wwPDB X-ray Structure Validation Summary Report (i)

Aug 29, 2023 – 05:32 AM EDT

PDB ID	:	3MTN
Title	:	Usp21 in complex with a ubiquitin-based, USP21-specific inhibitor
Authors	:	Walker, J.R.; Avvakumov, G.V.; Xue, S.; Li, Y.; Ernst, A.; Sidhu, S.; Weigelt, J.; Bountra, C.; Edwards, A.M.; Arrowsmith, C.H.; Bochkarev, A.; Dhe-Paganon, S.; Structural Genomics Consortium (SGC)
Deposited on	:	2010-04-30
Resolution	:	2.70 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 (1)) were used in the production of this report:

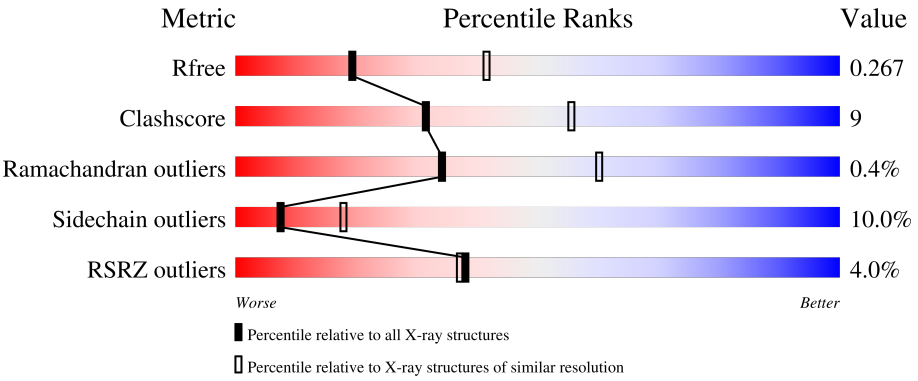
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 2.70 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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	373	2% 62% 19% • 15%
1	C	373	5% 64% 17% • 16%
2	B	85	% 69% 19% • 9%
2	D	85	2% 68% 21% • 9%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 6349 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 Ubiquitin carboxyl-terminal hydrolase 21.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	316	Total	C	N	O	S	0	1	0
			2527	1590	458	462	17			
1	C	314	Total	C	N	O	S	0	1	0
			2514	1580	458	459	17			

There are 38 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	190	MET	-	expression tag	UNP Q9UK80
A	191	GLY	-	expression tag	UNP Q9UK80
A	192	SER	-	expression tag	UNP Q9UK80
A	193	SER	-	expression tag	UNP Q9UK80
A	194	HIS	-	expression tag	UNP Q9UK80
A	195	HIS	-	expression tag	UNP Q9UK80
A	196	HIS	-	expression tag	UNP Q9UK80
A	197	HIS	-	expression tag	UNP Q9UK80
A	198	HIS	-	expression tag	UNP Q9UK80
A	199	HIS	-	expression tag	UNP Q9UK80
A	200	SER	-	expression tag	UNP Q9UK80
A	201	SER	-	expression tag	UNP Q9UK80
A	202	GLY	-	expression tag	UNP Q9UK80
A	203	LEU	-	expression tag	UNP Q9UK80
A	204	VAL	-	expression tag	UNP Q9UK80
A	205	PRO	-	expression tag	UNP Q9UK80
A	206	ARG	-	expression tag	UNP Q9UK80
A	207	GLY	-	expression tag	UNP Q9UK80
A	208	SER	-	expression tag	UNP Q9UK80
C	190	MET	-	expression tag	UNP Q9UK80
C	191	GLY	-	expression tag	UNP Q9UK80
C	192	SER	-	expression tag	UNP Q9UK80
C	193	SER	-	expression tag	UNP Q9UK80
C	194	HIS	-	expression tag	UNP Q9UK80
C	195	HIS	-	expression tag	UNP Q9UK80

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	196	HIS	-	expression tag	UNP Q9UK80
C	197	HIS	-	expression tag	UNP Q9UK80
C	198	HIS	-	expression tag	UNP Q9UK80
C	199	HIS	-	expression tag	UNP Q9UK80
C	200	SER	-	expression tag	UNP Q9UK80
C	201	SER	-	expression tag	UNP Q9UK80
C	202	GLY	-	expression tag	UNP Q9UK80
C	203	LEU	-	expression tag	UNP Q9UK80
C	204	VAL	-	expression tag	UNP Q9UK80
C	205	PRO	-	expression tag	UNP Q9UK80
C	206	ARG	-	expression tag	UNP Q9UK80
C	207	GLY	-	expression tag	UNP Q9UK80
C	208	SER	-	expression tag	UNP Q9UK80

- Molecule 2 is a protein called UBIQUITIN VARIANT UBV.21.4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	77	Total	C	N	O	S	0	0	0
			618	394	107	116	1			
2	D	77	Total	C	N	O	S	0	1	0
			629	400	111	117	1			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-2	GLY	-	expression tag	UNP P62988
B	-1	SER	-	expression tag	UNP P62988
B	0	HIS	-	expression tag	UNP P62988
B	64	TRP	GLU	engineered mutation	UNP P62988
B	68	PHE	HIS	engineered mutation	UNP P62988
B	70	LEU	VAL	engineered mutation	UNP P62988
B	77	GLY	-	expression tag	UNP P62988
B	78	GLY	-	expression tag	UNP P62988
B	79	GLY	-	expression tag	UNP P62988
B	80	GLY	-	expression tag	UNP P62988
B	81	SER	-	expression tag	UNP P62988
B	82	GLY	-	expression tag	UNP P62988
D	-2	GLY	-	expression tag	UNP P62988
D	-1	SER	-	expression tag	UNP P62988
D	0	HIS	-	expression tag	UNP P62988
D	64	TRP	GLU	engineered mutation	UNP P62988
D	68	PHE	HIS	engineered mutation	UNP P62988

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	70	LEU	VAL	engineered mutation	UNP P62988
D	77	GLY	-	expression tag	UNP P62988
D	78	GLY	-	expression tag	UNP P62988
D	79	GLY	-	expression tag	UNP P62988
D	80	GLY	-	expression tag	UNP P62988
D	81	SER	-	expression tag	UNP P62988
D	82	GLY	-	expression tag	UNP P62988

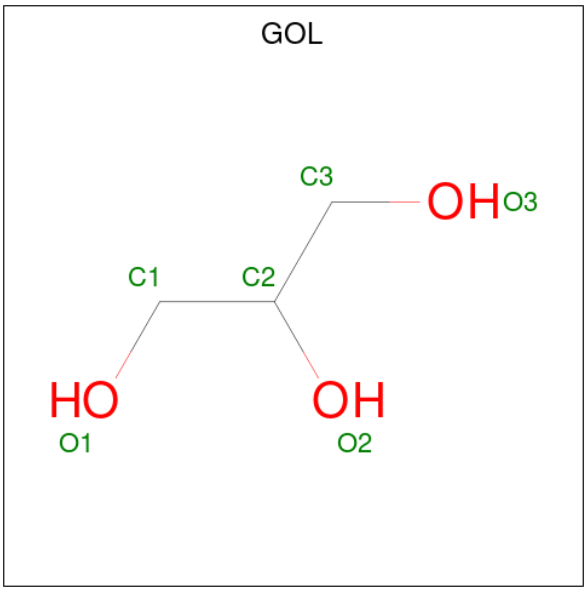
- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Zn 1 1	0	0
3	C	1	Total Zn 1 1	0	0

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	3	Total Cl 3 3	0	0
4	C	1	Total Cl 1 1	0	0

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C O 6 3 3	0	0
5	C	1	Total C O 6 3 3	0	0

- Molecule 6 is water.

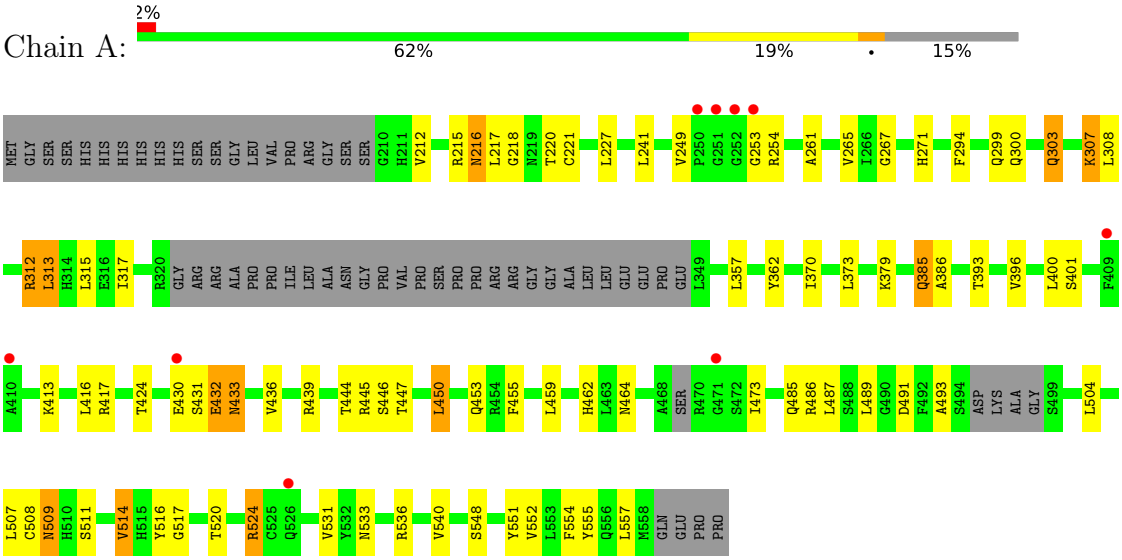
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	24	Total O 24 24	0	0
6	B	2	Total O 2 2	0	0
6	C	16	Total O 16 16	0	0
6	D	1	Total O 1 1	0	0

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.

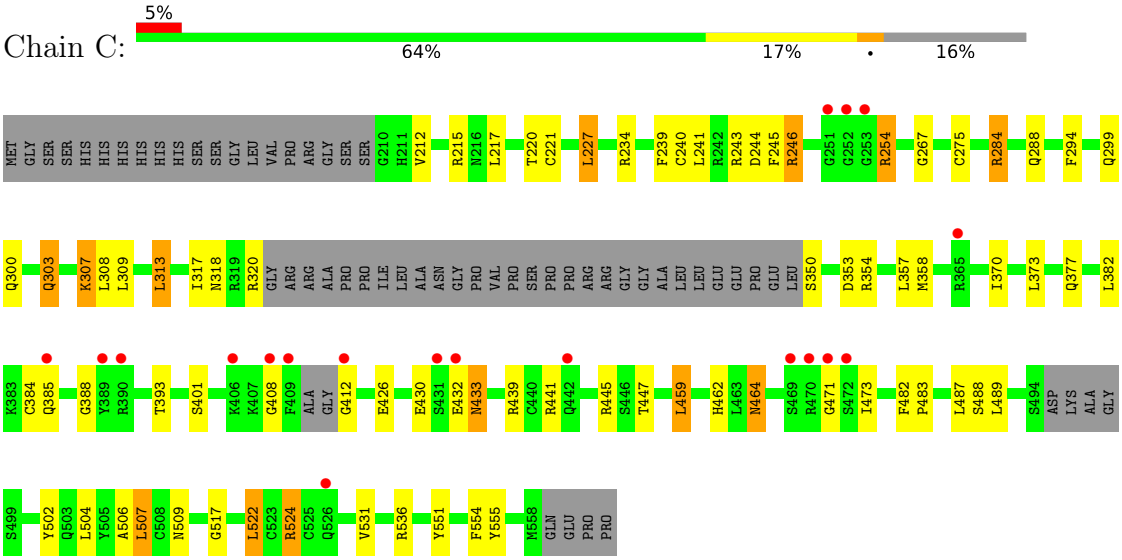
• Molecule 1: Ubiquitin carboxyl-terminal hydrolase 21

Chain A:



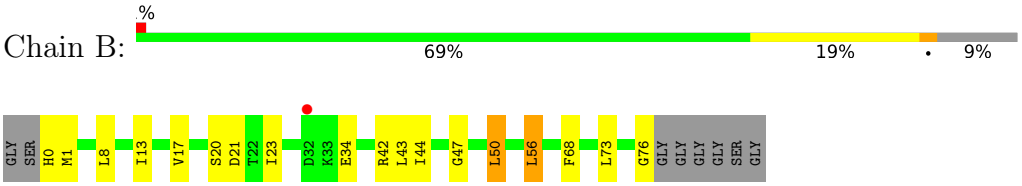
• Molecule 1: Ubiquitin carboxyl-terminal hydrolase 21

Chain C:



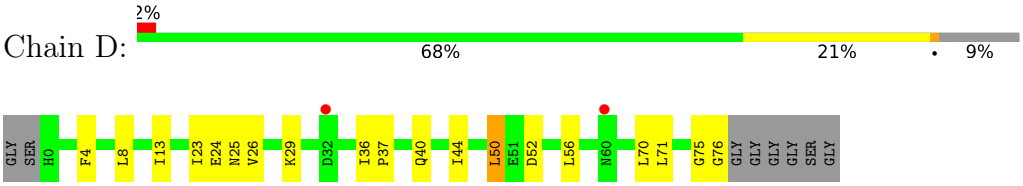
• Molecule 2: UBIQUITIN VARIANT UBV.21.4

Chain B:



• Molecule 2: UBIQUITIN VARIANT UBV.21.4

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	169.06Å 58.23Å 133.60Å 90.00° 121.19° 90.00°	Depositor
Resolution (Å)	19.61 – 2.70 19.61 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.9 (19.61-2.70) 99.9 (19.61-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.11	Depositor
$< I/\sigma(I) >$ [1]	2.46 (at 2.71Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.218 , 0.273 0.216 , 0.267	Depositor DCC
R_{free} test set	1553 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	50.1	Xtriage
Anisotropy	0.076	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 42.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	6349	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.74% 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: ZN, GOL, CL

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	0.47	0/2579	0.60	0/3477
1	C	0.45	0/2566	0.61	2/3460 (0.1%)
2	B	0.40	0/627	0.68	0/844
2	D	0.36	0/638	0.65	0/858
All	All	0.44	0/6410	0.62	2/8639 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	522	LEU	CA-CB-CG	5.36	127.62	115.30
1	C	507	LEU	CA-CB-CG	5.09	127.00	115.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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	2527	0	2469	56	0
1	C	2514	0	2451	45	0
2	B	618	0	644	13	0
2	D	629	0	656	14	0
3	A	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	1	0	0	0	0
4	A	3	0	0	0	0
4	C	1	0	0	0	0
5	A	6	0	8	0	0
5	C	6	0	8	0	0
6	A	24	0	0	1	0
6	B	2	0	0	0	0
6	C	16	0	0	1	0
6	D	1	0	0	0	0
All	All	6349	0	6236	113	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 9.

The worst 5 of 113 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:524:ARG:HH11	1:A:524:ARG:HG2	1.22	1.03
1:C:294:PHE:CE2	1:C:300:GLN:HG2	2.16	0.80
1:A:424:THR:HG23	1:A:450:LEU:O	1.88	0.74
1:C:524:ARG:HG2	1:C:524:ARG:HH11	1.54	0.72
1:A:294:PHE:CE2	1:A:300:GLN:HG2	2.26	0.71

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	309/373 (83%)	285 (92%)	22 (7%)	2 (1%)	25 50
1	C	307/373 (82%)	278 (91%)	28 (9%)	1 (0%)	41 66

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
2	B	75/85 (88%)	73 (97%)	2 (3%)	0	100 100
2	D	76/85 (89%)	72 (95%)	4 (5%)	0	100 100
All	All	767/916 (84%)	708 (92%)	56 (7%)	3 (0%)	34 60

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	432	GLU
1	A	253	GLY
1	C	471	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	277/324 (86%)	242 (87%)	35 (13%)	4 10
1	C	276/324 (85%)	250 (91%)	26 (9%)	8 20
2	B	69/71 (97%)	62 (90%)	7 (10%)	7 17
2	D	70/71 (99%)	68 (97%)	2 (3%)	42 71
All	All	692/790 (88%)	622 (90%)	70 (10%)	7 17

5 of 70 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	377	GLN
1	C	393	THR
1	C	489	LEU
1	A	459	LEU
1	A	450	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	385	GLN
1	C	464	ASN
1	C	556	GLN
1	C	543	ASN
1	A	464	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 8 ligands modelled in this entry, 6 are monoatomic - leaving 2 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	$\#\|Z\| > 2$
5	GOL	A	2	-	5,5,5	0.46	0	5,5,5	0.35	0
5	GOL	C	3	-	5,5,5	0.32	0	5,5,5	0.49	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
5	GOL	A	2	-	-	2/4/4/4	-
5	GOL	C	3	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	2	GOL	O1-C1-C2-C3
5	C	3	GOL	C1-C2-C3-O3
5	A	2	GOL	O1-C1-C2-O2
5	C	3	GOL	O2-C2-C3-O3

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^{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	316/373 (84%)	-0.10	9 (2%)	53	54	27, 45, 81, 95	0
1	C	314/373 (84%)	0.08	19 (6%)	21	20	30, 54, 101, 115	0
2	B	77/85 (90%)	-0.31	1 (1%)	77	78	31, 51, 65, 69	0
2	D	77/85 (90%)	-0.12	2 (2%)	56	57	42, 69, 92, 99	0
All	All	784/916 (85%)	-0.05	31 (3%)	38	37	27, 50, 91, 115	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	253	GLY	5.4
1	C	470	ARG	5.4
1	A	252	GLY	5.3
1	C	409	PHE	4.5
1	A	251	GLY	4.0

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
4	CL	C	1	1/1	0.83	0.38	109,109,109,109	0
5	GOL	C	3	6/6	0.87	0.25	77,78,79,79	0
4	CL	A	563	1/1	0.89	0.47	88,88,88,88	0
5	GOL	A	2	6/6	0.92	0.17	56,56,57,57	0
4	CL	A	564	1/1	0.95	0.25	72,72,72,72	0
4	CL	A	1	1/1	0.96	0.10	59,59,59,59	0
3	ZN	C	700	1/1	0.96	0.06	105,105,105,105	0
3	ZN	A	700	1/1	0.98	0.06	83,83,83,83	0

6.5 Other polymers (i)

There are no such residues in this entry.