



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 22, 2017 – 03:04 PM EDT

PDB ID : 4U1L  
Title : HLA class I micropolymorphisms determine peptide-HLA landscape and dictate differential HIV-1 escape through identical epitopes  
Authors : Rizkallah, P.J.; Cole, D.K.; Fuller, A.; Sewell, A.K.  
Deposited on : unknown  
Resolution : 2.06 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20029824  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20029824

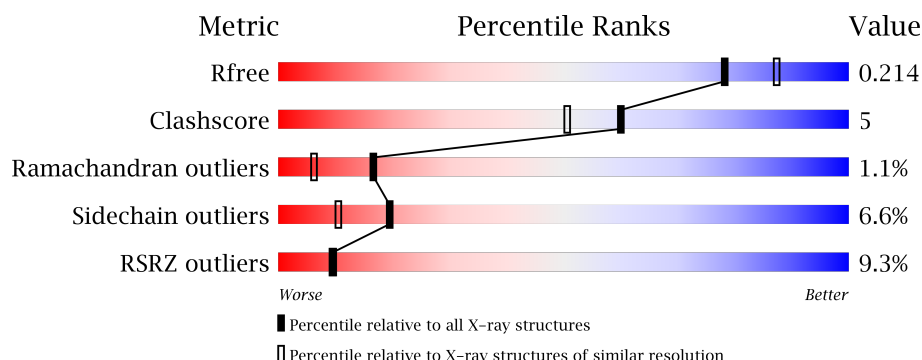
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.06 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



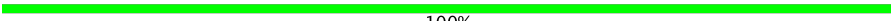
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	100719	2028 (2.08-2.04)
Clashscore	112137	2143 (2.08-2.04)
Ramachandran outliers	110173	2126 (2.08-2.04)
Sidechain outliers	110143	2126 (2.08-2.04)
RSRZ outliers	101464	2035 (2.08-2.04)

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	277	<div> <div>14%</div> <div> <div></div> <div>82%</div> <div>13%</div> <div>••</div> </div> </div>
1	D	277	<div> <div>9%</div> <div> <div></div> <div>82%</div> <div>17%</div> <div>•</div> </div> </div>
2	B	100	<div> <div>3%</div> <div> <div></div> <div>90%</div> <div>10%</div> </div> </div>
2	E	100	<div> <div>6%</div> <div> <div></div> <div>87%</div> <div>12%</div> <div>•</div> </div> </div>
3	C	9	<div> <div></div> <div>100%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	F	9	 100%

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	GOL	A	301	-	X	X	X
4	GOL	A	302	-	-	-	X
4	GOL	D	301	-	-	X	X
5	EDO	A	305	-	-	-	X
5	EDO	D	306	-	-	-	X
5	EDO	D	307	-	-	-	X
5	EDO	D	310	-	-	-	X
5	EDO	E	101	-	-	-	X
5	EDO	E	102	-	-	-	X
5	EDO	F	101	-	-	-	X
6	SO4	A	308	-	-	-	X
6	SO4	A	309	-	-	-	X

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 6806 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 HLA class I histocompatibility antigen, B-81 alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	277	Total	C	N	O	S	0	1	0
			2270	1409	414	440	7			
1	D	277	Total	C	N	O	S	0	3	0
			2284	1417	416	443	8			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	initiating methionine	UNP Q31610
D	0	MET	-	initiating methionine	UNP Q31610

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	100	Total	C	N	O	S	0	0	0
			837	533	141	159	4			
2	E	100	Total	C	N	O	S	0	0	0
			837	533	141	159	4			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	MET	-	initiating methionine	UNP P61769
E	0	MET	-	initiating methionine	UNP P61769

- Molecule 3 is a protein called Protein Nef.

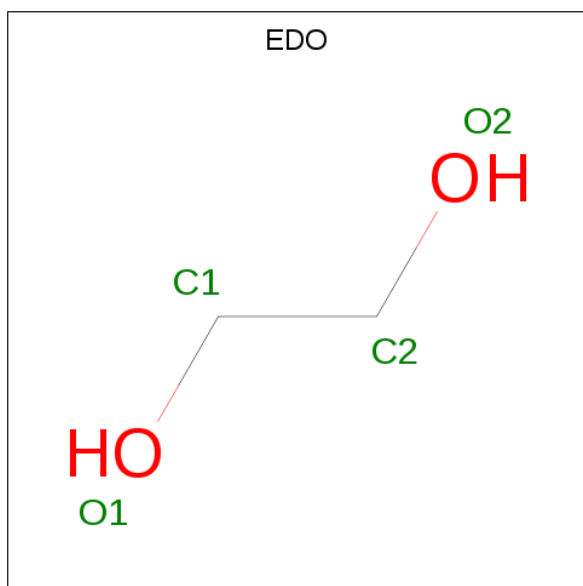
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	9	Total	C	N	O	S	0	0	0
			76	48	16	11	1			
3	F	9	Total	C	N	O	S	0	0	0
			76	48	16	11	1			

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			6	3	3		
4	A	1	Total	C	O	0	0
			6	3	3		
4	D	1	Total	C	O	0	0
			6	3	3		

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).



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

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



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

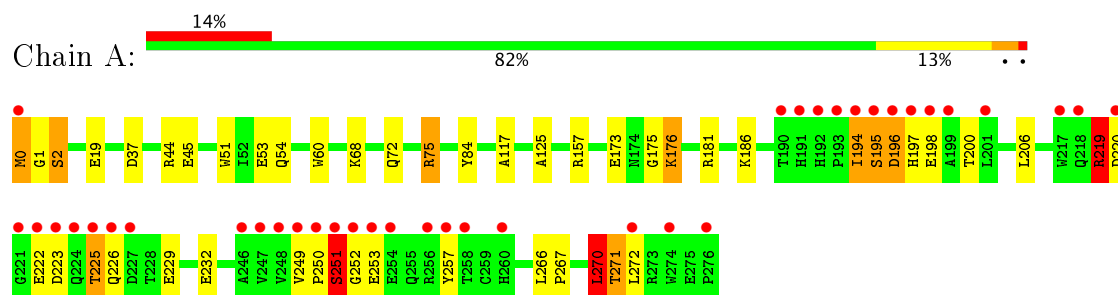
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	138	Total	O	0	0
			138	138		
7	B	28	Total	O	0	0
			28	28		
7	C	7	Total	O	0	0
			7	7		
7	D	125	Total	O	0	0
			125	125		
7	E	22	Total	O	0	0
			22	22		
7	F	6	Total	O	0	0
			6	6		

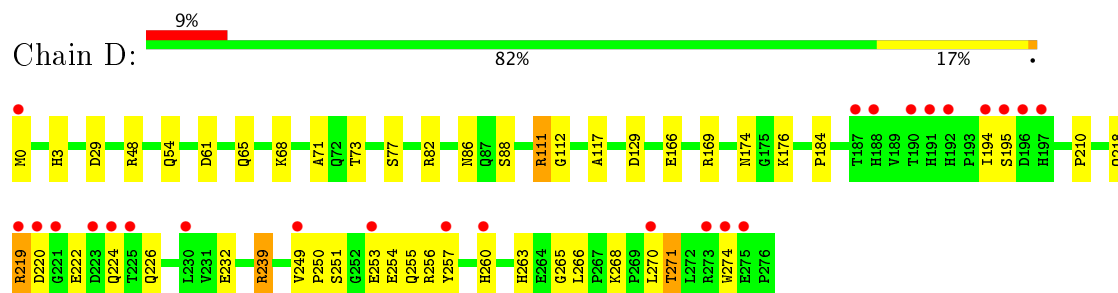
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

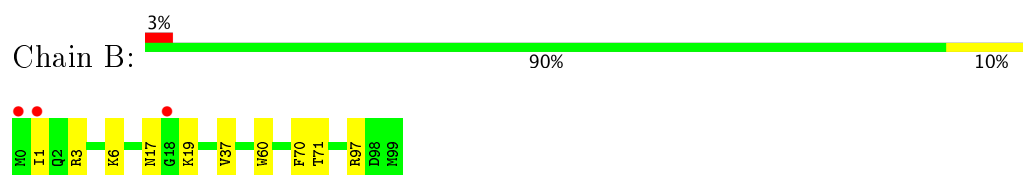
- Molecule 1: HLA class I histocompatibility antigen, B-81 alpha chain



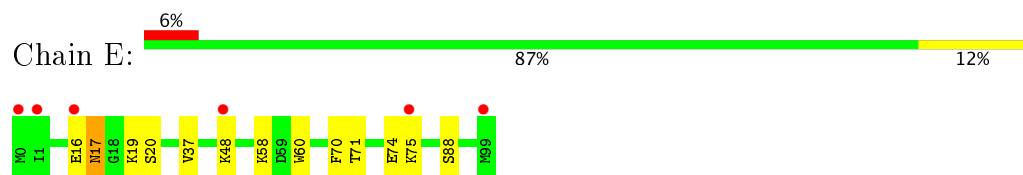
- Molecule 1: HLA class I histocompatibility antigen, B-81 alpha chain



- Molecule 2: Beta-2-microglobulin



- Molecule 2: Beta-2-microglobulin



- Molecule 3: Protein Nef





There are no outlier residues recorded for this chain.

- Molecule 3: Protein Nef

Chain F:  100%

There are no outlier residues recorded for this chain.

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	46.02Å 49.06Å 107.84Å 91.13° 93.53° 95.70°	Depositor
Resolution (Å)	53.80 – 2.06 53.80 – 2.06	Depositor EDS
% Data completeness (in resolution range)	97.7 (53.80-2.06) 96.9 (53.80-2.06)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.49 (at 2.07Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, $R_{free}$	0.173 , 0.209 0.181 , 0.214	Depositor DCC
$R_{free}$ test set	2874 reflections (5.35%)	DCC
Wilson B-factor (Å <sup>2</sup> )	25.1	Xtriage
Anisotropy	0.682	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 53.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	6806	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.24% of the height of the origin peak. No significant pseudotranslation is detected.*

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

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, SO4, EDO

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	0.92	1/2331 (0.0%)	0.97	7/3165 (0.2%)
1	D	0.95	0/2345	0.97	6/3183 (0.2%)
2	B	0.68	0/860	0.78	0/1162
2	E	0.63	0/860	0.77	0/1162
3	C	0.99	0/78	1.17	0/104
3	F	1.06	0/78	0.86	0/104
All	All	0.87	1/6552 (0.0%)	0.92	13/8880 (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	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	45	GLU	CD-OE1	6.63	1.32	1.25

All (13) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	111	ARG	NE-CZ-NH2	8.76	124.68	120.30
1	A	157	ARG	NE-CZ-NH1	7.80	124.20	120.30
1	D	239	ARG	NE-CZ-NH1	-5.91	117.34	120.30
1	D	54	GLN	CB-CA-C	-5.76	98.87	110.40
1	A	157	ARG	NE-CZ-NH2	-5.68	117.46	120.30
1	A	252	GLY	N-CA-C	-5.68	98.90	113.10

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	169	ARG	NE-CZ-NH1	5.54	123.07	120.30
1	D	239	ARG	NE-CZ-NH2	5.49	123.05	120.30
1	A	270	LEU	CA-CB-CG	5.30	127.50	115.30
1	A	219	ARG	NE-CZ-NH1	5.21	122.91	120.30
1	D	129	ASP	CB-CG-OD1	5.12	122.91	118.30
1	A	37	ASP	CB-CG-OD1	5.07	122.86	118.30
1	A	75	ARG	NE-CZ-NH1	5.02	122.81	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	194	ILE	Peptide
1	A	251	SER	Peptide

## 5.2 Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2270	0	2125	30	0
1	D	2284	0	2137	27	0
2	B	837	0	803	5	0
2	E	837	0	803	4	0
3	C	76	0	86	0	0
3	F	76	0	86	0	0
4	A	12	0	15	9	0
4	D	6	0	8	6	0
5	A	20	0	30	1	0
5	B	4	0	6	0	0
5	D	36	0	53	2	0
5	E	8	0	12	0	0
5	F	4	0	6	0	0
6	A	10	0	0	0	0
7	A	138	0	0	2	0
7	B	28	0	0	2	0
7	C	7	0	0	0	0
7	D	125	0	0	4	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	E	22	0	0	0	0
7	F	6	0	0	0	0
All	All	6806	0	6170	67	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 5.

All (67) 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)
7:A:621:HOH:O	4:D:301:GOL:H11	1.79	0.81
4:A:301:GOL:H11	7:D:414:HOH:O	1.87	0.74
1:D:71:ALA:HB3	4:D:301:GOL:H32	1.70	0.72
1:A:219:ARG:HG2	1:A:257:TYR:CE1	2.26	0.71
1:A:125:ALA:HB2	5:A:307:EDO:H22	1.74	0.68
1:A:84:TYR:CE1	4:A:302:GOL:H12	2.29	0.68
1:D:77:SER:OG	5:D:305:EDO:H12	1.91	0.68
1:D:71:ALA:HB3	4:D:301:GOL:C3	2.24	0.67
1:A:84:TYR:HE1	4:A:302:GOL:H12	1.60	0.67
2:E:17:ASN:OD1	2:E:17:ASN:N	2.30	0.64
1:A:232:GLU:OE1	2:B:6:LYS:NZ	2.31	0.64
1:D:249:VAL:HG12	1:D:257:TYR:CE2	2.34	0.63
1:A:270:LEU:C	1:A:270:LEU:HD23	2.19	0.62
7:A:621:HOH:O	4:D:301:GOL:C1	2.42	0.62
1:A:0:MET:O	1:A:2:SER:N	2.32	0.62
1:A:270:LEU:HD23	1:A:271:THR:N	2.17	0.60
2:B:3:ARG:HD3	7:B:205:HOH:O	2.00	0.60
1:D:73:THR:HG22	5:D:305:EDO:H11	1.82	0.59
1:D:219:ARG:N	1:D:222:GLU:O	2.35	0.58
1:A:19:GLU:HG2	1:A:75:ARG:HD3	1.86	0.56
1:A:270:LEU:HD21	1:A:272:LEU:HG	1.87	0.56
1:D:219:ARG:HB3	1:D:222:GLU:O	2.06	0.55
4:A:301:GOL:C1	7:D:414:HOH:O	2.50	0.55
1:D:111:ARG:NH2	1:D:112:GLY:O	2.41	0.54
1:D:48:ARG:O	1:D:239:ARG:NH1	2.34	0.54
1:A:249:VAL:O	1:A:251:SER:N	2.41	0.54
1:D:218:GLN:O	1:D:257:TYR:HA	2.09	0.53
1:A:44:ARG:HB2	4:A:301:GOL:H31	1.89	0.53
1:A:117:ALA:HB2	2:B:60:TRP:CE2	2.44	0.53
1:D:220:ASP:N	1:D:256:ARG:O	2.37	0.53
2:E:48:LYS:O	2:E:48:LYS:HG3	2.09	0.52

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:181:ARG:HH11	1:A:181:ARG:HB2	1.75	0.51
1:A:225:THR:OG1	1:A:225:THR:O	2.22	0.50
1:A:219:ARG:HA	1:A:219:ARG:NE	2.27	0.49
1:D:3:HIS:HD2	1:D:29:ASP:OD2	1.96	0.49
1:D:254:GLU:HG3	1:D:274:TRP:CD1	2.48	0.48
2:B:17:ASN:HD21	2:B:97:ARG:HH22	1.59	0.48
1:D:61:ASP:O	1:D:65[A]:GLN:HG2	2.13	0.48
1:A:194:ILE:O	1:A:196:ASP:N	2.46	0.48
1:D:218:GLN:OE1	1:D:260:HIS:CD2	2.67	0.48
2:E:17:ASN:HD22	2:E:74:GLU:N	2.12	0.48
1:A:270:LEU:HD21	1:A:272:LEU:CD1	2.44	0.47
1:A:51:TRP:O	1:A:54:GLN:HG2	2.15	0.47
1:A:44:ARG:HD3	4:A:301:GOL:H31	1.97	0.47
1:A:175:GLY:O	1:A:176:LYS:C	2.54	0.46
1:D:117:ALA:HB2	2:E:60:TRP:CE2	2.50	0.45
1:D:194:ILE:O	1:D:195:SER:HB3	2.16	0.45
1:D:174:ASN:ND2	7:D:403:HOH:O	2.39	0.45
1:D:210:PRO:O	1:D:263:HIS:HE1	1.99	0.45
1:D:266:LEU:HD13	1:D:270:LEU:HG	1.99	0.45
1:A:60:TRP:CB	4:A:301:GOL:H2	2.47	0.44
1:D:82:ARG:HD2	7:D:448:HOH:O	2.16	0.44
1:D:184:PRO:HB3	1:D:265:GLY:O	2.18	0.44
1:A:219:ARG:HH11	1:A:219:ARG:HG3	1.82	0.43
1:D:82:ARG:NH1	1:D:88:SER:O	2.44	0.43
1:D:260:HIS:CE1	1:D:271:THR:HG22	2.54	0.42
1:D:219:ARG:O	1:D:220:ASP:C	2.58	0.42
1:A:44:ARG:HD3	4:A:301:GOL:C3	2.50	0.42
1:A:60:TRP:HB3	4:A:301:GOL:H2	2.00	0.42
2:B:3:ARG:HA	7:B:221:HOH:O	2.20	0.42
1:A:270:LEU:C	1:A:270:LEU:CD2	2.86	0.42
1:A:173:GLU:OE1	1:A:176:LYS:NZ	2.40	0.41
1:A:68:LYS:O	1:A:72[B]:GLN:HG3	2.20	0.41
1:D:71:ALA:HB3	4:D:301:GOL:H31	2.01	0.41
1:A:266:LEU:HA	1:A:267:PRO:HD2	1.89	0.41
1:D:68:LYS:O	4:D:301:GOL:H32	2.20	0.41
1:A:219:ARG:N	1:A:222:GLU:O	2.52	0.41

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	276/277 (100%)	261 (95%)	10 (4%)	5 (2%)	10	2
1	D	278/277 (100%)	263 (95%)	12 (4%)	3 (1%)	17	6
2	B	98/100 (98%)	96 (98%)	2 (2%)	0	100	100
2	E	98/100 (98%)	96 (98%)	2 (2%)	0	100	100
3	C	7/9 (78%)	7 (100%)	0	0	100	100
3	F	7/9 (78%)	7 (100%)	0	0	100	100
All	All	764/772 (99%)	730 (96%)	26 (3%)	8 (1%)	17	7

All (8) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	176	LYS
1	D	176	LYS
1	D	251	SER
1	A	1	GLY
1	A	195	SER
1	A	196	ASP
1	A	250	PRO
1	D	250	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.

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	237/236 (100%)	218 (92%)	19 (8%)	14	6
1	D	239/236 (101%)	228 (95%)	11 (5%)	31	22
2	B	95/95 (100%)	90 (95%)	5 (5%)	26	17
2	E	95/95 (100%)	85 (90%)	10 (10%)	8	2
3	C	9/9 (100%)	9 (100%)	0	100	100
3	F	9/9 (100%)	9 (100%)	0	100	100
All	All	684/680 (101%)	639 (93%)	45 (7%)	19	10

All (45) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	0	MET
1	A	2	SER
1	A	53	GLU
1	A	186	LYS
1	A	195	SER
1	A	197	HIS
1	A	198	GLU
1	A	200	THR
1	A	206	LEU
1	A	219	ARG
1	A	220	ASP
1	A	223	ASP
1	A	225	THR
1	A	226	GLN
1	A	229	GLU
1	A	251	SER
1	A	253	GLU
1	A	270	LEU
1	A	271	THR
2	B	1	ILE
2	B	19	LYS
2	B	37	VAL
2	B	70	PHE
2	B	71	THR
1	D	0	MET
1	D	86	ASN
1	D	166	GLU

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	D	219	ARG
1	D	224	GLN
1	D	226	GLN
1	D	232	GLU
1	D	253	GLU
1	D	255	GLN
1	D	268	LYS
1	D	271	THR
2	E	16	GLU
2	E	17	ASN
2	E	19	LYS
2	E	20	SER
2	E	37	VAL
2	E	58	LYS
2	E	70	PHE
2	E	71	THR
2	E	75	LYS
2	E	88	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (12) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	65	GLN
1	A	115	GLN
1	A	260	HIS
2	B	17	ASN
2	B	24	ASN
1	D	3	HIS
1	D	86	ASN
1	D	87	GLN
1	D	174	ASN
1	D	255	GLN
1	D	263	HIS
2	E	24	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 ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

23 ligands are modelled in this entry.

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

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
4	GOL	A	301	-	5,5,5	2.69	3 (60%)	5,5,5	2.32	4 (80%)
4	GOL	A	302	-	5,5,5	1.91	1 (20%)	5,5,5	3.58	3 (60%)
5	EDO	A	303	-	3,3,3	0.28	0	2,2,2	0.52	0
5	EDO	A	304	-	3,3,3	0.38	0	2,2,2	0.75	0
5	EDO	A	305	-	3,3,3	0.87	0	2,2,2	0.64	0
5	EDO	A	306	-	3,3,3	0.79	0	2,2,2	0.50	0
5	EDO	A	307	-	3,3,3	0.96	0	2,2,2	1.55	0
6	SO4	A	308	-	4,4,4	0.50	0	6,6,6	0.30	0
6	SO4	A	309	-	4,4,4	0.65	0	6,6,6	0.28	0
5	EDO	B	101	-	3,3,3	0.75	0	2,2,2	0.17	0
4	GOL	D	301	-	5,5,5	0.94	0	5,5,5	2.37	1 (20%)
5	EDO	D	302	-	3,3,3	0.43	0	2,2,2	1.29	0
5	EDO	D	303	-	3,3,3	0.58	0	2,2,2	0.45	0
5	EDO	D	304	-	3,3,3	0.51	0	2,2,2	0.25	0
5	EDO	D	305	-	3,3,3	0.80	0	2,2,2	1.48	0
5	EDO	D	306	-	3,3,3	0.40	0	2,2,2	0.92	0
5	EDO	D	307	-	3,3,3	0.88	0	2,2,2	0.73	0
5	EDO	D	308	-	3,3,3	0.53	0	2,2,2	0.48	0
5	EDO	D	309	-	3,3,3	0.43	0	2,2,2	0.58	0
5	EDO	D	310	-	3,3,3	0.47	0	2,2,2	0.11	0
5	EDO	E	101	-	3,3,3	0.27	0	2,2,2	1.00	0
5	EDO	E	102	-	3,3,3	0.53	0	2,2,2	0.18	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
5	EDO	F	101	-	3,3,3	0.72	0	2,2,2	0.30	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	GOL	A	301	-	-	0/4/4/4	0/0/0/0
4	GOL	A	302	-	-	0/4/4/4	0/0/0/0
5	EDO	A	303	-	-	0/1/1/1	0/0/0/0
5	EDO	A	304	-	-	0/1/1/1	0/0/0/0
5	EDO	A	305	-	-	0/1/1/1	0/0/0/0
5	EDO	A	306	-	-	0/1/1/1	0/0/0/0
5	EDO	A	307	-	-	0/1/1/1	0/0/0/0
6	SO4	A	308	-	-	0/0/0/0	0/0/0/0
6	SO4	A	309	-	-	0/0/0/0	0/0/0/0
5	EDO	B	101	-	-	0/1/1/1	0/0/0/0
4	GOL	D	301	-	-	0/4/4/4	0/0/0/0
5	EDO	D	302	-	-	0/1/1/1	0/0/0/0
5	EDO	D	303	-	-	0/1/1/1	0/0/0/0
5	EDO	D	304	-	-	0/1/1/1	0/0/0/0
5	EDO	D	305	-	-	0/1/1/1	0/0/0/0
5	EDO	D	306	-	-	0/1/1/1	0/0/0/0
5	EDO	D	307	-	-	0/1/1/1	0/0/0/0
5	EDO	D	308	-	-	0/1/1/1	0/0/0/0
5	EDO	D	309	-	-	0/1/1/1	0/0/0/0
5	EDO	D	310	-	-	0/1/1/1	0/0/0/0
5	EDO	E	101	-	-	0/1/1/1	0/0/0/0
5	EDO	E	102	-	-	0/1/1/1	0/0/0/0
5	EDO	F	101	-	-	0/1/1/1	0/0/0/0

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	302	GOL	O2-C2	-4.06	1.31	1.43
4	A	301	GOL	O1-C1	2.26	1.51	1.42
4	A	301	GOL	O3-C3	2.79	1.54	1.42
4	A	301	GOL	O2-C2	4.51	1.56	1.43

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
4	A	302	GOL	O2-C2-C1	-5.02	85.14	108.84
4	A	301	GOL	O2-C2-C1	2.05	118.54	108.84
4	A	301	GOL	O2-C2-C3	2.32	119.80	108.84
4	A	301	GOL	O1-C1-C2	2.53	122.80	110.07
4	A	301	GOL	O3-C3-C2	3.04	125.41	110.07
4	A	302	GOL	O1-C1-C2	3.37	127.03	110.07
4	D	301	GOL	O3-C3-C2	4.51	132.78	110.07
4	A	302	GOL	C3-C2-C1	4.97	131.26	111.52

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5 monomers are involved in 18 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	301	GOL	7	0
4	A	302	GOL	2	0
5	A	307	EDO	1	0
4	D	301	GOL	6	0
5	D	305	EDO	2	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	277/277 (100%)	0.53	38 (13%) 3 3	16, 35, 120, 140	0
1	D	277/277 (100%)	0.33	25 (9%) 10 10	15, 31, 105, 150	0
2	B	100/100 (100%)	0.13	3 (3%) 51 54	25, 57, 87, 113	0
2	E	100/100 (100%)	0.09	6 (6%) 23 23	23, 51, 85, 97	0
3	C	9/9 (100%)	-0.16	0 100 100	17, 19, 31, 37	0
3	F	9/9 (100%)	-0.23	0 100 100	16, 18, 31, 44	0
All	All	772/772 (100%)	0.33	72 (9%) 9 9	15, 40, 109, 150	0

All (72) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	221	GLY	10.4
1	D	221	GLY	8.7
1	A	195	SER	8.7
1	D	194	ILE	8.3
1	A	199	ALA	6.6
1	D	220	ASP	6.1
1	A	257	TYR	5.9
1	A	0	MET	5.5
1	A	194	ILE	5.4
1	D	0	MET	5.3
2	B	0	MET	4.9
1	A	222	GLU	4.9
1	D	196	ASP	4.8
1	A	191	HIS	4.5
1	A	250	PRO	4.4
2	E	1	ILE	4.4
1	A	274	TRP	4.3
1	A	248	VAL	4.2
1	A	218	GLN	4.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	196	ASP	4.1
1	D	257	TYR	4.0
1	D	275	GLU	3.9
1	A	246	ALA	3.9
1	D	249	VAL	3.9
1	A	226	GLN	3.9
1	A	197	HIS	3.7
1	A	249	VAL	3.7
1	A	220	ASP	3.7
1	A	201	LEU	3.6
1	D	195	SER	3.6
1	D	190	THR	3.6
1	D	274	TRP	3.5
1	D	191	HIS	3.5
1	A	192	HIS	3.3
1	A	198	GLU	3.3
1	A	223	ASP	3.2
1	D	187	THR	3.2
1	A	256	ARG	3.1
1	A	253	GLU	3.1
1	A	227	ASP	3.0
1	A	260	HIS	3.0
1	D	192	HIS	3.0
1	D	253	GLU	2.9
2	E	0	MET	2.9
1	D	270	LEU	2.9
1	A	276	PRO	2.8
1	D	223	ASP	2.8
2	E	16	GLU	2.7
1	D	230	LEU	2.7
2	B	18	GLY	2.7
1	A	252	GLY	2.6
2	B	1	ILE	2.6
1	D	224	GLN	2.6
1	A	224	GLN	2.6
1	A	190	THR	2.5
1	D	260	HIS	2.5
1	A	254	GLU	2.4
1	D	225	THR	2.4
2	E	75	LYS	2.3
1	A	272	LEU	2.3
2	E	99	MET	2.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	D	188	HIS	2.2
1	A	251	SER	2.2
1	A	225	THR	2.1
1	A	193	PRO	2.1
1	D	197	HIS	2.1
1	A	258	THR	2.1
2	E	48	LYS	2.1
1	A	247	VAL	2.0
1	A	217	TRP	2.0
1	D	219	ARG	2.0
1	D	273	ARG	2.0

## 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

## 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. 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	SO4	A	308	5/5	0.84	0.33	17.54	89,107,110,114	0
5	EDO	D	306	4/4	0.84	0.18	10.67	52,56,60,64	0
4	GOL	A	302	6/6	0.94	0.25	9.44	8,9,9,10	6
5	EDO	E	102	4/4	0.87	0.20	9.30	41,44,46,48	0
5	EDO	F	101	4/4	0.84	0.23	8.47	45,51,53,54	0
5	EDO	D	307	4/4	0.63	0.24	8.06	42,48,50,53	0
5	EDO	E	101	4/4	0.94	0.19	7.16	41,43,45,47	0
6	SO4	A	309	5/5	0.88	0.20	6.26	80,83,89,94	0
4	GOL	A	301	6/6	0.85	0.22	5.76	26,36,37,38	0
5	EDO	A	305	4/4	0.94	0.15	3.33	28,42,47,48	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
4	GOL	D	301	6/6	0.94	0.18	2.55	20,32,36,49	0
5	EDO	D	310	4/4	0.95	0.18	2.44	39,44,51,53	0
5	EDO	A	307	4/4	0.85	0.14	1.68	26,30,35,37	0
5	EDO	D	304	4/4	0.96	0.13	1.41	30,34,37,39	0
5	EDO	A	303	4/4	0.98	0.12	1.07	24,24,26,27	0
5	EDO	B	101	4/4	0.92	0.13	0.90	38,40,40,41	0
5	EDO	D	309	4/4	0.96	0.15	0.90	48,50,51,55	0
5	EDO	D	305	4/4	0.95	0.13	0.88	23,25,28,34	0
5	EDO	D	302	4/4	0.94	0.15	0.49	38,40,43,45	0
5	EDO	A	304	4/4	0.91	0.14	0.38	35,36,36,43	0
5	EDO	A	306	4/4	0.96	0.12	-0.12	21,27,27,27	0
5	EDO	D	308	4/4	0.96	0.11	-0.14	39,45,45,45	0
5	EDO	D	303	4/4	0.96	0.10	-0.36	26,32,32,34	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.