



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 13, 2017 – 02:16 pm GMT

PDB ID : 2BVM  
Title : CRYSTAL STRUCTURE OF THE CATALYTIC DOMAIN OF TOXIN B FROM CLOSTRIDIUM DIFFICILE IN COMPLEX WITH UDP, GLC AND MANGANESE ION  
Authors : Reinert, D.J.; Jank, T.; Aktories, K.; Schulz, G.E.  
Deposited on : 2005-06-30  
Resolution : 2.55 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

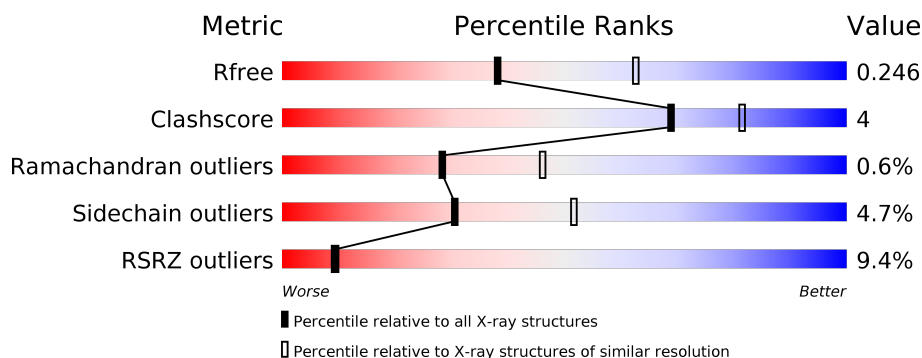
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.55 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	100719	4993 (2.58-2.50)
Clashscore	112137	5755 (2.58-2.50)
Ramachandran outliers	110173	5652 (2.58-2.50)
Sidechain outliers	110143	5654 (2.58-2.50)
RSRZ outliers	101464	5026 (2.58-2.50)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	542	<div> <div>9%</div> <div>87%</div> <div>11%</div> </div>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 4593 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called TOXIN B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	542	Total	C	N	O	S	0	0	1
			4421	2801	719	883	18			

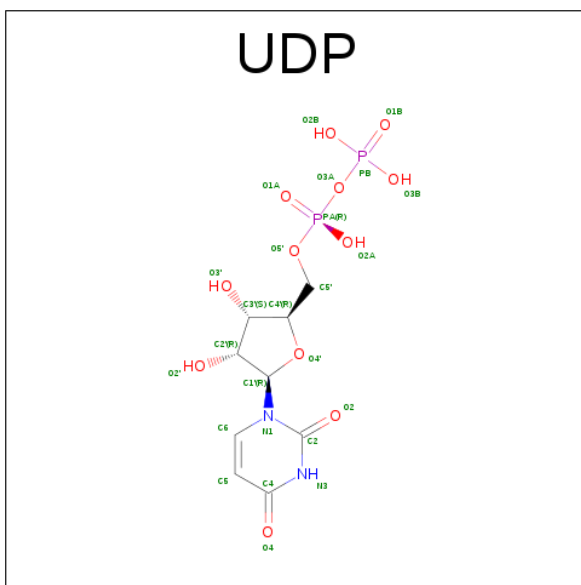
There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	59	CYS	TYR	ENGINEERED MUTATION	UNP P18177
A	244	GLY	GLU	ENGINEERED MUTATION	UNP P18177
A	542	ALA	SER	ENGINEERED MUTATION	UNP P18177

- Molecule 2 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

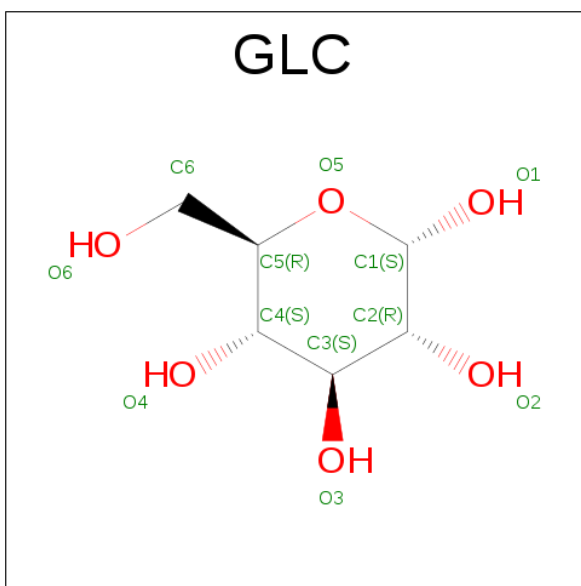
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Mn	0	0
			1	1		

- Molecule 3 is URIDINE-5'-DIPHOSPHATE (three-letter code: UDP) (formula: C<sub>9</sub>H<sub>14</sub>N<sub>2</sub>O<sub>12</sub>P<sub>2</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	N	O	P	0	0
			25	9	2	12	2		

- Molecule 4 is SUGAR (ALPHA-D-GLUCOSE) (three-letter code: GLC) (formula:  $C_6H_{12}O_6$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			12	6	6		

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula:  $O_4S$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	O	S	0	0
			5	4	1		

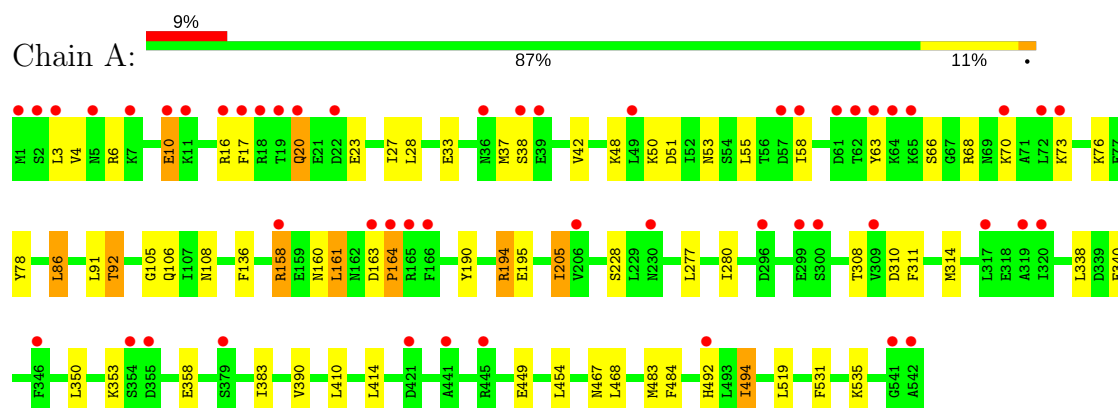
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	129	Total	O	0	0
			129	129		

### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

#### • Molecule 1: TOXIN B



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	65.78Å 97.30Å 114.45Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.55 19.99 – 2.55	Depositor EDS
% Data completeness (in resolution range)	100.0 (20.00-2.55) 100.0 (19.99-2.55)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.94 (at 2.56Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.199 , 0.245 0.200 , 0.246	Depositor DCC
$R_{free}$ test set	769 reflections (3.24%)	DCC
Wilson B-factor (Å <sup>2</sup> )	51.8	Xtriage
Anisotropy	0.043	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 44.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	4593	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.65% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: UDP, GLC, SO4, MN

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\# Z  > 5$	RMSZ	$\# Z  > 5$
1	A	0.68	7/4501 (0.2%)	0.67	5/6073 (0.1%)

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	10	GLU	CG-CD	19.53	1.81	1.51
1	A	10	GLU	CD-OE2	13.10	1.40	1.25
1	A	6	ARG	CZ-NH1	8.93	1.44	1.33
1	A	70	LYS	CE-NZ	7.25	1.67	1.49
1	A	10	GLU	C-O	6.09	1.34	1.23
1	A	70	LYS	CD-CE	5.56	1.65	1.51
1	A	20	GLN	CD-OE1	5.30	1.35	1.24

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	10	GLU	OE1-CD-OE2	10.71	136.15	123.30
1	A	6	ARG	NE-CZ-NH2	-9.42	115.59	120.30
1	A	10	GLU	CG-CD-OE2	-5.37	107.57	118.30
1	A	70	LYS	CD-CE-NZ	-5.36	99.36	111.70
1	A	6	ARG	NE-CZ-NH1	5.11	122.85	120.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen



atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	4421	0	4304	38	0
2	A	1	0	0	0	0
3	A	25	0	11	0	0
4	A	12	0	12	0	0
5	A	5	0	0	0	0
6	A	129	0	0	1	0
All	All	4593	0	4327	38	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 4.

All (38) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:10:GLU:CD	1:A:10:GLU:CG	1.81	1.48
1:A:194:ARG:HH11	1:A:194:ARG:HG3	1.11	1.12
1:A:449:GLU:HG2	1:A:467:ASN:ND2	1.82	0.94
1:A:163:ASP:HB2	1:A:164:PRO:HA	1.65	0.79
1:A:194:ARG:NH1	1:A:194:ARG:HG3	1.91	0.79
1:A:10:GLU:CD	1:A:10:GLU:CB	2.60	0.70
1:A:194:ARG:HH11	1:A:194:ARG:CG	1.98	0.70
1:A:483:MET:SD	1:A:492:HIS:HE1	2.20	0.64
1:A:3:LEU:HD13	1:A:78:TYR:O	1.99	0.63
1:A:449:GLU:HG2	1:A:467:ASN:CG	2.23	0.59
1:A:16:ARG:HH21	1:A:17:PHE:HZ	1.49	0.59
1:A:410:LEU:HD22	1:A:468:LEU:HD21	1.85	0.58
1:A:308:THR:HB	1:A:311:PHE:H	1.68	0.56
1:A:163:ASP:CB	1:A:164:PRO:HA	2.32	0.54
1:A:42:VAL:HG13	1:A:86:LEU:HD23	1.91	0.53
1:A:20:GLN:HA	1:A:68:ARG:NH2	2.25	0.52
1:A:163:ASP:HB2	1:A:164:PRO:CA	2.38	0.52
1:A:50:LYS:HA	1:A:484:PHE:CD1	2.45	0.52
1:A:414:LEU:HD11	1:A:454:LEU:HD21	1.93	0.50
1:A:353:LYS:HD2	1:A:358:GLU:HB3	1.92	0.50
1:A:92:THR:HG23	1:A:390:VAL:HG11	1.94	0.49
1:A:308:THR:HG22	1:A:310:ASP:H	1.76	0.49
1:A:105:GLY:HA2	1:A:136:PHE:O	2.13	0.48
1:A:53:ASN:OD1	1:A:76:LYS:HE3	2.13	0.48
1:A:33:GLU:O	1:A:37:MET:HG3	2.14	0.47

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:531:PHE:CZ	1:A:535:LYS:HD2	2.48	0.47
1:A:63:TYR:HB3	1:A:66:SER:HB2	1.98	0.46
1:A:20:GLN:NE2	6:A:2002:HOH:O	2.49	0.45
1:A:106:GLN:HG3	1:A:228:SER:HA	1.99	0.44
1:A:494:ILE:H	1:A:494:ILE:HG13	1.64	0.43
1:A:205:ILE:HG12	1:A:205:ILE:H	1.61	0.42
1:A:33:GLU:HG3	1:A:48:LYS:HE2	2.01	0.42
1:A:27:ILE:HD13	1:A:55:LEU:HB3	2.02	0.42
1:A:23:GLU:CD	1:A:23:GLU:H	2.23	0.41
1:A:194:ARG:NH1	1:A:194:ARG:CG	2.66	0.41
1:A:190:TYR:HA	1:A:205:ILE:HG21	2.02	0.41
1:A:449:GLU:CG	1:A:467:ASN:ND2	2.68	0.41
1:A:158:ARG:HA	1:A:161:LEU:HD13	2.02	0.41

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	540/542 (100%)	510 (94%)	27 (5%)	3 (1%)	28	46

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	158	ARG
1	A	164	PRO
1	A	4	VAL

### 5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	494/494 (100%)	471 (95%)	23 (5%)	30 52

All (23) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	28	LEU
1	A	38	SER
1	A	51	ASP
1	A	58	ILE
1	A	73	LYS
1	A	86	LEU
1	A	91	LEU
1	A	92	THR
1	A	108	ASN
1	A	160	ASN
1	A	161	LEU
1	A	194	ARG
1	A	195	GLU
1	A	205	ILE
1	A	277	LEU
1	A	280	ILE
1	A	314	MET
1	A	338	LEU
1	A	340	GLU
1	A	350	LEU
1	A	383	ILE
1	A	494	ILE
1	A	519	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (10) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	97	ASN
1	A	113	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	126	ASN
1	A	128	ASN
1	A	385	GLN
1	A	415	ASN
1	A	467	ASN
1	A	492	HIS
1	A	507	ASN
1	A	510	GLN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

### 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 1 is monoatomic - leaving 3 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
3	UDP	A	1543	2	21,26,26	1.89	5 (23%)	22,40,40	1.91	3 (13%)
4	GLC	A	1544	-	12,12,12	0.80	0	17,17,17	1.66	3 (17%)
5	SO4	A	1545	-	4,4,4	0.25	0	6,6,6	0.18	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral

centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	UDP	A	1543	2	-	0/12/32/32	0/2/2/2
4	GLC	A	1544	-	-	0/2/22/22	0/1/1/1
5	SO4	A	1545	-	-	0/0/0/0	0/0/0/0

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	1543	UDP	PB-O1B	2.84	1.60	1.50
3	A	1543	UDP	PB-O3A	2.87	1.64	1.60
3	A	1543	UDP	O4'-C1'	3.79	1.46	1.41
3	A	1543	UDP	C6-N1	3.85	1.41	1.35
3	A	1543	UDP	C4-N3	4.34	1.40	1.33

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1543	UDP	C4'-O4'-C1'	-3.41	106.14	109.77
3	A	1543	UDP	O4'-C1'-N1	2.48	113.06	108.08
4	A	1544	GLC	C1-O5-C5	3.01	118.82	113.39
4	A	1544	GLC	C1-C2-C3	3.11	116.27	110.65
4	A	1544	GLC	O5-C1-C2	4.16	116.94	110.04
3	A	1543	UDP	C4-N3-C2	6.76	119.94	114.13

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 6 Fit of model and data

### 6.1 Protein, DNA and RNA chains

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	542/542 (100%)	0.50	51 (9%) <b>9</b> <b>9</b>	36, 55, 66, 81	0

All (51) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	65	LYS	6.7
1	A	1	MET	6.4
1	A	58	ILE	5.2
1	A	16	ARG	5.1
1	A	17	PHE	4.9
1	A	11	LYS	4.6
1	A	163	ASP	4.4
1	A	354	SER	4.3
1	A	542	ALA	4.0
1	A	39	GLU	4.0
1	A	165	ARG	3.9
1	A	57	ASP	3.7
1	A	61	ASP	3.4
1	A	3	LEU	3.4
1	A	64	LYS	3.3
1	A	36	ASN	3.3
1	A	541	GLY	3.3
1	A	441	ALA	3.2
1	A	7	LYS	3.2
1	A	445	ARG	3.2
1	A	317	LEU	3.1
1	A	63	TYR	3.1
1	A	158	ARG	3.1
1	A	18	ARG	3.0
1	A	319	ALA	3.0
1	A	20	GLN	3.0
1	A	38	SER	2.9

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	300	SER	2.9
1	A	72	LEU	2.9
1	A	10	GLU	2.7
1	A	2	SER	2.7
1	A	166	PHE	2.7
1	A	19	THR	2.6
1	A	320	ILE	2.6
1	A	299	GLU	2.6
1	A	70	LYS	2.6
1	A	355	ASP	2.6
1	A	206	VAL	2.5
1	A	230	ASN	2.5
1	A	346	PHE	2.5
1	A	5	ASN	2.5
1	A	164	PRO	2.4
1	A	421	ASP	2.4
1	A	309	VAL	2.3
1	A	73	LYS	2.3
1	A	379	SER	2.2
1	A	62	THR	2.2
1	A	296	ASP	2.2
1	A	492	HIS	2.1
1	A	22	ASP	2.1
1	A	49	LEU	2.0

## 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

## 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
5	SO4	A	1545	5/5	0.96	0.16	1.15	62,63,63,64	0
4	GLC	A	1544	12/12	0.86	0.18	0.59	51,55,59,61	0
3	UDP	A	1543	25/25	0.98	0.07	-2.14	32,42,44,46	0
2	MN	A	1542	1/1	1.00	0.01	-4.02	40,40,40,40	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.