Full wwPDB X-ray Structure Validation Report (i)

Oct 15, 2023 – 07:04 AM EDT

PDB ID	:	7N2S
Title	:	AS3.1-PRPF3-HLA*B27
Authors	:	Yang, X.; Jude, K.M.; Garcia, K.C.
Deposited on	:	2021-05-29
Resolution	:	2.37 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*
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:

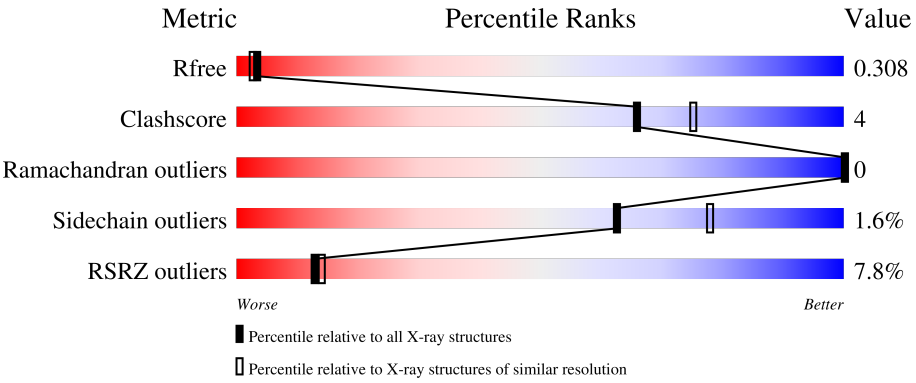
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.37 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	5509 (2.40-2.36)
Clashscore	141614	6082 (2.40-2.36)
Ramachandran outliers	138981	5973 (2.40-2.36)
Sidechain outliers	138945	5975 (2.40-2.36)
RSRZ outliers	127900	5397 (2.40-2.36)

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	C	9	89% 11%
2	A	278	6% 86% 13%
3	B	100	87% 12%
4	D	209	17% 76% 11% 13%
5	F	242	5% 88% 11%

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 6462 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 Pre-MRNA Processing Factor 3.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
1	C	9	69	45	13	11	0	0	0

- Molecule 2 is a protein called Human leukocyte antigen (HLA) B27.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	276	2251	1401	408	436	6	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	67	SER	CYS	conflict	UNP A3F718

- Molecule 3 is a protein called Beta-2-microglobulin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	B	100	837	533	141	159	4	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	MET	-	initiating methionine	UNP P61769

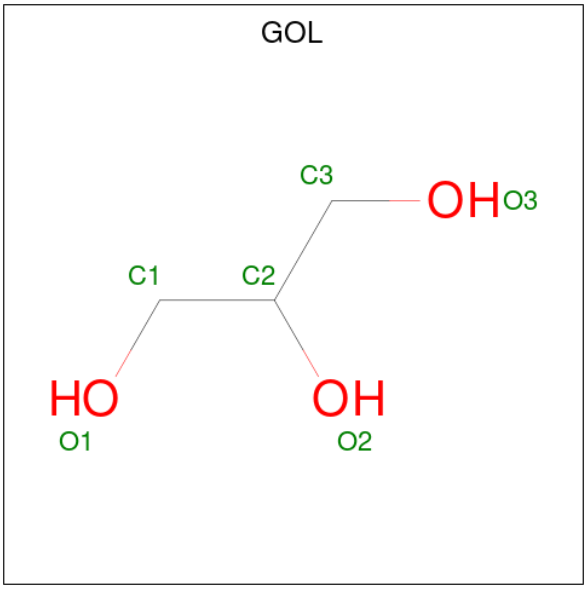
- Molecule 4 is a protein called T cell receptor alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	181	1376	863	229	279	5	0	0	0

- Molecule 5 is a protein called T cell receptor beta chain.

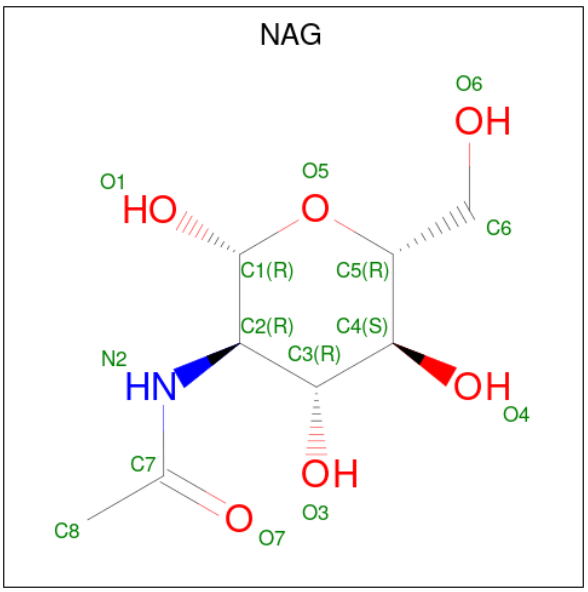
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	F	239	1905	1205	327	368	5	0	0	0

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

- Molecule 7 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	D	1	Total C N O 14 8 1 5	0	0

- Molecule 8 is water.

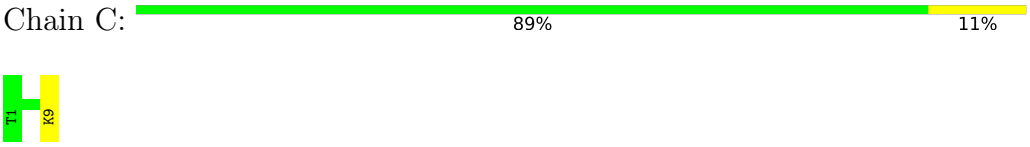
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	A	3	Total O 3 3	0	0
8	D	1	Total O 1 1	0	0

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.

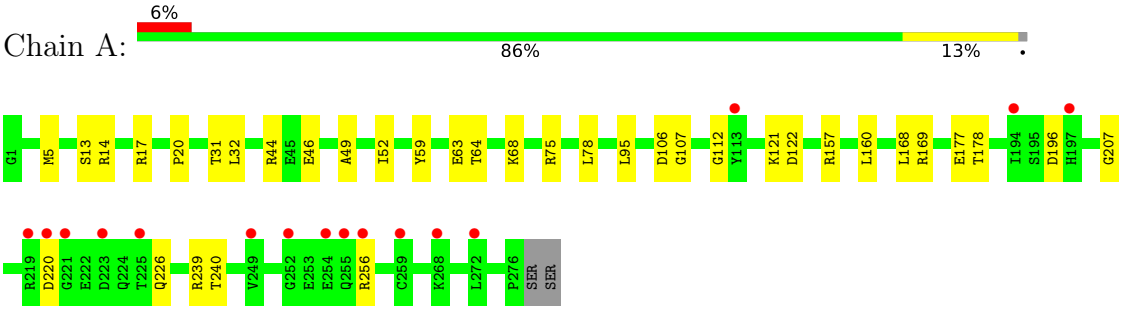
• Molecule 1: Pre-MRNA Processing Factor 3

Chain C:



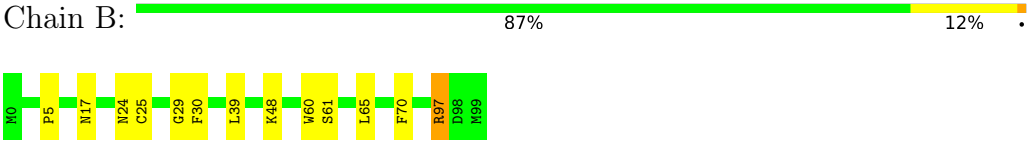
• Molecule 2: Human leukocyte antigen (HLA) B27

Chain A:



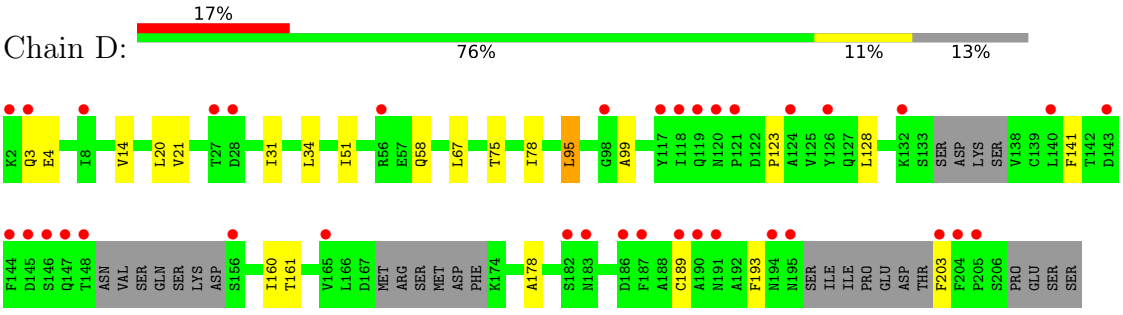
• Molecule 3: Beta-2-microglobulin

Chain B:



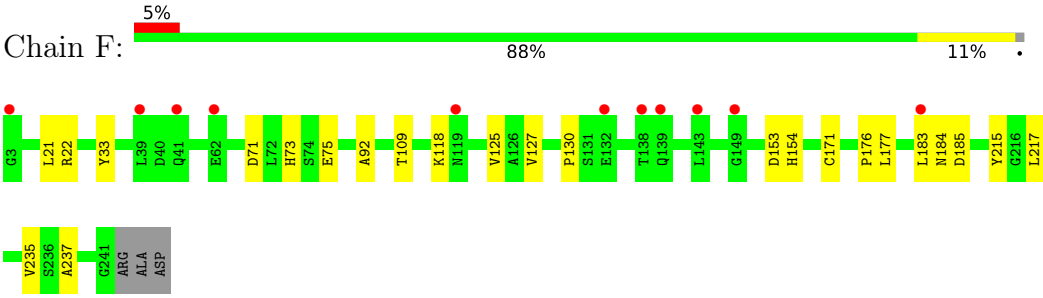
• Molecule 4: T cell receptor alpha chain

Chain D:



• Molecule 5: T cell receptor beta chain

Chain F:


5%
88%
11%
G3
L21
R22
Y33
L39
D40
Q41
E62
D71
L72
H73
S74
E75
A92
T109
K118
N119
V125
A126
V127
P130
S131
E132
T138
Q139
L143
G149
D153
H154
C171
P176
L177
L183
N184
D185
Y215
G216
L217
Q225
P230
V235
S236
A237
G241
ARG
ALA
ASP

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	83.21Å 52.79Å 106.29Å 90.00° 98.28° 90.00°	Depositor
Resolution (Å)	47.18 – 2.37 47.54 – 2.37	Depositor EDS
% Data completeness (in resolution range)	50.6 (47.18-2.37) 50.6 (47.54-2.37)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.77 (at 2.37Å)	Xtriage
Refinement program	PHENIX 1.19.1_4122+SVN	Depositor
R, R_{free}	0.251 , 0.311 0.250 , 0.308	Depositor DCC
R_{free} test set	1623 reflections (8.53%)	wwPDB-VP
Wilson B-factor (Å^2)	33.3	Xtriage
Anisotropy	0.052	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 24.0	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	0.86	EDS
Total number of atoms	6462	wwPDB-VP
Average B, all atoms (Å^2)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.99% 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: GOL, NAG

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	C	0.24	0/69	0.61	0/91
2	A	0.23	0/2313	0.50	0/3147
3	B	0.24	0/860	0.45	0/1162
4	D	0.25	0/1399	0.47	0/1895
5	F	0.24	0/1956	0.48	0/2666
All	All	0.24	0/6597	0.48	0/8961

There are no bond length outliers.

There are no bond angle outliers.

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	C	69	0	85	1	0
2	A	2251	0	2109	19	0
3	B	837	0	803	7	0
4	D	1376	0	1326	14	0
5	F	1905	0	1811	14	0
6	A	6	0	8	0	0
7	D	14	0	13	0	0
8	A	3	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
8	D	1	0	0	0	0
All	All	6462	0	6155	50	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 (50) 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)
4:D:31:ILE:HG22	4:D:95:LEU:HD23	1.78	0.65
2:A:32:LEU:O	2:A:239:ARG:NH2	2.30	0.64
2:A:107:GLY:O	2:A:169:ARG:NH1	2.33	0.60
3:B:17:ASN:OD1	3:B:97:ARG:NH2	2.33	0.60
2:A:14:ARG:HB3	2:A:17:ARG:HB2	1.84	0.59
4:D:123:PRO:O	4:D:203:PHE:N	2.36	0.58
4:D:20:LEU:HB2	4:D:78:ILE:HB	1.86	0.58
1:C:9:LYS:HE3	2:A:95:LEU:HD11	1.86	0.57
4:D:34:LEU:HD13	4:D:67:LEU:HD13	1.86	0.56
2:A:177:GLU:HG2	2:A:178:THR:HG23	1.88	0.55
2:A:5:MET:HB2	2:A:168:LEU:HD13	1.89	0.54
2:A:122:ASP:OD1	3:B:60:TRP:NE1	2.37	0.53
3:B:5:PRO:HB3	3:B:30:PHE:HB3	1.89	0.53
5:F:118:LYS:NZ	5:F:225:GLN:OE1	2.41	0.52
2:A:31:THR:HG23	2:A:239:ARG:HE	1.74	0.52
2:A:68:LYS:HD3	4:D:99:ALA:HB1	1.91	0.52
4:D:128:LEU:HD23	5:F:130:PRO:HA	1.91	0.52
5:F:153:ASP:HB2	5:F:176:PRO:HG2	1.91	0.51
5:F:183:LEU:HD23	5:F:185:ASP:H	1.76	0.51
5:F:33:TYR:HB2	5:F:92:ALA:HB3	1.93	0.51
4:D:160:ILE:HD11	4:D:193:PHE:HE2	1.77	0.50
5:F:184:ASN:OD1	5:F:184:ASN:N	2.44	0.50
3:B:29:GLY:HA2	3:B:61:SER:HB3	1.94	0.49
5:F:22:ARG:NH1	5:F:75:GLU:OE2	2.45	0.49
5:F:154:HIS:HB3	5:F:215:TYR:HB2	1.95	0.48
5:F:125:VAL:HG12	5:F:235:VAL:HG12	1.97	0.47
4:D:14:VAL:HG21	4:D:20:LEU:HD21	1.97	0.47
4:D:141:PHE:HB2	4:D:178:ALA:HB3	1.96	0.47
2:A:112:GLY:HA3	2:A:160:LEU:HD13	1.97	0.46
4:D:14:VAL:HG11	4:D:20:LEU:HD21	1.96	0.46
4:D:21:VAL:HG23	4:D:75:THR:HG23	1.98	0.46
3:B:24:ASN:HB3	3:B:65:LEU:HD21	1.99	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:D:161:THR:HG22	5:F:177:LEU:HD21	1.99	0.45
4:D:51:ILE:HG12	4:D:58:GLN:HB2	2.00	0.44
2:A:13:SER:HB3	2:A:78:LEU:HD13	1.99	0.44
2:A:106:ASP:OD1	2:A:106:ASP:N	2.45	0.43
3:B:25:CYS:HB2	3:B:39:LEU:HD21	2.00	0.43
2:A:59:TYR:O	2:A:63:GLU:HG2	2.17	0.43
5:F:127:VAL:HG23	5:F:237:ALA:HB3	2.01	0.43
3:B:48:LYS:HD2	3:B:48:LYS:HA	1.81	0.42
2:A:220:ASP:N	2:A:256:ARG:O	2.48	0.42
5:F:71:ASP:CG	5:F:73:HIS:HD1	2.23	0.42
2:A:44:ARG:HG3	2:A:64:THR:OG1	2.19	0.41
4:D:3:GLN:HG2	4:D:95:LEU:HD13	2.02	0.41
2:A:44:ARG:O	2:A:46:GLU:HG2	2.21	0.41
2:A:49:ALA:O	2:A:52:ILE:HG22	2.21	0.41
2:A:13:SER:HA	2:A:20:PRO:HB3	2.02	0.41
2:A:207:GLY:HA2	2:A:240:THR:HB	2.01	0.41
5:F:21:LEU:HD22	5:F:109:THR:HG21	2.01	0.41
5:F:217:LEU:HD12	5:F:230:PRO:HD2	2.03	0.40

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	C	7/9 (78%)	7 (100%)	0	0	100	100
2	A	274/278 (99%)	260 (95%)	14 (5%)	0	100	100
3	B	98/100 (98%)	95 (97%)	3 (3%)	0	100	100
4	D	171/209 (82%)	164 (96%)	7 (4%)	0	100	100
5	F	237/242 (98%)	229 (97%)	8 (3%)	0	100	100
All	All	787/838 (94%)	755 (96%)	32 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	C	7/7 (100%)	7 (100%)	0	100	100
2	A	235/237 (99%)	230 (98%)	5 (2%)	53	70
3	B	95/95 (100%)	93 (98%)	2 (2%)	53	70
4	D	155/183 (85%)	152 (98%)	3 (2%)	57	73
5	F	209/211 (99%)	208 (100%)	1 (0%)	88	95
All	All	701/733 (96%)	690 (98%)	11 (2%)	62	78

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	A	75	ARG
2	A	121	LYS
2	A	157	ARG
2	A	196	ASP
2	A	226	GLN
3	B	70	PHE
3	B	97	ARG
4	D	4	GLU
4	D	95	LEU
4	D	189	CYS
5	F	171	CYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
5	F	202	GLN

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

2 ligands are modelled in this entry.

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	$\#\|Z\| > 2$
7	NAG	D	301	4	14,14,15	0.20	0	17,19,21	0.41	0
6	GOL	A	301	-	5,5,5	0.91	0	5,5,5	1.01	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
7	NAG	D	301	4	-	0/6/23/26	0/1/1/1
6	GOL	A	301	-	-	0/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

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^{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	C	9/9 (100%)	0.11	0 100 100	18, 23, 30, 39	0
2	A	276/278 (99%)	0.67	16 (5%) 23 25	19, 36, 80, 109	0
3	B	100/100 (100%)	0.25	0 100 100	19, 35, 50, 62	0
4	D	181/209 (86%)	1.31	36 (19%) 1 1	26, 49, 110, 149	0
5	F	239/242 (98%)	0.53	11 (4%) 32 35	26, 43, 79, 126	0
All	All	805/838 (96%)	0.71	63 (7%) 13 14	18, 41, 87, 149	0

All (63) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	D	148	THR	11.9
4	D	190	ALA	10.8
4	D	204	PHE	9.2
4	D	119	GLN	8.8
4	D	117	TYR	7.9
4	D	145	ASP	6.7
4	D	165	VAL	6.1
4	D	120	ASN	5.4
4	D	195	ASN	5.1
5	F	183	LEU	5.1
5	F	3	GLY	5.0
4	D	183	ASN	4.9
4	D	144	PHE	4.7
2	A	255	GLN	4.6
2	A	194	ILE	4.4
2	A	249	VAL	4.1
5	F	62	GLU	4.1
4	D	146	SER	3.9
4	D	187	PHE	3.6
4	D	28	ASP	3.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	A	272	LEU	3.4
4	D	147	GLN	3.4
2	A	220	ASP	3.3
4	D	191	ASN	3.2
4	D	118	ILE	3.2
4	D	2	LYS	3.2
4	D	140	LEU	3.2
5	F	138	THR	3.1
4	D	182	SER	3.1
2	A	259	CYS	2.9
4	D	121	PRO	2.9
4	D	203	PHE	2.9
2	A	219	ARG	2.9
4	D	132	LYS	2.9
2	A	256	ARG	2.8
4	D	124	ALA	2.7
4	D	205	PRO	2.7
4	D	186	ASP	2.7
5	F	41	GLN	2.6
2	A	197	HIS	2.6
4	D	56	ARG	2.5
5	F	39	LEU	2.4
5	F	132	GLU	2.4
2	A	221	GLY	2.4
2	A	225	THR	2.4
2	A	223	ASP	2.3
4	D	98	GLY	2.3
4	D	3	GLN	2.3
5	F	139	GLN	2.3
2	A	254	GLU	2.3
5	F	119	ASN	2.3
4	D	194	ASN	2.3
4	D	27	THR	2.2
4	D	189	CYS	2.2
2	A	268	LYS	2.2
4	D	143	ASP	2.1
4	D	8	ILE	2.1
5	F	143	LEU	2.1
2	A	113	TYR	2.1
5	F	149	GLY	2.1
4	D	156	SER	2.1
4	D	126	TYR	2.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	A	252	GLY	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 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, 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
7	NAG	D	301	14/15	0.84	0.18	54,61,71,73	0
6	GOL	A	301	6/6	0.89	0.15	28,34,42,44	0

6.5 Other polymers (i)

There are no such residues in this entry.