



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 06:41 PM GMT

PDB ID : 4MBZ  
Title : Structure of B-Lymphotropic Polyomavirus VP1 in complex with 3'-sialyllactosamine  
Authors : Khan, Z.M.; Neu, U.; Stehle, T.  
Deposited on : 2013-08-21  
Resolution : 1.75 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

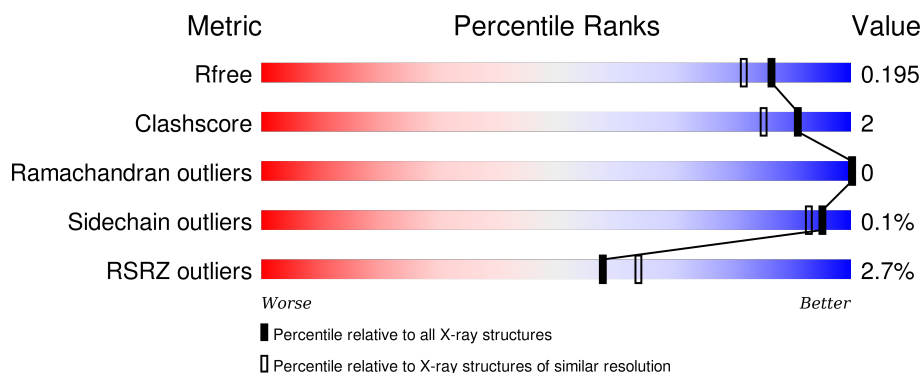
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.75 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))
$R_{free}$	91344	1609 (1.76-1.76)
Clashscore	102246	1730 (1.76-1.76)
Ramachandran outliers	100387	1711 (1.76-1.76)
Sidechain outliers	100360	1711 (1.76-1.76)
RSRZ outliers	91569	1610 (1.76-1.76)

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	278	<div> <div style="width: 95%;"></div> <div>95%</div> </div>
1	B	278	<div> <div style="width: 93%;"></div> <div>93%</div> <div style="width: 5%;"></div> <div>5%</div> </div>
1	C	278	<div> <div style="width: 94%;"></div> <div>94%</div> <div style="width: 5%;"></div> <div>5%</div> </div>
1	D	278	<div> <div style="width: 95%;"></div> <div>95%</div> <div style="width: 5%;"></div> <div>5%</div> </div>
1	E	278	<div> <div style="width: 96%;"></div> <div>96%</div> <div style="width: 5%;"></div> <div>5%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	278	
1	G	278	
1	H	278	
1	I	278	
1	J	278	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	EDO	A	403	-	-	-	X
4	EDO	B	404	-	-	-	X
4	EDO	B	406	-	-	X	-
4	EDO	B	407	-	-	-	X
4	EDO	C	403	-	-	-	X
4	EDO	E	404	-	-	-	X
4	EDO	F	403	-	-	-	X
4	EDO	J	405	-	-	-	X
5	IPA	A	405	-	-	-	X
5	IPA	B	408	-	-	-	X
5	IPA	C	405	-	-	-	X
5	IPA	D	405	-	-	-	X
5	IPA	E	406	-	-	-	X
5	IPA	F	404	-	-	-	X
5	IPA	G	405	-	-	-	X
5	IPA	H	404	-	-	-	X
5	IPA	I	405	-	-	-	X
5	IPA	J	406	-	-	-	X
6	SIA	A	406	-	-	-	X
6	GAL	A	407	-	-	-	X
6	SIA	B	409	-	-	-	X
6	GAL	B	410	-	-	-	X
6	SIA	D	406	-	-	-	X
6	GAL	D	407	-	-	-	X
6	SIA	E	407	-	-	-	X
6	GAL	E	408	-	-	-	X
6	SIA	H	405	-	-	-	X
6	GAL	H	406	-	-	-	X

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
6	SIA	I	406	-	-	-	X
6	GAL	I	407	-	-	-	X
7	SIA	J	407	-	-	-	X

## 2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 24643 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 Major Capsid Protein VP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	277	Total	C	N	O	S	0	10	0
			2168	1374	354	427	13			
1	B	274	Total	C	N	O	S	0	8	0
			2125	1345	350	417	13			
1	C	275	Total	C	N	O	S	0	12	0
			2156	1367	354	422	13			
1	D	274	Total	C	N	O	S	0	11	0
			2137	1356	349	419	13			
1	E	277	Total	C	N	O	S	0	7	0
			2144	1358	351	422	13			
1	F	274	Total	C	N	O	S	0	7	0
			2119	1341	349	416	13			
1	G	271	Total	C	N	O	S	0	7	0
			2097	1329	345	410	13			
1	H	273	Total	C	N	O	S	0	10	0
			2127	1347	347	420	13			
1	I	274	Total	C	N	O	S	0	9	0
			2138	1350	352	423	13			
1	J	277	Total	C	N	O	S	0	6	0
			2142	1354	353	422	13			

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	G	1	Total	Ca	0	0
			1	1		
2	J	1	Total	Ca	0	0
			1	1		
2	D	1	Total	Ca	0	0
			1	1		
2	E	1	Total	Ca	0	0
			1	1		

*Continued on next page...*

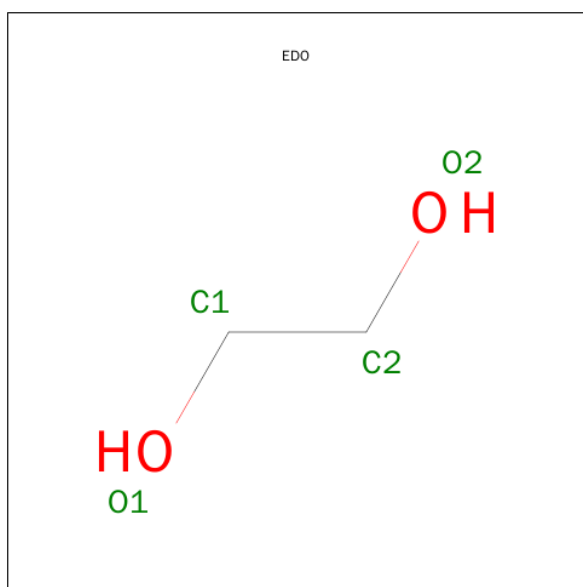
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	H	1	Total 1	Ca 1	0	0
2	B	1	Total 1	Ca 1	0	0
2	I	1	Total 1	Ca 1	0	0
2	C	1	Total 1	Ca 1	0	0
2	A	1	Total 1	Ca 1	0	0
2	F	1	Total 1	Ca 1	0	0

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	G	1	Total 1	Cl 1	0	0
3	J	1	Total 1	Cl 1	0	0
3	D	1	Total 1	Cl 1	0	0
3	E	1	Total 1	Cl 1	0	0
3	H	1	Total 1	Cl 1	0	0
3	B	1	Total 1	Cl 1	0	0
3	I	1	Total 1	Cl 1	0	0
3	C	1	Total 1	Cl 1	0	0
3	A	1	Total 1	Cl 1	0	0
3	F	1	Total 1	Cl 1	0	0

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



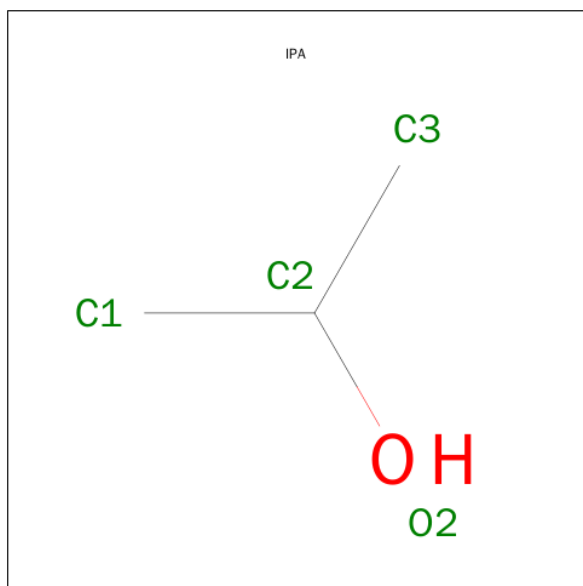
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			4	2	2		
4	A	1	Total	C	O	0	0
			4	2	2		
4	B	1	Total	C	O	0	0
			4	2	2		
4	B	1	Total	C	O	0	0
			4	2	2		
4	B	1	Total	C	O	0	0
			4	2	2		
4	B	1	Total	C	O	0	0
			4	2	2		
4	C	1	Total	C	O	0	0
			4	2	2		
4	C	1	Total	C	O	0	0
			4	2	2		
4	D	1	Total	C	O	0	0
			4	2	2		
4	D	1	Total	C	O	0	0
			4	2	2		
4	E	1	Total	C	O	0	0
			4	2	2		
4	E	1	Total	C	O	0	0
			4	2	2		
4	E	1	Total	C	O	0	0
			4	2	2		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	F	1	Total	C	O	0	0
			4	2	2		
4	G	1	Total	C	O	0	0
			4	2	2		
4	G	1	Total	C	O	0	0
			4	2	2		
4	H	1	Total	C	O	0	0
			4	2	2		
4	I	1	Total	C	O	0	0
			4	2	2		
4	I	1	Total	C	O	0	0
			4	2	2		
4	J	1	Total	C	O	0	0
			4	2	2		
4	J	1	Total	C	O	0	0
			4	2	2		
4	J	1	Total	C	O	0	0
			4	2	2		

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



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			4	3	1		
5	B	1	Total	C	O	0	0
			4	3	1		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	C	1	Total	C	O	0	0
			4	3	1		
5	D	1	Total	C	O	0	0
			4	3	1		
5	E	1	Total	C	O	0	0
			4	3	1		
5	F	1	Total	C	O	0	0
			4	3	1		
5	G	1	Total	C	O	0	0
			4	3	1		
5	H	1	Total	C	O	0	0
			4	3	1		
5	I	1	Total	C	O	0	0
			4	3	1		
5	J	1	Total	C	O	0	0
			4	3	1		

- Molecule 6 is a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	A	2	Total	C	N	O	0	0
			32	17	1	14		
6	B	2	Total	C	N	O	0	0
			32	17	1	14		
6	D	2	Total	C	N	O	0	0
			32	17	1	14		
6	E	2	Total	C	N	O	0	0
			32	17	1	14		
6	H	2	Total	C	N	O	0	0
			32	17	1	14		
6	I	2	Total	C	N	O	0	0
			32	17	1	14		

- Molecule 7 is a polymer of unknown type called SUGAR (3-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
7	J	3	Total	C	N	O	0	0
			46	25	2	19		

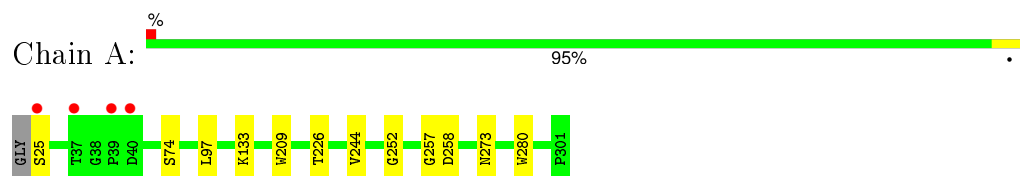
- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	359	Total 359	O 359	0	0
8	B	328	Total 328	O 328	0	0
8	C	330	Total 330	O 330	0	0
8	D	283	Total 283	O 283	0	0
8	E	346	Total 346	O 346	0	0
8	F	275	Total 275	O 275	0	0
8	G	230	Total 230	O 230	0	0
8	H	235	Total 235	O 235	0	0
8	I	246	Total 246	O 246	0	0
8	J	268	Total 268	O 268	0	0

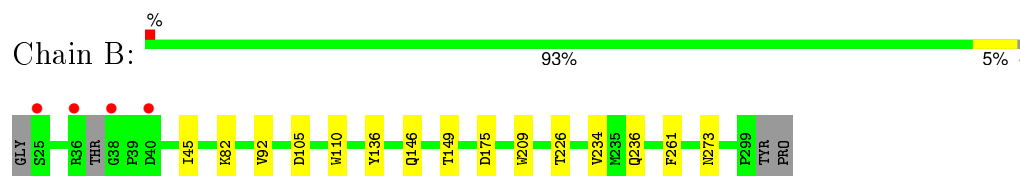
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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.

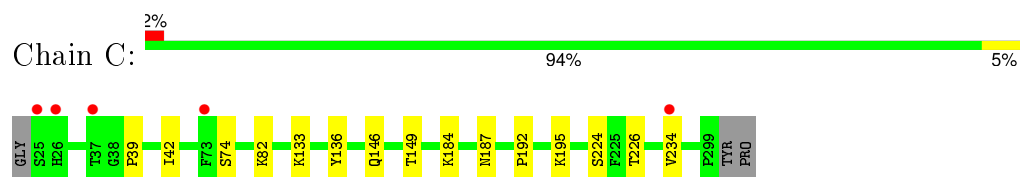
- Molecule 1: Major Capsid Protein VP1



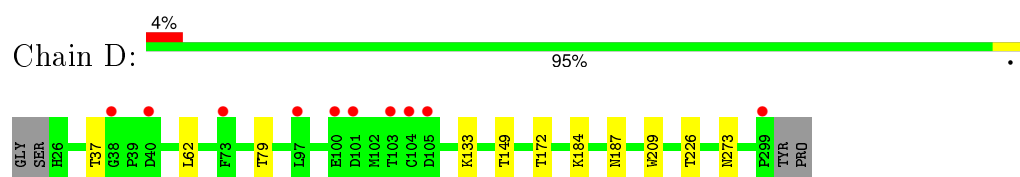
- Molecule 1: Major Capsid Protein VP1



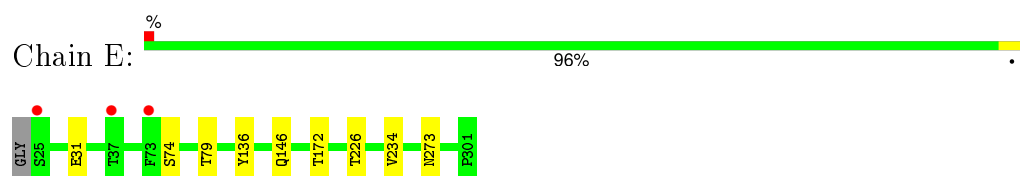
- Molecule 1: Major Capsid Protein VP1



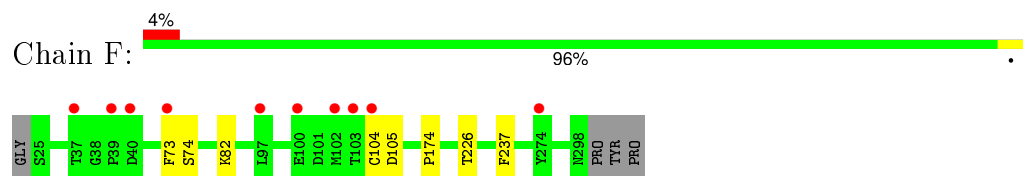
- Molecule 1: Major Capsid Protein VP1



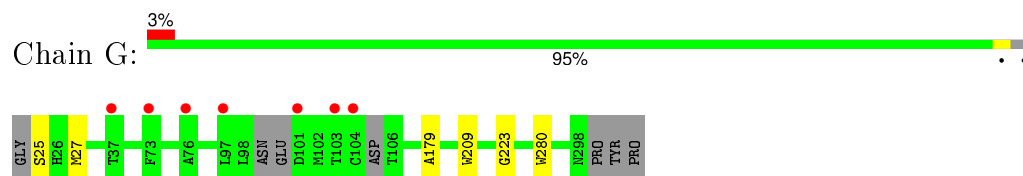
- Molecule 1: Major Capsid Protein VP1



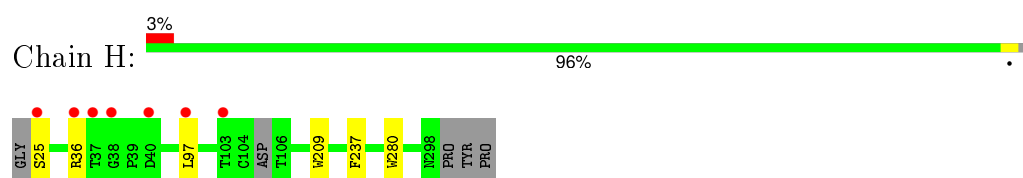
- Molecule 1: Major Capsid Protein VP1



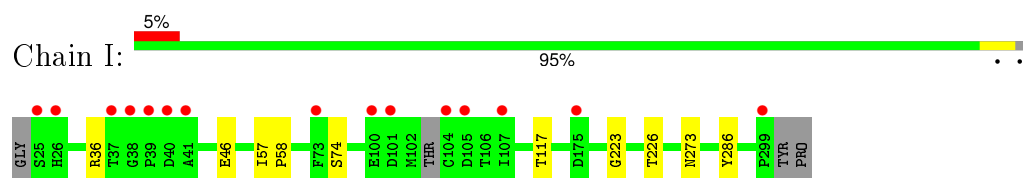
- Molecule 1: Major Capsid Protein VP1



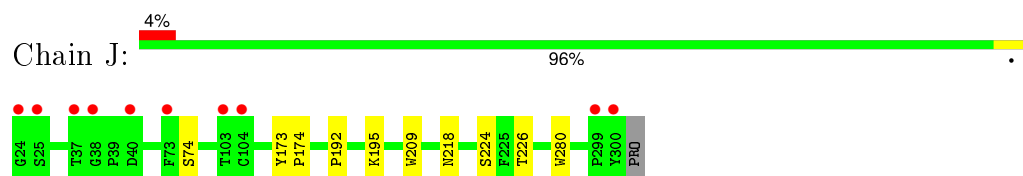
- Molecule 1: Major Capsid Protein VP1



- Molecule 1: Major Capsid Protein VP1



- Molecule 1: Major Capsid Protein VP1



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	150.16Å 97.24Å 234.22Å 90.00° 96.21° 90.00°	Depositor
Resolution (Å)	48.75 – 1.75 48.75 – 1.75	Depositor EDS
% Data completeness (in resolution range)	99.0 (48.75-1.75) 99.1 (48.75-1.75)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.22 (at 1.75Å)	Xtriage
Refinement program	REFMAC 5.8.0025	Depositor
R, $R_{free}$	0.165 , 0.195 0.165 , 0.195	Depositor DCC
$R_{free}$ test set	16696 reflections (5.27%)	DCC
Wilson B-factor (Å <sup>2</sup> )	18.4	Xtriage
Anisotropy	0.501	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 43.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	7 of 333434 reflections (0.002%)	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	24643	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.04% 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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: IPA, NAG, CL, CA, EDO, SIA, GAL

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	0.56	2/2242 (0.1%)	0.66	0/3056
1	B	0.55	2/2192 (0.1%)	0.65	0/2983
1	C	0.52	0/2233	0.63	0/3040
1	D	0.51	1/2217 (0.0%)	0.62	0/3022
1	E	0.54	0/2214	0.64	0/3018
1	F	0.49	0/2186	0.60	0/2976
1	G	0.50	2/2159 (0.1%)	0.61	0/2937
1	H	0.49	2/2203 (0.1%)	0.61	0/2998
1	I	0.48	0/2205	0.58	0/3002
1	J	0.50	2/2202 (0.1%)	0.58	0/3000
All	All	0.51	11/22053 (0.0%)	0.62	0/30032

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	J	209	TRP	CD2-CE2	5.31	1.47	1.41
1	H	209	TRP	CD2-CE2	5.29	1.47	1.41
1	A	209	TRP	CD2-CE2	5.18	1.47	1.41
1	J	280	TRP	CD2-CE2	5.15	1.47	1.41
1	B	110	TRP	CD2-CE2	5.14	1.47	1.41

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	2168	0	2148	11	0
1	B	2125	0	2115	16	0
1	C	2156	0	2161	14	0
1	D	2137	0	2141	9	0
1	E	2144	0	2129	9	0
1	F	2119	0	2112	7	0
1	G	2097	0	2088	4	0
1	H	2127	0	2114	3	0
1	I	2138	0	2112	16	0
1	J	2142	0	2116	6	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	F	1	0	0	0	0
2	G	1	0	0	0	0
2	H	1	0	0	0	0
2	I	1	0	0	0	0
2	J	1	0	0	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
3	E	1	0	0	0	0
3	F	1	0	0	0	0
3	G	1	0	0	0	0
3	H	1	0	0	0	0
3	I	1	0	0	0	0
3	J	1	0	0	0	0
4	A	8	0	12	2	0
4	B	20	0	30	6	0
4	C	8	0	12	0	0
4	D	8	0	12	1	0
4	E	12	0	18	1	0
4	F	4	0	6	0	0
4	G	8	0	12	1	0
4	H	4	0	6	1	0
4	I	8	0	12	2	0
4	J	12	0	18	1	0
5	A	4	0	8	2	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	B	4	0	8	1	0
5	C	4	0	8	1	0
5	D	4	0	8	1	0
5	E	4	0	8	0	0
5	F	4	0	8	0	0
5	G	4	0	8	1	0
5	H	4	0	8	1	0
5	I	4	0	8	1	0
5	J	4	0	8	1	0
6	A	32	0	28	2	0
6	B	32	0	28	5	0
6	D	32	0	28	3	0
6	E	32	0	28	2	0
6	H	32	0	28	3	0
6	I	32	0	28	2	0
7	J	46	0	40	2	0
8	A	359	0	0	6	0
8	B	328	0	0	3	0
8	C	330	0	0	4	0
8	D	283	0	0	1	0
8	E	346	0	0	4	0
8	F	275	0	0	1	0
8	G	230	0	0	1	0
8	H	235	0	0	0	0
8	I	246	0	0	1	0
8	J	268	0	0	1	0
All	All	24643	0	21662	95	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 2.

The worst 5 of 95 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:I:36[B]:ARG:HH11	1:I:36[B]:ARG:CG	1.71	1.02
1:I:36[B]:ARG:HG3	1:I:36[B]:ARG:HH11	1.25	0.99
1:I:36[B]:ARG:CB	1:I:36[B]:ARG:NH1	2.31	0.94
1:I:36[B]:ARG:HB3	1:I:36[B]:ARG:NH1	1.83	0.92
1:I:226[B]:THR:HG23	8:I:713:HOH:O	1.69	0.91

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	285/278 (102%)	280 (98%)	5 (2%)	0	100	100
1	B	278/278 (100%)	270 (97%)	8 (3%)	0	100	100
1	C	285/278 (102%)	278 (98%)	7 (2%)	0	100	100
1	D	283/278 (102%)	276 (98%)	7 (2%)	0	100	100
1	E	282/278 (101%)	276 (98%)	6 (2%)	0	100	100
1	F	279/278 (100%)	271 (97%)	8 (3%)	0	100	100
1	G	272/278 (98%)	263 (97%)	9 (3%)	0	100	100
1	H	279/278 (100%)	272 (98%)	7 (2%)	0	100	100
1	I	279/278 (100%)	273 (98%)	6 (2%)	0	100	100
1	J	281/278 (101%)	275 (98%)	6 (2%)	0	100	100
All	All	2803/2780 (101%)	2734 (98%)	69 (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	247/237 (104%)	246 (100%)	1 (0%)	93	90
1	B	242/237 (102%)	242 (100%)	0	100	100
1	C	247/237 (104%)	247 (100%)	0	100	100
1	D	245/237 (103%)	245 (100%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	E	244/237 (103%)	244 (100%)	0	100	100
1	F	241/237 (102%)	241 (100%)	0	100	100
1	G	238/237 (100%)	238 (100%)	0	100	100
1	H	243/237 (102%)	241 (99%)	2 (1%)	86	77
1	I	243/237 (102%)	243 (100%)	0	100	100
1	J	242/237 (102%)	242 (100%)	0	100	100
All	All	2432/2370 (103%)	2429 (100%)	3 (0%)	95	93

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	25	SER
1	H	25	SER
1	H	97	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	182	ASN
1	F	44	GLN
1	J	182	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 ⓘ

15 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$
6	SIA	A	406	6	16,20,21	0.42	0	18,28,31	0.85	0
6	GAL	A	407	6	12,12,12	0.50	0	17,17,17	0.81	0
6	SIA	B	409	6	16,20,21	0.54	0	18,28,31	1.25	1 (5%)
6	GAL	B	410	6	12,12,12	0.49	0	17,17,17	1.56	3 (17%)
6	SIA	D	406	6	16,20,21	0.48	0	18,28,31	0.87	0
6	GAL	D	407	6	12,12,12	0.49	0	17,17,17	0.77	0
6	SIA	E	407	6	16,20,21	0.53	0	18,28,31	0.97	1 (5%)
6	GAL	E	408	6	12,12,12	0.55	0	17,17,17	0.84	0
6	SIA	H	405	6	16,20,21	0.40	0	18,28,31	0.73	0
6	GAL	H	406	6	12,12,12	0.54	0	17,17,17	0.58	0
6	SIA	I	406	6	16,20,21	0.55	0	18,28,31	0.74	0
6	GAL	I	407	6	12,12,12	0.54	0	17,17,17	0.75	0
7	SIA	J	407	7	16,20,21	0.49	0	18,28,31	0.93	2 (11%)
7	NAG	J	408	7	15,15,15	0.51	0	17,21,21	1.10	1 (5%)
7	GAL	J	409	7	11,11,12	0.59	0	14,15,17	1.05	1 (7%)

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	SIA	A	406	6	-	0/14/34/38	0/1/1/1
6	GAL	A	407	6	-	0/2/22/22	0/1/1/1
6	SIA	B	409	6	-	0/14/34/38	0/1/1/1
6	GAL	B	410	6	-	0/2/22/22	0/1/1/1
6	SIA	D	406	6	-	0/14/34/38	0/1/1/1
6	GAL	D	407	6	-	0/2/22/22	0/1/1/1
6	SIA	E	407	6	-	0/14/34/38	0/1/1/1
6	GAL	E	408	6	-	0/2/22/22	0/1/1/1
6	SIA	H	405	6	-	0/14/34/38	0/1/1/1
6	GAL	H	406	6	-	0/2/22/22	0/1/1/1
6	SIA	I	406	6	-	0/14/34/38	0/1/1/1
6	GAL	I	407	6	-	0/2/22/22	0/1/1/1
7	SIA	J	407	7	-	0/14/34/38	0/1/1/1
7	NAG	J	408	7	-	0/6/26/26	0/1/1/1
7	GAL	J	409	7	-	0/2/19/22	0/1/1/1

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	B	409	SIA	C7-C6-C5	-4.02	108.23	114.32
7	J	407	SIA	C8-C7-C6	-2.00	108.98	113.01
7	J	407	SIA	C3-C4-C5	2.01	113.71	111.47
6	B	410	GAL	O3-C3-C4	2.11	115.09	110.34
6	E	407	SIA	C3-C4-C5	2.28	114.01	111.47

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

12 monomers are involved in 19 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	A	406	SIA	2	0
6	B	409	SIA	4	0
6	B	410	GAL	2	0
6	D	406	SIA	2	0
6	D	407	GAL	1	0
6	E	407	SIA	1	0
6	E	408	GAL	1	0
6	H	406	GAL	3	0
6	I	406	SIA	1	0
6	I	407	GAL	1	0
7	J	408	NAG	1	0
7	J	409	GAL	1	0

## 5.6 Ligand geometry

Of 53 ligands modelled in this entry, 20 are monoatomic - leaving 33 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
4	EDO	A	403	-	3,3,3	0.41	0	2,2,2	0.52	0
4	EDO	A	404	-	3,3,3	0.55	0	2,2,2	0.32	0
5	IPA	A	405	-	3,3,3	0.52	0	3,3,3	0.58	0
4	EDO	B	403	-	3,3,3	0.42	0	2,2,2	0.62	0
4	EDO	B	404	-	3,3,3	0.49	0	2,2,2	0.35	0
4	EDO	B	405	-	3,3,3	0.42	0	2,2,2	0.61	0
4	EDO	B	406	-	3,3,3	0.51	0	2,2,2	0.53	0
4	EDO	B	407	-	3,3,3	0.46	0	2,2,2	0.61	0
5	IPA	B	408	-	3,3,3	0.50	0	3,3,3	0.70	0
4	EDO	C	403	-	3,3,3	0.45	0	2,2,2	0.29	0
4	EDO	C	404	-	3,3,3	0.43	0	2,2,2	0.48	0
5	IPA	C	405	-	3,3,3	0.40	0	3,3,3	0.52	0
4	EDO	D	403	-	3,3,3	0.46	0	2,2,2	0.64	0
4	EDO	D	404	-	3,3,3	0.45	0	2,2,2	0.45	0
5	IPA	D	405	-	3,3,3	0.44	0	3,3,3	0.61	0
4	EDO	E	403	-	3,3,3	0.53	0	2,2,2	0.29	0
4	EDO	E	404	-	3,3,3	0.46	0	2,2,2	0.54	0
4	EDO	E	405	-	3,3,3	0.55	0	2,2,2	0.31	0
5	IPA	E	406	-	3,3,3	0.44	0	3,3,3	0.53	0
4	EDO	F	403	-	3,3,3	0.40	0	2,2,2	0.74	0
5	IPA	F	404	-	3,3,3	0.48	0	3,3,3	0.40	0
4	EDO	G	403	-	3,3,3	0.52	0	2,2,2	0.36	0
4	EDO	G	404	-	3,3,3	0.48	0	2,2,2	0.44	0
5	IPA	G	405	-	3,3,3	0.42	0	3,3,3	0.52	0
4	EDO	H	403	-	3,3,3	0.50	0	2,2,2	0.45	0
5	IPA	H	404	-	3,3,3	0.47	0	3,3,3	0.45	0
4	EDO	I	403	-	3,3,3	0.43	0	2,2,2	0.53	0
4	EDO	I	404	-	3,3,3	0.43	0	2,2,2	0.35	0
5	IPA	I	405	-	3,3,3	0.52	0	3,3,3	0.51	0
4	EDO	J	403	-	3,3,3	0.50	0	2,2,2	0.47	0
4	EDO	J	404	-	3,3,3	0.44	0	2,2,2	0.42	0
4	EDO	J	405	-	3,3,3	0.54	0	2,2,2	0.43	0
5	IPA	J	406	-	3,3,3	0.44	0	3,3,3	0.48	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	EDO	A	403	-	-	0/1/1/1	0/0/0/0
4	EDO	A	404	-	-	0/1/1/1	0/0/0/0
5	IPA	A	405	-	-	0/0/0/0	0/0/0/0
4	EDO	B	403	-	-	0/1/1/1	0/0/0/0
4	EDO	B	404	-	-	0/1/1/1	0/0/0/0
4	EDO	B	405	-	-	0/1/1/1	0/0/0/0
4	EDO	B	406	-	-	0/1/1/1	0/0/0/0
4	EDO	B	407	-	-	0/1/1/1	0/0/0/0
5	IPA	B	408	-	-	0/0/0/0	0/0/0/0
4	EDO	C	403	-	-	0/1/1/1	0/0/0/0
4	EDO	C	404	-	-	0/1/1/1	0/0/0/0
5	IPA	C	405	-	-	0/0/0/0	0/0/0/0
4	EDO	D	403	-	-	0/1/1/1	0/0/0/0
4	EDO	D	404	-	-	0/1/1/1	0/0/0/0
5	IPA	D	405	-	-	0/0/0/0	0/0/0/0
4	EDO	E	403	-	-	0/1/1/1	0/0/0/0
4	EDO	E	404	-	-	0/1/1/1	0/0/0/0
4	EDO	E	405	-	-	0/1/1/1	0/0/0/0
5	IPA	E	406	-	-	0/0/0/0	0/0/0/0
4	EDO	F	403	-	-	0/1/1/1	0/0/0/0
5	IPA	F	404	-	-	0/0/0/0	0/0/0/0
4	EDO	G	403	-	-	0/1/1/1	0/0/0/0
4	EDO	G	404	-	-	0/1/1/1	0/0/0/0
5	IPA	G	405	-	-	0/0/0/0	0/0/0/0
4	EDO	H	403	-	-	0/1/1/1	0/0/0/0
5	IPA	H	404	-	-	0/0/0/0	0/0/0/0
4	EDO	I	403	-	-	0/1/1/1	0/0/0/0
4	EDO	I	404	-	-	0/1/1/1	0/0/0/0
5	IPA	I	405	-	-	0/0/0/0	0/0/0/0
4	EDO	J	403	-	-	0/1/1/1	0/0/0/0
4	EDO	J	404	-	-	0/1/1/1	0/0/0/0
4	EDO	J	405	-	-	0/1/1/1	0/0/0/0
5	IPA	J	406	-	-	0/0/0/0	0/0/0/0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

18 monomers are involved in 18 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	403	EDO	1	0
4	A	404	EDO	1	0
5	A	405	IPA	2	0
4	B	405	EDO	1	0
4	B	406	EDO	5	0
5	B	408	IPA	1	0
5	C	405	IPA	1	0
4	D	404	EDO	1	0
5	D	405	IPA	1	0
4	E	405	EDO	1	0
4	G	403	EDO	1	0
5	G	405	IPA	1	0
4	H	403	EDO	1	0
5	H	404	IPA	1	0
4	I	404	EDO	2	0
5	I	405	IPA	1	0
4	J	405	EDO	1	0
5	J	406	IPA	1	0

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	277/278 (99%)	-0.31	4 (1%) 78 84	12, 17, 34, 77	0
1	B	274/278 (98%)	-0.21	4 (1%) 76 82	11, 16, 41, 59	0
1	C	275/278 (98%)	-0.26	5 (1%) 71 78	12, 18, 40, 80	0
1	D	274/278 (98%)	-0.22	10 (3%) 46 52	13, 20, 49, 97	0
1	E	277/278 (99%)	-0.29	3 (1%) 82 87	12, 18, 35, 83	0
1	F	274/278 (98%)	-0.19	10 (3%) 46 52	18, 25, 46, 90	0
1	G	271/278 (97%)	-0.12	7 (2%) 59 65	18, 27, 49, 76	0
1	H	273/278 (98%)	-0.14	7 (2%) 59 65	18, 27, 49, 81	0
1	I	274/278 (98%)	-0.04	15 (5%) 29 34	17, 25, 52, 97	0
1	J	277/278 (99%)	-0.18	10 (3%) 46 52	16, 24, 54, 86	0
All	All	2746/2780 (98%)	-0.20	75 (2%) 58 64	11, 22, 46, 97	0

The worst 5 of 75 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	73	PHE	5.7
1	I	101	ASP	5.7
1	J	24	GLY	5.4
1	I	73	PHE	5.2
1	I	104	CYS	4.7

### 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 [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
6	SIA	I	406	20/21	0.85	0.22	12.40	33,43,60,61	0
6	SIA	D	406	20/21	0.83	0.25	9.96	24,40,55,56	20
6	SIA	E	407	20/21	0.89	0.22	9.44	28,38,55,68	0
6	SIA	B	409	20/21	0.76	0.20	8.42	20,31,37,40	20
7	SIA	J	407	20/21	0.90	0.19	6.67	30,40,57,63	0
6	SIA	A	406	20/21	0.85	0.20	6.53	21,30,45,52	20
6	SIA	H	405	20/21	0.83	0.20	4.82	28,36,45,45	20
6	GAL	E	408	12/12	0.87	0.21	4.78	39,49,54,55	0
6	GAL	B	410	12/12	0.72	0.26	4.54	32,35,40,40	12
6	GAL	A	407	12/12	0.82	0.22	4.16	28,34,38,45	12
6	GAL	I	407	12/12	0.83	0.27	3.12	40,50,59,60	0
6	GAL	D	407	12/12	0.74	0.25	2.45	36,45,48,48	12
6	GAL	H	406	12/12	0.75	0.26	2.03	34,40,47,50	12
7	GAL	J	409	11/12	0.87	0.24	1.54	35,42,47,47	0
7	NAG	J	408	15/15	0.86	0.28	-	53,63,86,88	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
5	IPA	A	405	4/4	0.87	0.23	11.26	18,28,29,40	0
5	IPA	B	408	4/4	0.90	0.22	7.74	20,27,28,29	0
5	IPA	C	405	4/4	0.96	0.18	7.05	28,29,29,30	0
5	IPA	J	406	4/4	0.92	0.20	5.50	30,34,35,39	0
4	EDO	A	403	4/4	0.94	0.18	5.17	23,33,36,46	0
4	EDO	J	405	4/4	0.83	0.18	4.98	32,38,44,50	0
5	IPA	G	405	4/4	0.88	0.18	4.73	29,34,35,35	0
4	EDO	E	404	4/4	0.86	0.11	4.44	39,40,44,45	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
5	IPA	D	405	4/4	0.93	0.17	4.44	26,32,33,36	0
4	EDO	C	403	4/4	0.96	0.16	4.06	23,32,40,45	0
5	IPA	F	404	4/4	0.92	0.18	3.93	37,37,38,40	0
4	EDO	B	407	4/4	0.92	0.14	3.57	32,35,48,50	0
4	EDO	B	404	4/4	0.95	0.18	3.48	23,30,43,45	0
5	IPA	H	404	4/4	0.93	0.16	3.44	33,34,34,40	0
4	EDO	F	403	4/4	0.92	0.17	2.94	28,40,46,49	0
5	IPA	I	405	4/4	0.89	0.17	2.60	29,31,32,33	0
5	IPA	E	406	4/4	0.95	0.16	2.36	25,27,28,28	0
4	EDO	I	403	4/4	0.93	0.21	1.58	36,43,50,51	0
4	EDO	E	405	4/4	0.85	0.14	1.32	34,36,37,43	0
4	EDO	A	404	4/4	0.96	0.12	1.10	18,35,37,39	0
4	EDO	J	403	4/4	0.93	0.14	0.77	27,38,46,46	0
4	EDO	J	404	4/4	0.98	0.12	0.71	35,37,46,48	0
4	EDO	C	404	4/4	0.96	0.11	0.64	30,36,38,39	0
4	EDO	D	403	4/4	0.93	0.15	0.59	31,34,44,59	0
4	EDO	B	403	4/4	0.96	0.10	0.41	26,33,34,37	0
4	EDO	D	404	4/4	0.96	0.12	0.41	23,32,33,46	0
4	EDO	G	404	4/4	0.91	0.09	0.27	41,45,47,50	0
4	EDO	G	403	4/4	0.97	0.10	-0.40	23,38,38,44	0
4	EDO	I	404	4/4	0.97	0.10	-0.42	24,32,35,35	0
3	CL	I	402	1/1	0.98	0.07	-0.51	33,33,33,33	0
4	EDO	B	405	4/4	0.96	0.12	-0.64	20,33,34,37	0
4	EDO	H	403	4/4	0.95	0.09	-0.65	28,36,39,42	0
4	EDO	E	403	4/4	0.96	0.11	-0.67	20,34,36,38	0
3	CL	G	402	1/1	0.95	0.07	-0.80	37,37,37,37	0
3	CL	E	402	1/1	0.99	0.06	-0.83	28,28,28,28	0
3	CL	H	402	1/1	0.98	0.06	-0.90	35,35,35,35	0
3	CL	J	402	1/1	0.99	0.03	-1.26	31,31,31,31	0
3	CL	D	402	1/1	0.99	0.04	-1.33	28,28,28,28	0
3	CL	F	402	1/1	0.99	0.05	-1.47	34,34,34,34	0
2	CA	G	401	1/1	0.97	0.06	-1.48	24,24,24,24	1
2	CA	D	401	1/1	0.98	0.04	-1.51	21,21,21,21	1
4	EDO	B	406	4/4	0.94	0.11	-1.56	27,32,33,39	0
3	CL	B	402	1/1	1.00	0.04	-1.62	27,27,27,27	0
3	CL	A	402	1/1	1.00	0.04	-1.69	25,25,25,25	0
3	CL	C	402	1/1	0.99	0.04	-3.62	26,26,26,26	0
2	CA	J	401	1/1	0.97	0.04	-	26,26,26,26	1
2	CA	B	401	1/1	0.99	0.06	-	20,20,20,20	1
2	CA	A	401	1/1	0.98	0.04	-	19,19,19,19	1
2	CA	H	401	1/1	0.97	0.07	-	40,40,40,40	0
2	CA	I	401	1/1	0.99	0.04	-	23,23,23,23	1

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	CA	F	401	1/1	0.97	0.04	-	23,23,23,23	1
2	CA	C	401	1/1	0.97	0.05	-	23,23,23,23	1
2	CA	E	401	1/1	0.99	0.04	-	20,20,20,20	1

## 6.5 Other polymers [i](#)

There are no such residues in this entry.