



wwPDB EM Map/Model Validation Report ⓘ

Apr 10, 2016 – 02:18 PM BST

PDB ID : 5FUR
EMDB ID: : EMD-3305
Title : Structure of human TFIID-IIA bound to core promoter DNA
Authors : Louder, R.K.; He, Y.; Lopez-Blanco, J.R.; Fang, J.; Chacon, P.; Nogales, E.
Deposited on : 2016-01-29
Resolution : 8.50 Å(reported)

This is a wwPDB EM Map/Model Validation Report for a publicly released PDB/EMDB entry.
For rigid body fitted models, validation errors reported here could stem from errors in the original structure(s) used in the fitting.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
<http://wwpdb.org/validation/2016/EMValidationReportHelp>

MolProbity : 4.02b-467
Mogul : 1.7.1 (RC1), CSD as537be (2016)
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : trunk27241

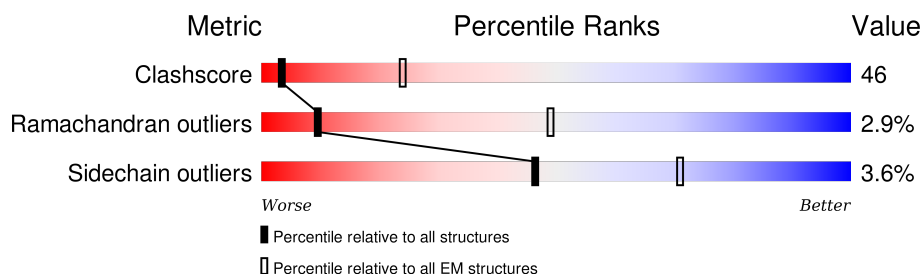
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 8.50 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	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686

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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	339	44% 8% • 47%
2	B	43	77% 21% •
3	C	47	53% 45% •
4	D	97	57% 41% •
5	E	89	• 69% 18% • 10%
6	F	93	• 68% 16% 14%
7	G	1893	19% • 79%
8	H	349	31% • 65%
9	I	1199	33% 31% 12% • 24%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
10	J	677	<div><div></div><div>22%8%..68%</div></div>
10	K	677	<div><div></div><div>18%8%..71%</div></div>
11	L	310	<div><div></div><div>• 5%91%</div></div>

2 Entry composition

There are 11 unique types of molecules in this entry. The entry contains 21485 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 TATA-BOX-BINDING PROTEIN.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	180	Total	C	N	O	S	0	0
			1429	927	252	243	7		

- Molecule 2 is a protein called TRANSCRIPTION INITIATION FACTOR IIA SUBUNIT 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	43	Total	C	N	O	S	0	0
			356	228	56	70	2		

- Molecule 3 is a protein called TRANSCRIPTION INITIATION FACTOR IIA SUBUNIT 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	47	Total	C	N	O	S	0	0
			393	250	70	71	2		

- Molecule 4 is a protein called TRANSCRIPTION INITIATION FACTOR IIA SUBUNIT 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	D	97	Total	C	N	O	S	0	0
			793	502	140	149	2		

- Molecule 5 is a DNA chain called SUPER CORE PROMOTER.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	E	80	Total	C	N	O	P	0	0
			1654	778	320	476	80		

- Molecule 6 is a DNA chain called SUPER CORE PROMOTER.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	F	80	Total	C	N	O	P	0	0
			1626	770	292	484	80		

- Molecule 7 is a protein called TRANSCRIPTION INITIATION FACTOR TFIID SUBUNIT 1.

Mol	Chain	Residues	Atoms						AltConf	Trace
7	G	406	Total	C	N	O	P	S	0	0
			3290	2090	580	596	2	22		

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	178	VAL	-	INSERTION	UNP P21675
G	179	SER	-	INSERTION	UNP P21675
G	180	GLU	-	INSERTION	UNP P21675
G	181	ASN	-	INSERTION	UNP P21675
G	182	GLY	-	INSERTION	UNP P21675
G	183	GLU	-	INSERTION	UNP P21675
G	184	GLY	-	INSERTION	UNP P21675
G	185	ILE	-	INSERTION	UNP P21675
G	186	ILE	-	INSERTION	UNP P21675
G	187	LEU	-	INSERTION	UNP P21675
G	188	PRO	-	INSERTION	UNP P21675
G	189	SER	-	INSERTION	UNP P21675
G	190	ILE	-	INSERTION	UNP P21675
G	191	ILE	-	INSERTION	UNP P21675
G	192	ALA	-	INSERTION	UNP P21675
G	193	PRO	-	INSERTION	UNP P21675
G	194	SER	-	INSERTION	UNP P21675
G	195	SER	-	INSERTION	UNP P21675
G	196	LEU	-	INSERTION	UNP P21675
G	197	ALA	-	INSERTION	UNP P21675
G	198	SER	-	INSERTION	UNP P21675

- Molecule 8 is a protein called TRANSCRIPTION INITIATION FACTOR TFIID SUBUNIT 7.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	H	123	Total	C	N	O	S	0	0
			998	638	184	172	4		

- Molecule 9 is a protein called TRANSCRIPTION INITIATION FACTOR TFIID SUBUNIT 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	I	914	Total	C	N	O	S	0	2
			7404	4761	1251	1336	56		

- Molecule 10 is a protein called TRANSCRIPTION INITIATION FACTOR TFIID SUBUNIT 6.

Mol	Chain	Residues	Atoms					AltConf	Trace
10	J	220	Total	C	N	O	S	0	0
			1741	1106	306	318	11		
10	K	198	Total	C	N	O	S	0	0
			1582	1006	276	290	10		

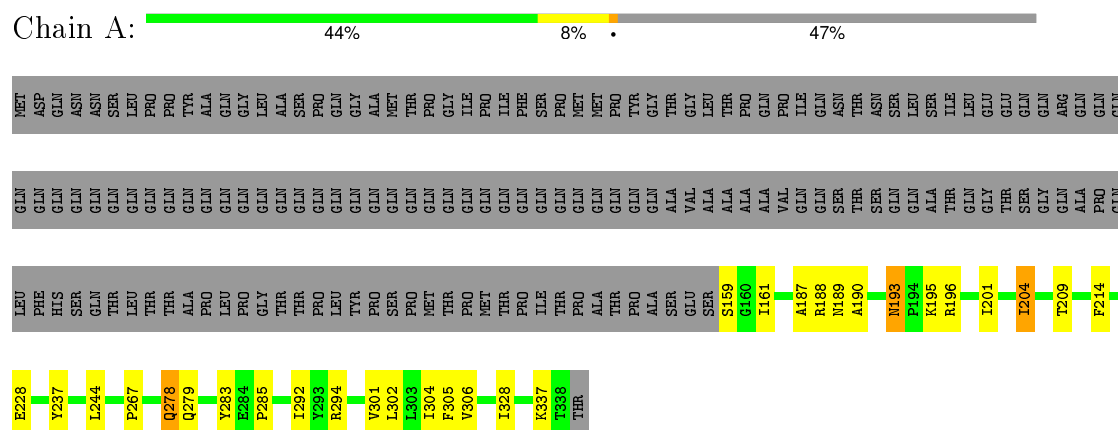
- Molecule 11 is a protein called TRANSCRIPTION INITIATION FACTOR TFIID SUBUNIT 8.

Mol	Chain	Residues	Atoms					AltConf	Trace
11	L	27	Total	C	N	O	S	0	1
			219	134	46	38	1		

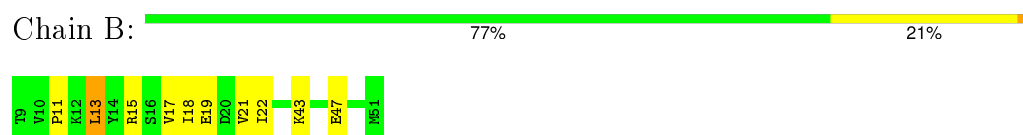
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

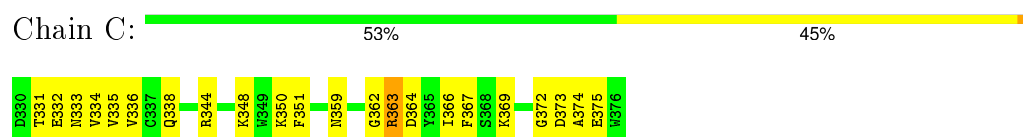
• Molecule 1: TATA-BOX-BINDING PROTEIN



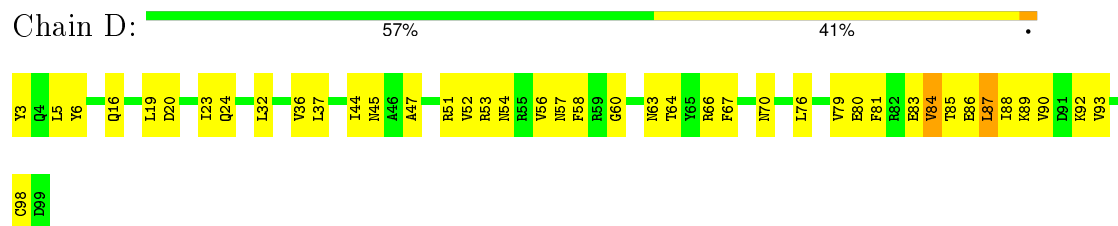
• Molecule 2: TRANSCRIPTION INITIATION FACTOR IIA SUBUNIT 1



• Molecule 3: TRANSCRIPTION INITIATION FACTOR IIA SUBUNIT 1



• Molecule 4: TRANSCRIPTION INITIATION FACTOR IIA SUBUNIT 2

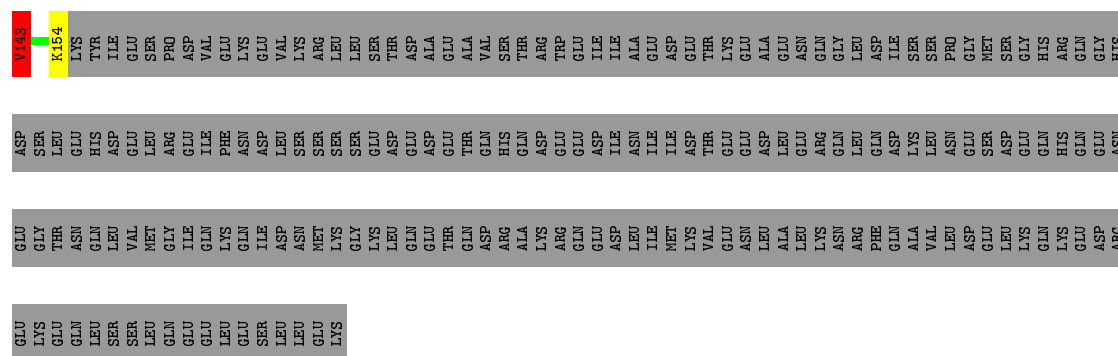


• Molecule 5: SUPER CORE PROMOTER

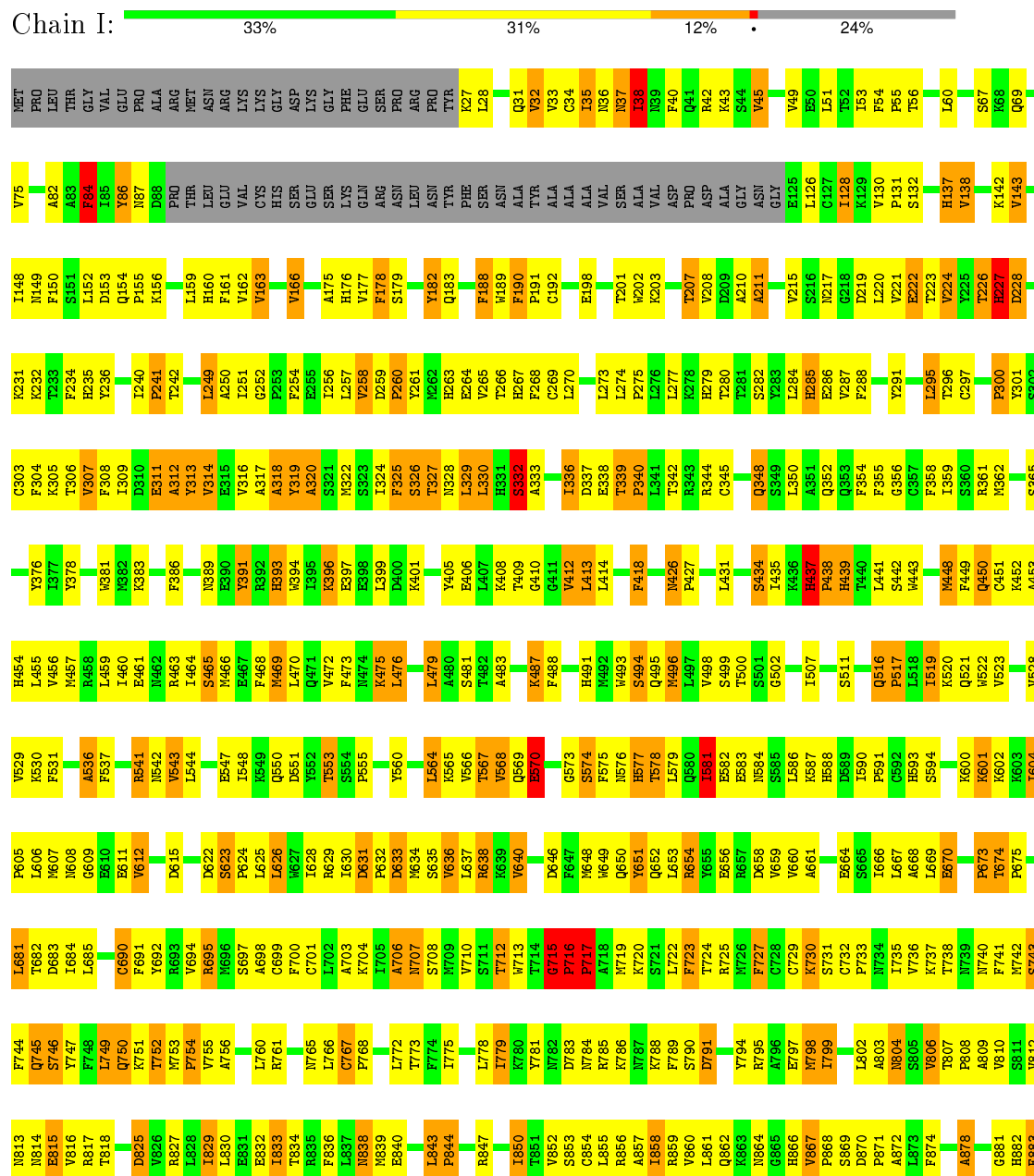
G60	C61	C62	G63	A64	G65	C66	A67	G68	A69	C70	G71	T72	G73	C74	C75	T76	A77	C78	G79	G80	DA	DC	DC	DA	DT	DG	DG
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	----	----	----	----	----	----

A130	C131	G132	C133	G134	C135	C136	C137	C138	C139	A140	C141	C142	C143	C144	C145	T146	T147	T148	T149	A150	T151	A152	G153	G154	C155	G156	C157	C158	C159	T160	T161	DC
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	----

LEU	ARG	THR	PRO	GLY	GLY	ASN	L609	V617	E618	L619	G647	L660	K666	K669	M670	R675	S678	G679	G680	L689	I700	L701	L711	A718	K722	T748	L752	L768	R789	Y793	E796	L797	P808	R819	A820	L824	S826
-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------



• Molecule 9: TRANSCRIPTION INITIATION FACTOR TFIIID SUBUNIT 2





4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of images	Not provided	Depositor
Resolution determination method	Not provided	Depositor
CTF correction method	EACH PARTICLE, Not provided	Depositor
Microscope	OTHER	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	Not provided	Depositor
Minimum defocus (nm)	2000	Depositor
Maximum defocus (nm)	4000	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT	Depositor

5 Model quality ⓘ

5.1 Standard geometry ⓘ

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

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	0.52	0/1455	0.69	0/1958
10	J	0.74	0/1773	1.42	26/2408 (1.1%)
10	K	0.75	0/1612	1.38	21/2188 (1.0%)
11	L	0.61	1/220 (0.5%)	1.10	0/292
2	B	0.34	0/360	0.53	0/487
3	C	0.40	0/402	0.83	1/539 (0.2%)
4	D	0.35	0/803	0.73	2/1088 (0.2%)
5	E	3.85	16/1806 (0.9%)	1.89	49/2658 (1.8%)
6	F	3.55	15/1764 (0.9%)	2.11	46/2582 (1.8%)
7	G	0.41	0/3349	0.55	0/4506
8	H	0.42	0/1017	0.59	1/1370 (0.1%)
9	I	0.84	8/7587 (0.1%)	0.82	5/10278 (0.0%)
All	All	1.61	40/22148 (0.2%)	1.16	151/30354 (0.5%)

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
10	J	2	16
10	K	1	12
5	E	0	2
7	G	0	1
9	I	0	235
All	All	3	266

The worst 5 of 40 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	E	43	DC	O3'-P	49.26	2.20	1.61

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	E	44	DG	O3'-P	48.90	2.19	1.61
5	E	62	DC	O3'-P	48.41	2.19	1.61
5	E	59	DA	O3'-P	46.14	2.16	1.61
6	F	105	DA	O3'-P	46.09	2.16	1.61

The worst 5 of 151 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	F	145	DC	O5'-P-OP2	-41.73	60.62	110.70
9	I	437	HIS	C-N-CD	-27.62	59.83	120.60
5	E	17	DG	P-O3'-C3'	-27.17	87.10	119.70
6	F	145	DC	P-O5'-C5'	-24.28	82.05	120.90
6	F	156	DG	O5'-P-OP1	-23.00	83.11	110.70

All (3) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
10	J	326	HIS	CA
10	J	411	VAL	CA
10	K	326	HIS	CA

5 of 266 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
5	E	10	DA	Sidechain
5	E	7	DC	Sidechain
7	G	1104	LEU	Peptide
9	I	32	VAL	Mainchain
9	I	33	VAL	Mainchain

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	1429	0	1521	49	0
2	B	356	0	360	7	0
3	C	393	0	380	26	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	793	0	801	39	0
5	E	1654	0	958	423	0
6	F	1626	0	962	396	0
7	G	3290	0	3276	61	0
8	H	998	0	1055	6	0
9	I	7404	0	7381	897	0
10	J	1741	0	1782	103	0
10	K	1582	0	1612	40	0
11	L	219	0	222	130	0
All	All	21485	0	20310	1916	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 46.

The worst 5 of 1916 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)
9:I:84:PHE:HZ	9:I:126:LEU:CG	1.10	1.64
5:E:70:DC:C5'	7:G:875:ARG:NH1	1.68	1.53
9:I:566:VAL:HB	9:I:579:LEU:CD1	1.36	1.51
9:I:309:ILE:CG2	9:I:312:ALA:HB2	1.42	1.48
9:I:84:PHE:CZ	9:I:126:LEU:CG	1.96	1.47

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 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	178/339 (52%)	175 (98%)	3 (2%)	0	100	100
2	B	41/43 (95%)	40 (98%)	1 (2%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	C	45/47 (96%)	43 (96%)	2 (4%)	0	100	100
4	D	95/97 (98%)	87 (92%)	7 (7%)	1 (1%)	17	63
7	G	398/1893 (21%)	391 (98%)	7 (2%)	0	100	100
8	H	119/349 (34%)	117 (98%)	2 (2%)	0	100	100
9	I	910/1199 (76%)	753 (83%)	122 (13%)	35 (4%)	4	37
10	J	218/677 (32%)	185 (85%)	17 (8%)	16 (7%)	1	21
10	K	196/677 (29%)	163 (83%)	20 (10%)	13 (7%)	1	24
11	L	25/310 (8%)	25 (100%)	0	0	100	100
All	All	2225/5631 (40%)	1979 (89%)	181 (8%)	65 (3%)	9	43

5 of 65 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	84	VAL
9	I	222	GLU
9	I	413	LEU
9	I	438	PRO
9	I	577	HIS

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 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	155/293 (53%)	152 (98%)	3 (2%)	65	86
2	B	42/42 (100%)	41 (98%)	1 (2%)	57	82
3	C	42/42 (100%)	41 (98%)	1 (2%)	57	82
4	D	89/89 (100%)	87 (98%)	2 (2%)	60	83
7	G	355/1680 (21%)	331 (93%)	24 (7%)	20	57
8	H	113/322 (35%)	104 (92%)	9 (8%)	15	50
9	I	832/1083 (77%)	830 (100%)	2 (0%)	95	97
10	J	194/574 (34%)	177 (91%)	17 (9%)	12	45

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
10	K	176/574 (31%)	162 (92%)	14 (8%)	15	50
11	L	22/270 (8%)	22 (100%)	0	100	100
All	All	2020/4969 (41%)	1947 (96%)	73 (4%)	46	74

5 of 73 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
8	H	99	VAL
9	I	84	PHE
10	K	326	HIS
8	H	141	LYS
10	J	219	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
9	I	542	ASN
9	I	838	ASN
10	K	220	GLN
4	D	13	ASN
10	K	254	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

2 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
7	SEP	G	1105	7	7,9,10	1.81	2 (28%)	8,12,14	1.17	1 (12%)
7	TPO	G	1106	7	7,10,11	1.25	0	10,14,16	1.20	1 (10%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
7	SEP	G	1105	7	-	0/5/8/10	0/0/0/0
7	TPO	G	1106	7	-	0/8/11/13	0/0/0/0

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	G	1105	SEP	P-O2P	2.05	1.61	1.54
7	G	1105	SEP	P-O1P	3.50	1.61	1.50

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	G	1106	TPO	P-OG1-CB	-2.73	109.50	121.42
7	G	1105	SEP	OG-P-O1P	2.18	112.57	107.08

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

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.