



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 13, 2017 – 01:01 PM EDT

PDB ID : 4XLQ
Title : Crystal structure of T.aquaticus transcription initiation complex containing upstream fork (-11 base-paired) promoter
Authors : Bae, B.; Darst, S.A.
Deposited on : unknown
Resolution : 4.60 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20029824
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	rb-20029824

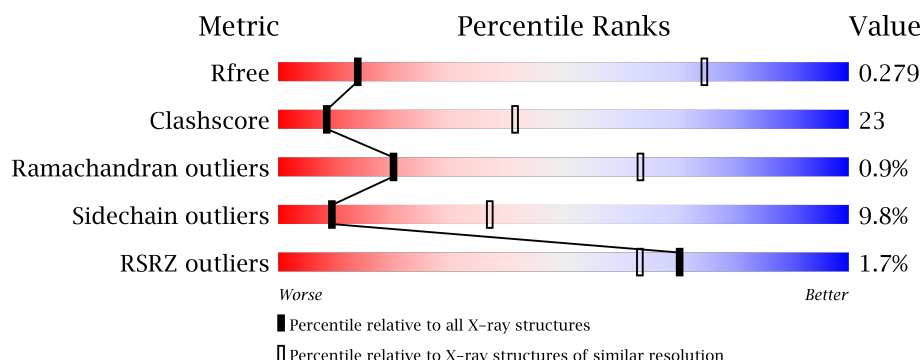
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 4.60 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}	100719	1005 (5.52-3.66)
Clashscore	112137	1052 (5.50-3.70)
Ramachandran outliers	110173	1041 (5.52-3.66)
Sidechain outliers	110143	1021 (5.50-3.66)
RSRZ outliers	101464	1014 (5.52-3.66)

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	314	<div> <div> <div>0.1%</div> <div>40%</div> <div>28%</div> <div>0.1%</div> <div>28%</div> </div> </div>
1	B	314	<div> <div>33%</div> <div>34%</div> <div>5%</div> <div>28%</div> </div>
1	G	314	<div> <div>5%</div> <div>36%</div> <div>31%</div> <div>5%</div> <div>28%</div> </div>
1	H	314	<div> <div>0.1%</div> <div>35%</div> <div>32%</div> <div>5%</div> <div>28%</div> </div>
2	C	1119	<div> <div>2%</div> <div>45%</div> <div>48%</div> <div>6%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	I	1119	<div><div></div><div>2%</div><div>46%</div><div>47%</div><div>6%</div><div></div></div>
3	D	1524	<div><div></div><div>%</div><div>49%</div><div>43%</div><div>6%</div><div></div></div>
3	J	1524	<div><div></div><div>%</div><div>46%</div><div>39%</div><div>5%</div><div>10%</div><div></div></div>
4	E	99	<div><div></div><div>2%</div><div>59%</div><div>34%</div><div></div><div>6%</div></div>
4	K	99	<div><div></div><div>3%</div><div>65%</div><div>28%</div><div></div><div>6%</div></div>
5	F	347	<div><div></div><div></div><div>56%</div><div>42%</div><div></div><div></div></div>
5	L	347	<div><div></div><div>2%</div><div>54%</div><div>43%</div><div></div><div></div></div>
6	O	30	<div><div></div><div>13%</div><div>17%</div><div>83%</div><div></div></div>
6	R	30	<div><div></div><div>13%</div><div>87%</div><div></div></div>
7	P	26	<div><div></div><div>19%</div><div>8%</div><div>88%</div><div></div><div></div></div>
7	S	26	<div><div></div><div>12%</div><div>88%</div><div></div></div>

2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 56477 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	227	Total	C	N	O	S	0	0	0
			1770	1130	303	334	3			
1	B	227	Total	C	N	O	S	0	0	0
			1770	1130	303	334	3			
1	G	227	Total	C	N	O	S	0	0	0
			1770	1130	303	334	3			
1	H	227	Total	C	N	O	S	0	0	0
			1770	1130	303	334	3			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1112	Total	C	N	O	S	0	0	0
			8739	5531	1553	1632	23			
2	I	1112	Total	C	N	O	S	0	0	0
			8739	5531	1553	1632	23			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1490	Total	C	N	O	S	0	0	0
			11761	7439	2088	2196	38			
3	J	1367	Total	C	N	O	S	0	0	0
			10779	6810	1923	2010	36			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	93	Total	C	N	O	S	0	0	0
			768	490	136	138	4			
4	K	93	Total	C	N	O	S	0	0	0
			768	490	136	138	4			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	345	Total	C	N	O	S	0	0	0
			2787	1758	502	523	4			
5	L	345	Total	C	N	O	S	0	0	0
			2787	1758	502	523	4			

- Molecule 6 is a DNA chain called DNA (30-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	O	30	Total	C	N	O	P	0	0	0
			613	296	109	179	29			
6	R	30	Total	C	N	O	P	0	0	0
			613	296	109	179	29			

- Molecule 7 is a DNA chain called DNA (26-MER).

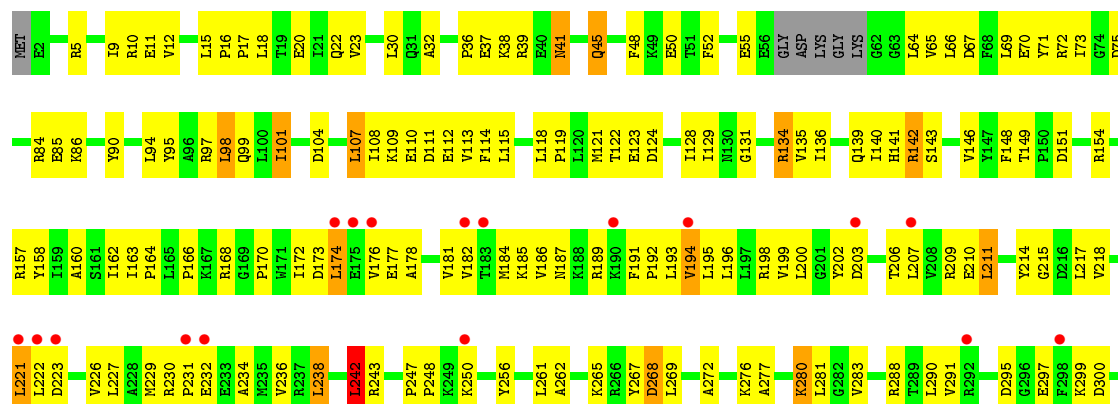
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	P	25	Total	C	N	O	P	0	0	0
			510	245	91	149	25			
7	S	26	Total	C	N	O	P	0	0	0
			527	255	93	154	25			

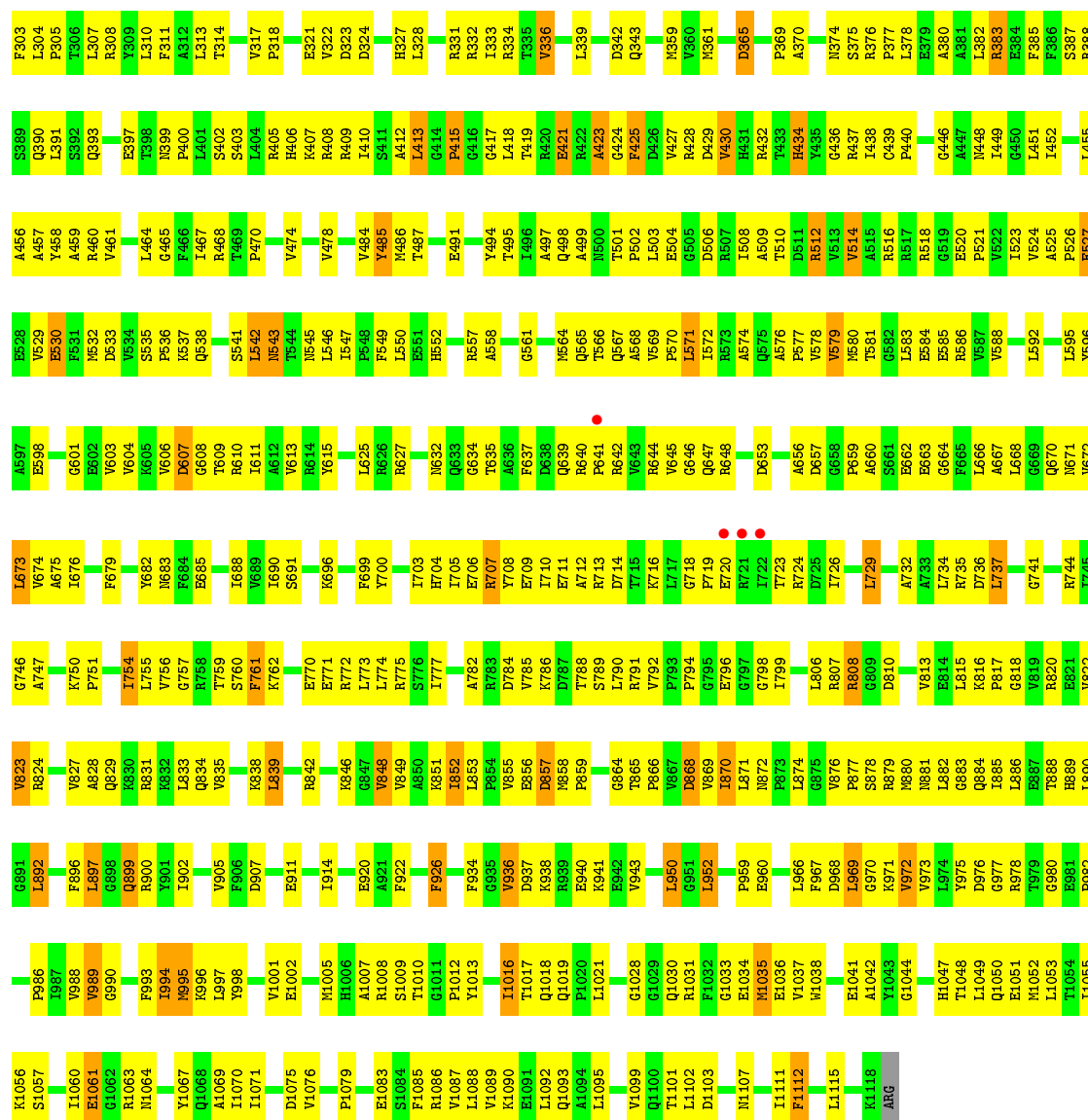
- Molecule 8 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

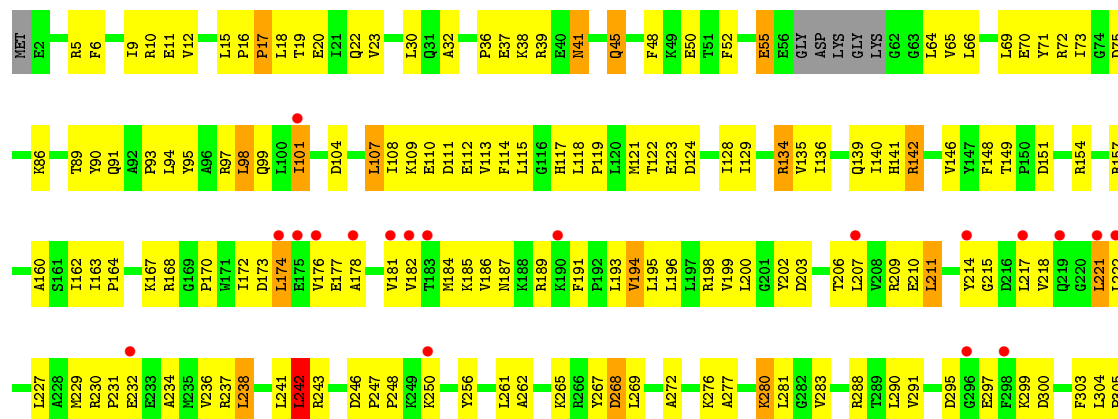
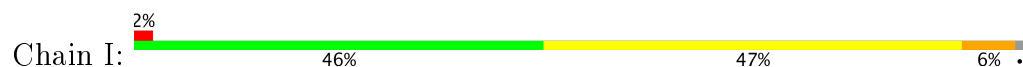
- Molecule 9 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

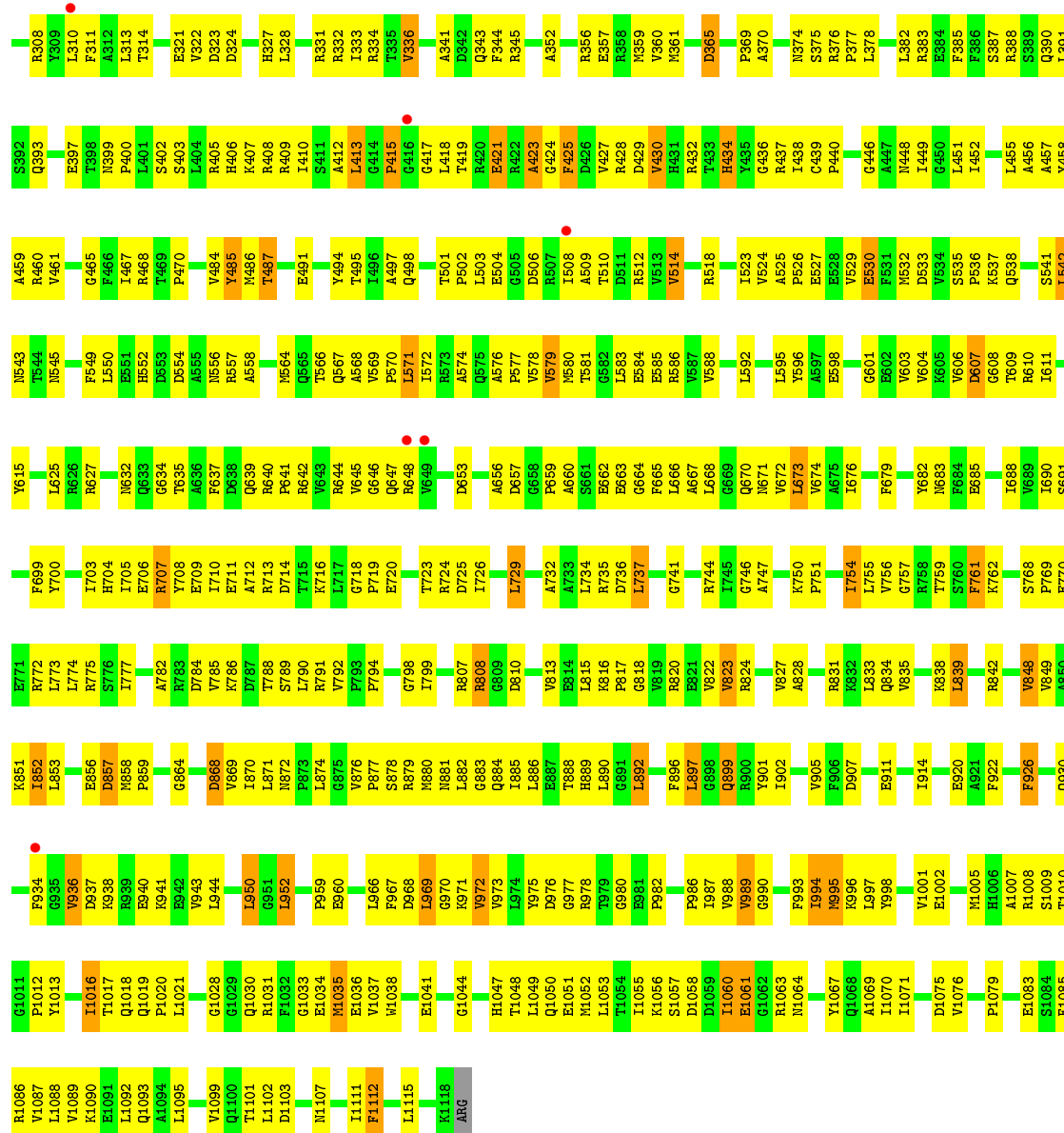
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	J	1	Total	Mg	0	0
			1	1		
9	D	1	Total	Mg	0	0
			1	1		



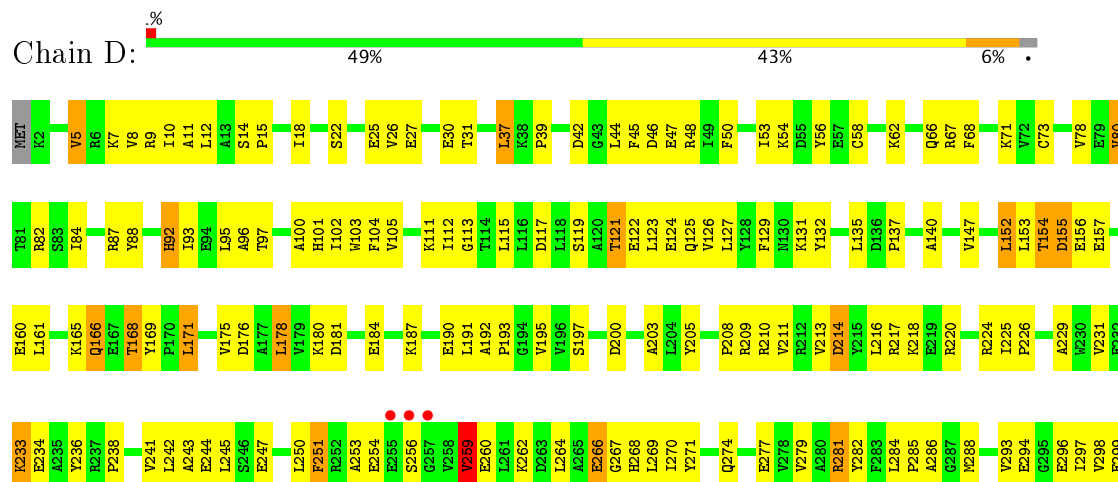


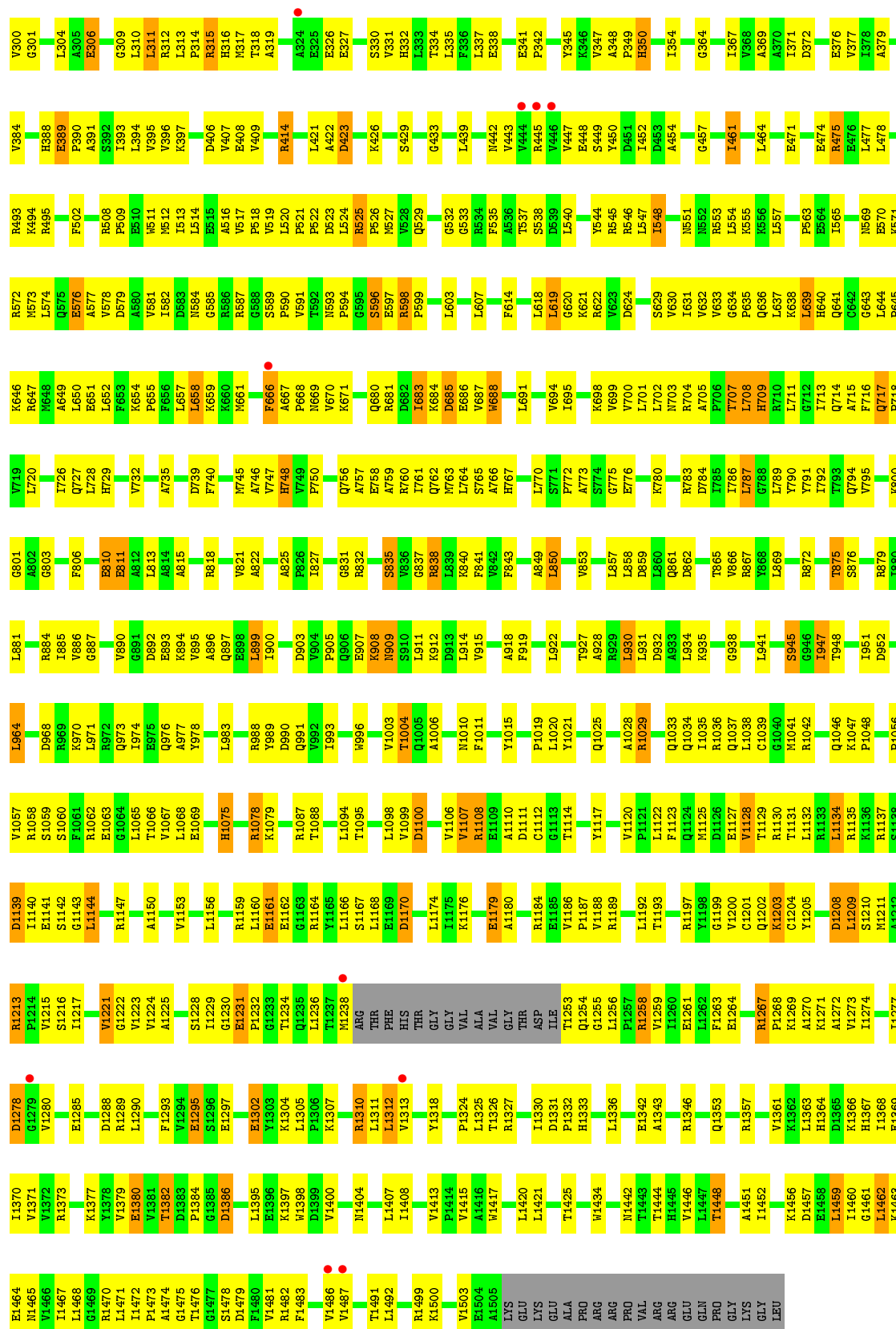
• Molecule 2: DNA-directed RNA polymerase subunit beta



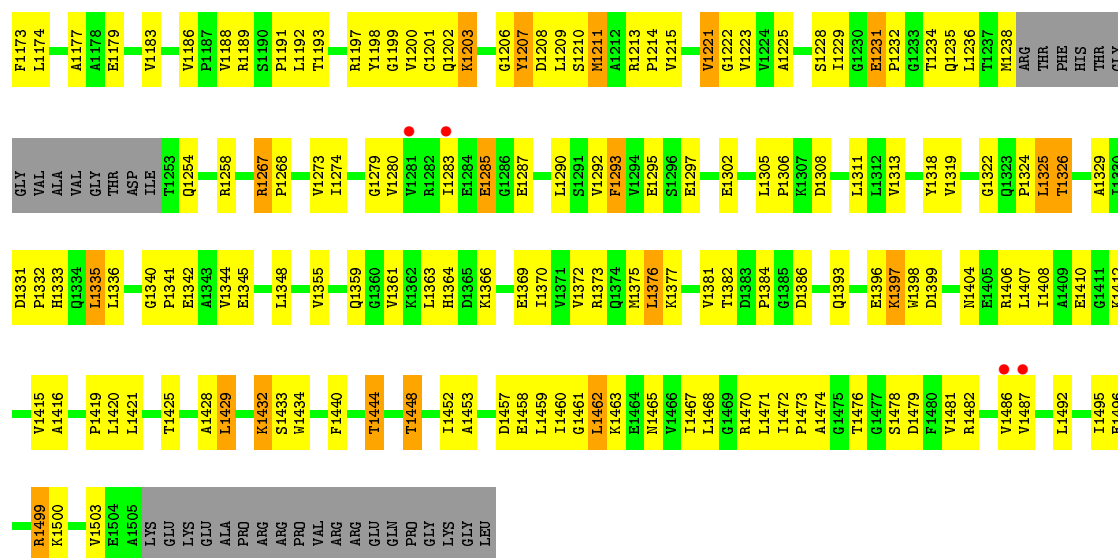


• Molecule 3: DNA-directed RNA polymerase subunit beta'

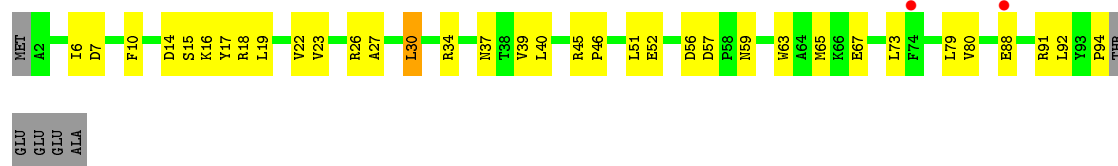




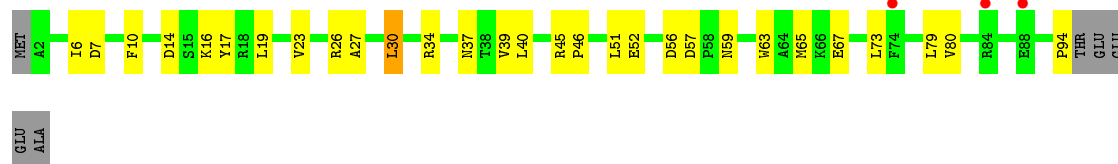




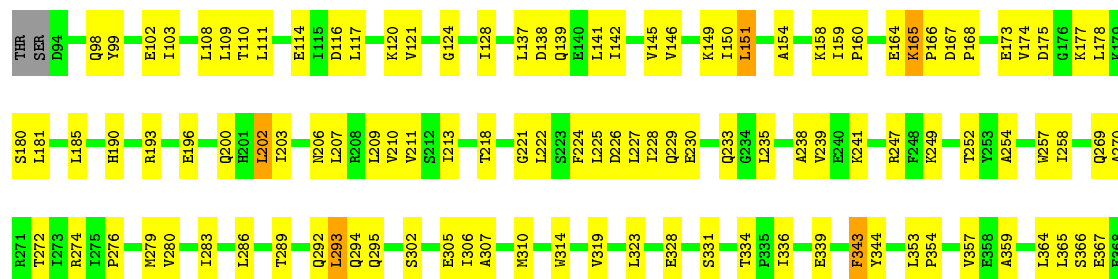
• Molecule 4: DNA-directed RNA polymerase subunit omega



• Molecule 4: DNA-directed RNA polymerase subunit omega

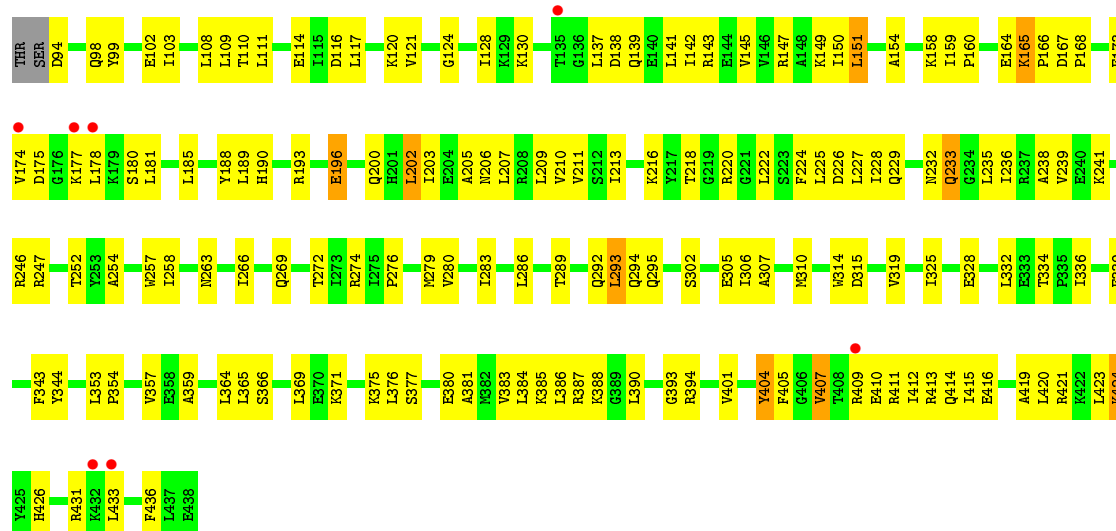


• Molecule 5: RNA polymerase sigma factor SigA

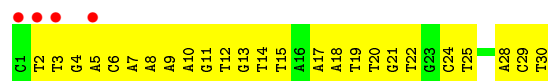




• Molecule 5: RNA polymerase sigma factor SigA



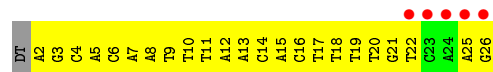
• Molecule 6: DNA (30-MER)



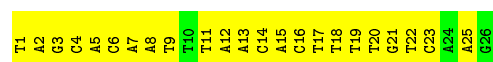
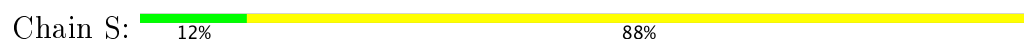
• Molecule 6: DNA (30-MER)



• Molecule 7: DNA (26-MER)



• Molecule 7: DNA (26-MER)



4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α , β , γ	288.23Å 288.23Å 535.25Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.81 – 4.60 49.81 – 4.60	Depositor EDS
% Data completeness (in resolution range)	98.6 (49.81-4.60) 98.6 (49.81-4.60)	Depositor EDS
R_{merge}	0.22	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.21 (at 4.64Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1839)	Depositor
R, R_{free}	0.245 , 0.281 0.240 , 0.279	Depositor DCC
R_{free} test set	6217 reflections (5.04%)	DCC
Wilson B-factor (Å ²)	154.6	Xtriage
Anisotropy	0.027	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 174.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.33$, $\langle L^2 \rangle = 0.17$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.87	EDS
Total number of atoms	56477	wwPDB-VP
Average B, all atoms (Å ²)	175.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.80% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

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	0.31	0/1804	0.64	1/2455 (0.0%)
1	B	0.30	0/1804	0.61	0/2455
1	G	0.31	0/1804	0.64	1/2455 (0.0%)
1	H	0.30	0/1804	0.61	0/2455
2	C	0.27	0/8905	0.55	2/12040 (0.0%)
2	I	0.27	0/8905	0.55	2/12040 (0.0%)
3	D	0.28	0/11963	0.55	3/16165 (0.0%)
3	J	0.28	0/10959	0.57	1/14802 (0.0%)
4	E	0.25	0/783	0.54	0/1054
4	K	0.25	0/783	0.53	0/1054
5	F	0.27	0/2829	0.55	1/3804 (0.0%)
5	L	0.27	0/2829	0.55	1/3804 (0.0%)
6	O	0.50	0/687	0.92	0/1059
6	R	0.50	0/687	0.91	0/1059
7	P	0.54	0/571	0.93	0/878
7	S	0.54	0/590	0.93	0/908
All	All	0.29	0/57707	0.59	12/78487 (0.0%)

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
2	C	0	3
2	I	0	3
3	D	0	1
3	J	0	1
All	All	0	8

There are no bond length outliers.

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	311	LEU	CA-CB-CG	7.45	132.43	115.30
3	D	1134	LEU	CA-CB-CG	6.98	131.36	115.30
2	I	417	GLY	N-CA-C	6.42	129.14	113.10
2	C	417	GLY	N-CA-C	6.40	129.09	113.10
3	J	1134	LEU	CA-CB-CG	5.63	128.26	115.30

There are no chirality outliers.

5 of 8 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	415	PRO	Peptide
2	C	423	ALA	Peptide
2	C	737	LEU	Peptide
3	D	1208	ASP	Peptide
2	I	415	PRO	Peptide

5.2 Too-close contacts [i](#)

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	1770	0	1799	89	0
1	B	1770	0	1799	101	0
1	G	1770	0	1799	103	0
1	H	1770	0	1799	95	0
2	C	8739	0	8841	499	0
2	I	8739	0	8841	485	0
3	D	11761	0	11976	585	0
3	J	10779	0	10993	503	0
4	E	768	0	784	37	0
4	K	768	0	784	29	0
5	F	2787	0	2866	120	0
5	L	2787	0	2866	133	0
6	O	613	0	343	28	0
6	R	613	0	343	26	0
7	P	510	0	284	27	0
7	S	527	0	297	25	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
8	D	2	0	0	0	0
8	J	2	0	0	0	0
9	D	1	0	0	0	0
9	J	1	0	0	0	0
All	All	56477	0	56414	2598	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 23.

The worst 5 of 2598 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:C:707:ARG:HE	2:C:824:ARG:HE	1.17	0.90
6:R:24:DC:H42	7:S:3:DG:H1	1.18	0.90
3:D:105:VAL:HA	3:D:112:ILE:HD11	1.55	0.88
4:E:30:LEU:HD12	4:E:37:ASN:HD21	1.39	0.88
4:K:30:LEU:HD12	4:K:37:ASN:HD21	1.39	0.87

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	225/314 (72%)	200 (89%)	23 (10%)	2 (1%)	20	63
1	B	225/314 (72%)	200 (89%)	20 (9%)	5 (2%)	8	46
1	G	225/314 (72%)	200 (89%)	23 (10%)	2 (1%)	20	63
1	H	225/314 (72%)	201 (89%)	18 (8%)	6 (3%)	6	41
2	C	1108/1119 (99%)	958 (86%)	139 (12%)	11 (1%)	18	61
2	I	1108/1119 (99%)	956 (86%)	140 (13%)	12 (1%)	17	60

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	D	1486/1524 (98%)	1315 (88%)	162 (11%)	9 (1%)	28	71
3	J	1361/1524 (89%)	1201 (88%)	150 (11%)	10 (1%)	25	68
4	E	91/99 (92%)	75 (82%)	16 (18%)	0	100	100
4	K	91/99 (92%)	75 (82%)	16 (18%)	0	100	100
5	F	343/347 (99%)	299 (87%)	42 (12%)	2 (1%)	28	71
5	L	343/347 (99%)	302 (88%)	40 (12%)	1 (0%)	44	81
All	All	6831/7434 (92%)	5982 (88%)	789 (12%)	60 (1%)	20	63

5 of 60 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	53	VAL
3	D	1128	VAL
3	D	1209	LEU
1	G	53	VAL
3	J	1128	VAL

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 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	194/270 (72%)	179 (92%)	15 (8%)	15	49
1	B	194/270 (72%)	172 (89%)	22 (11%)	7	31
1	G	194/270 (72%)	178 (92%)	16 (8%)	13	46
1	H	194/270 (72%)	171 (88%)	23 (12%)	6	29
2	C	931/936 (100%)	840 (90%)	91 (10%)	9	37
2	I	931/936 (100%)	840 (90%)	91 (10%)	9	37
3	D	1252/1281 (98%)	1114 (89%)	138 (11%)	7	32
3	J	1150/1281 (90%)	1028 (89%)	122 (11%)	8	34
4	E	83/88 (94%)	79 (95%)	4 (5%)	30	63
4	K	83/88 (94%)	79 (95%)	4 (5%)	30	63

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
5	F	296/299 (99%)	276 (93%)	20 (7%)	18	53
5	L	296/299 (99%)	276 (93%)	20 (7%)	18	53
All	All	5798/6288 (92%)	5232 (90%)	566 (10%)	9	37

5 of 566 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	1382	THR
1	H	113	ASP
3	J	1293	PHE
3	D	1459	LEU
5	F	409	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 85 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	1235	GLN
1	G	63	HIS
4	K	37	ASN
3	D	1359	GLN
5	F	294	GLN

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](#)

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 [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

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	227/314 (72%)	0.08	2 (0%) 84 78	107, 189, 236, 264	0
1	B	227/314 (72%)	-0.23	0 100 100	94, 161, 214, 264	0
1	G	227/314 (72%)	0.39	16 (7%) 17 16	116, 199, 240, 276	0
1	H	227/314 (72%)	-0.13	2 (0%) 84 78	116, 173, 216, 267	0
2	C	1112/1119 (99%)	-0.03	21 (1%) 67 60	90, 179, 243, 314	0
2	I	1112/1119 (99%)	0.01	25 (2%) 62 55	94, 185, 244, 315	0
3	D	1490/1524 (97%)	-0.13	13 (0%) 84 78	64, 149, 204, 259	0
3	J	1367/1524 (89%)	-0.08	18 (1%) 77 69	79, 159, 217, 264	0
4	E	93/99 (93%)	0.05	2 (2%) 62 55	100, 159, 205, 238	0
4	K	93/99 (93%)	0.03	3 (3%) 48 40	112, 168, 216, 253	0
5	F	345/347 (99%)	-0.10	1 (0%) 93 91	115, 185, 255, 300	0
5	L	345/347 (99%)	-0.11	7 (2%) 65 58	121, 188, 252, 300	0
6	O	30/30 (100%)	0.70	4 (13%) 4 6	154, 221, 293, 311	0
6	R	30/30 (100%)	0.22	0 100 100	164, 221, 257, 267	0
7	P	25/26 (96%)	0.81	5 (20%) 1 3	172, 235, 306, 326	0
7	S	26/26 (100%)	0.07	0 100 100	184, 224, 263, 283	0
All	All	6976/7546 (92%)	-0.05	119 (1%) 70 63	64, 171, 235, 326	0

The worst 5 of 119 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	176	VAL	5.1
2	I	175	GLU	4.6
2	C	221	LEU	4.4
1	G	13	ALA	4.1
2	C	175	GLU	3.7

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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
8	ZN	D	2001	1/1	0.99	0.13	-0.59	107,107,107,107	0
8	ZN	J	2001	1/1	0.97	0.12	-1.01	166,166,166,166	0
8	ZN	D	2002	1/1	0.96	0.16	-1.20	182,182,182,182	0
8	ZN	J	2002	1/1	0.93	0.07	-1.44	147,147,147,147	0
9	MG	D	2003	1/1	0.82	0.45	-	286,286,286,286	0
9	MG	J	2003	1/1	0.84	0.41	-	331,331,331,331	0

6.5 Other polymers [i](#)

There are no such residues in this entry.