wwPDB X-ray Structure Validation Summary Report (i)

Jun 23, 2024 – 01:13 AM EDT

PDB ID	:	6ERQ
Title	:	Structure of the human mitochondrial transcription initiation complex at the HSP promoter
Authors	:	Hillen, H.S.; Morozov, Y.I.; Sarfallah, A.; Temiakov, D.; Cramer, P.
Deposited on	:	2017-10-18
Resolution	:	4.50 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

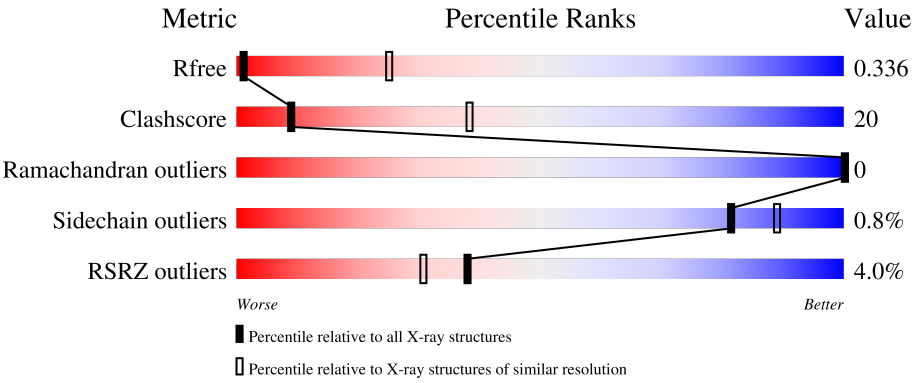
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 4.50 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)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1055 (5.20-3.80)
Clashscore	141614	1123 (5.20-3.80)
Ramachandran outliers	138981	1069 (5.20-3.80)
Sidechain outliers	138945	1050 (5.20-3.80)
RSRZ outliers	127900	1101 (5.30-3.70)

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 of 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	C	205	9% 67% 26% 6%
1	G	205	5% 59% 35% 6%
2	D	50	4% 16% 76% 8%
2	H	50	8% 30% 62% 8%
3	E	50	4% 18% 68% . 10%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	I	50	2% 24% 62% 10%
4	A	1128	4% 59% 30% 10%
4	B	1128	3% 57% 31% 11%
5	F	377	3% 52% 24% 24%
5	J	377	% 49% 27% 24%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 27643 atoms, of which 42 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 factor A, mitochondrial.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	192	Total	C	N	O	S	0	0	0
			1615	1019	291	300	5			
1	G	192	Total	C	N	O	S	0	0	0
			1615	1019	291	300	5			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	41	ASN	-	expression tag	UNP Q00059
C	42	ALA	-	expression tag	UNP Q00059
C	49	SER	CYS	conflict	UNP Q00059
G	41	ASN	-	expression tag	UNP Q00059
G	42	ALA	-	expression tag	UNP Q00059
G	49	SER	CYS	conflict	UNP Q00059

- Molecule 2 is a DNA chain called Non-Template DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	46	Total	C	N	O	P	0	0	0
			924	440	175	264	45			
2	H	46	Total	C	N	O	P	0	0	0
			924	440	175	264	45			

- Molecule 3 is a DNA chain called Template DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	45	Total	C	N	O	P	0	0	0
			938	444	171	279	44			
3	I	45	Total	C	N	O	P	0	0	0
			938	444	171	279	44			

- Molecule 4 is a protein called DNA-directed RNA polymerase, mitochondrial.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	A	1011	Total	C	N	O	S	0	0	0
			8028	5107	1452	1418	51			
4	B	1002	Total	C	N	O	S	0	0	0
			7967	5069	1443	1404	51			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	103	ASN	-	expression tag	UNP O00411
A	104	ALA	-	expression tag	UNP O00411
A	555	ALA	GLU	conflict	UNP O00411
B	103	ASN	-	expression tag	UNP O00411
B	104	ALA	-	expression tag	UNP O00411
B	555	ALA	GLU	conflict	UNP O00411

- Molecule 5 is a protein called Dimethyladenosine transferase 2, mitochondrial.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
5	F	288	Total	C	H	N	O	S	0	0	0
			2347	1506	21	397	408	15			
5	J	288	Total	C	H	N	O	S	0	0	0
			2347	1506	21	397	408	15			

There are 4 discrepancies between the modelled and reference sequences:

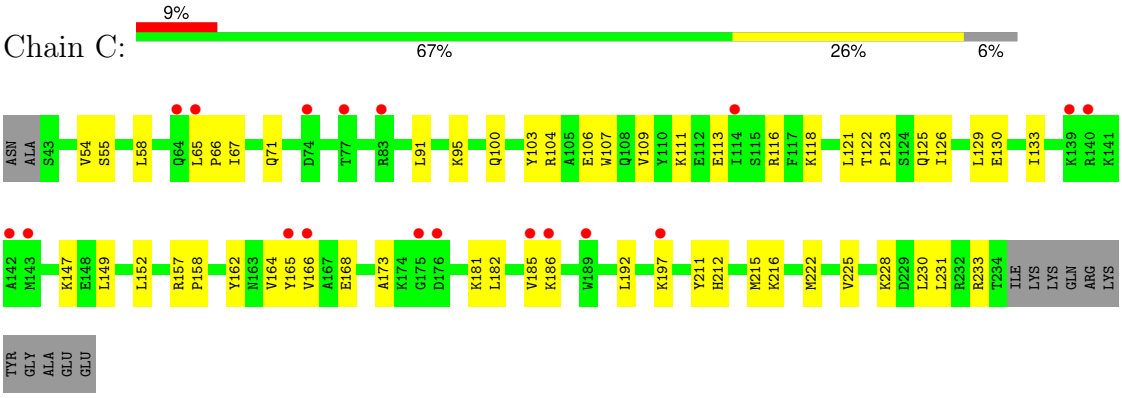
Chain	Residue	Modelled	Actual	Comment	Reference
F	20	ASN	-	expression tag	UNP Q9H5Q4
F	21	ALA	-	expression tag	UNP Q9H5Q4
J	20	ASN	-	expression tag	UNP Q9H5Q4
J	21	ALA	-	expression tag	UNP Q9H5Q4

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

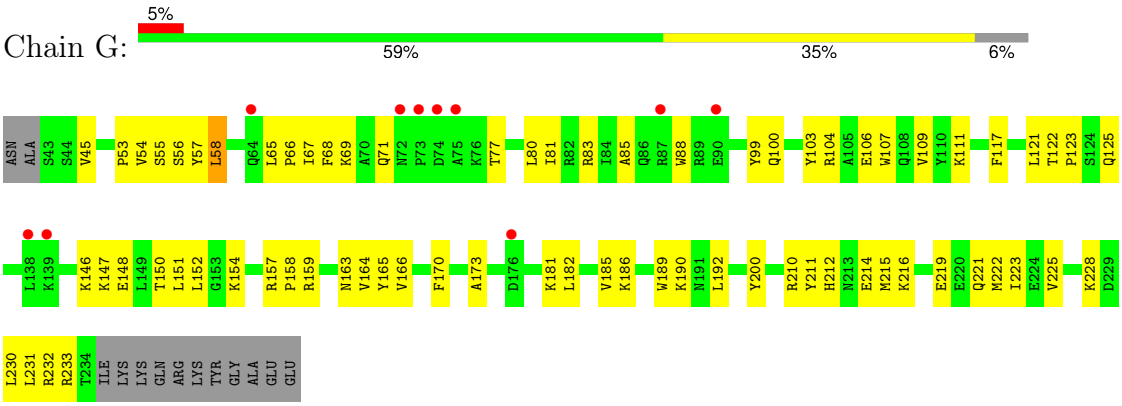
• Molecule 1: Transcription factor A, mitochondrial

Chain C:



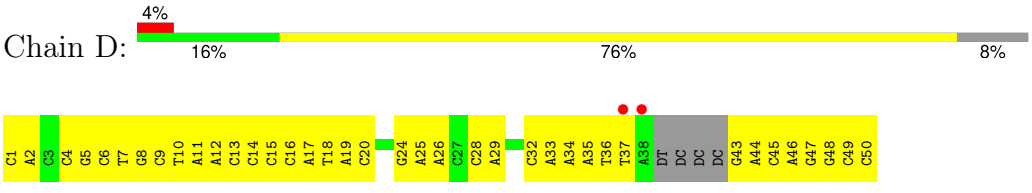
• Molecule 1: Transcription factor A, mitochondrial

Chain G:



• Molecule 2: Non-Template DNA

Chain D:

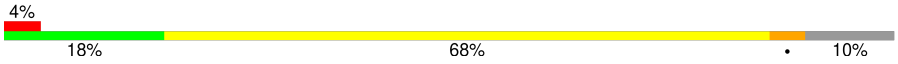


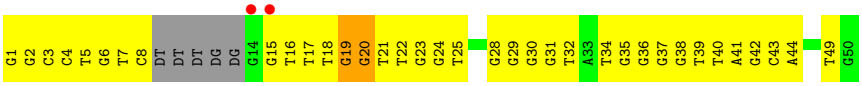
• Molecule 2: Non-Template DNA

Chain H:

• Molecule 3: Template DNA

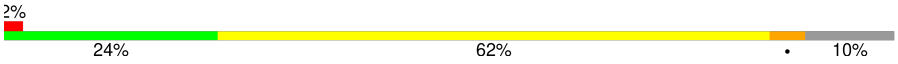
Chain E:





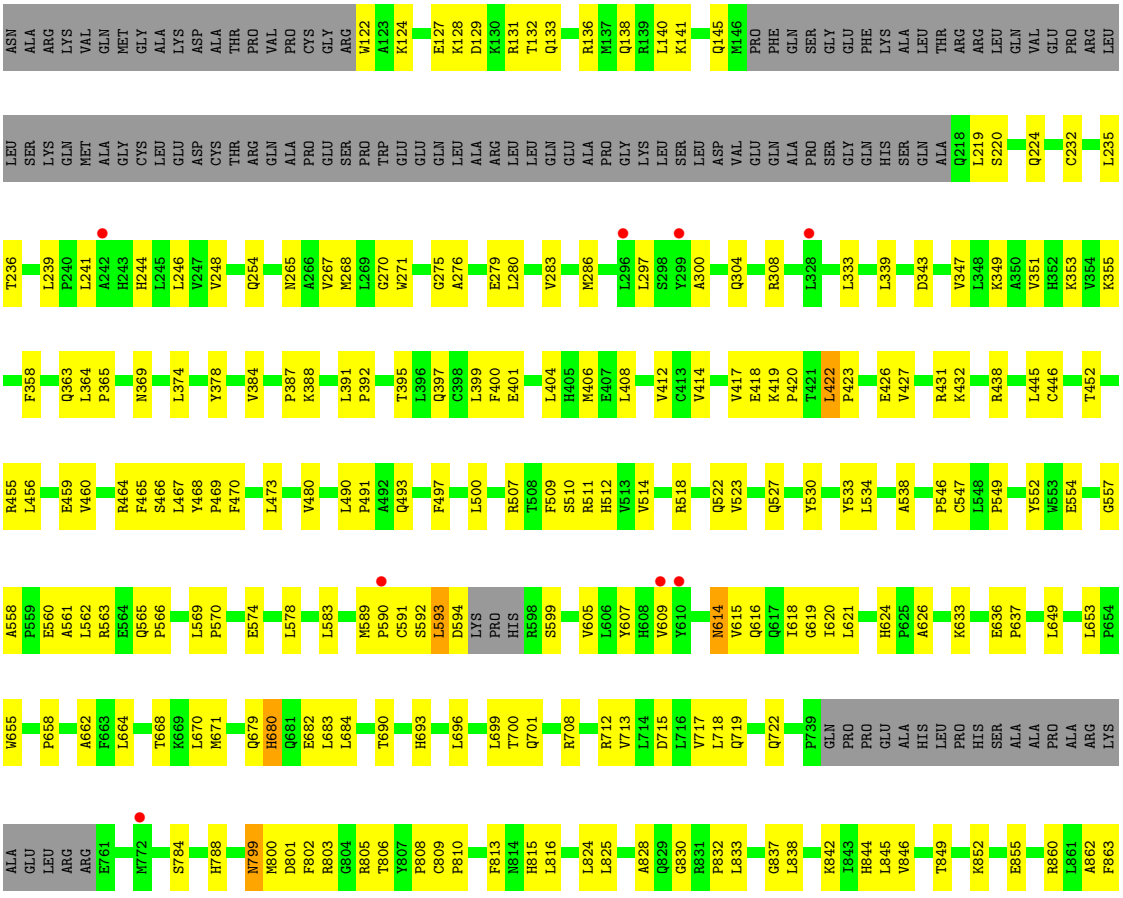
• Molecule 3: Template DNA

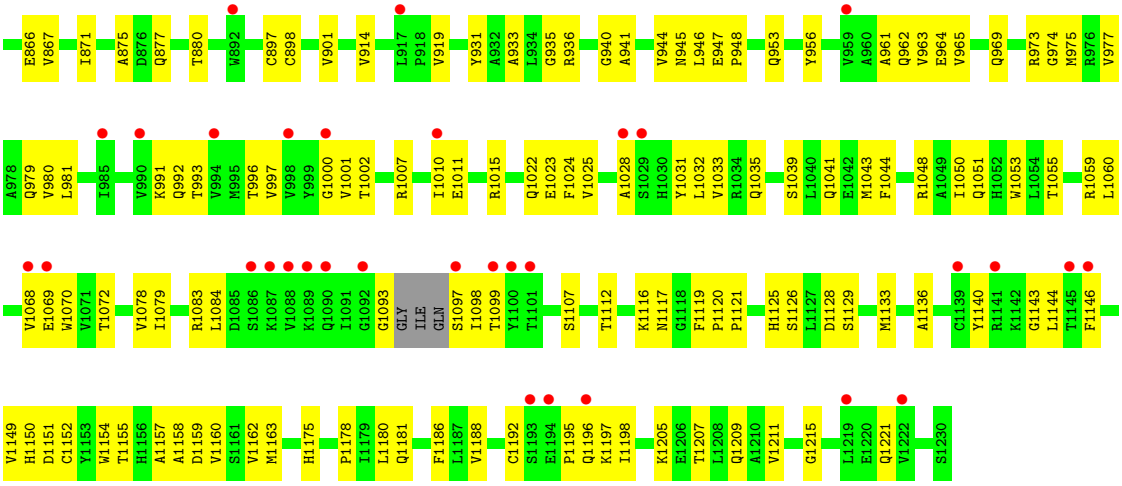
Chain I:



• Molecule 4: DNA-directed RNA polymerase, mitochondrial

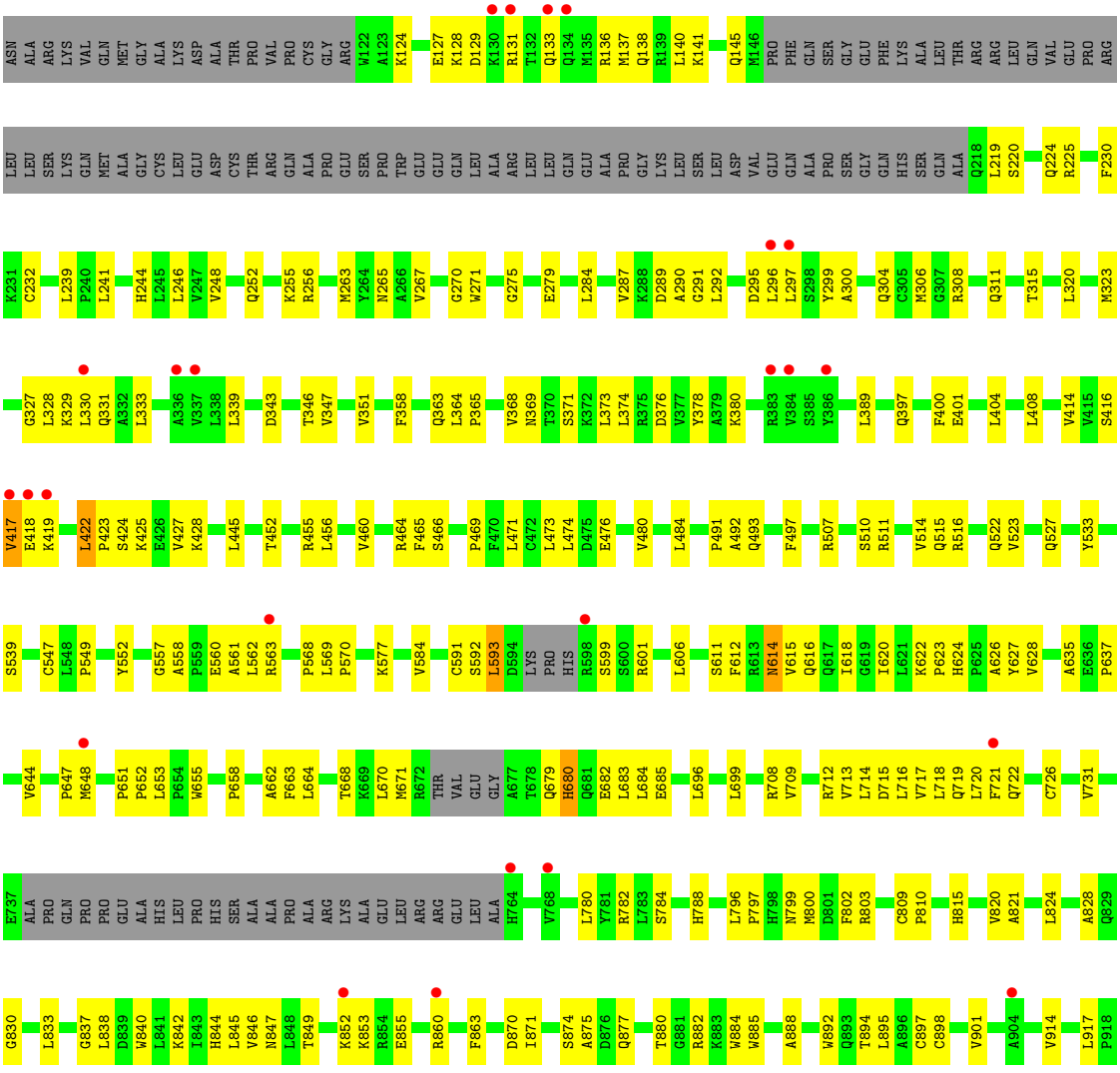
Chain A:

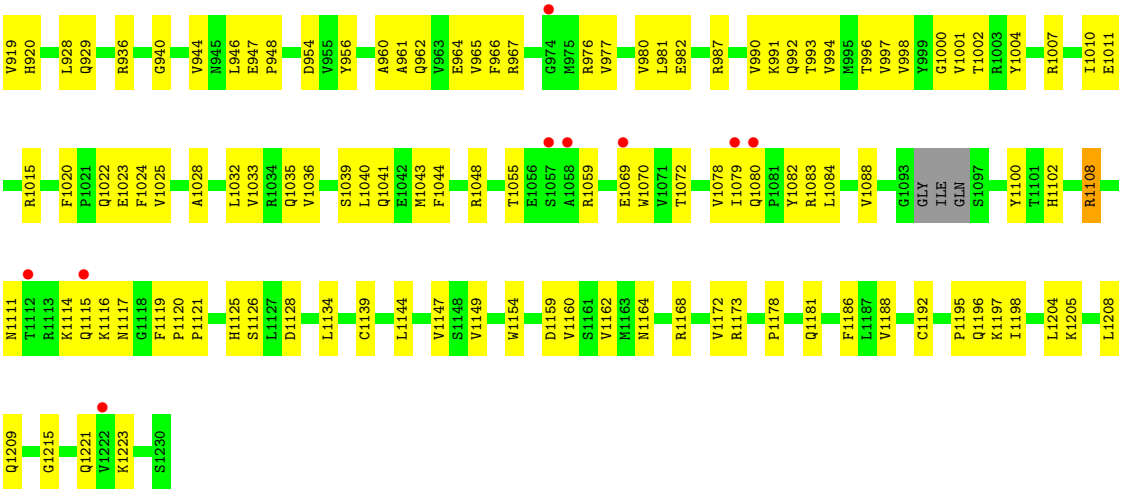




• Molecule 4: DNA-directed RNA polymerase, mitochondrial

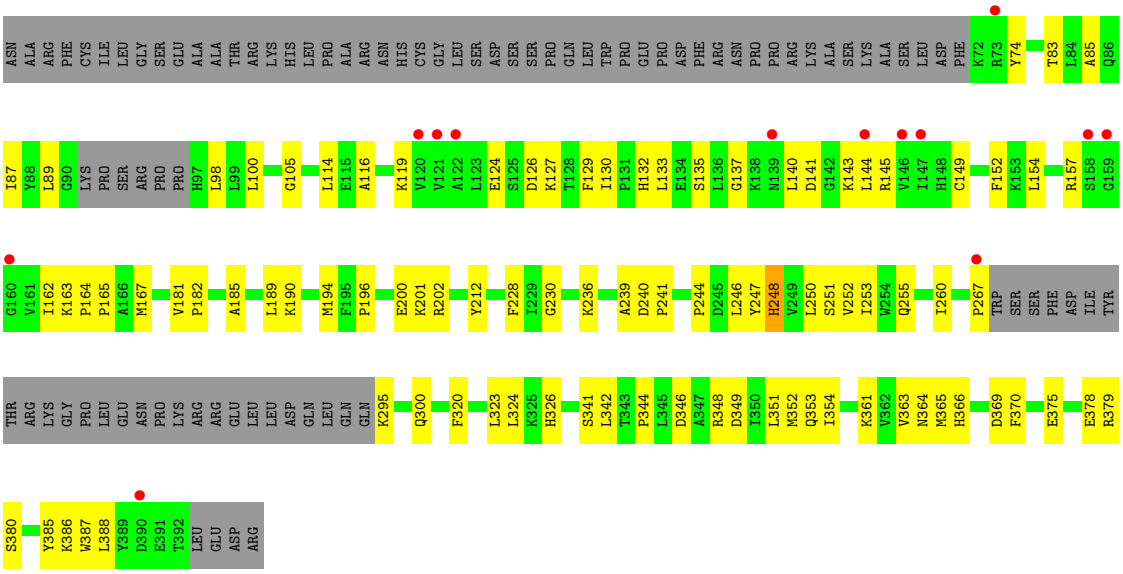
Chain B:





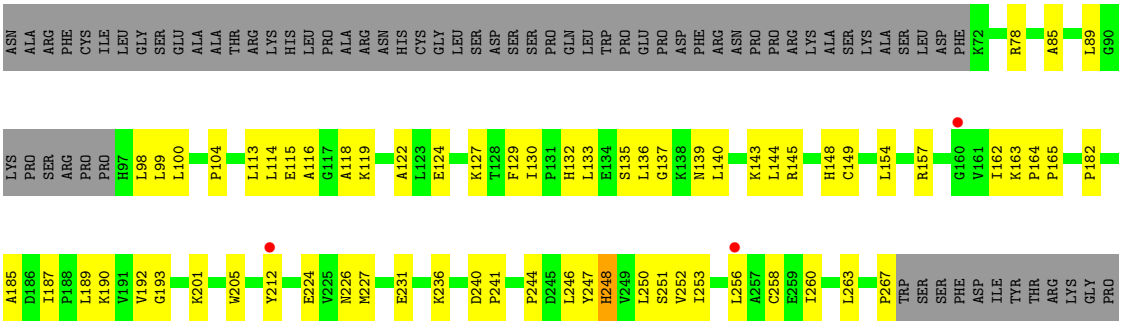
• Molecule 5: Dimethyladenosine transferase 2, mitochondrial

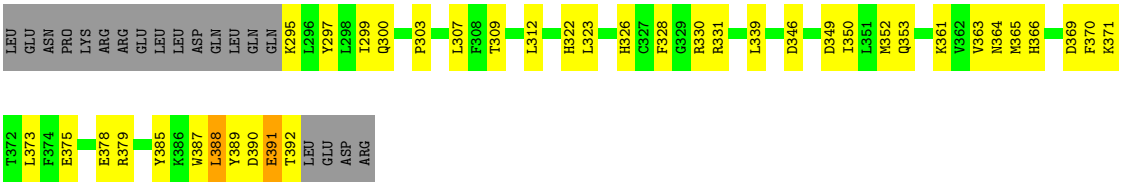
Chain F:



• Molecule 5: Dimethyladenosine transferase 2, mitochondrial

Chain J:




LEU GLU ASN PRO LYS ARG ARG GLU LEU LEU ASP GLN LEU GLN GLN
K295 L296 Y297 L298 I299 Q300 P303 L307 F308 T309 L312 H322 L323 H326 C327 F328 G329 R330 R331 L339 D346 D349 I350 L351 M352 Q353 K361 V362 V363 N364 M365 H366 D369 F370 K371
T372 L373 F374 E375 E378 R379 Y385 K386 W387 L388 Y389 D390 E391 T392 LEU GLU ASP ARG

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	103.62Å 197.73Å 134.34Å 90.00° 99.39° 90.00°	Depositor
Resolution (Å)	49.58 – 4.50 49.58 – 4.50	Depositor EDS
% Data completeness (in resolution range)	99.1 (49.58-4.50) 99.2 (49.58-4.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.43 (at 4.45Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, R_{free}	0.291 , 0.333 0.293 , 0.336	Depositor DCC
R_{free} test set	1573 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	248.7	Xtriage
Anisotropy	0.204	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 252.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	27643	wwPDB-VP
Average B, all atoms (Å^2)	333.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.71% 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	C	0.23	0/1646	0.35	0/2202
1	G	0.23	0/1646	0.35	0/2202
2	D	0.54	0/1035	0.87	0/1587
2	H	0.54	0/1035	0.88	0/1587
3	E	0.56	0/1051	1.07	4/1625 (0.2%)
3	I	0.57	0/1051	1.06	4/1625 (0.2%)
4	A	0.23	0/8213	0.38	0/11132
4	B	0.23	0/8150	0.38	0/11043
5	F	0.24	0/2380	0.38	0/3213
5	J	0.23	0/2380	0.38	0/3213
All	All	0.30	0/28587	0.52	8/39429 (0.0%)

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	E	19	DG	OP2-P-O3'	-11.21	80.54	105.20
3	I	19	DG	OP2-P-O3'	-11.08	80.81	105.20
3	E	20	DG	OP1-P-OP2	9.01	133.11	119.60
3	I	19	DG	OP1-P-O3'	-8.61	86.26	105.20
3	I	20	DG	OP1-P-OP2	8.61	132.51	119.60

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	C	1615	0	1655	47	0
1	G	1615	0	1655	83	1
2	D	924	0	514	54	1
2	H	924	0	514	50	0
3	E	938	0	512	66	0
3	I	938	0	512	66	1
4	A	8028	0	8092	310	1
4	B	7967	0	8032	323	0
5	F	2326	21	2375	81	0
5	J	2326	21	2375	94	0
All	All	27601	42	26236	1087	2

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 20.

The worst 5 of 1087 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)
4:A:363:GLN:HB3	4:A:364:LEU:HG	1.30	1.13
4:B:422:LEU:HD12	4:B:427:VAL:HG23	1.19	1.12
4:A:422:LEU:HD12	4:A:427:VAL:HG23	1.29	1.11
3:E:16:DT:H4'	4:A:618:ILE:HG22	1.31	1.11
4:B:363:GLN:HB3	4:B:364:LEU:HG	1.26	1.09

All (2) 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)
2:D:1:DC:O5'	1:G:233:ARG:O[2_654]	2.02	0.18
3:I:40:DT:OP1	4:A:432:LYS:NZ[2_644]	2.12	0.08

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	C	190/205 (93%)	185 (97%)	5 (3%)	0	100	100
1	G	190/205 (93%)	186 (98%)	4 (2%)	0	100	100
4	A	1001/1128 (89%)	962 (96%)	39 (4%)	0	100	100
4	B	990/1128 (88%)	949 (96%)	41 (4%)	0	100	100
5	F	282/377 (75%)	272 (96%)	10 (4%)	0	100	100
5	J	282/377 (75%)	272 (96%)	10 (4%)	0	100	100
All	All	2935/3420 (86%)	2826 (96%)	109 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	C	178/188 (95%)	178 (100%)	0	100	100
1	G	178/188 (95%)	177 (99%)	1 (1%)	86	92
4	A	865/976 (89%)	859 (99%)	6 (1%)	84	90
4	B	859/976 (88%)	852 (99%)	7 (1%)	81	89
5	F	255/335 (76%)	253 (99%)	2 (1%)	81	89
5	J	255/335 (76%)	251 (98%)	4 (2%)	62	79
All	All	2590/2998 (86%)	2570 (99%)	20 (1%)	81	89

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	F	157	ARG
5	J	248	HIS
5	J	391	GLU
5	J	388	LEU
4	A	799	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
4	A	979	GLN
4	B	624	HIS
4	B	1209	GLN
4	B	1111	ASN
4	A	815	HIS

5.3.3 RNA

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no monosaccharides 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	C	192/205 (93%)	0.35	18 (9%) 8 8	292, 364, 421, 433	0
1	G	192/205 (93%)	0.11	10 (5%) 27 24	335, 409, 458, 512	0
2	D	46/50 (92%)	-0.03	2 (4%) 35 29	344, 408, 562, 599	0
2	H	46/50 (92%)	0.31	4 (8%) 10 9	342, 419, 482, 500	0
3	E	45/50 (90%)	-0.02	2 (4%) 34 28	329, 373, 481, 547	0
3	I	45/50 (90%)	0.15	1 (2%) 62 52	336, 408, 504, 549	0
4	A	1011/1128 (89%)	0.18	40 (3%) 38 31	225, 300, 360, 402	0
4	B	1002/1128 (88%)	0.10	33 (3%) 46 37	245, 308, 366, 420	0
5	F	288/377 (76%)	0.03	13 (4%) 33 28	270, 328, 415, 455	0
5	J	288/377 (76%)	-0.06	3 (1%) 82 74	245, 321, 447, 510	0
All	All	3155/3620 (87%)	0.12	126 (3%) 38 31	225, 320, 426, 599	0

The worst 5 of 126 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	A	610	TYR	6.5
4	B	418	GLU	6.4
5	F	146	VAL	6.3
4	A	1090	GLN	5.6
4	A	1089	LYS	5.5

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 monosaccharides 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.