



# Full wwPDB X-ray Structure Validation Report ⓘ

Nov 19, 2023 – 06:05 PM JST

PDB ID : 6LWL  
Title : Crystal structure of human NEIL1(R242) bound to duplex DNA containing 2'-fluoro-2'-deoxy-5,6-dihydrouridine  
Authors : Liu, M.H.; Zhang, J.; Zhu, C.X.; Zhang, X.X.; Gao, Y.Q.; Yi, C.Q.  
Deposited on : 2020-02-07  
Resolution : 2.55 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.36  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.36

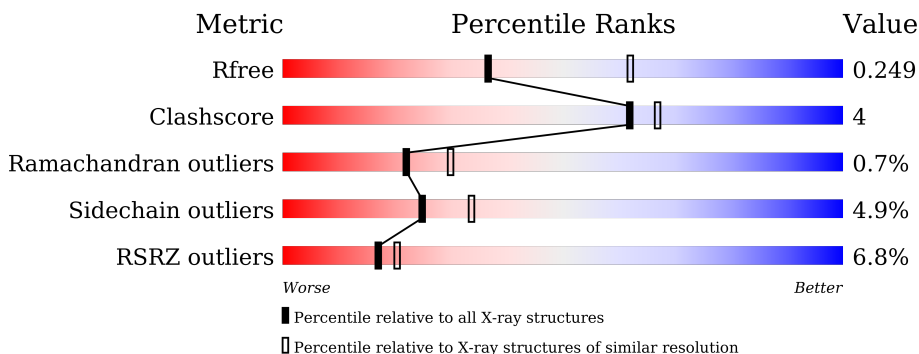
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.55 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))
$R_{free}$	130704	1284 (2.56-2.52)
Clashscore	141614	1332 (2.56-2.52)
Ramachandran outliers	138981	1315 (2.56-2.52)
Sidechain outliers	138945	1315 (2.56-2.52)
RSRZ outliers	127900	1272 (2.56-2.52)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. 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	295	 8% 81% 8% 8% 8%
1	D	295	 4% 79% 9% 10% 10%
1	G	295	 12% 71% 7% 21% 21%
2	B	13	 8% 85% 8% 8% 8%
2	E	13	 8% 77% 8% 15% 15%
2	H	13	 23% 92% 8% 8%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	C	13	 92% 8%
3	F	13	 54% 46%
3	I	13	 23% 46% 54%

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	D	301	-	-	-	X

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 7719 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 Endonuclease 8-like 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	270	Total 2142	C 1365	N 390	O 377	S 10	0	0	0
1	D	266	Total 2107	C 1346	N 385	O 366	S 10	0	0	0
1	G	233	Total 1667	C 1033	N 318	O 308	S 8	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	242	ARG	LYS	variant	UNP Q96FI4
D	242	ARG	LYS	variant	UNP Q96FI4
G	242	ARG	LYS	variant	UNP Q96FI4

- Molecule 2 is a DNA chain called DNA (5'-D(\*CP\*GP\*TP\*CP\*CP\*AP\*(FDU)P\*GP\*TP\*CP\*TP\*AP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	F	N	O				P
2	B	13	Total 258	C 124	F 1	N 43	O 78	P 12	0	0	0
2	E	13	Total 258	C 124	F 1	N 43	O 78	P 12	0	0	0
2	H	13	Total 258	C 124	F 1	N 43	O 78	P 12	0	0	0

- Molecule 3 is a DNA chain called DNA (5'-D(\*TP\*AP\*GP\*AP\*CP\*CP\*TP\*GP\*GP\*AP\*CP\*GP\*G)-3').

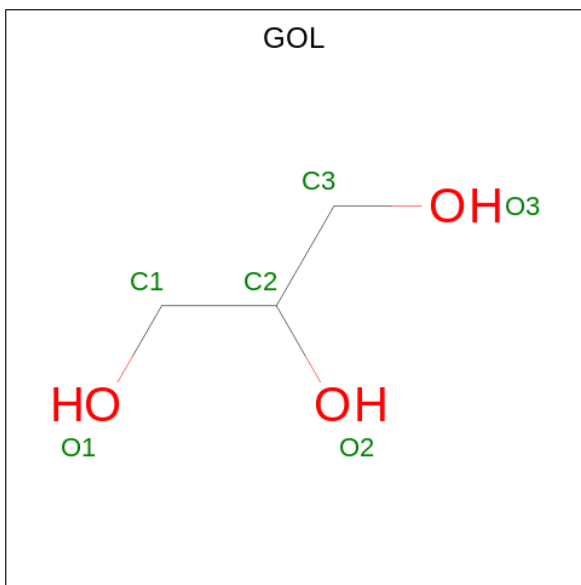
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	13	Total 267	C 127	N 53	O 75	P 12	0	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	13	Total	C	N	O	P	0	0	0
			267	127	53	75	12			
3	I	13	Total	C	N	O	P	0	0	0
			267	127	53	75	12			

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



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

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	114	Total	O	0	0
			114	114		
5	B	14	Total	O	0	0
			14	14		
5	C	7	Total	O	0	0
			7	7		
5	D	69	Total	O	0	0
			69	69		
5	E	9	Total	O	0	0
			9	9		
5	F	5	Total	O	0	0
			5	5		

Continued on next page...

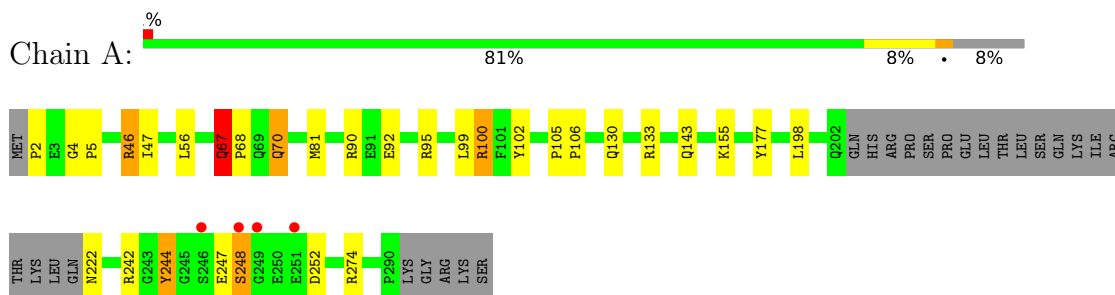
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>	<b>ZeroOcc</b>	<b>AltConf</b>
5	G	1	Total O 1 1	0	0
5	H	1	Total O 1 1	0	0
5	I	2	Total O 2 2	0	0

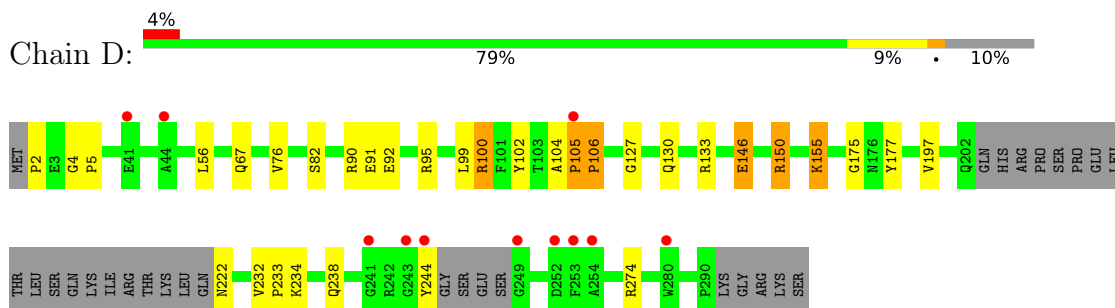
### 3 Residue-property plots i

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

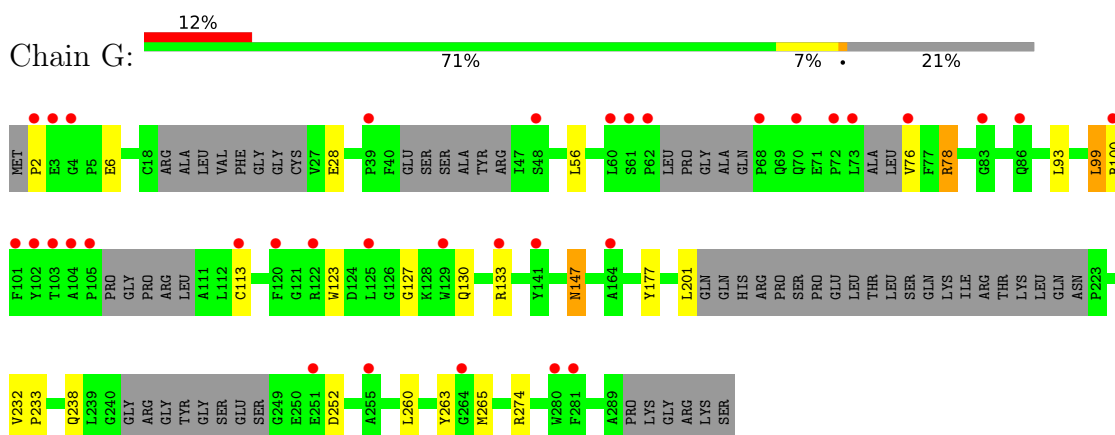
- Molecule 1: Endonuclease 8-like 1



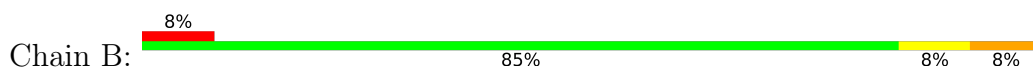
- Molecule 1: Endonuclease 8-like 1



- Molecule 1: Endonuclease 8-like 1

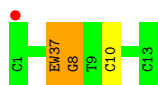
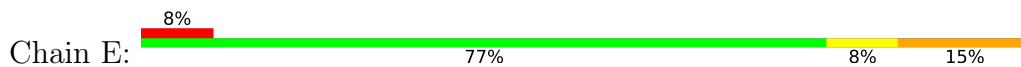


- Molecule 2: DNA (5'-D(\*CP\*GP\*TP\*CP\*CP\*AP\*(FDU)P\*GP\*TP\*CP\*TP\*AP\*C)-3')

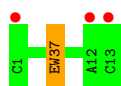
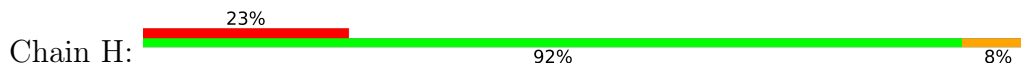




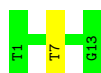
- Molecule 2: DNA (5'-D(\*CP\*GP\*TP\*CP\*CP\*AP\*(FDU)P\*GP\*TP\*CP\*TP\*AP\*C)-3')



- Molecule 2: DNA (5'-D(\*CP\*GP\*TP\*CP\*CP\*AP\*(FDU)P\*GP\*TP\*CP\*TP\*AP\*C)-3')



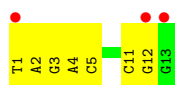
- Molecule 3: DNA (5'-D(\*TP\*AP\*GP\*AP\*CP\*CP\*TP\*GP\*GP\*AP\*CP\*GP\*G)-3')



- Molecule 3: DNA (5'-D(\*TP\*AP\*GP\*AP\*CP\*CP\*TP\*GP\*GP\*AP\*CP\*GP\*G)-3')



- Molecule 3: DNA (5'-D(\*TP\*AP\*GP\*AP\*CP\*CP\*TP\*GP\*GP\*AP\*CP\*GP\*G)-3')





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	73.77Å 108.65Å 170.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.08 – 2.55 36.05 – 2.54	Depositor EDS
% Data completeness (in resolution range)	99.3 (36.08-2.55) 99.4 (36.05-2.54)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.37 (at 2.54Å)	Xtrriage
Refinement program	REFMAC 5.8.0267	Depositor
R, $R_{free}$	0.209 , 0.253 0.211 , 0.249	Depositor DCC
$R_{free}$ test set	2310 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	61.5	Xtrriage
Anisotropy	0.430	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 56.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	7719	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	87.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.03% 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 i

### 5.1 Standard geometry i

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, EW3

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	0.93	2/2201 (0.1%)	0.91	3/2977 (0.1%)
1	D	0.95	4/2165 (0.2%)	0.89	1/2927 (0.0%)
1	G	0.70	0/1701	0.80	0/2277
2	B	0.51	0/264	0.76	0/402
2	E	0.59	0/264	1.50	3/402 (0.7%)
2	H	0.39	0/264	0.73	0/402
3	C	0.66	0/300	0.99	1/462 (0.2%)
3	F	0.61	0/300	0.89	0/462
3	I	0.44	0/300	0.88	1/462 (0.2%)
All	All	0.82	6/7759 (0.1%)	0.90	9/10773 (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	3
1	D	0	2
All	All	0	5

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	92	GLU	CD-OE1	19.61	1.47	1.25
1	D	92	GLU	CD-OE2	19.11	1.46	1.25
1	D	92	GLU	CD-OE1	18.41	1.45	1.25
1	A	92	GLU	CD-OE2	16.82	1.44	1.25
1	D	91	GLU	CD-OE2	-5.37	1.19	1.25
1	D	146	GLU	CD-OE2	-5.24	1.19	1.25

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	8	DG	O5'-P-OP1	-22.87	83.26	110.70
3	C	7	DT	O5'-P-OP2	-7.37	99.07	105.70
1	D	92	GLU	OE1-CD-OE2	6.99	131.69	123.30
2	E	8	DG	O5'-P-OP2	6.92	119.01	110.70
2	E	10	DC	O5'-P-OP2	-5.86	100.43	105.70
1	A	92	GLU	OE1-CD-OE2	5.63	130.06	123.30
3	I	3	DG	P-O3'-C3'	5.30	126.06	119.70
1	A	67	GLN	CB-CA-C	5.29	120.97	110.40
1	A	70	GLN	CB-CA-C	-5.09	100.21	110.40

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	105	PRO	Peptide
1	A	67	GLN	Peptide,Mainchain
1	D	105	PRO	Peptide
1	D	67	GLN	Peptide

## 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	2142	0	2110	22	0
1	D	2107	0	2078	17	0
1	G	1667	0	1478	17	0
2	B	258	0	137	7	0
2	E	258	0	137	2	0
2	H	258	0	137	1	0
3	C	267	0	147	0	0
3	F	267	0	147	3	0
3	I	267	0	147	3	0
4	D	6	0	8	1	0
5	A	114	0	0	3	0
5	B	14	0	0	0	0
5	C	7	0	0	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	D	69	0	0	1	0
5	E	9	0	0	0	0
5	F	5	0	0	0	0
5	G	1	0	0	0	0
5	H	1	0	0	0	0
5	I	2	0	0	0	0
All	All	7719	0	6526	62	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 4.

All (62) 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:G:76:VAL:HG23	1:G:127:GLY:HA2	1.60	0.83
1:A:81:MET:HE2	2:B:8:DG:C5'	2.16	0.74
1:A:242:ARG:HG3	1:A:247:GLU:O	1.87	0.74
1:G:2:PRO:N	2:H:7:EW3:F2	2.11	0.74
1:A:143:GLN:OE1	5:A:301:HOH:O	2.07	0.71
3:F:11:DC:H2'	3:F:12:DG:C8	2.30	0.67
1:A:81:MET:HE3	2:B:8:DG:C8	2.30	0.66
1:G:130:GLN:HE21	1:G:133:ARG:HD2	1.61	0.64
1:A:81:MET:HE2	2:B:8:DG:H5'	1.79	0.64
1:G:78:ARG:NH2	1:G:130:GLN:OE1	2.26	0.62
3:I:11:DC:H2'	3:I:12:DG:C8	2.36	0.61
1:G:93:LEU:HD13	1:G:100:ARG:HH11	1.64	0.61
1:D:2:PRO:N	2:E:7:EW3:F2	2.24	0.60
1:A:155:LYS:HE2	1:A:155:LYS:H	1.67	0.59
1:A:81:MET:CE	2:B:8:DG:C8	2.88	0.56
1:A:47:ILE:O	1:D:150:ARG:NH2	2.34	0.56
1:A:67:GLN:HA	1:A:67:GLN:NE2	2.24	0.53
1:A:81:MET:HE2	2:B:8:DG:O5'	2.09	0.52
1:D:104:ALA:HB1	1:D:105:PRO:HD2	1.92	0.52
1:A:81:MET:HE3	2:B:8:DG:N9	2.26	0.51
1:D:130:GLN:NE2	1:D:133:ARG:HD2	2.27	0.50
1:G:130:GLN:NE2	1:G:133:ARG:HD2	2.27	0.49
1:A:90:ARG:HH12	1:A:102:TYR:HB3	1.76	0.49
1:A:248:SER:HB2	1:A:252:ASP:OD2	2.13	0.49
3:I:4:DA:H4'	3:I:5:DC:OP1	2.12	0.48
3:F:4:DA:H4'	3:F:5:DC:OP1	2.14	0.48
1:G:99:LEU:HG	1:G:123:TRP:CZ2	2.48	0.47

Continued on next page...

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:D:301:GOL:H11	5:D:443:HOH:O	2.14	0.47
1:A:2:PRO:N	2:B:7:EW3:F2	2.38	0.47
1:D:234:LYS:O	1:D:238:GLN:HG3	2.15	0.46
1:D:146:GLU:OE2	1:D:146:GLU:HA	2.16	0.46
1:D:130:GLN:HE21	1:D:133:ARG:HD2	1.80	0.46
1:D:155:LYS:CE	1:D:155:LYS:H	2.28	0.46
1:G:93:LEU:HD13	1:G:100:ARG:NH1	2.31	0.45
1:G:28:GLU:OE1	1:G:100:ARG:HD2	2.16	0.45
1:G:2:PRO:HB3	1:G:6:GLU:HB2	1.99	0.45
1:G:147:ASN:O	1:G:147:ASN:ND2	2.49	0.45
3:I:1:DT:C2'	3:I:2:DA:C8	3.00	0.45
1:A:130:GLN:NE2	1:A:133:ARG:HD2	2.32	0.45
1:G:76:VAL:CG2	1:G:127:GLY:HA2	2.40	0.44
1:D:105:PRO:HA	1:D:106:PRO:HA	1.73	0.43
1:D:99:LEU:C	1:D:100:ARG:HG2	2.38	0.43
1:G:2:PRO:HB3	1:G:6:GLU:CB	2.49	0.43
1:A:46:ARG:HG3	5:A:345:HOH:O	2.18	0.43
1:A:56:LEU:C	1:A:56:LEU:HD23	2.39	0.43
1:A:90:ARG:NH1	1:A:102:TYR:HB3	2.33	0.43
3:F:1:DT:C2'	3:F:2:DA:C8	3.02	0.43
1:D:155:LYS:H	1:D:155:LYS:HE3	1.83	0.43
1:A:90:ARG:NH1	5:A:306:HOH:O	2.45	0.43
1:A:155:LYS:H	1:A:155:LYS:CE	2.32	0.42
1:D:175:GLY:HA3	2:E:8:DG:OP1	2.19	0.42
1:A:99:LEU:C	1:A:100:ARG:HG2	2.40	0.42
1:D:232:VAL:HB	1:D:233:PRO:HD3	2.02	0.42
1:D:90:ARG:HH12	1:D:102:TYR:HB3	1.84	0.42
1:G:130:GLN:HB2	1:G:133:ARG:CD	2.49	0.42
1:A:4:GLY:N	1:A:5:PRO:CD	2.83	0.41
1:D:56:LEU:HD23	1:D:56:LEU:C	2.41	0.41
1:D:76:VAL:HG23	1:D:127:GLY:HA2	2.03	0.41
1:G:232:VAL:HB	1:G:233:PRO:HD3	2.03	0.41
1:D:4:GLY:N	1:D:5:PRO:CD	2.84	0.41
1:G:56:LEU:C	1:G:56:LEU:HD23	2.41	0.40
1:G:260:LEU:HD22	1:G:263:TYR:HB3	2.04	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	266/295 (90%)	253 (95%)	9 (3%)	4 (2%)	10	14
1	D	260/295 (88%)	248 (95%)	11 (4%)	1 (0%)	34	46
1	G	217/295 (74%)	207 (95%)	10 (5%)	0	100	100
All	All	743/885 (84%)	708 (95%)	30 (4%)	5 (1%)	22	30

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	106	PRO
1	A	68	PRO
1	A	244	TYR
1	A	248	SER
1	A	106	PRO

### 5.3.2 Protein sidechains [i](#)

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	224/248 (90%)	215 (96%)	9 (4%)	31	43
1	D	219/248 (88%)	209 (95%)	10 (5%)	27	36
1	G	153/248 (62%)	143 (94%)	10 (6%)	17	23
All	All	596/744 (80%)	567 (95%)	29 (5%)	25	34

All (29) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	46	ARG
1	A	70	GLN
1	A	95	ARG
1	A	100	ARG
1	A	177	TYR
1	A	198	LEU
1	A	222	ASN
1	A	244	TYR
1	A	274	ARG
1	D	82	SER
1	D	95	ARG
1	D	100	ARG
1	D	150	ARG
1	D	155	LYS
1	D	177	TYR
1	D	197	VAL
1	D	222	ASN
1	D	244	TYR
1	D	274	ARG
1	G	78	ARG
1	G	99	LEU
1	G	113	CYS
1	G	147	ASN
1	G	177	TYR
1	G	201	LEU
1	G	238	GLN
1	G	252	ASP
1	G	265	MET
1	G	274	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (11) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	67	GLN
1	A	130	GLN
1	A	139	GLN
1	A	202	GLN
1	A	222	ASN
1	D	70	GLN
1	D	130	GLN
1	D	202	GLN
1	D	222	ASN
1	D	238	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	G	238	GLN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

3 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	EW3	B	7	2	18,21,22	1.57	3 (16%)	22,30,33	2.32	8 (36%)
2	EW3	H	7	2	18,21,22	1.08	2 (11%)	22,30,33	2.41	8 (36%)
2	EW3	E	7	2	18,21,22	1.06	1 (5%)	22,30,33	2.29	8 (36%)

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	EW3	B	7	2	-	1/7/38/39	0/2/2/2
2	EW3	H	7	2	-	1/7/38/39	0/2/2/2
2	EW3	E	7	2	-	1/7/38/39	0/2/2/2

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	7	EW3	F2-C2'	-4.54	1.30	1.40
2	B	7	EW3	O4'-C1'	2.86	1.48	1.42
2	E	7	EW3	F2-C2'	-2.48	1.34	1.40
2	H	7	EW3	F2-C2'	-2.32	1.35	1.40

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	H	7	EW3	C2-N3	-2.10	1.34	1.38
2	B	7	EW3	C2-N3	-2.07	1.34	1.38

All (24) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	H	7	EW3	O4'-C1'-N1	6.38	117.99	109.30
2	E	7	EW3	O4'-C1'-N1	5.55	116.86	109.30
2	B	7	EW3	O4'-C1'-N1	5.50	116.79	109.30
2	H	7	EW3	C5-C4-N3	4.45	121.65	116.65
2	B	7	EW3	C5-C4-N3	4.41	121.60	116.65
2	E	7	EW3	C5-C4-N3	4.24	121.41	116.65
2	H	7	EW3	C2'-C1'-N1	-4.03	107.52	114.08
2	E	7	EW3	C2'-C1'-N1	-3.80	107.89	114.08
2	H	7	EW3	C4-N3-C2	-3.63	122.78	125.79
2	B	7	EW3	C4-N3-C2	-3.59	122.81	125.79
2	B	7	EW3	C2'-C1'-N1	-3.44	108.48	114.08
2	H	7	EW3	F2-C2'-C3'	3.33	116.19	109.22
2	E	7	EW3	C4-N3-C2	-3.20	123.14	125.79
2	E	7	EW3	N3-C2-N1	3.11	119.94	116.65
2	B	7	EW3	N3-C2-N1	3.06	119.89	116.65
2	E	7	EW3	F2-C2'-C3'	3.05	115.61	109.22
2	H	7	EW3	N3-C2-N1	3.04	119.87	116.65
2	B	7	EW3	F2-C2'-C3'	3.01	115.53	109.22
2	E	7	EW3	F2-C2'-C1'	-2.66	103.54	109.08
2	B	7	EW3	F2-C2'-C1'	-2.65	103.56	109.08
2	H	7	EW3	O4-C4-C5	-2.22	117.43	122.17
2	E	7	EW3	O4-C4-C5	-2.16	117.56	122.17
2	B	7	EW3	C3'-C2'-C1'	-2.13	100.55	103.13
2	H	7	EW3	O2-C2-N1	-2.11	120.46	123.11

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	H	7	EW3	O4'-C4'-C5'-O5'
2	B	7	EW3	O4'-C4'-C5'-O5'
2	E	7	EW3	O4'-C4'-C5'-O5'

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	7	EW3	1	0
2	H	7	EW3	1	0
2	E	7	EW3	1	0

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

1 ligand is 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	D	301	-	5,5,5	0.20	0	5,5,5	0.45	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	D	301	-	-	0/4/4/4	-

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	D	301	GOL	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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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	270/295 (91%)	0.10	4 (1%) 73 79	38, 60, 99, 114	0
1	D	266/295 (90%)	0.44	11 (4%) 37 44	41, 76, 123, 149	0
1	G	233/295 (78%)	1.03	34 (14%) 2 3	78, 118, 143, 155	0
2	B	12/13 (92%)	0.31	1 (8%) 11 13	61, 93, 142, 167	0
2	E	12/13 (92%)	0.17	1 (8%) 11 13	53, 85, 111, 130	0
2	H	12/13 (92%)	0.93	3 (25%) 0 0	100, 117, 136, 152	0
3	C	13/13 (100%)	-0.15	0 100 100	74, 93, 130, 143	0
3	F	13/13 (100%)	0.31	0 100 100	56, 82, 125, 138	0
3	I	13/13 (100%)	0.83	3 (23%) 0 0	95, 124, 143, 145	0
All	All	844/963 (87%)	0.49	57 (6%) 17 20	38, 85, 135, 167	0

All (57) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	105	PRO	5.8
1	G	281	PHE	5.0
1	G	60	LEU	4.4
1	G	62	PRO	4.3
1	G	141	TYR	4.3
1	G	280	TRP	4.3
1	G	113	CYS	4.2
1	D	253	PHE	4.1
1	G	100	ARG	3.9
1	G	101	PHE	3.9
1	A	246	SER	3.9
1	G	72	PRO	3.8
1	G	61	SER	3.8
2	H	13	DC	3.6
1	G	264	GLY	3.6

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
2	B	13	DC	3.5
1	G	251	GLU	3.4
1	G	122	ARG	3.3
1	G	76	VAL	3.3
2	H	1	DC	3.2
1	A	248	SER	3.2
1	G	102	TYR	3.2
2	E	1	DC	3.1
1	G	73	LEU	2.9
3	I	1	DT	2.9
1	G	104	ALA	2.9
1	A	249	GLY	2.9
1	D	244	TYR	2.7
1	G	70	GLN	2.7
1	G	48	SER	2.6
1	D	243	GLY	2.6
3	I	12	DG	2.6
1	G	125	LEU	2.5
1	G	86	GLN	2.5
1	D	252	ASP	2.4
1	G	4	GLY	2.4
1	D	241	GLY	2.4
1	G	164	ALA	2.4
1	G	68	PRO	2.4
1	D	105	PRO	2.4
1	G	3	GLU	2.3
1	A	251	GLU	2.3
1	G	2	PRO	2.2
1	D	254	ALA	2.2
1	G	129	TRP	2.2
1	G	83	GLY	2.2
1	G	133	ARG	2.2
1	D	249	GLY	2.2
1	D	280	TRP	2.2
1	G	39	PRO	2.2
1	G	120	PHE	2.1
1	D	41	GLU	2.1
3	I	13	DG	2.1
1	G	103	THR	2.1
2	H	12	DA	2.1
1	D	44	ALA	2.0
1	G	255	ALA	2.0

## 6.2 Non-standard residues in protein, DNA, RNA chains [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. 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	B-factors(Å <sup>2</sup> )	Q<0.9
2	EW3	H	7	20/21	0.94	0.20	99,111,129,132	0
2	EW3	E	7	20/21	0.96	0.19	66,74,114,122	0
2	EW3	B	7	20/21	0.97	0.18	58,67,87,96	0

## 6.3 Carbohydrates [i](#)

There are no monosaccharides 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. 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	B-factors(Å <sup>2</sup> )	Q<0.9
4	GOL	D	301	6/6	0.58	0.40	77,86,93,95	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.