Structures of the transcriptional regulator BgaR, a lactose sensor

Janet Newman,[a] Karine Caron,[b] Tom Nebl[a] and Thomas S. Peat[a]*

[a]Biomedical Program, CSIRO, 343 Royal Parade, Parkville, VIC 3052, Australia, and [b]Health and Biosecurity, CSIRO, Clunies Ross Street, Black Mountain, ACT 2601, Australia. *Correspondence e-mail: tom.peat@csiro.au

The structure of BgaR, a transcriptional regulator of the lactose operon in *Clostridium perfringens*, has been solved by SAD phasing using a mercury derivative. BgaR is an exquisite sensor of lactose, with a binding affinity in the low-micromolar range. This sensor and regulator has been captured bound to lactose and to lactulose as well as in a nominal apo form, and was compared with AraC, another saccharide-binding transcriptional regulator. It is shown that the saccharides bind in the N-terminal region of a jelly-roll fold, but that part of the saccharide is exposed to bulk solvent. This differs from the classical AraC saccharide-binding site, which is mostly sequestered from the bulk solvent. The structures of BgaR bound to lactose and to lactulose highlight how specific and nonspecific interactions lead to a higher binding affinity of BgaR for lactose compared with lactulose. Moreover, solving multiple structures of BgaR in different space groups, both bound to saccharides and unbound, verified that the dimer interface along a C-terminal helix is similar to the dimer interface observed in AraC.

1. Introduction

Living organisms have evolved a broad range of sensing systems for the detection of small molecules, such as metabolites and nutrients, and also of temperature and pH variations. Transcriptional regulators (TRs) play a major role in sensing small molecules and are a broad family of proteins that control cell development, cell differentiation and cell growth through the regulation of gene expression, and these proteins are found throughout all biological kingdoms. The mechanism by which TRs regulate gene transcription varies amongst this family of proteins. Often, a small molecule such as a metabolite or a nutrient binds to the effector-binding domain, leading to a conformation change that alters the interaction of the DNA-binding domain with the target DNA, ultimately affecting the efficiency of gene transcription. TRs have been divided into a broad range of families based on structural and binding similarities. GntR-family members (named after the gluconate operon repressor in *Bacillus subtilis*) are characterized by having a somewhat conserved DNA-binding domain linked to a diverse globular regulatory effector-binding domain. The structural divergence amongst the TRs of the GntR superfamily gives rise to six subfamilies: four main subfamilies (FadR, HutC, MocR and YtrA) and two minor subfamilies (AraC and PlmA) (Jain, 2015; Rigali *et al.*, 2002).

To date, only a few AraC-family members have been structurally characterized. X-ray structures are available for the MarA (Rhee *et al.*, 1998), Rob (Kwon *et al.*, 2000), ToxT (Lowden *et al.*, 2010) and AraC TRs. For the AraC regulator the structures of the regulatory (Soisson *et al.*, 1997) and DNA-binding (Rodgers & Schleif, 2009) domains were solved
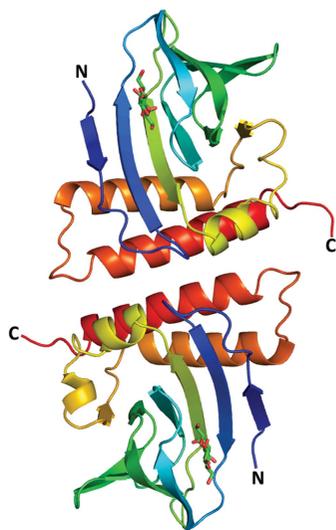
separately. TRs of the AraC subfamily are mainly involved in the regulation of carbon metabolism (AraC), stress response (SoxS, Rob and MarA) and the regulation of virulence gene expression [for example Rns from enterotoxigenic *Escherichia coli* (Munson & Scott, 1999), BfpT from enteropathogenic *E. coli* (Tobe *et al.*, 1996) and ExsA from *Pseudomonas aeruginosa* (Hovey & Frank, 1995)].

BgaR from *Clostridium perfringens* is a putative AraC-family TR with lactose (Hartman *et al.*, 2011; Caron & Trowell, 2018) and, less efficiently, lactulose as effectors (Caron & Trowell, 2018). Lactose is a commonly found disaccharide which consists of galactose and glucose moieties joined by a β-1→4 glycosidic link. Lactulose is a synthetic disaccharide that is formed from lactose by heat, and consists of linked galactose and fructose monosaccharides. Lactulose is an important product that is cited on the World Health Organization's list of *Essential medicines and health products* (https://www.who.int/medicines/news/2017/20th_essential_med-list/en/) and which has been used as a heat-treatment indicator in milk for over half a century (Adachi & Patton, 1961). Little is known about the BgaR protein or its interaction with its known effectors. It is of interest to uncover the structure of BgaR and to understand the ligand-binding mechanism of saccharides to this regulatory domain, which would allow its comparison with AraC and a more extensive characterization of this lactose biosensor. In turn, this would expand our understanding of this family of carbon-metabolism regulators and our understanding of the differences in binding affinity between these regulatory proteins. This would then allow the engineering of the protein to be a sensor for lactose in products that are to be labelled 'lactose-free'. This paper describes the purification, crystallization and structure determination by SAD phasing of the regulatory domain of BgaR bound to lactose and to lactulose, as well as a nominal apo form of the protein, in several different space groups. We then compare this new structure with the 'classical' TR regulatory binding domain of AraC.

2. Materials and methods

2.1. Cloning of BgaR$_{1-170}$-thrombin-His$_6$ into pRSET

The DNA sequence encoding the regulatory domain of BgaR (*C. perfringens* strain 13) was codon-optimized for expression in *E. coli* and synthesized commercially (GenScript, USA). A fragment, BgaR$_{1-170}$-thrombin-His$_6$, was amplified from the pUC-BgaR vector using the primers shown in Table 1. The BgaR$_{1-170}$-thrombin-His$_6$ fragment was cloned into the pRSET vector (BioLabs, Australia) using the NdeI and HindIII endonucleases. The integrity and orientation of the clone were verified by sequencing.

2.2. Expression and purification of BgaR$_{1-170}$

BgaR$_{1-170}$-thrombin-His$_6$ was expressed in *E. coli* BL21 (DE3) cells (New England Biolabs). 50 ml lysogeny broth (LB) supplemented with 2%(m/v) glucose and 100 µg ml^{-1} ampicillin was inoculated with a single colony and cultured at

Table 1
Macromolecule-production information.

Source organism	*C. perfringens* strain 13
DNA source	Synthetic
Forward primer	AAAAAACATATGCAGATTCTGTGGAAA
Reverse primer	TATATAAAGCTTTTAATGATGATGATGATG ATGGCTGCCGCGCGGCACCAGATGCAGT TCTTTATC
Cloning vector	pUC
Expression vector	pRSET
Expression host	*E. coli* BL21 (DE3)
Complete amino-acid sequence of the construct produced	MQILWKKYVKENFEMNVDECGIEQGIPGLG YNYEVLKNAVIHYVTKGYGTFKFNGKVY NLKQGDIFILLKGMQVEYVASIDDPWEY YWIGFSGSNANEYLNRTSITNSCVANCE ENSKIPQIILNMCEISKTYNPSRSDDIL LLKELYSLLYALIEEFPKPFEYKDKELH LVPRGSHHHHHH

37°C and 200 rev min^{-1} until it reached an Abs$_{600\,nm}$ of 0.8. The starter culture was used to inoculate 250 ml LB supplemented with 100 µg ml^{-1} ampicillin at an Abs$_{600\,nm}$ of 0.05, which was incubated at 28°C and 200 rev min^{-1} for 24 h. No addition of IPTG or any other induction compound was needed. The cells were harvested and were lysed using a homogenizer (Microfluidics M-110P, 137 MPa). The His$_6$-tagged protein was purified using TALON Superflow Metal Affinity Resin (Takara Clontech, Australia) according to the manufacturer's instructions [the crude lysate was applied onto the resin, washed with equilibration buffer (50 m*M* sodium phosphate, 300 m*M* sodium chloride pH 7.0), washed with 10 m*M* imidazole in equilibration buffer (pH 7.0) and eluted with 150 m*M* imidazole in equilibration buffer (pH 7.4)]. The eluted protein was dialysed against 50 m*M* bis-Tris pH 6.5, 50 m*M* sodium chloride buffer using a D-Tube Dialyzer (Merck; 6–8 kDa molecular-weight cutoff). The purity was assessed by SDS–PAGE (Supplementary Fig. S1) with an approximate size of 20 kDa. Protein concentrations were estimated by measurement of Abs$_{280\,nm}$. The protein was flash-frozen in 100 µl aliquots in thin-walled PCR tubes and stored at −80°C.

2.3. Crystallization

All crystallization trials were performed at the C3 centre (http://crystal.csiro.au) in SD-2 ('MRC') sitting-drop plates using a Phoenix (Art Robbins Instruments, USA) or a Mosquito (TTP Labtech, UK) robot. Experiments were stored and imaged using an RI1000 incubator system (Formulatrix, USA). Initial experiments (150 nl protein solution plus 150 nl reservoir solution equilibrated against 50 µl reservoir solution) were set up with the protein at 5 mg ml^{-1} in 50 m*M* sodium chloride, 50 m*M* bis-Tris pH 6.5 with 1 m*M* lactulose with or without thrombin (100 µl protein solution added to 10 µg thrombin). Screens tested included Shotgun (Fazio *et al.*, 2014) at both 20 and 8°C, as well as PACT and PSgradient both at 20°C. Detailed descriptions of these screens are available from the C6 website (http://c6.csiro.au; Newman *et al.*, 2010). Spherulites from the (thrombin) trials were optimized using a combination of fine screening, microseeding and the substitution of either trypsin or chymotrypsin for the thrombin

Table 2
Crystallization.

Method	Vapour diffusion, sitting drop
Plate type	SD-2
Temperature (K)	293
Protein concentration (mg ml^{-1})	5
Buffer composition of protein solution	50 mM bis-Tris pH 6.5, 50 mM NaCl
Composition of reservoir solution	See text
Volume and ratio of drop	200 nl + 200 nl drops
Volume of reservoir (µl)	50

protease. After refinement, crystals grew over the course of several days in many different conditions. Trials were set up with protein plus 1 mM lactose, protein plus 1 mM lactulose or protein with no added sugar.

Three different mercury compounds (mercuric cyanide, mersolyl acid and p-hydroxymercuribenzoic acid) were tested as phasing vehicles; in each case a 2 mM solution of the heavy atom in the reservoir mixture was added to the crystal and left to soak for 1–3 days. Crystallization information is given in Table 2.

2.4. Crystallography

Before harvesting, crystals grown in PEG 3350 were cryo-protected by adding glycerol to the reservoir to give a final concentration of 20% and then layering 1 µl of this solution over the crystal-containing drops. Crystals harvested from sodium malonate-containing conditions were cryoprotected by the addition of 3.4 M sodium malonate to the reservoir to a final concentration of 1.7 M and layering 1 µl of the modified reservoir solution over the drops before fishing out the crystals. Data were collected on the microfocus (MX2) beamline at the Australian Synchrotron using an EIGER 16M detector. For each crystal, a data set (360°) was collected over 36 s. For the mercury derivative, the energy of the X-ray beam was set to 12 300 eV (1.008 Å; *i.e.* close to the Hg peak) for SAD phasing and two data sets were collected from the same crystal, moving down the crystal for the second data set (see Table 3 for data-collection statistics). The monoclinic crystal (PDB entry 6nx3) had been soaked with sodium iodide and a data set was obtained with the wavelength set to 1.45865 Å (8500 eV) to obtain phase information from a possible iodine derivative (no anomalous signal was observed). The beam energy was set to 13 000 eV for the other data sets. The data were indexed using *XDS* (Kabsch, 2010) and scaled using *AIMLESS* (Evans & Murshudov, 2013); *CRANK*2 (Skubák & Pannu, 2013) was used to obtain the Hg sites, the initial phasing and the initial structure. The structure was rebuilt manually using *Coot* (Emsley *et al.*, 2010) and refined using *REFMAC* (Murshudov *et al.*, 2011). All subsequent structures were phased using *Phaser* (McCoy *et al.*, 2007) using PDB entry 6nwh as the starting model, manually rebuilt using *Coot* and refined using *REFMAC*.

Monoclinic crystals (space groups $I2$ and $P2_1$) were harvested from drops where the reservoir contained 20–22% PEG 3350, 200–220 mM MgCl$_2$ with either 100 mM bis-Tris pH 5.7 or no added buffer. Orthorhombic crystals (space group $P2_12_12$) were harvested from drops where the reservoir

contained 1.25–1.33 M sodium malonate in a variety of buffers [100 mM citrate pH 5.5 (PDB entry 6nwh), 100 mM HEPES pH 7.5 (PDB entry 6nwm) or 100 mM glycylglycine pH 8.2 (PDB entry 6nwj)]. Crystals grown in sodium malonate and citrate buffer and soaked with mercuric cyanide were used to collect the derivative data set used in phasing.

2.5. Mass spectrometry

Samples of BgaR$_{1-170}$-thrombin-His$_6$ treated with thrombin, trypsin or chymotrypsin, as well as the untreated protein, were analysed by mass spectrometry. For accurate mass determination, protein samples were spiked with formic acid (FA) to a final concentration of 0.1%(v/v) and separated by reverse-phase liquid chromatography on an UltiMate 3000 RSLCnano system (Thermo Fisher Scientific) fitted with a 50 × 4.6 mm, 5 µm particle-size, 300 Å pore-size PLRP-S column (Agilent). Proteins were eluted at a flow rate of 250 µl min^{-1} by applying a linear 30 min gradient from 0 to 80% solvent B [mobile phase A, 0.1%(v/v) FA; mobile phase B, 90%(v/v) acetonitrile/0.1%(v/v) FA] and were ionized using an Apollo II electro-spray ion source coupled to a micrOTOF-Q II mass spectrometer (Bruker). The native protein appeared as multiple species in the analysis, corresponding to various C-terminal truncations of the full-length construct (see Supplementary Fig. S2). The thrombin-treated sample was very clean by comparison, showing a single peak corresponding to cleavage after the arginine in the thrombin cleavage site. Both the trypsin-treated and the chymotrypsin-treated samples gave results consistent with cleavage close to the C-terminus: after Lys165 and Tyr164, respectively. Chymotrypsin treatment gave two equal-sized peaks: one consistent with cleavage after Tyr164 and one that was 1091 mass units smaller (data not shown).

Native mass spectra were acquired using a maXis II mass spectrometer (Bruker Daltonics) equipped with a standard ESI source in positive-ion mode. Protein samples were exchanged into 200 mM ammonium acetate buffer using Zeba Spin desalting columns (Thermo Fisher Scientific), and native MS was performed at a flow rate of 3 µl min^{-1} using a direct infusion mode. To guarantee the ionization effect of the sample solution, the capillary voltage and the end-plate offset were maintained at 3200 and −500 V, respectively. The temperature and the flow rate of the drying gas were maintained at 120°C and 3 l min^{-1}, respectively. The nebulizer gas pressure was 0.4 bar. To achieve a favourable desolvation effect and increase the sensitivity, the isCID and collision energies were set to 100 and 7 eV, respectively. Quadrupole frequencies, RF amplitudes and transfer times were adjusted to achieve the best ion-transmission efficiency. Native MS data were acquired at 1 Hz and were processed using parsimonious charge deconvolution (Bern *et al.*, 2018).

2.6. Dimerization of the regulatory domain

The protein was diluted to 0.2 mg ml^{-1} in 50 mM bis-Tris pH 6.5, 50 mM NaCl for dynamic light-scattering experiments. A Wyatt DynaPro DLS plate reader was used for data

Table 3
Data collection and refinement.

Values in parentheses are for the outer shell.

Structure	Lactose-bound	Lactose-bound	Lactulose-bound	Partial apo	Mercury derivative
PDB code	6nwj	6nx3	6nwm	6nwo	6nwh
Data collection					
Space group	$P22_12_1$	$P2_1$	$P22_12_1$	$I2$	$P22_12_1$
a, b, c (Å)	46.4, 87.0, 119.4	99.2, 46.3, 118.9	46.4, 87.5, 119.0	68.4, 40.5, 261.7	46.2, 86.6, 119.8
α, β, γ (°)	90, 90, 90	90, 100.7, 90	90, 90, 90	90, 90.3, 90	90, 90, 90
Wavelength (Å)	0.953657	1.45865	0.953723	0.953723	1.00802
Resolution (Å)	46.4–2.16	48.3–1.87	46.4–1.94	43.6–2.11	46.2–2.03
Completeness (%)	99.0 (92.6)	99.5 (97.2)	99.8 (99.4)	98.7 (91.7)	99.4 (92.5)
R_{merge} (%)	0.208 (1.167)	0.087 (0.783)	0.165 (1.776)	0.103 (0.672)	0.196 (0.749)
$R_{p.i.m.}$ (%)	0.060 (0.335)	0.037 (0.352)	0.048 (0.498)	0.043 (0.290)	0.040 (0.230)
Mean $I/\sigma(I)$	11.6 (3.5)	9.4 (2.4)	9.4 (1.8)	9.4 (2.7)	14.0 (4.2)
No. of unique reflections	26514	87904	36663	41261	31671
Multiplicity	13.0 (12.7)	6.6 (5.9)	13.1 (13.5)	6.7 (6.3)	24.2 (11.3)
$CC_{1/2}$	0.996 (0.827)	0.998 (0.922)	0.998 (0.761)	0.997 (0.744)	0.997 (0.859)
Anomalous completeness (%)					99.1 (89.9)
Anomalous multiplicity					12.8 (5.9)
No. of Hg atoms					9
Refinement					
Resolution (Å)	43.5–2.16	46.9–1.87	43.8–1.94	43.6–2.11	46.2–2.03
No. of reflections	25218	83468	34833	39182	30030
R_{work} (%)	18.8	22.4	20.4	19.4	17.9
R_{free} (%)	21.8	25.4	23.3	23.1	20.6
No. of atoms					
Total	2837	8514	2869	5276	3003
Sugar	46	138	46	23	46
No. of waters	138	499	133	148	211
Wilson B value (Å^2)	28.2	28.3	34.3	40.0	20.9
Mean B values (Å^2)					
Protein	30.3	30.5	37.2	43.9	20.3
Waters	33.2	34.8	40.5	41.3	28.9
Sugars	26.2	27.0	46.9	45.2	16.8
R.m.s.d.s					
Bond lengths (Å^2)	0.008	0.008	0.008	0.008	0.009
Bond angles (°)	1.484	1.470	1.474	1.427	1.451
Ramachandran analysis (%)					
Preferred	98.4	99.0	98.1	98.7	99.1
Allowed	1.6	1.0	1.9	1.3	0.9
Outliers	0	0	0	0	0

collection, which was performed at 25°C with 20 × 5 s readings. The average size of the particles was 2.6–2.7 nm, giving an estimated molecular weight of 32–35 kDa (data not shown). Additionally, 400 µl of protein solution (about 2 mg) was subjected to size-exclusion chromatography using a Superdex 200 column (GE Healthcare) in phosphate-buffered saline solution (9 mM sodium potassium phosphate pH 7.4 with 137 mM NaCl and 3 mM KCl) with the protein eluting at 16.5 ml, close to the standard chicken ovalbumin (44 kDa), which elutes at 15.9 ml (see Supplementary Fig. S3). Using linear interpolation, this gives a molecular weight of 36–37 kDa for the BgaR peak.

3. Results and discussion

The full-length BgaR protein yielded crystals, but these did not diffract beyond 7–8 Å resolution. The shorter regulatory (ligand-binding) domain construct (BgaR$_{1-170}$-thrombin-His$_6$) only gave crystals when treated with a protease, which resulted in the removal of the C-terminal tag and five or six residues from the C-terminus. Thrombin, chymotrypsin and trypsin were all tried in *in situ* proteolysis experiments. Thrombin treatment, which cleaved off only the tag, gave drops that contained promising spherulites but that would occasionally produce single crystals with seeding. Trypsin and chymotrypsin both removed the tag and a small number of C-terminal residues and reliably gave single crystals, although trypsin treatment generally produced better diffracting crystals, probably as the *in situ* trypsin treatment produced essentially a single product, whereas chymotrypsin treatment gave two major products according to the mass-spectrometric analysis. Mass spectrometry showed a mass of 20 312 Da for monomeric BgaR under denaturing conditions and also showed that the protein flew as a 40 624 Da dimer under native conditions (see Supplementary Fig. S2). The most complete structures showed density from residue 1 through to residue 162, and no density was seen for residues 163–166.

The dimer seems to be the prevalent form in solution based on size-exclusion chromatography (see Supplementary Fig. S3), dynamic light scattering (see Section 2) and crystallography (see below), as well as native-state mass-spectrometric data (Supplementary Fig. S2). The BgaR dimer interface is dominated by a long C-terminal helix which forms the majority of the interactions, having both hydrophilic and

hydrophobic components (Fig. 1). The hydrophobic residues include Ile141, Leu142, Leu144, Tyr99, Tyr148 and Tyr152, whereas the hydrophilic residues include Glu14, Asn12, Asn95, Arg102 and Asp140. Lys145 also sits at the interface, but it is the aliphatic chain of this residue which is presented to its neighbour instead of the amino head, which forms a same-chain bond with Glu146. Analysis using *PISA* (Krissinel & Henrick, 2007) suggests that the protein forms a dimer in solution with almost 20% of the surface area buried in the interface (2940 Å^2 buried out of a total of 15 380 Å^2). This dimer is also seen in the two monoclinic crystal forms and in the orthorhombic crystals over a broad range of pH values (5.5–8.2).

The most similar protein in the PDB is the regulatory domain of AraC (Soisson *et al.*, 1997), which shares 17% sequence identity with that of BgaR, and the best AraC/BgaR structure alignment gives an r.m.s.d. of 2.2 Å over 133 residues (out of 161/162). The available AraC structures did not provide useful solutions in molecular-replacement trials, even though the overall fold is quite similar: a jelly roll of eight (or nine) β-strands with C-terminal helices (Fig. 1), where the C-terminal helix forms the dimer interface. The crystal structures of BgaR with added saccharides gave models in which there was clear density from residue 1 through to residue 161 or 162. Although extensive hydrogen-bond networks and stacking interactions are made with the
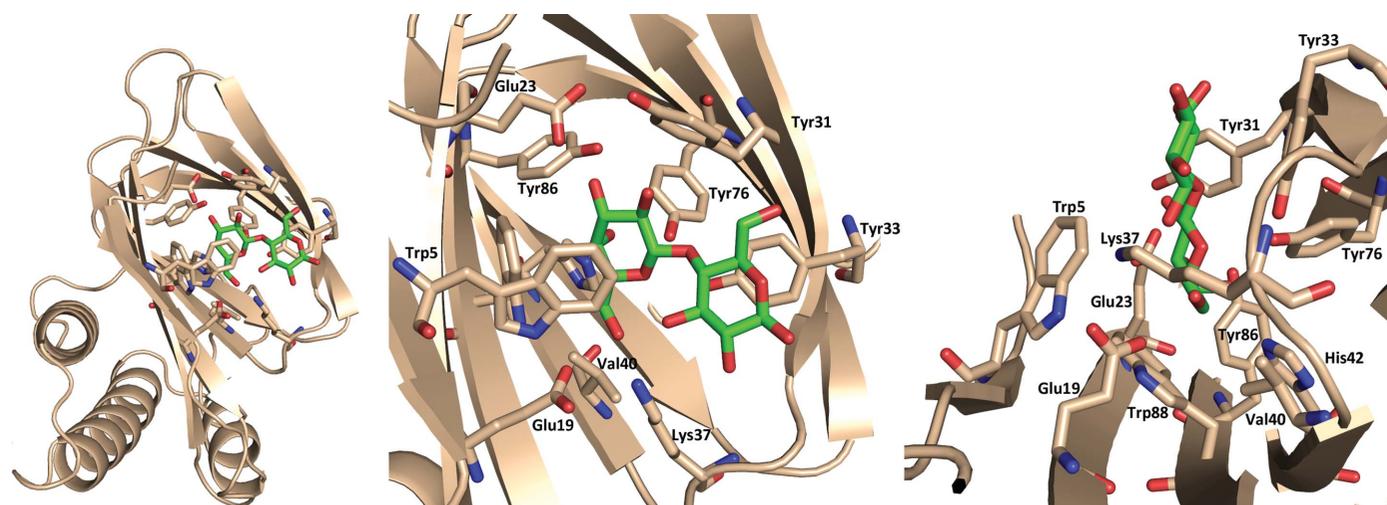

Figure 1
The BgaR dimer bound to lactose. The image on the left depicts the dimer as a cartoon coloured as a Jones' rainbow from the N-terminus to the C-terminus. Lactose is shown as a stick model with C atoms in green. The image on the right is an approximate 90° rotation and the colours have been changed to highlight the secondary structure (β-strands in magenta and α-helices in cyan). Lactose is shown as a stick model with C atoms in green. The dimer is held together by the long C-terminal helices and the saccharide-binding site is between the β-sheets of the jelly-roll fold.

Figure 2
Monomer of BgaR bound to lactose: detailed view. The image on the left shows lactose in a stick representation with green C atoms within a monomer of BgaR. The residues involved in binding to the lactose are also shown in stick representation. The middle image is an enlarged view highlighting the individual residues (labelled) and the image on the right is rotated to show that although one half of the saccharide (galactose) is deeply engaged in the binding site, the other half (the glucose ring) is mostly exposed to solvent.

saccharide (Fig. 2), part of the lactose (or lactulose) moiety is open to bulk solvent, with only the common galactose ring being sequestered significantly for both sugars.

The saccharide-free crystals showed some disorder in the first ten residues, and these residues are not modelled in two of the four protomers. Although no lactose/lactulose was added to the protein for the apo crystals, there is some density in one

of the binding sites (four protomers are seen in the asymmetric unit). Native mass spectrometry detected a +342.2 Da mass addition consistent with lactose binding to the BgaR dimer (Supplementary Fig. S2), and we have therefore modelled this density as lactose. In this case the galactose ring has clear density, with the glucose ring having clear but weaker density. No obvious saccharide density is seen in two of the other four binding sites and some weak density for a glycerol molecule occupies one side of the binding site in the last protomer. There is some unmodelled density in one of the two 'empty' saccharide-binding sites of the apo structure. Comparing the saccharide-free structure with that with lactose bound shows an overall r.m.s.d. of 0.3–0.4 Å for C^α atoms (151 residues aligned), with the largest changes in the binding site being a different rotamer position for Trp88 and shifts of 0.5–1.0 Å in the C^α positions of Tyr31 (depending on the protomer compared). As above, the first ten residues of the saccharide-free protomers are not modelled, including Trp5, which is also part of the binding site.

Comparing the structures of the lactose-bound, the lactulose-bound and the partial apo structure and the different space groups allow us to more fully
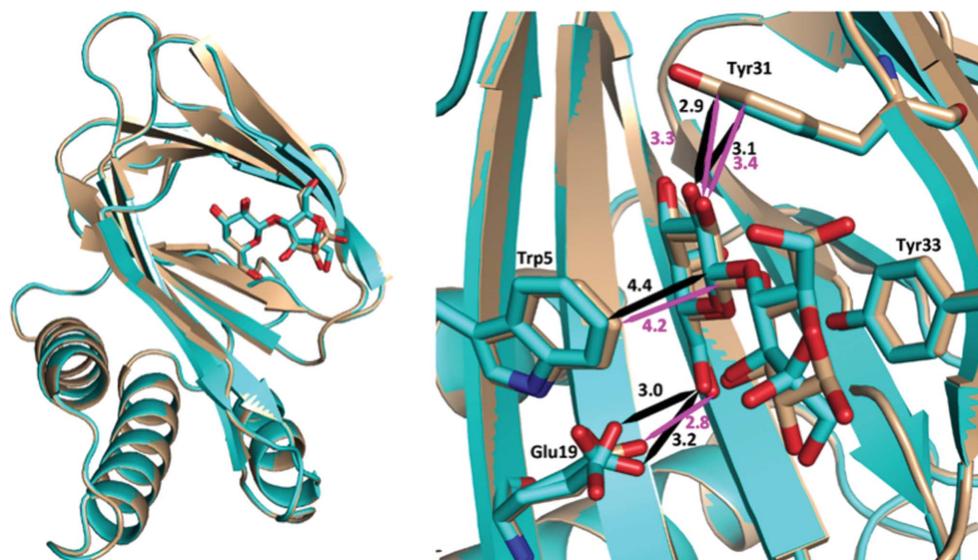

Figure 3
Lactose and lactulose binding in BgaR. The image on the left shows a superposition of the monomers of the lactose-bound and lactulose-bound BgaR structures, showing that the saccharides bind in the same pocket in a very similar way; the lactulose-bound structure is shown in cyan and the lactose-bound structure is in wheat. The image on the right is an enlarged perspective with distances shown as black arrows/type for the lactulose-bound structure and in magenta for the lactose-bound structure. A slight shift in the galactose moiety towards the viewer is seen in the lactose-bound structure. Slight rearrangements in the glucose/fructose ring cause a galactose hydroxyl to move into close proximity to the aromatic C atoms of Tyr31, cause another hydroxyl to make a weaker hydrogen bond to Glu19 and reduce the stacking interactions of the saccharide between Trp5 and Tyr33.
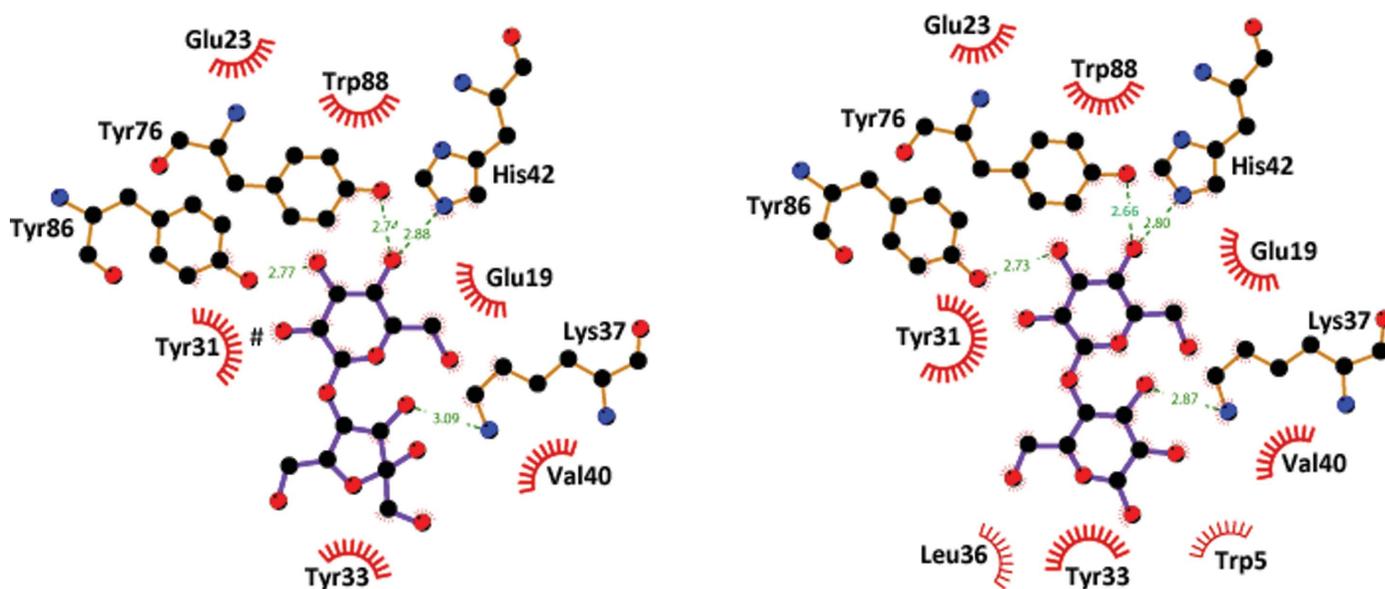

Figure 4
LigPlot+ representation of lactose and lactulose binding in BgaR. Lactulose binding is shown on the left and lactose binding is shown on the right. The # mark in the lactulose-binding figure indicates a steric clash between one of the galactose O atoms and Tyr31. The hydrogen bonds made by lactose are shorter/stronger than those made by lactulose. This figure was produced using LigPlot+ (Laskowski & Swindells, 2011).

understand the saccharide-binding characteristics of this regulatory domain. From previous work, we know that lactose binds to the BgaR protein about 200 times more tightly than lactulose, glucose or galactose, with lactulose being the next best binder (Caron & Trowell, 2018). BgaR binds lactose and lactulose through the same mechanism, and the same amino acids are involved (Figs. 3 and 4). Although many of the hydrogen bonds are shared between the lactose-bound and lactulose-bound structures, on average the bonds are slightly

longer in the lactulose-bound structure. As seen in Fig. 4, two residues make contact with lactose (Leu36 and Trp5) which do not make contact with lactulose (according to *LigPlot+* with default parameters; Laskowski & Swindells, 2011). Additionally, there are some stacking interactions of the saccharide with the protein, and these also seem to be weaker (further apart) in the lactulose-bound structure than in the lactose-bound structures (for example stacking with Tyr33). There are some close contacts with the aromatic C atoms of Tyr31 and
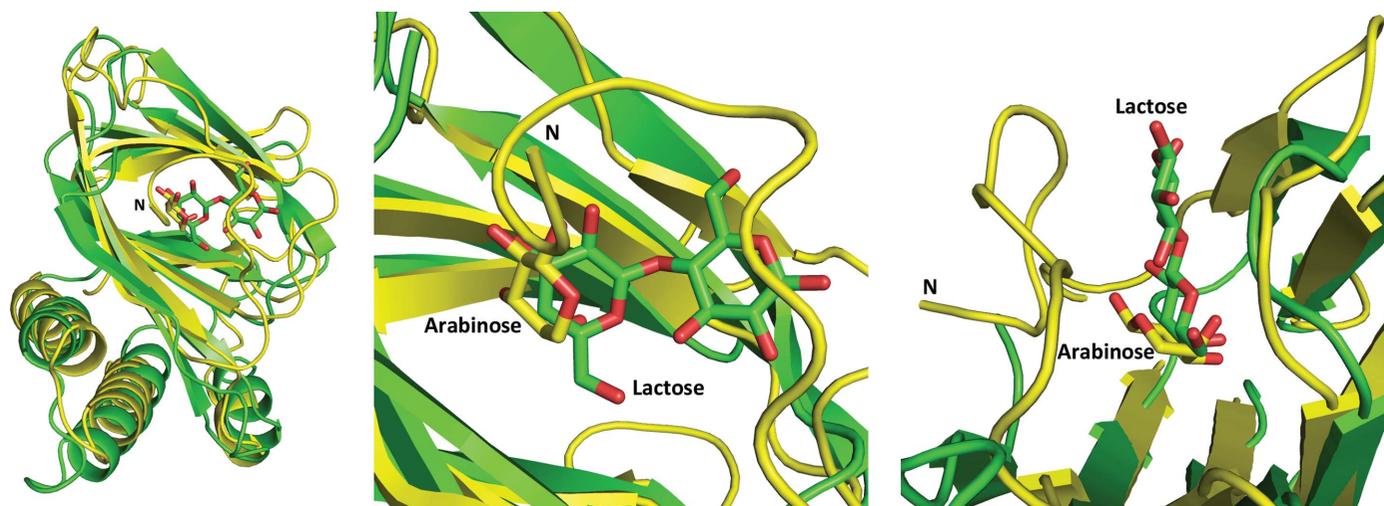

Figure 5
Comparison of lactose binding to BgaR and arabinose binding to AraC. The left image shows a comparison of the AraC structure with arabinose bound with the BgaR structure with lactose bound, showing that the overall folds are quite similar between the proteins and that the binding sites are both found between the β-sheets of the jelly roll; the AraC structure is in yellow and the BgaR structure is in green. The middle and right images are enlarged perspectives of the left image to show in detail how the arabinose sits at almost 90° in the pocket compared with the lactose moiety. The N-terminus of AraC does not participate in the β-sheets formed but instead comes up to enclose the binding site for the saccharide, unlike the BgaR structure. The N-terminus is labelled in each case.
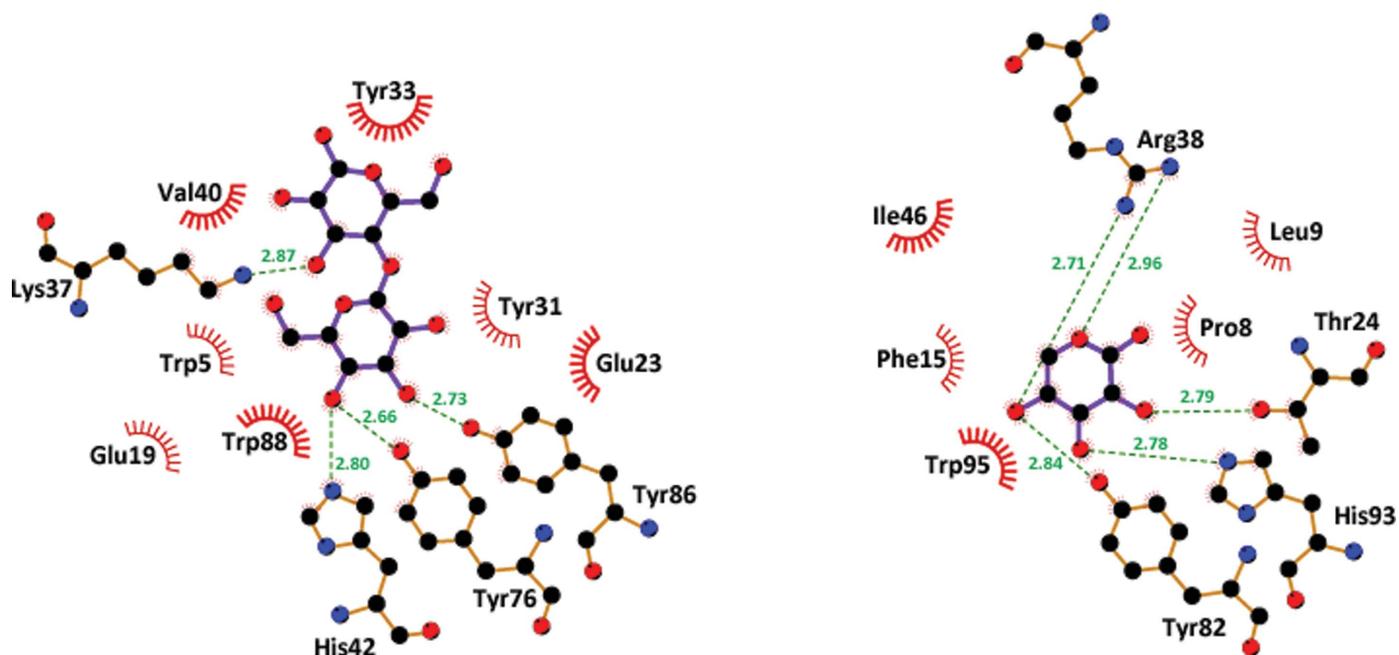

Figure 6
LigPlot+ representation of lactose binding to BgaR and arabinose binding to AraC. Lactose bound to BgaR is shown on the left and arabinose bound to AraC is shown on the right. In both cases, specific hydrogen bonds are made to the saccharides, as well as hydrophobic and stacking interactions. The residues at the N-terminus are involved in binding; when no saccharide is present the binding site opens and the N-terminus becomes disordered.

one of the hydroxyl O atoms on the galactose ring of lactulose which are not seen in the lactose-bound structure (2.9–3.1 Å instead of 3.2–3.4 Å). This shift between the saccharides is subtle, changing hydrogen-bond distances, stacking distances and van der Waals interactions, which in total cause the lower binding affinity of lactulose compared with lactose.

The closest structures from the PDB are those of AraC. The overall fold is quite similar, with the major difference being at the N-terminus, where the AraC structure more effectively covers the saccharide-binding site with a large loop (Figs. 5 and 6). The N-terminus of the BgaR structure instead forms an extra strand in the β-sheet which passes through the middle of the protein, and this creates a more open saccharide-binding site. In both proteins there is a long C-terminal helix which forms the majority of the dimerization face, which has both hydrophilic and hydrophobic components. The saccharide is oriented differently in these two proteins. The arabinose is almost perpendicular to the galactose moiety of the lactose in the binding site of AraC when the two structures are superposed (Fig. 5). The arabinose is effectively sequestered from the bulk solvent by the AraC protein, whereas the lactose is only covered effectively on one side (the galactose moiety) and the glucose (or fructose in the case of lactulose) moiety is partially free to interact with the bulk solvent in the BgaR structures.

4. Conclusion

Transcriptional regulator-based biosensors have been developed and used broadly for decades owing to their inherent function in triggering a reporter signal in the presence of metabolites, nutrients and other solutes. However, whole-cell detection-system applications are limited to research settings owing to their slow response times and their requirements for skilled staff and specialized equipment. A few examples of modular TR-based biosensors have been reported in which the analyte-induced conformational change in the TR was captured by resonance energy transfer (RET). Caron and Trowell designed and built a lactose biosensor in which the BgaR protein was flanked with RET reporter proteins, namely GFP and the *Renilla* luciferase variant RLuc8. BgaR had the highest affinity for lactose ($EC_{50} = 12 \pm 1 \ \mu M$) and a 200-fold lower affinity for the analogue lactulose ($EC_{50} = 2.4 \pm 0.2 \ mM$) (Caron & Trowell, 2018).

The crystal structures presented here show that lactose and lactulose bind in similar ways to the same residues in BgaR, but subtle differences in the binding modalities lead to different binding efficiencies. The higher affinity observed for lactose compared with lactulose may in part be owing to the short distances of Tyr31 to one of the hydroxyl atoms on the galactose ring of lactulose, but the stronger hydrogen bonds (*e.g.* Glu19), better stacking and additional residues making contact with lactose almost certainly contribute to stronger binding as well. We have shown using crystallography and

native-state mass spectrometry that despite having little sequence similarity, the structure and dimer of BgaR are quite similar to those of AraC. Both proteins are transcriptional regulators that control the carbon metabolism of different saccharides and thus provide similar functions in their respective organisms. The current structure allows more rational engineering of this protein for future biosensor applications.

References

Adachi, S. & Patton, S. (1961). *J. Dairy Sci.* **44**, 1375–1393.
Bern, M., Caval, T., Kil, Y. J., Tang, W., Becker, C., Carlson, E., Kletter, D., Sen, K. I., Galy, N., Hagemans, D., Franc, V. & Heck, A. J. R. (2018). *J. Proteome Res.* **17**, 1216–1226.
Caron, K. & Trowell, S. C. (2018). *Anal. Chem.* **90**, 12986–12993.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.
Fazio, V. J., Peat, T. S. & Newman, J. (2014). *Acta Cryst.* F**70**, 1303–1311.
Hartman, A. H., Liu, H. & Melville, S. B. (2011). *Appl. Environ. Microbiol.* **77**, 471–478.
Hovey, A. K. & Frank, D. W. (1995). *J. Bacteriol.* **177**, 4427–4436.
Jain, D. (2015). *IUBMB Life*, **67**, 556–563.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Kwon, H. J., Bennik, M. H. J., Demple, B. & Ellenberger, T. (2000). *Nature Struct. Biol.* **7**, 424–430.
Laskowski, R. A. & Swindells, M. B. (2011). *J. Chem. Inf. Model.* **51**, 2778–2786.
Lowden, M. J., Skorupski, K., Pellegrini, M., Chiorazzo, M. G., Taylor, R. K. & Kull, F. J. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 2860–2865.
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.
Munson, G. P. & Scott, J. R. (1999). *J. Bacteriol.* **181**, 2110–2117.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Newman, J., Fazio, V. J., Lawson, B. & Peat, T. S. (2010). *Cryst. Growth Des.* **10**, 2785–2792.
Rhee, S., Martin, R. G., Rosner, J. L. & Davies, D. R. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 10413–10418.
Rigali, S., Derouaux, A., Giannotta, F. & Dusart, J. (2002). *J. Biol. Chem.* **277**, 12507–12515.
Rodgers, M. E. & Schleif, R. (2009). *Proteins*, **77**, 202–208.
Skubák, P. & Pannu, N. S. (2013). *Nature Commun.* **4**, 2777.
Soisson, S. M., MacDougall-Shackleton, B., Schleif, R. & Wolberger, C. (1997). *Science*, **276**, 421–425.
Tobe, T., Schoolnik, G. K., Sohel, I., Bustamante, V. H. & Puente, J. L. (1996). *Mol. Microbiol.* **21**, 963–975.