wwPDB X-ray Structure Validation Summary Report (i)

Nov 1, 2021 – 06:44 PM EDT

PDB ID	:	3FHA
Title	:	Structure of endo-beta-N-acetylglucosaminidase A
Authors	:	Yin, J.; Li, L.; Shaw, N.; Li, Y.; Song, J.K.; Zhang, W.; Xia, C.; Zhang, R.; Joachimiak, A.; Zhang, H.C.; Wang, L.X.; Wang, P.; Liu, Z.J.
Deposited on	:	2008-12-09
Resolution	:	2.00 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 following versions of software and data (see references (1)) were used in the production of this report:

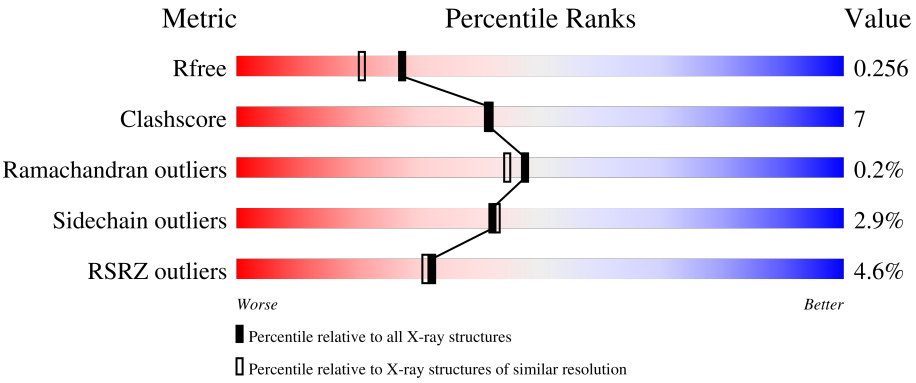
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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	621	3% 86% 9%
1	B	621	3% 82% 10% 6%
1	C	621	5% 80% 13% 5%
1	D	621	6% 77% 13% 8%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard

residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	GOL	B	700[B]	-	-	X	-

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 20919 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 Endo-beta-N-acetylglucosaminidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	598	Total	C	N	O	S	0	0	0
			4755	3014	796	934	11			
1	B	583	Total	C	N	O	S	0	0	0
			4660	2955	781	913	11			
1	C	592	Total	C	N	O	S	0	1	0
			4722	2994	793	924	11			
1	D	572	Total	C	N	O	S	0	0	0
			4577	2910	766	890	11			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	455	ASP	GLY	engineered mutation	UNP Q9ZB22
A	518	THR	ILE	engineered mutation	UNP Q9ZB22
A	583	ILE	LEU	engineered mutation	UNP Q9ZB22
B	455	ASP	GLY	engineered mutation	UNP Q9ZB22
B	518	THR	ILE	engineered mutation	UNP Q9ZB22
B	583	ILE	LEU	engineered mutation	UNP Q9ZB22
C	455	ASP	GLY	engineered mutation	UNP Q9ZB22
C	518	THR	ILE	engineered mutation	UNP Q9ZB22
C	583	ILE	LEU	engineered mutation	UNP Q9ZB22
D	455	ASP	GLY	engineered mutation	UNP Q9ZB22
D	518	THR	ILE	engineered mutation	UNP Q9ZB22
D	583	ILE	LEU	engineered mutation	UNP Q9ZB22

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

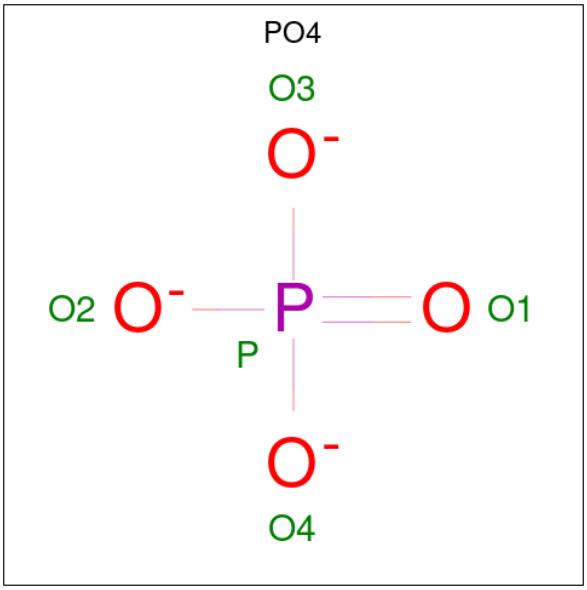
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Ca	0	0
			1	1		
2	B	1	Total	Ca	0	0
			1	1		

Continued on next page...

Continued from previous page...

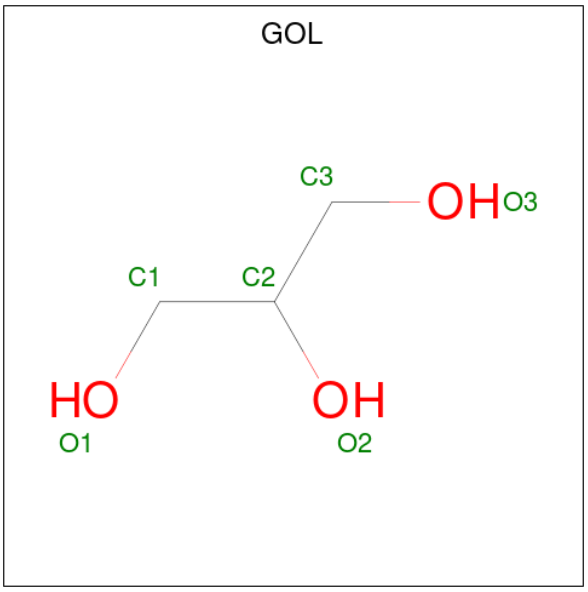
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	C	1	Total Ca 1 1	0	0
2	D	1	Total Ca 1 1	0	0

- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



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

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



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

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

- Molecule 6 is water.

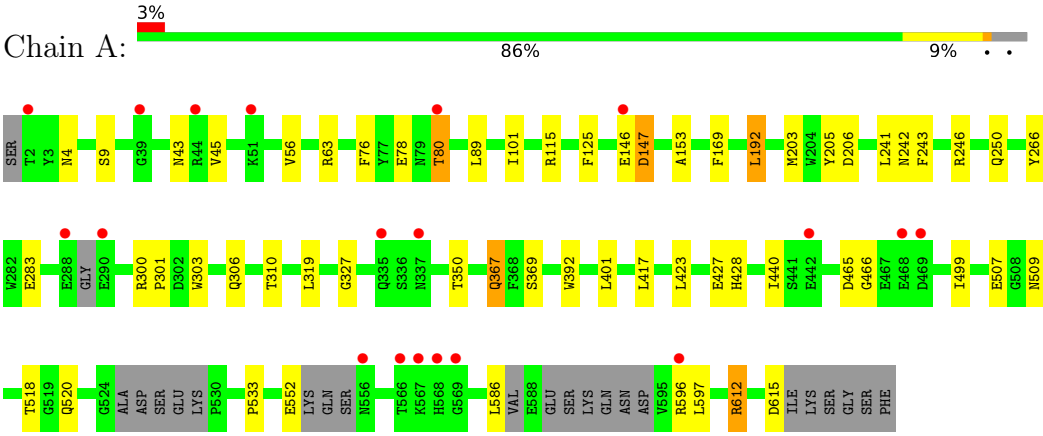
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	598	Total O 598 598	0	0
6	B	602	Total O 602 602	0	0
6	C	499	Total O 499 499	0	0
6	D	472	Total O 472 472	0	0

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.

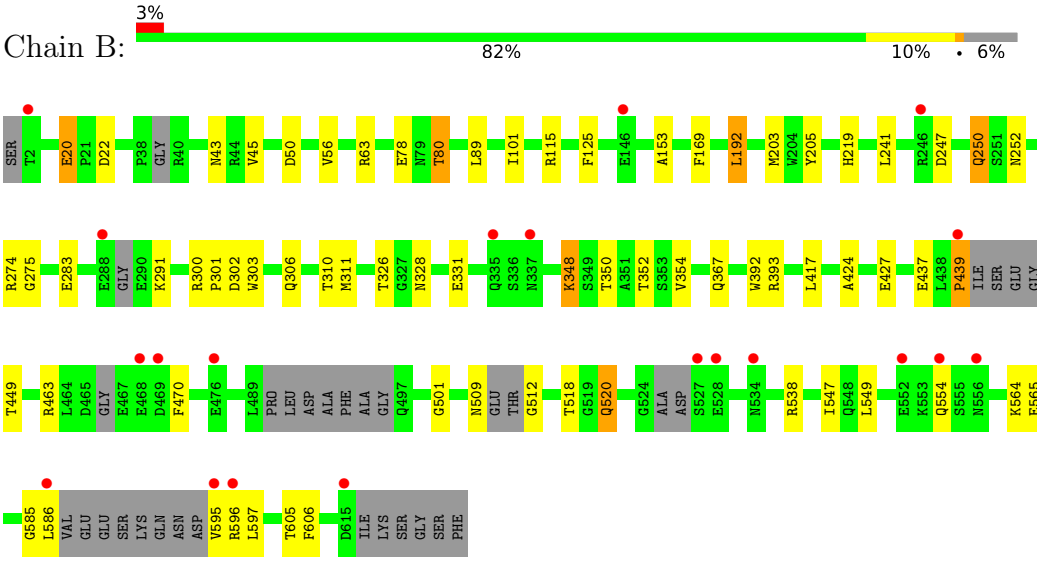
• Molecule 1: Endo-beta-N-acetylglucosaminidase

Chain A:



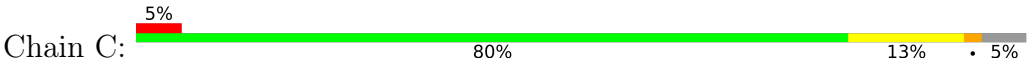
• Molecule 1: Endo-beta-N-acetylglucosaminidase

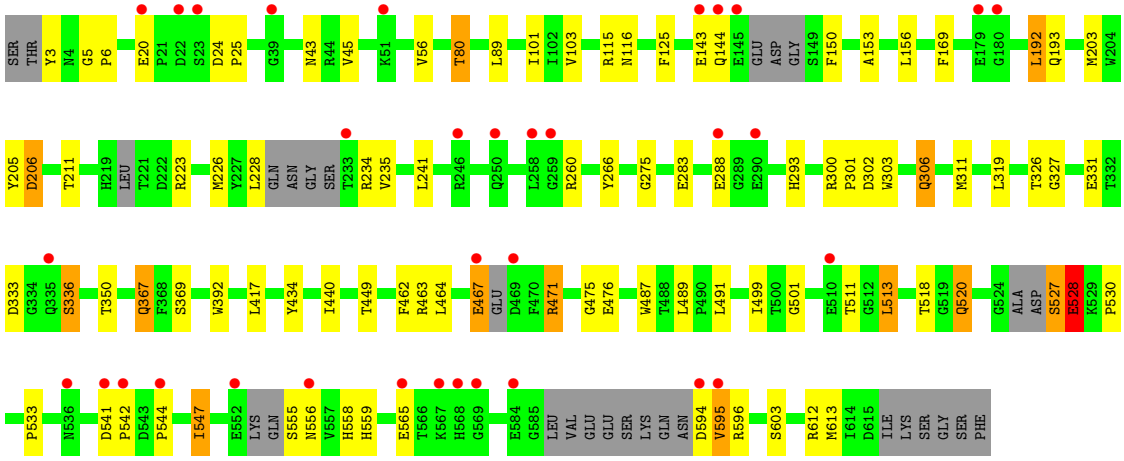
Chain B:



• Molecule 1: Endo-beta-N-acetylglucosaminidase

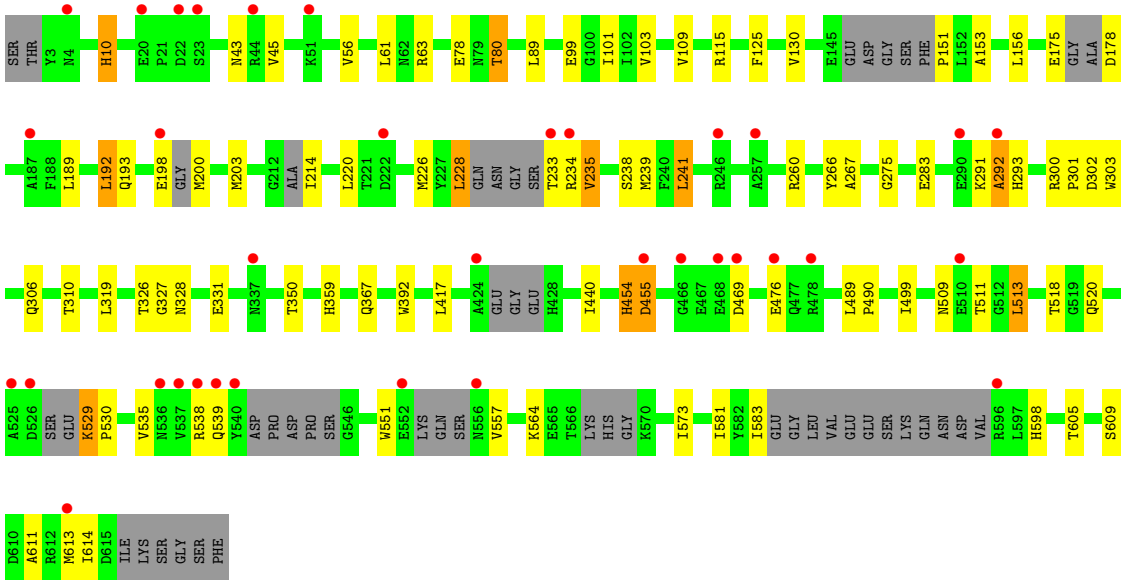
Chain C:





• Molecule 1: Endo-beta-N-acetylglucosaminidase

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	79.89Å 78.97Å 117.27Å 84.23° 80.77° 64.01°	Depositor
Resolution (Å)	20.00 – 2.00 19.92 – 2.00	Depositor EDS
% Data completeness (in resolution range)	(Not available) (20.00-2.00) 67.9 (19.92-2.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	7.85 (at 2.01Å)	Xtriage
Refinement program	REFMAC 5.4.0066	Depositor
R, R_{free}	0.203 , 0.251 0.213 , 0.256	Depositor DCC
R_{free} test set	5824 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	20.2	Xtriage
Anisotropy	0.333	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 49.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	20919	wwPDB-VP
Average B, all atoms (Å^2)	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.12% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MG, CA, PO4, GOL

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	0.56	2/4899 (0.0%)	0.58	1/6677 (0.0%)
1	B	0.53	0/4797	0.56	1/6531 (0.0%)
1	C	0.59	2/4864 (0.0%)	0.62	1/6626 (0.0%)
1	D	0.54	4/4711 (0.1%)	0.57	3/6413 (0.0%)
All	All	0.56	8/19271 (0.0%)	0.59	6/26247 (0.0%)

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	552	GLU	CB-CG	-7.86	1.37	1.52
1	D	292	ALA	CA-CB	-6.68	1.38	1.52
1	D	455	ASP	C-N	-5.87	1.20	1.34
1	C	306	GLN	CB-CG	-5.55	1.37	1.52
1	D	193	GLN	C-N	-5.47	1.21	1.34

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	20	GLU	OE1-CD-OE2	-5.49	116.71	123.30
1	D	454	HIS	O-C-N	5.26	131.12	122.70
1	C	528	GLU	CA-CB-CG	5.25	124.95	113.40
1	D	469	ASP	CB-CG-OD2	5.22	123.00	118.30
1	D	455	ASP	CB-CG-OD2	5.17	122.96	118.30

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	4755	0	4351	40	0
1	B	4660	0	4269	66	0
1	C	4722	0	4322	71	0
1	D	4577	0	4202	75	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	5	0	0	0	0
4	B	12	0	16	7	0
4	D	12	0	16	3	0
5	B	1	0	0	0	0
6	A	598	0	0	2	0
6	B	602	0	0	13	0
6	C	499	0	0	10	0
6	D	472	0	0	6	0
All	All	20919	0	17176	252	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 7.

The worst 5 of 252 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:586:LEU:HD22	1:B:595:VAL:HG21	1.26	1.15
1:D:43:ASN:HD22	1:D:392:TRP:HE1	1.05	1.01
1:B:43:ASN:HD22	1:B:392:TRP:HE1	1.02	1.00
1:B:586:LEU:HD22	1:B:595:VAL:CG2	1.92	1.00
1:C:558:HIS:HD2	1:C:559:HIS:HD1	1.10	0.99

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	587/621 (94%)	571 (97%)	15 (3%)	1 (0%)	47	44
1	B	563/621 (91%)	546 (97%)	16 (3%)	1 (0%)	47	44
1	C	577/621 (93%)	558 (97%)	16 (3%)	3 (0%)	29	23
1	D	548/621 (88%)	530 (97%)	18 (3%)	0	100	100
All	All	2275/2484 (92%)	2205 (97%)	65 (3%)	5 (0%)	47	44

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	595	VAL
1	C	288	GLU
1	A	147	ASP
1	B	585	GLY
1	C	206	ASP

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	494/515 (96%)	484 (98%)	10 (2%)	55	58
1	B	486/515 (94%)	473 (97%)	13 (3%)	44	46
1	C	490/515 (95%)	470 (96%)	20 (4%)	30	28
1	D	476/515 (92%)	461 (97%)	15 (3%)	39	38
All	All	1946/2060 (94%)	1888 (97%)	58 (3%)	42	41

5 of 58 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	234	ARG
1	D	529	LYS
1	C	471[B]	ARG
1	D	513	LEU
1	D	228	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	558	HIS
1	D	10	HIS
1	D	520	GLN
1	D	482	ASN
1	C	43	ASN

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 10 ligands modelled in this entry, 5 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	$\#\|Z\| > 2$
3	PO4	A	707	-	4,4,4	0.86	0	6,6,6	0.48	0
4	GOL	B	700[A]	-	5,5,5	0.33	0	5,5,5	0.30	0
4	GOL	D	701[A]	-	5,5,5	0.27	0	5,5,5	0.31	0
4	GOL	B	700[B]	-	5,5,5	0.41	0	5,5,5	0.41	0
4	GOL	D	701[B]	-	5,5,5	0.32	0	5,5,5	0.29	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	D	701[A]	-	-	4/4/4/4	-
4	GOL	B	700[B]	-	-	2/4/4/4	-
4	GOL	D	701[B]	-	-	4/4/4/4	-
4	GOL	B	700[A]	-	-	3/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	700[A]	GOL	C1-C2-C3-O3
4	B	700[B]	GOL	O1-C1-C2-C3
4	D	701[A]	GOL	O1-C1-C2-O2
4	D	701[A]	GOL	O1-C1-C2-C3
4	D	701[B]	GOL	O1-C1-C2-O2

There are no ring outliers.

4 monomers are involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	700[A]	GOL	1	0
4	D	701[A]	GOL	2	0
4	B	700[B]	GOL	6	0
4	D	701[B]	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

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^{th} 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	598/621 (96%)	0.01	19 (3%)	47	46	10, 19, 32, 42	0
1	B	583/621 (93%)	0.06	20 (3%)	45	44	10, 19, 32, 41	0
1	C	592/621 (95%)	0.16	34 (5%)	23	23	10, 21, 35, 55	0
1	D	572/621 (92%)	0.22	35 (6%)	21	20	10, 22, 35, 47	0
All	All	2345/2484 (94%)	0.11	108 (4%)	32	31	10, 20, 34, 55	0

The worst 5 of 108 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	466	GLY	7.1
1	C	510	GLU	5.5
1	C	288	GLU	5.3
1	C	250	GLN	5.1
1	C	569	GLY	5.1

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

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, 95^{th} 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
3	PO4	A	707	5/5	0.85	0.18	47,57,58,58	0
5	MG	B	706	1/1	0.85	0.26	43,43,43,43	0
4	GOL	D	701[B]	6/6	0.86	0.36	13,13,15,15	6
4	GOL	D	701[A]	6/6	0.86	0.36	12,13,13,15	6
4	GOL	B	700[B]	6/6	0.89	0.38	8,9,10,10	6
4	GOL	B	700[A]	6/6	0.89	0.38	8,9,10,11	6
2	CA	A	705	1/1	0.98	0.05	29,29,29,29	0
2	CA	D	703	1/1	0.98	0.05	32,32,32,32	0
2	CA	C	702	1/1	0.99	0.07	31,31,31,31	0
2	CA	B	704	1/1	1.00	0.03	28,28,28,28	0

6.5 Other polymers (i)

There are no such residues in this entry.