



wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 08:27 am BST

PDB ID : 5UI8
Title : structure of sigmaN-holoenzyme
Authors : Darst, S.A.; Campbell, E.A.
Deposited on : 2017-01-13
Resolution : 3.76 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
Xtriage (Phenix)	:	1.13
EDS	:	2.11
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)	:	2.11

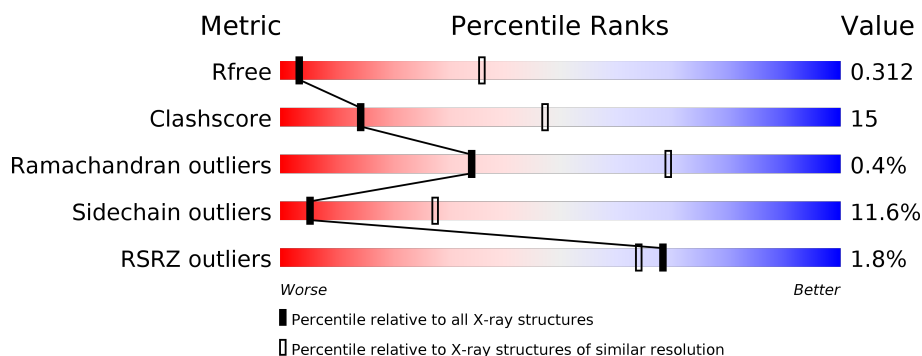
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.76 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	1039 (3.94-3.58)
Clashscore	141614	1051 (3.92-3.60)
Ramachandran outliers	138981	1015 (3.92-3.60)
Sidechain outliers	138945	1011 (3.92-3.60)
RSRZ outliers	127900	1050 (3.96-3.56)

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 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	G	329	<div> <div>%</div> <div> <div></div> <div>61%</div> <div>29%</div> <div>•</div> <div>6%</div> </div> </div>
1	H	329	<div> <div>2%</div> <div> <div></div> <div>42%</div> <div>21%</div> <div>•</div> <div>35%</div> </div> </div>
2	I	1342	<div> <div>%</div> <div> <div></div> <div>55%</div> <div>31%</div> <div>•</div> <div>9%</div> </div> </div>
3	J	1407	<div> <div>2%</div> <div> <div></div> <div>59%</div> <div>32%</div> <div>•</div> <div>5%</div> </div> </div>
4	K	91	<div> <div>%</div> <div> <div></div> <div>54%</div> <div>26%</div> <div>7%</div> <div>13%</div> </div> </div>
5	M	477	<div> <div></div> <div> <div></div> <div>47%</div> <div>28%</div> <div>5%</div> <div>•</div> <div>20%</div> </div> </div>

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 27004 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 DNA-directed RNA polymerase subunit alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	G	309	Total	C	N	O	S	0	0	0
			2337	1463	408	459	7			
1	H	215	Total	C	N	O	S	0	0	0
			1619	1013	279	321	6			

- Molecule 2 is a protein called DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	I	1217	Total	C	N	O	S	0	0	0
			9382	5883	1632	1827	40			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
I	516	VAL	ASP	conflict	UNP B7MIX3

- Molecule 3 is a protein called DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	J	1336	Total	C	N	O	S	0	0	0
			10148	6364	1812	1923	49			

- Molecule 4 is a protein called DNA-directed RNA polymerase subunit omega.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	K	79	Total	C	N	O	S	0	0	0
			620	377	118	124	1			

- Molecule 5 is a protein called RNA polymerase sigma-54 factor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	M	380	Total	C	N	O	S	0	0	0
			2895	1825	499	561	10			

- Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	J	1	Total	Mg	0	0
			1	1		

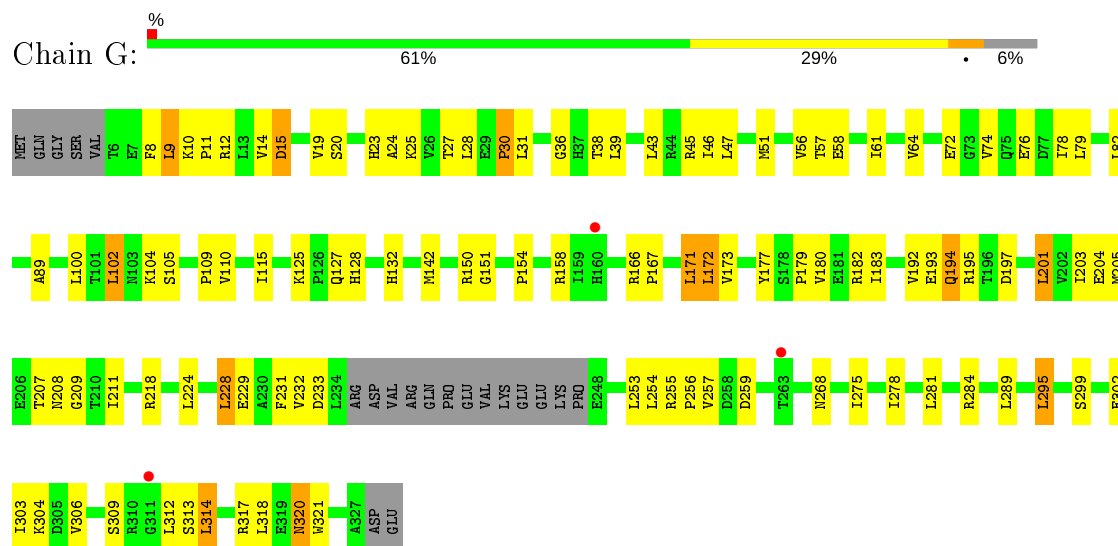
- Molecule 7 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	J	2	Total	Zn	0	0
			2	2		

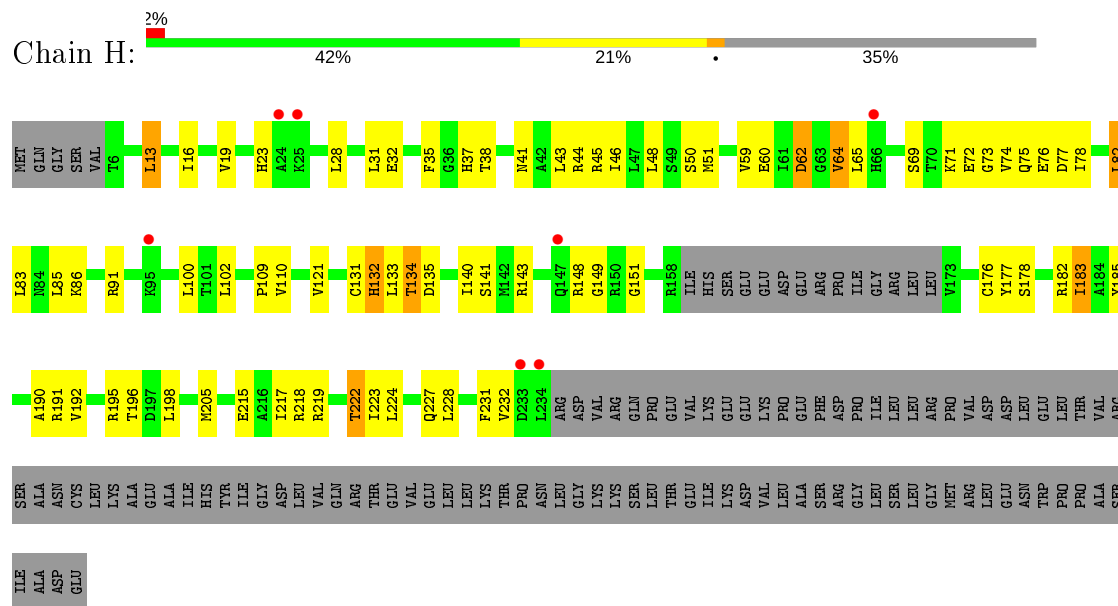
3 Residue-property plots [i](#)

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: DNA-directed RNA polymerase subunit alpha

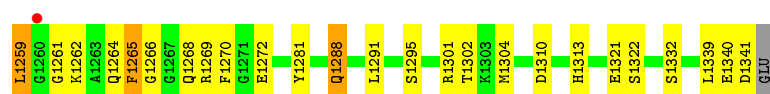


- Molecule 1: DNA-directed RNA polymerase subunit alpha

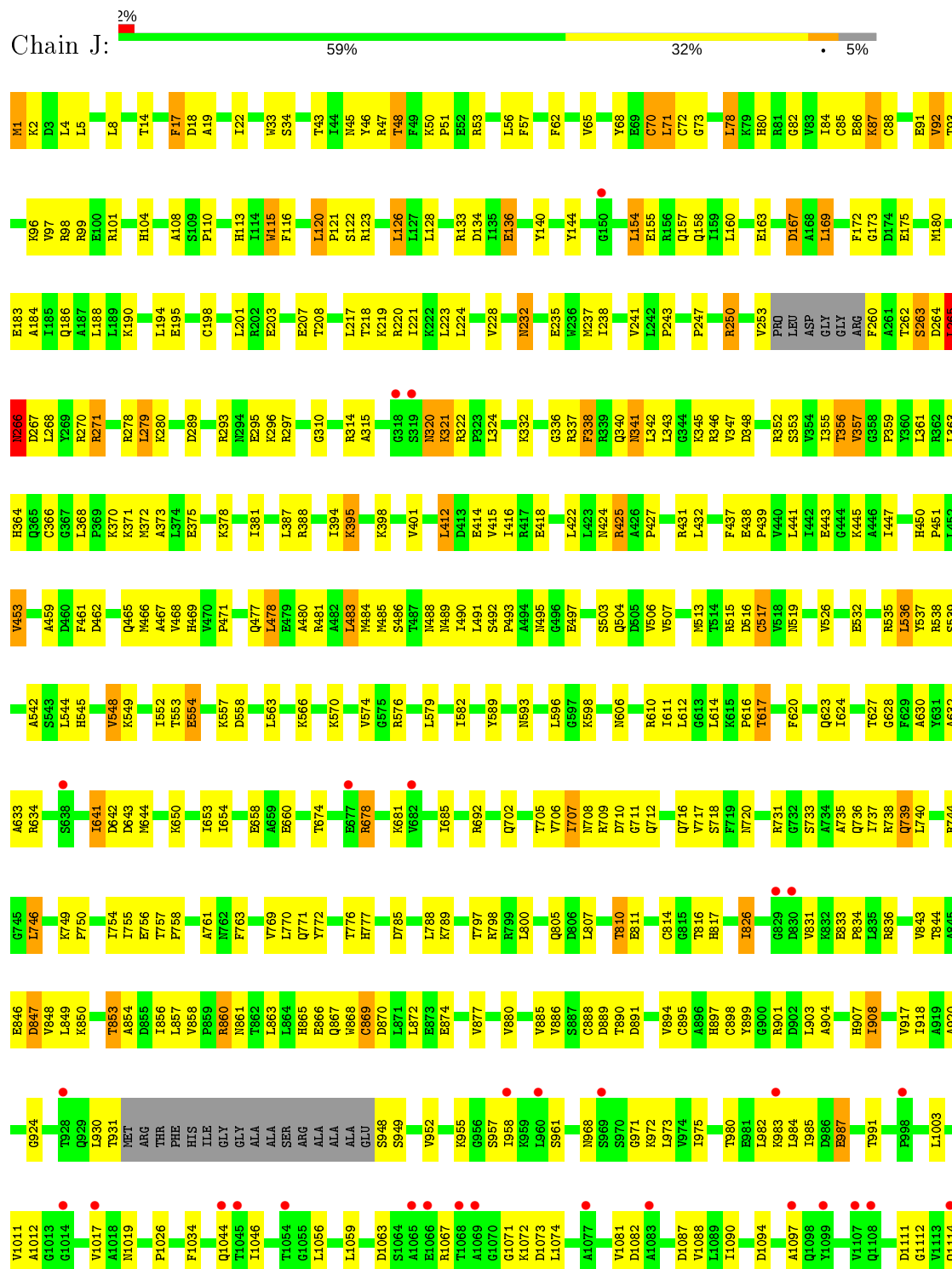


- Molecule 2: DNA-directed RNA polymerase subunit beta



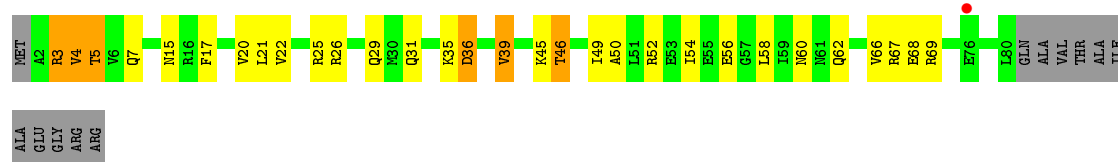


• Molecule 3: DNA-directed RNA polymerase subunit beta'

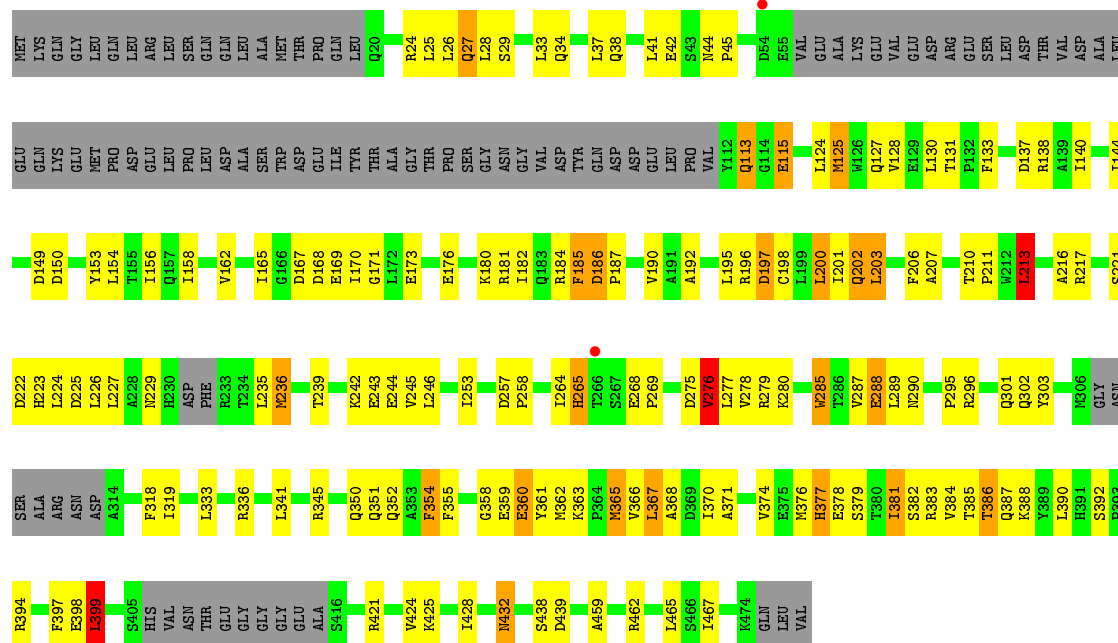




• Molecule 4: DNA-directed RNA polymerase subunit omega



• Molecule 5: RNA polymerase sigma-54 factor



4 Data and refinement statistics

Property	Value	Source
Space group	P 2 ₁ 2 ₁ 2	Depositor
Cell constants a, b, c, α , β , γ	208.48 Å 151.52 Å 195.28 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.64 – 3.76 29.64 – 3.76	Depositor EDS
% Data completeness (in resolution range)	98.4 (29.64-3.76) 98.4 (29.64-3.76)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.18 (at 3.75 Å)	Xtriage
Refinement program	PHENIX 1.10.1 _2155	Depositor
R, R_{free}	0.262 , 0.312 0.262 , 0.312	Depositor DCC
R_{free} test set	3074 reflections (4.93%)	wwPDB-VP
Wilson B-factor (Å ²)	143.4	Xtriage
Anisotropy	0.735	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 103.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	27004	wwPDB-VP
Average B, all atoms (Å ²)	178.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.33% 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

5.1 Standard geometry

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

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	G	0.27	0/2366	0.55	0/3215
1	H	0.25	0/1637	0.49	0/2224
2	I	0.28	1/9533 (0.0%)	0.49	0/12893
3	J	0.26	1/10294 (0.0%)	0.48	3/13920 (0.0%)
4	K	0.24	0/622	0.43	0/838
5	M	0.30	0/2935	0.57	4/3989 (0.1%)
All	All	0.27	2/27387 (0.0%)	0.50	7/37079 (0.0%)

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	G	0	1
2	I	0	5
3	J	0	4
5	M	0	1
All	All	0	11

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	I	109	ALA	C-N	11.59	1.56	1.34
3	J	120	LEU	C-N	6.71	1.47	1.34

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	M	213	LEU	CA-CB-CG	8.34	134.47	115.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	J	425	ARG	NE-CZ-NH2	-7.53	116.53	120.30
5	M	399	LEU	CA-CB-CG	6.10	129.34	115.30
3	J	265	LEU	N-CA-C	-5.19	97.00	111.00
3	J	78	LEU	CA-CB-CG	5.13	127.10	115.30

There are no chirality outliers.

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	G	194	GLN	Peptide
2	I	1040	ASP	Peptide
2	I	1154	ASP	Peptide
2	I	1202	GLY	Peptide
2	I	889	PRO	Peptide

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	G	2337	0	2343	75	0
1	H	1619	0	1619	54	0
2	I	9382	0	9211	315	0
3	J	10148	0	10169	325	1
4	K	620	0	621	20	0
5	M	2895	0	2828	115	1
6	J	1	0	0	0	0
7	J	2	0	0	0	0
All	All	27004	0	26791	831	1

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 15.

The worst 5 of 831 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)
2:I:49:LEU:O	2:I:53:PHE:HB2	1.68	0.93

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:G:43:LEU:O	1:G:47:LEU:HB2	1.74	0.88
5:M:216:ALA:HB1	5:M:253:ILE:HD11	1.56	0.87
3:J:289:ASP:O	3:J:293:ARG:HB2	1.77	0.84
5:M:127:GLN:HE22	5:M:182:ILE:HA	1.45	0.82

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:J:82:GLY:O	5:M:438:SER:OG[3_745]	2.18	0.02

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	G	305/329 (93%)	270 (88%)	33 (11%)	2 (1%)	22	59
1	H	211/329 (64%)	191 (90%)	20 (10%)	0	100	100
2	I	1211/1342 (90%)	1094 (90%)	110 (9%)	7 (1%)	25	61
3	J	1326/1407 (94%)	1198 (90%)	125 (9%)	3 (0%)	47	78
4	K	77/91 (85%)	71 (92%)	6 (8%)	0	100	100
5	M	370/477 (78%)	345 (93%)	22 (6%)	3 (1%)	19	56
All	All	3500/3975 (88%)	3169 (90%)	316 (9%)	15 (0%)	34	69

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	I	166	SER
3	J	266	ASN
5	M	276	VAL
2	I	170	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	I	437	ASN

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	G	252/286 (88%)	229 (91%)	23 (9%)	9	37
1	H	175/286 (61%)	158 (90%)	17 (10%)	8	34
2	I	999/1157 (86%)	884 (88%)	115 (12%)	5	28
3	J	1058/1168 (91%)	934 (88%)	124 (12%)	5	27
4	K	65/75 (87%)	55 (85%)	10 (15%)	2	17
5	M	306/423 (72%)	265 (87%)	41 (13%)	4	22
All	All	2855/3395 (84%)	2525 (88%)	330 (12%)	5	27

5 of 330 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	I	1265	PHE
3	J	241	VAL
5	M	236	MET
2	I	1304	MET
3	J	87	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 50 such sidechains are listed below:

Mol	Chain	Res	Type
2	I	1175	ASN
3	J	157	GLN
5	M	322	ASN
2	I	1220	GLN
2	I	1288	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 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 [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	G	309/329 (93%)	-0.22	3 (0%) 82 79	115, 174, 216, 248	0
1	H	215/329 (65%)	0.16	7 (3%) 46 40	144, 209, 239, 261	0
2	I	1217/1342 (90%)	-0.07	19 (1%) 72 67	97, 179, 242, 273	0
3	J	1336/1407 (94%)	0.00	33 (2%) 57 51	88, 182, 243, 270	0
4	K	79/91 (86%)	-0.16	1 (1%) 77 73	131, 168, 207, 230	0
5	M	380/477 (79%)	-0.34	2 (0%) 91 89	103, 150, 191, 226	0
All	All	3536/3975 (88%)	-0.07	65 (1%) 68 64	88, 177, 239, 273	0

The worst 5 of 65 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	J	1054	THR	4.8
2	I	197	ARG	4.0
3	J	1108	GLN	4.0
3	J	1099	TYR	3.9
2	I	106	GLU	3.6

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. 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
7	ZN	J	1503	1/1	0.93	0.16	179,179,179,179	0
6	MG	J	1501	1/1	0.94	0.15	192,192,192,192	0
7	ZN	J	1502	1/1	0.98	0.20	225,225,225,225	0

6.5 Other polymers [i](#)

There are no such residues in this entry.