wwPDB X-ray Structure Validation Summary Report (i)

Feb 5, 2024 – 09:54 PM EST

PDB ID	:	1YNN
Title	:	Taq RNA polymerase-rifampicin complex
Authors	:	Campbell, E.A.; Pavlova, O.; Zenkin, N.; Leon, F.; Irschik, H.; Jansen, R.; Severinov, K.; Darst, S.A.
Deposited on	:	2005-01-24
Resolution	:	3.30 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 (i)) were used in the production of this report:

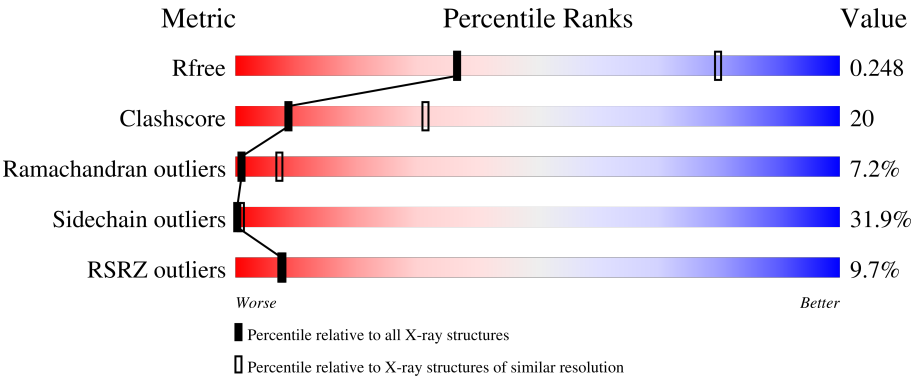
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.30 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	1149 (3.34-3.26)
Clashscore	141614	1205 (3.34-3.26)
Ramachandran outliers	138981	1183 (3.34-3.26)
Sidechain outliers	138945	1182 (3.34-3.26)
RSRZ outliers	127900	1115 (3.34-3.26)

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	A	314	4% 36% 25% 10% • 27%
1	B	314	5% 31% 29% 10% • 28%
2	C	1119	5% 43% 40% 15% •
3	D	1524	12% 37% 32% 11% • 19%
3	J	1524	% 8% 5% • 84%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	K	99	

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 24368 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 alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	230	Total	C	N	O	S	0	0	0
			1763	1126	300	334	3			
1	B	225	Total	C	N	O	S	0	0	0
			1750	1118	300	329	3			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1114	Total	C	N	O	S	0	0	0
			8576	5430	1513	1609	24			

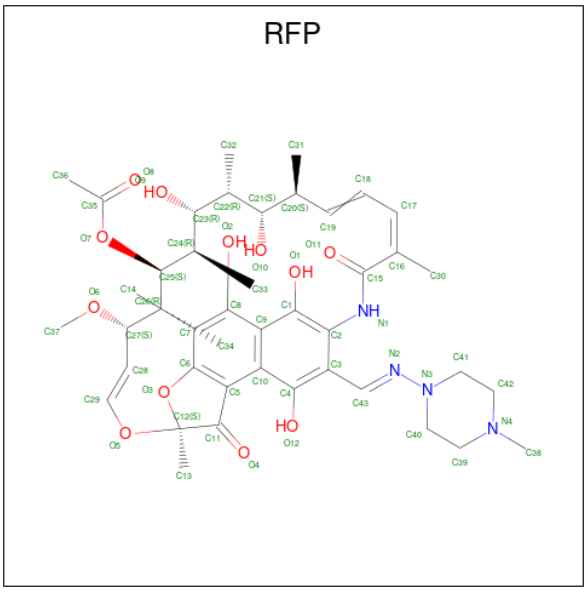
- Molecule 3 is a protein called DNA-directed RNA polymerase beta' chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1238	Total	C	N	O	S	0	0	0
			9602	6065	1703	1798	36			
3	J	249	Total	C	N	O	S	0	0	0
			1869	1191	320	356	2			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	K	95	Total	C	N	O	S	0	0	0
			747	476	134	132	5			

- Molecule 5 is RIFAMPICIN (three-letter code: RFP) (formula: $C_{43}H_{58}N_4O_{12}$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	C	1	Total C N O 59 43 4 12	0	0

- Molecule 6 is ZINC ION (three-letter code: ZN) (formula: Zn).

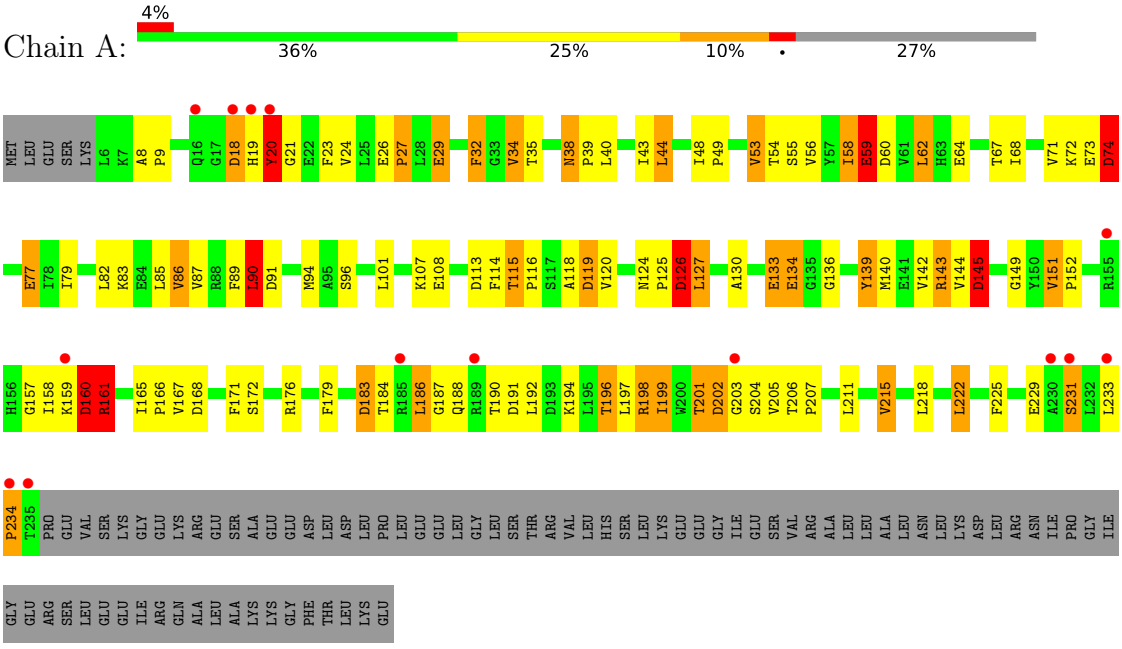
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	D	2	Total Zn 2 2	0	0

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.

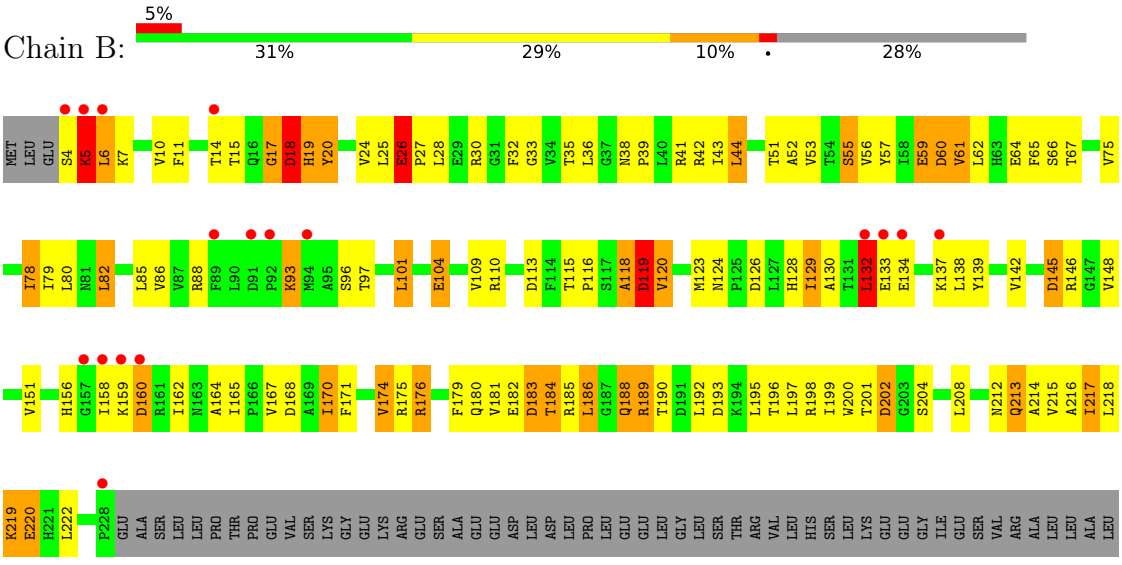
• Molecule 1: DNA-directed RNA polymerase alpha chain

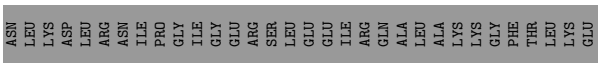
Chain A:



• Molecule 1: DNA-directed RNA polymerase alpha chain

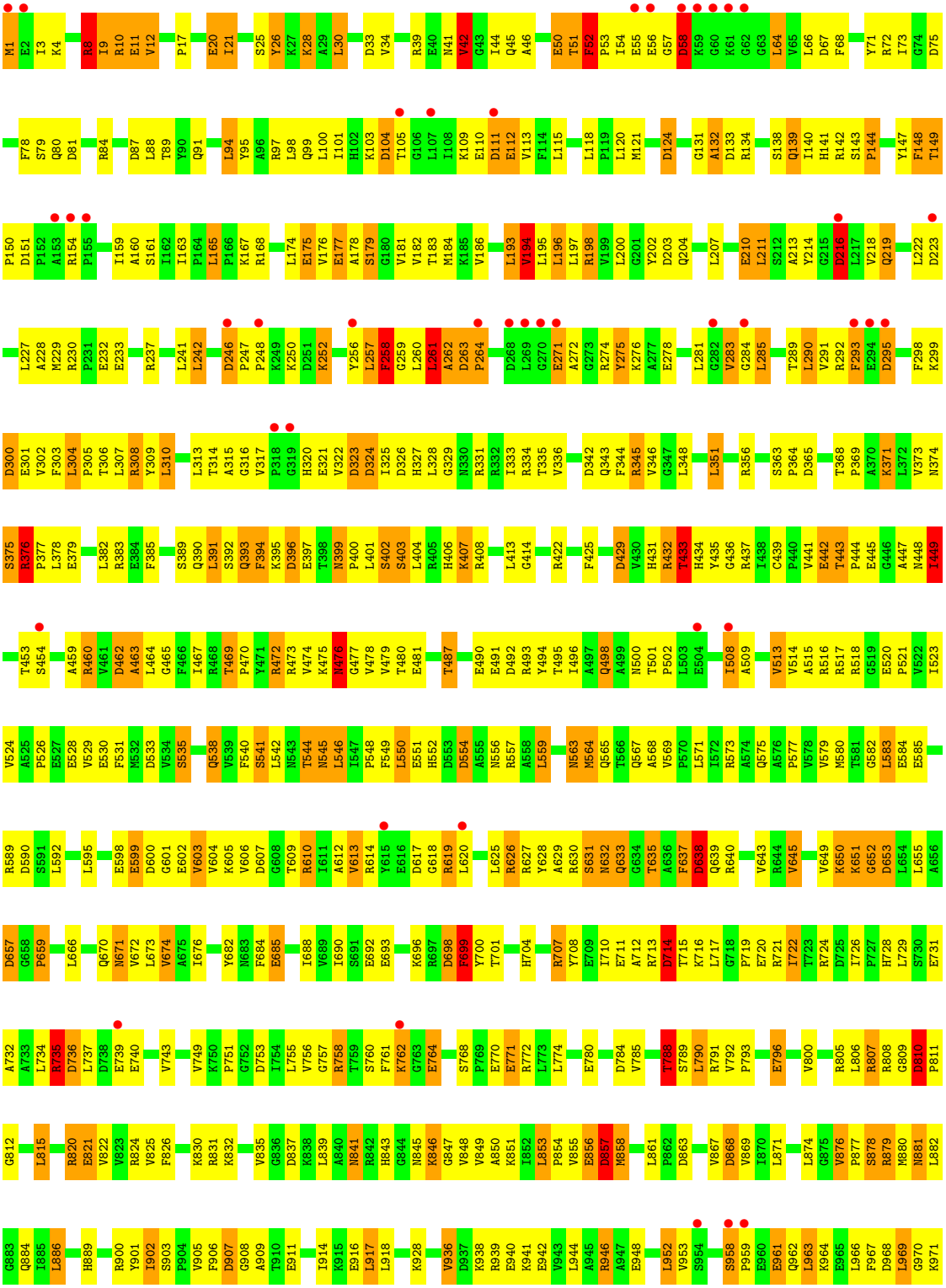
Chain B:

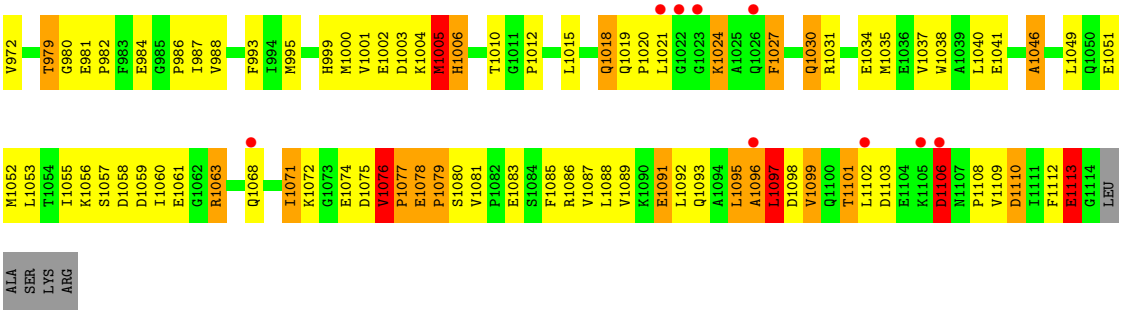




• Molecule 2: DNA-directed RNA polymerase beta chain

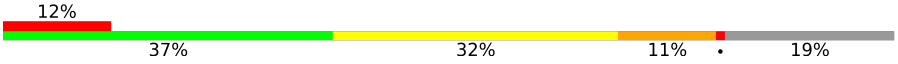
Chain C:

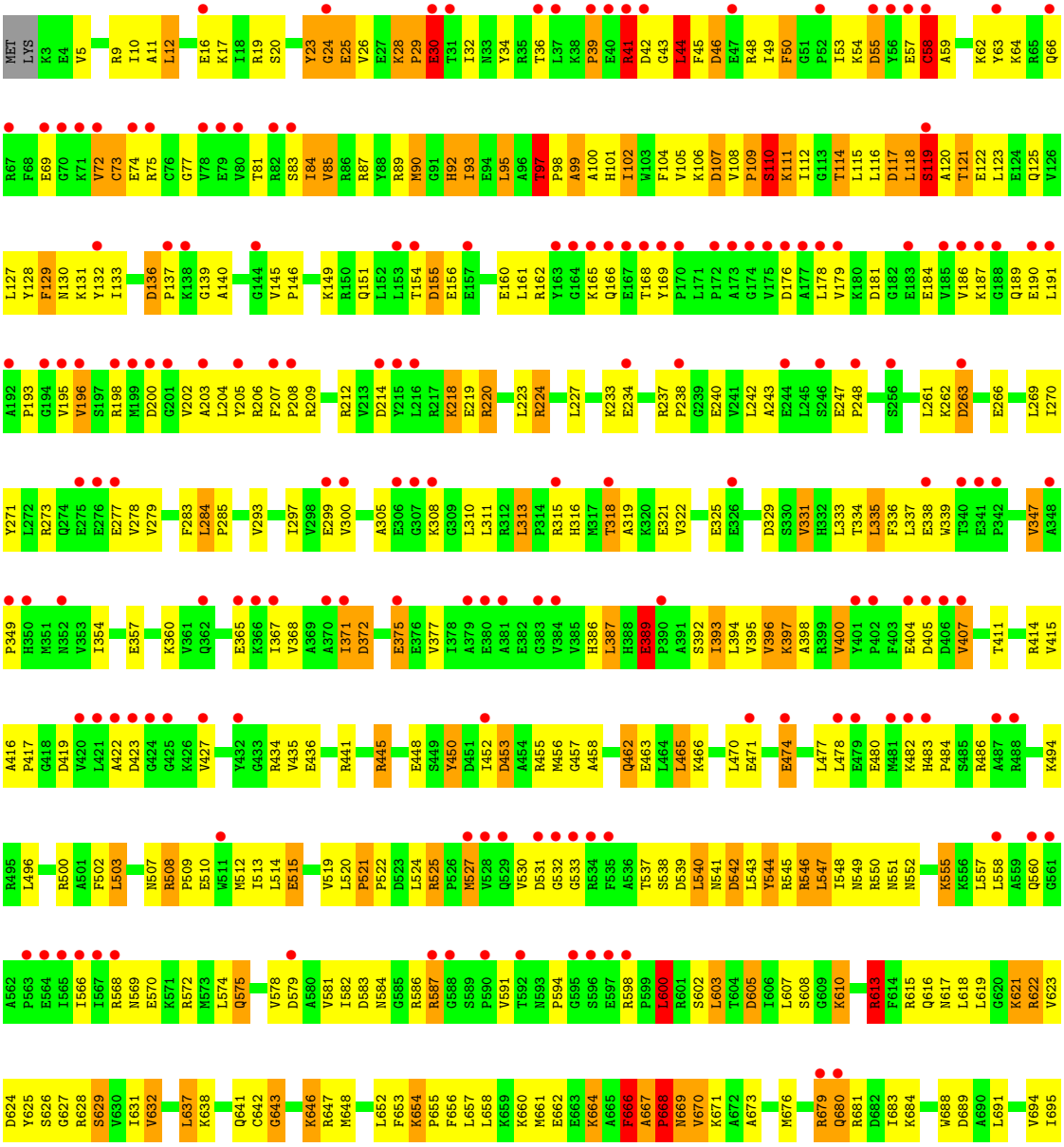


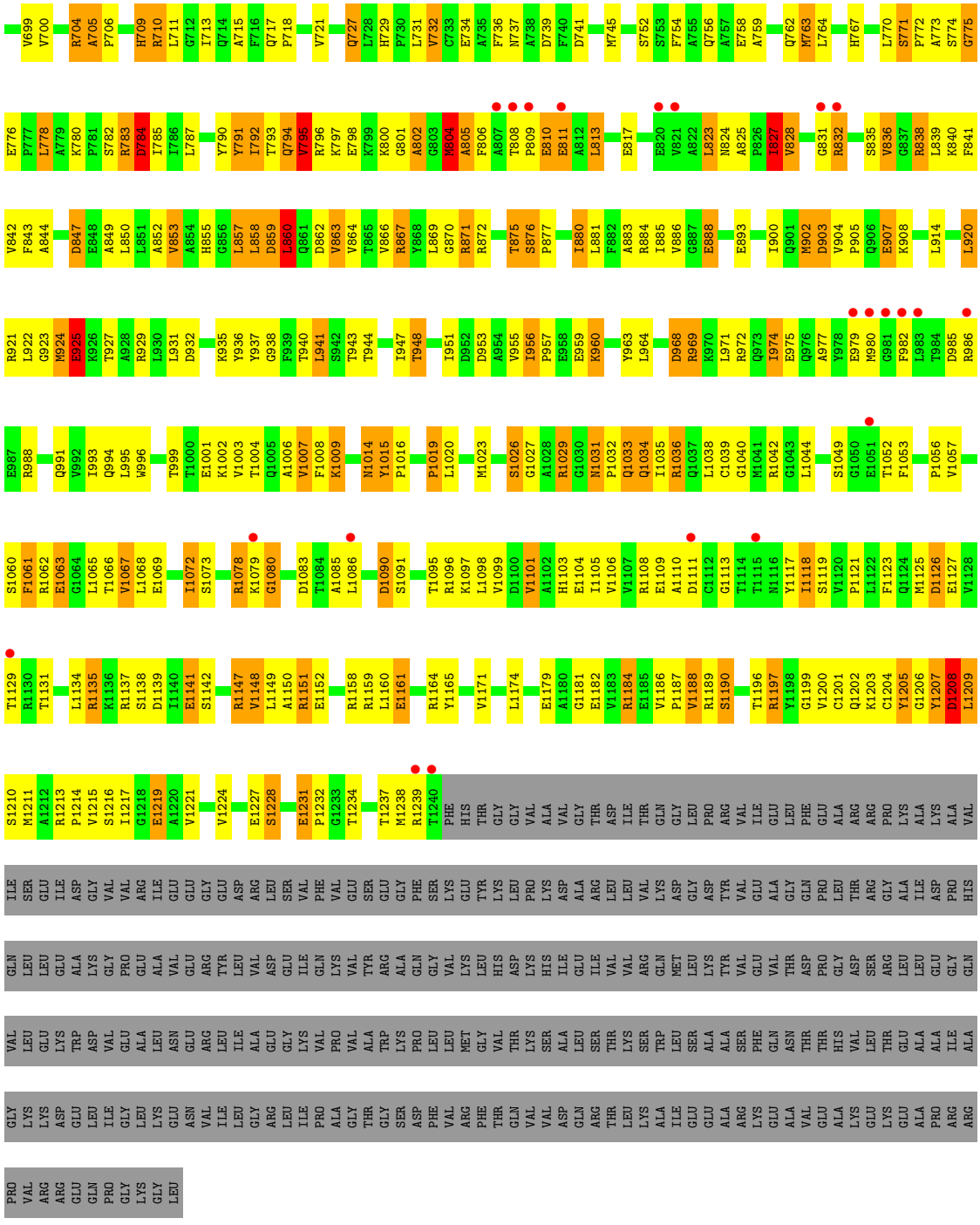


• Molecule 3: DNA-directed RNA polymerase beta' chain

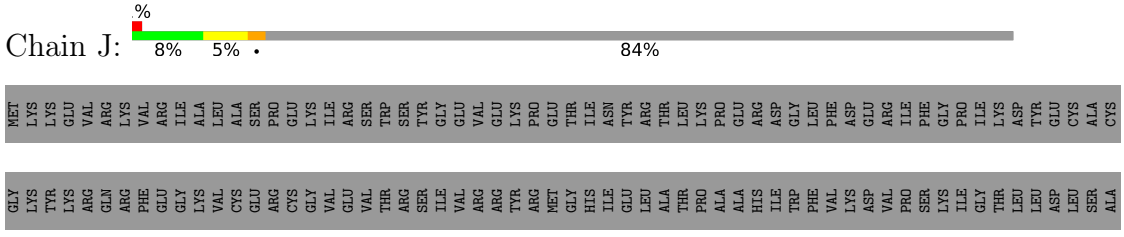
Chain D:

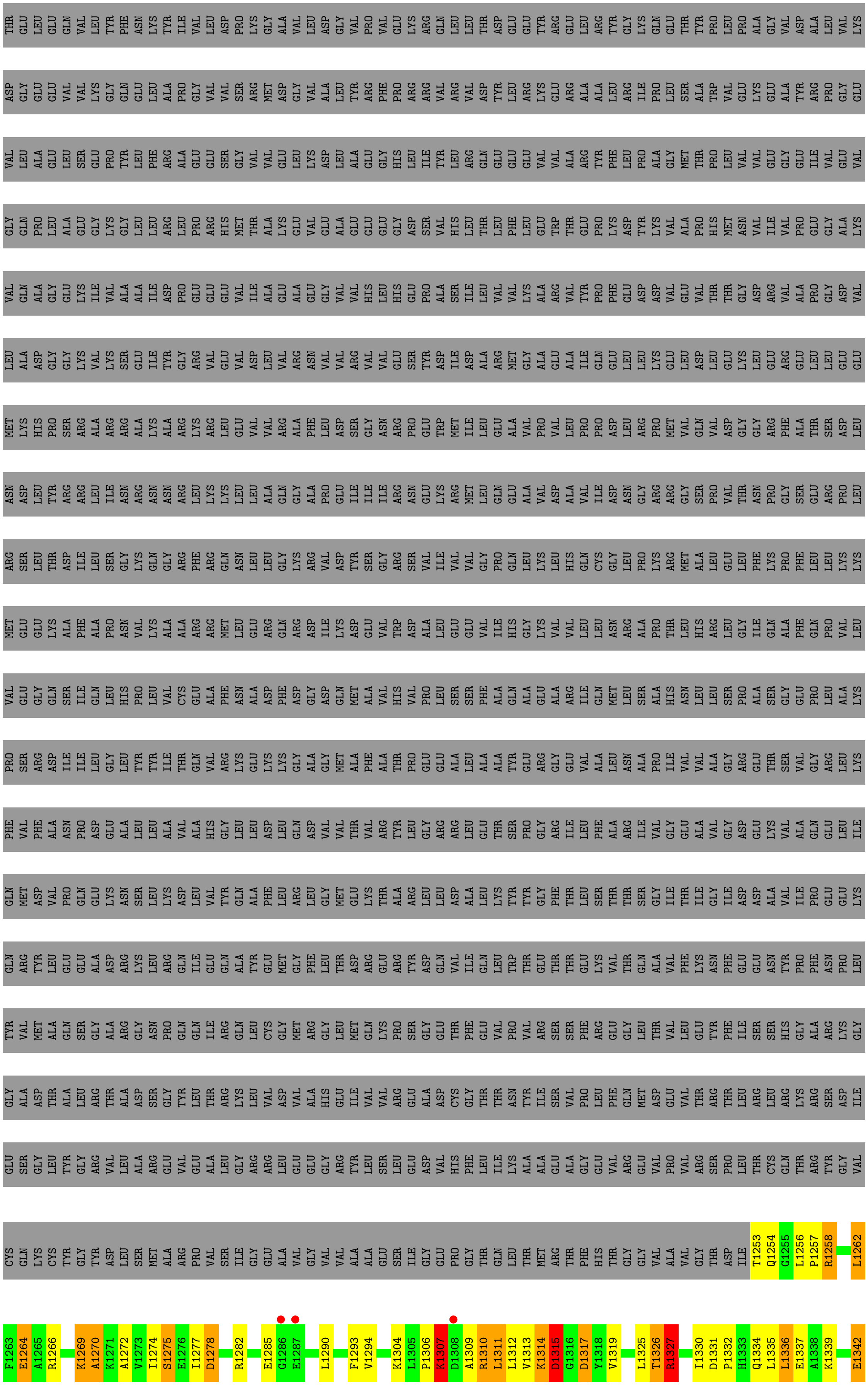


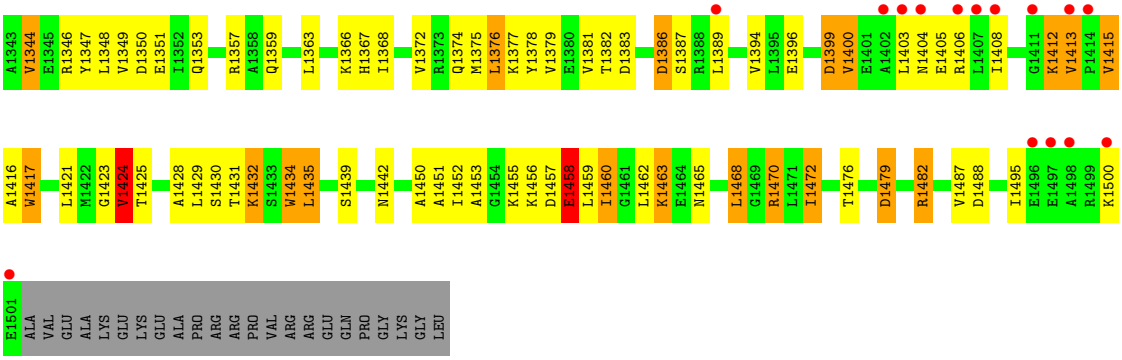




• Molecule 3: DNA-directed RNA polymerase beta' chain

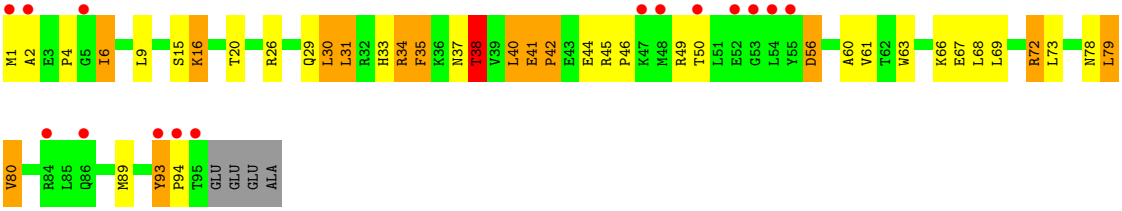



T1253
Q1254
G1255
L1256
P1257
R1258
L1262
F1263
E1264
A1265
R1266
K1269
A1270
K1271
A1272
V1273
I1274
S1275
E1276
I1277
D1278
R1282
E1285
G1286
E1287
L1290
F1293
V1294
K1304
L1305
P1306
K1307
D1308
A1309
R1310
L1311
L1312
V1313
K1314
D1315
G1316
D1317
Y1318
V1319
L1325
T1326
R1327
I1330
D1331
P1332
H1333
Q1334
L1335
L1336
E1337
A1338
K1339
E1342



• Molecule 4: DNA-directed RNA polymerase omega chain

Chain K:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	200.76Å 200.76Å 292.94Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 3.30 30.19 – 3.23	Depositor EDS
% Data completeness (in resolution range)	84.9 (30.00-3.30) 83.4 (30.19-3.23)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.73 (at 3.24Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.271 , 0.331 0.261 , 0.248	Depositor DCC
R_{free} test set	4075 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	87.9	Xtriage
Anisotropy	0.035	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 88.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.86	EDS
Total number of atoms	24368	wwPDB-VP
Average B, all atoms (Å^2)	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.76% 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

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

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	0.43	0/1798	0.84	11/2453 (0.4%)
1	B	0.41	0/1784	0.81	10/2428 (0.4%)
2	C	0.44	0/8742	0.82	41/11848 (0.3%)
3	D	0.43	0/9772	0.77	24/13234 (0.2%)
3	J	0.43	0/1897	0.75	6/2570 (0.2%)
4	K	0.45	0/762	0.74	1/1029 (0.1%)
All	All	0.43	0/24755	0.79	93/33562 (0.3%)

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
1	A	0	1
2	C	0	1
All	All	0	2

There are no bond length outliers.

The worst 5 of 93 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	75	ASP	CB-CG-OD2	7.29	124.86	118.30
2	C	1103	ASP	CB-CG-OD2	6.74	124.37	118.30
2	C	863	ASP	CB-CG-OD2	6.74	124.36	118.30
1	A	202	ASP	CB-CG-OD2	6.27	123.94	118.30
2	C	492	ASP	CB-CG-OD2	6.27	123.94	118.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	125	PRO	Peptide
2	C	51	THR	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	1763	0	1760	65	0
1	B	1750	0	1775	69	0
2	C	8576	0	8510	406	0
3	D	9602	0	9558	427	0
3	J	1869	0	1876	75	0
4	K	747	0	735	22	0
5	C	59	0	55	10	0
6	D	2	0	0	0	0
All	All	24368	0	24269	995	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 20.

The worst 5 of 995 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:D:136:ASP:HB3	3:D:137:PRO:HD3	1.14	1.10
2:C:432:ARG:O	2:C:433:THR:HB	1.49	1.09
2:C:263:ASP:HB3	2:C:264:PRO:HD3	1.12	1.09
1:B:26:GLU:HB3	1:B:27:PRO:HD3	1.29	1.09
2:C:650:LYS:HB3	2:C:653:ASP:HB2	1.35	1.07

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	228/314 (73%)	173 (76%)	38 (17%)	17 (8%)	1 7
1	B	223/314 (71%)	177 (79%)	31 (14%)	15 (7%)	1 8
2	C	1112/1119 (99%)	871 (78%)	153 (14%)	88 (8%)	1 6
3	D	1236/1524 (81%)	968 (78%)	187 (15%)	81 (7%)	1 8
3	J	247/1524 (16%)	194 (78%)	40 (16%)	13 (5%)	2 12
4	K	93/99 (94%)	74 (80%)	8 (9%)	11 (12%)	0 2
All	All	3139/4894 (64%)	2457 (78%)	457 (15%)	225 (7%)	1 7

5 of 225 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	90	LEU
1	A	161	ARG
1	A	203	GLY
1	B	18	ASP
1	B	26	GLU

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	189/270 (70%)	132 (70%)	57 (30%)	0 1
1	B	191/270 (71%)	130 (68%)	61 (32%)	0 1
2	C	889/936 (95%)	610 (69%)	279 (31%)	0 1
3	D	992/1281 (77%)	666 (67%)	326 (33%)	0 1
3	J	191/1281 (15%)	125 (65%)	66 (35%)	0 1
4	K	75/88 (85%)	59 (79%)	16 (21%)	1 4
All	All	2527/4126 (61%)	1722 (68%)	805 (32%)	0 1

5 of 805 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	263	ASP
3	D	670	VAL
4	K	69	LEU
3	D	313	LEU
3	D	261	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 53 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	845	ASN
3	D	507	ASN
3	D	1195	GLN
2	C	881	ASN
2	C	1018	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 3 ligands modelled in this entry, 2 are monoatomic - leaving 1 for Mogul analysis.

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	$\#\|Z\|>2$
5	RFP	C	1120	-	63,63,63	1.50	10 (15%)	94,94,94	2.35	20 (21%)

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
5	RFP	C	1120	-	-	12/60/85/85	0/5/5/5

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	C	1120	RFP	C43-N2	5.95	1.44	1.27
5	C	1120	RFP	O7-C35	4.43	1.45	1.35
5	C	1120	RFP	O2-C8	-3.20	1.24	1.35
5	C	1120	RFP	O5-C29	3.11	1.47	1.39
5	C	1120	RFP	N3-N2	-3.06	1.31	1.39

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	C	1120	RFP	C42-N4-C39	13.50	128.40	109.52
5	C	1120	RFP	C3-C43-N2	-7.93	110.01	121.54
5	C	1120	RFP	C41-N3-C40	5.83	128.93	113.74
5	C	1120	RFP	C40-C39-N4	4.80	116.23	110.80
5	C	1120	RFP	C38-N4-C39	4.23	116.98	110.66

There are no chirality outliers.

5 of 12 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	C	1120	RFP	C16-C15-N1-C2
5	C	1120	RFP	O11-C15-N1-C2
5	C	1120	RFP	C36-C35-O7-C25
5	C	1120	RFP	C43-N2-N3-C40
5	C	1120	RFP	C43-N2-N3-C41

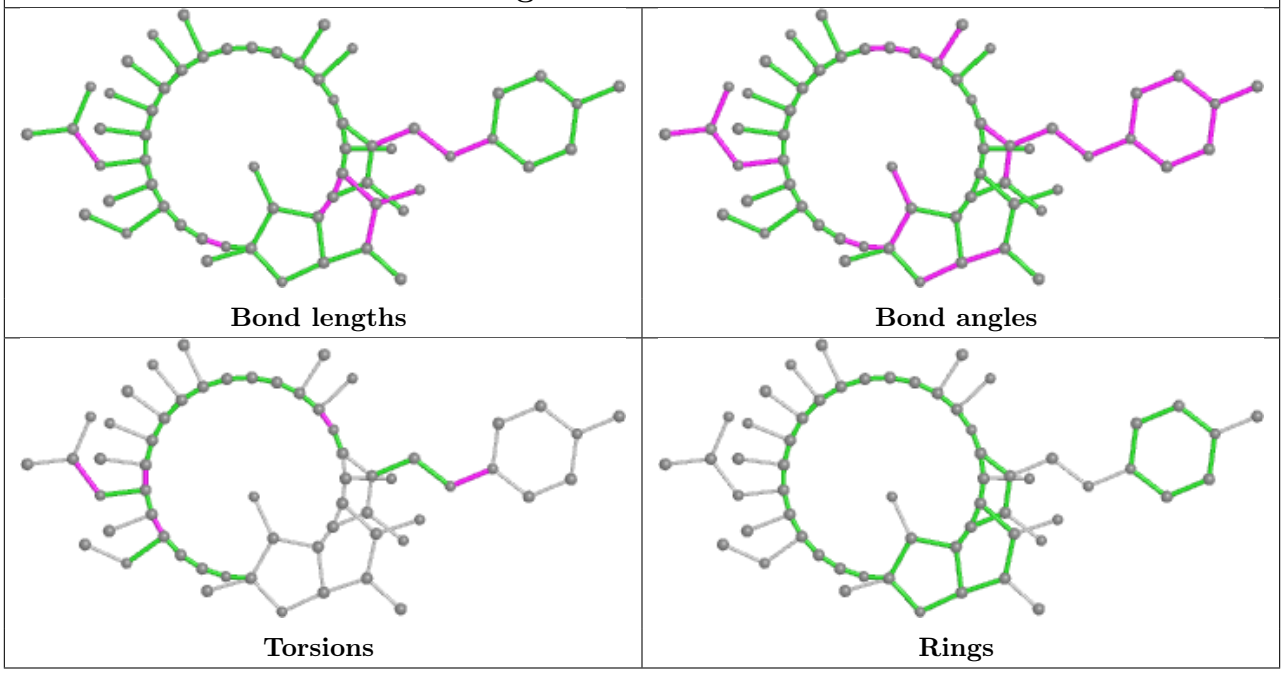
There are no ring outliers.

1 monomer is involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	C	1120	RFP	10	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand RFP C 1120



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, 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	230/314 (73%)	-0.03	14 (6%) 21 20	17, 38, 68, 98	0
1	B	225/314 (71%)	0.09	17 (7%) 13 13	24, 48, 73, 76	0
2	C	1114/1119 (99%)	0.02	51 (4%) 32 30	17, 44, 76, 93	0
3	D	1238/1524 (81%)	0.54	190 (15%) 2 2	15, 49, 87, 130	0
3	J	249/1524 (16%)	0.19	18 (7%) 15 15	17, 45, 114, 122	0
4	K	95/99 (95%)	0.51	15 (15%) 2 2	20, 59, 140, 154	0
All	All	3151/4894 (64%)	0.25	305 (9%) 7 8	15, 46, 85, 154	0

The worst 5 of 305 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	349	PRO	11.9
3	J	1403	LEU	11.7
3	D	57	GLU	11.1
3	D	256	SER	10.0
3	D	191	LEU	9.5

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

There are no monosaccharides in this entry.

6.4 Ligands (i)

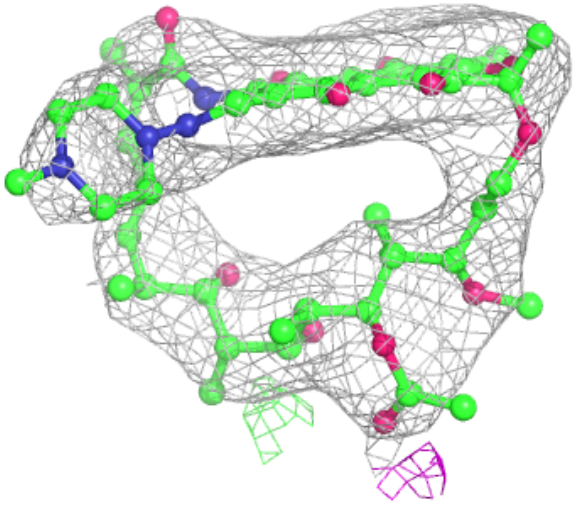
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

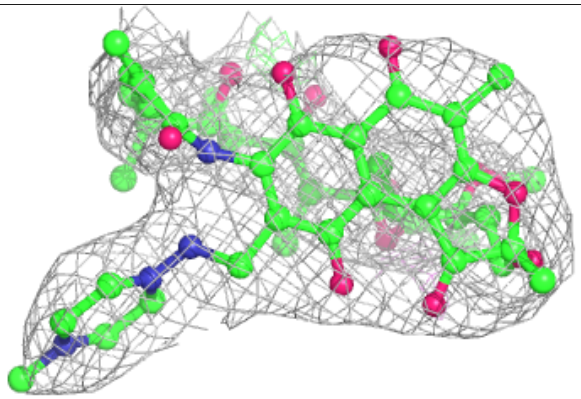
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
5	RFP	C	1120	59/59	0.92	0.26	44,47,60,63	0
6	ZN	D	1525	1/1	0.93	0.11	90,90,90,90	0
6	ZN	D	1526	1/1	0.96	0.18	52,52,52,52	0

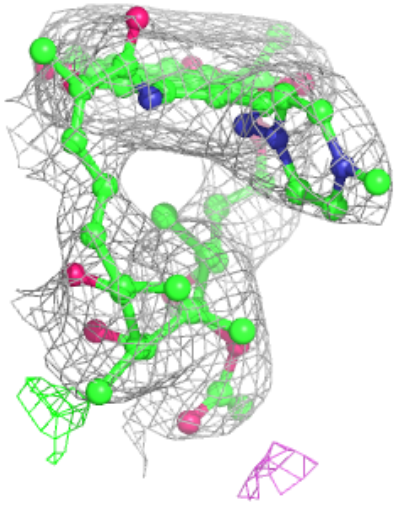
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around RFP C 1120:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)







6.5 Other polymers (i)

There are no such residues in this entry.