



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 1, 2021 – 02:10 PM JST

PDB ID : 6LWA
Title : Crystal structure of human NEIL1(P2G, E3Q, K242) bound to duplex DNA containing 5-hydroxyuracil (5-OHU)
Authors : Liu, M.H.; Zhang, J.; Zhu, C.X.; Zhang, X.X.; Gao, Y.Q.; Yi, C.Q.
Deposited on : 2020-02-07
Resolution : 2.76 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

A user guide is available at

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

with specific help available everywhere you see the ⓘ symbol.

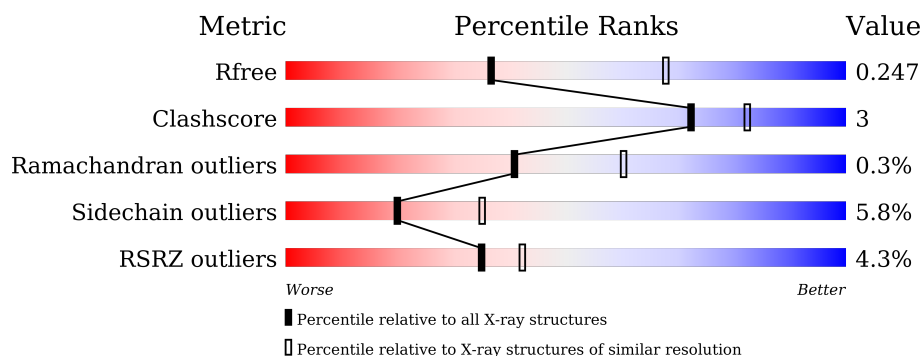
The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.19
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)	:	2.19

i

X-RAY DIFFRACTION

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	1235 (2.78-2.74)
Clashscore	141614	1277 (2.78-2.74)
Ramachandran outliers	138981	1257 (2.78-2.74)
Sidechain outliers	138945	1257 (2.78-2.74)
RSRZ outliers	127900	1207 (2.78-2.74)

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 ≥ 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	<div> <div></div> <div>78%</div> <div>12%</div> <div>10%</div> </div>
1	D	295	<div> <div></div> <div>78%</div> <div>12%</div> <div>9%</div> </div>
1	G	295	<div> <div>11%</div> <div>70%</div> <div>6%</div> <div>23%</div> </div>
2	B	13	<div> <div>77%</div> <div>23%</div> </div>
2	E	13	<div> <div>62%</div> <div>31%</div> <div>8%</div> </div>
2	H	13	<div> <div>69%</div> <div>8%</div> <div>23%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	C	13	 85%15%
3	F	13	 100%
3	I	13	 92%8%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 7504 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
1	A	266	Total	C	N	O	S	0	0	0
			2112	1349	385	368	10			
1	D	267	Total	C	N	O	S	0	0	0
			2111	1348	386	367	10			
1	G	227	Total	C	N	O	S	0	0	0
			1568	969	301	290	8			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2	GLY	PRO	engineered mutation	UNP Q96FI4
A	3	GLN	GLU	engineered mutation	UNP Q96FI4
D	2	GLY	PRO	engineered mutation	UNP Q96FI4
D	3	GLN	GLU	engineered mutation	UNP Q96FI4
G	2	GLY	PRO	engineered mutation	UNP Q96FI4
G	3	GLN	GLU	engineered mutation	UNP Q96FI4

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	13	Total	C	N	O	P	0	0	0
			258	124	43	79	12			
2	E	13	Total	C	N	O	P	0	0	0
			258	124	43	79	12			
2	H	13	Total	C	N	O	P	0	0	0
			258	124	43	79	12			

- 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
3	C	13	Total 267	C 127	N 53	O 75	P 12	0	0	0
3	F	13	Total 267	C 127	N 53	O 75	P 12	0	0	0
3	I	13	Total 267	C 127	N 53	O 75	P 12	0	0	0

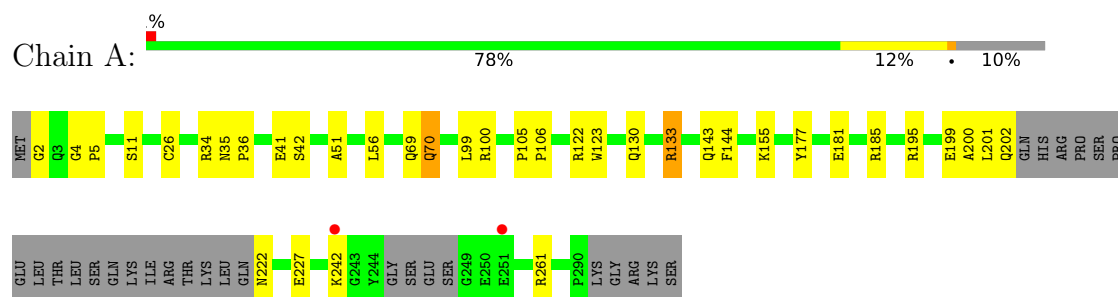
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	71	Total 71	O 71	0	0
4	B	14	Total 14	O 14	0	0
4	C	1	Total 1	O 1	0	0
4	D	43	Total 43	O 43	0	0
4	E	3	Total 3	O 3	0	0
4	F	3	Total 3	O 3	0	0
4	G	1	Total 1	O 1	0	0
4	H	1	Total 1	O 1	0	0
4	I	1	Total 1	O 1	0	0

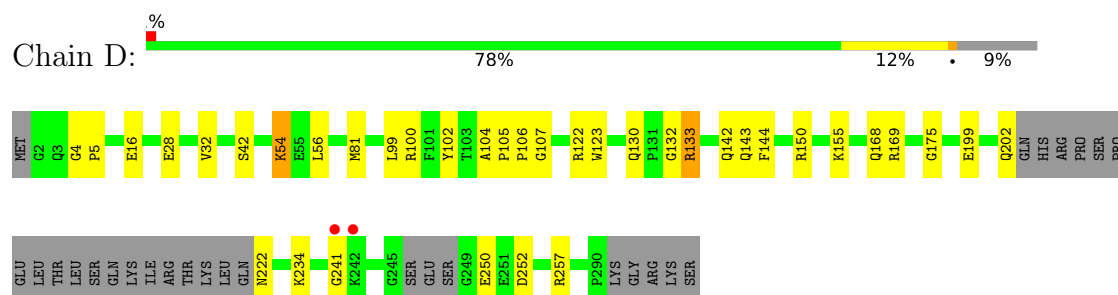
3 Residue-property plots [i](#)

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.

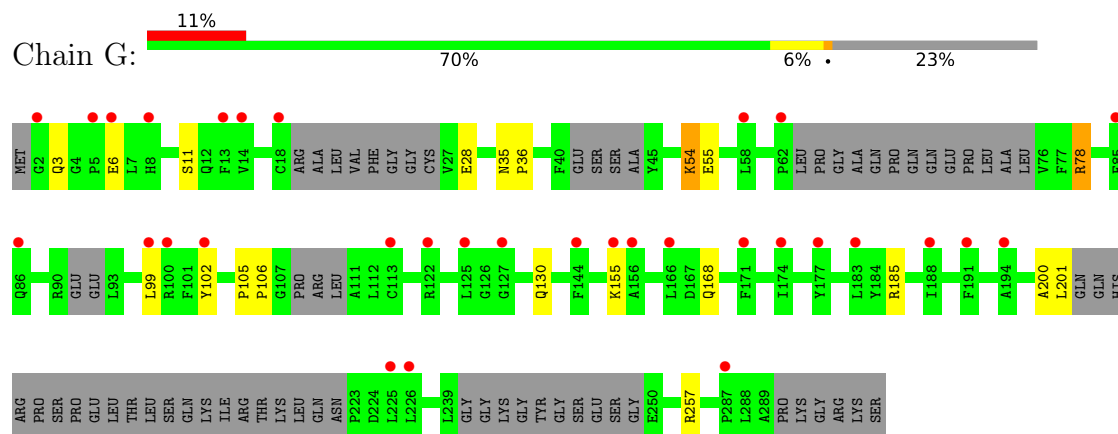
- Molecule 1: Endonuclease 8-like 1



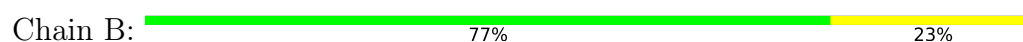
- Molecule 1: Endonuclease 8-like 1

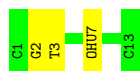


- Molecule 1: Endonuclease 8-like 1

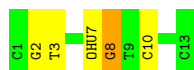


- Molecule 2: DNA (5'-D(*CP*GP*TP*CP*CP*AP*(OHU)P*GP*TP*CP*TP*AP*C)-3')

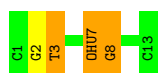




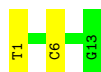
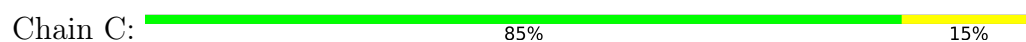
- Molecule 2: DNA (5'-D(*CP*GP*TP*CP*CP*AP*(OHU)P*GP*TP*CP*TP*AP*C)-3')



- Molecule 2: DNA (5'-D(*CP*GP*TP*CP*CP*AP*(OHU)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')



There are no outlier residues recorded for this chain.

- 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, α , β , γ	73.56Å 109.77Å 169.17Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.58 – 2.76 49.53 – 2.76	Depositor EDS
% Data completeness (in resolution range)	99.5 (49.58-2.76) 99.6 (49.53-2.76)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.86 (at 2.77Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.207 , 0.246 0.209 , 0.247	Depositor DCC
R_{free} test set	1765 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å ²)	65.3	Xtriage
Anisotropy	0.552	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 53.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	7504	wwPDB-VP
Average B, all atoms (Å ²)	88.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.38% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: OHU

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	0.78	1/2169 (0.0%)	0.91	1/2932 (0.0%)
1	D	0.77	0/2168	0.92	3/2930 (0.1%)
1	G	0.72	0/1599	0.83	1/2136 (0.0%)
2	B	0.53	0/264	0.87	0/402
2	E	0.69	1/264 (0.4%)	1.07	1/402 (0.2%)
2	H	0.47	0/264	0.99	3/402 (0.7%)
3	C	0.66	1/300 (0.3%)	0.83	0/462
3	F	0.61	0/300	0.81	0/462
3	I	0.53	0/300	0.80	0/462
All	All	0.72	3/7628 (0.0%)	0.89	9/10590 (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
1	D	0	1
All	All	0	2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	227	GLU	CD-OE1	8.53	1.35	1.25
2	E	8	DG	O3'-P	6.87	1.69	1.61
3	C	6	DC	O3'-P	-5.04	1.55	1.61

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	8	DG	O5'-P-OP1	-12.66	94.31	105.70
1	D	150	ARG	NE-CZ-NH1	-7.08	116.76	120.30
2	H	3	DT	C4'-C3'-O3'	6.84	126.80	109.70
1	D	150	ARG	NE-CZ-NH2	6.41	123.50	120.30
1	G	105	PRO	C-N-CA	5.65	145.73	122.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	105	PRO	Peptide
1	D	107	GLY	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	2112	0	2087	18	0
1	D	2111	0	2085	17	0
1	G	1568	0	1324	9	0
2	B	258	0	146	1	0
2	E	258	0	146	3	0
2	H	258	0	146	3	0
3	C	267	0	147	2	0
3	F	267	0	147	0	0
3	I	267	0	147	2	0
4	A	71	0	0	4	0
4	B	14	0	0	0	0
4	C	1	0	0	0	0
4	D	43	0	0	0	0
4	E	3	0	0	1	0
4	F	3	0	0	0	0
4	G	1	0	0	0	0
4	H	1	0	0	0	0
4	I	1	0	0	0	0
All	All	7504	0	6375	46	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 3.

The worst 5 of 46 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:A:143:GLN:NE2	4:A:301:HOH:O	1.72	1.06
1:A:51:ALA:H	1:D:142:GLN:HE22	1.22	0.87
1:A:41:GLU:O	1:A:70:GLN:NE2	2.10	0.83
1:A:2:GLY:O	4:A:302:HOH:O	2.02	0.78
2:H:2:DG:H2''	2:H:3:DT:O5'	1.94	0.66

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	260/295 (88%)	250 (96%)	9 (4%)	1 (0%)	34	53
1	D	261/295 (88%)	251 (96%)	10 (4%)	0	100	100
1	G	211/295 (72%)	199 (94%)	11 (5%)	1 (0%)	29	47
All	All	732/885 (83%)	700 (96%)	30 (4%)	2 (0%)	41	60

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	G	106	PRO
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	220/247 (89%)	207 (94%)	13 (6%)	19	34
1	D	219/247 (89%)	206 (94%)	13 (6%)	19	34
1	G	132/247 (53%)	125 (95%)	7 (5%)	22	38
All	All	571/741 (77%)	538 (94%)	33 (6%)	20	35

5 of 33 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	78	ARG
1	G	99	LEU
1	G	257	ARG
1	A	242	LYS
1	A	222	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	202	GLN
1	D	143	GLN
1	D	130	GLN
1	A	202	GLN
1	D	142	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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	OHU	E	7	2	15,21,22	1.13	1 (6%)	15,30,33	1.91	2 (13%)
2	OHU	B	7	2	15,21,22	1.19	1 (6%)	15,30,33	2.08	4 (26%)
2	OHU	H	7	2	15,21,22	1.08	1 (6%)	15,30,33	1.98	2 (13%)

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	OHU	E	7	2	-	2/4/21/22	0/2/2/2
2	OHU	B	7	2	-	3/4/21/22	0/2/2/2
2	OHU	H	7	2	-	2/4/21/22	0/2/2/2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	7	OHU	C4-C5	3.27	1.48	1.40
2	E	7	OHU	C4-C5	2.95	1.47	1.40
2	H	7	OHU	C4-C5	2.95	1.47	1.40

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	H	7	OHU	C4-N3-C2	6.62	120.73	115.14
2	E	7	OHU	C4-N3-C2	6.27	120.44	115.14
2	B	7	OHU	C4-N3-C2	5.54	119.82	115.14
2	B	7	OHU	C2'-C1'-N1	3.62	122.61	114.27
2	B	7	OHU	C4'-O4'-C1'	2.24	114.86	109.45

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	7	OHU	O4'-C1'-N1-C6
2	E	7	OHU	O4'-C4'-C5'-O5'
2	H	7	OHU	O4'-C4'-C5'-O5'
2	E	7	OHU	C3'-C4'-C5'-O5'
2	H	7	OHU	C3'-C4'-C5'-O5'

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	H	7	OHU	1	0

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

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, 95th 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 ²)	Q<0.9
1	A	266/295 (90%)	0.03	2 (0%) 86 90	41, 64, 106, 151	0
1	D	267/295 (90%)	0.12	2 (0%) 87 91	40, 71, 120, 166	0
1	G	227/295 (76%)	0.79	32 (14%) 2 3	79, 117, 155, 165	0
2	B	12/13 (92%)	-0.37	0 100 100	64, 100, 144, 146	0
2	E	12/13 (92%)	-0.23	0 100 100	74, 99, 112, 115	0
2	H	12/13 (92%)	-0.41	0 100 100	104, 120, 131, 137	0
3	C	13/13 (100%)	-0.47	0 100 100	78, 97, 133, 156	0
3	F	13/13 (100%)	-0.28	0 100 100	60, 105, 152, 167	0
3	I	13/13 (100%)	-0.27	0 100 100	78, 116, 142, 143	0
All	All	835/963 (86%)	0.23	36 (4%) 35 42	40, 84, 144, 167	0

The worst 5 of 36 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	100	ARG	5.9
1	G	125	LEU	5.3
1	G	58	LEU	4.5
1	G	287	PRO	3.9
1	G	194	ALA	3.5

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, 95th 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(\AA^2)	Q<0.9
2	OHU	E	7	20/21	0.93	0.24	75,94,121,121	0
2	OHU	H	7	20/21	0.94	0.22	113,125,144,145	0
2	OHU	B	7	20/21	0.96	0.16	60,70,114,126	0

6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.