



# Full wwPDB X-ray Structure Validation Report

Feb 28, 2014 – 09:59 AM GMT

PDB ID : 4ELM  
Title : Crystal structure of the mouse CD1d-lysosulfatide-Hy19.3TCR complex  
Authors : Girardi, E.; Maricic, I.; Wang, J.; Mac, T.T.; Iyer, P.; Kumar, V.; Zajonc, D.M.  
Deposited on : 2012-04-11  
Resolution : 3.48 Å(reported)

This is a full wwPDB 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/ValidationPDFNotes.html>

---

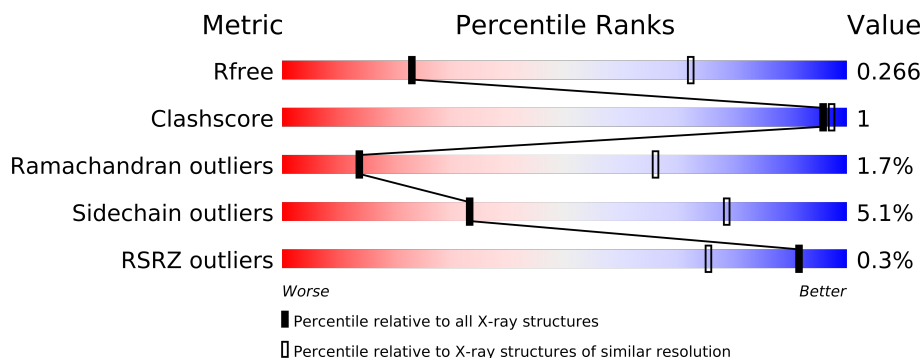
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 3.48 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))
$R_{free}$	66092	1175 (3.66-3.30)
Clashscore	79885	1001 (3.62-3.34)
Ramachandran outliers	78287	1001 (3.64-3.32)
Sidechain outliers	78261	1002 (3.64-3.32)
RSRZ outliers	66119	1175 (3.66-3.30)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	285	
1	C	285	
2	B	99	
2	D	99	
3	E	208	
3	G	208	
4	F	244	
4	H	244	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
7	SGF	A	305	-	X

Continued on next page...

*Continued from previous page...*

Mol	Type	Chain	Res	Geometry	Electron density
7	SGF	C	507	-	X
8	PLM	A	306	-	X
8	PLM	C	508	-	X

## 2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 11360 atoms, of which 0 are hydrogen and 0 are deuterium.

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 Antigen-presenting glycoprotein CD1d1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	266	Total	C	N	O	S	0	0	0
			2047	1310	343	382	12			
1	C	269	Total	C	N	O	S	0	0	0
			2093	1340	358	383	12			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	201	HIS	ASP	SEE REMARK 999	UNP P11609
A	280	HIS	-	EXPRESSION TAG	UNP P11609
A	281	HIS	-	EXPRESSION TAG	UNP P11609
A	282	HIS	-	EXPRESSION TAG	UNP P11609
A	283	HIS	-	EXPRESSION TAG	UNP P11609
A	284	HIS	-	EXPRESSION TAG	UNP P11609
A	285	HIS	-	EXPRESSION TAG	UNP P11609
C	201	HIS	ASP	SEE REMARK 999	UNP P11609
C	280	HIS	-	EXPRESSION TAG	UNP P11609
C	281	HIS	-	EXPRESSION TAG	UNP P11609
C	282	HIS	-	EXPRESSION TAG	UNP P11609
C	283	HIS	-	EXPRESSION TAG	UNP P11609
C	284	HIS	-	EXPRESSION TAG	UNP P11609
C	285	HIS	-	EXPRESSION TAG	UNP P11609

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	98	Total	C	N	O	S	0	0	0
			747	471	126	143	7			
2	D	99	Total	C	N	O	S	0	0	0
			777	495	129	146	7			

- Molecule 3 is a protein called Hy19.3 TCR alpha chain (mouse variable domain, human

constant domain).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	179	Total	C	N	O	S	0	0	0
			1234	772	217	239	6			
3	G	171	Total	C	N	O	S	0	0	0
			1185	735	207	236	7			

- Molecule 4 is a protein called Hy19.3 TCR beta chain (mouse variable domain, human constant domain).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	F	224	Total	C	N	O	S	0	0	0
			1578	1007	266	296	9			
4	H	211	Total	C	N	O	S	0	0	0
			1459	929	247	274	9			

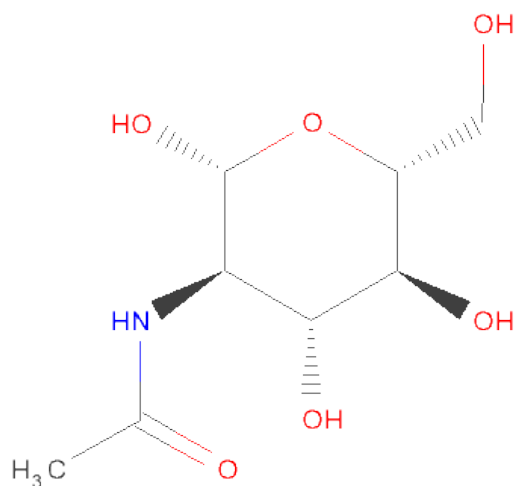
- Molecule 5 is a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	2	Total	C	N	O	0	0
			28	16	2	10		

There are 7 discrepancies between the modelled and reference sequences:

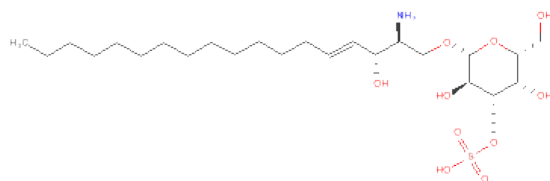
Chain	Residue	Modelled	Actual	Comment	Reference
A	201	HIS	ASP	SEE REMARK 999	UNP P11609
A	280	HIS	-	EXPRESSION TAG	UNP P11609
A	281	HIS	-	EXPRESSION TAG	UNP P11609
A	282	HIS	-	EXPRESSION TAG	UNP P11609
A	283	HIS	-	EXPRESSION TAG	UNP P11609
A	284	HIS	-	EXPRESSION TAG	UNP P11609
A	285	HIS	-	EXPRESSION TAG	UNP P11609

- Molecule 6 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	A	1	Total	C	N	O	0	0
			13	8	1	4		
6	A	1	Total	C	N	O	0	0
			14	8	1	5		
6	C	1	Total	C	N	O	0	0
			14	8	1	5		
6	C	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 7 is (2S,3R,4E)-2-AMINO-3-HYDROXYOCTADEC-4-EN-1-YL3-O-SULFO-BETA-D-GALACTOPYRANOSIDE (three-letter code: SGF) (formula: C<sub>24</sub>H<sub>47</sub>NO<sub>10</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
7	A	1	Total	C	N	O	S	0	0
			36	24	1	10	1		
7	C	1	Total	C	N	O	S	0	0
			36	24	1	10	1		

- Molecule 8 is PALMITIC ACID (three-letter code: PLM) (formula: C<sub>16</sub>H<sub>32</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
8	A	1	Total	C	O	0	0
			18	16	2		
8	C	1	Total	C	O	0	0
			18	16	2		

- Molecule 9 is a polymer of unknown type called SUGAR (4-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
9	C	4	Total	C	N	O	0	0
			49	28	2	19		

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	201	HIS	ASP	SEE REMARK 999	UNP P11609
C	280	HIS	-	EXPRESSION TAG	UNP P11609
C	281	HIS	-	EXPRESSION TAG	UNP P11609
C	282	HIS	-	EXPRESSION TAG	UNP P11609
C	283	HIS	-	EXPRESSION TAG	UNP P11609

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	284	HIS	-	EXPRESSION TAG	UNP P11609
C	285	HIS	-	EXPRESSION TAG	UNP P11609

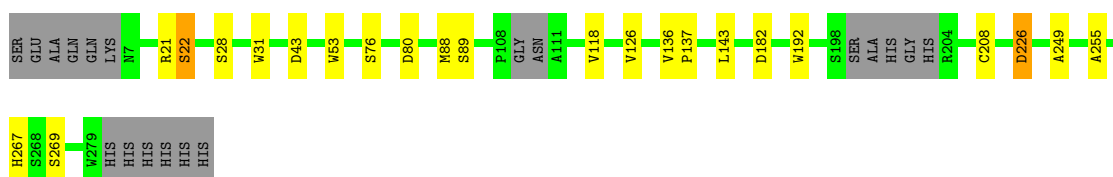


### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: Antigen-presenting glycoprotein CD1d1

Chain A: 



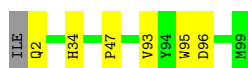
- Molecule 1: Antigen-presenting glycoprotein CD1d1

Chain C: 

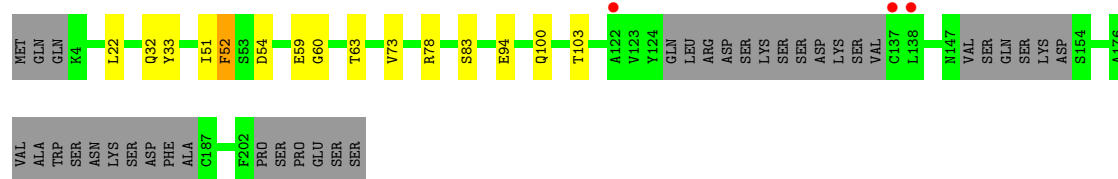


- Molecule 2: Beta-2 microglobulin

Chain B: 

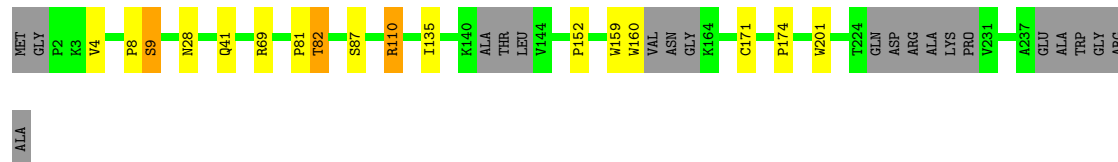


Chain G: 



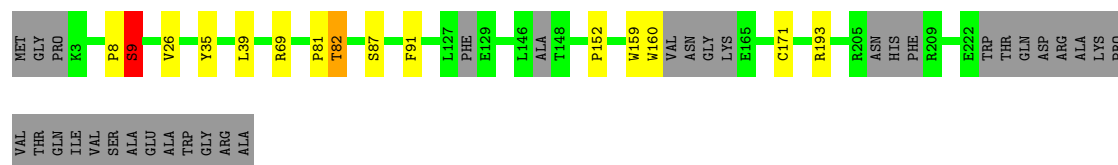
- Molecule 4: Hy19.3 TCR beta chain (mouse variable domain, human constant domain)

Chain F: 



- Molecule 4: Hy19.3 TCR beta chain (mouse variable domain, human constant domain)

Chain H: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	98.51Å 127.00Å 104.35Å 90.00° 110.53° 90.00°	Depositor
Resolution (Å)	92.26 – 3.48 83.22 – 3.48	Depositor EDS
% Data completeness (in resolution range)	99.0 (92.26-3.48) 99.1 (83.22-3.48)	Depositor EDS
$R_{merge}$	0.24	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.63 (at 3.49Å)	Xtriage
Refinement program	REFMAC 5.6.0104	Depositor
R, $R_{free}$	0.209 , 0.265 0.209 , 0.266	Depositor DCC
$R_{free}$ test set	1541 reflections (5.30%)	DCC
Wilson B-factor (Å <sup>2</sup> )	73.3	Xtriage
Anisotropy	0.076	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 38.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.27$	Xtriage
Outliers	0 of 30592 reflections	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	11360	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	82.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 16.32% of the height of the origin peak. No significant pseudotranslation is detected.*

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

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: FUC, PLM, BMA, NAG, SGF

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	0.63	3/2108 (0.1%)	0.55	0/2882
1	C	0.62	4/2155 (0.2%)	0.55	0/2943
2	B	0.54	1/773 (0.1%)	0.51	0/1061
2	D	0.55	0/803	0.54	0/1099
3	E	0.48	0/1263	0.54	0/1727
3	G	0.47	0/1211	0.53	0/1654
4	F	0.54	3/1623 (0.2%)	0.60	2/2230 (0.1%)
4	H	0.53	2/1501 (0.1%)	0.57	0/2063
All	All	0.56	13/11437 (0.1%)	0.56	2/15659 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
4	F	0	1

All (13) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	219	TRP	CD2-CE2	5.30	1.47	1.41
4	H	160	TRP	CD2-CE2	5.16	1.47	1.41
1	A	53	TRP	CD2-CE2	5.15	1.47	1.41
4	H	159	TRP	CD2-CE2	5.14	1.47	1.41
4	F	160	TRP	CD2-CE2	5.10	1.47	1.41
1	C	53	TRP	CD2-CE2	5.06	1.47	1.41
4	F	201	TRP	CD2-CE2	5.04	1.47	1.41
2	B	95	TRP	CD2-CE2	5.03	1.47	1.41
1	C	40	TRP	CD2-CE2	5.02	1.47	1.41
4	F	159	TRP	CD2-CE2	5.02	1.47	1.41

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	31	TRP	CD2-CE2	5.02	1.47	1.41
1	C	222	TRP	CD2-CE2	5.01	1.47	1.41
1	A	192	TRP	CD2-CE2	5.00	1.47	1.41

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	F	9	SER	N-CA-C	-5.79	95.36	111.00
4	F	82	THR	N-CA-CB	5.08	119.95	110.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	F	81	PRO	Peptide

## 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2047	0	0	5	0
1	C	2093	0	0	2	0
2	B	747	0	0	0	0
2	D	777	0	0	0	0
3	E	1234	0	0	2	0
3	G	1185	0	0	3	0
4	F	1578	0	0	1	0
4	H	1459	0	0	2	0
5	A	28	0	0	0	0
6	A	27	0	0	0	0
6	C	28	0	0	0	0
7	A	36	0	0	0	0
7	C	36	0	0	0	0
8	A	18	0	31	0	0
8	C	18	0	31	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
9	C	49	0	0	0	0
All	All	11360	0	62	14	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 1.

All (14) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
3:G:33:TYR:CD1	3:G:52:PHE:CB	2.59	0.85
4:F:110:ARG:NH1	4:F:110:ARG:CG	2.67	0.57
3:G:94:GLU:N	3:G:100:GLN:OE1	2.43	0.51
1:C:167:THR:OG1	3:E:98:TYR:OH	2.29	0.51
1:A:267:HIS:ND1	1:A:269:SER:OG	2.44	0.49
1:C:226:ASP:OD1	1:C:226:ASP:N	2.45	0.49
3:E:94:GLU:N	3:E:100:GLN:OE1	2.46	0.49
4:H:8:PRO:CA	4:H:9:SER:CB	2.92	0.48
1:A:21:ARG:O	1:A:22:SER:CB	2.62	0.48
1:A:118:VAL:N	1:A:126:VAL:O	2.47	0.48
1:A:208:CYS:N	1:A:249:ALA:O	2.50	0.45
3:G:60:GLY:O	3:G:78:ARG:NH1	2.50	0.45
1:A:226:ASP:OD1	1:A:226:ASP:N	2.50	0.44
4:H:35:TYR:O	4:H:91:PHE:N	2.51	0.44

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	260/285 (91%)	252 (97%)	3 (1%)	5 (2%)	12	63
1	C	263/285 (92%)	252 (96%)	9 (3%)	2 (1%)	27	82
2	B	96/99 (97%)	86 (90%)	9 (9%)	1 (1%)	22	78
2	D	97/99 (98%)	91 (94%)	6 (6%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	E	171/208 (82%)	146 (85%)	23 (14%)	2 (1%)	19	74
3	G	163/208 (78%)	142 (87%)	20 (12%)	1 (1%)	33	85
4	F	216/244 (88%)	181 (84%)	27 (12%)	8 (4%)	5	47
4	H	201/244 (82%)	173 (86%)	22 (11%)	6 (3%)	7	53
All	All	1467/1672 (88%)	1323 (90%)	119 (8%)	25 (2%)	14	66

All (25) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	22	SER
4	F	82	THR
4	F	152	PRO
4	H	81	PRO
4	H	152	PRO
4	F	9	SER
4	H	9	SER
1	A	88	MET
1	A	255	ALA
1	C	255	ALA
4	F	41	GLN
4	F	87	SER
4	H	39	LEU
4	H	87	SER
3	E	52	PHE
4	F	8	PRO
4	F	28	ASN
4	H	82	THR
1	A	137	PRO
1	C	137	PRO
1	A	89	SER
4	F	174	PRO
3	G	83	SER
3	E	116	ILE
2	B	47	PRO

### 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	208/249 (84%)	200 (96%)	8 (4%)	44	85
1	C	215/249 (86%)	206 (96%)	9 (4%)	40	83
2	B	75/93 (81%)	71 (95%)	4 (5%)	32	77
2	D	83/93 (89%)	81 (98%)	2 (2%)	61	91
3	E	104/185 (56%)	94 (90%)	10 (10%)	12	50
3	G	104/185 (56%)	95 (91%)	9 (9%)	15	56
4	F	131/217 (60%)	126 (96%)	5 (4%)	44	85
4	H	118/217 (54%)	112 (95%)	6 (5%)	33	79
All	All	1038/1488 (70%)	985 (95%)	53 (5%)	33	79

All (53) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	28	SER
1	A	43	ASP
1	A	76	SER
1	A	80	ASP
1	A	136	VAL
1	A	143	LEU
1	A	182	ASP
1	A	226	ASP
1	C	43	ASP
1	C	100	LEU
1	C	101	SER
1	C	131	THR
1	C	136	VAL
1	C	143	LEU
1	C	157	SER
1	C	196	VAL
1	C	226	ASP
2	B	2	GLN
2	B	34	HIS
2	B	93	VAL
2	B	96	ASP
2	D	28	THR
2	D	70	PHE
3	E	8	SER
3	E	13	SER

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
3	E	27	SER
3	E	48	LEU
3	E	51	ILE
3	E	59	GLU
3	E	68	LYS
3	E	73	VAL
3	E	83	SER
3	E	100	GLN
4	F	4	VAL
4	F	69	ARG
4	F	110	ARG
4	F	135	ILE
4	F	171	CYS
3	G	22	LEU
3	G	32	GLN
3	G	51	ILE
3	G	52	PHE
3	G	54	ASP
3	G	59	GLU
3	G	63	THR
3	G	73	VAL
3	G	103	THR
4	H	9	SER
4	H	26	VAL
4	H	69	ARG
4	H	82	THR
4	H	171	CYS
4	H	193	ARG

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 chains 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 ⓘ

6 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	NAG	A	301	5	12,14,15	0.57	0	15,19,21	1.60	3 (20%)
5	NAG	A	302	1,5	12,14,15	0.65	0	15,19,21	0.84	1 (6%)
9	NAG	C	503	1,9	12,14,15	0.71	0	15,19,21	0.78	0
9	NAG	C	504	9	12,14,15	0.69	0	15,19,21	2.10	5 (33%)
9	BMA	C	505	9	10,11,12	0.57	0	11,15,17	1.53	3 (27%)
9	FUC	C	506	9	9,10,11	0.71	0	10,14,16	1.02	1 (10%)

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	NAG	A	301	5	-	0/6/23/26	0/1/1/1
5	NAG	A	302	1,5	-	0/6/23/26	0/1/1/1
9	NAG	C	503	1,9	-	0/6/23/26	0/1/1/1
9	NAG	C	504	9	-	0/6/23/26	0/1/1/1
9	BMA	C	505	9	-	0/2/19/22	0/1/1/1
9	FUC	C	506	9	-	0/0/17/20	0/1/1/1

There are no bond length outliers.

All (13) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
9	C	504	NAG	C2-N2-C7	4.11	129.99	123.09
5	A	301	NAG	O5-C5-C4	3.92	115.63	110.65
9	C	504	NAG	C3-C4-C5	3.52	116.48	110.20
9	C	505	BMA	C4-C3-C2	3.27	114.90	110.50
5	A	301	NAG	C3-C4-C5	3.12	115.78	110.20
9	C	504	NAG	O5-C5-C4	3.06	114.54	110.65

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
9	C	504	NAG	C8-C7-N2	3.05	122.06	116.11
9	C	506	FUC	O5-C5-C6	2.63	112.26	108.03
9	C	505	BMA	O5-C5-C6	2.49	109.59	106.98
9	C	505	BMA	C3-C4-C5	2.48	114.64	110.20
5	A	301	NAG	C2-N2-C7	2.42	127.15	123.09
5	A	302	NAG	O5-C5-C6	2.39	109.49	106.98
9	C	504	NAG	C4-C3-C2	2.06	116.35	111.32

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.6 Ligand geometry ⓘ

8 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$
6	NAG	A	303	1	11,13,15	0.77	0	9,17,21	1.14	0
6	NAG	A	304	1	12,14,15	0.58	0	15,19,21	0.92	0
7	SGF	A	305	-	36,36,36	1.91	2 (5%)	46,46,46	1.33	5 (10%)
8	PLM	A	306	-	17,17,17	0.52	0	17,17,17	0.62	0
6	NAG	C	501	1	12,14,15	0.71	1 (8%)	15,19,21	0.88	1 (6%)
6	NAG	C	502	1	12,14,15	0.76	0	15,19,21	0.71	0
7	SGF	C	507	-	36,36,36	1.86	3 (8%)	46,46,46	1.29	5 (10%)
8	PLM	C	508	-	17,17,17	0.55	0	17,17,17	0.58	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
6	NAG	A	303	1	-	0/6/19/26	0/1/1/1
6	NAG	A	304	1	-	0/6/23/26	0/1/1/1
7	SGF	A	305	-	-	0/31/51/51	0/1/1/1
8	PLM	A	306	-	-	0/15/15/15	0/0/0/0
6	NAG	C	501	1	-	0/6/23/26	0/1/1/1
6	NAG	C	502	1	-	0/6/23/26	0/1/1/1
7	SGF	C	507	-	-	0/31/51/51	0/1/1/1
8	PLM	C	508	-	-	0/15/15/15	0/0/0/0

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	A	305	SGF	O4-S	-10.22	1.44	1.60
7	C	507	SGF	O4-S	-9.86	1.44	1.60
7	A	305	SGF	C5-C4	-2.51	1.39	1.50
7	C	507	SGF	C5-C4	-2.51	1.39	1.50
7	C	507	SGF	C2-C3	2.22	1.53	1.49
6	C	501	NAG	O5-C5	-2.11	1.41	1.45

All (11) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	A	305	SGF	C2-C3-C4	-4.17	118.98	125.12
7	A	305	SGF	C45-O4-S	-3.74	112.27	118.37
7	C	507	SGF	C-O-C43	-3.54	106.76	113.81
7	C	507	SGF	C2-C3-C4	-3.44	120.05	125.12
7	A	305	SGF	C-O-C43	-3.15	107.53	113.81
7	C	507	SGF	O-C-C1	3.13	115.00	108.12
7	A	305	SGF	O-C43-C44	3.07	112.09	108.18
7	C	507	SGF	C45-O4-S	-2.75	113.88	118.37
7	C	507	SGF	O-C43-C44	2.57	111.45	108.18
7	A	305	SGF	O-C-C1	2.48	113.57	108.12
6	C	501	NAG	O5-C5-C6	2.41	109.51	106.98

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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	266/285 (93%)	-0.20	0	100 100	46, 61, 99, 152	0
1	C	269/285 (94%)	-0.15	0	100 100	38, 54, 95, 133	0
2	B	98/99 (98%)	-0.16	0	100 100	55, 85, 108, 117	0
2	D	99/99 (100%)	-0.20	0	100 100	36, 49, 72, 78	0
3	E	179/208 (86%)	0.03	2 (1%)	77 42	40, 62, 171, 193	0
3	G	171/208 (82%)	0.08	3 (1%)	65 32	52, 81, 174, 203	0
4	F	224/244 (91%)	-0.12	0	100 100	40, 101, 169, 221	0
4	H	211/244 (86%)	-0.06	0	100 100	53, 127, 189, 236	0
All	All	1517/1672 (90%)	-0.10	5 (0%)	91 75	36, 71, 162, 236	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	G	122	ALA	3.0
3	E	138	LEU	2.7
3	G	138	LEU	2.4
3	E	176	ALA	2.4
3	G	137	CYS	2.2

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates ⓘ

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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
9	BMA	C	505	11/12	0.17	1.60	105,111,116,117	0
9	FUC	C	506	10/11	0.25	1.07	72,76,80,84	0
5	NAG	A	301	14/15	0.18	0.89	72,76,85,85	0
9	NAG	C	503	14/15	0.18	-0.75	63,67,77,79	0
5	NAG	A	302	14/15	0.16	-0.93	60,62,67,69	0
9	NAG	C	504	14/15	0.14	-1.14	82,86,95,99	0

## 6.4 Ligands

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. 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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
8	PLM	C	508	18/18	0.73	12.33	52,69,98,100	0
8	PLM	A	306	18/18	0.61	8.73	64,69,87,87	0
7	SGF	A	305	36/36	0.53	5.46	169,222,273,280	0
7	SGF	C	507	36/36	0.49	3.46	126,173,216,221	0
6	NAG	A	304	14/15	0.20	1.20	77,83,86,90	0
6	NAG	A	303	13/15	0.24	1.05	102,109,118,118	0
6	NAG	C	501	14/15	0.24	0.76	83,88,98,99	0
6	NAG	C	502	14/15	0.20	-0.40	53,58,61,62	0

## 6.5 Other polymers

There are no such residues in this entry.