



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 8, 2020 – 07:05 AM BST

PDB ID : 4YDL  
Title : Crystal structure of broadly and potently neutralizing antibody C38-VRC18.02  
in complex with HIV-1 clade AE strain 93TH057gp120  
Authors : Zhou, T.; Srivatsan, S.; Kwong, P.D.  
Deposited on : 2015-02-22  
Resolution : 1.80 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

<https://www.wwpdb.org/validation/2017/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 : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.13.1  
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) : 2.13.1

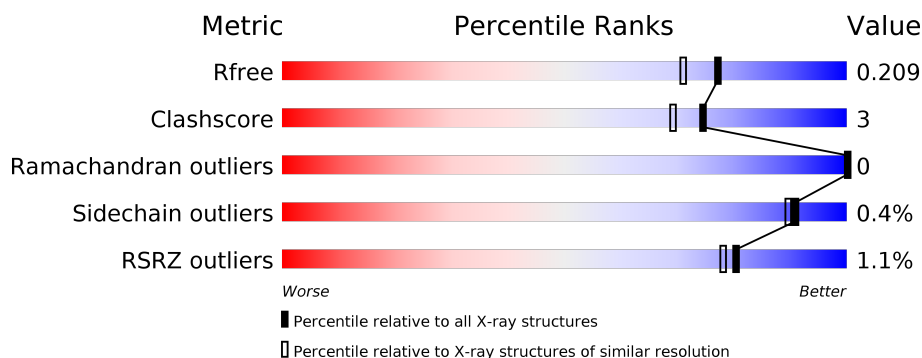
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.80 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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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 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	353	<div> <div>2%</div> <div> <div></div> <div>92%</div> <div>5%</div> <div></div> </div> </div>
1	G	353	<div> <div>%</div> <div> <div></div> <div>91%</div> <div>7%</div> <div></div> </div> </div>
2	B	226	<div> <div>%</div> <div> <div></div> <div>94%</div> <div>5%</div> <div></div> </div> </div>
2	H	226	<div> <div></div> <div> <div></div> <div>94%</div> <div>5%</div> <div></div> </div> </div>
3	C	211	<div> <div></div> <div> <div></div> <div>89%</div> <div>11%</div> <div></div> </div> </div>
3	L	211	<div> <div></div> <div> <div></div> <div>92%</div> <div>8%</div> <div></div> </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
6	SO4	C	303	-	-	X	-

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 13805 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 Envelope glycoprotein gp160,Envelope glycoprotein gp160.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	G	344	Total	C	N	O	S	0	0	0
			2689	1687	468	511	23			
1	A	343	Total	C	N	O	S	0	0	0
			2685	1685	467	510	23			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	124	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
A	124	GLY	-	linker	UNP Q0ED31
A	198	GLY	-	linker	UNP Q0ED31
A	318	GLY	-	linker	UNP Q0ED31
A	319	GLY	-	linker	UNP Q0ED31
A	320	SER	-	linker	UNP Q0ED31
A	321	GLY	-	linker	UNP Q0ED31
A	322	SER	-	linker	UNP Q0ED31
A	323	GLY	-	linker	UNP Q0ED31

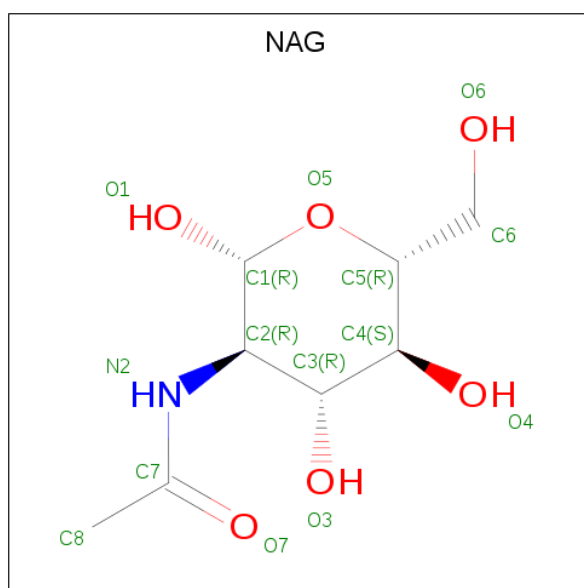
- Molecule 2 is a protein called HEAVY CHAIN OF ANTIBODY C38-VRC18.02.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	224	Total	C	N	O	S	0	0	0
			1702	1083	292	320	7			
2	B	224	Total	C	N	O	S	0	0	0
			1702	1083	292	320	7			

- Molecule 3 is a protein called LIGHT CHAIN OF ANTIBODY C38-VRC18.02.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	L	211	Total	C	N	O	S	0	0	0
			1636	1019	288	323	6			
3	C	211	Total	C	N	O	S	0	0	0
			1636	1019	288	323	6			

- Molecule 4 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



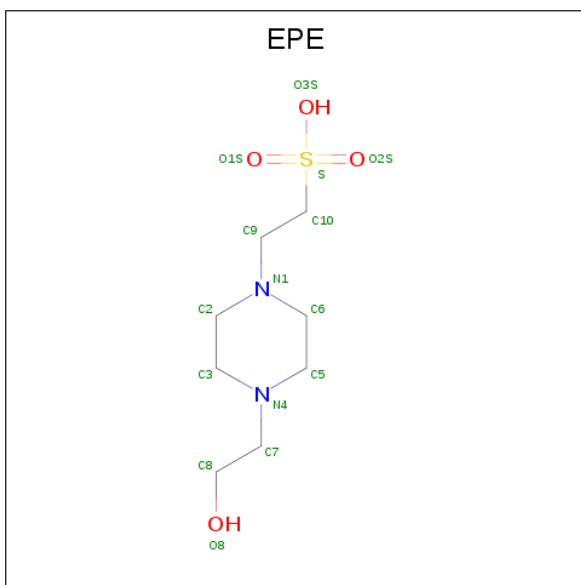
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	G	1	Total	C	N	O	0	0
			14	8	1	5		
4	G	1	Total	C	N	O	0	0
			14	8	1	5		
4	G	1	Total	C	N	O	0	0
			14	8	1	5		
4	G	1	Total	C	N	O	0	0
			14	8	1	5		
4	G	1	Total	C	N	O	0	0
			14	8	1	5		
4	G	1	Total	C	N	O	0	0
			14	8	1	5		
4	G	1	Total	C	N	O	0	0
			14	8	1	5		

*Continued on next page...*

*Continued from previous page...*

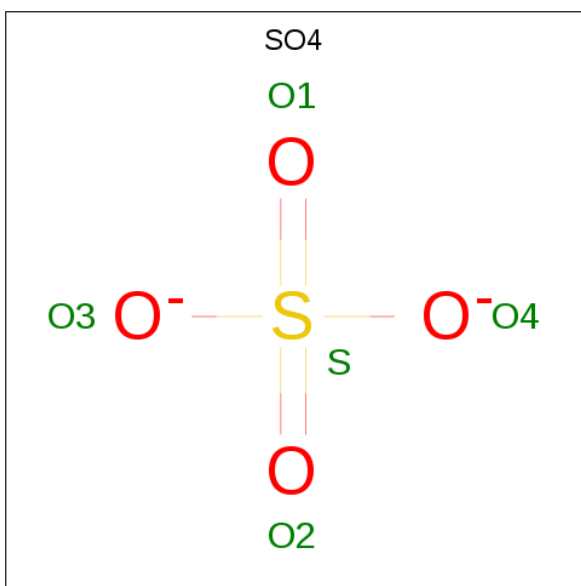
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	G	1	Total	C	N	O	0	0
			14	8	1	5		
4	G	1	Total	C	N	O	0	0
			14	8	1	5		
4	L	1	Total	C	N	O	0	0
			14	8	1	5		
4	A	1	Total	C	N	O	0	0
			14	8	1	5		
4	A	1	Total	C	N	O	0	0
			14	8	1	5		
4	A	1	Total	C	N	O	0	0
			14	8	1	5		
4	A	1	Total	C	N	O	0	0
			14	8	1	5		
4	A	1	Total	C	N	O	0	0
			14	8	1	5		
4	A	1	Total	C	N	O	0	0
			14	8	1	5		
4	A	1	Total	C	N	O	0	0
			14	8	1	5		
4	C	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 5 is 4-(2-HYDROXYETHYL)-1-PIPERAZINE ETHANESULFONIC ACID (three-letter code: EPE) (formula: C<sub>8</sub>H<sub>18</sub>N<sub>2</sub>O<sub>4</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	G	1	Total	C	N	O	S	0	0
			15	8	2	4	1		
5	A	1	Total	C	N	O	S	0	0
			15	8	2	4	1		

- Molecule 6 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



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

*Continued on next page...*

*Continued from previous page...*

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

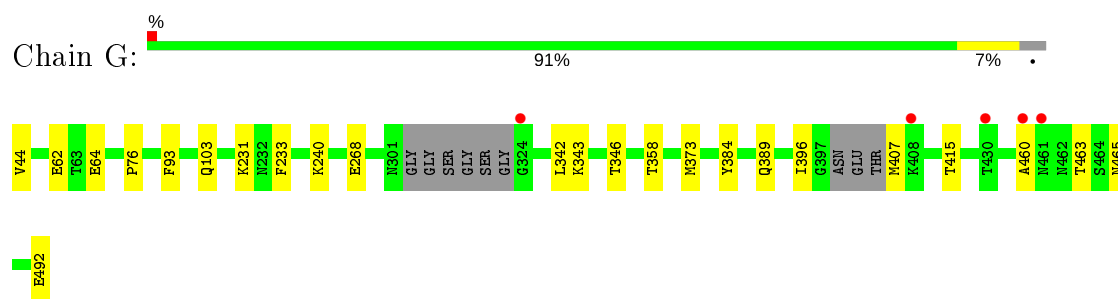
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	G	206	Total	O	0	0
			206	206		
7	H	251	Total	O	0	0
			251	251		
7	L	239	Total	O	0	0
			239	239		
7	A	215	Total	O	0	0
			215	215		
7	B	263	Total	O	0	0
			263	263		
7	C	228	Total	O	0	0
			228	228		

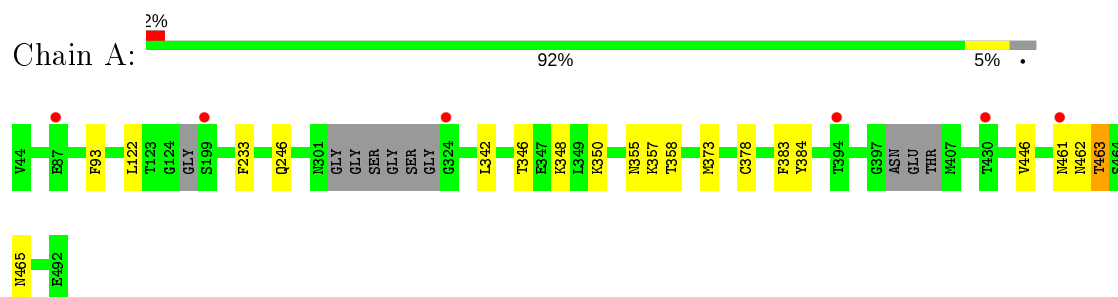
### 3 Residue-property plots [i](#)

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.

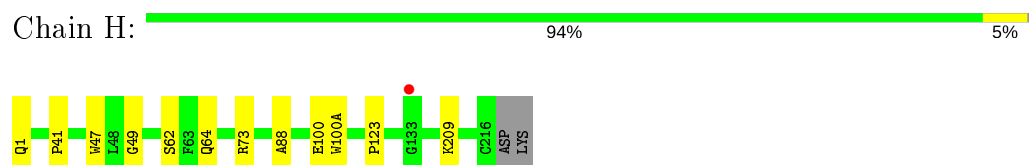
- Molecule 1: Envelope glycoprotein gp160,Envelope glycoprotein gp160



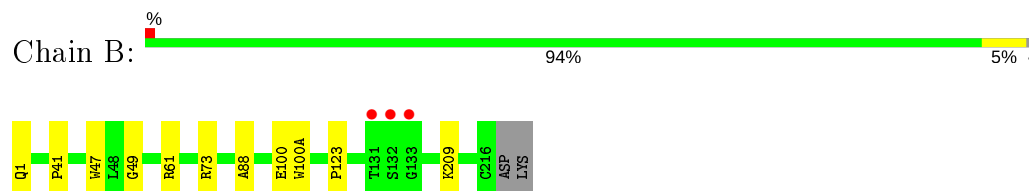
- Molecule 1: Envelope glycoprotein gp160,Envelope glycoprotein gp160



- Molecule 2: HEAVY CHAIN OF ANTIBODY C38-VRC18.02



- Molecule 2: HEAVY CHAIN OF ANTIBODY C38-VRC18.02

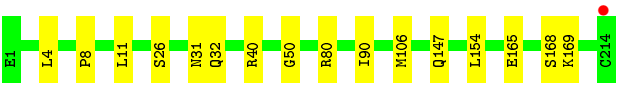


- Molecule 3: LIGHT CHAIN OF ANTIBODY C38-VRC18.02

Chain L: 

92%

8%

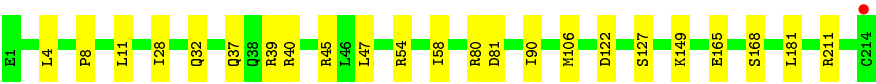


● Molecule 3: LIGHT CHAIN OF ANTIBODY C38-VRC18.02

Chain C: 

89%

11%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	66.33Å 213.09Å 68.19Å 90.00° 90.17° 90.00°	Depositor
Resolution (Å)	35.85 – 1.80 35.85 – 1.79	Depositor EDS
% Data completeness (in resolution range)	98.4 (35.85-1.80) 91.7 (35.85-1.79)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.88 (at 1.78Å)	Xtriage
Refinement program	PHENIX dev_1702	Depositor
R, $R_{free}$	0.174 , 0.207 0.177 , 0.209	Depositor DCC
$R_{free}$ test set	8793 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	28.5	Xtriage
Anisotropy	0.213	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 36.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.013 for l,k,-h 0.477 for h,-k,-l 0.018 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	13805	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.13% of the height of the origin peak. No significant pseudotranslation is detected.*

<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: EPE, PCA, NAG, 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	0/2740	0.46	0/3716
1	G	0.30	0/2745	0.46	0/3724
2	B	0.38	0/1742	0.55	0/2372
2	H	0.39	0/1742	0.53	0/2372
3	C	0.41	0/1669	0.57	0/2262
3	L	0.42	0/1669	0.56	0/2262
All	All	0.36	0/12307	0.51	0/16708

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	A	2685	0	2619	15	0
1	G	2689	0	2623	20	0
2	B	1702	0	1663	7	0
2	H	1702	0	1663	7	0
3	C	1636	0	1599	19	0
3	L	1636	0	1599	12	0
4	A	140	0	130	2	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	C	14	0	13	0	0
4	G	140	0	130	1	0
4	L	14	0	13	0	0
5	A	15	0	17	0	0
5	G	15	0	17	2	0
6	C	10	0	0	3	0
6	L	5	0	0	0	0
7	A	215	0	0	2	0
7	B	263	0	0	1	0
7	C	228	0	0	8	0
7	G	206	0	0	3	0
7	H	251	0	0	2	0
7	L	239	0	0	4	0
All	All	13805	0	12086	78	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 3.

All (78) 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)
6:C:303:SO4:O1	7:C:595:HOH:O	1.98	0.80
3:C:40:ARG:HD2	3:C:165:GLU:HG2	1.65	0.77
3:C:127:SER:OG	7:C:593:HOH:O	2.09	0.71
3:L:80:ARG:HD3	3:L:168:SER:HB2	1.73	0.69
1:G:231:LYS:HB3	1:G:268:GLU:HG3	1.75	0.68
1:G:407:MET:N	7:G:798:HOH:O	2.28	0.66
3:C:80:ARG:HD2	3:C:168:SER:HB2	1.80	0.63
3:C:122:ASP:N	6:C:303:SO4:O1	2.32	0.62
1:G:460:ALA:O	2:H:62:SER:OG	2.16	0.62
2:H:100:GLU:HG2	2:H:100(A):TRP:CD1	2.35	0.62
3:L:106:MET:SD	7:L:551:HOH:O	2.56	0.61
1:G:44:VAL:N	7:G:781:HOH:O	2.33	0.61
3:L:4:LEU:HG	3:L:90:ILE:HD12	1.83	0.61
2:B:100:GLU:HG2	2:B:100(A):TRP:CD1	2.37	0.60
3:L:169:LYS:NZ	7:L:401:HOH:O	2.34	0.59
2:H:64:GLN:NE2	7:H:488:HOH:O	2.31	0.58
3:C:39:ARG:NH2	3:C:81:ASP:OD2	2.36	0.58
3:L:40:ARG:NH1	7:L:637:HOH:O	2.38	0.57
3:L:80:ARG:NH2	7:L:402:HOH:O	2.36	0.57
1:A:350:LYS:HG2	1:A:355:ASN:HA	1.87	0.56

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:G:389:GLN:NE2	1:G:415:THR:O	2.39	0.55
1:G:463:THR:HG23	1:G:465:ASN:O	2.06	0.55
1:G:373:MET:HE3	1:G:384:TYR:HB3	1.89	0.55
1:A:463:THR:OG1	1:A:463:THR:O	2.19	0.54
3:C:165:GLU:HG3	7:C:457:HOH:O	2.08	0.54
1:G:240:LYS:NZ	7:G:802:HOH:O	2.30	0.52
1:A:358:THR:HB	1:A:465:ASN:HB3	1.92	0.51
2:B:47:TRP:CZ2	2:B:49:GLY:HA2	2.46	0.51
1:A:373:MET:HE3	1:A:384:TYR:HB3	1.93	0.51
3:C:4:LEU:HG	3:C:90:ILE:HD12	1.94	0.50
3:C:8:PRO:HG3	3:C:11:LEU:HD13	1.94	0.50
3:C:45:ARG:HD2	7:C:615:HOH:O	2.12	0.49
1:G:343:LYS:HG3	1:G:396:ILE:HG23	1.94	0.49
1:G:93:PHE:HB2	1:G:233:PHE:HZ	1.78	0.48
2:H:123:PRO:HD3	2:H:209:LYS:HE2	1.97	0.47
2:H:47:TRP:CZ2	2:H:49:GLY:HA2	2.50	0.47
3:C:40:ARG:HD2	3:C:165:GLU:CG	2.42	0.47
2:H:100:GLU:HB2	7:H:529:HOH:O	2.15	0.46
1:A:461:ASN:HB3	2:B:61:ARG:HB3	1.96	0.46
4:A:504:NAG:H61	3:C:28:ILE:HA	1.97	0.45
1:G:44:VAL:HG13	1:G:492:GLU:HB2	1.98	0.45
6:C:303:SO4:O3	7:C:566:HOH:O	2.21	0.45
2:B:100:GLU:HB2	7:B:350:HOH:O	2.17	0.45
3:L:40:ARG:HG2	3:L:165:GLU:CD	2.38	0.44
1:G:389:GLN:HG2	4:G:509:NAG:O7	2.17	0.44
1:A:461:ASN:ND2	7:A:794:HOH:O	2.43	0.44
1:G:93:PHE:HB2	1:G:233:PHE:CZ	2.52	0.44
3:C:211:ARG:NH1	7:C:573:HOH:O	2.37	0.43
1:A:246:GLN:NE2	7:A:603:HOH:O	2.50	0.43
1:A:348:LYS:HA	1:A:348:LYS:HD2	1.82	0.43
1:A:93:PHE:HB2	1:A:233:PHE:HZ	1.82	0.43
3:C:39:ARG:NH1	7:C:512:HOH:O	2.50	0.43
3:C:106:MET:HB3	3:C:106:MET:HE2	1.91	0.43
3:C:4:LEU:HD12	3:C:4:LEU:N	2.33	0.43
1:G:76:PRO:HD2	1:A:122:LEU:HD22	1.99	0.43
1:G:346:THR:HG21	1:G:396:ILE:HD13	2.01	0.43
3:L:147:GLN:CD	3:L:154:LEU:HG	2.39	0.42
3:C:54:ARG:HG2	3:C:58:ILE:HB	2.01	0.42
3:C:149:LYS:NZ	7:C:405:HOH:O	2.53	0.42
1:G:103:GLN:HB3	5:G:511:EPE:C8	2.49	0.42
3:L:8:PRO:HG2	3:L:11:LEU:HG	2.01	0.42

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:G:342:LEU:O	1:G:346:THR:HG23	2.20	0.42
1:G:358:THR:HB	1:G:465:ASN:HB3	2.02	0.41
1:A:446:VAL:HG21	4:A:506:NAG:H82	2.02	0.41
1:A:465:ASN:OD1	2:B:61:ARG:NH2	2.47	0.41
2:B:123:PRO:HD3	2:B:209:LYS:HE2	2.02	0.41
1:A:357:LYS:HD3	1:A:462:ASN:HD21	1.86	0.41
3:L:32:GLN:O	3:L:90:ILE:HA	2.19	0.41
2:H:41:PRO:HD3	2:H:88:ALA:HA	2.03	0.41
3:L:31:ASN:HA	3:L:50:GLY:HA2	2.02	0.41
1:G:103:GLN:HB3	5:G:511:EPE:O8	2.21	0.41
2:B:41:PRO:HD3	2:B:88:ALA:HA	2.02	0.40
1:G:62:GLU:HG3	1:G:64:GLU:H	1.85	0.40
3:C:37:GLN:HB2	3:C:47:LEU:HD11	2.03	0.40
1:A:342:LEU:O	1:A:346:THR:HG23	2.22	0.40
1:A:378:CYS:HB3	1:A:383:PHE:CE1	2.57	0.40
3:C:32:GLN:O	3:C:90:ILE:HA	2.22	0.40
3:L:147:GLN:OE1	3:L:154:LEU:HG	2.22	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	335/353 (95%)	326 (97%)	9 (3%)	0	100	100
1	G	338/353 (96%)	328 (97%)	10 (3%)	0	100	100
2	B	222/226 (98%)	219 (99%)	3 (1%)	0	100	100
2	H	222/226 (98%)	221 (100%)	1 (0%)	0	100	100
3	C	209/211 (99%)	208 (100%)	1 (0%)	0	100	100
3	L	209/211 (99%)	207 (99%)	2 (1%)	0	100	100
All	All	1535/1580 (97%)	1509 (98%)	26 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains ⓘ

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	306/311 (98%)	305 (100%)	1 (0%)	92	91
1	G	306/311 (98%)	306 (100%)	0	100	100
2	B	188/190 (99%)	187 (100%)	1 (0%)	88	87
2	H	188/190 (99%)	187 (100%)	1 (0%)	88	87
3	C	184/184 (100%)	183 (100%)	1 (0%)	88	87
3	L	184/184 (100%)	183 (100%)	1 (0%)	88	87
All	All	1356/1370 (99%)	1351 (100%)	5 (0%)	91	89

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	H	73	ARG
3	L	26	SER
1	A	463	THR
2	B	73	ARG
3	C	181	LEU

Some 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 ⓘ

2 non-standard protein/DNA/RNA residues 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	$\# Z  > 2$
2	PCA	B	1	2	7,8,9	1.79	1 (14%)	9,10,12	2.26	5 (55%)
2	PCA	H	1	2	7,8,9	1.81	1 (14%)	9,10,12	2.26	5 (55%)

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	PCA	B	1	2	-	0/0/11/13	0/1/1/1
2	PCA	H	1	2	-	0/0/11/13	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	H	1	PCA	CD-N	4.61	1.46	1.34
2	B	1	PCA	CD-N	4.57	1.46	1.34

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	H	1	PCA	CB-CA-C	-3.34	108.10	112.70
2	B	1	PCA	CB-CA-C	-3.16	108.35	112.70
2	B	1	PCA	OE-CD-CG	-3.08	121.38	126.76
2	H	1	PCA	CA-N-CD	-3.00	103.31	113.58
2	B	1	PCA	CA-N-CD	-2.98	103.38	113.58
2	H	1	PCA	OE-CD-CG	-2.97	121.59	126.76
2	B	1	PCA	CB-CA-N	2.68	110.98	103.30
2	H	1	PCA	CB-CA-N	2.56	110.64	103.30
2	B	1	PCA	CG-CD-N	2.49	114.84	108.39
2	H	1	PCA	CG-CD-N	2.47	114.78	108.39

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.5 Carbohydrates

There are no monosaccharides in this entry.

## 5.6 Ligand geometry

27 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	$\# Z  > 2$
4	NAG	G	504	1	14,14,15	0.57	1 (7%)	17,19,21	0.38	0
4	NAG	G	508	1	14,14,15	0.35	0	17,19,21	0.47	0
4	NAG	A	501	1	14,14,15	0.32	0	17,19,21	0.36	0
4	NAG	G	501	1	14,14,15	0.28	0	17,19,21	0.35	0
5	EPE	G	511	-	15,15,15	0.73	1 (6%)	18,20,20	1.91	5 (27%)
4	NAG	A	507	1	14,14,15	0.32	0	17,19,21	0.51	0
4	NAG	A	503	1	14,14,15	0.36	0	17,19,21	0.49	0
6	SO4	L	302	-	4,4,4	0.14	0	6,6,6	0.10	0
4	NAG	A	508	1	14,14,15	0.36	0	17,19,21	0.54	0
4	NAG	A	506	1	14,14,15	0.38	0	17,19,21	0.52	0
4	NAG	A	509	1	14,14,15	0.30	0	17,19,21	0.62	1 (5%)
4	NAG	C	301	3	14,14,15	0.18	0	17,19,21	0.53	0
4	NAG	G	510	1	14,14,15	0.28	0	17,19,21	0.66	1 (5%)
4	NAG	A	504	1	14,14,15	0.31	0	17,19,21	0.36	0
4	NAG	L	301	3	14,14,15	0.20	0	17,19,21	0.45	0
4	NAG	G	505	1	14,14,15	0.32	0	17,19,21	0.54	0
4	NAG	G	509	1	14,14,15	0.36	0	17,19,21	0.61	1 (5%)
4	NAG	A	502	1	14,14,15	0.58	0	17,19,21	0.48	0
4	NAG	A	510	1	14,14,15	0.28	0	17,19,21	0.96	1 (5%)
5	EPE	A	511	-	15,15,15	0.82	1 (6%)	18,20,20	1.77	4 (22%)
6	SO4	C	302	-	4,4,4	0.14	0	6,6,6	0.05	0
4	NAG	G	503	1	14,14,15	0.41	0	17,19,21	0.50	0
6	SO4	C	303	-	4,4,4	0.14	0	6,6,6	0.10	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
4	NAG	G	502	1	14,14,15	0.36	0	17,19,21	0.46	0
4	NAG	A	505	1	14,14,15	0.37	0	17,19,21	0.61	1 (5%)
4	NAG	G	506	1	14,14,15	0.27	0	17,19,21	0.60	0
4	NAG	G	507	1	14,14,15	0.24	0	17,19,21	0.61	1 (5%)

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	G	504	1	-	0/6/23/26	0/1/1/1
4	NAG	G	508	1	-	2/6/23/26	0/1/1/1
4	NAG	A	501	1	-	0/6/23/26	0/1/1/1
4	NAG	G	501	1	-	0/6/23/26	0/1/1/1
5	EPE	G	511	-	-	2/9/19/19	0/1/1/1
4	NAG	A	507	1	-	0/6/23/26	0/1/1/1
4	NAG	A	503	1	-	0/6/23/26	0/1/1/1
4	NAG	A	508	1	-	0/6/23/26	0/1/1/1
4	NAG	A	506	1	-	2/6/23/26	0/1/1/1
4	NAG	A	509	1	-	2/6/23/26	0/1/1/1
4	NAG	C	301	3	-	2/6/23/26	0/1/1/1
4	NAG	G	510	1	-	2/6/23/26	0/1/1/1
4	NAG	A	504	1	-	1/6/23/26	0/1/1/1
4	NAG	L	301	3	-	0/6/23/26	0/1/1/1
4	NAG	G	505	1	-	1/6/23/26	0/1/1/1
4	NAG	G	509	1	-	2/6/23/26	0/1/1/1
4	NAG	A	502	1	-	4/6/23/26	0/1/1/1
4	NAG	A	510	1	-	0/6/23/26	0/1/1/1
5	EPE	A	511	-	-	5/9/19/19	0/1/1/1
4	NAG	G	503	1	-	2/6/23/26	0/1/1/1
4	NAG	G	502	1	-	2/6/23/26	0/1/1/1
4	NAG	A	505	1	-	0/6/23/26	0/1/1/1
4	NAG	G	506	1	-	1/6/23/26	0/1/1/1
4	NAG	G	507	1	-	0/6/23/26	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	511	EPE	C10-S	2.75	1.81	1.77
5	G	511	EPE	C10-S	2.42	1.80	1.77
4	G	504	NAG	O5-C1	-2.08	1.40	1.43

All (15) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	G	511	EPE	C5-N4-C3	5.02	120.14	108.83
5	A	511	EPE	C5-N4-C3	4.48	118.92	108.83
4	A	510	NAG	C1-O5-C5	3.63	117.11	112.19
5	G	511	EPE	C5-C6-N1	-3.42	103.63	110.64
5	A	511	EPE	O2S-S-C10	2.87	110.38	106.92
5	A	511	EPE	C6-N1-C2	2.60	114.69	108.83
5	A	511	EPE	C2-C3-N4	2.42	115.61	110.64
5	G	511	EPE	O3S-S-C10	2.41	109.67	105.77
4	G	510	NAG	C1-O5-C5	2.29	115.29	112.19
4	A	509	NAG	C1-O5-C5	2.15	115.11	112.19
5	G	511	EPE	C7-N4-C5	2.13	116.69	111.23
4	G	507	NAG	C1-O5-C5	2.13	115.07	112.19
4	G	509	NAG	C1-O5-C5	2.05	114.97	112.19
5	G	511	EPE	C7-N4-C3	2.04	116.45	111.23
4	A	505	NAG	C1-O5-C5	2.03	114.94	112.19

There are no chirality outliers.

All (30) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	C	301	NAG	O5-C5-C6-O6
4	A	506	NAG	O5-C5-C6-O6
4	A	506	NAG	C4-C5-C6-O6
4	G	509	NAG	C4-C5-C6-O6
4	G	509	NAG	O5-C5-C6-O6
4	A	502	NAG	C4-C5-C6-O6
4	C	301	NAG	C4-C5-C6-O6
4	G	510	NAG	O5-C5-C6-O6
4	A	502	NAG	C8-C7-N2-C2
4	A	502	NAG	O7-C7-N2-C2
4	G	502	NAG	C4-C5-C6-O6
4	G	503	NAG	O5-C5-C6-O6
4	G	510	NAG	C4-C5-C6-O6
4	A	502	NAG	O5-C5-C6-O6
4	G	503	NAG	C4-C5-C6-O6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms
5	A	511	EPE	C9-C10-S-O3S
4	G	502	NAG	O5-C5-C6-O6
5	A	511	EPE	C8-C7-N4-C3
4	A	509	NAG	O5-C5-C6-O6
4	A	504	NAG	O5-C5-C6-O6
4	G	505	NAG	O5-C5-C6-O6
4	A	509	NAG	C4-C5-C6-O6
5	A	511	EPE	C9-C10-S-O1S
5	A	511	EPE	C9-C10-S-O2S
4	G	508	NAG	O5-C5-C6-O6
4	G	508	NAG	C4-C5-C6-O6
5	G	511	EPE	N4-C7-C8-O8
4	G	506	NAG	C3-C2-N2-C7
5	G	511	EPE	C10-C9-N1-C2
5	A	511	EPE	C10-C9-N1-C6

There are no ring outliers.

5 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	G	511	EPE	2	0
4	A	506	NAG	1	0
4	A	504	NAG	1	0
4	G	509	NAG	1	0
6	C	303	SO4	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

### 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	A	343/353 (97%)	-0.28	6 (1%) 70 66	29, 50, 85, 127	0
1	G	344/353 (97%)	-0.27	5 (1%) 73 70	29, 50, 88, 122	0
2	B	223/226 (98%)	-0.55	3 (1%) 77 74	22, 33, 60, 95	0
2	H	223/226 (98%)	-0.55	1 (0%) 92 90	23, 33, 58, 96	0
3	C	211/211 (100%)	-0.55	1 (0%) 91 89	21, 33, 54, 119	0
3	L	211/211 (100%)	-0.55	1 (0%) 91 89	22, 34, 54, 107	0
All	All	1555/1580 (98%)	-0.43	17 (1%) 80 78	21, 40, 77, 127	0

All (17) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	324	GLY	6.8
2	H	133	GLY	5.2
2	B	133	GLY	4.8
3	C	214	CYS	4.7
1	A	430	THR	4.7
1	G	460	ALA	4.6
1	A	461	ASN	3.4
2	B	132	SER	3.3
1	G	461	ASN	2.6
3	L	214	CYS	2.6
1	A	87	GLU	2.5
1	A	394	THR	2.3
1	G	408	LYS	2.2
1	G	430	THR	2.2
1	A	199	SER	2.2
2	B	131	THR	2.1
1	G	324	GLY	2.1

## 6.2 Non-standard residues in protein, DNA, RNA chains [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<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	B-factors(Å <sup>2</sup> )	Q<0.9
2	PCA	H	1	8/9	0.95	0.10	41,57,63,71	0
2	PCA	B	1	8/9	0.96	0.11	40,55,62,69	0

## 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<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	B-factors(Å <sup>2</sup> )	Q<0.9
4	NAG	A	502	14/15	0.81	0.34	84,97,106,108	0
4	NAG	A	510	14/15	0.81	0.36	83,97,107,110	0
6	SO4	L	302	5/5	0.83	0.35	128,129,129,129	0
4	NAG	G	510	14/15	0.84	0.22	64,77,79,81	0
4	NAG	A	509	14/15	0.85	0.23	80,92,97,99	0
4	NAG	G	507	14/15	0.85	0.23	54,67,68,70	0
4	NAG	G	504	14/15	0.86	0.15	39,55,63,64	0
4	NAG	G	502	14/15	0.87	0.16	66,78,87,89	0
4	NAG	G	509	14/15	0.87	0.19	78,90,96,99	0
4	NAG	L	301	14/15	0.89	0.13	42,58,63,64	0
4	NAG	C	301	14/15	0.89	0.12	41,53,61,62	0
4	NAG	A	504	14/15	0.89	0.14	42,59,69,70	0
4	NAG	G	506	14/15	0.90	0.13	52,66,69,74	0
4	NAG	A	506	14/15	0.90	0.14	54,65,74,77	0
6	SO4	C	302	5/5	0.91	0.27	118,119,120,121	0
4	NAG	A	507	14/15	0.91	0.14	49,62,68,70	0
4	NAG	A	508	14/15	0.91	0.11	47,54,60,64	0
4	NAG	G	505	14/15	0.91	0.10	43,57,64,64	0
4	NAG	A	505	14/15	0.92	0.12	44,54,66,67	0
4	NAG	G	503	14/15	0.93	0.08	38,47,53,55	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
6	SO4	C	303	5/5	0.93	0.21	115,116,117,118	0
4	NAG	G	508	14/15	0.93	0.09	43,54,64,65	0
4	NAG	A	503	14/15	0.94	0.07	34,45,51,53	0
4	NAG	G	501	14/15	0.95	0.10	44,53,56,56	0
4	NAG	A	501	14/15	0.95	0.11	42,51,55,55	0
5	EPE	G	511	15/15	0.96	0.09	45,49,55,63	0
5	EPE	A	511	15/15	0.97	0.11	43,50,59,60	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.