



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 23, 2024 – 01:27 AM EDT

PDB ID : 5IOB
Title : Crystal structure of beta-N-acetylglucosaminidase-like protein from *Corynebacterium glutamicum*
Authors : Chang, C.; Mack, J.; Endres, M.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on : 2016-03-08
Resolution : 2.25 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 ⓘ 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 ⓘ](#)) 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	:	2.37.1
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)	:	2.37.1

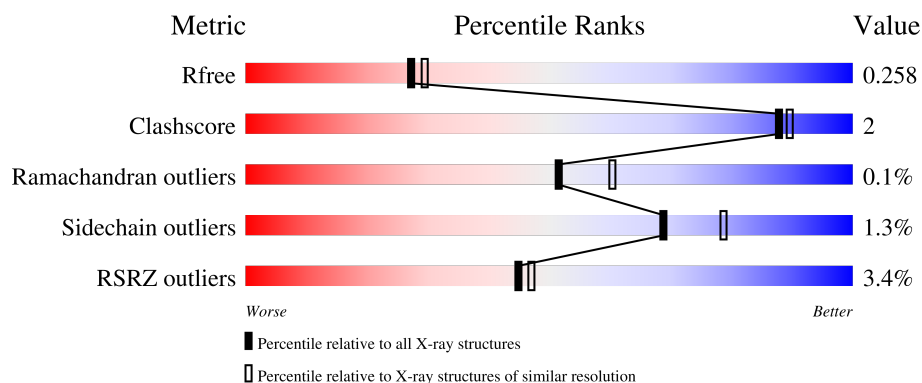
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.25 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	1377 (2.26-2.26)
Clashscore	141614	1487 (2.26-2.26)
Ramachandran outliers	138981	1449 (2.26-2.26)
Sidechain outliers	138945	1450 (2.26-2.26)
RSRZ outliers	127900	1356 (2.26-2.26)

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 $\leq 5\%$. 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	348	<div> <div>4%</div> <div>86%</div> <div>9%</div> </div>
1	B	348	<div> <div>3%</div> <div>87%</div> <div>10%</div> </div>
1	C	348	<div> <div>2%</div> <div>89%</div> <div>8%</div> </div>
1	D	348	<div> <div>2%</div> <div>86%</div> <div>5%</div> <div>9%</div> </div>
1	E	348	<div> <div>4%</div> <div>81%</div> <div>8%</div> <div>10%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	348	<div><div>%</div><div><div></div><div>85%</div><div>5%</div><div>10%</div></div></div>
1	G	348	<div><div>4%</div><div><div></div><div>89%</div><div>7%</div></div></div>
1	H	348	<div><div>3%</div><div><div></div><div>82%</div><div>7%</div><div>11%</div></div></div>

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 20143 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 Beta-glucosidase-related glycosidases.

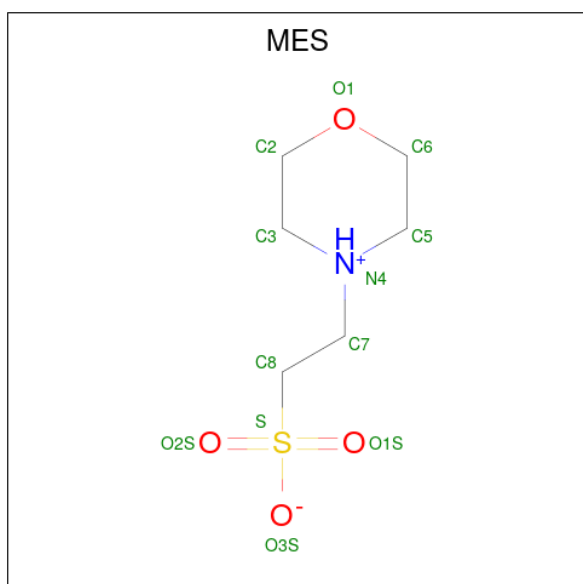
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	316	Total	C	N	O	Se	0	1	0
			2317	1460	386	463	8			
1	B	313	Total	C	N	O	Se	0	2	0
			2331	1468	388	467	8			
1	C	321	Total	C	N	O	Se	0	2	0
			2363	1486	397	472	8			
1	D	316	Total	C	N	O	Se	0	2	0
			2334	1472	388	466	8			
1	E	312	Total	C	N	O	Se	0	1	0
			2300	1448	384	460	8			
1	F	313	Total	C	N	O	Se	0	0	0
			2311	1457	386	460	8			
1	G	323	Total	C	N	O	Se	0	2	0
			2380	1497	401	474	8			
1	H	309	Total	C	N	O	Se	0	0	0
			2281	1437	379	457	8			

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



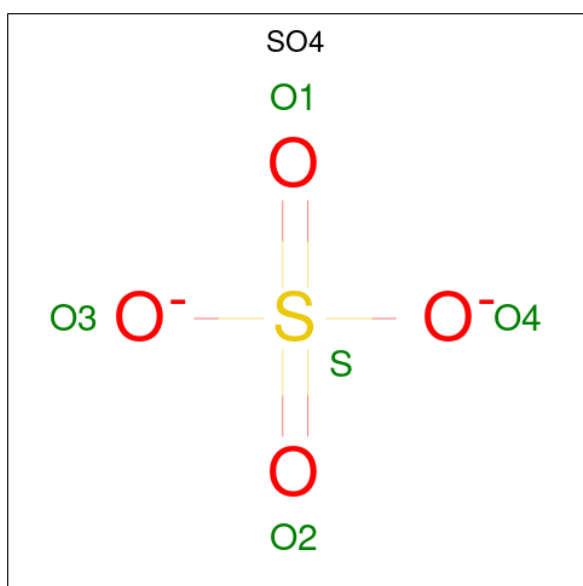
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			6	3	3		
2	C	1	Total	C	O	0	0
			6	3	3		
2	E	1	Total	C	O	0	0
			6	3	3		
2	G	1	Total	C	O	0	0
			6	3	3		

- Molecule 3 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: $C_6H_{13}NO_4S$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	C	1	Total	C	N	O	S	0	0
			12	6	1	4	1		
3	D	1	Total	C	N	O	S	0	0
			12	6	1	4	1		
3	E	1	Total	C	N	O	S	0	0
			12	6	1	4	1		
3	F	1	Total	C	N	O	S	0	0
			12	6	1	4	1		

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



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

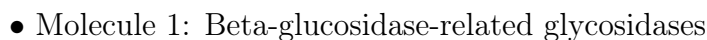
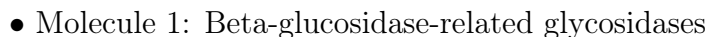
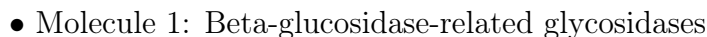
- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

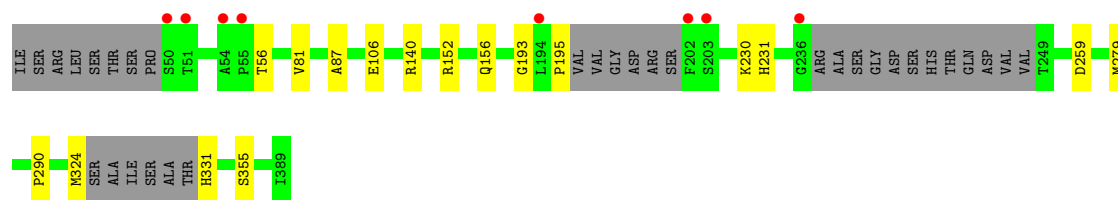
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	G	1	Total	Cl	0	0
			1	1		

- Molecule 6 is water.

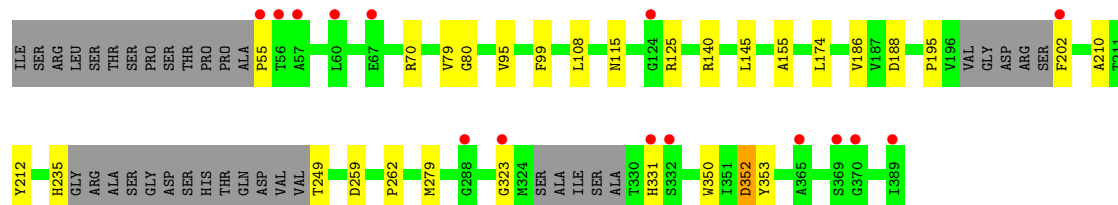
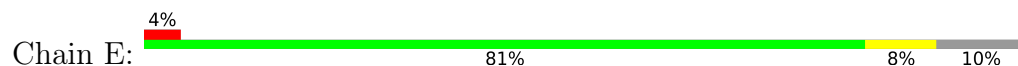
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	173	Total 173	O 173	0	0
6	B	181	Total 181	O 181	0	0
6	C	215	Total 215	O 215	0	0
6	D	186	Total 186	O 186	0	0
6	E	155	Total 155	O 155	0	0
6	F	198	Total 198	O 198	0	0
6	G	192	Total 192	O 192	0	0
6	H	143	Total 143	O 143	0	0

- Molecule 1: Beta-glucosidase-related glycosidases

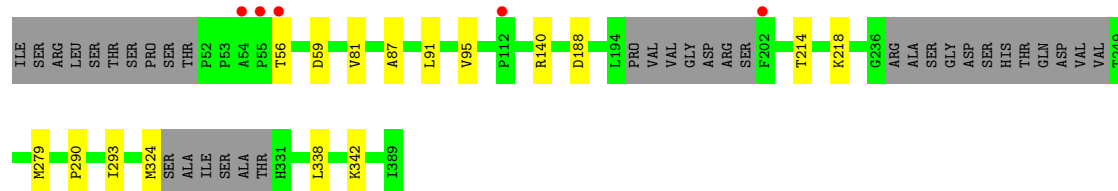
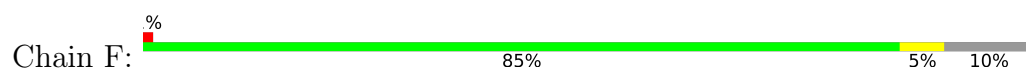




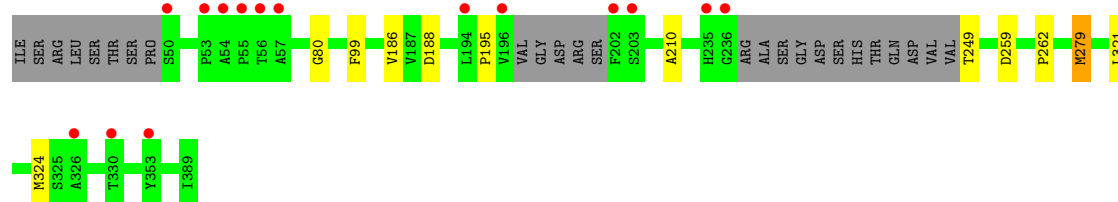
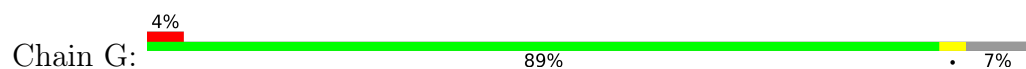
- Molecule 1: Beta-glucosidase-related glycosidases



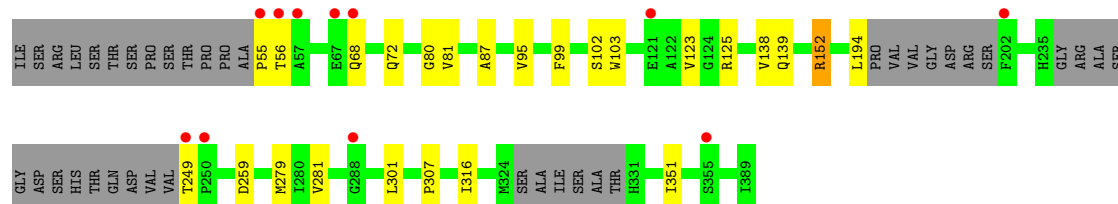
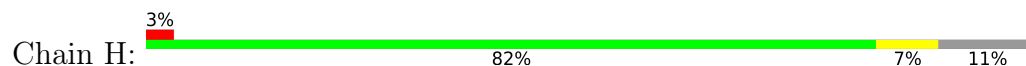
- Molecule 1: Beta-glucosidase-related glycosidases



- Molecule 1: Beta-glucosidase-related glycosidases



- Molecule 1: Beta-glucosidase-related glycosidases



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	308.07Å 91.36Å 120.02Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	37.28 – 2.25 37.28 – 2.25	Depositor EDS
% Data completeness (in resolution range)	84.1 (37.28-2.25) 96.6 (37.28-2.25)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.64 (at 2.24Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.209 , 0.257 0.211 , 0.258	Depositor DCC
R_{free} test set	7767 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	30.8	Xtriage
Anisotropy	0.249	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 45.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.56$, $\langle L^2 \rangle = 0.41$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	20143	wwPDB-VP
Average B, all atoms (Å ²)	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 44.15 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.5901e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

²Theoretical 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: GOL, CL, MES, SO4

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	0.29	0/2352	0.51	1/3196 (0.0%)
1	B	0.30	0/2369	0.49	0/3218
1	C	0.30	0/2402	0.49	0/3268
1	D	0.31	0/2373	0.50	1/3227 (0.0%)
1	E	0.31	0/2335	0.49	2/3171 (0.1%)
1	F	0.30	0/2350	0.46	0/3193
1	G	0.29	0/2420	0.48	1/3292 (0.0%)
1	H	0.29	0/2317	0.48	1/3147 (0.0%)
All	All	0.30	0/18918	0.49	6/25712 (0.0%)

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	H	55	PRO	N-CA-CB	5.94	110.43	103.30
1	E	55	PRO	N-CA-CB	5.92	110.40	103.30
1	A	55	PRO	N-CA-CB	5.89	110.38	103.30
1	E	195	PRO	N-CA-CB	5.88	110.35	103.30
1	D	195	PRO	N-CA-CB	5.73	110.18	103.30

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	2317	0	2220	7	0
1	B	2331	0	2249	5	0
1	C	2363	0	2275	10	0
1	D	2334	0	2233	9	0
1	E	2300	0	2203	15	0
1	F	2311	0	2232	9	0
1	G	2380	0	2287	7	0
1	H	2281	0	2196	11	0
2	A	6	0	8	0	0
2	C	6	0	8	3	0
2	E	6	0	8	0	0
2	G	6	0	8	0	0
3	C	12	0	12	0	0
3	D	12	0	12	1	0
3	E	12	0	12	1	0
3	F	12	0	12	1	0
4	C	5	0	0	0	0
4	G	5	0	0	0	0
5	G	1	0	0	0	0
6	A	173	0	0	1	0
6	B	181	0	0	0	0
6	C	215	0	0	2	0
6	D	186	0	0	2	0
6	E	155	0	0	6	0
6	F	198	0	0	2	0
6	G	192	0	0	1	0
6	H	143	0	0	1	0
All	All	20143	0	17975	74	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 2.

The worst 5 of 74 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:C:120[A]:ARG:NH1	6:C:601:HOH:O	2.18	0.77
1:G:249:THR:N	6:G:601:HOH:O	2.23	0.71
1:H:259:ASP:OD2	6:H:401:HOH:O	2.14	0.65
1:E:202:PHE:N	6:E:1203:HOH:O	2.29	0.65
1:F:188:ASP:OD2	6:F:601:HOH:O	2.14	0.65

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	309/348 (89%)	302 (98%)	6 (2%)	1 (0%)	41	46
1	B	307/348 (88%)	300 (98%)	7 (2%)	0	100	100
1	C	317/348 (91%)	312 (98%)	5 (2%)	0	100	100
1	D	310/348 (89%)	305 (98%)	5 (2%)	0	100	100
1	E	305/348 (88%)	298 (98%)	6 (2%)	1 (0%)	41	46
1	F	305/348 (88%)	298 (98%)	7 (2%)	0	100	100
1	G	319/348 (92%)	315 (99%)	4 (1%)	0	100	100
1	H	301/348 (86%)	294 (98%)	7 (2%)	0	100	100
All	All	2473/2784 (89%)	2424 (98%)	47 (2%)	2 (0%)	51	60

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	323	GLY
1	A	323	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	235/264 (89%)	230 (98%)	5 (2%)	53	62
1	B	243/264 (92%)	240 (99%)	3 (1%)	71	80
1	C	243/264 (92%)	242 (100%)	1 (0%)	91	94

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	239/264 (90%)	236 (99%)	3 (1%)	69	79
1	E	235/264 (89%)	230 (98%)	5 (2%)	53	62
1	F	239/264 (90%)	238 (100%)	1 (0%)	91	94
1	G	244/264 (92%)	243 (100%)	1 (0%)	91	94
1	H	236/264 (89%)	231 (98%)	5 (2%)	53	62
All	All	1914/2112 (91%)	1890 (99%)	24 (1%)	69	79

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	331	HIS
1	F	279	MSE
1	E	353	TYR
1	G	279	MSE
1	B	196	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 11 ligands modelled in this entry, 1 is monoatomic - leaving 10 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$
2	GOL	E	502	-	5,5,5	0.40	0	5,5,5	0.31	0
2	GOL	A	401	-	5,5,5	0.31	0	5,5,5	0.37	0
3	MES	D	501	-	12,12,12	2.33	1 (8%)	14,16,16	2.07	3 (21%)
4	SO4	C	502	-	4,4,4	0.11	0	6,6,6	0.18	0
3	MES	F	501	-	12,12,12	2.40	1 (8%)	14,16,16	2.10	8 (57%)
3	MES	E	501	-	12,12,12	2.31	1 (8%)	14,16,16	1.93	4 (28%)
2	GOL	G	503	-	5,5,5	0.37	0	5,5,5	0.22	0
4	SO4	G	501	-	4,4,4	0.15	0	6,6,6	0.06	0
2	GOL	C	503	-	5,5,5	0.39	0	5,5,5	0.23	0
3	MES	C	501	-	12,12,12	2.37	1 (8%)	14,16,16	2.49	7 (50%)

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
2	GOL	E	502	-	-	2/4/4/4	-
2	GOL	A	401	-	-	4/4/4/4	-
3	MES	D	501	-	-	1/6/14/14	0/1/1/1
3	MES	F	501	-	-	0/6/14/14	0/1/1/1
3	MES	E	501	-	-	0/6/14/14	0/1/1/1
2	GOL	G	503	-	-	0/4/4/4	-
2	GOL	C	503	-	-	2/4/4/4	-
3	MES	C	501	-	-	1/6/14/14	0/1/1/1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	F	501	MES	C8-S	-8.01	1.66	1.77
3	C	501	MES	C8-S	-8.00	1.66	1.77
3	D	501	MES	C8-S	-7.87	1.66	1.77
3	E	501	MES	C8-S	-7.75	1.66	1.77

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	501	MES	C5-N4-C3	5.76	121.80	108.83
3	D	501	MES	C5-N4-C3	5.24	120.61	108.83
3	E	501	MES	C5-N4-C3	4.77	119.57	108.83
3	F	501	MES	C5-N4-C3	4.23	118.35	108.83
3	D	501	MES	O2S-S-C8	3.39	111.00	106.92

There are no chirality outliers.

5 of 10 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	503	GOL	O1-C1-C2-C3
2	E	502	GOL	O1-C1-C2-C3
3	D	501	MES	C8-C7-N4-C5
2	C	503	GOL	O1-C1-C2-O2
2	E	502	GOL	O1-C1-C2-O2

There are no ring outliers.

4 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	D	501	MES	1	0
3	F	501	MES	1	0
3	E	501	MES	1	0
2	C	503	GOL	3	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	308/348 (88%)	-0.15	13 (4%) 36 38	23, 37, 70, 140	0
1	B	305/348 (87%)	-0.18	11 (3%) 42 44	23, 34, 66, 112	0
1	C	313/348 (89%)	-0.25	6 (1%) 66 69	19, 33, 64, 74	0
1	D	308/348 (88%)	-0.30	8 (2%) 56 59	19, 31, 66, 96	0
1	E	304/348 (87%)	-0.05	15 (4%) 29 32	29, 41, 75, 117	0
1	F	305/348 (87%)	-0.29	5 (1%) 72 74	27, 38, 73, 110	0
1	G	315/348 (90%)	-0.09	15 (4%) 30 33	28, 39, 73, 114	0
1	H	301/348 (86%)	-0.06	11 (3%) 41 44	31, 41, 73, 115	0
All	All	2459/2784 (88%)	-0.17	84 (3%) 45 47	19, 38, 71, 140	0

The worst 5 of 84 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	326	ALA	9.2
1	H	55	PRO	7.3
1	B	55	PRO	6.4
1	D	55	PRO	5.7
1	D	202	PHE	5.6

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
4	SO4	C	502	5/5	0.79	0.24	80,90,98,101	0
2	GOL	G	503	6/6	0.80	0.36	54,67,75,77	0
2	GOL	A	401	6/6	0.84	0.23	46,55,62,64	0
4	SO4	G	501	5/5	0.86	0.22	120,123,124,126	0
2	GOL	C	503	6/6	0.88	0.24	56,61,70,71	0
2	GOL	E	502	6/6	0.89	0.27	51,62,68,68	0
3	MES	E	501	12/12	0.94	0.17	57,76,83,83	0
3	MES	F	501	12/12	0.95	0.16	60,66,75,80	0
3	MES	D	501	12/12	0.96	0.12	52,57,60,62	0
5	CL	G	502	1/1	0.96	0.06	57,57,57,57	0
3	MES	C	501	12/12	0.97	0.15	42,56,62,64	0

6.5 Other polymers [i](#)

There are no such residues in this entry.