Full wwPDB X-ray Structure Validation Report (i)

Nov 14, 2023 – 03:58 PM JST

PDB ID	:	5ZKJ
Title	:	Human EXOG-H140A in complex with RNA/DNA hybrid duplex
Authors	:	Wu, C.C.; Lin, J.L.J.; Yuan, H.S.
Deposited on	:	2018-03-24
Resolution	:	2.80 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (i)) were used in the production of this report:

|		
MolProbity	:	4.02b-467
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
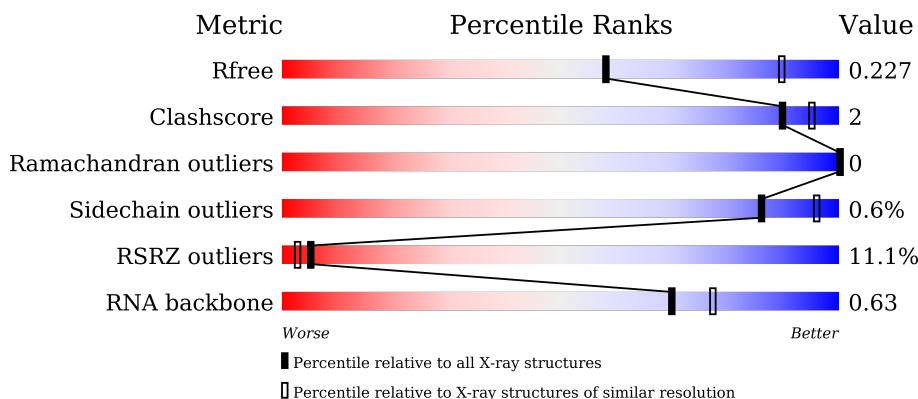
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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	Percentile Ranks	Value
Rfree		0.227
Clashscore		2
Ramachandran outliers		0
Sidechain outliers		0.6%
RSRZ outliers		11.1%
RNA backbone		0.63

Worse   *Better*

▌Percentile relative to all X-ray structures

▯Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)
RNA backbone	3102	1227 (3.10-2.50)

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 <=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	348	11% / 81% / 15%
1	B	348	7% / 85% / 13%
2	C	12	8% / 83% / 17%
2	E	12	42% / 67% / 33%

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	D	12	8% 67% 33%
3	F	12	50% 50%

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 11162 atoms, of which 5284 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 Nuclease EXOG, mitochondrial.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	297	Total 4732	C 1518	H 2345	N 408	O 450	S 11	0	1	0
1	B	303	Total 4843	C 1552	H 2405	N 414	O 461	S 11	0	2	0

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	21	MET	-	expression tag	UNP Q9Y2C4
A	22	GLY	-	expression tag	UNP Q9Y2C4
A	23	SER	-	expression tag	UNP Q9Y2C4
A	24	SER	-	expression tag	UNP Q9Y2C4
A	25	HIS	-	expression tag	UNP Q9Y2C4
A	26	HIS	-	expression tag	UNP Q9Y2C4
A	27	HIS	-	expression tag	UNP Q9Y2C4
A	28	HIS	-	expression tag	UNP Q9Y2C4
A	29	HIS	-	expression tag	UNP Q9Y2C4
A	30	HIS	-	expression tag	UNP Q9Y2C4
A	31	SER	-	expression tag	UNP Q9Y2C4
A	32	SER	-	expression tag	UNP Q9Y2C4
A	33	GLY	-	expression tag	UNP Q9Y2C4
A	34	LEU	-	expression tag	UNP Q9Y2C4
A	35	VAL	-	expression tag	UNP Q9Y2C4
A	36	PRO	-	expression tag	UNP Q9Y2C4
A	37	ARG	-	expression tag	UNP Q9Y2C4
A	38	GLY	-	expression tag	UNP Q9Y2C4
A	39	SER	-	expression tag	UNP Q9Y2C4
A	40	HIS	-	expression tag	UNP Q9Y2C4
A	41	MET	-	expression tag	UNP Q9Y2C4
A	140	ALA	HIS	engineered mutation	UNP Q9Y2C4
B	21	MET	-	expression tag	UNP Q9Y2C4
B	22	GLY	-	expression tag	UNP Q9Y2C4
B	23	SER	-	expression tag	UNP Q9Y2C4

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	24	SER	-	expression tag	UNP Q9Y2C4
B	25	HIS	-	expression tag	UNP Q9Y2C4
B	26	HIS	-	expression tag	UNP Q9Y2C4
B	27	HIS	-	expression tag	UNP Q9Y2C4
B	28	HIS	-	expression tag	UNP Q9Y2C4
B	29	HIS	-	expression tag	UNP Q9Y2C4
B	30	HIS	-	expression tag	UNP Q9Y2C4
B	31	SER	-	expression tag	UNP Q9Y2C4
B	32	SER	-	expression tag	UNP Q9Y2C4
B	33	GLY	-	expression tag	UNP Q9Y2C4
B	34	LEU	-	expression tag	UNP Q9Y2C4
B	35	VAL	-	expression tag	UNP Q9Y2C4
B	36	PRO	-	expression tag	UNP Q9Y2C4
B	37	ARG	-	expression tag	UNP Q9Y2C4
B	38	GLY	-	expression tag	UNP Q9Y2C4
B	39	SER	-	expression tag	UNP Q9Y2C4
B	40	HIS	-	expression tag	UNP Q9Y2C4
B	41	MET	-	expression tag	UNP Q9Y2C4
B	140	ALA	HIS	engineered mutation	UNP Q9Y2C4

- Molecule 2 is a RNA chain called RNA (5'-R(P*CP*GP*GP*GP*AP*UP*GP*UP*CP*AP *CP*G)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	C	12	Total 391	C 115	H 131	N 48	O 85	P 12	0	0	0
2	E	12	Total 391	C 115	H 131	N 48	O 85	P 12	0	0	0

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	D	12	Total 376	C 115	H 136	N 44	O 70	P 11	0	0	0
3	F	12	Total 376	C 115	H 136	N 44	O 70	P 11	0	0	0

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total 1	Mg 1	0	0
4	C	1	Total 1	Mg 1	0	0

- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total 1	Cl 1	0	0
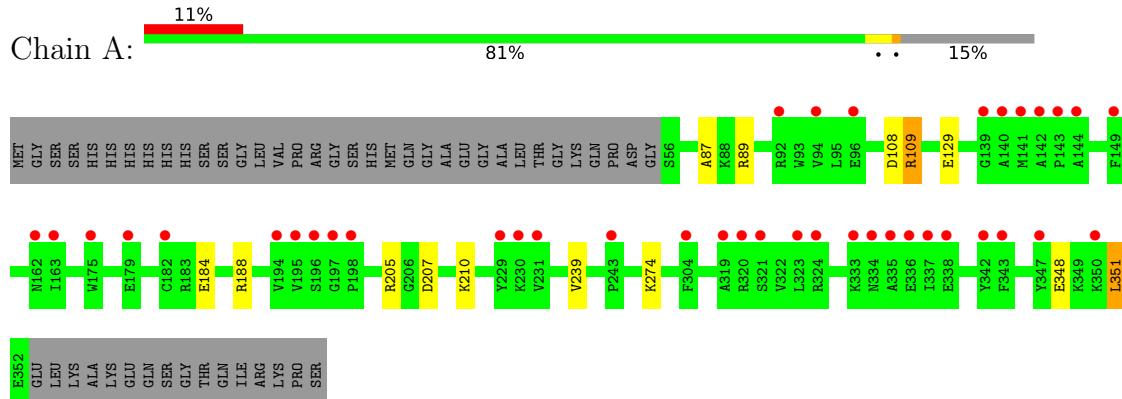
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	24	Total 24	O 24	0	0
6	B	24	Total 24	O 24	0	0
6	C	2	Total 2	O 2	0	0
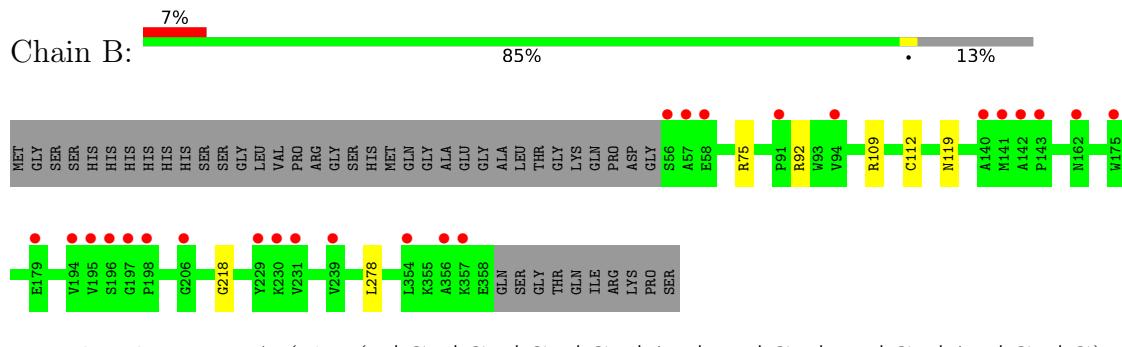
3 Residue-property plots (i)

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.
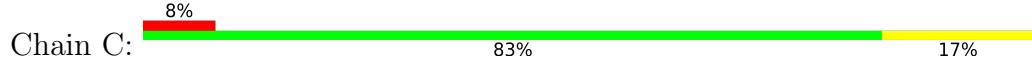
• Molecule 1: Nuclease EXOG, mitochondrial
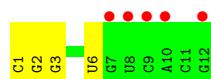
Chain A:

• Molecule 1: Nuclease EXOG, mitochondrial

Chain B:

• Molecule 2: RNA (5'-R(P*CP*GP*GP*GP*AP*UP*GP*UP*CP*AP*CP*G)-3')

Chain C:

• Molecule 2: RNA (5'-R(P*CP*GP*GP*GP*AP*UP*GP*UP*CP*AP*CP*G)-3')

Chain E:

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

Chain D:

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

Chain F:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α, β, γ	167.58Å 167.58Å 105.44Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	28.43 – 2.80 28.43 – 2.80	Depositor EDS
% Data completeness (in resolution range)	95.4 (28.43-2.80) 95.4 (28.43-2.80)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.89 (at 2.80Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, R_{free}	0.181 , 0.227 0.181 , 0.227	Depositor DCC
R_{free} test set	1340 reflections (5.16%)	wwPDB-VP
Wilson B-factor (Å^2)	75.3	Xtriage
Anisotropy	0.323	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 56.1	EDS
L-test for twinning[2]	$< \lvert L \rvert > = 0.49, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.000 for -1/3*h+1/3*k+4/3*l,-k,2/3*h+1/3*k+1/3*l 0.011 for -2/3*h-1/3*k-4/3*l,-1/3*h-2/3*k+4/3*l,-1/3*h+1/3*k+1/3*l 0.004 for -h,1/3*h-1/3*k-4/3*l,-1/3*h-2/3*k+1/3*l 0.016 for -1/3*h-2/3*k+4/3*l,-2/3*h-1/3*k-4/3*l,1/3*h-1/3*k-1/3*l 0.009 for -h,2/3*h+1/3*k+4/3*l,1/3*h+2/3*k-1/3*l 0.002 for 1/3*h+2/3*k-4/3*l,-k,-2/3*h-1/3*k-1/3*l 0.017 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	11162	wwPDB-VP
Average B, all atoms (Å^2)	93.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.00% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< \lvert L \rvert >$, $< L^2 >$ 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: MG, CL

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	0.24	0/2446	0.40	0/3309
1	B	0.25	0/2500	0.41	0/3382
2	C	0.66	1/290 (0.3%)	0.72	0/449
2	E	0.65	1/290 (0.3%)	0.73	0/449
3	D	0.44	0/268	0.86	0/411
3	F	0.50	0/268	0.87	0/411
All	All	0.33	2/6062 (0.0%)	0.51	0/8411

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	1	C	OP3-P	-10.59	1.48	1.61
2	C	1	C	OP3-P	-10.55	1.48	1.61

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	2387	2345	2347	10	0
1	B	2438	2405	2405	4	0
2	C	260	131	131	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	260	131	131	2	0
3	D	240	136	136	2	0
3	F	240	136	136	5	0
4	A	1	0	0	0	0
4	C	1	0	0	0	0
5	B	1	0	0	0	0
6	A	24	0	0	0	0
6	B	24	0	0	0	0
6	C	2	0	0	0	0
All	All	5878	5284	5286	20	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 2.

All (20) 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:B:75:ARG:NH1	1:B:119:ASN:O	2.32	0.61
1:A:109:ARG:NH1	2:E:3:G:OP2	2.36	0.58
1:A:210:LYS:NZ	1:B:218:GLY:O	2.40	0.51
3:F:2:DG:H2'	3:F:3:DT:H5'	1.91	0.51
1:A:205:ARG:HG3	1:A:207:ASP:OD2	2.14	0.47
1:A:184:GLU:OE2	1:A:188:ARG:NH2	2.44	0.46
3:F:11:DC:H2'	3:F:12:DG:C8	2.51	0.45
1:A:87:ALA:O	1:A:89:ARG:NH1	2.51	0.44
3:F:4:DG:O6	3:F:5:DA:N6	2.51	0.43
3:D:4:DG:O6	3:D:5:DA:N6	2.51	0.43
1:A:108:ASP:OD1	1:A:109:ARG:N	2.52	0.42
1:A:109:ARG:NH1	2:E:2:G:H3'	2.34	0.42
3:F:2:DG:C2'	3:F:3:DT:C5'	2.98	0.41
1:B:92:ARG:HD3	1:B:278:LEU:HD21	2.02	0.41
1:B:109:ARG:O	1:B:112:CYS:N	2.52	0.41
3:D:2:DG:H2'	3:D:3:DT:H5'	2.03	0.40
3:F:3:DT:H5'	3:F:3:DT:H6	1.86	0.40
1:A:348:GLU:HA	1:A:351:LEU:HD12	2.03	0.40
1:A:129:GLU:OE1	1:A:129:GLU:N	2.55	0.40
1:A:239:VAL:HG22	1:A:239:VAL:O	2.20	0.40

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	296/348 (85%)	286 (97%)	10 (3%)	0	100	100
1	B	303/348 (87%)	296 (98%)	7 (2%)	0	100	100
All	All	599/696 (86%)	582 (97%)	17 (3%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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

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

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	263/302 (87%)	260 (99%)	3 (1%)	73	92
1	B	269/302 (89%)	269 (100%)	0	100	100
All	All	532/604 (88%)	529 (99%)	3 (1%)	86	96

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

Mol	Chain	Res	Type
1	A	109	ARG
1	A	274	LYS
1	A	351	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	C	11/12 (91%)	1 (9%)	0
2	E	11/12 (91%)	1 (9%)	0
All	All	22/24 (91%)	2 (9%)	0

All (2) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	C	6	U
2	E	6	U

There are no RNA pucker outliers to report.

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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

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.

No monomer is involved in short contacts.

5.7 Other polymers ⓘ

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, 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	297/348 (85%)	0.80	40 (13%)	3	1	54, 79, 152, 173	0
1	B	303/348 (87%)	0.45	25 (8%)	11	6	46, 68, 109, 136	0
2	C	12/12 (100%)	0.28	1 (8%)	11	6	57, 79, 129, 140	0
2	E	12/12 (100%)	1.72	5 (41%)	0	0	86, 131, 160, 164	0
3	D	12/12 (100%)	-0.14	1 (8%)	11	6	67, 99, 131, 146	0
3	F	12/12 (100%)	0.74	0	100	100	101, 149, 155, 166	0
All	All	648/744 (87%)	0.62	72 (11%)	5	3	46, 74, 150, 173	0

All (72) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	197	GLY	5.2
1	A	320	ARG	4.8
1	A	196	SER	4.8
1	B	196	SER	4.4
1	A	140	ALA	4.4
1	A	142	ALA	4.3
1	A	337	ILE	4.0
1	A	143	PRO	4.0
1	A	229	TYR	3.9
1	A	141	MET	3.9
1	A	319	ALA	3.7
1	B	57	ALA	3.7
2	E	9	C	3.7
2	E	10	A	3.6
1	B	197	GLY	3.6
1	A	324	ARG	3.5
2	E	8	U	3.5
1	B	206	GLY	3.4
1	A	231	VAL	3.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	142	ALA	3.3
1	A	182	CYS	3.3
1	B	143	PRO	3.3
1	A	194	VAL	3.2
1	A	336	GLU	3.2
1	A	347	TYR	3.2
1	B	56	SER	3.2
1	A	323	LEU	3.1
1	B	229	TYR	3.1
1	A	243	PRO	3.1
1	A	94	VAL	3.0
1	A	335	ALA	3.0
1	A	350	LYS	3.0
1	A	338	GLU	2.9
1	B	141	MET	2.9
1	A	334	ASN	2.8
1	B	357	LYS	2.8
1	A	195	VAL	2.8
1	A	139	GLY	2.8
2	E	7	G	2.8
1	B	195	VAL	2.7
1	B	231	VAL	2.7
1	A	333	LYS	2.7
1	B	140	ALA	2.7
1	A	304	PHE	2.6
3	D	1	DC	2.6
1	B	356	ALA	2.6
1	A	163	ILE	2.5
1	B	194	VAL	2.5
2	C	12	G	2.5
1	B	239	VAL	2.5
1	A	230	LYS	2.4
1	A	321	SER	2.4
1	A	96	GLU	2.4
1	B	58	GLU	2.4
1	A	162	ASN	2.4
1	A	342	TYR	2.4
2	E	12	G	2.3
1	A	198	PRO	2.3
1	B	198	PRO	2.2
1	A	144	ALA	2.2
1	B	94	VAL	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	162	ASN	2.1
1	A	92[A]	ARG	2.1
1	B	179	GLU	2.1
1	A	343	PHE	2.1
1	A	175	TRP	2.1
1	B	354	LEU	2.1
1	A	149	PHE	2.1
1	B	91	PRO	2.0
1	B	175	TRP	2.0
1	A	179	GLU	2.0
1	B	230	LYS	2.0

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

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

6.3 Carbohydrates ⓘ

There are no monosaccharides 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. 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(Å^2)	Q<0.9
4	MG	A	401	1/1	0.90	0.45	89,89,89,89	0
5	CL	B	401	1/1	0.93	0.17	78,78,78,78	0
4	MG	C	101	1/1	0.98	0.30	58,58,58,58	0

6.5 Other polymers ⓘ

There are no such residues in this entry.