



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 06:44 PM GMT

PDB ID : 4MEY
Title : Crystal structure of Escherichia coli RNA polymerase holoenzyme
Authors : Feng, Y.; Zhang, Y.; Arnold, E.; Ebright, R.H.
Deposited on : 2013-08-27
Resolution : 3.95 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
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk26865

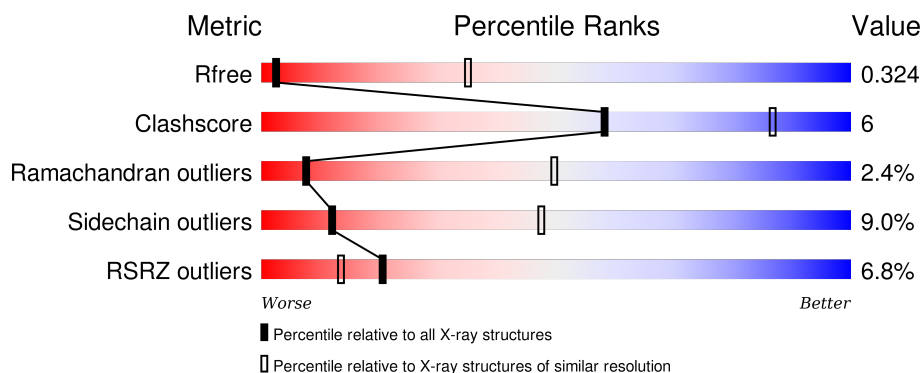
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.95 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}	91344	1007 (4.34-3.54)
Clashscore	102246	1042 (4.30-3.58)
Ramachandran outliers	100387	1000 (4.30-3.58)
Sidechain outliers	100360	1021 (4.32-3.56)
RSRZ outliers	91569	1011 (4.34-3.54)

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. 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 $\leq 5\%$. 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	335	<div> <div>7%</div> <div>69%</div> <div>17%</div> <div>•</div> <div>11%</div> </div>
1	B	335	<div> <div>12%</div> <div>65%</div> <div>17%</div> <div>•</div> <div>14%</div> </div>
1	G	335	<div> <div>3%</div> <div>52%</div> <div>11%</div> <div>•</div> <div>36%</div> </div>
1	H	335	<div> <div>10%</div> <div>49%</div> <div>13%</div> <div>•</div> <div>36%</div> </div>
2	C	1342	<div> <div>7%</div> <div>82%</div> <div>17%</div> <div>•</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	I	1342	<div><div></div><div>11%</div><div>82%</div><div>17%</div><div></div></div>
3	D	1407	<div><div></div><div>%</div><div>66%</div><div>14%</div><div>18%</div><div></div></div>
3	J	1407	<div><div></div><div>2%</div><div>66%</div><div>13%</div><div>19%</div><div></div></div>
4	E	91	<div><div></div><div></div><div>78%</div><div>16%</div><div></div></div>
4	K	91	<div><div></div><div>%</div><div>68%</div><div>12%</div><div>18%</div><div></div></div>
5	F	613	<div><div></div><div>8%</div><div>61%</div><div>14%</div><div>24%</div><div></div></div>
5	L	613	<div><div></div><div>7%</div><div>60%</div><div>14%</div><div>24%</div><div></div></div>

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 49826 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 DNA-directed RNA polymerase subunit alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	298	Total	C	N	O	S	0	0	0
			2236	1405	391	432	8			
1	B	287	Total	C	N	O	S	0	0	0
			2160	1359	374	419	8			
1	G	216	Total	C	N	O	S	0	0	0
			1618	1013	282	317	6			
1	H	215	Total	C	N	O	S	0	0	0
			1605	1005	278	316	6			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	HIS	-	EXPRESSION TAG	UNP P0A7Z4
A	-3	HIS	-	EXPRESSION TAG	UNP P0A7Z4
A	-2	HIS	-	EXPRESSION TAG	UNP P0A7Z4
A	-1	HIS	-	EXPRESSION TAG	UNP P0A7Z4
A	0	HIS	-	EXPRESSION TAG	UNP P0A7Z4
A	1	HIS	-	EXPRESSION TAG	UNP P0A7Z4
B	-4	HIS	-	EXPRESSION TAG	UNP P0A7Z4
B	-3	HIS	-	EXPRESSION TAG	UNP P0A7Z4
B	-2	HIS	-	EXPRESSION TAG	UNP P0A7Z4
B	-1	HIS	-	EXPRESSION TAG	UNP P0A7Z4
B	0	HIS	-	EXPRESSION TAG	UNP P0A7Z4
B	1	HIS	-	EXPRESSION TAG	UNP P0A7Z4
G	-4	HIS	-	EXPRESSION TAG	UNP P0A7Z4
G	-3	HIS	-	EXPRESSION TAG	UNP P0A7Z4
G	-2	HIS	-	EXPRESSION TAG	UNP P0A7Z4
G	-1	HIS	-	EXPRESSION TAG	UNP P0A7Z4
G	0	HIS	-	EXPRESSION TAG	UNP P0A7Z4
G	1	HIS	-	EXPRESSION TAG	UNP P0A7Z4
H	-4	HIS	-	EXPRESSION TAG	UNP P0A7Z4
H	-3	HIS	-	EXPRESSION TAG	UNP P0A7Z4
H	-2	HIS	-	EXPRESSION TAG	UNP P0A7Z4

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
H	-1	HIS	-	EXPRESSION TAG	UNP P0A7Z4
H	0	HIS	-	EXPRESSION TAG	UNP P0A7Z4
H	1	HIS	-	EXPRESSION TAG	UNP P0A7Z4

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1340	Total	C	N	O	S	3	0	0
			9522	5999	1675	1829	19			
2	I	1340	Total	C	N	O	S	3	0	0
			9544	6013	1676	1835	20			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1147	Total	C	N	O	S	0	0	0
			7549	4756	1355	1411	27			
3	J	1140	Total	C	N	O	S	0	0	0
			7512	4733	1348	1404	27			

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	E	89	Total	C	N	O	0	0	0
			482	299	93	90			
4	K	75	Total	C	N	O	0	0	0
			408	253	79	76			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	464	Total	C	N	O	S	0	0	0
			3592	2253	643	682	14			
5	L	464	Total	C	N	O	S	0	0	0
			3592	2253	643	682	14			

- Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	J	1	Total	Mg	0	0
			1	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	D	1	Total	Mg	0	0
			1	1		

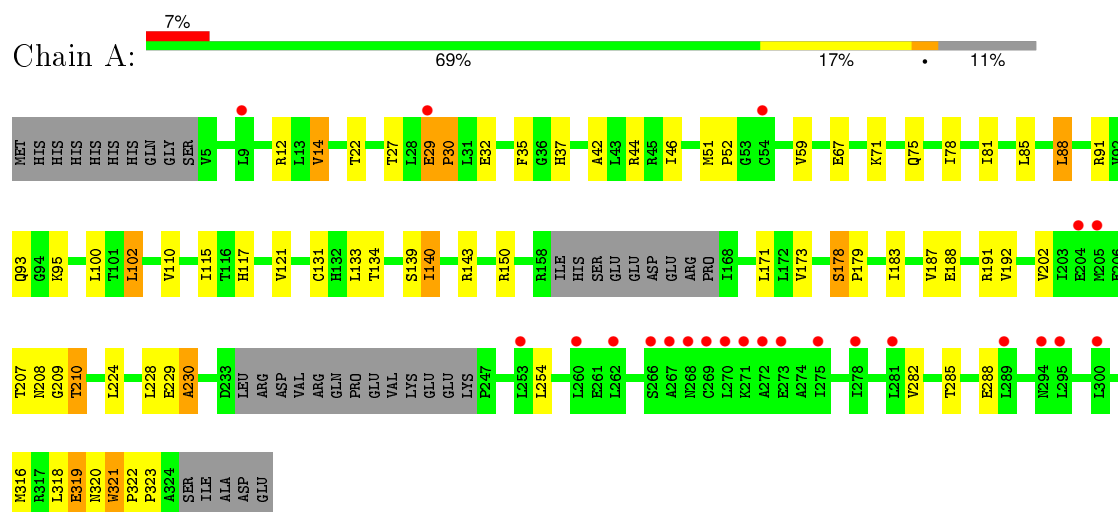
- Molecule 7 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	J	2	Total	Zn	0	0
			2	2		
7	D	2	Total	Zn	0	0
			2	2		

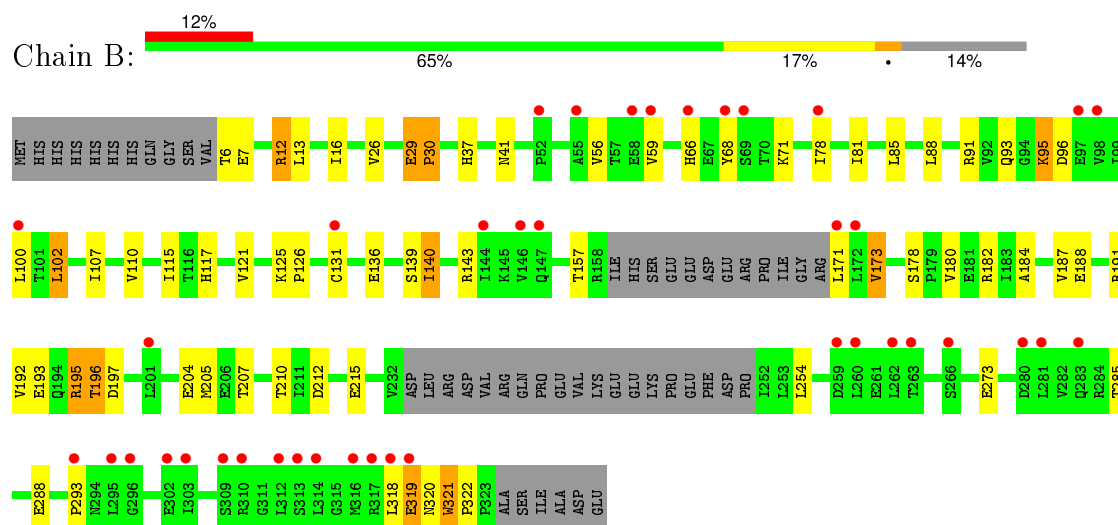
3 Residue-property plots [i](#)

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

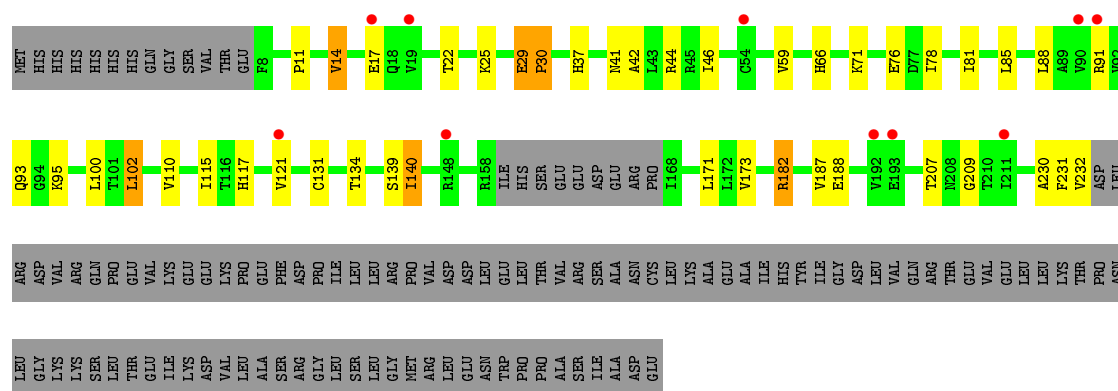


• Molecule 1: DNA-directed RNA polymerase subunit alpha

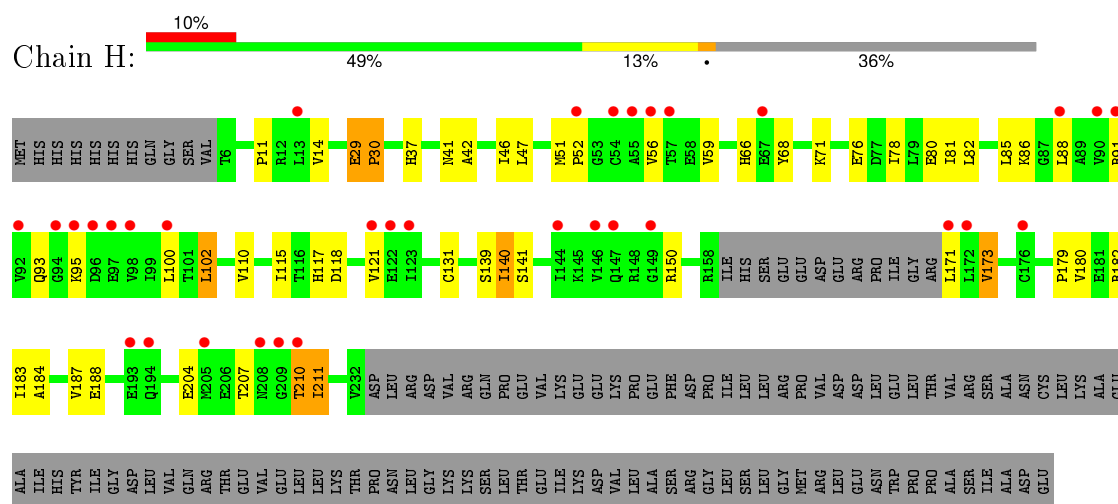


• Molecule 1: DNA-directed RNA polymerase subunit alpha

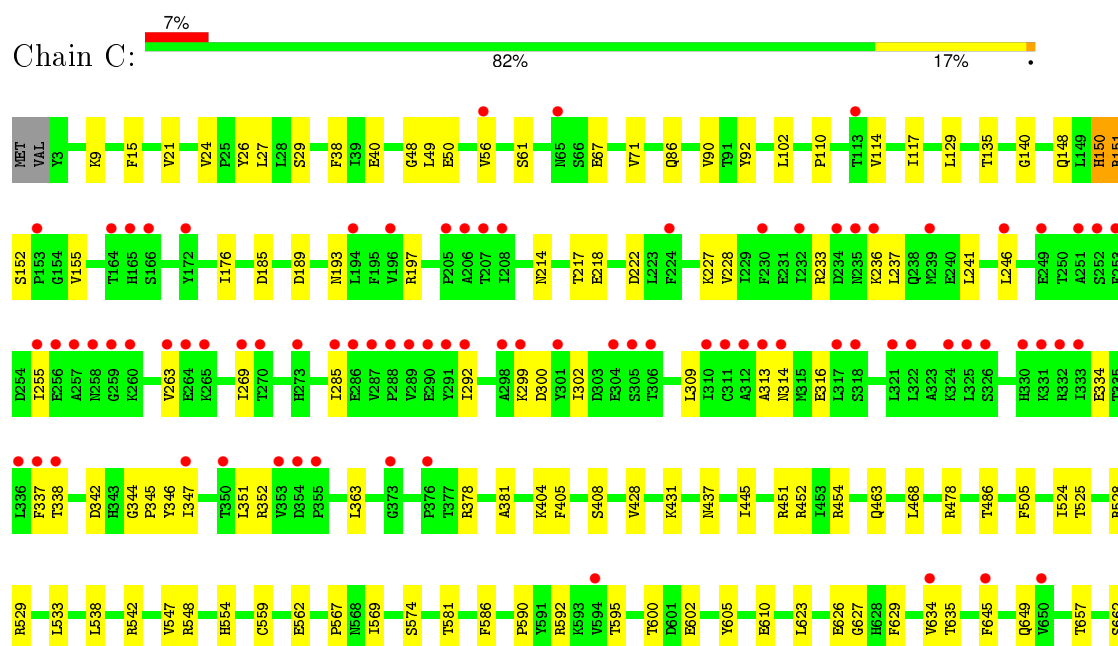


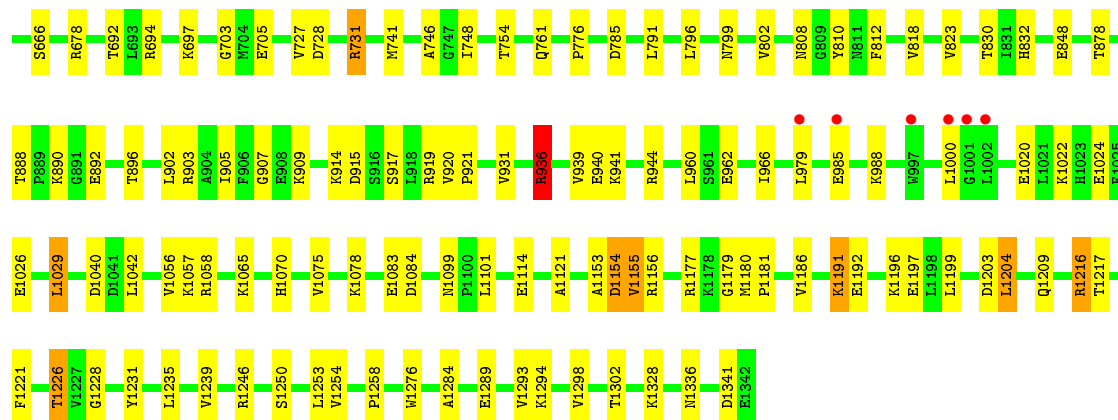


• Molecule 1: DNA-directed RNA polymerase subunit alpha

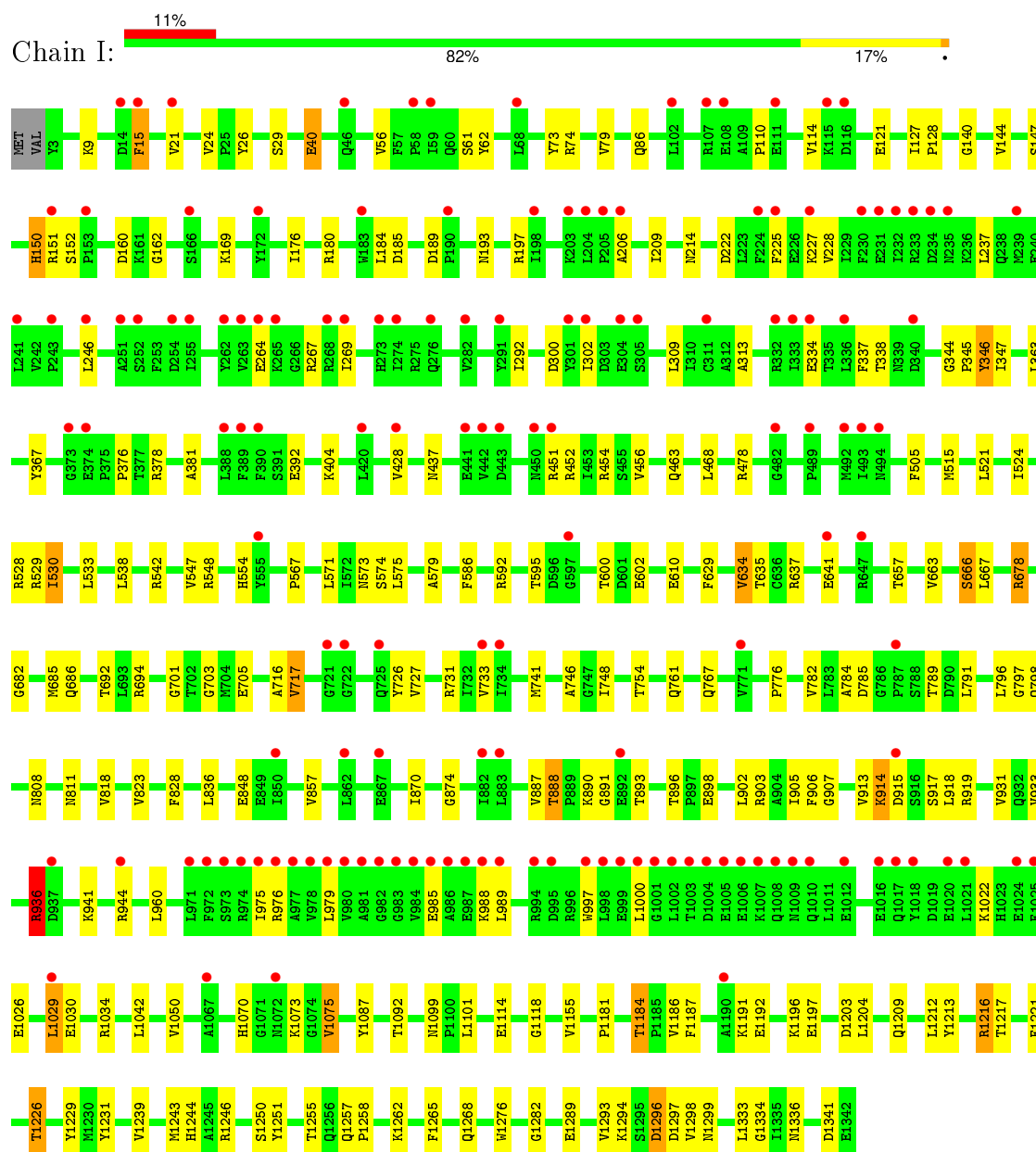


• Molecule 2: DNA-directed RNA polymerase subunit beta

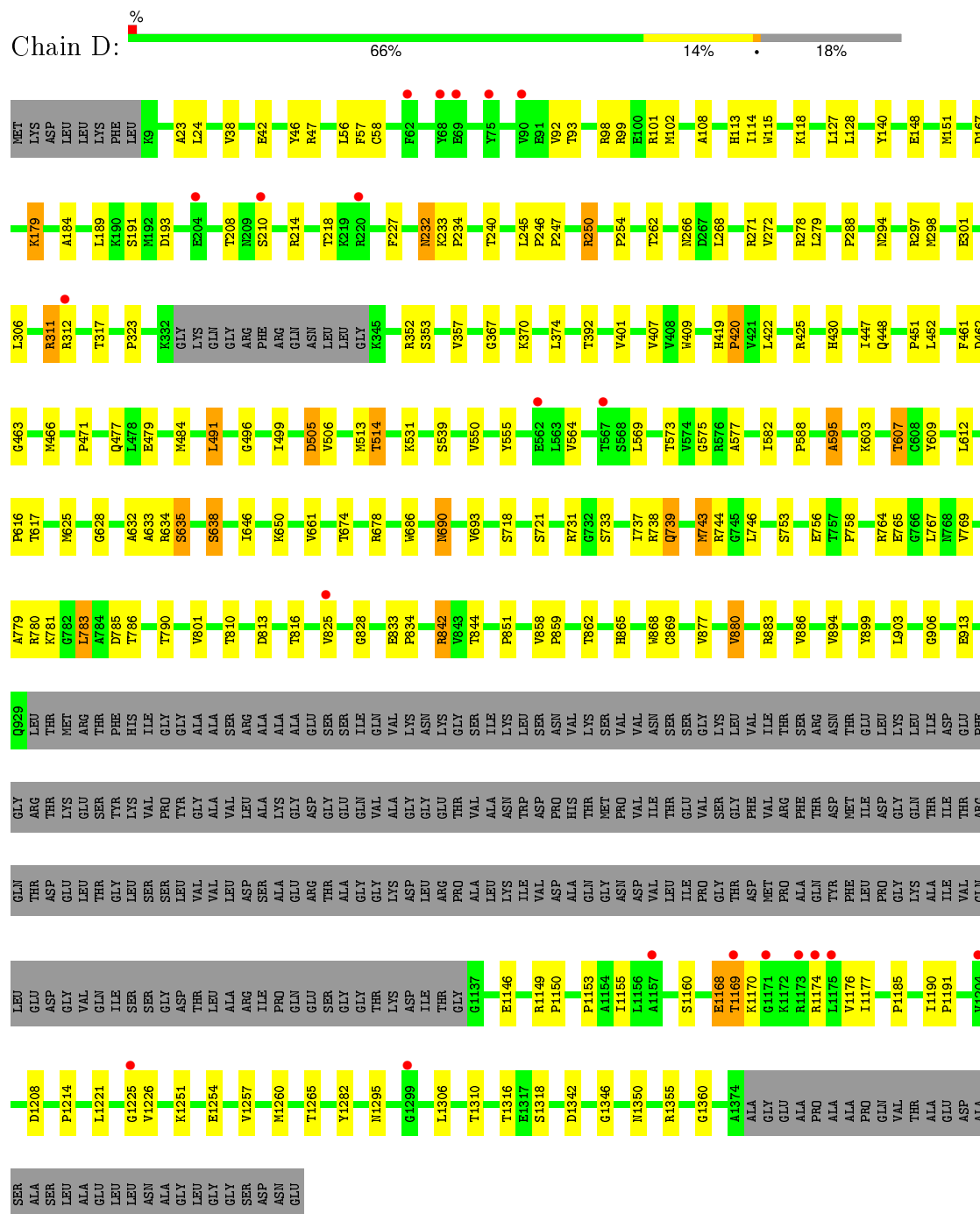


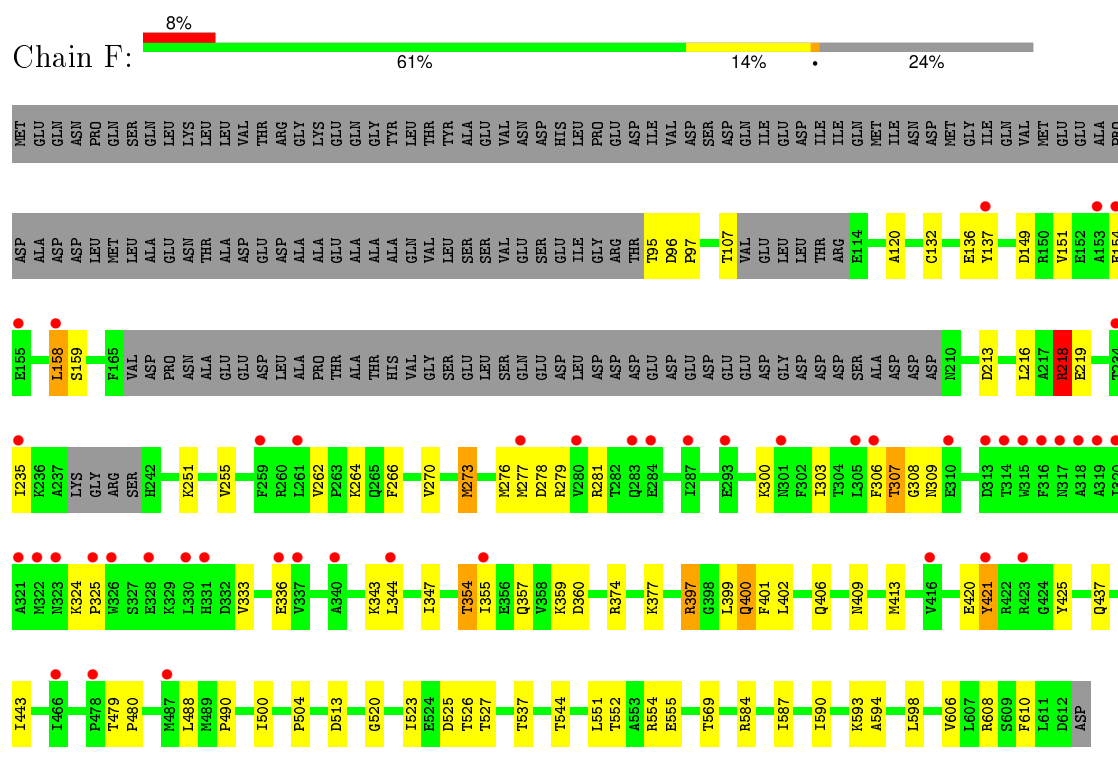


• Molecule 2: DNA-directed RNA polymerase subunit beta

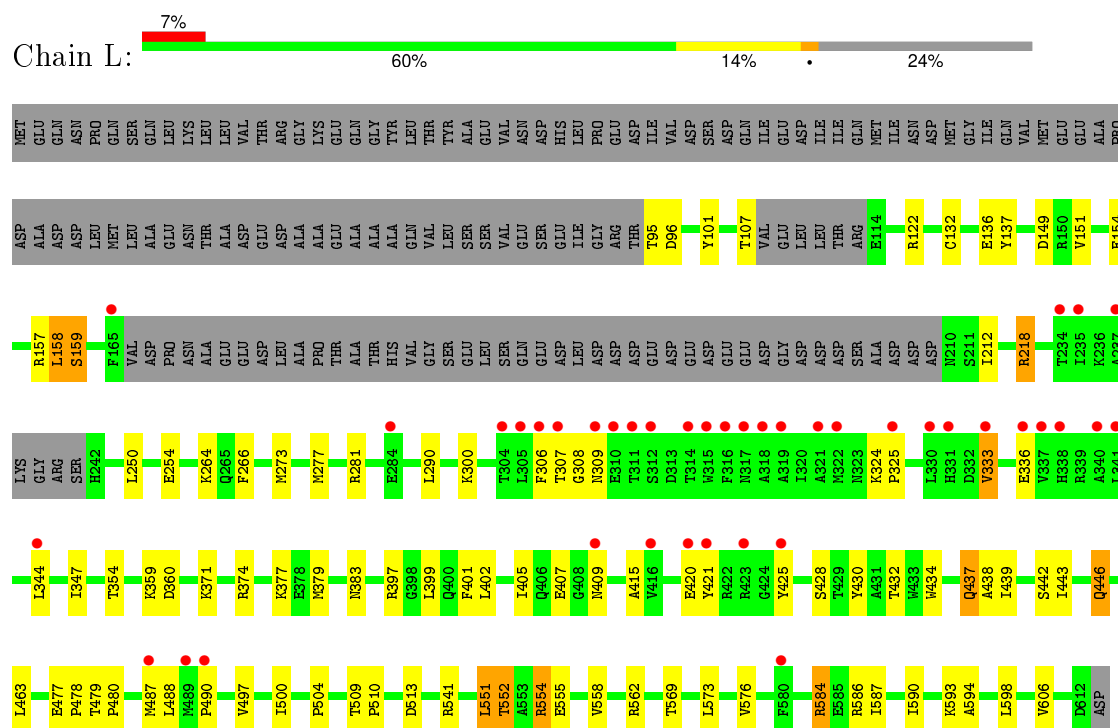


• Molecule 3: DNA-directed RNA polymerase subunit beta'





- Molecule 5: RNA polymerase sigma factor RpoD



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	186.39Å 207.24Å 308.30Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.92 – 3.95 49.92 – 3.95	Depositor EDS
% Data completeness (in resolution range)	96.6 (49.92-3.95) 96.2 (49.92-3.95)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	0.12	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.45 (at 4.00Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, R_{free}	0.276 , 0.325 0.273 , 0.324	Depositor DCC
R_{free} test set	2022 reflections (1.99%)	DCC
Wilson B-factor (Å ²)	142.5	Xtriage
Anisotropy	0.187	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.25 , 153.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.44$, $\langle L^2 \rangle = 0.27$	Xtriage
Outliers	0 of 103788 reflections	Xtriage
F_o, F_c correlation	0.88	EDS
Total number of atoms	49826	wwPDB-VP
Average B, all atoms (Å ²)	148.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.67% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

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 > 5$	RMSZ	# $ Z > 5$
1	A	0.25	0/2263	0.49	0/3073
1	B	0.25	0/2185	0.49	0/2967
1	G	0.24	0/1636	0.47	0/2221
1	H	0.23	0/1623	0.46	0/2205
2	C	0.26	0/9653	0.47	1/13062 (0.0%)
2	I	0.25	0/9676	0.45	1/13089 (0.0%)
3	D	0.28	0/7644	0.50	0/10385
3	J	0.25	0/7607	0.47	0/10334
4	E	0.33	0/482	0.63	1/662 (0.2%)
4	K	0.24	0/407	0.48	0/558
5	F	0.25	0/3636	0.47	2/4892 (0.0%)
5	L	0.25	0/3636	0.47	2/4892 (0.0%)
All	All	0.26	0/50448	0.48	7/68340 (0.0%)

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	E	39	VAL	C-N-CA	7.62	153.99	122.00
2	I	936	ARG	NE-CZ-NH2	6.31	123.45	120.30
2	C	936	ARG	NE-CZ-NH2	5.72	123.16	120.30
5	F	149	ASP	CB-CG-OD2	5.23	123.00	118.30
5	L	149	ASP	CB-CG-OD2	5.17	122.96	118.30

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	2236	0	2254	31	0
1	B	2160	0	2184	30	0
1	G	1618	0	1622	22	0
1	H	1605	0	1599	28	0
2	C	9522	0	8569	113	0
2	I	9544	0	8601	125	0
3	D	7549	0	6266	88	0
3	J	7512	0	6245	87	0
4	E	482	0	301	7	0
4	K	408	0	255	9	0
5	F	3592	0	3433	40	0
5	L	3592	0	3433	46	0
6	D	1	0	0	0	0
6	J	1	0	0	0	0
7	D	2	0	0	0	0
7	J	2	0	0	0	0
All	All	49826	0	44762	558	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 6.

The worst 5 of 558 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)
3:J:910:ASN:CG	4:K:15:ASN:HB2	1.65	1.17
2:I:936:ARG:HH21	2:I:936:ARG:HG3	1.29	0.98
2:C:936:ARG:HG3	2:C:936:ARG:HH21	1.33	0.92
2:C:808:ASN:H	3:D:633:ALA:HB2	1.44	0.82
1:G:11:PRO:HA	1:G:30:PRO:HD2	1.65	0.79

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	292/335 (87%)	268 (92%)	14 (5%)	10 (3%)	5	43
1	B	281/335 (84%)	257 (92%)	12 (4%)	12 (4%)	3	35
1	G	212/335 (63%)	193 (91%)	14 (7%)	5 (2%)	7	50
1	H	211/335 (63%)	195 (92%)	12 (6%)	4 (2%)	10	54
2	C	1338/1342 (100%)	1259 (94%)	60 (4%)	19 (1%)	14	59
2	I	1338/1342 (100%)	1257 (94%)	65 (5%)	16 (1%)	16	62
3	D	1141/1407 (81%)	1030 (90%)	75 (7%)	36 (3%)	5	44
3	J	1134/1407 (81%)	1026 (90%)	71 (6%)	37 (3%)	5	43
4	E	87/91 (96%)	72 (83%)	6 (7%)	9 (10%)	1	12
4	K	73/91 (80%)	61 (84%)	5 (7%)	7 (10%)	1	14
5	F	456/613 (74%)	437 (96%)	14 (3%)	5 (1%)	17	63
5	L	456/613 (74%)	436 (96%)	14 (3%)	6 (1%)	15	60
All	All	7019/8246 (85%)	6491 (92%)	362 (5%)	166 (2%)	7	50

5 of 166 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	93	GLN
1	A	188	GLU
1	A	229	GLU
1	A	319	GLU
1	A	320	ASN

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	239/292 (82%)	220 (92%)	19 (8%)	15	54
1	B	233/292 (80%)	215 (92%)	18 (8%)	16	55
1	G	173/292 (59%)	163 (94%)	10 (6%)	25	64
1	H	171/292 (59%)	159 (93%)	12 (7%)	19	58
2	C	822/1157 (71%)	763 (93%)	59 (7%)	18	57
2	I	828/1157 (72%)	770 (93%)	58 (7%)	19	58
3	D	517/1168 (44%)	451 (87%)	66 (13%)	5	31
3	J	515/1168 (44%)	447 (87%)	68 (13%)	5	30
4	E	10/75 (13%)	8 (80%)	2 (20%)	1	13
4	K	9/75 (12%)	8 (89%)	1 (11%)	8	38
5	F	348/540 (64%)	312 (90%)	36 (10%)	9	41
5	L	348/540 (64%)	317 (91%)	31 (9%)	12	48
All	All	4213/7048 (60%)	3833 (91%)	380 (9%)	12	47

5 of 380 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	F	262	VAL
1	H	95	LYS
5	L	154	GLU
5	F	333	VAL
5	F	527	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 65 such sidechains are listed below:

Mol	Chain	Res	Type
5	F	128	ASN
5	F	472	GLN
5	L	129	GLN
5	F	129	GLN
5	F	357	GLN

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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

Of 6 ligands modelled in this entry, 6 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 ⓘ

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, 95th 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	298/335 (88%)	0.26	23 (7%) 16 11	27, 114, 296, 459	0
1	B	287/335 (85%)	0.66	40 (13%) 4 4	36, 191, 420, 550	0
1	G	216/335 (64%)	0.33	10 (4%) 36 26	33, 168, 289, 370	0
1	H	215/335 (64%)	0.65	33 (15%) 3 3	68, 188, 341, 524	0
2	C	1340/1342 (99%)	0.16	88 (6%) 22 14	2, 102, 402, 550	1 (0%)
2	I	1340/1342 (99%)	0.40	146 (10%) 7 6	3, 144, 388, 550	1 (0%)
3	D	1147/1407 (81%)	-0.18	21 (1%) 71 61	3, 72, 236, 550	0
3	J	1140/1407 (81%)	-0.09	33 (2%) 55 42	3, 96, 268, 550	0
4	E	89/91 (97%)	-0.66	0 100 100	6, 76, 206, 285	0
4	K	75/91 (82%)	0.07	1 (1%) 79 70	47, 171, 393, 535	0
5	F	464/613 (75%)	0.35	46 (9%) 9 7	21, 149, 369, 550	0
5	L	464/613 (75%)	0.35	41 (8%) 12 9	31, 169, 397, 550	0
All	All	7075/8246 (85%)	0.17	482 (6%) 20 13	2, 121, 352, 550	2 (0%)

The worst 5 of 482 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	I	1000	LEU	17.4
5	L	315	TRP	15.7
2	C	311	CYS	15.4
2	C	305	SER	12.6
2	I	999	GLU	12.2

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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(\AA^2)	Q<0.9
7	ZN	D	2002	1/1	0.99	0.20	-0.72	68,68,68,68	0
7	ZN	J	2002	1/1	1.00	0.19	-1.09	63,63,63,63	0
7	ZN	J	2001	1/1	0.99	0.09	-1.57	104,104,104,104	0
7	ZN	D	2001	1/1	0.99	0.07	-2.75	83,83,83,83	0
6	MG	D	2000	1/1	0.75	0.35	-	320,320,320,320	0
6	MG	J	2000	1/1	0.39	0.36	-	396,396,396,396	0

6.5 Other polymers [i](#)

There are no such residues in this entry.