wwPDB/EMDataBank EM Map/Model Validation Summary Report (i)

Oct 29, 2019 – 04:27 PM EDT

PDB ID	:	6K4Y
EMDB ID:	:	EMD-9916
Title	:	CryoEM structure of sigma appropriation complex
Authors	:	Shi, J.; Wen, A.; Feng, Y.
Deposited on	:	2019-05-27
Resolution	:	3.79 Å(reported)

This is a wwPDB/EMDataBank EM Map/Model Validation Summary Report for a publicly released PDB/EMDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the (i) symbol.

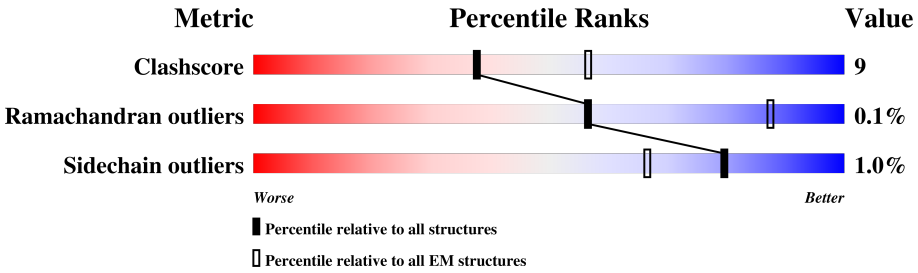
MolProbity	:	4.02b-467
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.4

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 3.79 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)	EM structures (#Entries)
Clashscore	136327	1886
Ramachandran outliers	132723	1663
Sidechain outliers	132532	1531

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the 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%

Mol	Chain	Length	Quality of chain
1	A	329	54% 12% 33%
1	B	329	54% 11% 34%
2	C	1342	80% 19%
3	D	1407	72% 23% 5%
4	E	91	59% 27% 13%
5	F	613	56% 19% 24%
6	I	110	69% 10% • 20%
7	M	211	67% 31% ••
8	N	60	68% 30% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	T	60	

2 Entry composition (i)

There are 11 unique types of molecules in this entry. The entry contains 33247 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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					AltConf	Trace
1	A	219	Total	C	N	O	S	0	0
			1686	1056	298	326	6		
1	B	217	Total	C	N	O	S	0	0
			1673	1046	296	325	6		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
2	C	1340	Total	C	N	O	S	0	0
			10567	6631	1841	2052	43		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
3	D	1335	Total	C	N	O	S	0	0
			10384	6524	1851	1959	50		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
4	E	79	Total	C	N	O	S	0	0
			627	382	118	126	1		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
5	F	463	Total	C	N	O	S	0	0
			3781	2372	676	710	23		

- Molecule 6 is a protein called 10 kDa anti-sigma factor.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	I	88	Total	C	N	O	S	0	0
			729	462	125	141	1		

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
I	-19	MET	-	initiating methionine	UNP P32267
I	-18	GLY	-	expression tag	UNP P32267
I	-17	SER	-	expression tag	UNP P32267
I	-16	SER	-	expression tag	UNP P32267
I	-15	HIS	-	expression tag	UNP P32267
I	-14	HIS	-	expression tag	UNP P32267
I	-13	HIS	-	expression tag	UNP P32267
I	-12	HIS	-	expression tag	UNP P32267
I	-11	HIS	-	expression tag	UNP P32267
I	-10	HIS	-	expression tag	UNP P32267
I	-9	SER	-	expression tag	UNP P32267
I	-8	SER	-	expression tag	UNP P32267
I	-7	GLY	-	expression tag	UNP P32267
I	-6	LEU	-	expression tag	UNP P32267
I	-5	VAL	-	expression tag	UNP P32267
I	-4	PRO	-	expression tag	UNP P32267
I	-3	ARG	-	expression tag	UNP P32267
I	-2	GLY	-	expression tag	UNP P32267
I	-1	SER	-	expression tag	UNP P32267
I	0	HIS	-	expression tag	UNP P32267

- Molecule 7 is a protein called Middle transcription regulatory protein motA.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	M	209	Total	C	N	O	S	0	0
			1638	1032	285	314	7		

- Molecule 8 is a DNA chain called DNA (60-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
8	N	59	Total	C	N	O	P	0	0
			1223	582	231	351	59		

- Molecule 9 is a DNA chain called DNA (60-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
9	T	46	Total	C	N	O	P	0	0
			936	449	163	278	46		

- Molecule 10 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
10	D	1	Total	Mg	0
			1	1	

- Molecule 11 is ZINC ION (three-letter code: ZN) (formula: Zn).

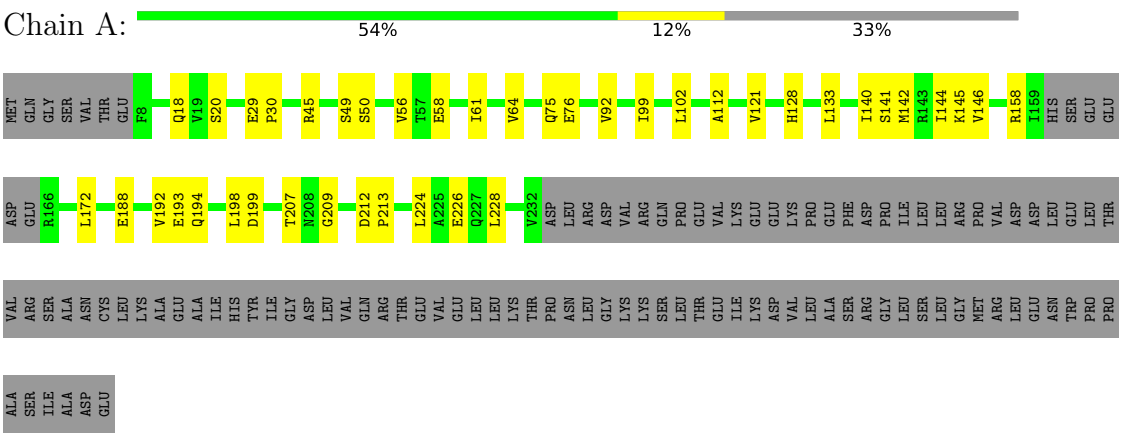
Mol	Chain	Residues	Atoms		AltConf
11	D	2	Total	Zn	0
			2	2	

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. 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. 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.

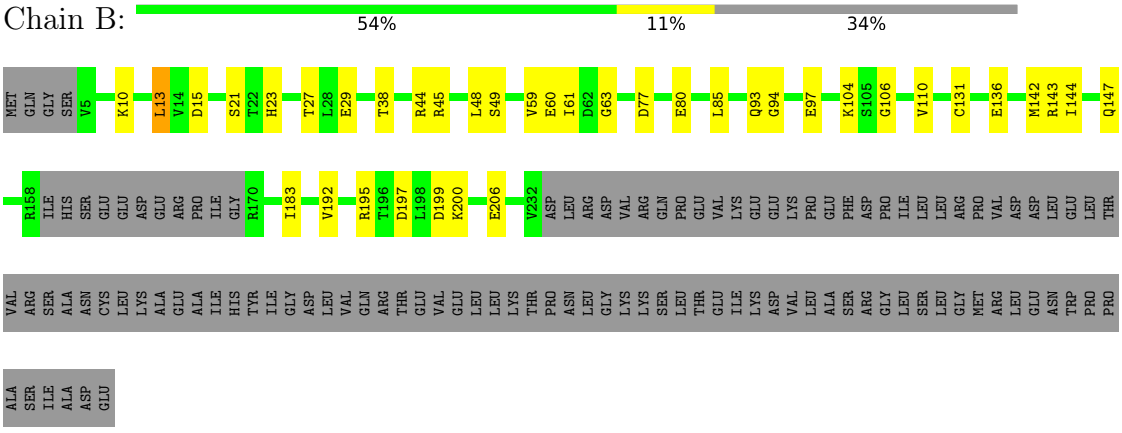
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain A:



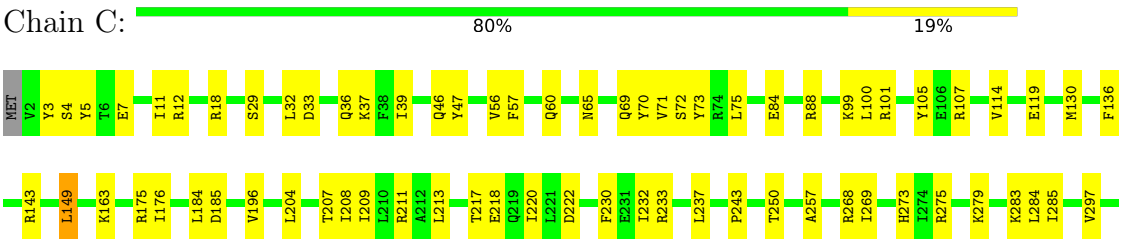
• Molecule 1: DNA-directed RNA polymerase subunit alpha

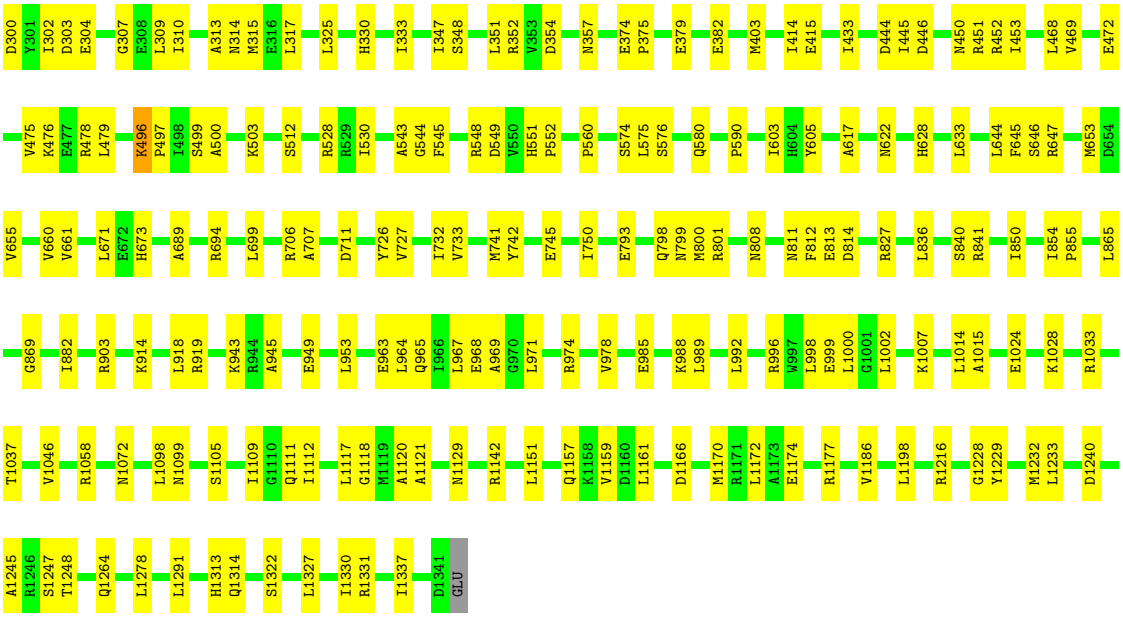
Chain B:



• Molecule 2: DNA-directed RNA polymerase subunit beta

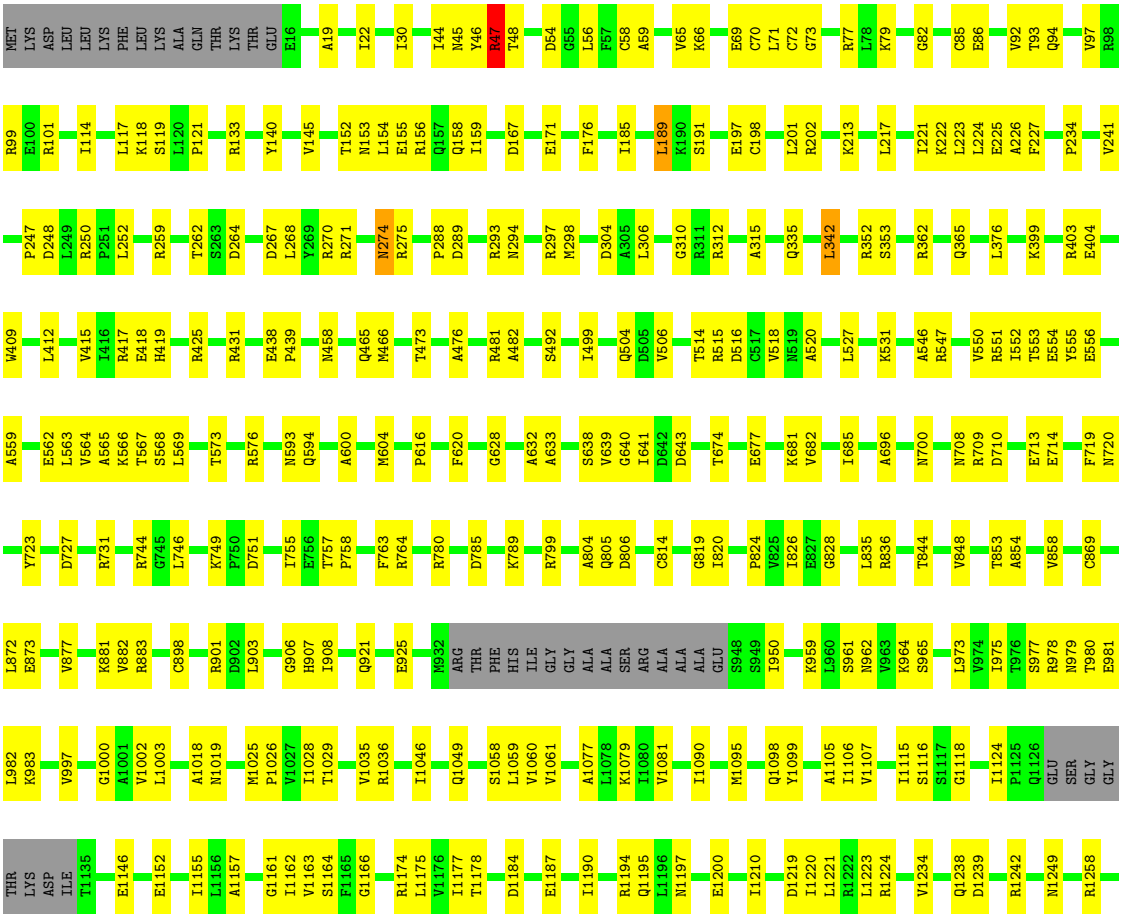
Chain C:

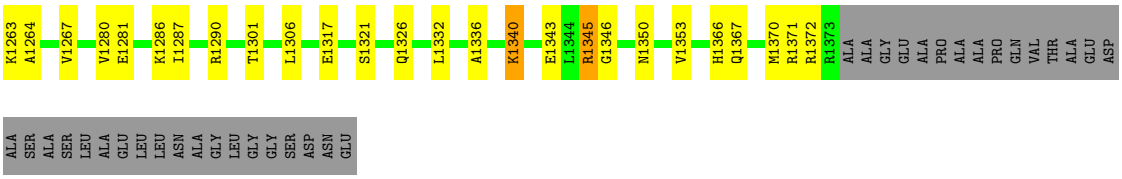




• Molecule 3: DNA-directed RNA polymerase subunit beta'

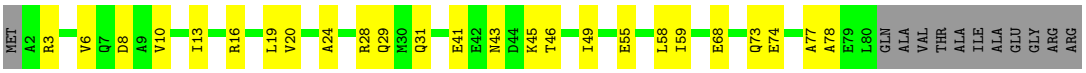
Chain D:





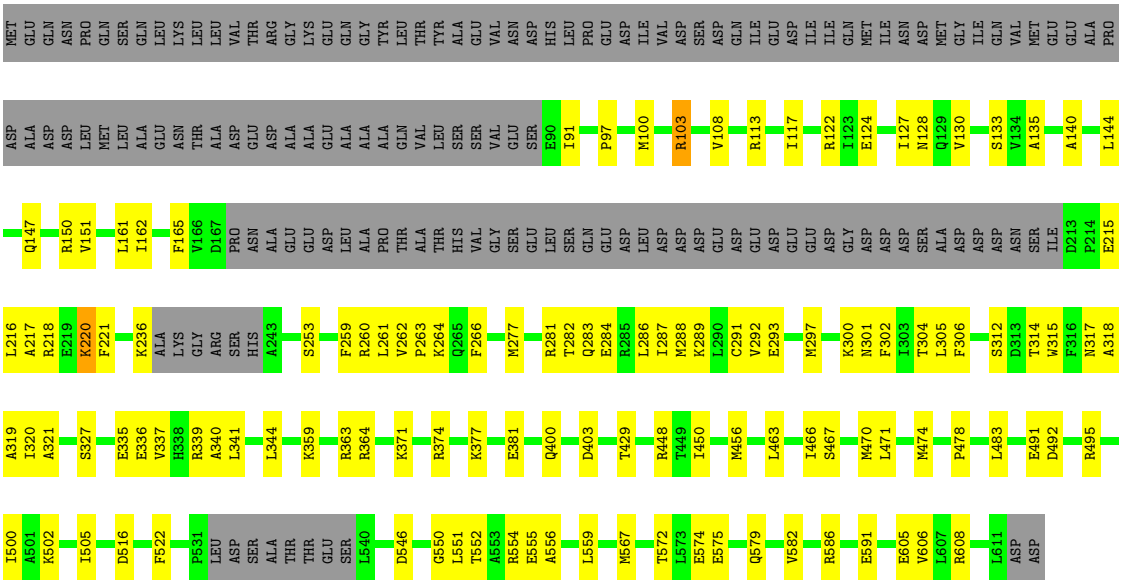
• Molecule 4: DNA-directed RNA polymerase subunit omega

Chain E:



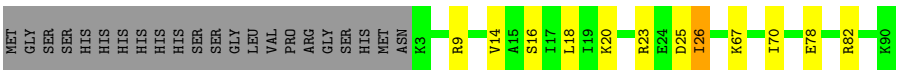
• Molecule 5: RNA polymerase sigma factor RpoD

Chain F:



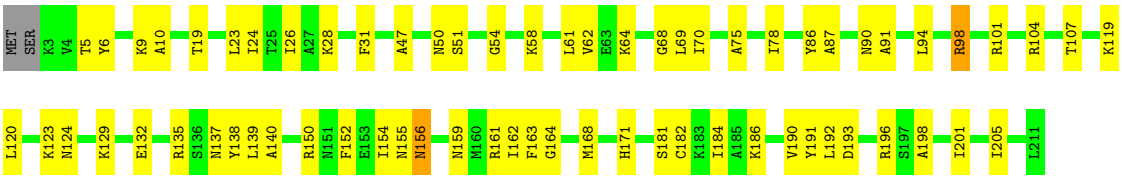
• Molecule 6: 10 kDa anti-sigma factor

Chain I:



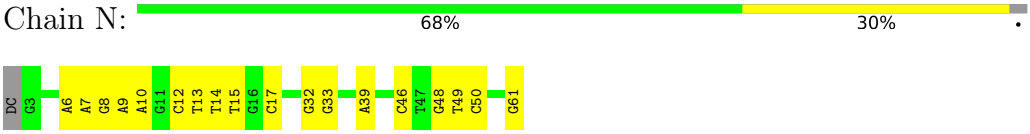
• Molecule 7: Middle transcription regulatory protein motA

Chain M:



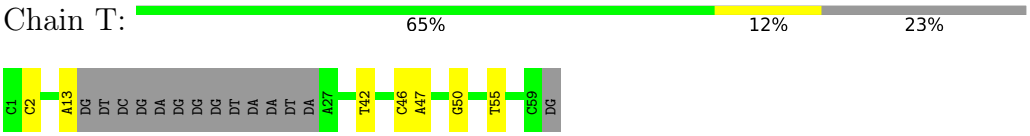
• Molecule 8: DNA (60-MER)

Chain N:



• Molecule 9: DNA (60-MER)

Chain T:



4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	105108	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	56	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

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	$\#\|Z\| > 2$	RMSZ	$\#\|Z\| > 2$
1	A	0.42	0/1706	0.59	0/2312
1	B	0.35	0/1692	0.59	0/2293
2	C	0.43	0/10736	0.58	0/14487
3	D	0.40	0/10541	0.61	0/14232
4	E	0.31	0/629	0.56	0/847
5	F	0.32	0/3831	0.53	0/5145
6	I	0.42	0/738	0.58	0/990
7	M	0.30	0/1653	0.54	0/2219
8	N	0.79	0/1375	0.99	0/2123
9	T	0.82	0/1046	1.06	0/1608
All	All	0.44	0/33947	0.63	0/46256

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

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	1686	0	1726	25	0
1	B	1673	0	1710	26	0
2	C	10567	0	10585	158	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	D	10384	0	10606	219	0
4	E	627	0	634	19	0
5	F	3781	0	3858	86	0
6	I	729	0	730	11	0
7	M	1638	0	1707	49	0
8	N	1223	0	666	15	0
9	T	936	0	523	6	0
10	D	1	0	0	0	0
11	D	2	0	0	0	0
All	All	33247	0	32745	571	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 9.

The worst 5 of 571 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:13:LEU:HD23	1:B:13:LEU:H	1.16	1.05
7:M:64:LYS:HA	7:M:68:GLY:O	1.65	0.95
2:C:1245:ALA:HB1	3:D:376:LEU:CD1	1.97	0.93
5:F:335:GLU:O	5:F:339:ARG:HB2	1.67	0.93
3:D:961:SER:O	3:D:980:THR:HA	1.68	0.92

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 PDB entries followed by that with respect to all EM entries.

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	215/329 (65%)	202 (94%)	13 (6%)	0	100 100
1	B	213/329 (65%)	198 (93%)	14 (7%)	1 (0%)	31 71

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	C	1338/1342 (100%)	1229 (92%)	109 (8%)	0	100	100
3	D	1329/1407 (94%)	1209 (91%)	117 (9%)	3 (0%)	49	83
4	E	77/91 (85%)	71 (92%)	6 (8%)	0	100	100
5	F	455/613 (74%)	427 (94%)	28 (6%)	0	100	100
6	I	86/110 (78%)	82 (95%)	4 (5%)	0	100	100
7	M	207/211 (98%)	200 (97%)	7 (3%)	0	100	100
All	All	3920/4432 (88%)	3618 (92%)	298 (8%)	4 (0%)	56	87

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	21	SER
3	D	710	ASP
3	D	1345	ARG
3	D	47	ARG

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 PDB entries followed by that with respect to all EM entries.

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	186/286 (65%)	186 (100%)	0	100	100
1	B	185/286 (65%)	182 (98%)	3 (2%)	65	85
2	C	1155/1157 (100%)	1145 (99%)	10 (1%)	81	91
3	D	1120/1168 (96%)	1110 (99%)	10 (1%)	81	91
4	E	67/75 (89%)	67 (100%)	0	100	100
5	F	412/540 (76%)	405 (98%)	7 (2%)	63	84
6	I	81/100 (81%)	80 (99%)	1 (1%)	74	88
7	M	179/181 (99%)	176 (98%)	3 (2%)	63	84
All	All	3385/3793 (89%)	3351 (99%)	34 (1%)	80	89

5 of 34 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	274	ASN
3	D	744	ARG
7	M	98	ARG
3	D	431	ARG
2	C	478	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	1313	HIS
3	D	157	GLN
5	F	409	ASN
2	C	1264	GLN
2	C	1268	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.