



# Full wwPDB X-ray Structure Validation Report

Oct 30, 2014 – 09:35 PM EDT

PDB ID : 4UQ7  
Title : Crystal Structure of Fucose binding lectin from Aspergillus Fumigatus (AFL) in complex with L-galactopyranose.  
Authors : houser, J.; cioci, g.; komarek, J.; wimmerowa, M.; kostlanova, N.; lahmman, M.; varrot, A.; Imberty, A.  
Deposited on : 2014-06-20  
Resolution : 1.76 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>

---

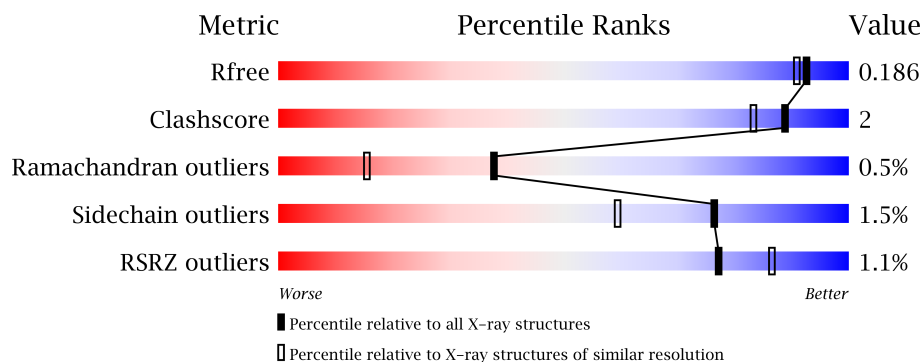
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.16 November 2013  
Xtriage (Phenix) : dev-1439  
EDS : stable24103  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.1.3  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable24103

# 1 Overall quality at a glance





The reported resolution of this entry is 1.76 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	1134 (1.76-1.76)
Clashscore	79885	1304 (1.76-1.76)
Ramachandran outliers	78287	1288 (1.76-1.76)
Sidechain outliers	78261	1288 (1.76-1.76)
RSRZ outliers	66119	1135 (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. 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	315	
2	B	315	
2	D	315	
3	C	315	

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

Mol	Type	Chain	Res	Geometry	Electron density
10	NA	B	800	-	X
5	GXL	D	932[B]	-	X
7	PG4	A	940	-	X
8	PEG	A	950	-	X

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Geometry	Electron density
8	PEG	C	919	-	X
8	PEG	C	959	-	X
8	PEG	D	959	-	X

## 2 Entry composition

There are 12 unique types of molecules in this entry. The entry contains 11349 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 FUCOSE-SPECIFIC LECTIN FLEA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	314	Total	C	N	O	S	0	1	0
			2449	1560	423	461	5			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	20	SER	LEU	CONFLICT	UNP Q4WW81
A	111	CYS	ARG	ENGINEERED MUTATION	UNP Q4WW81

- Molecule 2 is a protein called FUCOSE-SPECIFIC LECTIN FLEA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	314	Total	C	N	O	S	0	3	0
			2455	1564	421	465	5			
2	D	314	Total	C	N	O	S	0	1	0
			2448	1559	421	464	4			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	20	SER	LEU	CONFLICT	UNP Q4WW81
B	111	CSD	ARG	ENGINEERED MUTATION	UNP Q4WW81
D	20	SER	LEU	CONFLICT	UNP Q4WW81
D	111	CSD	ARG	ENGINEERED MUTATION	UNP Q4WW81

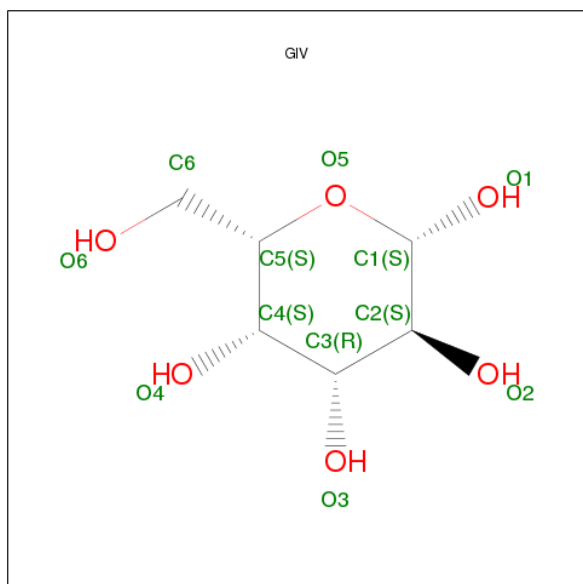
- Molecule 3 is a protein called FUCOSE-SPECIFIC LECTIN FLEA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	314	Total	C	N	O	S	0	2	0
			2449	1558	421	466	4			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	20	SER	LEU	CONFLICT	UNP Q4WW81
C	111	CSD	ARG	ENGINEERED MUTATION	UNP Q4WW81

- Molecule 4 is SUGAR (BETA-L-GALACTOPYRANOSE) (three-letter code: GIV) (formula:  $C_6H_{12}O_6$ ).



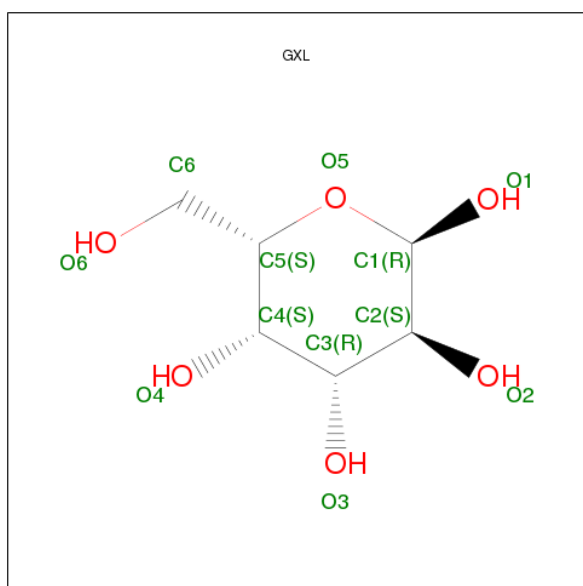
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 12 6 6	0	1
4	A	1	Total C O 12 6 6	0	1
4	A	1	Total C O 12 6 6	0	1
4	B	1	Total C O 12 6 6	0	1
4	B	1	Total C O 12 6 6	0	1
4	B	1	Total C O 12 6 6	0	1
4	C	1	Total C O 12 6 6	0	1
4	C	1	Total C O 12 6 6	0	1
4	C	1	Total C O 12 6 6	0	1
4	C	1	Total C O 12 6 6	0	1

*Continued on next page...*

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	D	1	Total	C	O	0	0
			12	6	6		
4	D	1	Total	C	O	0	1
			12	6	6		
4	D	1	Total	C	O	0	1
			12	6	6		

- Molecule 5 is SUGAR (ALPHA-L-GALACTOPYRANOSE) (three-letter code: GXL) (formula: C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>).



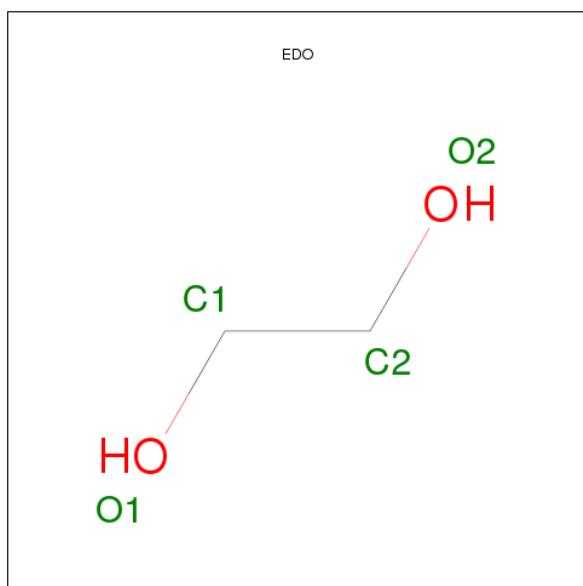
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	1
			12	6	6		
5	A	1	Total	C	O	0	0
			12	6	6		
5	A	1	Total	C	O	0	1
			12	6	6		
5	A	1	Total	C	O	0	1
			12	6	6		
5	B	1	Total	C	O	0	1
			12	6	6		
5	B	1	Total	C	O	0	0
			12	6	6		
5	B	1	Total	C	O	0	1
			12	6	6		
5	B	1	Total	C	O	0	1
			12	6	6		

Continued on next page...

Continued from previous page...

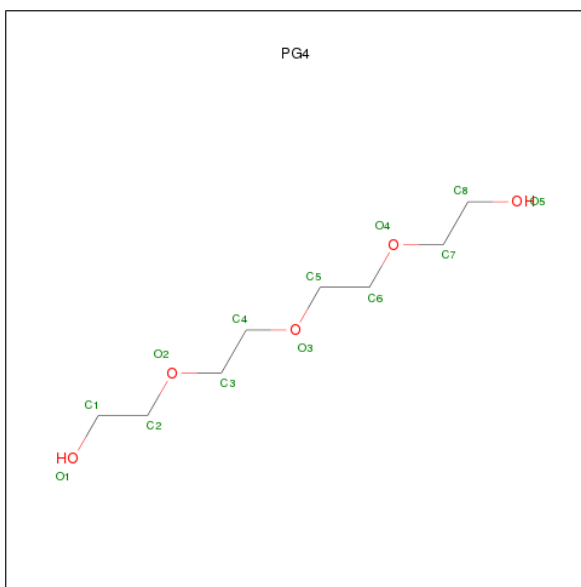
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	C	1	Total	C	O	0	1
			12	6	6		
5	C	1	Total	C	O	0	1
			12	6	6		
5	C	1	Total	C	O	0	1
			12	6	6		
5	C	1	Total	C	O	0	1
			12	6	6		
5	D	1	Total	C	O	0	0
			12	6	6		
5	D	1	Total	C	O	0	1
			12	6	6		
5	D	1	Total	C	O	0	1
			12	6	6		

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



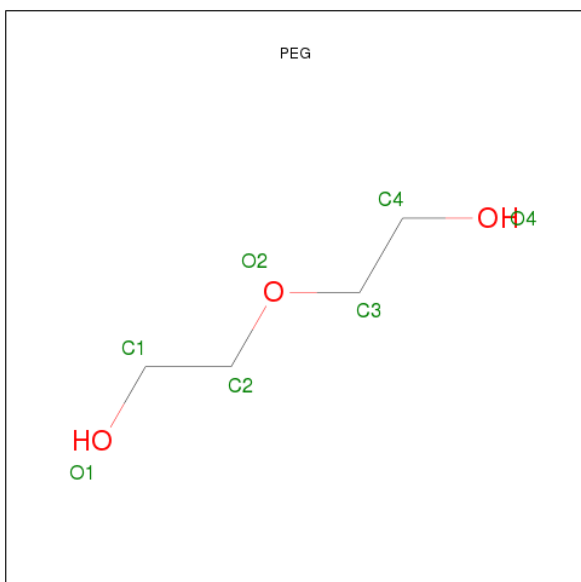
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	C	O	0	0
			4	2	2		
6	B	1	Total	C	O	0	0
			4	2	2		

- Molecule 7 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: C<sub>8</sub>H<sub>18</sub>O<sub>5</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	A	1	Total	C	O	0	0
			13	8	5		
7	B	1	Total	C	O	0	0
			13	8	5		
7	C	1	Total	C	O	0	0
			13	8	5		
7	C	1	Total	C	O	0	0
			13	8	5		
7	D	1	Total	C	O	0	0
			13	8	5		

- Molecule 8 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula:  $C_4H_{10}O_3$ ).





Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	A	1	Total C O 7 4 3	0	0
8	B	1	Total C O 7 4 3	0	0
8	C	1	Total C O 7 4 3	0	0
8	C	1	Total C O 7 4 3	0	0
8	D	1	Total C O 7 4 3	0	0

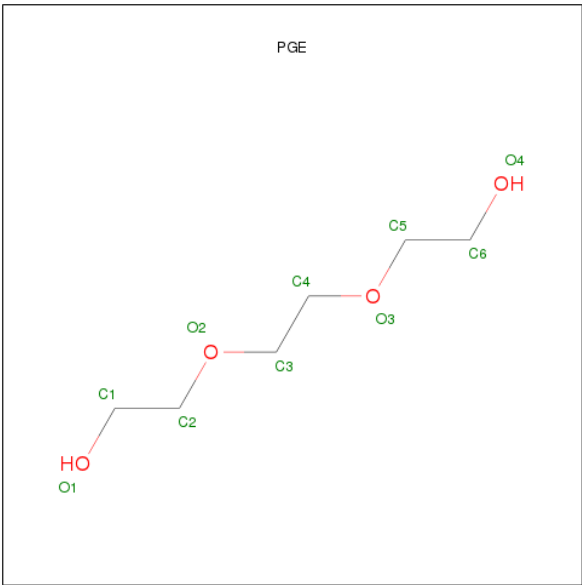
- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

- Molecule 10 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
10	B	1	Total Na 1 1	0	0
10	D	1	Total Na 1 1	0	0
10	C	1	Total Na 1 1	0	0

- Molecule 11 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: C<sub>6</sub>H<sub>14</sub>O<sub>4</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
11	B	1	Total	C	O	0	0
			10	6	4		
11	C	1	Total	C	O	0	0
			10	6	4		
11	D	1	Total	C	O	0	0
			10	6	4		

- Molecule 12 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
12	A	348	Total	O	0	0
			348	348		
12	B	248	Total	O	0	0
			248	248		
12	C	255	Total	O	0	0
			255	255		
12	D	219	Total	O	0	0
			219	219		

### 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.

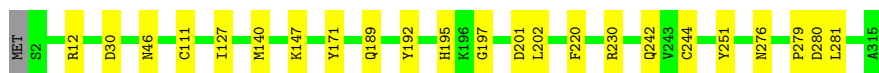
- Molecule 1: FUCOSE-SPECIFIC LECTIN FLEA

Chain A: 



- Molecule 2: FUCOSE-SPECIFIC LECTIN FLEA

Chain B: 



- Molecule 2: FUCOSE-SPECIFIC LECTIN FLEA

Chain D: 



- Molecule 3: FUCOSE-SPECIFIC LECTIN FLEA

Chain C: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	80.03Å 70.44Å 117.80Å 90.00° 108.34° 90.00°	Depositor
Resolution (Å)	111.82 – 1.76 33.59 – 1.76	Depositor EDS
% Data completeness (in resolution range)	97.3 (111.82-1.76) 97.3 (33.59-1.76)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.66 (at 1.76Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, $R_{free}$	0.152 , 0.177 0.164 , 0.186	Depositor DCC
$R_{free}$ test set	6078 reflections (5.29%)	DCC
Wilson B-factor (Å <sup>2</sup> )	13.6	Xtriage
Anisotropy	0.796	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 36.1	EDS
Estimated twinning fraction	0.025 for h,-k,-h-l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 121067 reflections	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	11349	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	18.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.86% 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: GIV, ZN, PGE, GXL, CSD, EDO, PG4, NA, PEG

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\# Z  > 5$	RMSZ	$\# Z  > 5$
1	A	0.67	0/2520	0.79	2/3435 (0.1%)
2	B	0.67	0/2527	0.78	3/3445 (0.1%)
2	D	0.62	0/2514	0.79	3/3428 (0.1%)
3	C	0.64	0/2507	0.80	2/3420 (0.1%)
All	All	0.65	0/10068	0.79	10/13728 (0.1%)

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
1	A	0	1

There are no bond length outliers.

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	30	ASP	CB-CG-OD1	8.67	126.10	118.30
3	C	39	LEU	CA-CB-CG	7.63	132.84	115.30
1	A	167	ARG	NE-CZ-NH1	6.44	123.52	120.30
2	B	12	ARG	NE-CZ-NH2	-6.40	117.10	120.30
2	D	39	LEU	CA-CB-CG	6.28	129.74	115.30
3	C	12	ARG	NE-CZ-NH2	-5.94	117.33	120.30
2	B	30	ASP	CB-CG-OD2	-5.76	113.12	118.30
2	D	12	ARG	NE-CZ-NH2	-5.73	117.44	120.30
2	D	30	ASP	CB-CG-OD2	-5.24	113.58	118.30
1	A	12	ARG	NE-CZ-NH2	-5.05	117.78	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	115	VAL	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	2449	0	2333	8	0
2	B	2455	0	2336	17	1
2	D	2448	0	2325	5	0
3	C	2449	0	2313	15	0
4	A	36	0	36	0	0
4	B	36	0	36	0	0
4	C	48	0	48	1	0
4	D	36	0	36	0	0
5	A	48	0	48	1	0
5	B	48	0	48	0	0
5	C	48	0	48	0	0
5	D	36	0	36	0	0
6	A	4	0	6	1	0
6	B	4	0	6	0	0
7	A	13	0	18	0	0
7	B	13	0	18	1	0
7	C	26	0	36	1	0
7	D	13	0	18	1	0
8	A	7	0	10	0	0
8	B	7	0	10	0	0
8	C	14	0	20	0	0
8	D	7	0	10	0	0
9	B	1	0	0	0	0
10	B	1	0	0	0	0
10	C	1	0	0	0	0
10	D	1	0	0	0	0
11	B	10	0	14	7	0
11	C	10	0	14	0	0
11	D	10	0	14	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
12	A	348	0	0	3	0
12	B	248	0	0	3	0
12	C	255	0	0	3	0
12	D	219	0	0	0	0
All	All	11349	0	9837	47	1

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 2.

All (47) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
2:B:281:LEU:HD11	11:B:950:PGE:H52	1.60	0.82
2:B:242:GLN:HE22	11:B:950:PGE:H3	1.45	0.81
2:B:201[A]:ASP:OD1	12:B:2162:HOH:O	1.99	0.80
2:D:246:ASP:OD2	2:D:252:HIS:HE1	1.66	0.79
3:C:111:CSD:OD2	12:C:2121:HOH:O	1.99	0.78
2:B:230:ARG:HE	11:B:950:PGE:H5	1.49	0.78
1:A:246:ASP:OD2	1:A:252:HIS:HE1	1.68	0.76
1:A:167:ARG:HH11	1:A:167:ARG:HG2	1.56	0.70
2:B:111:CSD:OD1	12:B:2109:HOH:O	2.08	0.69
3:C:277:ASN:OD1	3:C:301:ARG:HD2	1.98	0.64
2:B:230:ARG:CZ	2:B:244[A]:CYS:SG	2.87	0.63
3:C:246:ASP:OD2	3:C:252:HIS:HE1	1.81	0.62
5:A:911[B]:GXL:O1	12:A:2133:HOH:O	2.18	0.55
2:B:140:MET:HE2	2:B:147:LYS:HE2	1.88	0.55
1:A:196:LYS:HD2	1:A:199:TYR:CE2	2.42	0.55
2:D:230:ARG:HE	11:D:950:PGE:H62	1.72	0.55
3:C:95:TYR:OH	4:C:920[A]:GIV:O1	2.25	0.54
2:B:189:GLN:NE2	2:B:251:TYR:OH	2.40	0.54
3:C:189:GLN:HE22	7:C:940:PG4:H42	1.73	0.54
2:B:230:ARG:NH1	2:B:244[A]:CYS:SG	2.81	0.54
2:B:189:GLN:HE22	7:B:940:PG4:H42	1.75	0.51
2:B:220:PHE:CZ	2:B:279:PRO:HG3	2.46	0.51
2:B:230:ARG:HE	11:B:950:PGE:C5	2.22	0.50
2:B:281:LEU:CD1	11:B:950:PGE:H52	2.37	0.50
3:C:301:ARG:HG2	12:C:2225:HOH:O	2.11	0.49
3:C:246:ASP:OD2	3:C:252:HIS:CE1	2.65	0.48
3:C:303:HIS:HB2	12:C:2236:HOH:O	2.14	0.47
2:D:246:ASP:OD2	2:D:252:HIS:CE1	2.56	0.47
2:B:230:ARG:NE	11:B:950:PGE:H5	2.25	0.46
1:A:167:ARG:HH11	1:A:167:ARG:CG	2.27	0.45
6:A:919:EDO:H12	12:A:2342:HOH:O	2.15	0.45

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Distance(Å)	Clash(Å)
3:C:34:SER:OG	3:C:55:ASN:ND2	2.49	0.45
1:A:186:LYS:HG2	1:A:207:ASP:O	2.16	0.45
3:C:297:TRP:CH2	3:C:299:TRP:HB2	2.52	0.44
2:D:189:GLN:HE22	7:D:940:PG4:H42	1.83	0.44
3:C:283:LEU:HD12	3:C:297:TRP:CE3	2.53	0.44
3:C:273:SER:C	3:C:275:ALA:N	2.71	0.43
2:B:171:TYR:HE2	2:B:195:HIS:CE1	2.36	0.43
3:C:274:PHE:HD1	3:C:301:ARG:NH2	2.17	0.43
1:A:152:LEU:HD13	1:A:178:ILE:HD11	2.00	0.43
2:B:192:TYR:CE2	2:B:197:GLY:HA2	2.54	0.42
2:D:34:SER:OG	2:D:55:ASN:ND2	2.52	0.42
3:C:126:ARG:C	3:C:127:ILE:HD13	2.41	0.41
1:A:123:LEU:HD13	12:A:2185:HOH:O	2.19	0.41
1:A:137:GLN:OE1	3:C:4:PRO:HG2	2.20	0.41
11:B:950:PGE:H1	12:B:2229:HOH:O	2.19	0.41

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

Atom-1	Atom-2	Distance(Å)	Clash(Å)
2:B:201[B]:ASP:OD1	2:B:276:ASN:ND2[2_746]	2.13	0.07

## 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	313/315 (99%)	304 (97%)	8 (3%)	1 (0%)	50	28
2	B	314/315 (100%)	307 (98%)	6 (2%)	1 (0%)	50	28
2	D	312/315 (99%)	304 (97%)	7 (2%)	1 (0%)	50	28
3	C	312/315 (99%)	300 (96%)	9 (3%)	3 (1%)	22	6
All	All	1251/1260 (99%)	1215 (97%)	30 (2%)	6 (0%)	38	15

All (6) Ramachandran outliers are listed below:



Mol	Chain	Res	Type
1	A	46	ASN
2	B	46	ASN
3	C	46	ASN
2	D	46	ASN
3	C	275	ALA
3	C	276	ASN

### 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	250/250 (100%)	244 (98%)	6 (2%)	61	34
2	B	250/249 (100%)	247 (99%)	3 (1%)	82	66
2	D	248/249 (100%)	243 (98%)	5 (2%)	68	43
3	C	246/248 (99%)	244 (99%)	2 (1%)	89	80
All	All	994/996 (100%)	978 (98%)	16 (2%)	76	55

All (16) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	39	LEU
1	A	111[A]	CYS
1	A	111[B]	CYS
1	A	167	ARG
1	A	202	LEU
1	A	280	ASP
2	B	127	ILE
2	B	202	LEU
2	B	280	ASP
3	C	39	LEU
3	C	280	ASP
2	D	39	LEU
2	D	202	LEU
2	D	253	ASP
2	D	254	LYS
2	D	280	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (17) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	46	ASN
1	A	55	ASN
1	A	252	HIS
2	B	46	ASN
2	B	55	ASN
2	B	105	GLN
2	B	189	GLN
2	B	242	GLN
3	C	46	ASN
3	C	55	ASN
3	C	189	GLN
3	C	252	HIS
2	D	46	ASN
2	D	55	ASN
2	D	189	GLN
2	D	252	HIS
2	D	303	HIS

### 5.3.3 RNA ⓘ

There are no RNA chains in this entry.

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

4 non-standard protein/DNA/RNA residues are modelled in this entry.

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

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
2	CSD	B	111	2	7,7,8	6.71	3 (42%)	6,8,10	4.78	4 (66%)
3	CSD	C	111	3	7,7,8	7.46	4 (57%)	6,8,10	3.86	3 (50%)
3	CSD	C	244	3	7,7,8	6.84	3 (42%)	6,8,10	1.93	2 (33%)
2	CSD	D	111	2	7,7,8	7.67	2 (28%)	6,8,10	2.85	3 (50%)

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

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	CSD	B	111	2	-	0/3/6/8	0/0/0/0
3	CSD	C	111	3	-	0/3/6/8	0/0/0/0
3	CSD	C	244	3	-	0/3/6/8	0/0/0/0
2	CSD	D	111	2	-	0/3/6/8	0/0/0/0

All (12) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	111	CSD	O-C	20.01	1.25	1.11
3	C	111	CSD	O-C	18.36	1.24	1.11
3	C	244	CSD	O-C	17.69	1.23	1.11
2	B	111	CSD	O-C	17.13	1.23	1.11
3	C	111	CSD	OD1-SG	-6.35	1.40	1.47
2	B	111	CSD	OD1-SG	-2.64	1.44	1.47
3	C	244	CSD	OD1-SG	-2.55	1.44	1.47
2	D	111	CSD	OD1-SG	2.36	1.50	1.47
3	C	111	CSD	CB-SG	-2.22	1.66	1.79
3	C	111	CSD	CB-CA	-2.15	1.49	1.53
2	B	111	CSD	CB-SG	-2.14	1.66	1.79
3	C	244	CSD	CB-SG	-2.09	1.67	1.79

All (12) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	111	CSD	OD1-SG-CB	-8.39	90.77	105.29
3	C	111	CSD	OD2-SG-OD1	7.04	123.07	109.75
2	B	111	CSD	OD2-SG-CB	6.11	120.58	98.82
3	C	111	CSD	CA-CB-SG	-5.38	103.14	110.82
2	D	111	CSD	C-CA-N	-5.10	108.74	113.83
2	B	111	CSD	C-CA-N	-4.71	109.13	113.83
2	D	111	CSD	OD2-SG-OD1	3.52	116.40	109.75
3	C	244	CSD	OD2-SG-OD1	-3.27	103.56	109.75
2	D	111	CSD	OD1-SG-CB	3.17	110.77	105.29
3	C	111	CSD	C-CA-N	-2.97	110.86	113.83
3	C	244	CSD	OD2-SG-CB	2.75	108.62	98.82
2	B	111	CSD	OD2-SG-OD1	-2.51	105.00	109.75

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 47 ligands modelled in this entry, 4 are monoatomic - leaving 43 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	GIV	A	910[A]	-	12,12,12	0.78	0	17,17,17	1.18	1 (5%)
5	GXL	A	911[B]	-	12,12,12	0.87	0	17,17,17	1.26	1 (5%)
6	EDO	A	919	-	3,3,3	0.37	0	2,2,2	0.48	0
5	GXL	A	920	-	12,12,12	0.73	0	17,17,17	0.70	0
4	GIV	A	931[A]	-	12,12,12	0.58	0	17,17,17	0.91	0
5	GXL	A	932[B]	-	12,12,12	0.62	0	17,17,17	1.01	0
7	PG4	A	940	-	12,12,12	0.56	0	11,11,11	1.16	1 (9%)
8	PEG	A	950	-	6,6,6	0.51	0	5,5,5	0.25	0
4	GIV	A	961[A]	-	12,12,12	0.55	0	17,17,17	1.02	1 (5%)
5	GXL	A	962[B]	-	12,12,12	0.57	0	17,17,17	1.02	1 (5%)
4	GIV	B	910[A]	-	12,12,12	0.63	0	17,17,17	0.84	0
5	GXL	B	911[B]	-	12,12,12	0.63	0	17,17,17	0.83	0
6	EDO	B	919	-	3,3,3	0.42	0	2,2,2	0.17	0
5	GXL	B	920	-	12,12,12	0.46	0	17,17,17	0.81	0
4	GIV	B	930[A]	-	12,12,12	0.61	0	17,17,17	0.66	0
5	GXL	B	931[B]	-	12,12,12	0.61	0	17,17,17	0.60	0
7	PG4	B	940	-	12,12,12	0.55	0	11,11,11	0.46	0
11	PGE	B	950	-	9,9,9	0.58	0	8,8,8	0.61	0
8	PEG	B	959	-	6,6,6	0.42	0	5,5,5	0.33	0
4	GIV	B	961[A]	-	12,12,12	0.59	0	17,17,17	1.26	2 (11%)
5	GXL	B	962[B]	-	12,12,12	0.64	0	17,17,17	1.15	1 (5%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
4	GIV	C	911[A]	-	12,12,12	0.53	0	17,17,17	1.08	1 (5%)
5	GXL	C	912[B]	-	12,12,12	0.56	0	17,17,17	1.07	1 (5%)
8	PEG	C	919	-	6,6,6	0.59	0	5,5,5	0.54	0
4	GIV	C	920[A]	-	12,12,12	0.46	0	17,17,17	1.32	3 (17%)
5	GXL	C	921[B]	-	12,12,12	0.57	0	17,17,17	1.07	2 (11%)
4	GIV	C	930[A]	-	12,12,12	0.76	0	17,17,17	0.98	1 (5%)
5	GXL	C	931[B]	-	12,12,12	0.76	0	17,17,17	0.78	0
7	PG4	C	940	-	12,12,12	0.52	0	11,11,11	0.46	0
7	PG4	C	949	-	12,12,12	0.43	0	11,11,11	0.60	0
11	PGE	C	950	-	9,9,9	0.53	0	8,8,8	0.66	0
8	PEG	C	959	-	6,6,6	0.37	0	5,5,5	0.35	0
4	GIV	C	961[A]	-	12,12,12	0.47	0	17,17,17	0.84	0
5	GXL	C	962[B]	-	12,12,12	0.44	0	17,17,17	0.88	0
4	GIV	D	910	-	12,12,12	0.61	0	17,17,17	1.47	4 (23%)
5	GXL	D	920	-	12,12,12	0.77	0	17,17,17	1.14	1 (5%)
4	GIV	D	931[A]	-	12,12,12	0.38	0	17,17,17	1.39	3 (17%)
5	GXL	D	932[B]	-	12,12,12	0.44	0	17,17,17	1.26	2 (11%)
7	PG4	D	940	-	12,12,12	0.62	0	11,11,11	0.49	0
11	PGE	D	950	-	9,9,9	0.59	0	8,8,8	0.37	0
8	PEG	D	959	-	6,6,6	0.34	0	5,5,5	0.45	0
4	GIV	D	960[A]	-	12,12,12	0.66	0	17,17,17	0.89	1 (5%)
5	GXL	D	961[B]	-	12,12,12	0.69	0	17,17,17	0.90	1 (5%)

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

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	GIV	A	910[A]	-	-	0/2/22/22	0/1/1/1
5	GXL	A	911[B]	-	-	0/2/22/22	0/1/1/1
6	EDO	A	919	-	-	0/1/1/1	0/0/0/0
5	GXL	A	920	-	-	0/2/22/22	0/1/1/1
4	GIV	A	931[A]	-	-	0/2/22/22	0/1/1/1
5	GXL	A	932[B]	-	-	0/2/22/22	0/1/1/1
7	PG4	A	940	-	-	0/10/10/10	0/0/0/0
8	PEG	A	950	-	-	0/4/4/4	0/0/0/0
4	GIV	A	961[A]	-	-	0/2/22/22	0/1/1/1
5	GXL	A	962[B]	-	-	0/2/22/22	0/1/1/1
4	GIV	B	910[A]	-	-	0/2/22/22	0/1/1/1

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	GXL	B	911[B]	-	-	0/2/22/22	0/1/1/1
6	EDO	B	919	-	-	0/1/1/1	0/0/0/0
5	GXL	B	920	-	-	0/2/22/22	0/1/1/1
4	GIV	B	930[A]	-	-	0/2/22/22	0/1/1/1
5	GXL	B	931[B]	-	-	0/2/22/22	0/1/1/1
7	PG4	B	940	-	-	0/10/10/10	0/0/0/0
11	PGE	B	950	-	-	0/7/7/7	0/0/0/0
8	PEG	B	959	-	-	0/4/4/4	0/0/0/0
4	GIV	B	961[A]	-	-	0/2/22/22	0/1/1/1
5	GXL	B	962[B]	-	-	0/2/22/22	0/1/1/1
4	GIV	C	911[A]	-	-	0/2/22/22	0/1/1/1
5	GXL	C	912[B]	-	-	0/2/22/22	0/1/1/1
8	PEG	C	919	-	-	0/4/4/4	0/0/0/0
4	GIV	C	920[A]	-	-	0/2/22/22	0/1/1/1
5	GXL	C	921[B]	-	-	0/2/22/22	0/1/1/1
4	GIV	C	930[A]	-	-	0/2/22/22	0/1/1/1
5	GXL	C	931[B]	-	-	0/2/22/22	0/1/1/1
7	PG4	C	940	-	-	0/10/10/10	0/0/0/0
7	PG4	C	949	-	-	0/10/10/10	0/0/0/0
11	PGE	C	950	-	-	0/7/7/7	0/0/0/0
8	PEG	C	959	-	-	0/4/4/4	0/0/0/0
4	GIV	C	961[A]	-	-	0/2/22/22	0/1/1/1
5	GXL	C	962[B]	-	-	0/2/22/22	0/1/1/1
4	GIV	D	910	-	-	0/2/22/22	0/1/1/1
5	GXL	D	920	-	-	0/2/22/22	0/1/1/1
4	GIV	D	931[A]	-	-	0/2/22/22	0/1/1/1
5	GXL	D	932[B]	-	-	0/2/22/22	0/1/1/1
7	PG4	D	940	-	-	0/10/10/10	0/0/0/0
11	PGE	D	950	-	-	0/7/7/7	0/0/0/0
8	PEG	D	959	-	-	0/4/4/4	0/0/0/0
4	GIV	D	960[A]	-	-	0/2/22/22	0/1/1/1
5	GXL	D	961[B]	-	-	0/2/22/22	0/1/1/1

There are no bond length outliers.

All (28) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	D	932[B]	GXL	O5-C1-C2	3.79	115.94	109.80
4	D	931[A]	GIV	O5-C1-C2	3.79	115.94	109.80
4	C	920[A]	GIV	O1-C1-O5	-3.30	101.15	110.29
7	A	940	PG4	C3-O2-C2	3.10	126.76	113.36
4	A	910[A]	GIV	O5-C5-C6	2.96	113.71	106.34

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	911[B]	GXL	O5-C5-C6	2.96	113.71	106.34
4	D	931[A]	GIV	O1-C1-O5	-2.89	102.28	110.29
5	D	920	GXL	C4-C3-C2	-2.77	105.69	110.80
5	C	912[B]	GXL	O5-C1-C2	2.73	114.22	109.80
4	C	911[A]	GIV	O5-C1-C2	2.73	114.22	109.80
4	D	910	GIV	C1-O5-C5	-2.65	108.50	113.43
4	D	910	GIV	O1-C1-C2	-2.64	102.33	109.15
5	D	932[B]	GXL	O2-C2-C3	-2.42	104.94	110.36
4	D	931[A]	GIV	O2-C2-C3	-2.42	104.94	110.36
4	D	910	GIV	O3-C3-C4	2.33	115.56	110.36
5	A	962[B]	GXL	O5-C1-C2	2.27	113.48	109.80
4	A	961[A]	GIV	O5-C1-C2	2.27	113.48	109.80
5	B	962[B]	GXL	O5-C1-C2	2.19	113.35	109.80
4	B	961[A]	GIV	O5-C1-C2	2.19	113.35	109.80
4	D	910	GIV	O1-C1-O5	-2.13	104.41	110.29
4	B	961[A]	GIV	O1-C1-O5	-2.10	104.49	110.29
4	C	920[A]	GIV	O5-C1-C2	2.07	113.15	109.80
5	C	921[B]	GXL	O5-C1-C2	2.07	113.15	109.80
4	D	960[A]	GIV	O3-C3-C4	2.05	114.94	110.36
5	D	961[B]	GXL	O3-C3-C4	2.05	114.94	110.36
4	C	920[A]	GIV	O2-C2-C1	2.03	114.13	109.78
5	C	921[B]	GXL	O2-C2-C1	2.03	114.13	109.78
4	C	930[A]	GIV	O1-C1-O5	-2.03	104.69	110.29

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	314/315 (99%)	-0.38	1 (0%) 91 94	10, 14, 25, 41	0
2	B	314/315 (99%)	-0.34	0 100 100	10, 15, 27, 38	0
2	D	314/315 (99%)	-0.13	8 (2%) 54 65	10, 17, 31, 52	0
3	C	314/315 (99%)	-0.38	5 (1%) 68 79	10, 15, 28, 65	1 (0%)
All	All	1256/1260 (99%)	-0.31	14 (1%) 77 86	10, 15, 28, 65	1 (0%)

All (14) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	275	ALA	4.3
2	D	275	ALA	4.2
2	D	274	PHE	3.6
2	D	121	ASP	3.4
3	C	274	PHE	3.3
3	C	273	SER	3.2
3	C	172	ASN	3.1
2	D	273	SER	2.5
1	A	121	ASP	2.3
2	D	276	ASN	2.2
2	D	91	GLY	2.1
2	D	144	ASP	2.1
2	D	303	HIS	2.1
3	C	276	ASN	2.0

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

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
3	CSD	C	111	8/9	0.10	1.15	13,17,32,40	0
2	CSD	D	111	8/9	0.10	0.53	12,15,27,28	0
3	CSD	C	244	8/9	0.08	0.22	16,17,25,25	2
2	CSD	B	111	8/9	0.09	0.20	11,14,31,35	0

### 6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

### 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	PEG	C	919	7/7	0.14	6.38	25,27,34,35	0
8	PEG	A	950	7/7	0.10	5.15	27,28,29,31	0
8	PEG	C	959	7/7	0.12	4.04	36,37,41,44	0
8	PEG	D	959	7/7	0.17	4.01	39,39,42,48	0
7	PG4	A	940	13/13	0.08	3.32	17,18,30,33	0
5	GXL	D	932[B]	12/12	0.10	2.74	18,21,26,32	12
10	NA	B	800	1/1	0.15	2.19	13,13,13,13	0
7	PG4	C	949	13/13	0.14	1.81	35,39,42,46	0
5	GXL	D	920	12/12	0.10	1.62	17,19,23,24	0
10	NA	D	700	1/1	0.12	1.42	15,15,15,15	0
4	GIV	D	931[A]	12/12	0.09	1.14	18,21,26,31	12
5	GXL	B	931[B]	12/12	0.08	1.12	15,15,18,26	12
4	GIV	B	930[A]	12/12	0.08	1.10	15,15,18,21	12
11	PGE	C	950	10/10	0.13	1.07	24,28,37,39	0
7	PG4	C	940	13/13	0.08	0.96	18,20,32,39	0
7	PG4	B	940	13/13	0.10	0.85	21,23,32,33	0
11	PGE	D	950	10/10	0.14	0.85	28,35,41,43	0
11	PGE	B	950	10/10	0.10	0.65	29,32,38,41	0
6	EDO	A	919	4/4	0.10	0.63	30,32,36,37	0
8	PEG	B	959	7/7	0.10	0.61	34,36,40,41	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
7	PG4	D	940	13/13	0.07	0.51	20,21,38,46	0
4	GIV	A	910[A]	12/12	0.07	0.49	11,12,13,15	12
5	GXL	A	920	12/12	0.07	-0.07	12,15,17,20	0
4	GIV	C	911[A]	12/12	0.06	-0.09	12,14,17,22	12
5	GXL	C	921[B]	12/12	0.06	-0.14	15,17,18,18	12
4	GIV	A	961[A]	12/12	0.06	-0.21	11,15,17,19	12
5	GXL	C	912[B]	12/12	0.06	-0.21	12,14,17,23	12
5	GXL	A	911[B]	12/12	0.06	-0.23	11,12,13,23	12
4	GIV	C	961[A]	12/12	0.06	-0.24	13,17,19,24	12
5	GXL	C	931[B]	12/12	0.07	-0.24	14,17,18,21	12
6	EDO	B	919	4/4	0.07	-0.25	26,27,27,28	0
4	GIV	A	931[A]	12/12	0.06	-0.27	12,16,19,25	12
5	GXL	C	962[B]	12/12	0.06	-0.32	13,17,19,24	12
5	GXL	A	932[B]	12/12	0.06	-0.40	12,16,19,25	12
5	GXL	A	962[B]	12/12	0.06	-0.41	11,15,17,24	12
4	GIV	C	920[A]	12/12	0.06	-0.52	15,17,18,28	12
4	GIV	B	961[A]	12/12	0.06	-0.55	16,18,21,26	12
4	GIV	D	910	12/12	0.07	-0.58	17,19,21,21	0
5	GXL	B	962[B]	12/12	0.06	-0.68	16,18,21,27	12
4	GIV	B	910[A]	12/12	0.06	-0.72	14,15,17,20	12
4	GIV	D	960[A]	12/12	0.06	-0.78	19,21,24,29	12
4	GIV	C	930[A]	12/12	0.05	-0.83	14,17,18,25	12
5	GXL	B	911[B]	12/12	0.06	-0.85	14,15,17,24	12
5	GXL	D	961[B]	12/12	0.06	-0.85	19,21,24,34	12
5	GXL	B	920	12/12	0.05	-1.11	13,15,16,17	0
10	NA	C	800	1/1	0.05	-2.23	21,21,21,21	0
9	ZN	B	700	1/1	0.02	-2.74	17,17,17,17	1

## 6.5 Other polymers ⓘ

There are no such residues in this entry.