



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 15, 2017 – 01:35 am GMT

PDB ID : 3HUJ  
Title : Crystal structure of human CD1d-alpha-Galactosylceramide in complex with semi-invariant NKT cell receptor  
Authors : Pang, S.S.  
Deposited on : 2009-06-14  
Resolution : 2.50 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

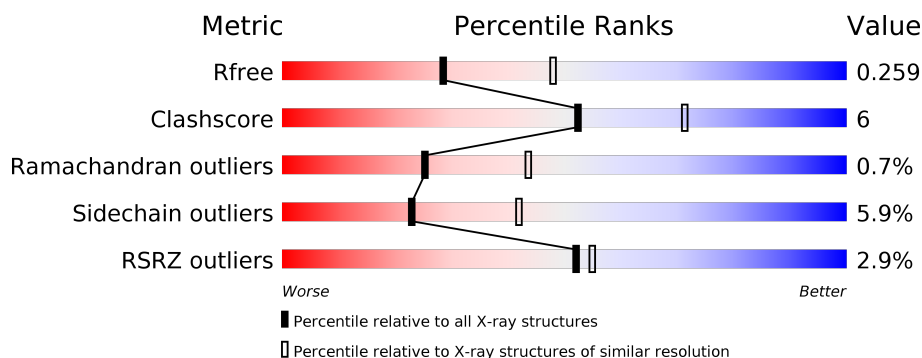
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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	3846 (2.50-2.50)
Clashscore	112137	4554 (2.50-2.50)
Ramachandran outliers	110173	4463 (2.50-2.50)
Sidechain outliers	110143	4465 (2.50-2.50)
RSRZ outliers	101464	3876 (2.50-2.50)

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	A	284	<div> <div>2%</div> <div> <div></div> <div>77%</div> <div>19%</div> <div>..</div> </div> </div>
1	C	284	<div> <div>7%</div> <div> <div></div> <div>79%</div> <div>16%</div> <div>..</div> </div> </div>
2	B	99	<div> <div></div> <div> <div></div> <div>80%</div> <div>17%</div> <div>..</div> </div> </div>
2	D	99	<div> <div>%</div> <div> <div></div> <div>82%</div> <div>16%</div> <div>.</div> </div> </div>
3	E	209	<div> <div>3%</div> <div> <div></div> <div>81%</div> <div>11%</div> <div>.. 5%</div> </div> </div>
3	G	209	<div> <div>6%</div> <div> <div></div> <div>79%</div> <div>13%</div> <div>..</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	F	246	<div><div></div><div>82%</div><div>15%</div><div>• •</div></div>
4	H	246	<div>%<div><div></div><div>84%</div><div>14%</div><div>•</div></div></div>

## 2 Entry composition

There are 11 unique types of molecules in this entry. The entry contains 13690 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 T-cell surface glycoprotein CD1d.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	275	Total	C	N	O	S	0	0	0
			2181	1398	384	392	7			
1	C	274	Total	C	N	O	S	0	0	0
			2147	1378	374	388	7			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	SER	-	EXPRESSION TAG	UNP P15813
A	1	PRO	-	EXPRESSION TAG	UNP P15813
A	2	GLY	-	EXPRESSION TAG	UNP P15813
A	278	HIS	-	EXPRESSION TAG	UNP P15813
A	279	HIS	-	EXPRESSION TAG	UNP P15813
A	280	HIS	-	EXPRESSION TAG	UNP P15813
A	281	HIS	-	EXPRESSION TAG	UNP P15813
A	282	HIS	-	EXPRESSION TAG	UNP P15813
A	283	HIS	-	EXPRESSION TAG	UNP P15813
C	0	SER	-	EXPRESSION TAG	UNP P15813
C	1	PRO	-	EXPRESSION TAG	UNP P15813
C	2	GLY	-	EXPRESSION TAG	UNP P15813
C	278	HIS	-	EXPRESSION TAG	UNP P15813
C	279	HIS	-	EXPRESSION TAG	UNP P15813
C	280	HIS	-	EXPRESSION TAG	UNP P15813
C	281	HIS	-	EXPRESSION TAG	UNP P15813
C	282	HIS	-	EXPRESSION TAG	UNP P15813
C	283	HIS	-	EXPRESSION TAG	UNP P15813

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	97	Total	C	N	O	S	0	0	0
			805	514	138	151	2			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	99	Total	C	N	O	S	0	1	0
			830	527	140	160	3			

- Molecule 3 is a protein called NKT15 T cell receptor alpha-chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	199	Total	C	N	O	S	0	1	0
			1562	967	265	321	9			
3	G	202	Total	C	N	O	S	0	1	0
			1566	973	261	323	9			

- Molecule 4 is a protein called NKT15 T cell receptor beta-chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	F	242	Total	C	N	O	S	0	1	0
			1940	1222	336	374	8			
4	H	245	Total	C	N	O	S	0	1	0
			1964	1236	339	380	9			

- Molecule 5 is a polymer of unknown type called SUGAR (2-MER).

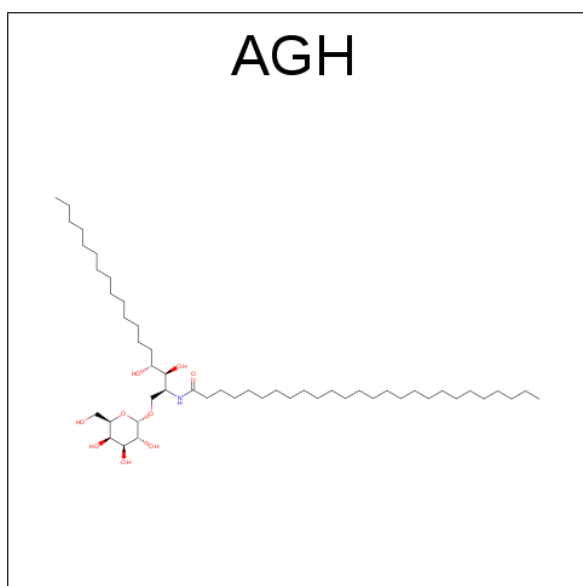
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	2	Total	C	N	O	0	0
			25	14	1	10		

- Molecule 6 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



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

- Molecule 7 is SUGAR (N-{(1S,2R,3S)-1-[(ALPHA-D-GALACTOPYRANOSYLOXY)MET HYL]-2,3-DIHYDROXYHEPTADECYL}HEXACOSANAMIDE) (three-letter code: AGH) (formula: C<sub>50</sub>H<sub>99</sub>NO<sub>9</sub>).

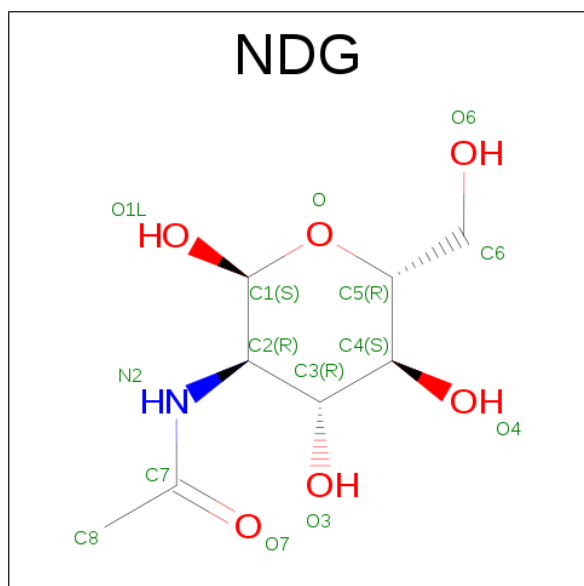


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
7	A	1	Total	C	N	O	0	0
			60	50	1	9		
7	C	1	Total	C	N	O	0	0
			60	50	1	9		

- Molecule 8 is a polymer of unknown type called SUGAR (3-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
8	C	3	Total	C	N	O	0	0
			39	22	2	15		

- Molecule 9 is SUGAR (2-(ACETYLAMINO)-2-DEOXY-A-D-GLUCOPYRANOSE) (three-letter code: NDG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



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

- Molecule 10 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	H	1	Total	Mg	0	0
			1	1		

- Molecule 11 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
11	A	95	Total	O	0	0
			95	95		
11	B	42	Total	O	0	0
			42	42		
11	C	88	Total	O	0	0
			88	88		

Continued on next page...

*Continued from previous page...*

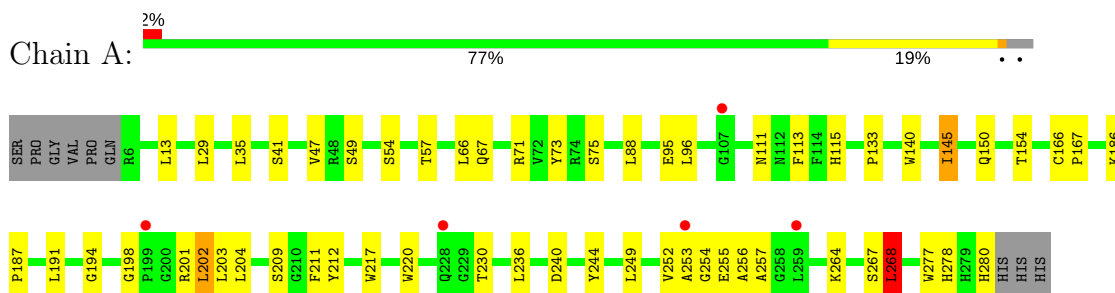
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
11	D	35	Total 35	O 35	0	0
11	E	40	Total 40	O 40	0	0
11	F	52	Total 52	O 52	0	0
11	G	62	Total 62	O 62	0	0
11	H	68	Total 68	O 68	0	0



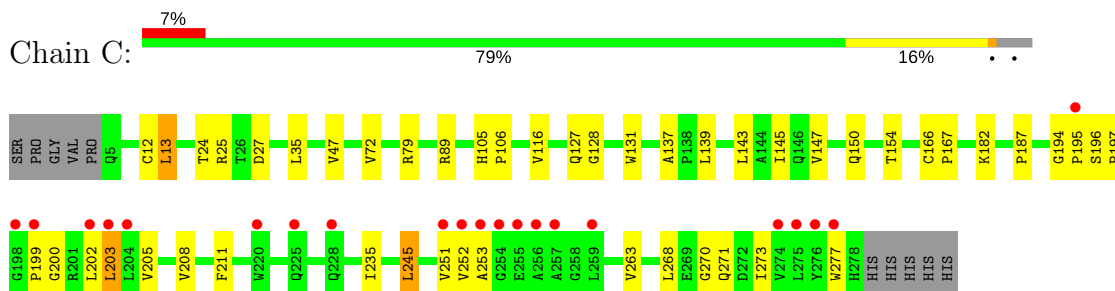
### 3 Residue-property plots [i](#)

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.

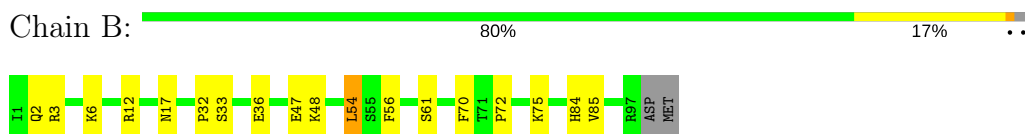
- Molecule 1: T-cell surface glycoprotein CD1d



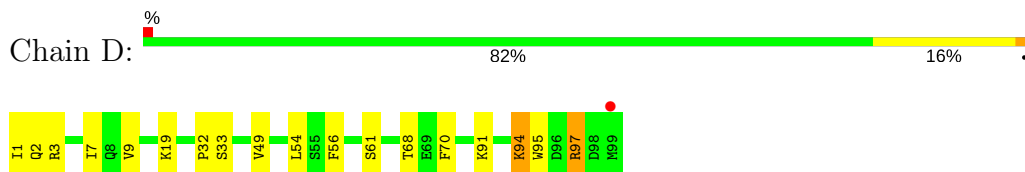
- Molecule 1: T-cell surface glycoprotein CD1d



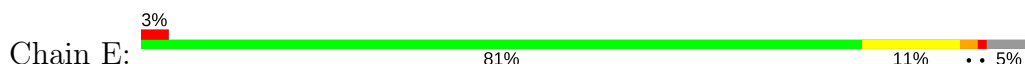
- Molecule 2: Beta-2-microglobulin

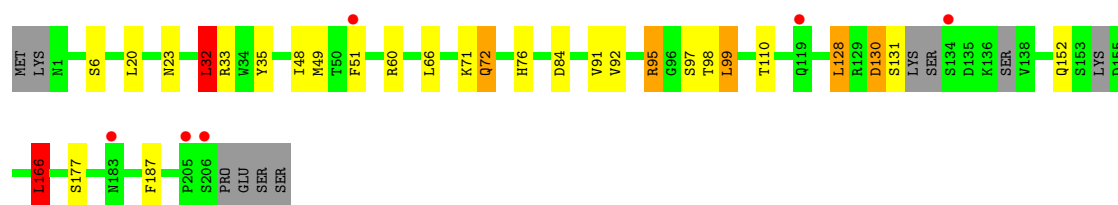


- Molecule 2: Beta-2-microglobulin

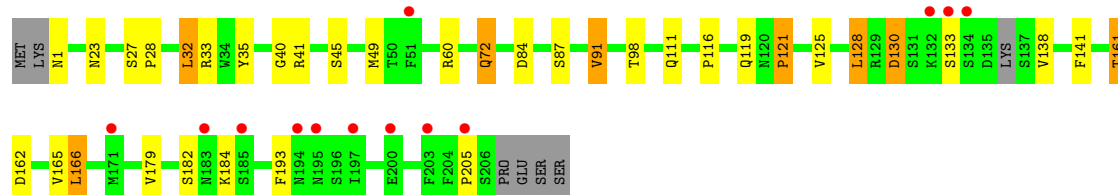
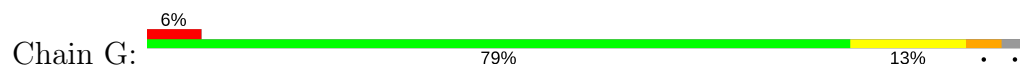


- Molecule 3: NKT15 T cell receptor alpha-chain

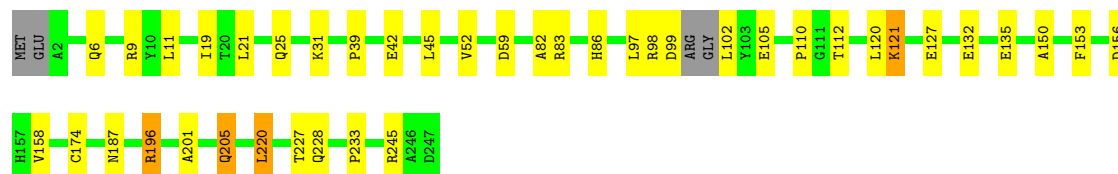
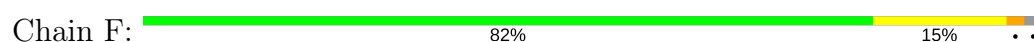




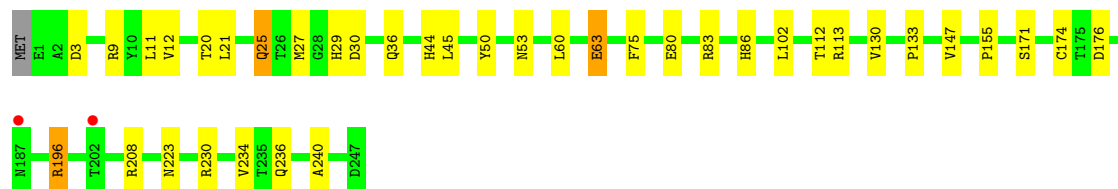
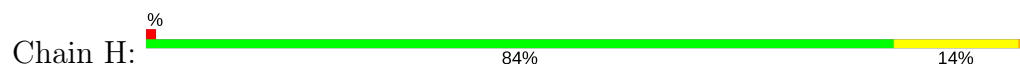
- Molecule 3: NKT15 T cell receptor alpha-chain



- Molecule 4: NKT15 T cell receptor beta-chain



- Molecule 4: NKT15 T cell receptor beta-chain



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	112.11Å 82.36Å 117.18Å 90.00° 101.26° 90.00°	Depositor
Resolution (Å)	47.28 – 2.50 47.28 – 2.50	Depositor EDS
% Data completeness (in resolution range)	90.5 (47.28-2.50) 90.5 (47.28-2.50)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.03 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.216 , 0.279 0.233 , 0.259	Depositor DCC
$R_{free}$ test set	3288 reflections (5.27%)	DCC
Wilson B-factor (Å <sup>2</sup> )	41.9	Xtriage
Anisotropy	0.207	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 18.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.057 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	13690	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.74% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MG, BMA, NAG, NDG, AGH

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	0.41	0/2249	0.57	0/3066
1	C	0.39	0/2213	0.53	0/3022
2	B	0.42	0/828	0.54	0/1122
2	D	0.43	0/853	0.57	0/1156
3	E	0.40	0/1590	0.59	3/2156 (0.1%)
3	G	0.41	0/1598	0.59	2/2176 (0.1%)
4	F	0.38	0/1993	0.53	0/2712
4	H	0.41	0/2018	0.54	0/2747
All	All	0.40	0/13342	0.56	5/18157 (0.0%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	E	32	LEU	CA-CB-CG	7.54	132.64	115.30
3	G	32	LEU	CA-CB-CG	7.18	131.82	115.30
3	G	166	LEU	CA-CB-CG	6.26	129.70	115.30
3	E	166	LEU	CA-CB-CG	5.42	127.76	115.30
3	E	128	LEU	CA-CB-CG	5.31	127.51	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	2181	0	2091	42	0
1	C	2147	0	2040	25	0
2	B	805	0	768	8	0
2	D	830	0	779	10	0
3	E	1562	0	1481	20	0
3	G	1566	0	1465	24	0
4	F	1940	0	1833	26	0
4	H	1964	0	1859	21	0
5	A	25	0	22	0	0
6	A	14	0	13	0	0
7	A	60	0	99	2	0
7	C	60	0	99	4	0
8	C	39	0	34	1	0
9	C	14	0	13	0	0
10	H	1	0	0	0	0
11	A	95	0	0	2	0
11	B	42	0	0	0	0
11	C	88	0	0	2	0
11	D	35	0	0	1	0
11	E	40	0	0	2	0
11	F	52	0	0	1	0
11	G	62	0	0	1	0
11	H	68	0	0	8	0
All	All	13690	0	12596	162	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 6.

All (162) 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:267:SER:HB2	1:A:268:LEU:HB2	1.29	1.15
1:A:253:ALA:HB1	1:A:254:GLY:HA3	1.11	1.05
1:A:253:ALA:CB	1:A:254:GLY:HA3	1.90	1.01
4:F:98:ARG:HB3	4:F:99:ASP:HB2	1.45	0.97
2:D:94:LYS:H	2:D:94:LYS:HE2	1.35	0.91
1:A:253:ALA:HB1	1:A:254:GLY:CA	2.02	0.86
1:C:105:HIS:HB3	1:C:106:PRO:HD2	1.59	0.85
3:G:161:THR:HG22	3:G:179:VAL:H	1.40	0.84
4:H:230:ARG:HG2	11:H:422:HOH:O	1.83	0.79
1:A:255:GLU:HA	1:A:257:ALA:HB3	1.64	0.79
1:A:267:SER:CB	1:A:268:LEU:HB2	2.11	0.78

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:154:THR:HG23	7:C:3000:AGH:HAB2	1.64	0.77
4:F:42:GLU:HG2	11:F:341:HOH:O	1.85	0.77
1:C:202:LEU:HG	1:C:253:ALA:HA	1.67	0.75
1:A:253:ALA:CB	1:A:254:GLY:CA	2.62	0.74
3:G:49:MET:SD	11:G:432:HOH:O	2.46	0.73
1:A:187:PRO:HB3	1:A:211:PHE:HB3	1.73	0.70
3:G:130:ASP:HB2	3:G:133:SER:O	1.91	0.69
3:G:72:GLN:HA	3:G:72:GLN:HE21	1.57	0.69
4:H:3:ASP:HB2	11:H:255:HOH:O	1.93	0.68
1:A:267:SER:H	1:A:268:LEU:HB3	1.57	0.68
1:C:72:VAL:HB	11:C:558:HOH:O	1.93	0.68
3:E:60:ARG:NH2	3:E:84:ASP:OD1	2.25	0.68
4:H:27[B]:MET:HB3	11:H:268:HOH:O	1.95	0.67
1:A:267:SER:HB2	1:A:268:LEU:CB	2.18	0.64
1:A:166:CYS:HB3	1:A:167:PRO:HD3	1.80	0.64
4:H:25:GLN:HE22	4:H:29:HIS:HB2	1.60	0.64
3:G:35:TYR:CE2	3:G:91:VAL:HG11	2.33	0.63
1:A:47:VAL:HB	1:A:67:GLN:HE22	1.63	0.63
4:F:83:ARG:H	4:F:86:HIS:HD2	1.47	0.62
3:E:33:ARG:HB2	3:E:91:VAL:HG13	1.82	0.61
3:G:23:ASN:ND2	3:G:72:GLN:HE22	1.97	0.61
1:A:217:TRP:HB3	1:A:264:LYS:HB2	1.81	0.61
4:H:83:ARG:HG2	4:H:86:HIS:CE1	2.35	0.61
1:A:267:SER:N	1:A:268:LEU:HB3	2.15	0.61
2:B:33:SER:HB2	2:B:54:LEU:HD11	1.82	0.60
1:A:209:SER:HB2	1:A:244:TYR:HD1	1.66	0.60
1:A:71:ARG:HD2	11:A:286:HOH:O	2.02	0.60
4:F:83:ARG:H	4:F:86:HIS:CD2	2.20	0.59
2:B:47:GLU:HG3	3:E:71:LYS:HG3	1.85	0.59
3:E:76:HIS:HB2	11:E:327:HOH:O	2.03	0.59
1:C:251:VAL:HG12	1:C:252:VAL:N	2.16	0.59
1:A:154:THR:HG23	7:A:3000:AGH:HAB2	1.85	0.58
4:F:98:ARG:H	4:F:99:ASP:C	2.07	0.57
2:D:7:ILE:HD12	2:D:91:LYS:HD3	1.88	0.56
1:C:166:CYS:HB3	1:C:167:PRO:HD3	1.87	0.56
1:C:105:HIS:HB3	1:C:106:PRO:CD	2.35	0.55
1:C:187:PRO:HB3	1:C:211:PHE:HB3	1.87	0.55
1:A:133:PRO:HD3	1:A:145:ILE:HG12	1.89	0.54
1:A:186:LYS:HG2	1:A:267:SER:HB3	1.89	0.54
4:F:98:ARG:HB3	4:F:99:ASP:CB	2.28	0.54
3:E:98:THR:O	4:F:31:LYS:NZ	2.31	0.54

Continued on next page...

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:198:GLY:HA3	1:A:201:ARG:H	1.72	0.54
1:C:251:VAL:HG12	1:C:252:VAL:H	1.72	0.54
4:H:63:GLU:HG3	11:H:254:HOH:O	2.08	0.54
3:G:33:ARG:HB2	3:G:91:VAL:HG13	1.90	0.53
3:E:35:TYR:HE1	3:E:91:VAL:HG11	1.74	0.53
4:F:120:LEU:HD22	4:F:220:LEU:HD11	1.91	0.53
1:A:240:ASP:OD1	2:B:12:ARG:NH1	2.43	0.52
3:E:51:PHE:HB2	11:E:213:HOH:O	2.09	0.52
1:A:150:GLN:NE2	3:G:98:THR:H	2.06	0.52
3:G:35:TYR:HD1	3:G:45:SER:HA	1.74	0.51
1:C:13:LEU:HG	2:D:56:PHE:CZ	2.46	0.51
3:E:35:TYR:CE1	3:E:91:VAL:HG11	2.45	0.51
1:A:95:GLU:HG3	2:B:32:PRO:HB3	1.92	0.51
2:D:1:ILE:CB	2:D:32:PRO:HD3	2.41	0.50
2:D:49:VAL:HG12	2:D:68:THR:HB	1.92	0.50
3:G:161:THR:HG23	3:G:162:ASP:O	2.10	0.50
2:D:33:SER:HB2	2:D:54:LEU:HD21	1.92	0.50
1:A:202:LEU:HD22	1:A:204:LEU:HD23	1.94	0.50
1:A:13:LEU:HG	2:B:56:PHE:CZ	2.47	0.50
1:A:95:GLU:HG2	11:A:304:HOH:O	2.11	0.49
1:C:25:ARG:NH1	1:C:27:ASP:OD1	2.42	0.49
3:G:40:GLY:HA2	11:H:425:HOH:O	2.11	0.49
1:A:113:PHE:HB2	1:A:115:HIS:CE1	2.47	0.49
4:F:220:LEU:HD22	4:F:233:PRO:HG2	1.93	0.49
3:E:23:ASN:OD1	3:E:72:GLN:NE2	2.46	0.49
4:F:21:LEU:HD22	4:F:112:THR:HG21	1.94	0.49
1:A:256:ALA:HA	1:A:257:ALA:HB3	1.93	0.49
1:C:147:VAL:HG11	3:E:99:LEU:HD13	1.94	0.49
4:H:130:VAL:HG23	4:H:240:ALA:HB3	1.94	0.49
1:C:196:SER:HB3	1:C:203:LEU:HD22	1.95	0.48
2:D:97:ARG:HB3	11:D:439:HOH:O	2.13	0.48
1:A:35:LEU:HD11	1:A:212:TYR:OH	2.14	0.48
3:G:182:SER:OG	3:G:184:LYS:HG2	2.13	0.47
1:A:191:LEU:HB2	1:A:280:HIS:HB3	1.96	0.47
2:B:17:ASN:HA	2:B:72:PRO:O	2.15	0.47
4:F:132:GLU:HG2	4:F:245:ARG:HH21	1.79	0.47
4:F:201:ALA:O	4:F:205:GLN:HB2	2.14	0.47
4:H:53:ASN:ND2	11:H:328:HOH:O	2.47	0.46
3:G:138:VAL:HG11	4:H:147:VAL:HG11	1.96	0.46
1:A:47:VAL:H	1:A:67:GLN:NE2	2.13	0.46
1:C:263:VAL:HB	1:C:273:ILE:HB	1.98	0.46

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:F:121:LYS:HE2	4:F:228:GLN:HE22	1.81	0.46
4:F:6:GLN:O	4:F:9:ARG:NH1	2.36	0.46
4:H:21:LEU:HD22	4:H:112:THR:HG21	1.97	0.46
1:A:220:TRP:HD1	1:A:230:THR:HG23	1.81	0.46
2:B:3:ARG:HH11	2:B:61:SER:HB3	1.81	0.46
1:A:73:TYR:CE2	7:A:3000:AGH:HAI2	2.50	0.45
1:A:49:SER:HB3	1:A:54:SER:HB2	1.98	0.45
1:A:267:SER:CA	1:A:268:LEU:CB	2.94	0.45
1:C:194:GLY:O	1:C:203:LEU:HB2	2.17	0.45
3:G:141:PHE:HB2	3:G:193:PHE:CE1	2.52	0.45
1:C:47:VAL:HG11	7:C:3000:AGH:HAQ1	1.98	0.44
1:C:199:PRO:HA	1:C:200:GLY:HA2	1.77	0.44
3:G:128:LEU:HD12	4:H:133:PRO:HA	1.99	0.44
3:E:32:LEU:HB2	3:E:92:VAL:HG22	1.98	0.44
3:E:177:SER:OG	4:F:196:ARG:HD2	2.17	0.44
1:A:194:GLY:HA3	1:A:203:LEU:HD23	2.00	0.44
2:D:95:TRP:CZ3	2:D:97:ARG:HD2	2.53	0.44
4:H:25:GLN:HE22	4:H:29:HIS:CB	2.30	0.44
4:F:99:ASP:O	4:F:102:LEU:N	2.51	0.44
3:G:130:ASP:OD1	3:G:130:ASP:N	2.51	0.44
8:C:1000:NAG:H62	8:C:1001:NDG:C1	2.48	0.43
4:F:153:PHE:HE2	4:F:156:ASP:HA	1.83	0.43
3:G:60:ARG:NH2	3:G:84:ASP:OD1	2.51	0.43
4:F:6:GLN:HB2	4:F:110:PRO:HD2	2.01	0.43
1:A:88:LEU:HD13	1:A:140:TRP:CE3	2.54	0.43
1:C:131:TRP:HB3	1:C:145:ILE:HD12	2.00	0.43
4:F:127:GLU:O	4:F:150:ALA:HA	2.18	0.43
3:G:119:GLN:O	3:G:121:PRO:HD3	2.19	0.43
3:E:166:LEU:H	3:E:166:LEU:HD13	1.84	0.43
3:E:49:MET:HE3	3:E:66:LEU:HB2	2.01	0.43
4:F:98:ARG:N	4:F:99:ASP:C	2.72	0.43
1:C:251:VAL:CG1	1:C:252:VAL:N	2.81	0.42
1:A:255:GLU:HA	1:A:256:ALA:HA	1.77	0.42
7:C:3000:AGH:O2A	3:E:95:ARG:HG3	2.19	0.42
4:F:19:ILE:HD11	4:F:82:ALA:HB2	2.01	0.42
4:F:97:LEU:H	4:F:97:LEU:HD23	1.84	0.42
4:H:113:ARG:HB3	4:H:113:ARG:HH11	1.84	0.42
1:C:12:CYS:HB3	7:C:3000:AGH:HAN1	2.00	0.42
1:C:251:VAL:CG1	1:C:252:VAL:H	2.32	0.42
4:F:121:LYS:HE2	4:F:228:GLN:NE2	2.34	0.42
3:G:116:PRO:HG3	3:G:165:VAL:HG13	2.02	0.42

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:277:TRP:HA	1:A:278:HIS:HA	1.53	0.42
1:C:202:LEU:HB2	1:C:251:VAL:O	2.19	0.42
2:D:9:VAL:HG11	2:D:95:TRP:HB2	2.02	0.42
2:B:84:HIS:CE1	2:B:85:VAL:HG12	2.54	0.42
3:E:20:LEU:HD22	3:E:110:THR:HG21	2.02	0.42
1:C:150:GLN:OE1	3:E:97:SER:HB2	2.20	0.42
4:F:98:ARG:HD3	4:F:99:ASP:OD2	2.20	0.42
3:E:48:ILE:HD11	4:F:105:GLU:HG3	2.02	0.42
2:D:3:ARG:HH11	2:D:61:SER:HB3	1.85	0.41
1:C:235:ILE:HD12	11:C:434:HOH:O	2.20	0.41
1:A:150:GLN:HE22	3:G:98:THR:H	1.69	0.41
1:A:267:SER:CB	1:A:268:LEU:CB	2.91	0.41
3:E:130:ASP:HB2	3:E:131:SER:H	1.68	0.41
3:G:27:SER:HA	3:G:28:PRO:C	2.41	0.41
4:H:12:VAL:HB	4:H:155:PRO:HG3	2.02	0.41
4:H:176:ASP:OD1	4:H:196:ARG:NH1	2.49	0.41
1:C:208:VAL:HB	1:C:245:LEU:HD23	2.03	0.41
3:G:87:SER:HB2	3:G:111:GLN:HE22	1.86	0.40
4:H:113:ARG:NH1	4:H:113:ARG:HB3	2.36	0.40
3:G:125:VAL:O	3:G:205:PRO:HD3	2.21	0.40
4:H:234:VAL:O	4:H:236:GLN:HG2	2.21	0.40
4:H:27[A]:MET:HB3	11:H:268:HOH:O	2.20	0.40
3:E:166:LEU:HB3	4:F:174:CYS:HB2	2.03	0.40
4:H:44:HIS:HB3	4:H:60:LEU:HB2	2.03	0.40
1:A:191:LEU:O	1:A:280:HIS:HB2	2.22	0.40
3:G:161:THR:CG2	3:G:179:VAL:H	2.20	0.40
4:H:30:ASP:HB3	4:H:50:TYR:O	2.20	0.40
4:H:80:GLU:HG2	11:H:318:HOH:O	2.20	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	A	273/284 (96%)	263 (96%)	9 (3%)	1 (0%)	38	59
1	C	272/284 (96%)	244 (90%)	21 (8%)	7 (3%)	6	9
2	B	95/99 (96%)	94 (99%)	1 (1%)	0	100	100
2	D	98/99 (99%)	93 (95%)	4 (4%)	1 (1%)	18	32
3	E	192/209 (92%)	184 (96%)	7 (4%)	1 (0%)	32	53
3	G	199/209 (95%)	190 (96%)	9 (4%)	0	100	100
4	F	239/246 (97%)	230 (96%)	8 (3%)	1 (0%)	38	59
4	H	244/246 (99%)	234 (96%)	10 (4%)	0	100	100
All	All	1612/1676 (96%)	1532 (95%)	69 (4%)	11 (1%)	25	43

All (11) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	268	LEU
1	C	271	GLN
3	E	187	PHE
1	C	137	ALA
1	C	195	PRO
1	C	277	TRP
2	D	2	GLN
4	F	39	PRO
1	C	128	GLY
1	C	197	PRO
1	C	270	GLY

### 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	233/249 (94%)	220 (94%)	13 (6%)	25	45
1	C	226/249 (91%)	212 (94%)	14 (6%)	21	39
2	B	90/94 (96%)	83 (92%)	7 (8%)	15	28
2	D	93/94 (99%)	89 (96%)	4 (4%)	33	58

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	E	182/191 (95%)	173 (95%)	9 (5%)	29	52
3	G	180/191 (94%)	170 (94%)	10 (6%)	25	45
4	F	212/217 (98%)	199 (94%)	13 (6%)	22	40
4	H	215/217 (99%)	201 (94%)	14 (6%)	20	37
All	All	1431/1502 (95%)	1347 (94%)	84 (6%)	23	42

All (84) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	29	LEU
1	A	41	SER
1	A	57	THR
1	A	66	LEU
1	A	75	SER
1	A	96	LEU
1	A	111	ASN
1	A	145	ILE
1	A	202	LEU
1	A	236	LEU
1	A	249	LEU
1	A	252	VAL
1	A	268	LEU
2	B	2	GLN
2	B	6	LYS
2	B	36	GLU
2	B	48	LYS
2	B	54	LEU
2	B	70	PHE
2	B	75	LYS
1	C	13	LEU
1	C	24	THR
1	C	35	LEU
1	C	79	ARG
1	C	89	ARG
1	C	116	VAL
1	C	127	GLN
1	C	139	LEU
1	C	143	LEU
1	C	182	LYS
1	C	203	LEU
1	C	205	VAL

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	245	LEU
1	C	268	LEU
2	D	19	LYS
2	D	70	PHE
2	D	94	LYS
2	D	97	ARG
3	E	6	SER
3	E	32	LEU
3	E	72	GLN
3	E	95	ARG
3	E	99	LEU
3	E	128	LEU
3	E	130	ASP
3	E	152	GLN
3	E	166	LEU
4	F	11	LEU
4	F	25	GLN
4	F	45	LEU
4	F	52	VAL
4	F	59	ASP
4	F	121	LYS
4	F	135	GLU
4	F	158	VAL
4	F	187	ASN
4	F	196	ARG
4	F	205	GLN
4	F	220	LEU
4	F	227	THR
3	G	1	ASN
3	G	32	LEU
3	G	41	ARG
3	G	72	GLN
3	G	91	VAL
3	G	121	PRO
3	G	128	LEU
3	G	130	ASP
3	G	161	THR
3	G	166	LEU
4	H	9	ARG
4	H	11	LEU
4	H	20	THR
4	H	25	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
4	H	36	GLN
4	H	45	LEU
4	H	63	GLU
4	H	75	PHE
4	H	102	LEU
4	H	171	SER
4	H	174	CYS
4	H	196	ARG
4	H	208	ARG
4	H	223	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (37) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	14	GLN
1	A	55	GLN
1	A	62	GLN
1	A	67	GLN
1	A	111	ASN
1	A	142	ASN
1	A	150	GLN
1	A	184	GLN
2	B	2	GLN
1	C	115	HIS
1	C	231	GLN
2	D	2	GLN
2	D	8	GLN
2	D	89	GLN
3	E	21	GLN
3	E	23	ASN
3	E	72	GLN
3	E	127	GLN
3	E	152	GLN
3	E	194	ASN
3	E	195	ASN
4	F	86	HIS
4	F	89	GLN
4	F	187	ASN
4	F	228	GLN
4	F	236	GLN
3	G	1	ASN
3	G	2	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	G	21	GLN
3	G	23	ASN
3	G	72	GLN
3	G	111	GLN
3	G	149	ASN
4	H	25	GLN
4	H	53	ASN
4	H	140	HIS
4	H	223	ASN

### 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 ⓘ

5 carbohydrates 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	$\# Z  > 2$
5	NDG	A	1000	1,5	14,14,15	0.55	0	15,19,21	1.14	1 (6%)
5	BMA	A	1001	5	11,11,12	0.54	0	13,15,17	0.73	0
8	NAG	C	1000	1,8	14,14,15	0.48	0	15,19,21	1.51	4 (26%)
8	NDG	C	1001	8	14,14,15	0.78	0	15,19,21	1.05	1 (6%)
8	BMA	C	1002	8	11,11,12	0.69	0	13,15,17	1.00	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	NDG	A	1000	1,5	-	0/6/23/26	0/1/1/1
5	BMA	A	1001	5	-	0/2/19/22	0/1/1/1
8	NAG	C	1000	1,8	-	0/6/23/26	0/1/1/1
8	NDG	C	1001	8	-	0/6/23/26	0/1/1/1
8	BMA	C	1002	8	-	0/2/19/22	0/1/1/1

There are no bond length outliers.

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	C	1000	NAG	O4-C4-C3	-2.81	104.23	110.36
8	C	1000	NAG	O5-C1-C2	-2.53	107.96	111.47
8	C	1000	NAG	O4-C4-C5	2.33	115.17	109.28
5	A	1000	NDG	C1-O-C5	2.77	115.99	112.17
8	C	1000	NAG	C1-O5-C5	2.98	116.28	112.17
8	C	1001	NDG	C4-C3-C2	3.44	116.07	111.02

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
8	C	1000	NAG	1	0
8	C	1001	NDG	1	0

## 5.6 Ligand geometry

Of 5 ligands modelled in this entry, 1 is monoatomic - leaving 4 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
6	NAG	A	2000	1	14,14,15	0.52	0	15,19,21	2.73	5 (33%)
7	AGH	A	3000	-	60,60,60	0.47	1 (1%)	64,69,69	0.76	1 (1%)
9	NDG	C	2000	1	14,14,15	0.49	0	15,19,21	1.29	2 (13%)
7	AGH	C	3000	-	60,60,60	0.41	0	64,69,69	0.79	1 (1%)

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
6	NAG	A	2000	1	-	0/6/23/26	0/1/1/1
7	AGH	A	3000	-	-	0/58/78/78	0/1/1/1
9	NDG	C	2000	1	-	0/6/23/26	0/1/1/1
7	AGH	C	3000	-	-	0/58/78/78	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	A	3000	AGH	O1A-C1A	2.58	1.44	1.40

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	A	2000	NAG	C4-C3-C2	-3.76	105.51	111.02
7	C	3000	AGH	C1-C2-N2	-3.33	105.16	109.76
7	A	3000	AGH	C1-C2-N2	-2.61	106.15	109.76
9	C	2000	NDG	O7-C7-C8	-2.06	118.31	122.06
6	A	2000	NAG	C8-C7-N2	2.27	120.20	116.11
6	A	2000	NAG	C1-C2-N2	2.57	114.88	110.49
9	C	2000	NDG	C1-O-C5	3.61	117.14	112.17
6	A	2000	NAG	C2-N2-C7	3.63	128.23	122.94
6	A	2000	NAG	C1-O5-C5	8.00	123.20	112.17

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 6 short contacts:



Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	A	3000	AGH	2	0
7	C	3000	AGH	4	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	275/284 (96%)	0.17	5 (1%) 69 70	22, 39, 59, 68	0
1	C	274/284 (96%)	0.48	21 (7%) 14 14	29, 43, 69, 74	0
2	B	97/99 (97%)	-0.11	0 100 100	23, 31, 40, 43	0
2	D	99/99 (100%)	0.01	1 (1%) 82 83	30, 35, 44, 55	0
3	E	199/209 (95%)	0.22	6 (3%) 51 53	27, 37, 70, 70	0
3	G	202/209 (96%)	0.32	13 (6%) 20 21	24, 36, 70, 73	0
4	F	242/246 (98%)	0.23	0 100 100	31, 47, 56, 61	0
4	H	245/246 (99%)	0.08	2 (0%) 86 86	22, 43, 54, 55	1 (0%)
All	All	1633/1676 (97%)	0.21	48 (2%) 52 55	22, 41, 64, 74	1 (0%)

All (48) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	228	GLN	6.6
1	C	198	GLY	4.8
3	E	206	SER	4.8
1	C	256	ALA	4.5
1	C	257	ALA	4.2
2	D	99	MET	4.1
1	C	225	GLN	3.9
1	C	199	PRO	3.8
3	G	205	PRO	3.8
1	C	253	ALA	3.6
1	A	253	ALA	3.5
1	C	251	VAL	3.4
1	C	277	TRP	3.4
3	G	183	ASN	3.2
1	C	252	VAL	3.1
1	C	203	LEU	3.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	C	202	LEU	3.0
3	G	197	ILE	3.0
1	C	254	GLY	3.0
4	H	202	THR	2.9
3	G	200	GLU	2.8
3	G	134	SER	2.8
4	H	187	ASN	2.7
1	C	259	LEU	2.6
3	G	133	SER	2.6
3	G	203	PHE	2.6
3	G	194	ASN	2.5
3	E	205	PRO	2.5
3	E	183	ASN	2.4
1	A	228	GLN	2.4
1	C	275	LEU	2.4
3	E	51	PHE	2.3
1	A	107	GLY	2.3
1	C	220	TRP	2.3
1	A	259	LEU	2.3
1	A	199	PRO	2.2
1	C	195	PRO	2.2
1	C	274	VAL	2.2
3	G	132	LYS	2.2
3	E	134	SER	2.2
1	C	255	GLU	2.2
1	C	204	LEU	2.2
3	G	185	SER	2.2
3	G	171	MET	2.1
3	G	195	ASN	2.1
3	E	119	GLN	2.1
1	C	276	TYR	2.1
3	G	51	PHE	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](#)

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	NAG	C	1000	14/15	0.95	0.13	-	46,49,53,57	0
5	NDG	A	1000	14/15	0.87	0.17	-	45,50,52,53	0
8	NDG	C	1001	14/15	0.93	0.15	-	60,63,63,64	0
5	BMA	A	1001	11/12	0.77	0.17	-	55,56,56,57	0
8	BMA	C	1002	11/12	0.84	0.15	-	65,65,65,65	0

## 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
7	AGH	C	3000	60/60	0.96	0.18	0.71	26,31,34,35	0
7	AGH	A	3000	60/60	0.96	0.17	0.38	22,26,37,38	0
6	NAG	A	2000	14/15	0.90	0.13	-	38,40,41,41	0
10	MG	H	1000	1/1	0.94	0.21	-	27,27,27,27	1
9	NDG	C	2000	14/15	0.92	0.13	-	43,44,45,46	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.