wwPDB EM Map/Model Validation Report (i)

Apr 10, 2016 – 01:44 PM BST

PDB ID	:	3IYD
EMDB ID:	:	EMD-5127
Title	:	Three-dimensional EM structure of an intact activator-dependent transcription initiation complex
Authors	:	Hudson, B.P.; Quispe, J.; Lara, S.; Kim, Y.; Berman, H.; Arnold, E.; Ebright, R.H.; Lawson, C.L.
Deposited on	:	2009-08-01
Resolution	:	19.80 Å(reported)
Based on PDB ID	:	2AUK, 1SIG, 1LB2, 3DXJ, 1BDF

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

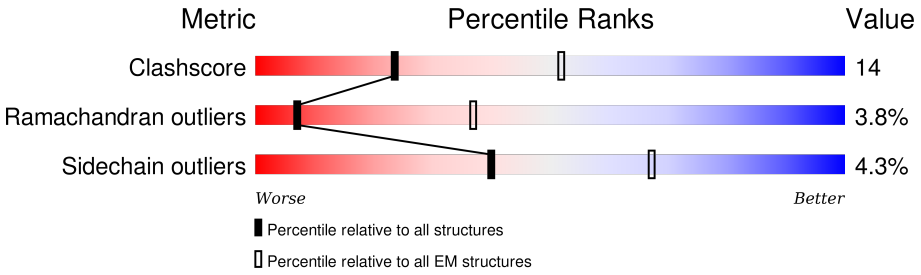
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk27241

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 19.80 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.



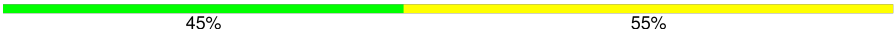
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686

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 <=5%

Mol	Chain	Length	Quality of chain
1	A	329	67% 28% • •
1	B	329	47% 22% • 29%
2	C	1342	53% 25% • 19%
3	D	1413	63% 29% 5% •
4	E	90	57% 33% 9% •
5	F	613	57% 18% • 22%
6	G	209	77% 19% •
6	H	209	72% 23% •
7	I	98	44% 56%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
8	J	98	

2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 35254 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	322	Total	C	N	O	S	0	0
			2496	1563	438	487	8		
1	B	235	Total	C	N	O	S	0	0
			1814	1129	320	358	7		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
2	C	1089	Total	C	N	O	S	0	0
			8524	5351	1481	1655	37		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
3	D	1368	Total	C	N	O	S	0	0
			10606	6657	1893	2006	50		

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	1408	HIS	-	EXPRESSION TAG	UNP P0A8T7
D	1409	HIS	-	EXPRESSION TAG	UNP P0A8T7
D	1410	HIS	-	EXPRESSION TAG	UNP P0A8T7
D	1411	HIS	-	EXPRESSION TAG	UNP P0A8T7
D	1412	HIS	-	EXPRESSION TAG	UNP P0A8T7
D	1413	HIS	-	EXPRESSION TAG	UNP P0A8T7

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

Mol	Chain	Residues	Atoms					AltConf	Trace
4	E	90	Total	C	N	O	S	0	0
			709	430	136	142	1		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
5	F	479	Total	C	N	O	S	0	0
			3877	2423	693	740	21		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	149	ASN	ASP	CONFLICT	UNP P00579
F	571	HIS	TYR	CONFLICT	UNP P00579

- Molecule 6 is a protein called Catabolite gene activator.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	G	201	Total	C	N	O	S	0	0
			1591	1007	280	295	9		
6	H	201	Total	C	N	O	S	0	0
			1591	1007	280	295	9		

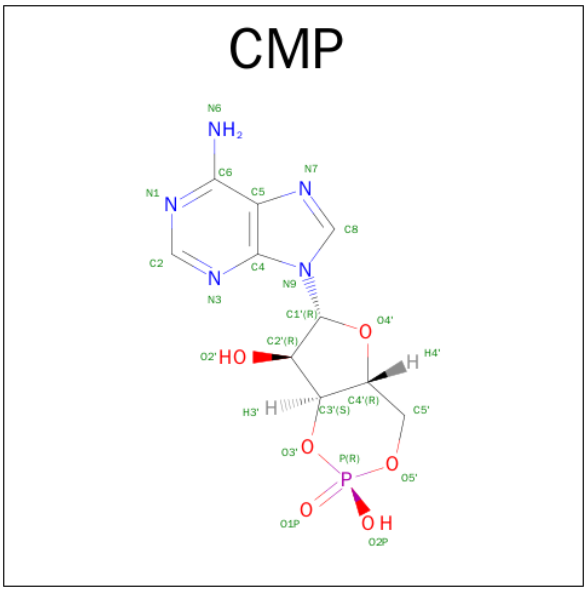
- Molecule 7 is a DNA chain called DNA (98-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
7	I	98	Total	C	N	O	P	0	0
			2006	960	366	583	97		

- Molecule 8 is a DNA chain called DNA (98-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
8	J	98	Total	C	N	O	P	0	0
			1996	958	353	588	97		

- Molecule 9 is ADENOSINE-3',5'-CYCLIC-MONOPHOSPHATE (three-letter code: CMP) (formula: $C_{10}H_{12}N_5O_6P$).



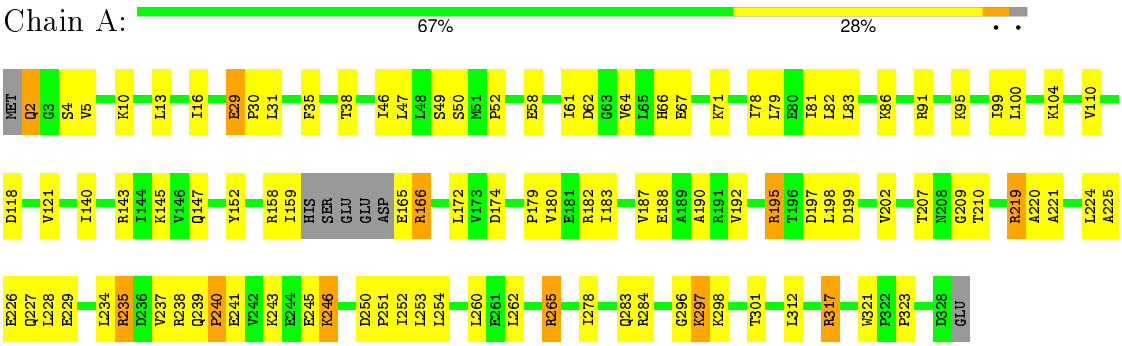
Mol	Chain	Residues	Atoms					AltConf
9	G	1	Total	C	N	O	P	0
			22	10	5	6	1	
9	H	1	Total	C	N	O	P	0
			22	10	5	6	1	

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.

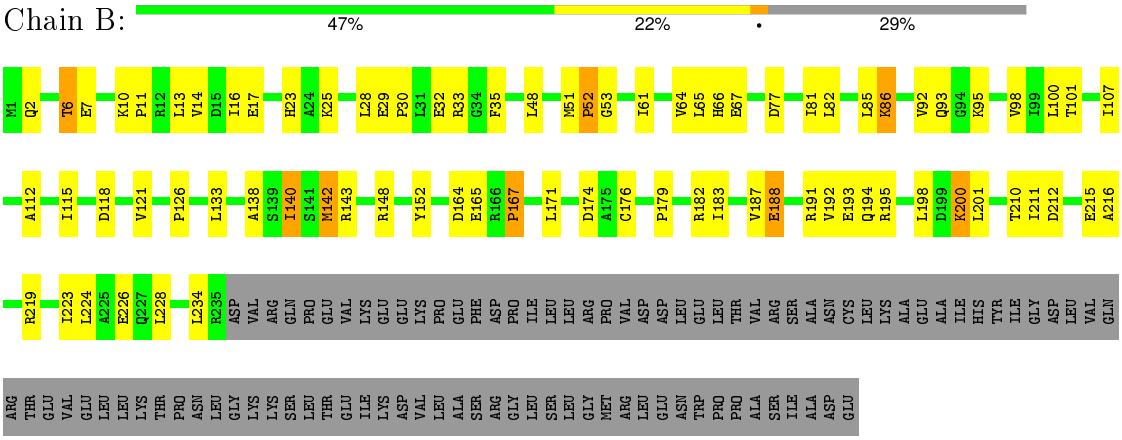
- Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain A:



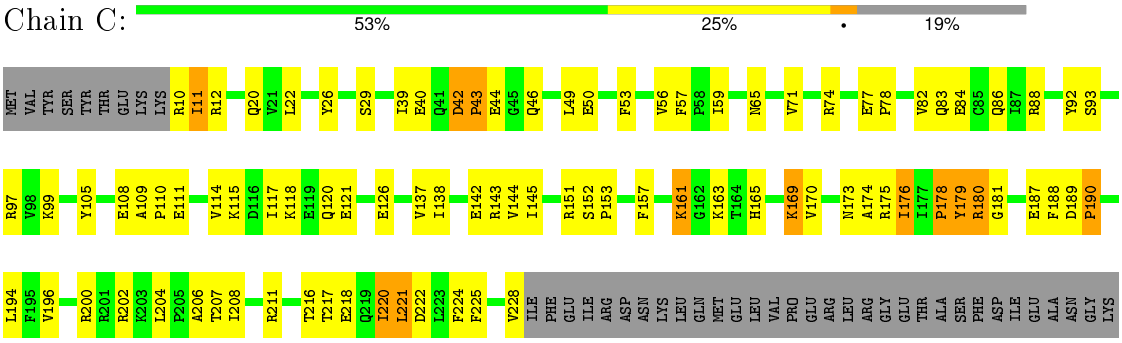
- Molecule 1: DNA-directed RNA polymerase subunit alpha

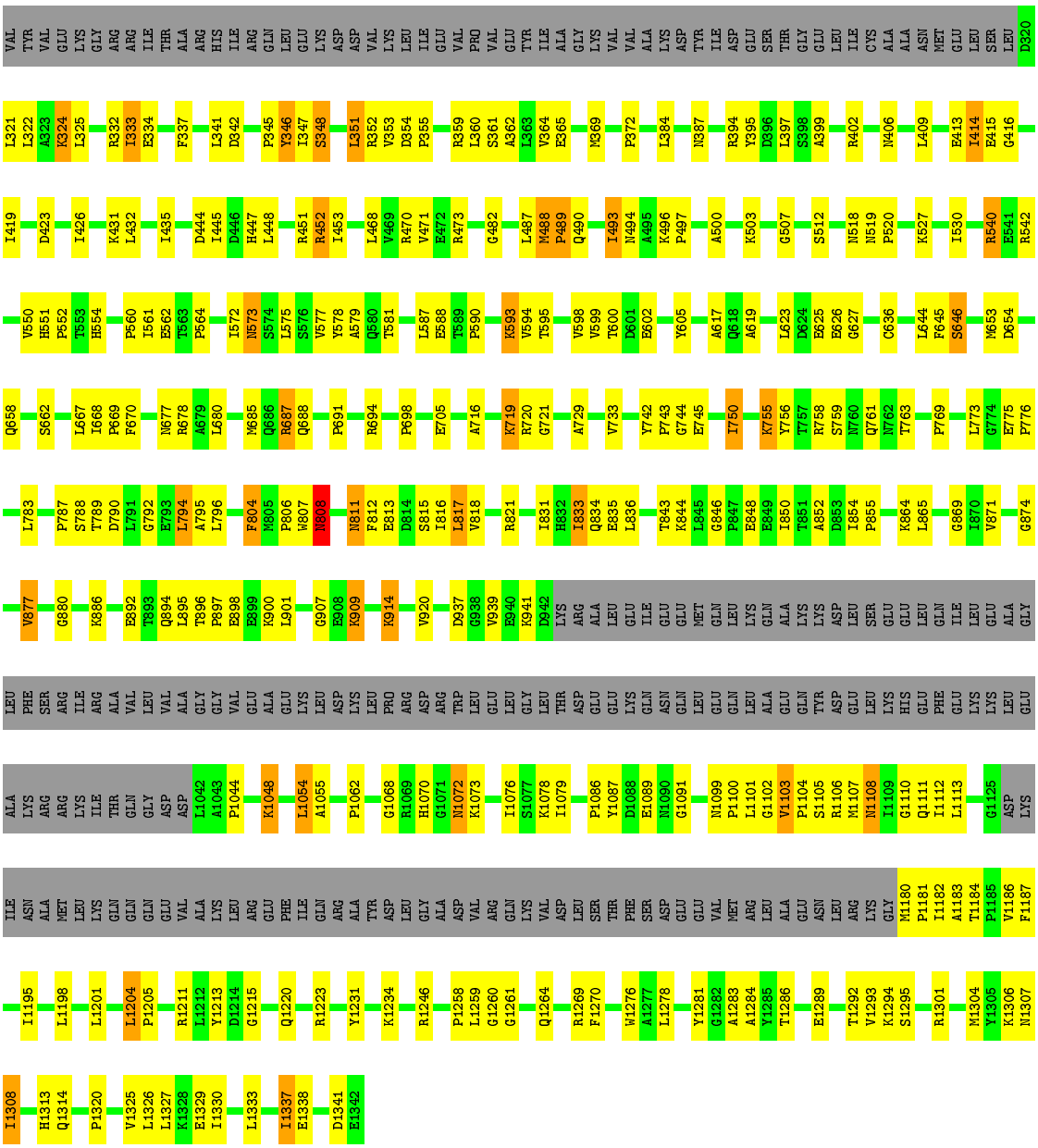
Chain B:



- Molecule 2: DNA-directed RNA polymerase subunit beta

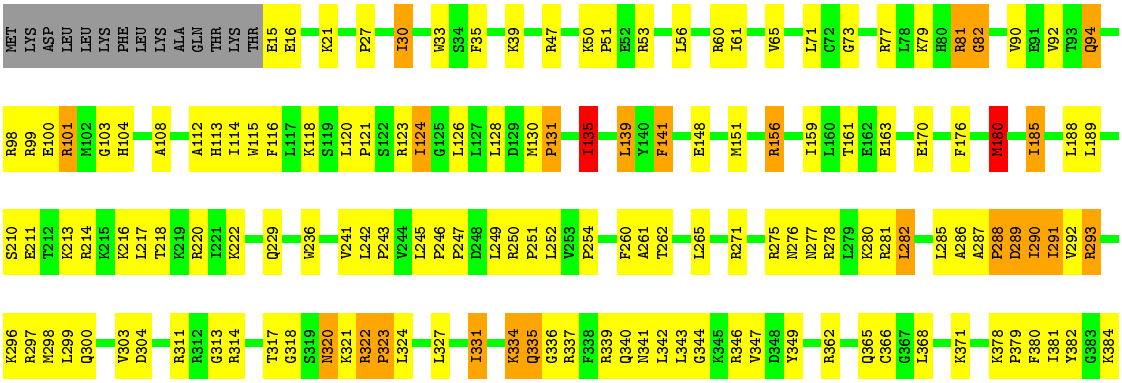
Chain C:



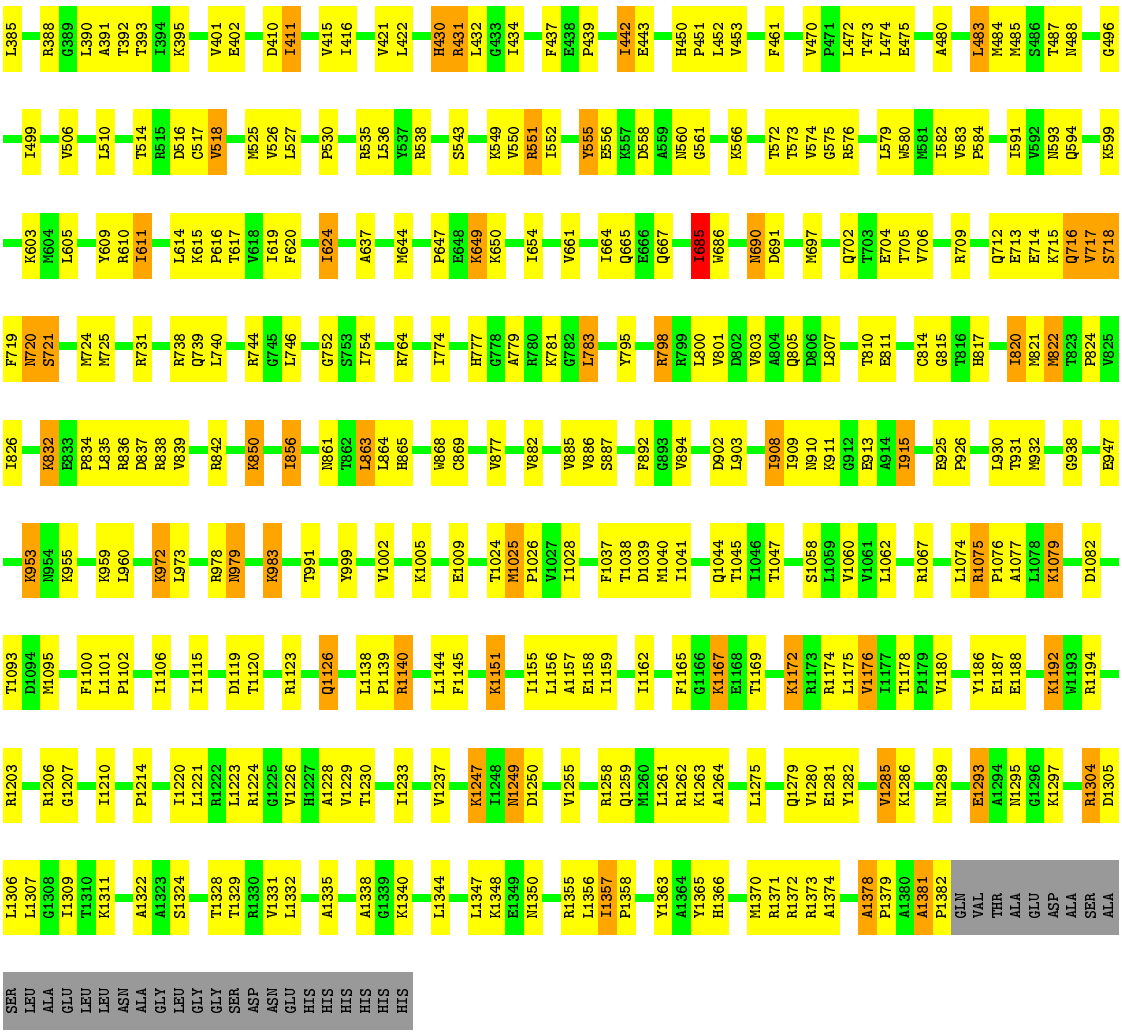


• Molecule 3: DNA-directed RNA polymerase subunit beta

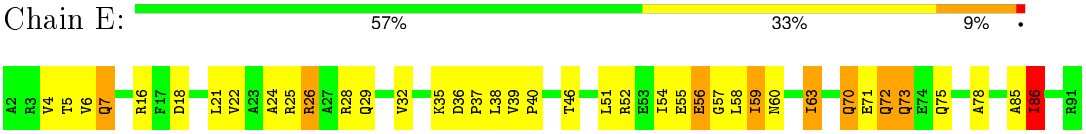
Chain D:



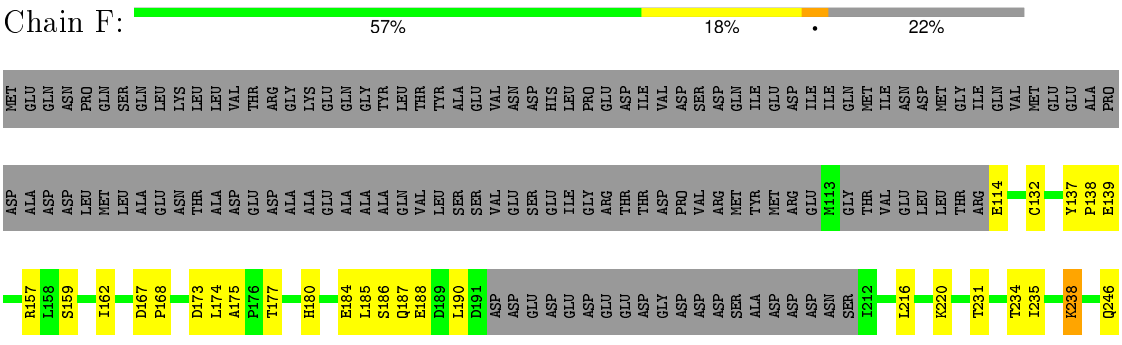
WORLDWIDE PDB PROTEIN DATA BANK
EMDataBank Unified Data Resource for 3DEM

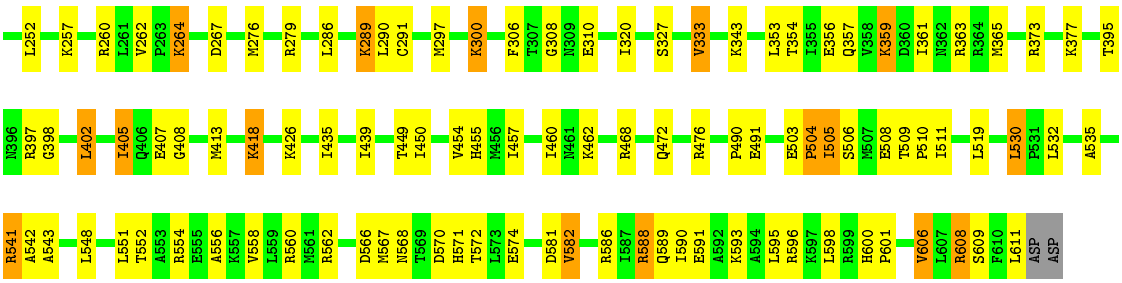


- Molecule 4: DNA-directed RNA polymerase subunit omega



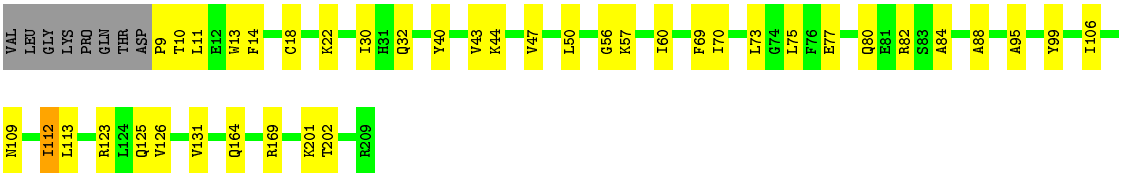
- Molecule 5: RNA polymerase sigma factor rpoD





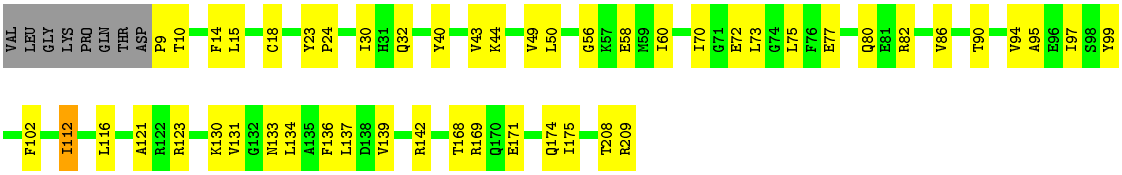
- Molecule 6: Catabolite gene activator

Chain G:



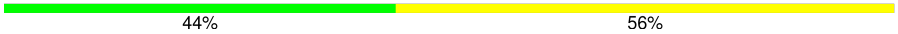
- Molecule 6: Catabolite gene activator

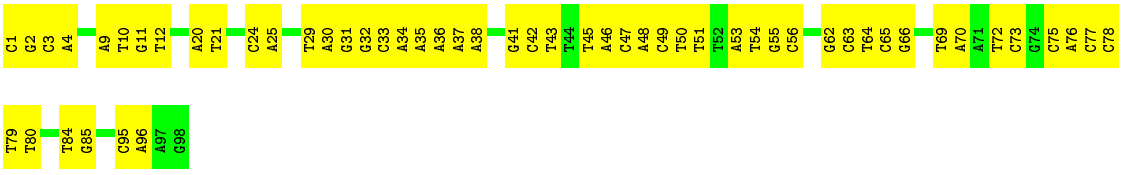
Chain H:



- Molecule 7: DNA (98-MER)

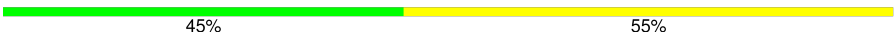
Chain I:

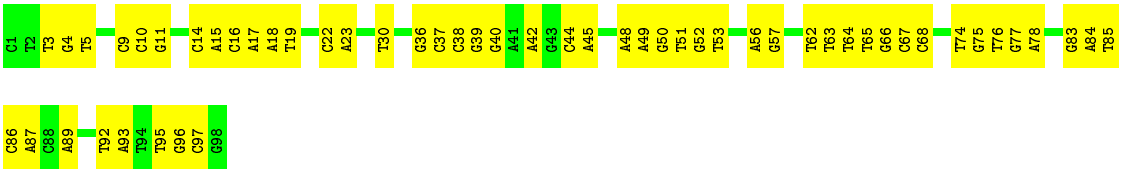




- Molecule 8: DNA (98-MER)

Chain J:





4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of images	14097	Depositor
Resolution determination method	FSC at 0.5 cut-off	Depositor
CTF correction method	ACE	Depositor
Microscope	FEI TECNAI F20	Depositor
Voltage (kV)	120	Depositor
Electron dose ($e^-/Å^2$)	16	Depositor
Minimum defocus (nm)	500	Depositor
Maximum defocus (nm)	1500	Depositor
Magnification	50000	Depositor
Image detector	Teitz F415 4k x 4k pixel CCD camera	Depositor

5 Model quality

5.1 Standard geometry

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

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	0.19	0/2528	0.36	0/3427
1	B	0.19	0/1836	0.35	0/2488
2	C	0.20	0/8672	0.35	0/11722
3	D	0.19	0/10769	0.35	0/14544
4	E	0.19	0/711	0.34	0/956
5	F	0.20	0/3931	0.34	0/5288
6	G	0.20	0/1616	0.33	0/2174
6	H	0.20	0/1616	0.33	0/2174
7	I	0.45	0/2251	0.93	0/3472
8	J	0.44	0/2236	0.93	0/3447
All	All	0.24	0/36166	0.47	0/49692

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	2496	0	2554	74	0
1	B	1814	0	1839	60	0
2	C	8524	0	8491	275	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	D	10606	0	10822	387	0
4	E	709	0	719	37	0
5	F	3877	0	3929	89	0
6	G	1591	0	1632	28	0
6	H	1591	0	1632	34	0
7	I	2006	0	1108	45	0
8	J	1996	0	1111	45	0
9	G	22	0	11	3	0
9	H	22	0	11	2	0
All	All	35254	0	33859	996	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 14.

The worst 5 of 996 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:G:679:CMP:H2	9:G:679:CMP:C2	0.97	1.49
9:H:680:CMP:C2	9:H:680:CMP:H2	0.97	1.49
4:E:59:ILE:HG12	4:E:60:ASN:H	1.45	0.81
3:D:1378:ALA:H	3:D:1379:PRO:HD3	1.45	0.81
1:A:235:ARG:HD2	1:A:235:ARG:H	1.45	0.80

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	318/329 (97%)	259 (81%)	46 (14%)	13 (4%)	3 35
1	B	233/329 (71%)	191 (82%)	38 (16%)	4 (2%)	11 55

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
2	C	1085/1342 (81%)	822 (76%)	217 (20%)	46 (4%)	3 34
3	D	1366/1413 (97%)	1075 (79%)	232 (17%)	59 (4%)	3 34
4	E	88/90 (98%)	68 (77%)	14 (16%)	6 (7%)	1 23
5	F	475/613 (78%)	401 (84%)	54 (11%)	20 (4%)	3 34
6	G	199/209 (95%)	172 (86%)	24 (12%)	3 (2%)	13 57
6	H	199/209 (95%)	182 (92%)	16 (8%)	1 (0%)	34 77
All	All	3963/4534 (87%)	3170 (80%)	641 (16%)	152 (4%)	7 37

5 of 152 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	52	PRO
1	A	240	PRO
1	B	52	PRO
2	C	180	ARG
2	C	332	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	278/286 (97%)	268 (96%)	10 (4%)	42 74
1	B	201/286 (70%)	193 (96%)	8 (4%)	38 71
2	C	936/1157 (81%)	896 (96%)	40 (4%)	35 70
3	D	1138/1174 (97%)	1075 (94%)	63 (6%)	27 63
4	E	74/74 (100%)	67 (90%)	7 (10%)	11 41
5	F	423/540 (78%)	407 (96%)	16 (4%)	40 73
6	G	173/180 (96%)	172 (99%)	1 (1%)	90 95
6	H	173/180 (96%)	172 (99%)	1 (1%)	90 95
All	All	3396/3877 (88%)	3250 (96%)	146 (4%)	40 70

5 of 146 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	151	MET
3	D	549	LYS
5	F	359	LYS
3	D	156	ARG
3	D	334	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 115 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	229	GLN
3	D	495	ASN
5	F	579	GLN
3	D	274	ASN
3	D	341	ASN

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)

2 ligands 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	$\#\|Z\|>2$
9	CMP	G	679	-	20,25,25	1.18	1 (5%)	18,39,39	2.19	3 (16%)
9	CMP	H	680	-	20,25,25	1.18	2 (10%)	18,39,39	2.21	3 (16%)

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
9	CMP	G	679	-	-	0/0/31/31	0/4/4/4
9	CMP	H	680	-	-	0/0/31/31	0/4/4/4

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
9	H	680	CMP	O4'-C1'	2.01	1.44	1.41
9	G	679	CMP	C5-C4	3.11	1.47	1.40
9	H	680	CMP	C5-C4	3.12	1.47	1.40

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
9	H	680	CMP	N3-C2-N1	-6.62	123.67	128.87
9	G	679	CMP	N3-C2-N1	-6.51	123.76	128.87
9	H	680	CMP	O3'-C3'-C4'	-4.81	106.88	110.72
9	G	679	CMP	O3'-C3'-C4'	-4.80	106.88	110.72
9	H	680	CMP	C4'-O4'-C1'	-2.03	107.49	109.64

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
9	G	679	CMP	3	0
9	H	680	CMP	2	0

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.