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

wwPDB X-ray Structure Validation Summary Report ⓘ

Jul 26, 2017 – 05:02 PM EDT

PDB ID	:	4R8P
Title	:	Crystal structure of the Ring1B/Bmi1/UbcH5c PRC1 ubiquitylation module bound to the nucleosome core particle
Authors	:	McGinty, R.K.; Henrici, R.C.; Tan, S.
Deposited on	:	unknown
Resolution	:	3.28 Å(reported)

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

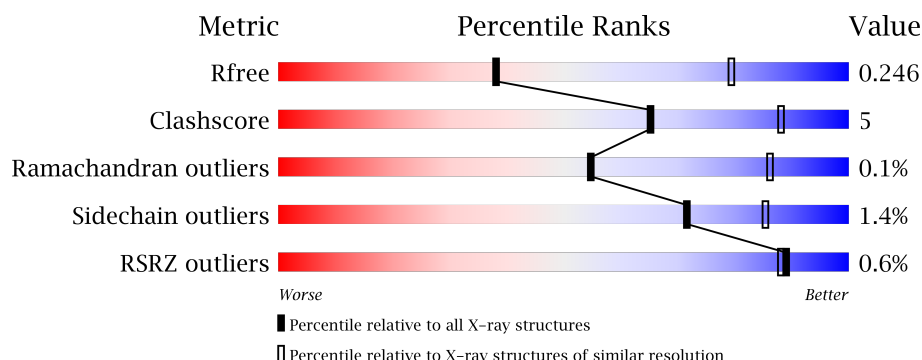|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20029824
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 3.28 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		0.246
Clashscore		5
Ramachandran outliers		0.1%
Sidechain outliers		1.4%
RSRZ outliers		0.6%

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}	100719	1006 (3.34-3.22)
Clashscore	112137	1070 (3.34-3.22)
Ramachandran outliers	110173	1050 (3.34-3.22)
Sidechain outliers	110143	1049 (3.34-3.22)
RSRZ outliers	101464	1011 (3.34-3.22)

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 >=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 <=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	135	.% 67% 6% 27%
1	E	135	.% 66% 6% 28%
2	B	102	76% 8% 16%
2	F	102	73% 8% 20%
3	C	129	67% 14% 19%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	129	.% 77% 6% 17%
4	D	122	71% 7% 22%
4	H	122	72% 6% 22%
5	I	147	76% 22%
6	J	147	.% 75% 24%
7	K	110	3% 83% 13% 5%
7	M	110	75% 13% 12%
8	L	268	.% 74% 19% 5%
8	N	268	81% 13% 6%

2 Entry composition (i)

There are 9 unique types of molecules in this entry. The entry contains 17134 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 Histone H3.2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	98	Total 789	C 500	N 147	O 139	S 3	0	0	0
1	E	97	Total 788	C 498	N 151	O 136	S 3	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	102	ALA	GLY	CONFLICT	UNP P84233
E	102	ALA	GLY	CONFLICT	UNP P84233

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	86	Total 675	C 424	N 133	O 117	S 1	0	0	0
2	F	82	Total 635	C 400	N 121	O 113	S 1	0	0	0

- Molecule 3 is a protein called Histone H2A.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	104	Total 766	C 483	N 143	O 140	0	0	0
3	G	107	Total 793	C 498	N 152	O 143	0	0	0

- Molecule 4 is a protein called Histone H2B 1.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	95	Total 721	C 453	N 127	O 139	S 2	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	H	95	719	451	126	140	2	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	32	THR	SER	CONFLICT	UNP P02281
H	32	THR	SER	CONFLICT	UNP P02281

- Molecule 5 is a DNA chain called DNA (147-mer).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
5	I	144	2935	1393	536	862	144	0	0	0

- Molecule 6 is a DNA chain called DNA (147-mer).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
6	J	145	2990	1415	559	871	145	0	0	0

- Molecule 7 is a protein called Polycomb complex protein BMI-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
7	K	105	819	523	141	146	9	0	0	0
7	M	97	744	477	128	131	8	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
K	0	GLY	-	EXPRESSION TAG	UNP P35226
K	1	SER	-	EXPRESSION TAG	UNP P35226
M	0	GLY	-	EXPRESSION TAG	UNP P35226
M	1	SER	-	EXPRESSION TAG	UNP P35226

- Molecule 8 is a protein called E3 ubiquitin-protein ligase RING2, Ubiquitin-conjugating enzyme E2 D3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	L	254	Total 1857	C 1179	N 315	O 348	S 15	0	0	0
8	N	253	Total 1895	C 1205	N 321	O 353	S 16	0	0	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
L	1	GLY	-	EXPRESSION TAG	UNP Q99496
L	117	GLY	-	LINKER	UNP Q99496
L	118	SER	-	LINKER	UNP Q99496
L	119	GLY	-	LINKER	UNP Q99496
L	120	SER	-	LINKER	UNP Q99496
L	121	ARG	-	LINKER	UNP Q99496
L	122	SER	-	LINKER	UNP Q99496
N	1	GLY	-	EXPRESSION TAG	UNP Q99496
N	117	GLY	-	LINKER	UNP Q99496
N	118	SER	-	LINKER	UNP Q99496
N	119	GLY	-	LINKER	UNP Q99496
N	120	SER	-	LINKER	UNP Q99496
N	121	ARG	-	LINKER	UNP Q99496
N	122	SER	-	LINKER	UNP Q99496

- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	L	2	Total 2	Zn 2	0	0
9	K	2	Total 2	Zn 2	0	0
9	N	2	Total 2	Zn 2	0	0
9	M	2	Total 2	Zn 2	0	0

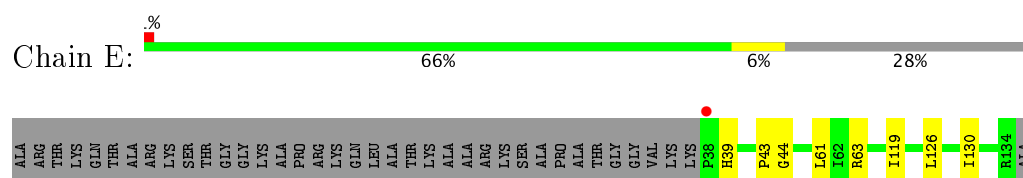# 3 Residue-property plots ⓘ

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: Histone H3.2
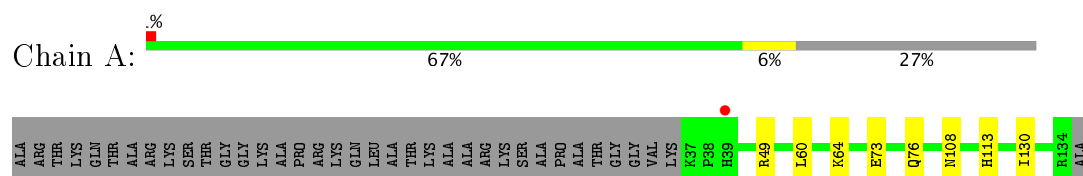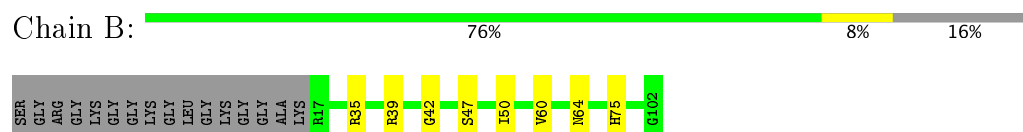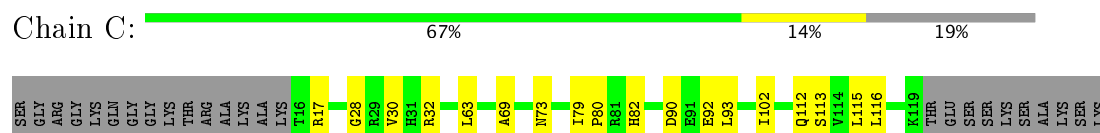
Chain A:

- Molecule 1: Histone H3.2

Chain E:

- Molecule 2: Histone H4

Chain B:

- Molecule 2: Histone H4

Chain F:
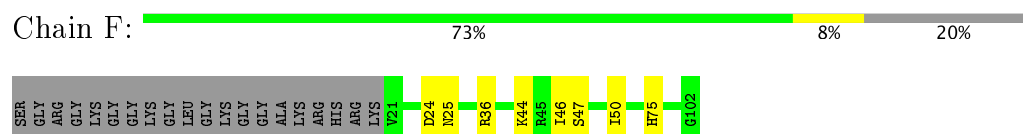
- Molecule 3: Histone H2A

Chain C:
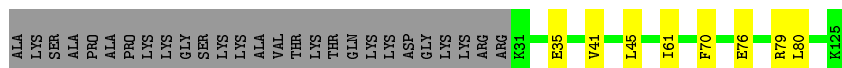
- Molecule 3: Histone H2A
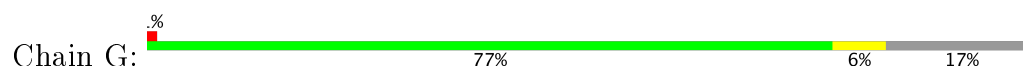
Chain G:
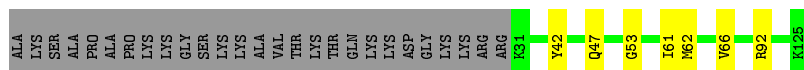
- Molecule 4: Histone H2B 1.1

Chain D:
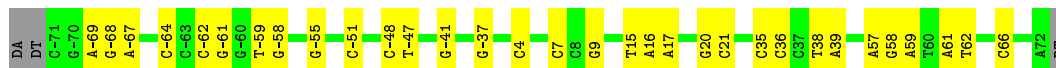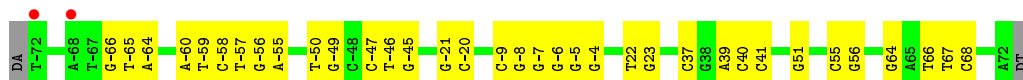

- Molecule 4: Histone H2B 1.1

Chain H:

- Molecule 5: DNA (147-mer)

Chain I:
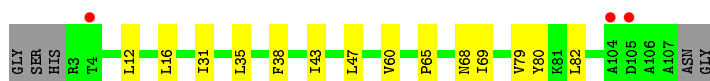

- Molecule 6: DNA (147-mer)

Chain J:

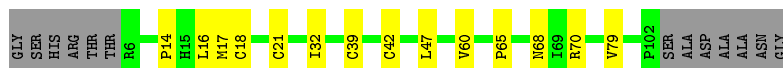- Molecule 7: Polycomb complex protein BMI-1

Chain K:

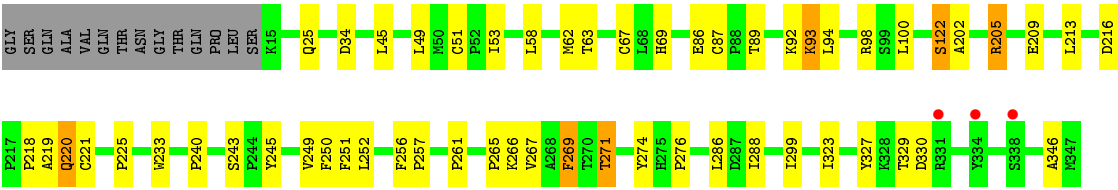- Molecule 7: Polycomb complex protein BMI-1

Chain M:

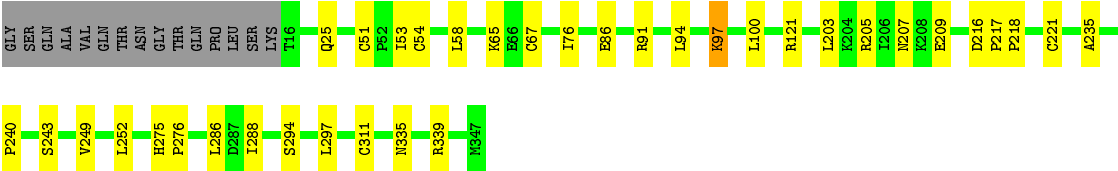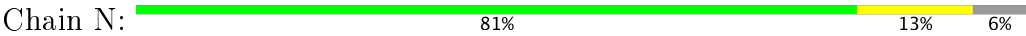- Molecule 8: E3 ubiquitin-protein ligase RING2, Ubiquitin-conjugating enzyme E2 D3

Chain L:

• Molecule 8: E3 ubiquitin-protein ligase RING2, Ubiquitin-conjugating enzyme E2 D3

Chain N:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	104.92Å 180.05Å 375.25Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.00 – 3.28 45.00 – 3.28	Depositor EDS
% Data completeness (in resolution range)	99.1 (45.00-3.28) 99.1 (45.00-3.28)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	1.66 (at 3.25Å)	Xtriage
Refinement program	PHENIX 1.8.4_1496	Depositor
R, R_{free}	0.197 , 0.245 0.198 , 0.246	Depositor DCC
R_{free} test set	2749 reflections (5.08%)	DCC
Wilson B-factor (Å^2)	105.6	Xtriage
Anisotropy	0.343	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 77.8	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.027 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.034 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	17134	wwPDB-VP
Average B, all atoms (Å^2)	137.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.30% 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: ZN

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	0.21	0/801	0.37	0/1079
1	E	0.20	0/800	0.35	0/1076
2	B	0.20	0/682	0.38	0/914
2	F	0.20	0/642	0.38	0/864
3	C	0.19	0/776	0.35	0/1058
3	G	0.19	0/803	0.35	0/1093
4	D	0.20	0/732	0.36	0/989
4	H	0.20	0/730	0.36	0/988
5	I	0.46	0/3289	0.85	0/5069
6	J	0.46	0/3357	0.85	0/5184
7	K	0.21	0/836	0.38	0/1136
7	M	0.20	0/761	0.38	0/1038
8	L	0.20	0/1905	0.38	0/2606
8	N	0.20	0/1945	0.37	0/2660
All	All	0.32	0/18059	0.61	0/25754

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	789	0	809	6	0
1	E	788	0	815	7	0
2	B	675	0	704	5	0
2	F	635	0	652	6	0
3	C	766	0	771	13	0
3	G	793	0	805	7	0
4	D	721	0	713	8	0
4	H	719	0	707	5	0
5	I	2935	0	1615	24	0
6	J	2990	0	1628	27	0
7	K	819	0	807	8	0
7	M	744	0	705	8	0
8	L	1857	0	1734	34	0
8	N	1895	0	1804	21	0
9	K	2	0	0	0	0
9	L	2	0	0	0	0
9	M	2	0	0	0	0
9	N	2	0	0	0	0
All	All	17134	0	14269	142	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.

The worst 5 of 142 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)
8:L:219:ALA:C	8:L:220:GLN:HG3	1.97	0.85
8:L:219:ALA:O	8:L:220:GLN:HG3	1.80	0.81
5:I:-41:DG:H1	6:J:41:DC:H42	1.40	0.70
8:L:53:ILE:O	8:L:205:ARG:NH1	2.26	0.67
8:L:329:THR:N	8:L:330:ASP:HA	2.11	0.64

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	96/135 (71%)	92 (96%)	4 (4%)	0	100	100
1	E	95/135 (70%)	91 (96%)	4 (4%)	0	100	100
2	B	84/102 (82%)	81 (96%)	3 (4%)	0	100	100
2	F	80/102 (78%)	75 (94%)	5 (6%)	0	100	100
3	C	102/129 (79%)	99 (97%)	3 (3%)	0	100	100
3	G	105/129 (81%)	102 (97%)	3 (3%)	0	100	100
4	D	93/122 (76%)	86 (92%)	7 (8%)	0	100	100
4	H	93/122 (76%)	90 (97%)	3 (3%)	0	100	100
7	K	103/110 (94%)	96 (93%)	7 (7%)	0	100	100
7	M	95/110 (86%)	88 (93%)	7 (7%)	0	100	100
8	L	252/268 (94%)	228 (90%)	23 (9%)	1 (0%)	38	72
8	N	251/268 (94%)	232 (92%)	19 (8%)	0	100	100
All	All	1449/1732 (84%)	1360 (94%)	88 (6%)	1 (0%)	55	86

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
8	L	122	SER

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	82/110 (74%)	82 (100%)	0	100	100
1	E	82/110 (74%)	81 (99%)	1 (1%)	75	86
2	B	67/78 (86%)	67 (100%)	0	100	100
2	F	63/78 (81%)	63 (100%)	0	100	100
3	C	75/101 (74%)	75 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	G	77/101 (76%)	77 (100%)	0	100	100
4	D	75/102 (74%)	75 (100%)	0	100	100
4	H	75/102 (74%)	74 (99%)	1 (1%)	73	86
7	K	90/99 (91%)	89 (99%)	1 (1%)	78	87
7	M	79/99 (80%)	79 (100%)	0	100	100
8	L	193/242 (80%)	183 (95%)	10 (5%)	27	62
8	N	204/242 (84%)	201 (98%)	3 (2%)	70	84
All	All	1162/1464 (79%)	1146 (99%)	16 (1%)	71	85

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
8	L	205	ARG
8	L	213	LEU
8	L	327	TYR
8	L	94	LEU
8	N	25	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	108	ASN
8	L	25	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)

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

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 8 ligands modelled in this entry, 8 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 (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	98/135 (72%)	-0.16	1 (1%)	82	81	60, 88, 149, 211	0
1	E	97/135 (71%)	-0.20	1 (1%)	82	81	66, 100, 155, 225	0
2	B	86/102 (84%)	-0.29	0	100	100	57, 81, 138, 207	0
2	F	82/102 (80%)	-0.13	0	100	100	66, 93, 136, 171	0
3	C	104/129 (80%)	-0.26	0	100	100	58, 96, 131, 167	0
3	G	107/129 (82%)	-0.21	1 (0%)	84	83	56, 87, 132, 207	0
4	D	95/122 (77%)	-0.29	0	100	100	59, 94, 148, 194	0
4	H	95/122 (77%)	-0.19	0	100	100	63, 93, 133, 169	0
5	I	144/147 (97%)	-0.30	0	100	100	111, 178, 240, 264	0
6	J	145/147 (98%)	-0.21	2 (1%)	75	73	101, 178, 258, 299	0
7	K	105/110 (95%)	-0.14	3 (2%)	52	49	66, 91, 185, 218	0
7	M	97/110 (88%)	-0.22	0	100	100	89, 126, 200, 251	0
8	L	254/268 (94%)	-0.20	3 (1%)	79	76	60, 124, 231, 286	0
8	N	253/268 (94%)	-0.21	0	100	100	78, 139, 202, 237	0
All	All	1762/2026 (86%)	-0.21	11 (0%)	89	88	56, 114, 216, 299	0

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
8	L	331	ARG	4.2
1	A	39	HIS	3.5
6	J	-72	DT	2.9
7	K	104	ALA	2.3
3	G	115	LEU	2.2

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, 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	LLDF	B-factors(Å^2)	Q<0.9
9	ZN	K	201	1/1	0.99	0.20	0.23	81,81,81,81	0
9	ZN	N	402	1/1	0.99	0.17	0.20	78,78,78,78	0
9	ZN	N	401	1/1	0.99	0.17	0.20	93,93,93,93	0
9	ZN	L	401	1/1	0.99	0.17	-0.21	88,88,88,88	0
9	ZN	K	202	1/1	0.99	0.15	-0.39	66,66,66,66	0
9	ZN	M	201	1/1	0.94	0.14	-0.50	133,133,133,133	0
9	ZN	M	202	1/1	0.99	0.14	-0.61	114,114,114,114	0
9	ZN	L	402	1/1	1.00	0.13	-0.65	58,58,58,58	0

6.5 Other polymers (i)

There are no such residues in this entry.