wwPDB X-ray Structure Validation Summary Report (i)

Sep 4, 2023 – 11:30 PM EDT

PDB ID	:	3TGO
Title	:	Crystal structure of the E. coli BamCD complex
Authors	:	Paetzel, M.; Kim, K.H.; Aulakh, S.
Deposited on	:	2011-08-17
Resolution	:	2.90 Å(reported)

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

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

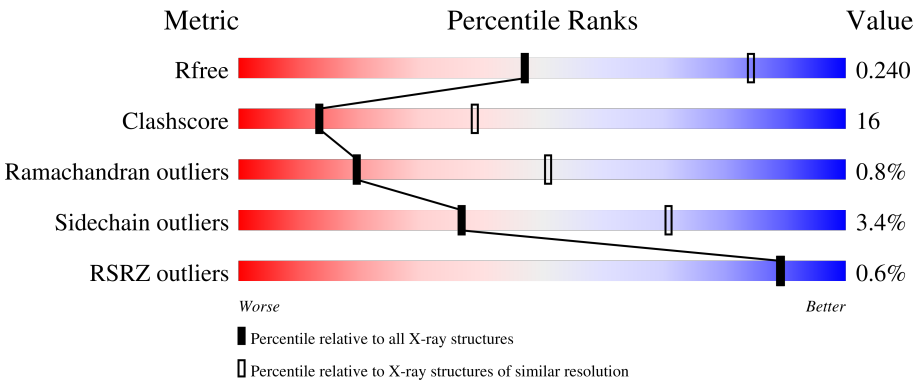
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.90 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}	130704	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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 of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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 <=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	229	69% 25% • 5%
1	B	229	71% 24% 5%
2	C	323	43% 12% • 44%
2	D	323	% 37% 20% • 42%

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 6466 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 UPF0169 lipoprotein yfiO.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	217	1754	1106	308	333	7	0	0	0
1	B	218	1761	1109	309	336	7	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	17	GLY	-	expression tag	UNP P0AC02
A	18	SER	-	expression tag	UNP P0AC02
A	19	HIS	-	expression tag	UNP P0AC02
A	20	MET	-	expression tag	UNP P0AC02
B	17	GLY	-	expression tag	UNP P0AC02
B	18	SER	-	expression tag	UNP P0AC02
B	19	HIS	-	expression tag	UNP P0AC02
B	20	MET	-	expression tag	UNP P0AC02

- Molecule 2 is a protein called Lipoprotein 34.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	182	1383	864	238	277	4	0	1	0
2	D	186	1412	881	244	283	4	0	1	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	22	GLY	-	expression tag	UNP P0A903
C	23	SER	-	expression tag	UNP P0A903
C	24	HIS	-	expression tag	UNP P0A903
C	25	MET	-	expression tag	UNP P0A903

Continued on next page...

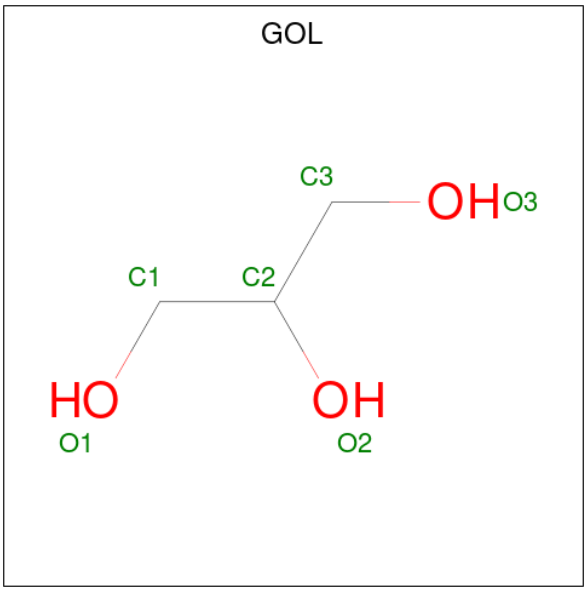
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	22	GLY	-	expression tag	UNP P0A903
D	23	SER	-	expression tag	UNP P0A903
D	24	HIS	-	expression tag	UNP P0A903
D	25	MET	-	expression tag	UNP P0A903

- Molecule 3 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	K		
3	A	1	1	1	0	0

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

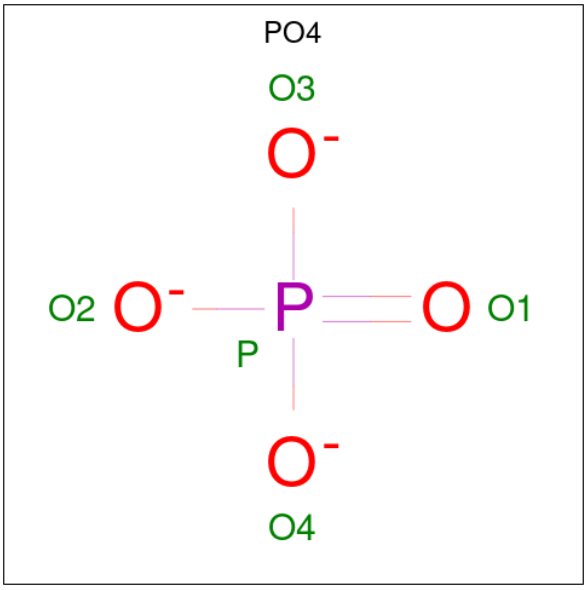


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
4	A	1	6	3	3	0	0
			Total	C	O		
4	B	1	6	3	3	0	0
			Total	C	O		
4	D	1	6	3	3	0	0
			Total	C	O		
4	D	1	6	3	3	0	0

- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	1	Total Cl 1 1	0	0

- Molecule 6 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	C	1	Total O P 5 4 1	0	0

- Molecule 7 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	D	1	Total Na 1 1	0	0

- Molecule 8 is water.

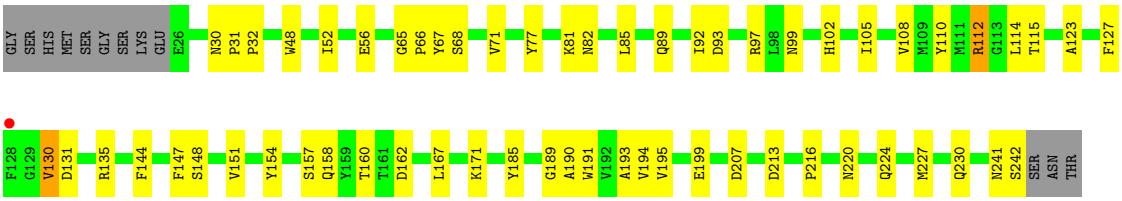
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	A	34	Total O 34 34	0	0
8	B	40	Total O 40 40	0	0
8	C	33	Total O 33 33	0	0
8	D	17	Total O 17 17	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

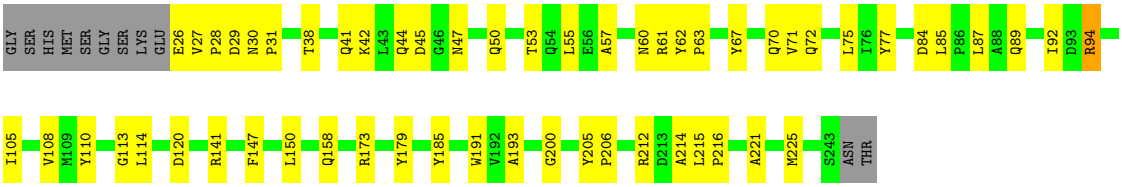
• Molecule 1: UPF0169 lipoprotein yfiO

Chain A: 69% 25% • 5%



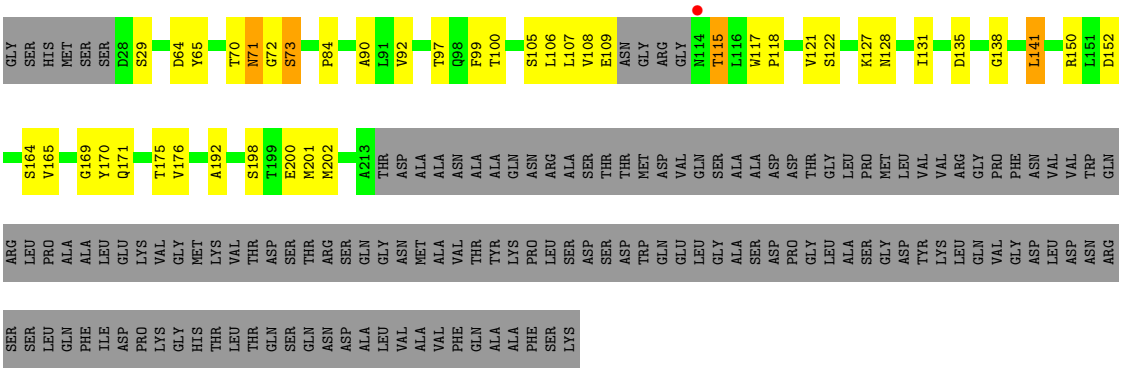
• Molecule 1: UPF0169 lipoprotein yfiO

Chain B: 71% 24% 5%



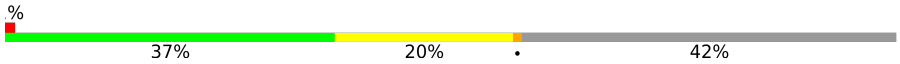
• Molecule 2: Lipoprotein 34

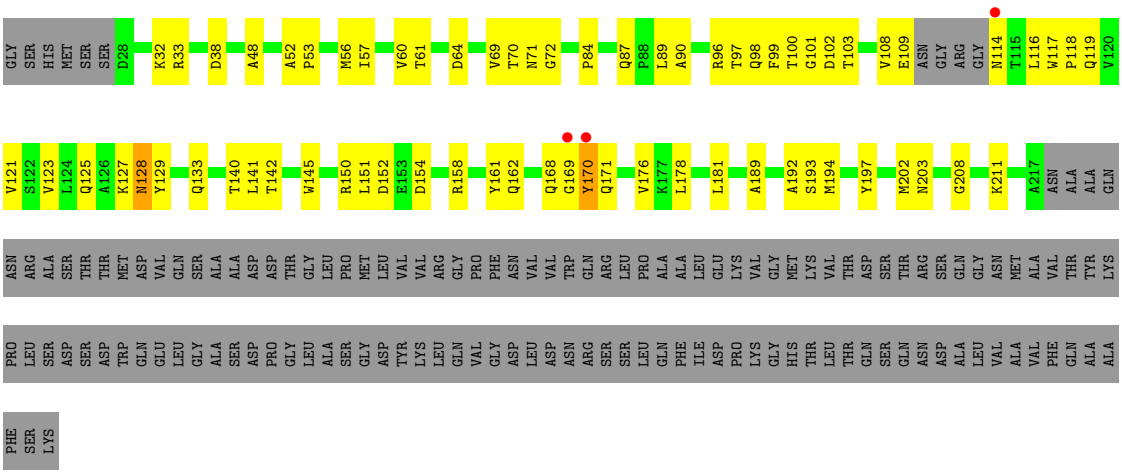
Chain C: 43% 12% • 44%



• Molecule 2: Lipoprotein 34

Chain D:


%
37%
20%
42%


GLY SER HIS MET SER SER D28 K32 R33 D38 A48 A52 P53 M56 I57 V60 T61 D64 V69 T70 N71 G72 P84 Q87 P88 L89 A90 R96 T97 Q98 F99 T100 G101 D102 T103 V108 E109 ASN GLY ARG GLY N114 T115 L116 W117 P118 Q119 V120
V121 S122 V123 L124 Q125 A126 K127 N128 Y129 Q133 T140 L141 T142 W145 R150 L151 D152 E153 D154 R158 Y161 Q162 Q168 G169 Y170 Q171 V176 K177 L178 L181 A189 A192 S193 M194 Y197 M202 N203 G208 K211 A217 ASN ALA ALA GLN
ASN ARG ALA SER THR THR MET ASP VAL GLN SER ALA ALA ASP ASP THR GLY LEU PRO MET LEU VAL VAL ARG GLY PRO PHE ASN VAL VAL TRP GLN ARG LEU PRO ALA ALA LEU GLU LYS VAL GLY MET LYS VAL THR ASP SER THR ARG SER GLN GLY ASN MET ALA VAL THR TYR LYS
PRO LEU SER ASP SER ASP TRP GLN GLU LEU GLY ALA SER ASP PRO GLY LEU ALA SER GLY ASP TYR LYS LEU GLN VAL GLY ASP LEU ASP ASN ARG SER SER LEU GLN PHE ILE ASP PRO LYS GLY HIS THR LEU THR GLN SER GLN ASN ASP ALA LEU VAL ALA VAL PHE GLN ALA ALA
PHE SER LYS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	73.78Å 133.38Å 144.99Å 90.00° 100.15° 90.00°	Depositor
Resolution (Å)	97.45 – 2.90 51.70 – 2.90	Depositor EDS
% Data completeness (in resolution range)	99.5 (97.45-2.90) 99.6 (51.70-2.90)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.73 (at 2.91Å)	Xtriage
Refinement program	REFMAC 5.5.0110	Depositor
R, R_{free}	0.177 , 0.243 0.177 , 0.240	Depositor DCC
R_{free} test set	1539 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	41.2	Xtriage
Anisotropy	0.231	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 50.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	6466	wwPDB-VP
Average B, all atoms (Å^2)	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.16% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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: GOL, K, NA, CL, PO4

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	0.70	0/1794	0.52	0/2438
1	B	0.67	0/1801	0.52	0/2447
2	C	0.70	0/1410	0.55	0/1923
2	D	0.75	3/1439 (0.2%)	0.52	0/1962
All	All	0.70	3/6444 (0.0%)	0.53	0/8770

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	60	VAL	CB-CG1	-5.96	1.40	1.52
2	D	60	VAL	CB-CG2	-5.50	1.41	1.52
2	D	129	TYR	CE2-CZ	-5.13	1.31	1.38

There are no bond angle outliers.

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	1754	0	1694	44	0
1	B	1761	0	1699	53	0
2	C	1383	0	1355	49	0
2	D	1412	0	1385	71	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	1	0	0	0	0
4	A	6	0	8	0	0
4	B	6	0	8	0	0
4	D	12	0	16	2	0
5	B	1	0	0	0	0
6	C	5	0	0	0	0
7	D	1	0	0	0	0
8	A	34	0	0	1	0
8	B	40	0	0	3	0
8	C	33	0	0	0	0
8	D	17	0	0	1	0
All	All	6466	0	6165	202	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 16.

The worst 5 of 202 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:B:191:TRP:HB2	1:B:225:MET:CE	1.62	1.28
2:D:70:THR:HA	2:D:71:ASN:HB2	1.16	1.13
1:B:191:TRP:CB	1:B:225:MET:HE1	1.80	1.12
1:B:191:TRP:HB2	1:B:225:MET:HE1	1.24	1.08
1:B:191:TRP:CB	1:B:225:MET:CE	2.33	1.03

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	215/229 (94%)	204 (95%)	9 (4%)	2 (1%)	17 48

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	216/229 (94%)	206 (95%)	10 (5%)	0	100	100
2	C	179/323 (55%)	161 (90%)	17 (10%)	1 (1%)	25	58
2	D	183/323 (57%)	162 (88%)	18 (10%)	3 (2%)	9	32
All	All	793/1104 (72%)	733 (92%)	54 (7%)	6 (1%)	19	51

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	170	TYR
2	C	71	ASN
2	D	128	ASN
2	D	61	THR
1	A	65	GLY

5.3.2 Protein sidechains (i)

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	182/192 (95%)	178 (98%)	4 (2%)	52	81
1	B	183/192 (95%)	180 (98%)	3 (2%)	62	86
2	C	145/260 (56%)	138 (95%)	7 (5%)	25	58
2	D	148/260 (57%)	140 (95%)	8 (5%)	22	54
All	All	658/904 (73%)	636 (97%)	22 (3%)	37	72

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	64	ASP
2	D	116	LEU
2	D	108	VAL
2	D	140	THR
1	B	94	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	119	GLN
2	D	114	ASN
2	D	171	GLN
2	D	133	GLN
1	B	158	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 8 ligands modelled in this entry, 3 are monoatomic - leaving 5 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	$\#\|Z\| > 2$
4	GOL	D	2	-	5,5,5	0.35	0	5,5,5	0.36	0
4	GOL	A	4	-	5,5,5	0.36	0	5,5,5	0.19	0
4	GOL	D	345	-	5,5,5	0.43	0	5,5,5	0.20	0
4	GOL	B	3	-	5,5,5	0.36	0	5,5,5	0.34	0
6	PO4	C	1	-	4,4,4	0.89	0	6,6,6	0.33	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
4	GOL	A	4	-	-	3/4/4/4	-
4	GOL	D	345	-	-	2/4/4/4	-
4	GOL	B	3	-	-	2/4/4/4	-
4	GOL	D	2	-	-	4/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	D	345	GOL	O1-C1-C2-C3
4	D	2	GOL	O1-C1-C2-C3
4	A	4	GOL	C1-C2-C3-O3
4	B	3	GOL	O1-C1-C2-C3
4	D	2	GOL	C1-C2-C3-O3

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	D	2	GOL	1	0
4	D	345	GOL	1	0

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 (i)

6.1 Protein, DNA and RNA chains (i)

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, 95th 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^2)	Q<0.9
1	A	217/229 (94%)	-0.46	1 (0%)	91	91	17, 31, 53, 70	0
1	B	218/229 (95%)	-0.54	0	100	100	15, 29, 58, 73	0
2	C	182/323 (56%)	-0.45	1 (0%)	91	91	17, 36, 63, 72	0
2	D	186/323 (57%)	-0.03	3 (1%)	72	71	17, 44, 76, 93	0
All	All	803/1104 (72%)	-0.38	5 (0%)	89	89	15, 34, 64, 93	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	114	ASN	3.1
2	C	114	ASN	2.6
2	D	170	TYR	2.6
2	D	169	GLY	2.3
1	A	128	PHE	2.3

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 monosaccharides 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. The B-factors column lists the minimum,

median, 95th 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	B-factors(Å^2)	Q<0.9
6	PO4	C	1	5/5	0.81	0.17	91,92,93,93	0
4	GOL	B	3	6/6	0.83	0.24	77,78,79,79	0
4	GOL	A	4	6/6	0.83	0.23	81,87,88,90	0
5	CL	B	1	1/1	0.90	0.26	40,40,40,40	0
4	GOL	D	345	6/6	0.91	0.19	58,61,63,64	0
4	GOL	D	2	6/6	0.91	0.24	56,63,66,67	0
3	K	A	1	1/1	0.93	0.10	61,61,61,61	0
7	NA	D	1	1/1	0.97	0.15	13,13,13,13	0

6.5 Other polymers (i)

There are no such residues in this entry.