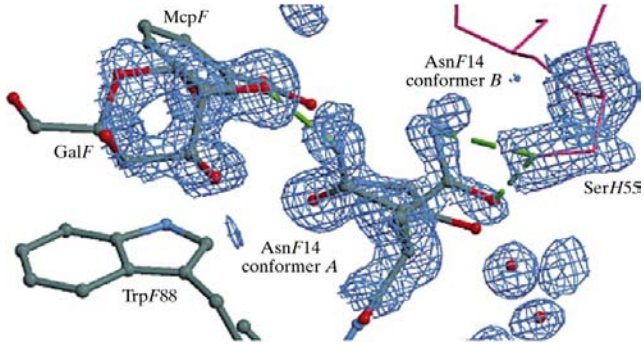


**Figure 5**
Asn14*F* with two alternative conformations in a $\sigma_A$-weighted ($2mF_o - DF_c$) electron-density map contoured at 1$\sigma$. For conformer *A* the OD1 and ND2 assignments were based on the electron density which forms a hydrogen bond with galactose and for conformer *B* bound to subunit *F* they were based on the hydrogen-bonding pattern of the interactions with Ser55 of subunit *H*. The two conformations have occupancies of 59 and 41%, respectively. This figure and Fig. 6 were produced using *XFIT* (McRee, 1992) and *Raster3D* (Merritt & Bacon, 1997).

As frequently observed in structures at sufficiently high resolution, several side chains adopt more than one discrete conformation. The following 20 residues have been interpreted to have two conformers and their positions and occupancies have been refined as such: Lys23*D*, Lys43*D*, Met101*D*, Glu7*E*, Lys43*E*, Glu79*E*, Lys102*E*, Glu7*F*, Asn14*F*, Asp59*F*, Ile82*F*, Lys62*G*, Lys84*G*, Met101*G*, Asn103*G*, Glu7*H*, Arg13*H*, Leu25*H*, Lys81*H* and Met101*H*. It is noticeable that none of these residues is disordered in all five subunits or even in four out of the five subunits. Only the disorder for Glu7 and Met101 was observed three times. For Glu7 atom OE1 forms a hydrogen bond to the backbone amide of Gln3 in both conformations, while OE2 is rather exposed to the solvent. For Met101 the disorder is hard to model, probably because the side chain is in a rather large hydrophobic cavity. Most of the other disordered residues are very solvent exposed and since they do not contribute to ligand binding they will not be discussed further. The most interesting disordered residue is Asn14*F* (Fig. 5). In one of its conformations ND2 forms a hydrogen bond with O2 of galactose bound to receptor site *F*, while in the other conformation OD1 forms a hydrogen bond to Ser55*H* of a symmetry-related pentamer.

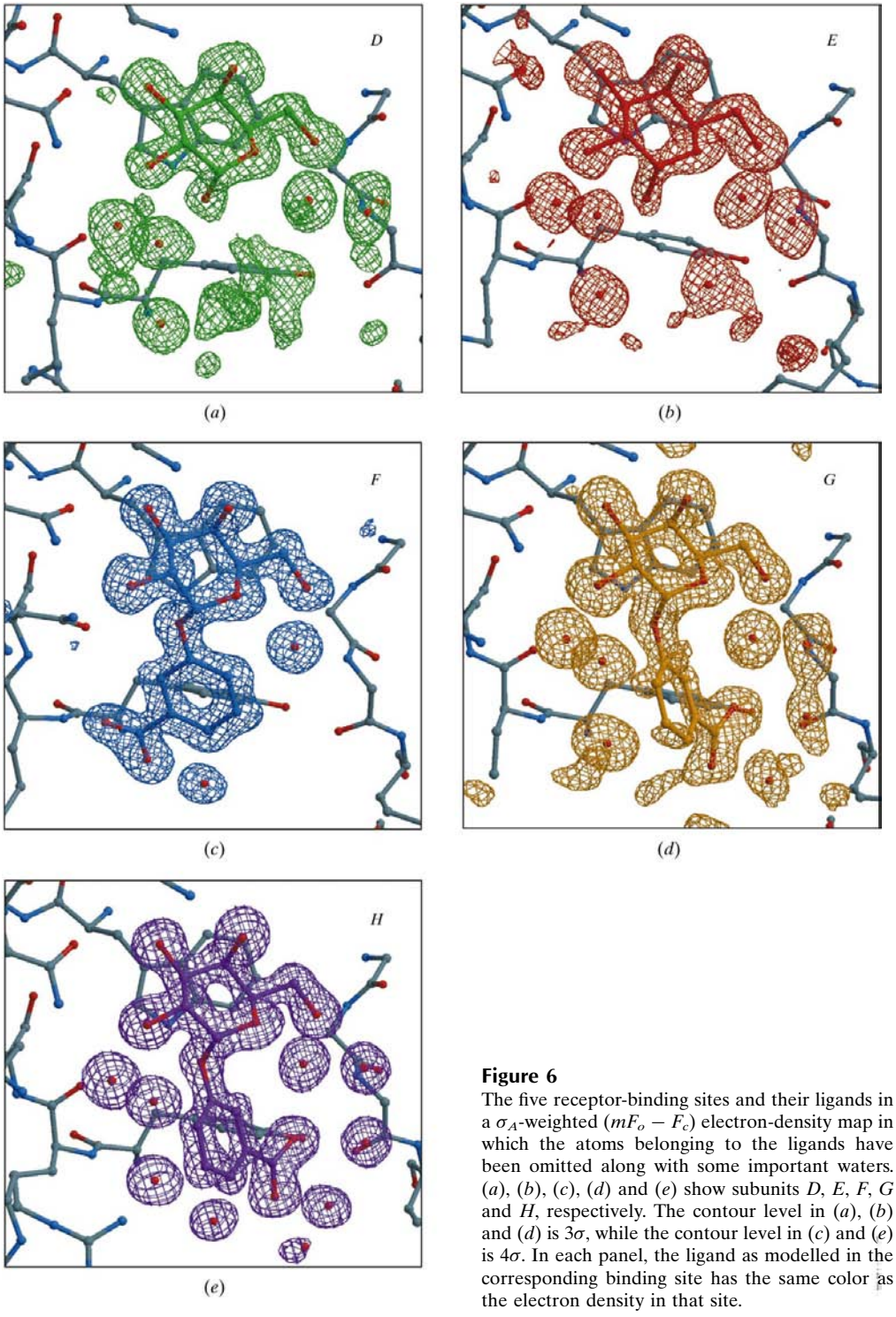


**Figure 6**
The five receptor-binding sites and their ligands in a $\sigma_A$-weighted ($mF_o - F_c$) electron-density map in which the atoms belonging to the ligands have been omitted along with some important waters. (*a*), (*b*), (*c*), (*d*) and (*e*) show subunits *D*, *E*, *F*, *G* and *H*, respectively. The contour level in (*a*), (*b*) and (*d*) is 3$\sigma$, while the contour level in (*c*) and (*e*) is 4$\sigma$. In each panel, the ligand as modelled in the corresponding binding site has the same color as the electron density in that site.

The atomic resolution structure of the pLT *B*5–MCPG structure indicates the presence of three bound glycerol molecules. The glycerol with the clearest density was found on the interface between two symmetry-related pentamers where $\beta$-strand 1*F* interacts with the loop between $\beta$-strands 3*H* and 4*H* of a symmetry-related pentamer. This glycerol molecule forms three hydrogen bonds to one pentamer and one direct and one water-mediated hydrogen bond to the symmetry-related pentamer. The other two glycerols were found in identical pockets formed by the same residues, one in subunit *D* and the other in subunit *G*, where they form three direct hydrogen bonds with the protein. The same pockets in the other subunits, *E*, *F* and *H*, also show extra density. However, the density in these sites is not clear enough to model glycerol molecules.

The most remarkable feature of the crystal structure is the variation in the binding mode of the ligand MCPG. The density for the galactose moiety of MCPG is very clear

in all binding sites and galactose could be refined with the same binding mode in all subunits. However, the electron density for the *m*-carboxyphenyl moiety differs tremendously in each of the five subunits and it is clear that this moiety has several different binding modes when bound to pLT $B_5$ (Fig. 6). Subunit *E* displays the poorest density for the *m*-carboxyphenyl moiety; there is no density for any of the phenyl atoms and only a minute amount of density that may be ascribed to the carboxylate group. Subunit *D* has some density for the *m*-carboxyphenyl, in which this moiety was initially modelled and refined. However, the density for the C3′ atom of the *m*-carboxyphenyl moiety in subunit *D* could never be observed and therefore this *m*-carboxyphenyl moiety was omitted in later stages of the refinement. Subunit *G* clearly exhibits density for the *m*-carboxyphenyl moiety and the ligand could be modelled and refined. Subunit *H* contains even better defined density for the *m*-carboxyphenyl group and the moiety could be modelled in the same conformation as in subunit *G* and subsequently refined. Finally, subunit *F* reveals the best electron density, but it was impossible to model the *m*-carboxyphenyl moiety in the same conformation as in binding sites *G* and *H*. In contrast, the phenyl group has rotated about 180° around the bond between O1 of galactose and C3′ of *m*-carboxyphenyl, resulting in a completely different binding mode.

## 4. Discussion

The near-atomic resolution obtained for the pLT $B_5$–MCPG complex structure described here allows a detailed comparison with the CT $B_5$–GM1 OS structure at a resolution of 1.25 Å (accession code 3chb in the Protein Data Bank; Merritt *et al.*, 1998). Superpositioning of the CA atoms of the pentamers of the two high-resolution structures shows that they are very similar, with an r.m.s.d. of 0.51 Å. The main coordinate differences occur in the 10–14 loop of subunit *F* of pLT $B_5$–MCPG and in the 42–46 loop of subunits *E* and *G* of pLT $B_5$–MCPG. The differences in the 10–14 loop are a result of crystal packing and since this loop forms part of the receptor-binding site it is important for ligand-design purposes to realise that crystal contacts can have an effect on the conformation of this part of the binding site. The deviations in the 42–46 loop originate from the sequence difference between LT and CT and are not relevant to ligand design, since this loop resides on the other face of the pentamer to the receptor-binding site.

It is remarkable that some precise geometric aspects of the high-resolution structures of LT and CT are almost exactly the same: (i) the outliers in the Ramachandran plot previously observed for CT $B_5$–GM1 OS are also outliers in the pLT $B_5$–MCPG structure (Fig. 2), (ii) the non-planar peptide of residue 49 observed previously in CT is also non-planar in this LT structure, (iii) the conserved waters are present in the ligand-binding sites of both structures, (iv) the distribution of anisotropy is similar for both CT $B_5$–GM1 OS and pLT $B_5$–MCPG, with mean anisotropy values of 0.46 and 0.45, respectively, as determined using *PARVATI* (Merritt, 1999) and (v) the 80 atoms defining the GM1 OS binding site in LT and CT superimpose within 0.29–0.48 Å for 20 pairwise superpositions (excluding binding site *F* of pLT $B_5$–MCPG, which is involved in crystal contacts). This is a favourable factor in the design of receptor-binding antagonists: ligands that are designed to inhibit receptor binding of LT may also be effective against CT.

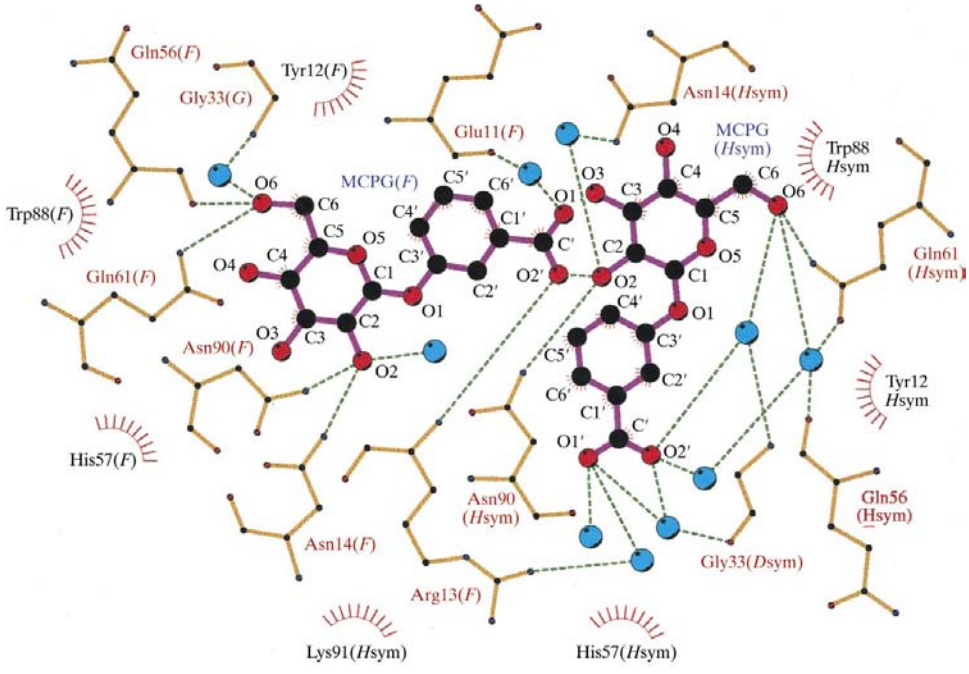


**Figure 7**
Schematic of the binding interactions of MCPG in binding site *F* and MCPG in binding site *H* of a symmetry-related pentamer. For clarity, the hydrogen bonds of O3 and O4 of the galactose moieties have been omitted. This figure was created using the program *LIGPLOT* (Wallace *et al.*, 1995).

The high-resolution pLT $B_5$–MCPG crystal structure revealed a number of entirely different conformations for part of the ligand bound to the receptor-binding site. In all five subunits the galactose moiety binds in a similar way as in all other LT and CT complex structures: the galactose stacks nicely on top of Trp88 and its hydroxyl groups form many hydrogen bonds with the protein. It is the position of the *m*-carboxyphenyl moiety that differs in the two well defined binding modes, of which one conformation is observed in subunit *F* and the other in both subunits *G* and *H* (Table 3 and Figs. 6*c*, 6*d* and 6*e*). The dihedral angles formed by atoms C2—C1—O1—C3′ (where the prime denotes an atom from the

**Table 3**
Properties of MCPG as observed in binding sites *F*, *G* and *H*.

| Feature | MCPGF | MCPGG | MCPGH |
|---|---|---|---|
| C2—C1—O1—C3′ torsion angle (°) | −165 | −144 | −150 |
| C1—O1—C3′—C2′ torsion angle (°) | 150 | −125 | −109 |
| LT—*X*†···O2′—MCPG hydrogen-bond distance (Å) | 2.75 | 2.63 | 2.78 |
| LT—*X*†···O2′—C7′—MCPG hydrogen-bond angle (°) | 120 | 153 | 144 |
| Buried solvent-accessible surface area of Tyr12 upon complex formation (Å$^2$) | 27.5 | 24.2 | 26.0 |
| Tyr12 CD2···C3′—MCPG distance (Å) | 3.75 | 5.03 | 4.59 |

† *X* is the backbone amide of Arg13 in subunit *F* and water 2 in subunits *G* and *H*.

*m*-carboxyphenyl moiety) are similar in the three fully observed MCPGs, while the dihedral angles formed by atoms C1—O—C3′—C2′ differ by almost 180° between the two binding modes. Consequently, the carboxylate groups are positioned differently and form hydrogen bonds with entirely different acceptors in the protein. In binding site *F* the O2′ of the carboxylate group forms a hydrogen bond with the backbone amide of Arg13, while in binding sites *G* and *H* this O atom forms a hydrogen bond to one of the conserved waters, water 2 (Table 3). In both binding modes a hydrophobic interaction takes place between the phenyl ring of MCPG and Tyr12, of which 25 Å$^2$ is buried upon binding of the ligand (Table 3). However, only for the binding mode as observed in subunit *F* is there a true van der Waals interaction between MCPG and Tyr12; the shortest distance between a phenyl atom of MCPG and a phenyl atom of Tyr12 is 3.75 Å, while in the other binding modes this shortest distance is more than 4.5 Å (Table 3). Besides the interaction of the MCPGs with their respective subunits, MCPG-*F* also interacts with MCPG-*H* in a symmetry-related pentamer: the O2′ of the carboxylate group of MCPG-*F* forms a hydrogen bond to the O2 of galactose of MCPG-*H* in a symmetry-related pentamer (Fig. 7).

Two of the five subunits (*D* and *E*) of the 1.3 Å pLT $B_5$–MCPG structure clearly reveal great flexibility of the *m*-carboxyphenyl moiety with respect to a well defined galactose moiety, as indicated by the absence of electron density for the carboxyphenyl group (Fig. 6). In subunit *G* a well defined conformation occurs. These three instances are the only receptor-binding sites not engaged in crystal contacts in the present structure. Therefore, together these three images of MCPG binding probably reflect most accurately the ensemble of conformations adopted by MCPG when bound to pLT in solution. The MCPG conformation in subunit *H* is similar to that in subunit *G* and thus confirms the favourable energy of the *G* conformation. The MCPG conformation in subunit *F* is distinctly different from all other conformations, with the carboxylate moiety in contact with a galactose hydroxyl group of a neighbouring subunit (Fig. 7). This '*F* conformation' probably represents a binding mode with a slightly higher energy, possibly owing to the need to displace loops 10–14 and 51–60 in the receptor-binding site. It is therefore likely that the LT $B_5$–MCPG complex in solution displays great rotational flexibility of the *m*-carboxyphenyl moiety with respect to a very well defined position of the galactose moiety. Part of the time the conformation seen in subunit *G* (and *H*) will be adopted and the *F* conformation will probably be adopted quite rarely.

The multiple binding modes of MCPG to five *B* subunits of LT show the power of crystallography to reveal different binding modes of the same ligand to the same polypeptide chain. At the same time, our results are a warning not to draw firm conclusions on the basis of structures of protein–ligand complexes with only one copy of the complex per asymmetric unit. For low-affinity ligands in particular, compounds may engage in more interaction modes with a target protein than a single copy per asymmetric unit of a single-crystal form might reveal.

We would like to thank Jungwoo Choe, David Chudzik, Mic Feese and the staff members at SSRL for their assistance with the data collection. This research was supported by the National Institutes of Health through grants to CLMJV (GM54618) and WGJH (AI34501). We are indebted to the Murdock Charitable Trust for financial support to the Biomolecular Structure Center.

## References

Akker, F. van den, Steensma, E. & Hol, W. G. (1996). *Protein Sci.* **5**, 1184–1188.
Allen, F. & Kennard, O. (1993*a*). *Chem. Design Autom. News*, **8**, 1.
Allen, F. & Kennard, O. (1993*b*). *Chem. Design Autom. News*, **8**, 31–37.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Cruickshank, D. W. (1999). *Acta Cryst.* D**55**, 583–601.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Feil, I. K., Reddy, R., de Haan, L., Merritt, E. A., van den Akker, F., Storm, D. R. & Hol, W. G. J. (1996). *Mol. Microbiol.* **20**, 823–832.
Hooft, R. W., Vriend, G., Sander, C. & Abola, E. E. (1996). *Nature (London)*, **381**, 272.
Kleywegt, G. J. & Jones, T. A. (1996). *Structure*, **4**, 1395–1400.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
MacArthur, M. W. & Thornton, J. M. (1996). *J. Mol. Biol.* **264**, 1180–1195.
McDonald, I. K. & Thornton, J. M. (1994). *J. Mol. Biol.* **238**, 777–793.
McRee, D. E. (1992). *J. Mol. Graph.* **10**, 44–46.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Merritt, E. A. (1999). *Acta Cryst.* D**55**, 1109–1117.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Merritt, E. A., Kuhn, P., Sarfaty, S., Erbe, J. L., Holmes, R. K. & Hol, W. G. J. (1998). *J. Mol. Biol.* **282**, 1043–1059.
Merritt, E. A., Sarfaty, S., Feil, I. K. & Hol, W. G. J. (1997). *Structure*, **5**, 1485–1499.
Merritt, E. A., Sixma, T. K., Kalk, K. H., van Zanten, B. A. & Hol, W. G. J. (1994). *Mol. Microbiol.* **13**, 745–753.
Minke, W. E., Diller, D. J., Hol, W. G. & Verlinde, C. L. (1999). *J. Med. Chem.* **42**, 1778–1788.
Minke, W. E., Hong, F., Verlinde, C. L. M. J., Hol, W. G. J. & Fan, E. (1999). *J. Biol. Chem.* **274**, 33469–33473.

Minke, W. E., Roach, C., Hol, W. G. J. & Verlinde, C. L. M. J. (1999). *Biochemistry*, **38**, 5684–5692.
Moews, P. C. & Kretsinger, R. H. (1975). *J. Mol. Biol.* **91**, 229–232.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–325.
Pickens, J., Minke, W. E., Verlinde, C. L. M. J., Hol, W. G. J. & Fan, E. (2000). In preparation.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Sixma, T. K., Kalk, K. H., van Zanten, B. A., Dauter, Z., Kingma, J., Witholt, B. & Hol, W. G. (1993). *J. Mol. Biol.* **230**, 890–918.
Sixma, T. K., Pronk, S. E., Kalk, K. H., van Zanten, B. A., Berghuis, A. M. & Hol, W. G. J. (1992). *Nature (London)*, **355**, 561–564.
Sixma, T. K., Pronk, S. E., Kalk, K. H., Wartna, E. S., van Zanten, B. A., Witholt, B. & Hol, W. G. J. (1991). *Nature (London)*, **351**, 371–377.
Uesaka, Y., Otsuka, Y., Lin, Z., Yamasaki, S., Yamaoka, J., Kurazono, H. & Takeda, Y. (1994). *Microb. Pathogen.* **16**, 71–76.
Vriend, G. & Sander, C. (1993). *J. Appl. Cryst.* **26**, 47–60.
Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). *Protein Eng.* **8**, 127–134.