wwPDB X-ray Structure Validation Summary Report (i)

Jun 17, 2024 – 07:00 PM EDT

PDB ID	:	5TMF
Title	:	Re-refinement of thermus thermophilus RNA polymerase
Authors	:	Wang, J.
Deposited on	:	2016-10-12
Resolution	:	3.00 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:

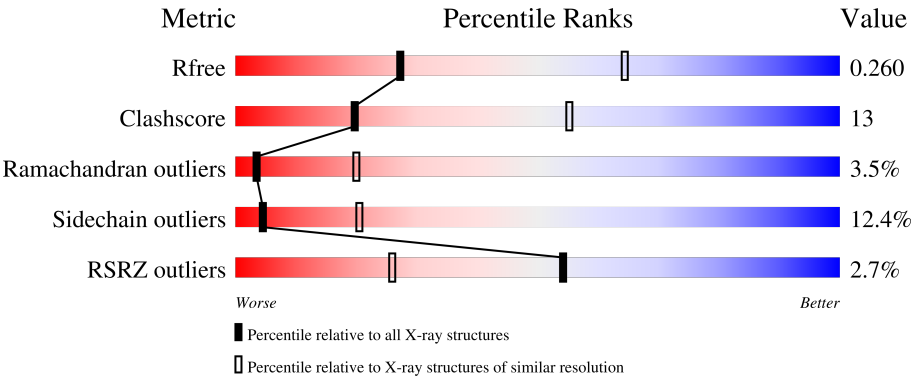
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 3.00 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.



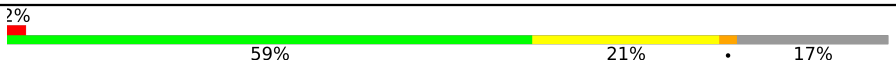
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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 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	315	53% 18% 27%
1	B	315	52% 22% 24%
2	C	1119	63% 29% 7%
3	D	1524	65% 29% 5%
4	E	99	72% 18% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	F	423	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
7	NE6	C	1201	X	-	-	-

2 Entry composition (i)

There are 10 unique types of molecules in this entry. The entry contains 28078 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	231	Total	C	N	O	S	0	0	0
			1816	1159	315	339	3			
1	B	238	Total	C	N	O	S	0	0	0
			1863	1188	322	350	3			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1119	Total	C	N	O	S	0	0	0
			8829	5581	1577	1647	24			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1504	Total	C	N	O	S	0	0	0
			11864	7518	2091	2219	36			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	95	Total	C	N	O	S	0	0	0
			769	488	133	144	4			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	61	GLU	VAL	conflict	UNP Q72ID6
E	92	ILE	LEU	conflict	UNP Q72ID6
E	95	GLY	VAL	conflict	UNP Q72ID6

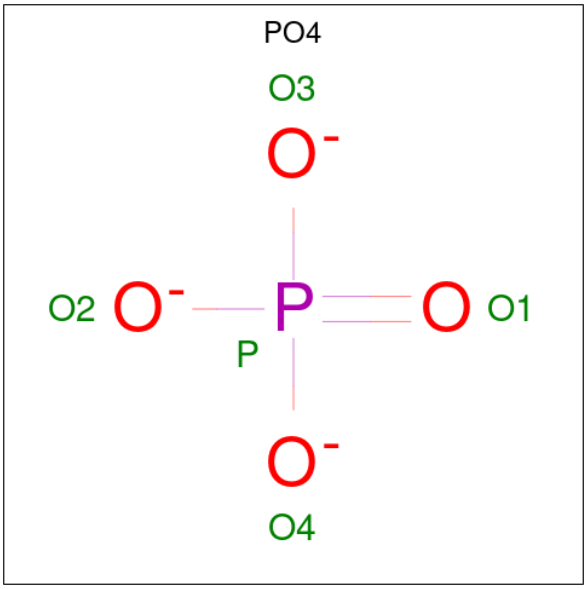
- Molecule 5 is a protein called RNA polymerase sigma factor SigA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	351	Total	C	N	O	S	0	0	0
			2844	1794	515	531	4			

There is a discrepancy between the modelled and reference sequences:

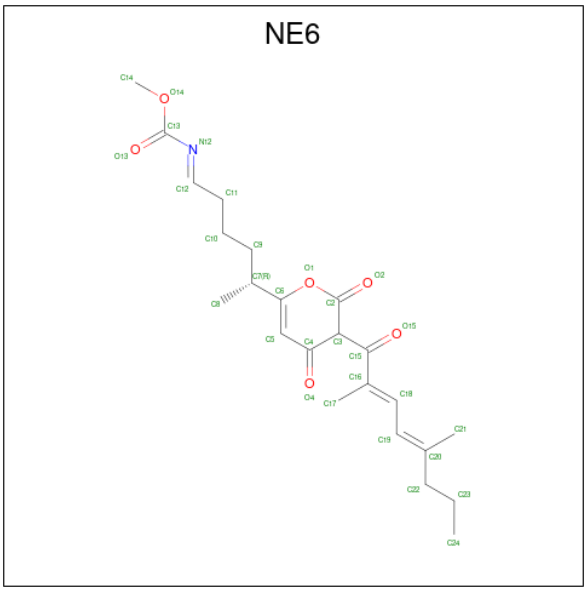
Chain	Residue	Modelled	Actual	Comment	Reference
F	46	THR	ALA	conflict	UNP Q72L95

- Molecule 6 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	O	P	0	0
			5	4	1		
6	D	1	Total	O	P	0	0
			5	4	1		

- Molecule 7 is methyl [(1E,5R)-5-{(3S)-3-[(2E,4E)-2,5-dimethylocta-2,4-dienoyl]-2,4-dioxo-3,4-dihydro-2H-pyran-6-yl}hexylidene]carbamate (three-letter code: NE6) (formula: $C_{23}H_{31}NO_6$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	C	1	Total C N O 30 23 1 6	0	0

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

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

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
9	D	2	Total Mg 2 2	0	0

- Molecule 10 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
10	A	2	Total O 2 2	0	0
10	B	3	Total O 3 3	0	0
10	C	23	Total O 23 23	0	0
10	D	20	Total O 20 20	0	0

Continued on next page...

Continued from previous page...

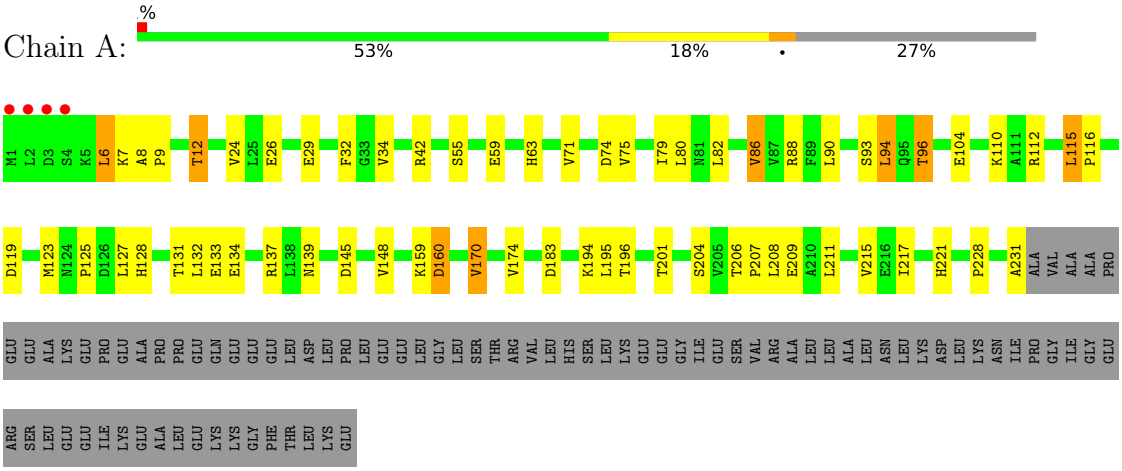
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
10	F	1	Total O 1 1	0	0

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.

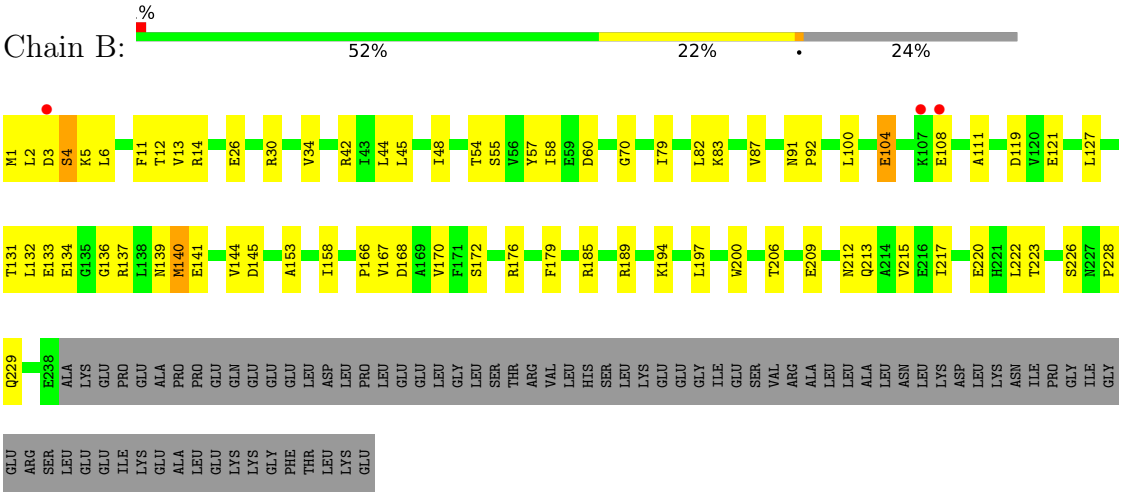
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain A:



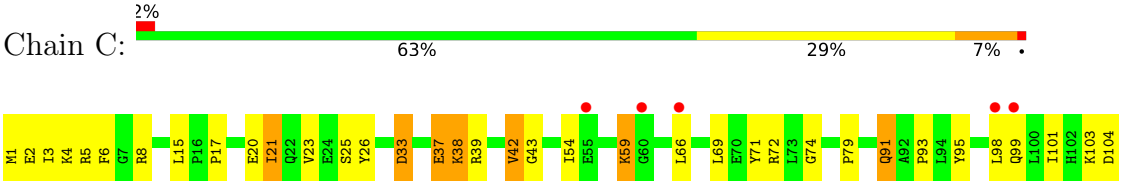
• Molecule 1: DNA-directed RNA polymerase subunit alpha

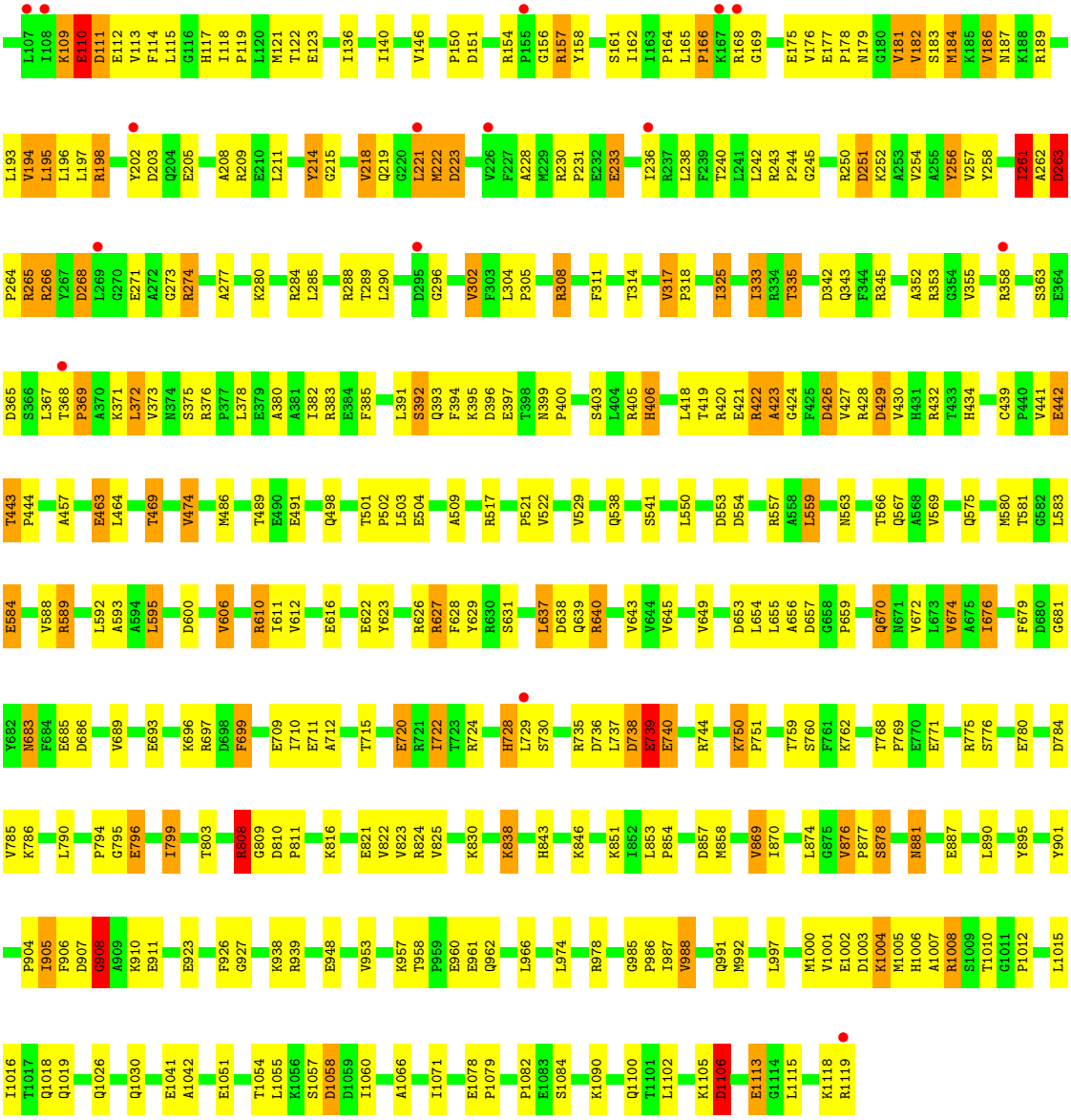
Chain B:



• Molecule 2: DNA-directed RNA polymerase subunit beta

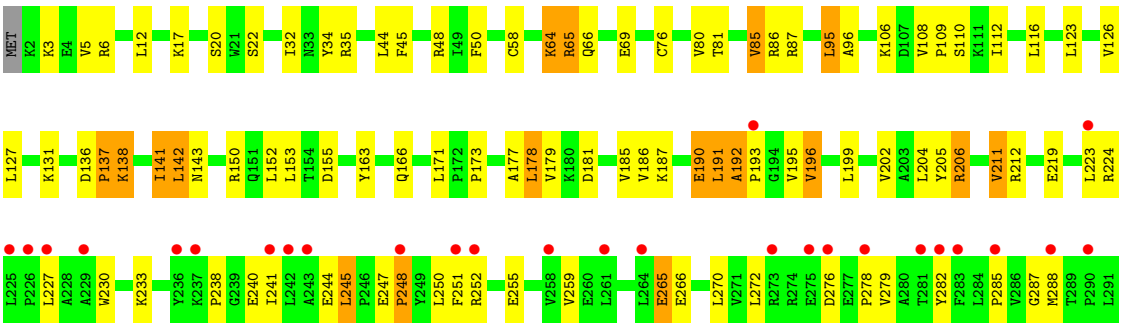
Chain C:

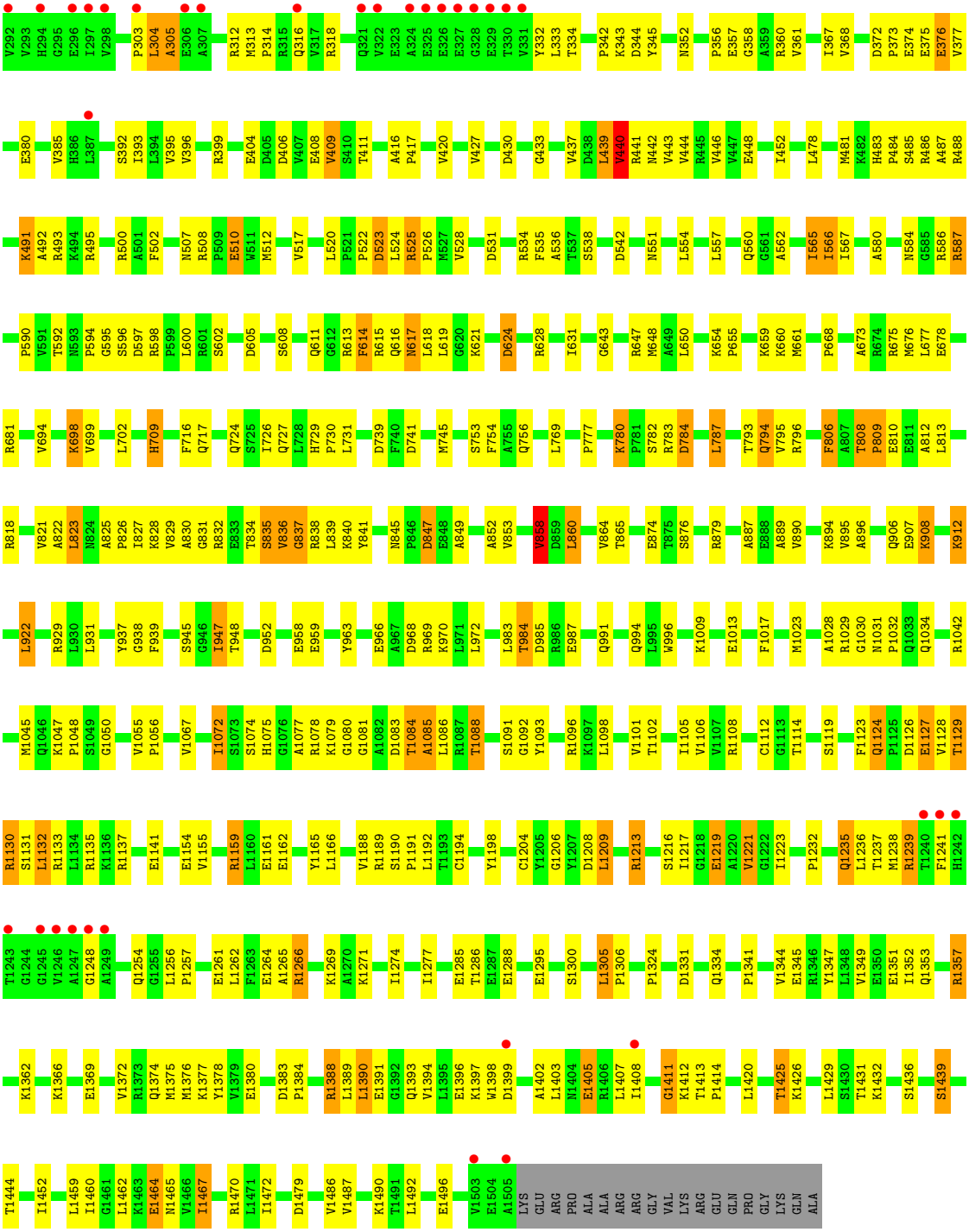




• Molecule 3: DNA-directed RNA polymerase subunit beta'

Chain D:



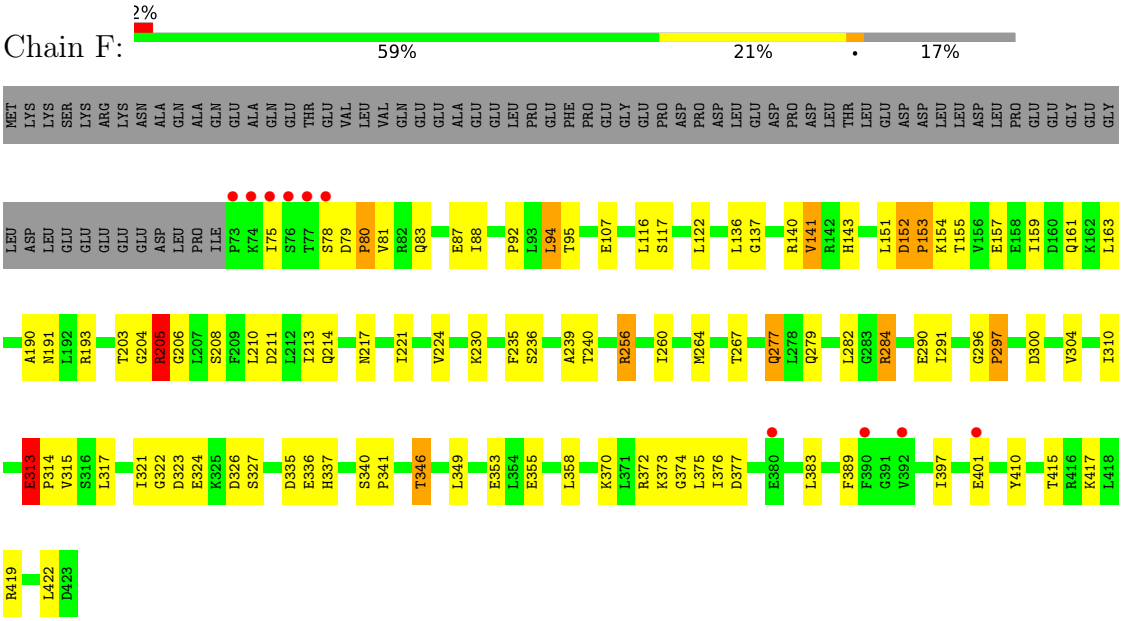


• Molecule 4: DNA-directed RNA polymerase subunit omega

Chain E:

• Molecule 5: RNA polymerase sigma factor SigA

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, α, β, γ	235.09Å 235.09Å 250.88Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	46.80 – 3.00 46.80 – 3.00	Depositor EDS
% Data completeness (in resolution range)	95.8 (46.80-3.00) 96.0 (46.80-3.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.52 (at 3.01Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.203 , 0.234 0.207 , 0.260	Depositor DCC
R_{free} test set	1177 reflections (0.78%)	wwPDB-VP
Wilson B-factor (Å^2)	83.9	Xtriage
Anisotropy	0.045	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.23 , 63.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	0.048 for h,-h-k,-l	Xtriage
Reported twinning fraction	0.956 for H, K, L 0.044 for -H-K, K, -L	Depositor
Outliers	0 of 150767 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	28078	wwPDB-VP
Average B, all atoms (Å^2)	129.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.60% 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: NE6, ZN, MG, PO4

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	0.44	0/1848	0.62	0/2512
1	B	0.42	0/1896	0.59	0/2579
2	C	0.47	0/8997	0.66	1/12164 (0.0%)
3	D	0.48	0/12073	0.66	3/16324 (0.0%)
4	E	0.46	0/783	0.63	1/1054 (0.1%)
5	F	0.42	0/2890	0.62	1/3888 (0.0%)
All	All	0.46	0/28487	0.65	6/38521 (0.0%)

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	E	51	LEU	CA-CB-CG	5.89	128.85	115.30
3	D	837	GLY	N-CA-C	-5.79	98.63	113.10
3	D	95	LEU	CA-CB-CG	5.68	128.36	115.30
3	D	618	LEU	CA-CB-CG	5.51	127.98	115.30
5	F	313	GLU	C-N-CD	-5.11	109.37	120.60

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	A	1816	0	1871	41	0
1	B	1863	0	1914	41	0
2	C	8829	0	8933	275	0
3	D	11864	0	12094	339	0
4	E	769	0	775	15	0
5	F	2844	0	2926	56	0
6	A	5	0	0	0	0
6	D	5	0	0	0	0
7	C	30	0	30	5	0
8	D	2	0	0	0	0
9	D	2	0	0	0	0
10	A	2	0	0	0	0
10	B	3	0	0	0	0
10	C	23	0	0	3	0
10	D	20	0	0	4	0
10	F	1	0	0	1	0
All	All	28078	0	28543	717	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 13.

The worst 5 of 717 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:1265:ALA:HA	3:D:1266:ARG:CB	1.72	1.19
3:D:613:ARG:HB3	3:D:614:PHE:HA	1.21	1.18
2:C:93:PRO:HA	2:C:117:HIS:HB2	1.31	1.06
2:C:93:PRO:HA	2:C:117:HIS:CB	1.85	1.06
3:D:1085:ALA:HA	3:D:1088:THR:HG22	1.39	1.04

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	229/315 (73%)	211 (92%)	17 (7%)	1 (0%)	34 72
1	B	236/315 (75%)	218 (92%)	16 (7%)	2 (1%)	19 57
2	C	1117/1119 (100%)	928 (83%)	151 (14%)	38 (3%)	3 20
3	D	1502/1524 (99%)	1228 (82%)	208 (14%)	66 (4%)	2 15
4	E	93/99 (94%)	71 (76%)	16 (17%)	6 (6%)	1 7
5	F	349/423 (82%)	302 (86%)	36 (10%)	11 (3%)	4 22
All	All	3526/3795 (93%)	2958 (84%)	444 (13%)	124 (4%)	3 20

5 of 124 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	42	VAL
2	C	194	VAL
2	C	195	LEU
2	C	223	ASP
2	C	261	ILE

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	202/273 (74%)	183 (91%)	19 (9%)	8 32
1	B	206/273 (76%)	192 (93%)	14 (7%)	16 48
2	C	941/941 (100%)	802 (85%)	139 (15%)	3 14
3	D	1264/1279 (99%)	1106 (88%)	158 (12%)	4 20
4	E	83/87 (95%)	76 (92%)	7 (8%)	11 38
5	F	306/371 (82%)	272 (89%)	34 (11%)	6 25
All	All	3002/3224 (93%)	2631 (88%)	371 (12%)	4 20

5 of 371 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	523	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	D	1124	GLN
3	D	597	ASP
3	D	832	ARG
3	D	1219	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 39 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	1034	GLN
5	F	83	GLN
3	D	1124	GLN
3	D	1465	ASN
5	F	277	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 7 ligands modelled in this entry, 4 are monoatomic - leaving 3 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	$\#\|Z\|>2$
6	PO4	A	401	-	4,4,4	0.85	0	6,6,6	0.66	0
6	PO4	D	1603	-	4,4,4	0.92	0	6,6,6	0.30	0
7	NE6	C	1201	-	29,30,30	3.76	8 (27%)	27,39,39	2.34	5 (18%)

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
7	NE6	C	1201	-	1/1/9/13	11/26/46/46	0/1/1/1

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	C	1201	NE6	C11-C12	-16.20	1.33	1.49
7	C	1201	NE6	O4-C4	9.04	1.36	1.22
7	C	1201	NE6	O14-C13	4.55	1.44	1.33
7	C	1201	NE6	C5-C4	-3.53	1.39	1.46
7	C	1201	NE6	C3-C15	-2.47	1.48	1.53

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
7	C	1201	NE6	O14-C13-N12	8.47	119.82	108.77
7	C	1201	NE6	O1-C2-O2	6.44	123.11	116.94
7	C	1201	NE6	C21-C20-C22	2.60	119.65	115.27
7	C	1201	NE6	O14-C13-O13	-2.60	115.92	122.36
7	C	1201	NE6	C2-O1-C6	2.17	123.91	120.67

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
7	C	1201	NE6	C3

5 of 11 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
7	C	1201	NE6	O1-C6-C7-C8
7	C	1201	NE6	O1-C6-C7-C9
7	C	1201	NE6	C9-C10-C11-C12

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
7	C	1201	NE6	C3-C15-C16-C17
7	C	1201	NE6	O15-C15-C16-C17

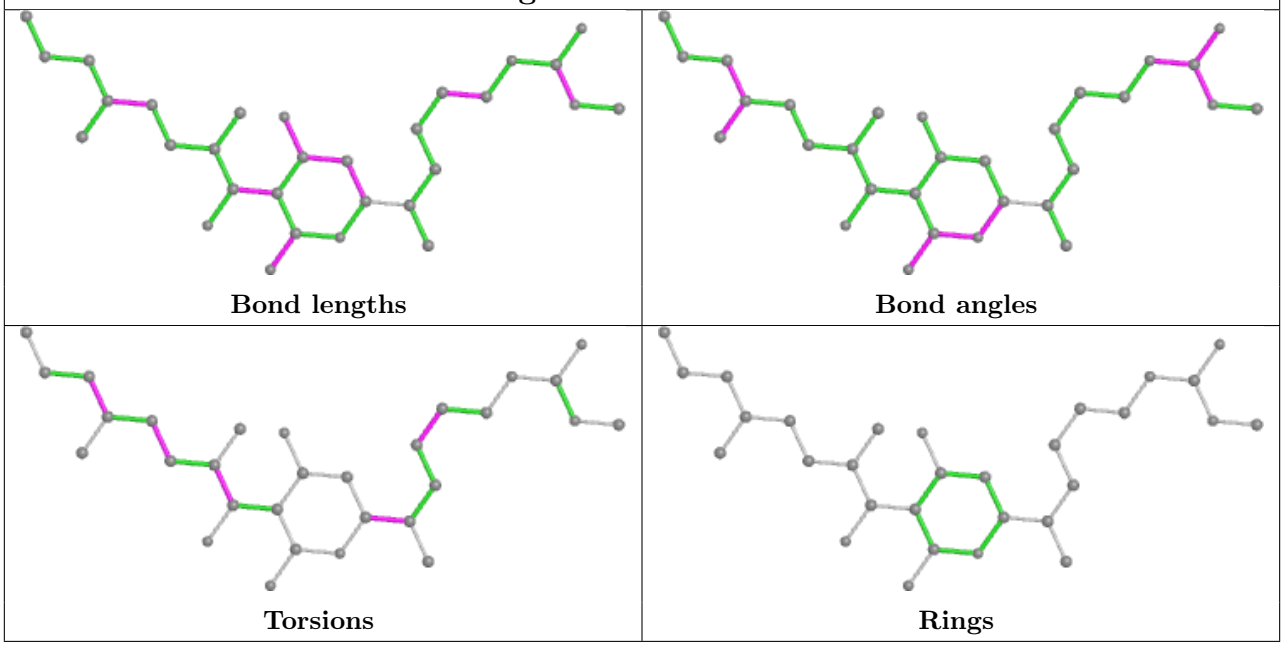
There are no ring outliers.

1 monomer is involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	C	1201	NE6	5	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 NE6 C 1201



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	231/315 (73%)	-0.53	4 (1%) 70 41	67, 96, 155, 259	0
1	B	238/315 (75%)	-0.40	3 (1%) 77 51	82, 145, 226, 248	0
2	C	1119/1119 (100%)	-0.36	20 (1%) 68 40	56, 104, 253, 323	0
3	D	1504/1524 (98%)	-0.26	60 (3%) 38 15	56, 106, 251, 301	0
4	E	95/99 (95%)	-0.44	0 100 100	82, 129, 203, 210	0
5	F	351/423 (82%)	-0.29	10 (2%) 53 25	78, 129, 236, 277	0
All	All	3538/3795 (93%)	-0.33	97 (2%) 54 26	56, 111, 246, 323	0

The worst 5 of 97 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	328	GLY	9.4
3	D	1248	GLY	7.4
3	D	243	ALA	6.1
3	D	297	ILE	6.0
3	D	329	GLU	5.9

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)

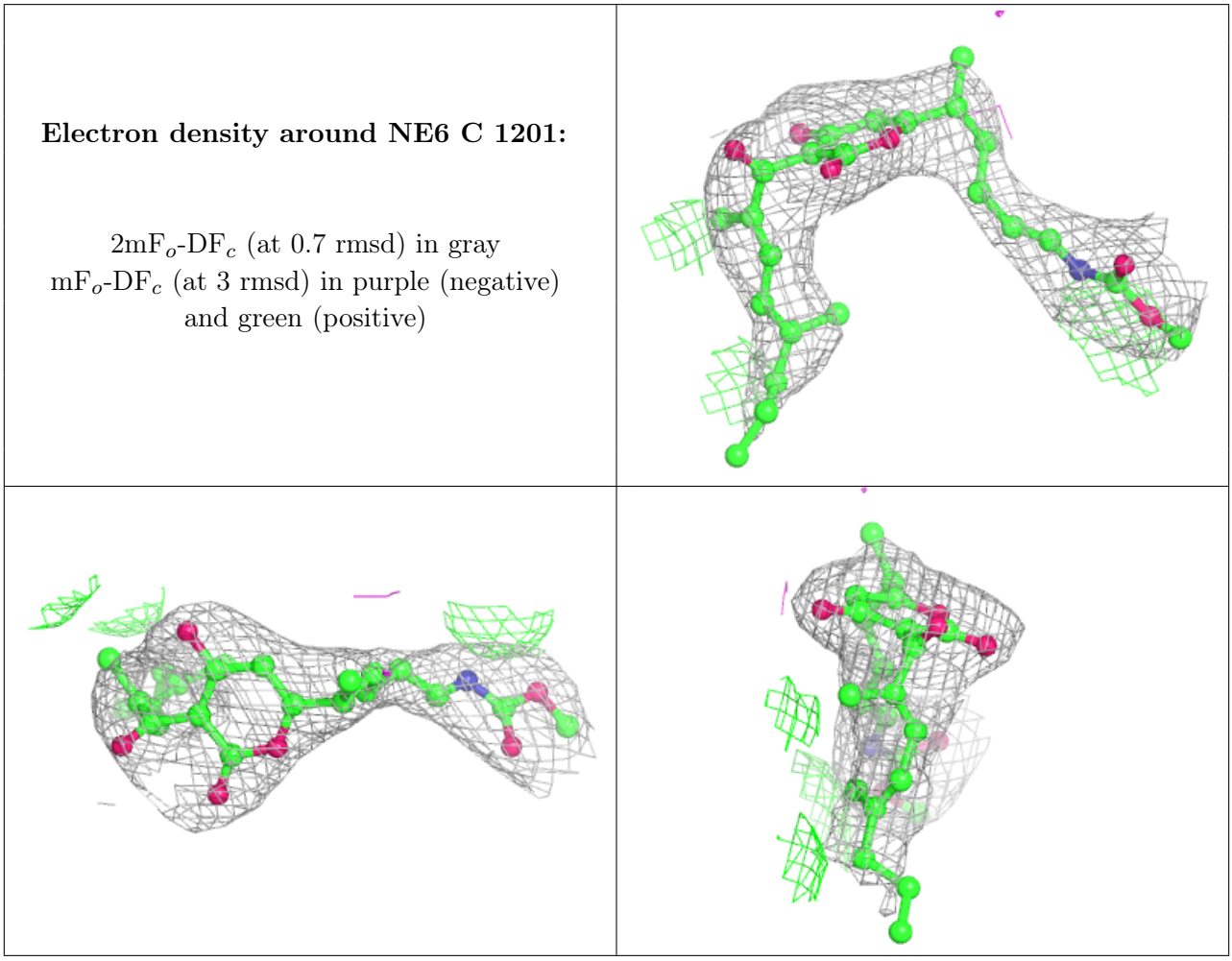
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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
6	PO4	D	1603	5/5	0.80	0.23	143,144,156,163	0
9	MG	D	1604	1/1	0.91	0.28	101,101,101,101	0
7	NE6	C	1201	30/30	0.96	0.21	76,92,105,115	0
6	PO4	A	401	5/5	0.98	0.23	122,124,129,130	0
9	MG	D	1605	1/1	0.98	0.27	98,98,98,98	0
8	ZN	D	1601	1/1	0.99	0.13	104,104,104,104	0
8	ZN	D	1602	1/1	1.00	0.15	84,84,84,84	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 NE6 C 1201:

$2mF_o$-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.