



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 15, 2017 – 01:23 am GMT

PDB ID : 4YDJ  
Title : Crystal structure of broadly and potently neutralizing antibody 44-VRC13.01  
in complex with HIV-1 clade AE strain 93TH057 gp120  
Authors : Zhou, T.; Moquin, S.; Kwong, P.D.  
Deposited on : 2015-02-22  
Resolution : 2.31 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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity	:	4.02b-467
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

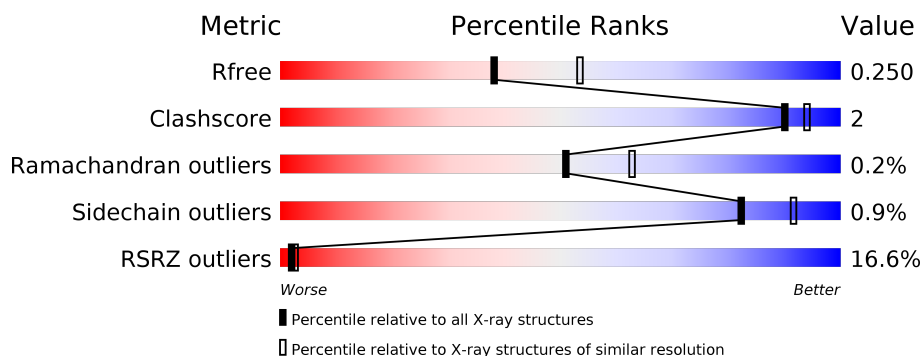
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.31 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))
$R_{free}$	100719	4130 (2.30-2.30)
Clashscore	112137	4751 (2.30-2.30)
Ramachandran outliers	110173	4705 (2.30-2.30)
Sidechain outliers	110143	4704 (2.30-2.30)
RSRZ outliers	101464	4156 (2.30-2.30)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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	G	353	<div> <div>28%</div> <div>79% 5% 16%</div> </div>
1	I	353	<div> <div>27%</div> <div>76% 6% 17%</div> </div>
2	A	238	<div> <div>5%</div> <div>97% .</div> </div>
2	H	238	<div> <div>7%</div> <div>95% .</div> </div>
3	B	206	<div> <div>4%</div> <div>94% 5%</div> </div>
3	L	206	<div> <div>7%</div> <div>94% 5%</div> </div>

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	NAG	B	301	-	-	-	X
4	NAG	G	509	-	-	-	X
4	NAG	I	501	-	-	-	X
4	NAG	I	502	-	-	-	X
4	NAG	I	506	-	-	-	X
4	NAG	I	507	-	-	-	X
4	NAG	L	301	-	-	-	X
6	BU3	B	302	-	-	-	X
8	PEG	A	304	-	-	-	X
8	PEG	B	304	-	-	-	X
8	PEG	B	305	-	-	-	X

## 2 Entry composition [i](#)

There are 10 unique types of molecules in this entry. The entry contains 23589 atoms, of which 11493 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 Envelope glycoprotein gp160,Envelope glycoprotein gp160.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	G	298	Total	C	H	N	O	S	0	0	0
			4646	1476	2295	411	445	19			
1	I	292	Total	C	H	N	O	S	0	0	0
			4555	1451	2248	402	436	18			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	197	GLY	-	linker	UNP Q0ED31
G	198	GLY	-	linker	UNP Q0ED31
G	318	GLY	-	linker	UNP Q0ED31
G	319	GLY	-	linker	UNP Q0ED31
G	320	SER	-	linker	UNP Q0ED31
G	321	GLY	-	linker	UNP Q0ED31
G	322	SER	-	linker	UNP Q0ED31
G	323	GLY	-	linker	UNP Q0ED31
I	197	GLY	-	linker	UNP Q0ED31
I	198	GLY	-	linker	UNP Q0ED31
I	318	GLY	-	linker	UNP Q0ED31
I	319	GLY	-	linker	UNP Q0ED31
I	320	SER	-	linker	UNP Q0ED31
I	321	GLY	-	linker	UNP Q0ED31
I	322	SER	-	linker	UNP Q0ED31
I	323	GLY	-	linker	UNP Q0ED31

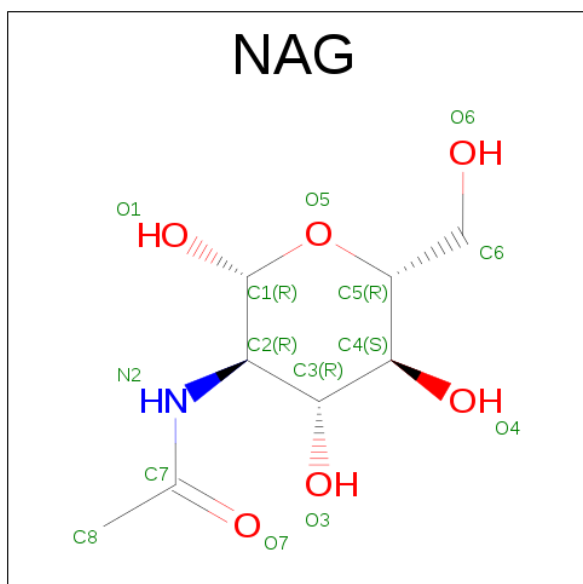
- Molecule 2 is a protein called HEAVY CHAIN OF ANTIBODY 44-VRC13.01.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	H	238	Total	C	H	N	O	S	0	0	0
			3583	1141	1787	308	339	8			
2	A	238	Total	C	H	N	O	S	0	0	0
			3582	1141	1786	308	339	8			

- Molecule 3 is a protein called LIGHT CHAIN OF ANTIBODY 44-VRC13.01.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	L	205	Total	C	H	N	O	S	0	0	0
			3052	967	1500	266	313	6			
3	B	205	Total	C	H	N	O	S	0	0	0
			3049	967	1497	266	313	6			

- Molecule 4 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



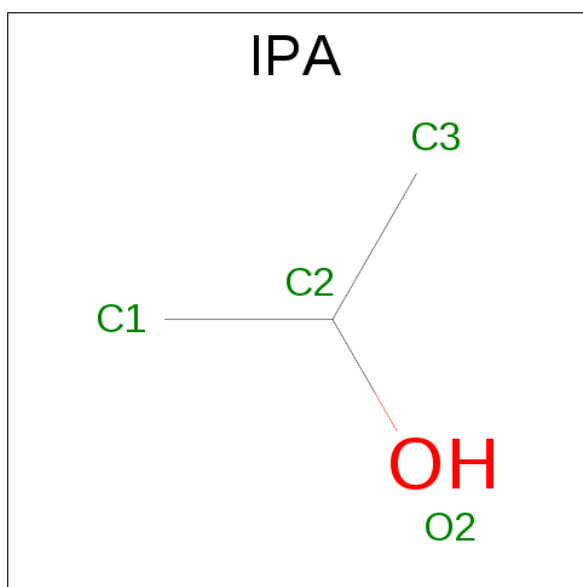
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	G	1	Total 27	C 8	H 13	N 1	O 5	0	0
4	G	1	Total 27	C 8	H 13	N 1	O 5	0	0
4	G	1	Total 27	C 8	H 13	N 1	O 5	0	0
4	G	1	Total 27	C 8	H 13	N 1	O 5	0	0
4	G	1	Total 27	C 8	H 13	N 1	O 5	0	0
4	G	1	Total 27	C 8	H 13	N 1	O 5	0	0
4	G	1	Total 27	C 8	H 13	N 1	O 5	0	0
4	G	1	Total 27	C 8	H 13	N 1	O 5	0	0
4	G	1	Total 27	C 8	H 13	N 1	O 5	0	0

*Continued on next page...*

*Continued from previous page...*

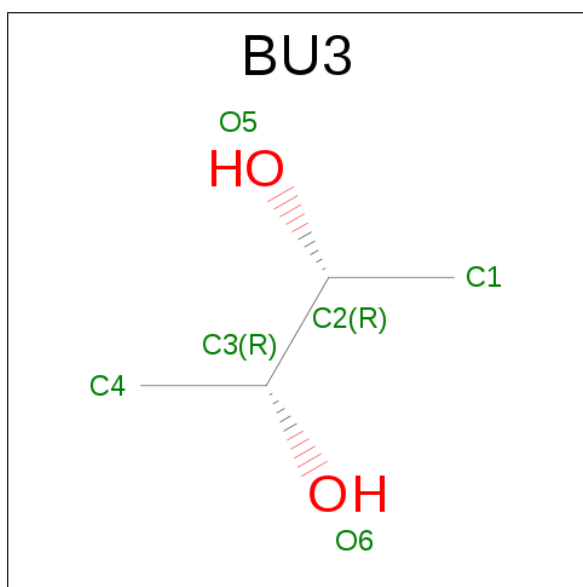
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	H	1	Total	C	H	N	O	0	0
			27	8	13	1	5		
4	L	1	Total	C	H	N	O	0	0
			27	8	13	1	5		
4	I	1	Total	C	H	N	O	0	0
			27	8	13	1	5		
4	I	1	Total	C	H	N	O	0	0
			27	8	13	1	5		
4	I	1	Total	C	H	N	O	0	0
			27	8	13	1	5		
4	I	1	Total	C	H	N	O	0	0
			27	8	13	1	5		
4	I	1	Total	C	H	N	O	0	0
			27	8	13	1	5		
4	I	1	Total	C	H	N	O	0	0
			27	8	13	1	5		
4	I	1	Total	C	H	N	O	0	0
			27	8	13	1	5		
4	A	1	Total	C	H	N	O	0	0
			27	8	13	1	5		
4	B	1	Total	C	H	N	O	0	0
			27	8	13	1	5		

- Molecule 5 is ISOPROPYL ALCOHOL (three-letter code: IPA) (formula: C<sub>3</sub>H<sub>8</sub>O).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	H	1	Total	C	H	O	0	0
			12	3	8	1		
5	L	1	Total	C	H	O	0	0
			12	3	8	1		
5	A	1	Total	C	H	O	0	0
			12	3	8	1		

- Molecule 6 is (R,R)-2,3-BUTANEDIOL (three-letter code: BU3) (formula: C<sub>4</sub>H<sub>10</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	H	1	Total	C	H	O	0	0
			16	4	10	2		

*Continued on next page...*

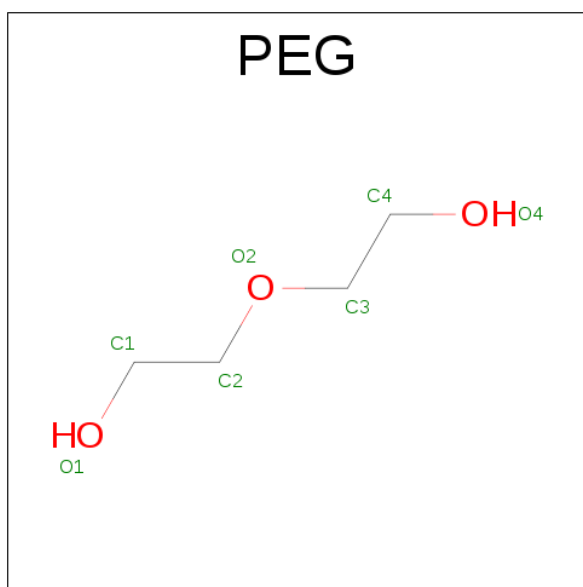
Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	A	1	Total	C	H	O	0	0
			16	4	10	2		
6	B	1	Total	C	H	O	0	0
			16	4	10	2		
6	B	1	Total	C	H	O	0	0
			16	4	10	2		

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

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

- Molecule 8 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: C<sub>4</sub>H<sub>10</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
8	A	1	Total	C	H	O	0	0
			17	4	10	3		
8	B	1	Total	C	H	O	0	0
			17	4	10	3		
8	B	1	Total	C	H	O	0	0
			17	4	10	3		

- Molecule 9 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

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

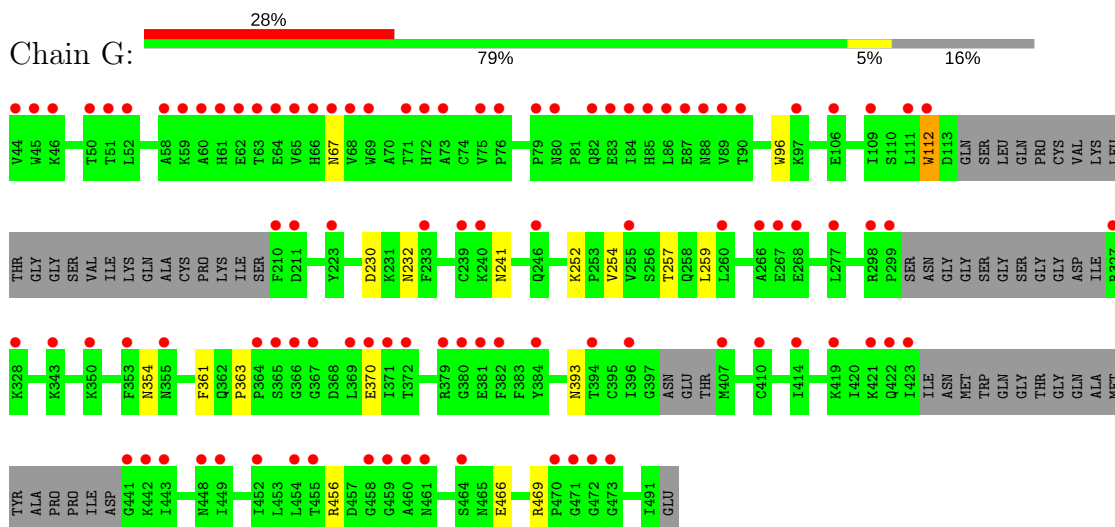
- Molecule 10 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
10	G	7	Total O 7 7	0	0
10	H	90	Total O 90 90	0	0
10	L	78	Total O 78 78	0	0
10	I	3	Total O 3 3	0	0
10	A	97	Total O 97 97	0	0
10	B	100	Total O 100 100	0	0

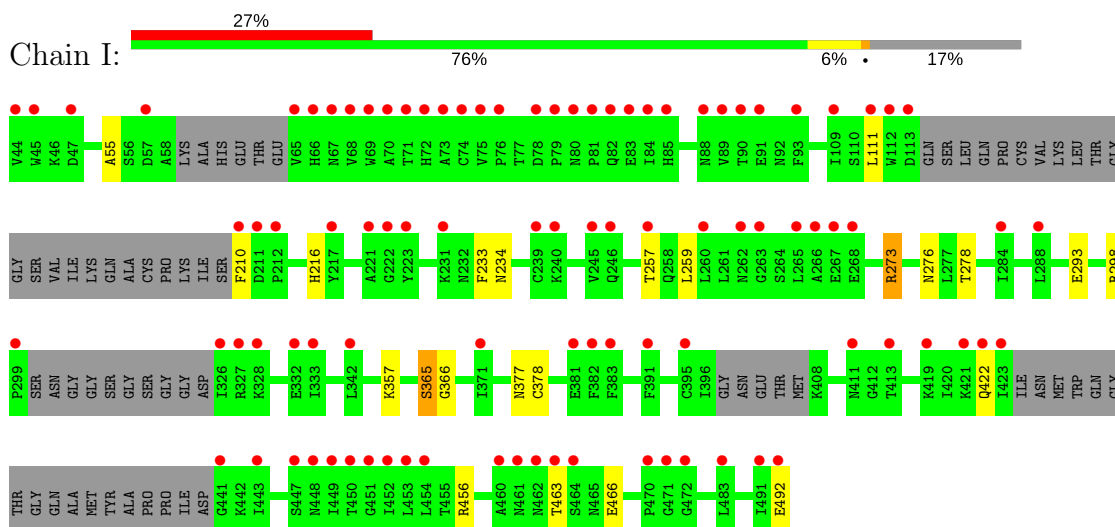
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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: Envelope glycoprotein gp160,Envelope glycoprotein gp160

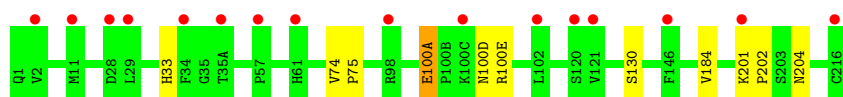


- Molecule 1: Envelope glycoprotein gp160,Envelope glycoprotein gp160

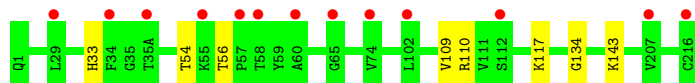


- Molecule 2: HEAVY CHAIN OF ANTIBODY 44-VRC13.01

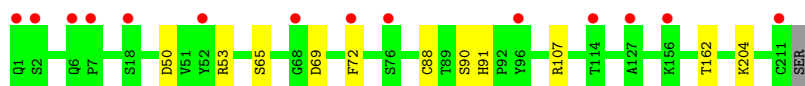




• Molecule 2: HEAVY CHAIN OF ANTIBODY 44-VRC13.01



• Molecule 3: LIGHT CHAIN OF ANTIBODY 44-VRC13.01



• Molecule 3: LIGHT CHAIN OF ANTIBODY 44-VRC13.01



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	293.80Å 67.05Å 93.64Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.30 – 2.31 38.30 – 2.31	Depositor EDS
% Data completeness (in resolution range)	98.9 (38.30-2.31) 92.8 (38.30-2.31)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.57 (at 2.31Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.223 , 0.249 0.224 , 0.250	Depositor DCC
$R_{free}$ test set	4073 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	50.3	Xtriage
Anisotropy	0.376	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 51.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	23589	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	98.0	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 28.58 % 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 1.7983e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: IPA, NAG, CL, BU3, NA, PEG

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	G	0.22	0/2398	0.37	0/3249
1	I	0.21	0/2352	0.38	0/3187
2	A	0.24	0/1844	0.42	0/2510
2	H	0.24	0/1844	0.42	0/2510
3	B	0.24	0/1591	0.41	0/2170
3	L	0.25	0/1591	0.41	0/2170
All	All	0.23	0/11620	0.40	0/15796

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 [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	G	2351	2295	2285	11	0
1	I	2307	2248	2245	11	0
2	A	1796	1786	1781	5	0
2	H	1796	1787	1782	7	0
3	B	1552	1497	1493	9	0
3	L	1552	1500	1494	7	0
4	A	14	13	13	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	14	13	13	0	0
4	G	126	117	117	1	0
4	H	14	13	13	0	0
4	I	126	117	117	0	0
4	L	14	13	13	0	0
5	A	4	8	8	1	0
5	H	4	8	8	0	0
5	L	4	8	8	1	0
6	A	6	10	10	0	0
6	B	12	20	20	4	0
6	H	6	10	10	0	0
7	L	1	0	0	0	0
8	A	7	10	10	0	0
8	B	14	20	20	0	0
9	B	1	0	0	0	0
10	A	97	0	0	3	0
10	B	100	0	0	1	0
10	G	7	0	0	0	0
10	H	90	0	0	1	1
10	I	3	0	0	0	0
10	L	78	0	0	1	1
All	All	12096	11493	11460	48	1

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.

All (48) 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:I:233:PHE:O	1:I:273:ARG:NH2	2.25	0.69
2:H:204:ASN:ND2	10:H:401:HOH:O	2.29	0.66
2:A:117:LYS:NZ	10:A:401:HOH:O	2.30	0.65
2:A:134:GLY:H	5:A:302:IPA:H12	1.67	0.60
1:I:276:ASN:OD1	1:I:278:THR:OG1	2.20	0.60
1:G:361:PHE:O	1:G:393:ASN:ND2	2.35	0.60
3:B:50:ASP:OD2	3:B:53:ARG:NH1	2.36	0.58
3:B:31:ASP:OD1	3:B:32:ARG:N	2.37	0.58
3:L:107:ARG:HD3	5:L:302:IPA:H31	1.86	0.57
3:B:138:ASP:OD1	3:B:167:GLN:NE2	2.39	0.56
3:B:149:LYS:CE	6:B:302:BU3:H11	2.37	0.55
1:I:492:GLU:OE1	1:I:492:GLU:N	2.41	0.54

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:I:377:ASN:OD1	1:I:378:CYS:N	2.41	0.54
1:I:357:LYS:NZ	1:I:463:THR:O	2.36	0.53
2:A:110:ARG:NH1	10:A:484:HOH:O	2.41	0.53
1:G:96:TRP:O	2:H:100(E):ARG:NH1	2.39	0.52
2:A:143:LYS:NZ	10:A:495:HOH:O	2.43	0.52
1:G:241:ASN:ND2	4:G:502:NAG:O7	2.45	0.50
1:G:370:GLU:OE2	1:G:370:GLU:N	2.43	0.50
1:G:230:ASP:O	1:G:232:ASN:N	2.45	0.50
3:B:201:THR:HB	6:B:303:BU3:H13	1.94	0.49
2:A:54:THR:O	2:A:56:THR:N	2.44	0.47
3:L:50:ASP:OD2	3:L:53:ARG:NH1	2.45	0.47
1:G:257:THR:O	1:G:259:LEU:N	2.46	0.46
3:B:203:GLU:OE1	6:B:302:BU3:O5	2.33	0.46
3:L:90:SER:OG	3:L:91:HIS:N	2.47	0.46
3:L:65:SER:O	3:L:72:PHE:N	2.48	0.46
1:I:293:GLU:N	1:I:293:GLU:OE1	2.47	0.46
1:G:456:ARG:NH1	1:G:466:GLU:OE2	2.49	0.45
3:B:126:GLN:NE2	10:B:403:HOH:O	2.49	0.45
2:H:74:VAL:HB	2:H:75:PRO:CD	2.47	0.45
1:I:55:ALA:N	1:I:216:HIS:O	2.51	0.44
1:I:456:ARG:NH1	1:I:466:GLU:OE2	2.49	0.44
3:B:194:GLN:NE2	3:B:203:GLU:OE2	2.49	0.44
2:H:74:VAL:HB	2:H:75:PRO:HD3	2.00	0.43
1:I:257:THR:O	1:I:259:LEU:N	2.46	0.43
3:B:149:LYS:HE3	6:B:302:BU3:H11	1.99	0.43
2:H:100(A):GLU:OE2	2:H:100(D):ASN:N	2.49	0.42
1:I:234:ASN:O	1:I:273:ARG:NE	2.52	0.42
1:I:365:SER:OG	1:I:366:GLY:N	2.51	0.42
1:G:252:LYS:O	1:G:254:VAL:N	2.50	0.42
1:G:67:ASN:OD1	1:G:67:ASN:N	2.52	0.41
2:H:201:LYS:N	2:H:202:PRO:HD2	2.35	0.41
3:L:69:ASP:OD1	3:L:69:ASP:N	2.53	0.41
3:L:162:THR:OG1	10:L:464:HOH:O	2.22	0.40
2:H:130:SER:OG	3:L:204:LYS:NZ	2.54	0.40
1:G:112:TRP:HA	1:G:112:TRP:CE3	2.56	0.40
1:G:363:PRO:O	1:G:469:ARG:NH1	2.54	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
10:H:425:HOH:O	10:L:412:HOH:O[3_453]	2.16	0.04

## 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	G	288/353 (82%)	266 (92%)	21 (7%)	1 (0%)	44	55
1	I	280/353 (79%)	253 (90%)	25 (9%)	2 (1%)	25	30
2	A	236/238 (99%)	229 (97%)	7 (3%)	0	100	100
2	H	236/238 (99%)	230 (98%)	6 (2%)	0	100	100
3	B	203/206 (98%)	191 (94%)	12 (6%)	0	100	100
3	L	203/206 (98%)	192 (95%)	11 (5%)	0	100	100
All	All	1446/1594 (91%)	1361 (94%)	82 (6%)	3 (0%)	51	63

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	G	354	ASN
1	I	422	GLN
1	I	365	SER

### 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	G	268/311 (86%)	267 (100%)	1 (0%)	93	97
1	I	264/311 (85%)	260 (98%)	4 (2%)	70	83
2	A	202/202 (100%)	200 (99%)	2 (1%)	80	90
2	H	202/202 (100%)	199 (98%)	3 (2%)	70	83
3	B	175/176 (99%)	175 (100%)	0	100	100
3	L	175/176 (99%)	174 (99%)	1 (1%)	89	95
All	All	1286/1378 (93%)	1275 (99%)	11 (1%)	82	91

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

Mol	Chain	Res	Type
1	G	112	TRP
2	H	33	HIS
2	H	100(A)	GLU
2	H	184	VAL
3	L	88	CYS
1	I	111	LEU
1	I	210	PHE
1	I	273	ARG
1	I	298	ARG
2	A	33	HIS
2	A	109	VAL

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

Mol	Chain	Res	Type
3	L	66	HIS

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry

Of 34 ligands modelled in this entry, 2 are monoatomic - leaving 32 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	NAG	A	301	2	14,14,15	0.19	0	15,19,21	0.55	0
5	IPA	A	302	-	3,3,3	0.55	0	3,3,3	0.19	0
6	BU3	A	303	-	4,5,5	1.31	0	6,6,6	0.37	0
8	PEG	A	304	-	6,6,6	0.56	0	5,5,5	0.29	0
4	NAG	B	301	3	14,14,15	0.26	0	15,19,21	0.44	0
6	BU3	B	302	-	4,5,5	1.28	0	6,6,6	0.30	0
6	BU3	B	303	-	4,5,5	1.33	0	6,6,6	0.39	0
8	PEG	B	304	-	6,6,6	0.56	0	5,5,5	0.40	0
8	PEG	B	305	-	6,6,6	0.55	0	5,5,5	0.48	0
4	NAG	G	501	1	14,14,15	0.19	0	15,19,21	0.49	0
4	NAG	G	502	1	14,14,15	0.53	0	15,19,21	0.70	1 (6%)
4	NAG	G	503	1	14,14,15	0.22	0	15,19,21	0.48	0
4	NAG	G	504	1	14,14,15	0.24	0	15,19,21	0.47	0
4	NAG	G	505	1	14,14,15	0.28	0	15,19,21	0.52	0
4	NAG	G	506	1	14,14,15	0.25	0	15,19,21	0.50	0
4	NAG	G	507	1	14,14,15	0.19	0	15,19,21	0.57	0
4	NAG	G	508	1	14,14,15	0.21	0	15,19,21	0.44	0
4	NAG	G	509	1	14,14,15	0.25	0	15,19,21	0.56	0
4	NAG	H	301	2	14,14,15	0.23	0	15,19,21	0.53	0
5	IPA	H	302	-	3,3,3	0.57	0	3,3,3	0.18	0
6	BU3	H	303	-	4,5,5	1.28	0	6,6,6	0.35	0
4	NAG	I	501	1	14,14,15	0.20	0	15,19,21	0.53	0
4	NAG	I	502	1	14,14,15	0.23	0	15,19,21	0.46	0
4	NAG	I	503	1	14,14,15	0.22	0	15,19,21	0.45	0
4	NAG	I	504	1	14,14,15	0.28	0	15,19,21	0.43	0
4	NAG	I	505	1	14,14,15	0.22	0	15,19,21	0.51	0
4	NAG	I	506	1	14,14,15	0.25	0	15,19,21	0.46	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
4	NAG	I	507	1	14,14,15	0.20	0	15,19,21	0.57	0
4	NAG	I	508	1	14,14,15	0.26	0	15,19,21	0.50	0
4	NAG	I	509	1	14,14,15	0.26	0	15,19,21	0.54	0
4	NAG	L	301	3	14,14,15	0.22	0	15,19,21	0.52	0
5	IPA	L	302	-	3,3,3	0.54	0	3,3,3	0.15	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	NAG	A	301	2	-	0/6/23/26	0/1/1/1
5	IPA	A	302	-	-	0/0/0/0	0/0/0/0
6	BU3	A	303	-	-	0/4/4/4	0/0/0/0
8	PEG	A	304	-	-	0/4/4/4	0/0/0/0
4	NAG	B	301	3	-	0/6/23/26	0/1/1/1
6	BU3	B	302	-	-	0/4/4/4	0/0/0/0
6	BU3	B	303	-	-	0/4/4/4	0/0/0/0
8	PEG	B	304	-	-	0/4/4/4	0/0/0/0
8	PEG	B	305	-	-	0/4/4/4	0/0/0/0
4	NAG	G	501	1	-	0/6/23/26	0/1/1/1
4	NAG	G	502	1	-	0/6/23/26	0/1/1/1
4	NAG	G	503	1	-	0/6/23/26	0/1/1/1
4	NAG	G	504	1	-	0/6/23/26	0/1/1/1
4	NAG	G	505	1	-	0/6/23/26	0/1/1/1
4	NAG	G	506	1	-	0/6/23/26	0/1/1/1
4	NAG	G	507	1	-	0/6/23/26	0/1/1/1
4	NAG	G	508	1	-	0/6/23/26	0/1/1/1
4	NAG	G	509	1	-	0/6/23/26	0/1/1/1
4	NAG	H	301	2	-	0/6/23/26	0/1/1/1
5	IPA	H	302	-	-	0/0/0/0	0/0/0/0
6	BU3	H	303	-	-	0/4/4/4	0/0/0/0
4	NAG	I	501	1	-	0/6/23/26	0/1/1/1
4	NAG	I	502	1	-	0/6/23/26	0/1/1/1
4	NAG	I	503	1	-	0/6/23/26	0/1/1/1
4	NAG	I	504	1	-	0/6/23/26	0/1/1/1
4	NAG	I	505	1	-	0/6/23/26	0/1/1/1
4	NAG	I	506	1	-	0/6/23/26	0/1/1/1
4	NAG	I	507	1	-	0/6/23/26	0/1/1/1
4	NAG	I	508	1	-	0/6/23/26	0/1/1/1
4	NAG	I	509	1	-	0/6/23/26	0/1/1/1
4	NAG	L	301	3	-	0/6/23/26	0/1/1/1

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	IPA	L	302	-	-	0/0/0/0	0/0/0/0

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	G	502	NAG	C1-O5-C5	2.25	115.27	112.17

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	302	IPA	1	0
6	B	302	BU3	3	0
6	B	303	BU3	1	0
4	G	502	NAG	1	0
5	L	302	IPA	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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	G	298/353 (84%)	1.72	99 (33%) 0 0	66, 119, 181, 232	0
1	I	292/353 (82%)	1.67	95 (32%) 0 0	65, 130, 194, 260	0
2	A	238/238 (100%)	0.41	13 (5%) 26 32	36, 57, 105, 153	0
2	H	238/238 (100%)	0.46	16 (6%) 19 25	37, 63, 109, 144	0
3	B	205/206 (99%)	0.31	8 (3%) 40 47	35, 56, 100, 163	0
3	L	205/206 (99%)	0.52	14 (6%) 18 24	41, 65, 132, 202	0
All	All	1476/1594 (92%)	0.93	245 (16%) 2 3	35, 78, 167, 260	0

All (245) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	72	HIS	17.6
1	G	63	THR	13.8
1	I	66	HIS	10.2
1	G	210	PHE	9.3
1	I	112	TRP	9.0
1	I	71	THR	8.8
3	B	211	CYS	8.5
1	I	210	PHE	8.4
1	G	67	ASN	8.0
1	G	112	TRP	7.9
1	I	462	ASN	7.9
1	I	67	ASN	7.9
1	G	65	VAL	7.7
1	G	60	ALA	7.6
1	I	70	ALA	7.1
1	G	61	HIS	7.1
1	I	44	VAL	6.9
1	I	68	VAL	6.8
2	A	216	CYS	6.5

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	G	423	ILE	6.1
1	I	491	ILE	5.9
1	G	44	VAL	5.9
1	G	460	ALA	5.8
1	G	71	THR	5.7
1	I	326	ILE	5.7
1	G	379	ARG	5.6
1	I	461	ASN	5.6
1	G	68	VAL	5.5
1	G	64	GLU	5.3
1	I	422	GLN	5.3
1	G	90	THR	5.2
3	L	2	SER	5.2
1	I	221	ALA	5.2
1	I	460	ALA	5.0
1	G	382	PHE	4.8
1	I	80	ASN	4.7
1	G	380	GLY	4.7
1	I	84	ILE	4.6
1	I	69	TRP	4.6
1	G	327	ARG	4.4
2	H	29	LEU	4.3
3	L	68	GLY	4.3
1	I	448	ASN	4.2
1	I	421	LYS	4.2
3	L	7	PRO	4.2
1	G	84	ILE	4.1
1	G	80	ASN	4.1
1	I	332	GLU	4.1
1	G	421	LYS	4.1
1	G	422	GLN	4.0
1	I	419	LYS	4.0
1	G	79	PRO	3.9
1	I	65	VAL	3.9
1	I	472	GLY	3.9
1	G	58	ALA	3.9
1	I	72	HIS	3.8
1	G	89	VAL	3.8
1	G	223	TYR	3.8
1	G	454	LEU	3.8
1	I	217	TYR	3.8
1	I	492	GLU	3.8

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	G	73	ALA	3.8
1	I	76	PRO	3.7
1	I	383	PHE	3.7
1	I	113	ASP	3.6
1	G	443	ILE	3.6
1	I	327	ARG	3.5
1	G	87	GLU	3.5
1	I	79	PRO	3.5
1	I	470	PRO	3.5
1	G	59	LYS	3.5
1	I	74	CYS	3.5
1	I	78	ASP	3.4
1	G	371	ILE	3.4
2	A	58	THR	3.4
3	L	1	GLN	3.3
1	G	266	ALA	3.3
1	G	239	CYS	3.3
1	I	450	THR	3.3
1	I	266	ALA	3.3
1	G	299	PRO	3.2
1	I	443	ILE	3.2
1	G	355	ASN	3.2
1	I	454	LEU	3.2
1	I	471	GLY	3.2
1	I	111	LEU	3.2
1	I	342	LEU	3.2
1	G	472	GLY	3.2
1	G	365	SER	3.1
1	G	51	THR	3.1
1	I	371	ILE	3.1
1	G	470	PRO	3.1
1	G	267	GLU	3.1
1	I	240	LYS	3.1
1	I	88	ASN	3.1
1	G	459	GLY	3.1
3	B	156	LYS	3.0
1	G	442	LYS	3.0
1	I	333	ILE	3.0
1	G	82	GLN	3.0
1	G	246	GLN	3.0
1	I	483	LEU	3.0
3	L	52	TYR	3.0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	G	211	ASP	2.9
1	I	382	PHE	2.9
1	I	239	CYS	2.9
1	I	381	GLU	2.9
1	G	75	VAL	2.9
1	I	89	VAL	2.9
1	G	367	GLY	2.9
1	G	45	TRP	2.9
1	G	381	GLU	2.8
1	G	69	TRP	2.8
1	G	50	THR	2.8
1	G	452	ILE	2.8
1	G	353	PHE	2.8
1	I	57	ASP	2.8
1	I	45	TRP	2.8
1	I	246	GLN	2.8
1	G	109	ILE	2.8
3	L	211	CYS	2.8
1	G	268	GLU	2.8
2	H	2	VAL	2.7
1	I	82	GLN	2.7
1	G	366	GLY	2.7
1	G	97	LYS	2.7
1	G	372	THR	2.7
2	H	35(A)	THR	2.7
3	L	76	SER	2.7
1	G	111	LEU	2.7
1	I	411	ASN	2.7
1	G	85	HIS	2.7
2	A	55	LYS	2.7
2	H	120	SER	2.7
1	I	73	ALA	2.7
3	B	107	ARG	2.6
2	A	112	SER	2.6
1	I	47	ASP	2.6
1	G	471	GLY	2.6
2	H	121	VAL	2.6
2	A	29	LEU	2.6
1	G	62	GLU	2.6
2	H	61	HIS	2.6
1	G	461	ASN	2.6
1	G	298	ARG	2.6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	I	75	VAL	2.6
1	G	76	PRO	2.5
1	I	268	GLU	2.5
1	G	83	GLU	2.5
1	I	211	ASP	2.5
1	I	265	LEU	2.5
2	H	11	MET	2.5
1	I	222	GLY	2.5
1	I	441	GLY	2.5
3	L	18	SER	2.5
1	G	240	LYS	2.5
1	I	90	THR	2.5
1	G	473	GLY	2.5
1	G	86	LEU	2.5
1	G	407	MET	2.5
1	I	93	PHE	2.5
3	B	96	TYR	2.5
3	B	19	ILE	2.5
1	G	419	LYS	2.4
1	I	231	LYS	2.4
1	I	223	TYR	2.4
1	I	449	ILE	2.4
1	I	85	HIS	2.4
1	G	88	ASN	2.4
2	A	102	LEU	2.4
1	I	91	GLU	2.4
2	A	34	PHE	2.4
1	G	396	ILE	2.4
1	G	66	HIS	2.4
2	H	28	ASP	2.4
3	L	156	LYS	2.4
1	G	233	PHE	2.4
3	B	91	HIS	2.3
2	A	35(A)	THR	2.3
3	L	72	PHE	2.3
1	G	458	GLY	2.3
1	I	263	GLY	2.3
1	I	267	GLU	2.3
1	G	343	LYS	2.3
2	H	57	PRO	2.3
3	B	113	PRO	2.3
1	G	394	THR	2.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	G	106	GLU	2.3
1	I	83	GLU	2.3
1	G	449	ILE	2.3
1	I	81	PRO	2.3
1	I	288	LEU	2.3
1	I	109	ILE	2.3
1	G	370	GLU	2.3
3	L	96	TYR	2.3
2	H	34	PHE	2.3
1	G	448	ASN	2.3
1	I	212	PRO	2.3
3	L	127	ALA	2.3
3	L	114	THR	2.3
1	G	441	GLY	2.3
2	A	65	GLY	2.3
1	I	299	PRO	2.3
1	G	46	LYS	2.2
1	G	255	VAL	2.2
1	I	260	LEU	2.2
2	A	207	VAL	2.2
1	I	413	THR	2.2
1	G	260	LEU	2.2
1	I	245	VAL	2.2
1	I	447	SER	2.2
1	I	464	SER	2.2
1	I	262	ASN	2.2
1	I	284	ILE	2.2
1	I	452	ILE	2.2
2	H	146	PHE	2.2
2	A	57	PRO	2.2
1	G	410	CYS	2.2
1	G	52	LEU	2.2
1	G	464	SER	2.2
1	G	350	LYS	2.1
2	H	102	LEU	2.1
1	I	257	THR	2.1
2	H	216	CYS	2.1
2	A	60	ALA	2.1
1	G	364	PRO	2.1
1	I	391	PHE	2.1
1	I	423	ILE	2.1
1	G	369	LEU	2.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	I	328	LYS	2.1
1	G	328	LYS	2.1
1	G	277	LEU	2.1
1	I	453	LEU	2.1
3	B	201	THR	2.1
1	I	395	CYS	2.1
2	H	100(C)	LYS	2.0
2	H	201	LYS	2.0
1	I	451	GLY	2.0
2	A	74	VAL	2.0
1	G	384	TYR	2.0
1	G	414	ILE	2.0
1	G	455	THR	2.0
3	L	6	GLN	2.0
1	I	463	THR	2.0
2	H	98	ARG	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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
8	PEG	A	304	7/7	0.53	0.32	11.87	89,107,109,109	0
8	PEG	B	305	7/7	0.70	0.54	9.40	93,112,117,117	0
4	NAG	L	301	14/15	0.74	0.48	5.19	94,103,124,125	0
4	NAG	B	301	14/15	0.75	0.35	3.11	79,92,109,112	0
4	NAG	I	502	14/15	0.86	0.51	3.06	93,100,120,120	0

*Continued on next page...*

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
4	NAG	I	507	14/15	0.82	0.40	2.82	96,107,128,129	0
8	PEG	B	304	7/7	0.45	0.45	2.73	111,133,134,134	0
4	NAG	I	501	14/15	0.81	0.27	2.58	84,92,110,111	0
4	NAG	G	509	14/15	0.73	0.32	2.31	81,91,109,109	0
6	BU3	B	302	6/6	0.82	0.23	2.01	80,97,98,98	0
4	NAG	I	506	14/15	0.75	0.41	1.82	104,110,132,133	0
5	IPA	L	302	4/4	0.77	0.21	1.38	78,96,97,97	0
4	NAG	I	509	14/15	0.82	0.36	1.37	81,88,105,107	0
4	NAG	G	507	14/15	0.81	0.33	1.22	79,90,107,107	0
6	BU3	A	303	6/6	0.86	0.17	1.20	105,126,127,127	0
4	NAG	G	506	14/15	0.80	0.21	0.81	81,86,103,106	0
6	BU3	B	303	6/6	0.87	0.23	0.78	104,125,126,126	0
4	NAG	G	501	14/15	0.87	0.21	0.65	77,83,99,101	0
5	IPA	A	302	4/4	0.81	0.15	0.58	64,77,79,81	0
6	BU3	H	303	6/6	0.79	0.15	0.54	96,116,117,117	0
4	NAG	I	508	14/15	0.81	0.27	0.53	72,79,94,96	0
7	NA	L	303	1/1	0.97	0.17	0.39	59,59,59,59	0
4	NAG	I	503	14/15	0.86	0.24	0.22	80,93,112,112	0
4	NAG	I	504	14/15	0.88	0.34	-0.09	83,93,113,113	0
4	NAG	G	508	14/15	0.89	0.17	-0.62	70,78,93,95	0
4	NAG	G	503	14/15	0.89	0.16	-0.77	75,82,99,100	0
5	IPA	H	302	4/4	0.90	0.15	-1.07	80,96,96,96	0
4	NAG	I	505	14/15	0.91	0.16	-1.20	75,79,94,96	0
4	NAG	G	505	14/15	0.86	0.14	-1.44	63,69,82,84	0
4	NAG	A	301	14/15	0.77	0.34	-	71,82,99,99	0
4	NAG	G	502	14/15	0.70	0.69	-	86,97,115,115	0
4	NAG	G	504	14/15	0.79	0.28	-	88,97,116,119	0
9	CL	B	306	1/1	0.78	0.12	-	106,106,106,106	0
4	NAG	H	301	14/15	0.89	0.36	-	80,91,109,110	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.