Full wwPDB X-ray Structure Validation Report (i)

Nov 14, 2023 – 03:30 PM JST

PDB ID	:	5ZKI
Title	:	Human EXOG-H140A in complex with duplex DNA
Authors	:	Wu, C.C.; Lin, J.L.J.; Yuan, H.S.
Deposited on	:	2018-03-24
Resolution	:	2.32 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:

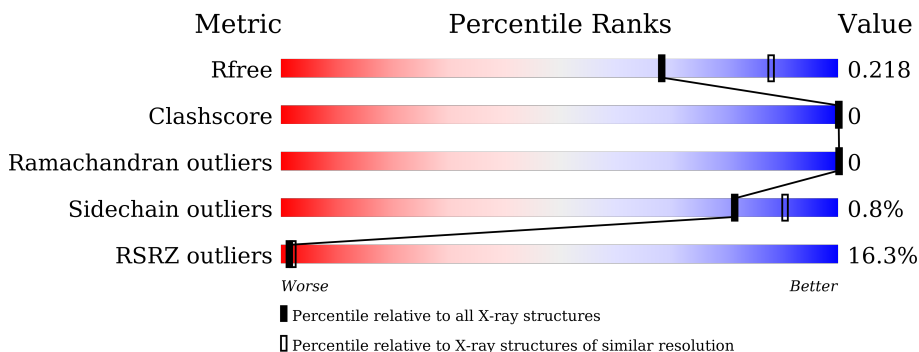|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1   Overall quality at a glance ⓘ

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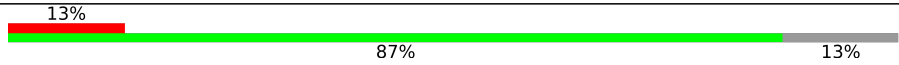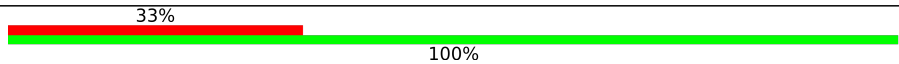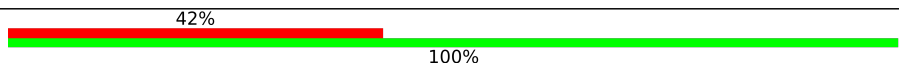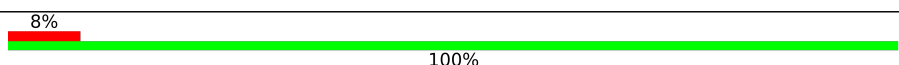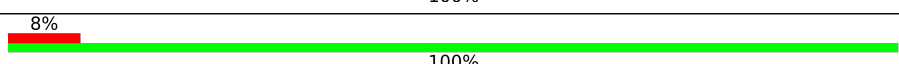*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.32 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		0.218
Clashscore		0
Ramachandran outliers		0
Sidechain outliers		0.8%
RSRZ outliers		16.3%

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	5974 (2.34-2.30)
Clashscore	141614	6604 (2.34-2.30)
Ramachandran outliers	138981	6523 (2.34-2.30)
Sidechain outliers	138945	6523 (2.34-2.30)
RSRZ outliers	127900	5855 (2.34-2.30)

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	14% / 85% · 14%
1	B	348	13% / 87% 13%
2	C	12	33% / 100%
2	D	12	42% / 100%
2	E	12	8% / 100%
2	F	12	8% / 100%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 11380 atoms, of which 5387 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	300	Total 4815	C 1542	H 2393	N 414	O 455	S 11	0	2	0
1	B	304	Total 4933	C 1579	H 2450	N 421	O 472	S 11	0	10	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

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	C	12	Total 379	C 116	H 136	N 46	O 70	P 11	0	0	0
2	D	12	Total 379	C 116	H 136	N 46	O 70	P 11	0	0	0
2	E	12	Total 379	C 116	H 136	N 46	O 70	P 11	0	0	0
2	F	12	Total 379	C 116	H 136	N 46	O 70	P 11	0	0	0

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total 1	Cl 1	0	0
3	B	1	Total 1	Cl 1	0	0

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

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

• Molecule 5 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	37	Total 37	O 37	0	0
5	C	6	Total 6	O 6	0	0
5	D	1	Total 1	O 1	0	0
5	E	8	Total 8	O 8	0	0
5	F	3	Total 3	O 3	0	0
5	B	56	Total 56	O 56	0	0

3 Residue-property plots ⓘ

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
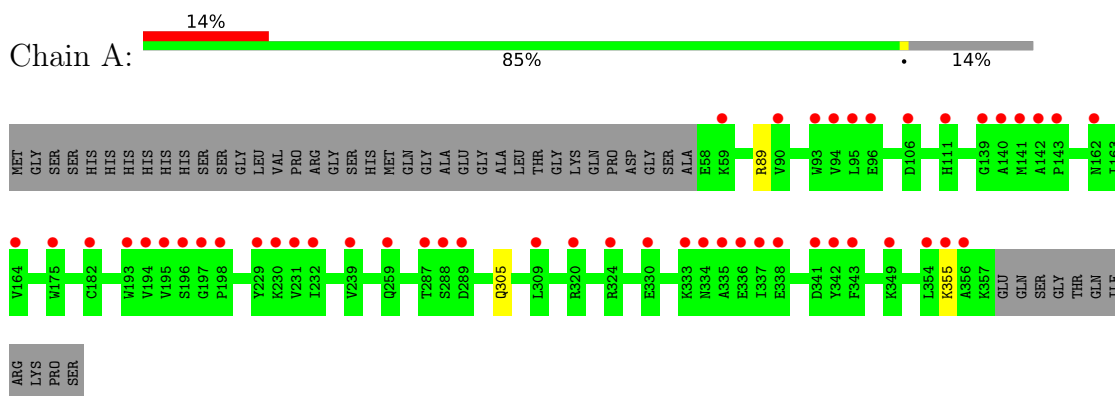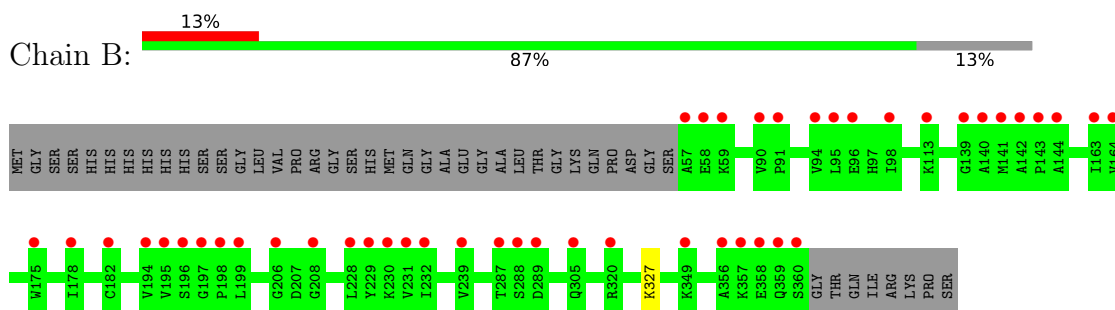
Chain A:
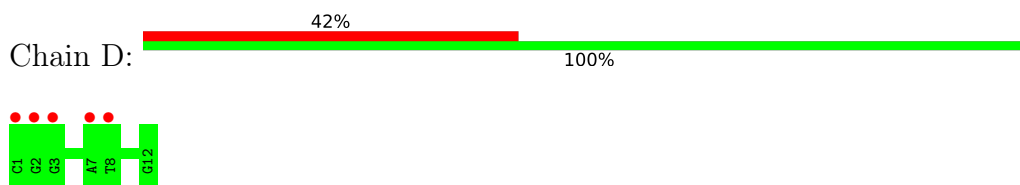

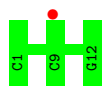• Molecule 1: Nuclease EXOG, mitochondrial

Chain B:

• Molecule 2: DNA (5'-D(*CP*GP*GP*GP*AP*TP*AP*TP*CP*CP*CP*G)-3')
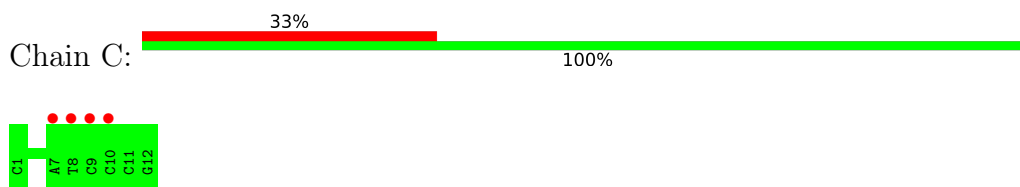
Chain C:

• Molecule 2: DNA (5'-D(*CP*GP*GP*GP*AP*TP*AP*TP*CP*CP*CP*G)-3')

Chain D:

• Molecule 2: DNA (5'-D(*CP*GP*GP*GP*AP*TP*AP*TP*CP*CP*CP*G)-3')

Chain E:

• Molecule 2: DNA (5'-D(*CP*GP*GP*GP*AP*TP*AP*TP*CP*CP*CP*G)-3')

Chain F:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	73.50Å 91.12Å 129.45Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.62 – 2.32 29.62 – 2.32	Depositor EDS
% Data completeness (in resolution range)	99.3 (29.62-2.32) 99.3 (29.62-2.32)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.86 (at 2.31Å)	Xtriage
Refinement program	PHENIX (1.12_2829: ???)	Depositor
R, R_{free}	0.180 , 0.218 0.180 , 0.218	Depositor DCC
R_{free} test set	1908 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	47.7	Xtriage
Anisotropy	0.780	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.45 , 59.8	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	11380	wwPDB-VP
Average B, all atoms (Å^2)	68.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.94% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

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	0.25	0/2484	0.41	0/3358
1	B	0.25	0/2571	0.40	0/3478
2	C	0.48	0/272	0.82	0/418
2	D	0.50	0/272	0.86	0/418
2	E	0.49	0/272	0.84	0/418
2	F	0.50	0/272	0.87	0/418
All	All	0.31	0/6143	0.52	0/8508

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	2422	2393	2393	0	0
1	B	2483	2450	2433	0	0
2	C	243	136	136	0	0
2	D	243	136	136	0	0
2	E	243	136	136	0	0
2	F	243	136	136	0	0
3	A	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	1	0	0	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
4	E	1	0	0	0	0
5	A	37	0	0	0	0
5	B	56	0	0	0	0
5	C	6	0	0	0	0
5	D	1	0	0	0	0
5	E	8	0	0	0	0
5	F	3	0	0	0	0
All	All	5993	5387	5370	0	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 0.

There are no clashes within the asymmetric unit.

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	300/348 (86%)	296 (99%)	4 (1%)	0	100	100
1	B	312/348 (90%)	308 (99%)	4 (1%)	0	100	100
All	All	612/696 (88%)	604 (99%)	8 (1%)	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	267/302 (88%)	264 (99%)	3 (1%)	73	85
1	B	276/302 (91%)	275 (100%)	1 (0%)	91	96
All	All	543/604 (90%)	539 (99%)	4 (1%)	81	92

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	89	ARG
1	A	305	GLN
1	A	355	LYS
1	B	327	LYS

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

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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 5 ligands modelled in this entry, 5 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 ⓘ

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(Å²)	Q<0.9
1	A	300/348 (86%)	0.89	49 (16%)	1	2	42, 55, 103, 144	0
1	B	304/348 (87%)	0.74	46 (15%)	2	3	37, 53, 83, 102	0
2	C	12/12 (100%)	1.36	4 (33%)	0	0	49, 77, 103, 113	0
2	D	12/12 (100%)	1.59	5 (41%)	0	0	64, 90, 99, 102	0
2	E	12/12 (100%)	0.73	1 (8%)	11	15	43, 72, 88, 91	0
2	F	12/12 (100%)	0.93	1 (8%)	11	15	55, 83, 89, 91	0
All	All	652/744 (87%)	0.84	106 (16%)	1	2	37, 55, 94, 144	0

All (106) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	320	ARG	8.6
1	A	356	ALA	6.3
1	B	57	ALA	5.8
2	C	9	DC	5.5
1	A	194	VAL	5.4
1	A	195	VAL	4.9
1	A	355	LYS	4.7
1	A	287	THR	4.7
2	D	1	DC	4.6
1	B	194	VAL	4.5
1	B	58	GLU	4.4
1	A	336	GLU	4.4
1	B	360	SER	4.3
1	A	239	VAL	4.2
1	A	333	LYS	4.2
1	B	196	SER	4.2
1	A	196	SER	3.9
1	A	354	LEU	3.9
2	D	7	DA	3.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	259	GLN	3.7
1	B	197	GLY	3.7
1	A	94	VAL	3.7
1	B	359	GLN	3.7
1	A	334	ASN	3.6
1	A	141	MET	3.6
1	B	141	MET	3.5
1	B	198	PRO	3.5
2	D	2	DG	3.5
1	B	229	TYR	3.5
1	A	341	ASP	3.5
1	B	356	ALA	3.4
1	A	229	TYR	3.4
1	A	338	GLU	3.2
1	A	231	VAL	3.2
1	B	140	ALA	3.2
1	B	164	VAL	3.2
1	A	111	HIS	3.2
2	F	1	DC	3.1
1	A	140	ALA	3.1
1	A	197	GLY	3.1
1	A	182	CYS	3.1
1	B	175	TRP	3.0
1	B	163	ILE	3.0
1	B	94	VAL	3.0
1	B	357	LYS	2.9
1	A	343	PHE	2.9
1	A	230	LYS	2.9
1	A	309	LEU	2.9
1	B	287	THR	2.8
1	B	206	GLY	2.8
1	A	342	TYR	2.8
1	B	239	VAL	2.8
1	A	95	LEU	2.8
1	A	288	SER	2.8
1	B	358	GLU	2.8
1	A	198	PRO	2.7
1	B	178	ILE	2.7
1	A	142	ALA	2.7
1	B	288	SER	2.7
1	B	90	VAL	2.7
2	C	8	DT	2.7

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	182	CYS	2.7
1	B	59	LYS	2.7
1	B	349	LYS	2.6
1	A	175	TRP	2.6
1	B	230	LYS	2.6
1	A	324	ARG	2.5
1	A	90	VAL	2.5
1	A	96	GLU	2.5
1	B	113	LYS	2.5
1	A	289	ASP	2.5
1	B	91	PRO	2.4
1	B	139	GLY	2.4
1	A	93	TRP	2.4
1	A	337	ILE	2.4
1	B	142	ALA	2.4
1	B	195	VAL	2.4
2	E	9	DC	2.4
1	B	231	VAL	2.4
1	B	228	LEU	2.4
1	A	164	VAL	2.3
1	A	193	TRP	2.3
2	D	3	DG	2.3
2	D	8	DT	2.2
1	B	143	PRO	2.2
1	B	144	ALA	2.2
1	B	320	ARG	2.2
1	A	349	LYS	2.2
1	A	139	GLY	2.2
1	A	335	ALA	2.2
2	C	10	DC	2.2
1	B	305[A]	GLN	2.2
2	C	7	DA	2.2
1	A	330	GLU	2.2
1	A	143	PRO	2.2
1	A	232	ILE	2.1
1	B	199	LEU	2.1
1	B	289	ASP	2.1
1	B	232	ILE	2.1
1	B	98	ILE	2.1
1	A	106	ASP	2.1
1	A	59	LYS	2.1
1	B	96	GLU	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	162	ASN	2.0
1	B	208	GLY	2.0
1	B	95	LEU	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	E	101	1/1	0.91	0.21	44,44,44,44	0
4	MG	B	402	1/1	0.92	0.13	68,68,68,68	0
4	MG	C	101	1/1	0.94	0.17	40,40,40,40	0
3	CL	A	401	1/1	0.95	0.10	59,59,59,59	0
3	CL	B	401	1/1	0.96	0.12	60,60,60,60	0

6.5 Other polymers ⓘ

There are no such residues in this entry.