wwPDB X-ray Structure Validation Summary Report (i)

May 14, 2020 – 06:28 pm BST

PDB ID	:	5OLA
Title	:	Structure of mitochondrial transcription elongation complex in complex with elongation factor TEFM
Authors	:	Hillen, H.S.; Parshin, A.V.; Agaronyan, K.; Morozov, Y.; Graber, J.J.; Chernev, A.; Schwinghammer, K.; Urlaub, H.; Anikin, M.; Cramer, P.; Temiakov, D.
Deposited on	:	2017-07-27
Resolution	:	3.90 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

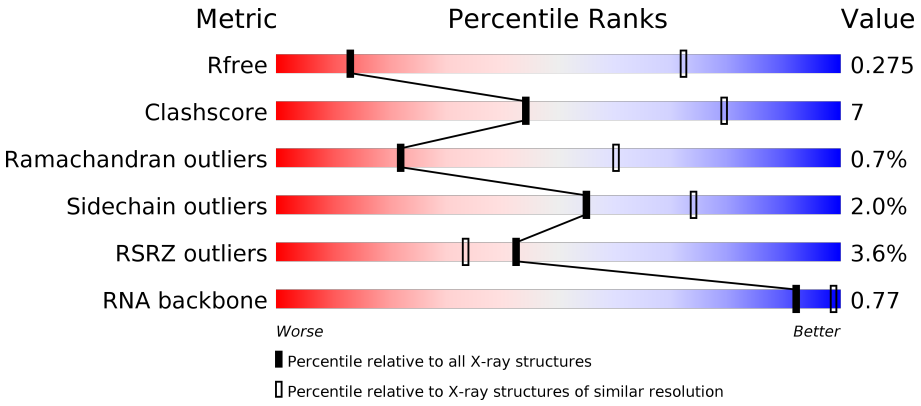
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.90 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.



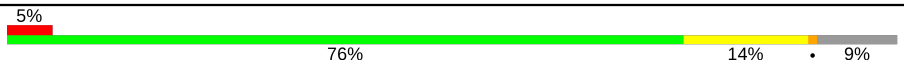
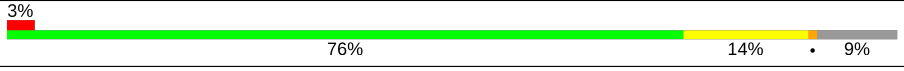
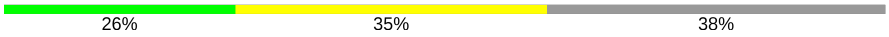
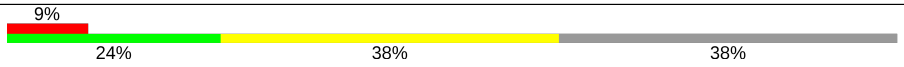
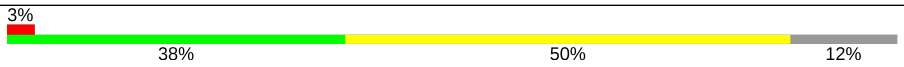
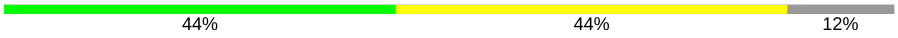
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1002 (4.14-3.66)
Clashscore	141614	1004 (4.12-3.68)
Ramachandran outliers	138981	1021 (4.14-3.66)
Sidechain outliers	138945	1014 (4.14-3.66)
RSRZ outliers	127900	1275 (4.20-3.60)
RNA backbone	3102	1040 (4.76-3.00)

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 <=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	234	76% 14% • 9%
1	B	234	74% 14% 12%
1	C	234	78% 10% • 9%
1	D	234	74% 14% 12%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	E	1088	
2	F	1088	
3	G	34	
3	N	34	
4	H	14	
4	R	14	
5	I	34	
5	T	34	

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 25122 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 Transcription elongation factor, mitochondrial.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	212	Total	C	N	O	S	0	0	0
			1746	1130	298	311	7			
1	B	207	Total	C	N	O	S	0	0	0
			1700	1101	289	303	7			
1	C	212	Total	C	N	O	S	0	0	0
			1746	1130	298	311	7			
1	D	207	Total	C	N	O	S	0	0	0
			1700	1101	289	303	7			

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	135	MET	-	initiating methionine	UNP Q96QE5
A	361	LEU	-	expression tag	UNP Q96QE5
A	362	GLU	-	expression tag	UNP Q96QE5
A	363	HIS	-	expression tag	UNP Q96QE5
A	364	HIS	-	expression tag	UNP Q96QE5
A	365	HIS	-	expression tag	UNP Q96QE5
A	366	HIS	-	expression tag	UNP Q96QE5
A	367	HIS	-	expression tag	UNP Q96QE5
A	368	HIS	-	expression tag	UNP Q96QE5
B	135	MET	-	initiating methionine	UNP Q96QE5
B	361	LEU	-	expression tag	UNP Q96QE5
B	362	GLU	-	expression tag	UNP Q96QE5
B	363	HIS	-	expression tag	UNP Q96QE5
B	364	HIS	-	expression tag	UNP Q96QE5
B	365	HIS	-	expression tag	UNP Q96QE5
B	366	HIS	-	expression tag	UNP Q96QE5
B	367	HIS	-	expression tag	UNP Q96QE5
B	368	HIS	-	expression tag	UNP Q96QE5
C	135	MET	-	initiating methionine	UNP Q96QE5
C	361	LEU	-	expression tag	UNP Q96QE5
C	362	GLU	-	expression tag	UNP Q96QE5

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	363	HIS	-	expression tag	UNP Q96QE5
C	364	HIS	-	expression tag	UNP Q96QE5
C	365	HIS	-	expression tag	UNP Q96QE5
C	366	HIS	-	expression tag	UNP Q96QE5
C	367	HIS	-	expression tag	UNP Q96QE5
C	368	HIS	-	expression tag	UNP Q96QE5
D	135	MET	-	initiating methionine	UNP Q96QE5
D	361	LEU	-	expression tag	UNP Q96QE5
D	362	GLU	-	expression tag	UNP Q96QE5
D	363	HIS	-	expression tag	UNP Q96QE5
D	364	HIS	-	expression tag	UNP Q96QE5
D	365	HIS	-	expression tag	UNP Q96QE5
D	366	HIS	-	expression tag	UNP Q96QE5
D	367	HIS	-	expression tag	UNP Q96QE5
D	368	HIS	-	expression tag	UNP Q96QE5

- Molecule 2 is a protein called DNA-directed RNA polymerase, mitochondrial.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	989	Total	C	N	O	S	0	0	0
			7876	5015	1423	1390	48			
2	F	989	Total	C	N	O	S	0	0	0
			7876	5015	1423	1390	48			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	143	MET	-	initiating methionine	UNP O00411
E	144	GLY	-	expression tag	UNP O00411
E	145	HIS	-	expression tag	UNP O00411
E	146	HIS	-	expression tag	UNP O00411
E	147	HIS	-	expression tag	UNP O00411
E	148	HIS	-	expression tag	UNP O00411
E	149	HIS	-	expression tag	UNP O00411
E	150	HIS	-	expression tag	UNP O00411
E	555	ALA	GLU	conflict	UNP O00411
F	143	MET	-	initiating methionine	UNP O00411
F	144	GLY	-	expression tag	UNP O00411
F	145	HIS	-	expression tag	UNP O00411
F	146	HIS	-	expression tag	UNP O00411
F	147	HIS	-	expression tag	UNP O00411
F	148	HIS	-	expression tag	UNP O00411

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	149	HIS	-	expression tag	UNP O00411
F	150	HIS	-	expression tag	UNP O00411
F	555	ALA	GLU	conflict	UNP O00411

- Molecule 3 is a DNA chain called DNA (5'-D(P*AP*TP*GP*GP*TP*GP*TP*AP*AP*CP *GP*CP*CP*AP*GP*AP*CP*GP*AP*AP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	N	21	Total	C	N	O	P	0	0	0
			434	205	86	122	21			
3	G	21	Total	C	N	O	P	0	0	0
			434	205	86	122	21			

- Molecule 4 is a RNA chain called RNA (5'-R(P*GP*CP*GP*GP*CP*GP*CP*GP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	R	9	Total	C	N	O	P	0	0	0
			195	86	37	63	9			
4	H	9	Total	C	N	O	P	0	0	0
			195	86	37	63	9			

- Molecule 5 is a DNA chain called DNA (30-MER).

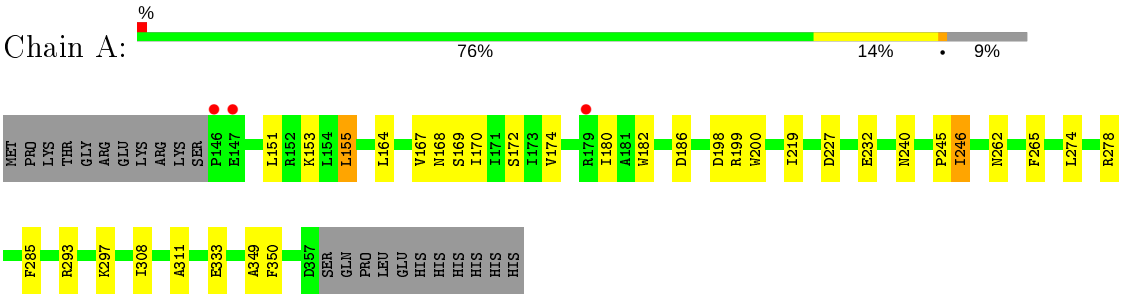
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	T	30	Total	C	N	O	P	0	0	0
			610	289	107	184	30			
5	I	30	Total	C	N	O	P	0	0	0
			610	289	107	184	30			

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.

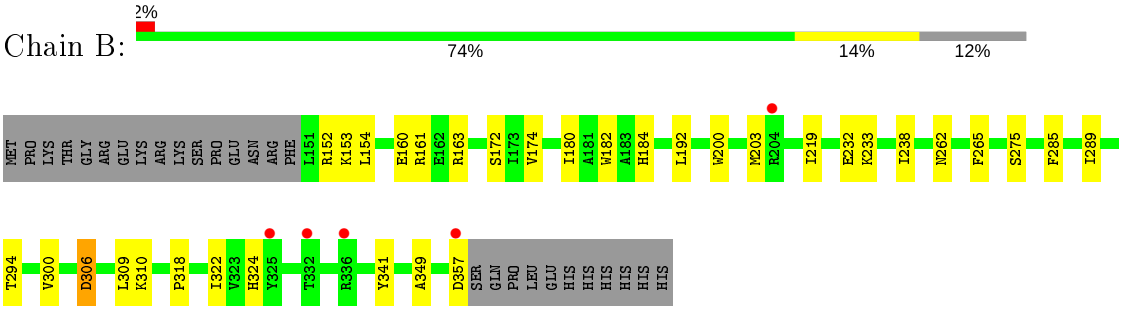
- Molecule 1: Transcription elongation factor, mitochondrial

Chain A:



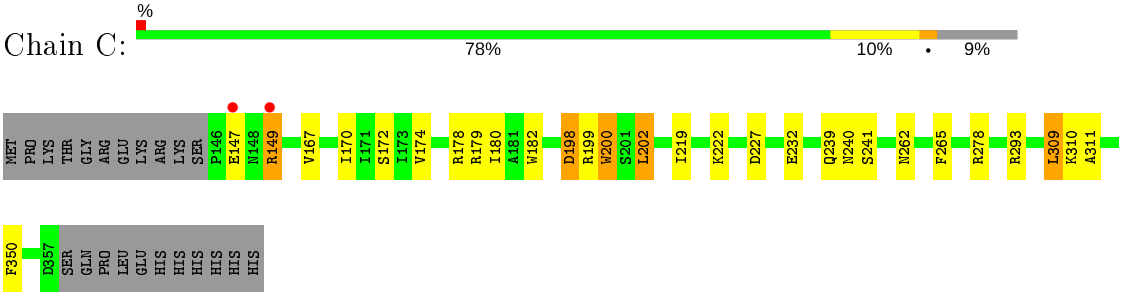
- Molecule 1: Transcription elongation factor, mitochondrial

Chain B:



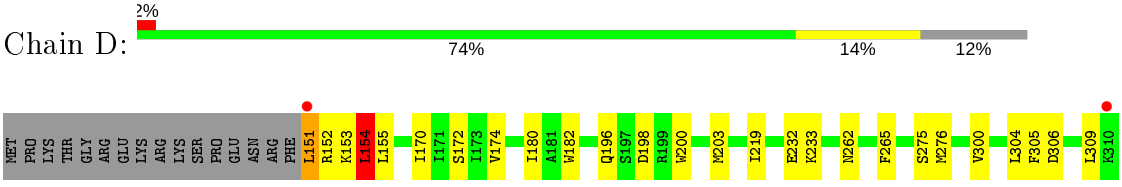
- Molecule 1: Transcription elongation factor, mitochondrial

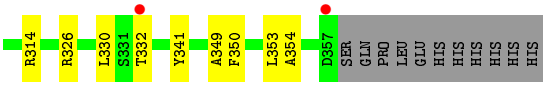
Chain C:



- Molecule 1: Transcription elongation factor, mitochondrial

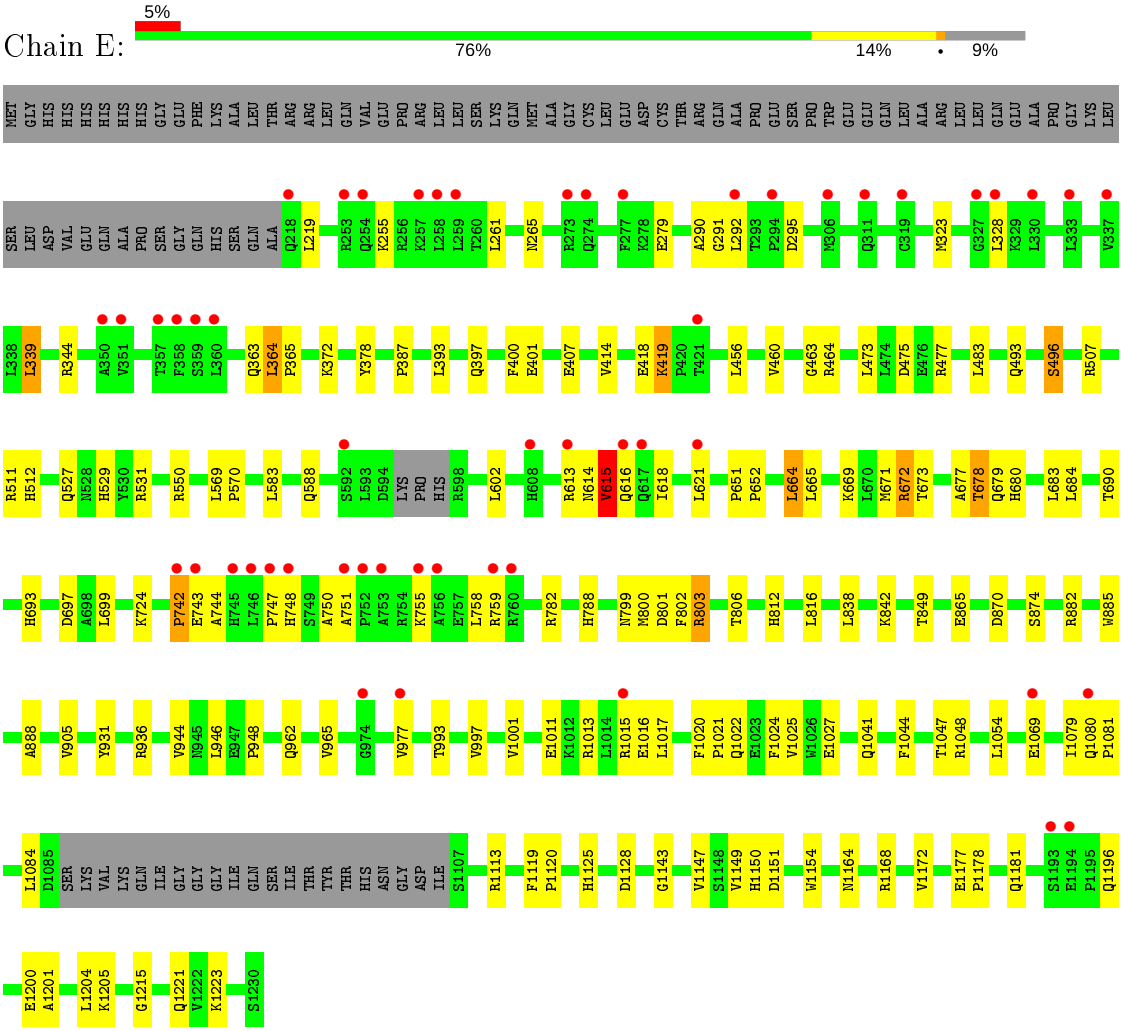
Chain D:





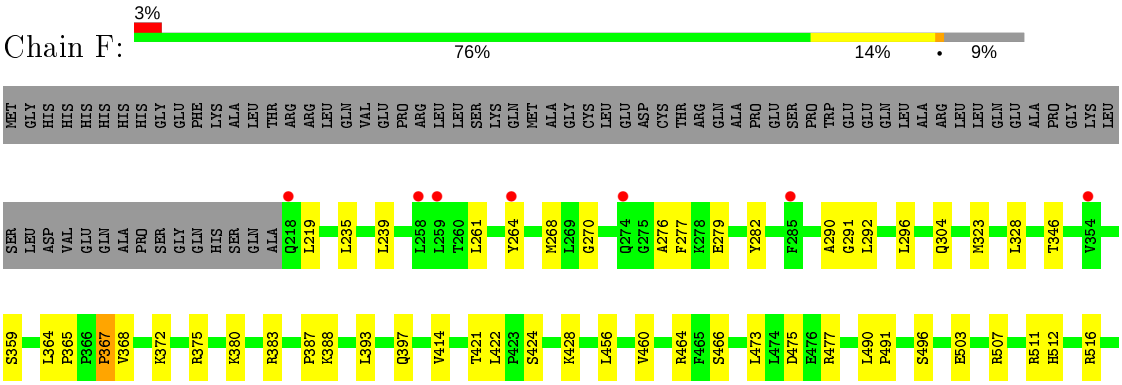
- Molecule 2: DNA-directed RNA polymerase, mitochondrial

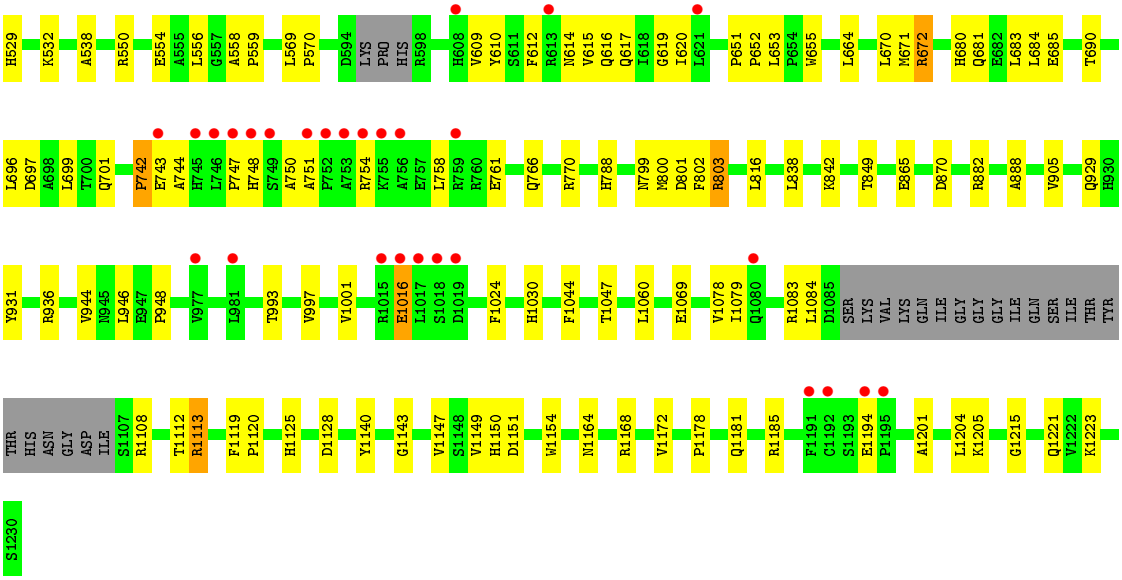
Chain E:



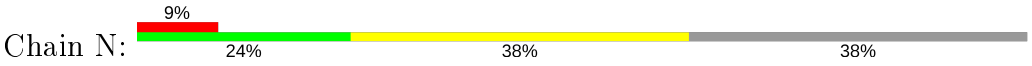
- Molecule 2: DNA-directed RNA polymerase, mitochondrial

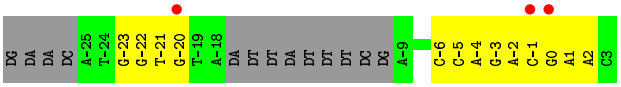
Chain F:



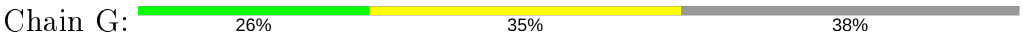


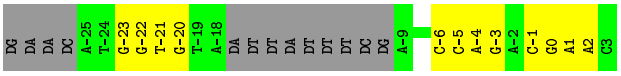
• Molecule 3: DNA (5'-D(P*AP*TP*GP*GP*TP*GP*TP*AP*AP*CP*GP*CP*CP*AP*GP*AP*CP*GP*AP*AP*C)-3')





• Molecule 3: DNA (5'-D(P*AP*TP*GP*GP*TP*GP*TP*AP*AP*CP*GP*CP*CP*AP*GP*AP*CP*GP*AP*AP*C)-3')



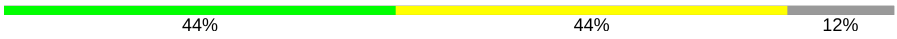


• Molecule 4: RNA (5'-R(P*GP*CP*GP*GP*CP*GP*CP*GP*C)-3')

• Molecule 4: RNA (5'-R(P*GP*CP*GP*GP*CP*GP*CP*GP*C)-3')

• Molecule 5: DNA (30-MER)

Chain T:



- Molecule 5: DNA (30-MER)

Chain I:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	224.53Å 155.55Å 164.19Å 90.00° 113.58° 90.00°	Depositor
Resolution (Å)	49.15 – 3.90 49.15 – 3.90	Depositor EDS
% Data completeness (in resolution range)	99.7 (49.15-3.90) 99.8 (49.15-3.90)	Depositor EDS
R_{merge}	0.26	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.42 (at 3.88Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, R_{free}	0.243 , 0.276 0.242 , 0.275	Depositor DCC
R_{free} test set	2347 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	141.8	Xtriage
Anisotropy	0.213	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 95.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	25122	wwPDB-VP
Average B, all atoms (Å^2)	160.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.17% 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

5.1 Standard geometry

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	0.28	0/1784	0.45	0/2403
1	B	0.28	0/1736	0.44	0/2339
1	C	0.30	0/1784	0.48	0/2403
1	D	0.30	0/1736	0.50	1/2339 (0.0%)
2	E	0.28	0/8067	0.42	0/10947
2	F	0.28	0/8067	0.43	0/10947
3	G	0.56	0/487	0.85	0/747
3	N	0.56	0/487	0.87	0/747
4	H	0.39	0/217	1.03	0/337
4	R	0.41	0/217	1.11	2/337 (0.6%)
5	I	0.70	0/681	0.97	0/1048
5	T	0.64	0/681	0.98	0/1048
All	All	0.33	0/25944	0.53	3/35642 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	154	LEU	CA-CB-CG	7.31	132.12	115.30
4	R	2	C	C6-N1-C2	-5.44	118.12	120.30
4	R	2	C	C5-C6-N1	5.41	123.70	121.00

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	1746	0	1774	19	0
1	B	1700	0	1732	17	0
1	C	1746	0	1774	20	0
1	D	1700	0	1732	20	0
2	E	7876	0	7965	100	0
2	F	7876	0	7965	100	0
3	G	434	0	236	15	0
3	N	434	0	236	13	0
4	H	195	0	100	3	0
4	R	195	0	100	5	0
5	I	610	0	338	16	0
5	T	610	0	338	15	0
All	All	25122	0	24290	324	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 7.

The worst 5 of 324 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:F:235:LEU:HD22	2:F:516:ARG:HH12	1.43	0.83
1:D:152:ARG:NH2	3:G:1:DA:OP2	2.13	0.82
2:E:1178:PRO:HB2	2:E:1181:GLN:HB2	1.67	0.77
2:F:842:LYS:NZ	2:F:865:GLU:OE2	2.16	0.76
1:D:153:LYS:NZ	5:I:-5:DT:OP1	2.18	0.76

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	210/234 (90%)	196 (93%)	12 (6%)	2 (1%)	15 52

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	205/234 (88%)	192 (94%)	13 (6%)	0	100	100
1	C	210/234 (90%)	197 (94%)	11 (5%)	2 (1%)	15	52
1	D	205/234 (88%)	190 (93%)	13 (6%)	2 (1%)	15	52
2	E	983/1088 (90%)	931 (95%)	43 (4%)	9 (1%)	17	54
2	F	983/1088 (90%)	929 (94%)	49 (5%)	5 (0%)	29	67
All	All	2796/3112 (90%)	2635 (94%)	141 (5%)	20 (1%)	22	60

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	308	ILE
1	C	240	ASN
1	C	310	LYS
1	D	314	ARG
2	E	419	LYS

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	193/214 (90%)	187 (97%)	6 (3%)	40	64
1	B	188/214 (88%)	182 (97%)	6 (3%)	39	63
1	C	193/214 (90%)	187 (97%)	6 (3%)	40	64
1	D	188/214 (88%)	184 (98%)	4 (2%)	53	73
2	E	856/942 (91%)	841 (98%)	15 (2%)	59	77
2	F	856/942 (91%)	844 (99%)	12 (1%)	67	81
All	All	2474/2740 (90%)	2425 (98%)	49 (2%)	55	74

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	E	339	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	E	618	ILE
2	F	1016	GLU
2	E	496	SER
2	E	664	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
2	E	614	ASN
2	E	1030	HIS
2	F	1041	GLN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
4	H	9/14 (64%)	1 (11%)	1 (11%)
4	R	9/14 (64%)	1 (11%)	1 (11%)
All	All	18/28 (64%)	2 (11%)	2 (11%)

All (2) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
4	R	2	C
4	H	2	C

All (2) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
4	R	1	G
4	H	1	G

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

There are no ligands in this entry.

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.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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, 95^{th} 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	212/234 (90%)	0.00	3 (1%) 75 66	115, 138, 177, 248	0
1	B	207/234 (88%)	0.03	5 (2%) 59 48	116, 140, 178, 204	0
1	C	212/234 (90%)	0.13	2 (0%) 84 77	99, 125, 167, 212	0
1	D	207/234 (88%)	0.01	4 (1%) 66 57	99, 132, 190, 223	0
2	E	989/1088 (90%)	0.22	52 (5%) 26 22	107, 153, 245, 283	0
2	F	989/1088 (90%)	0.16	35 (3%) 44 34	109, 149, 208, 286	0
3	G	21/34 (61%)	0.50	0 100 100	206, 258, 301, 303	0
3	N	21/34 (61%)	0.87	3 (14%) 2 2	213, 272, 299, 306	0
4	H	9/14 (64%)	0.01	0 100 100	149, 159, 199, 201	0
4	R	9/14 (64%)	0.39	0 100 100	146, 152, 190, 195	0
5	I	30/34 (88%)	0.09	1 (3%) 46 36	151, 219, 309, 317	0
5	T	30/34 (88%)	0.28	0 100 100	155, 227, 303, 311	0
All	All	2936/3276 (89%)	0.16	105 (3%) 42 33	99, 146, 238, 317	0

The worst 5 of 105 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	E	258	LEU	6.1
2	E	756	ALA	4.7
2	F	756	ALA	4.5
2	E	311	GLN	4.5
2	E	752	PRO	4.4

6.2 Non-standard residues in protein, DNA, RNA chains

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.