



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 22, 2024 – 09:36 PM EDT

PDB ID : 4XLP
Title : Crystal structure of T.aquaticus transcription initiation complex containing upstream fork promoter
Authors : Bae, B.; Darst, S.A.
Deposited on : 2015-01-13
Resolution : 4.00 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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
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

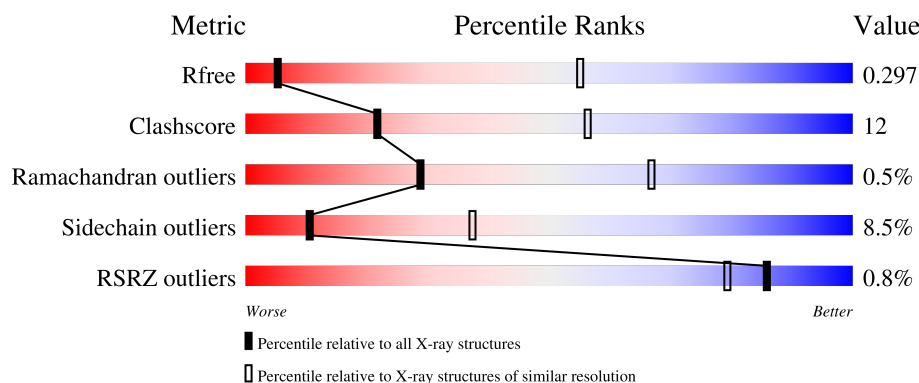
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION






The reported resolution of this entry is 4.00 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	1087 (4.30-3.70)
Clashscore	141614	1148 (4.30-3.70)
Ramachandran outliers	138981	1108 (4.30-3.70)
Sidechain outliers	138945	1099 (4.30-3.70)
RSRZ outliers	127900	1028 (4.34-3.66)

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 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 $\leq 5\%$. 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	
1	B	314	
1	G	314	
1	H	314	
2	C	1119	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	I	1119	
3	D	1524	
3	J	1524	
4	E	99	
4	K	99	
5	F	347	
5	L	347	
6	O	30	
6	R	30	
7	P	25	
7	S	25	

2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 56454 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 (25-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	P	25	Total	C	N	O	P	0	0	0
			507	245	91	147	24			
7	S	25	Total	C	N	O	P	0	0	0
			507	245	91	147	24			

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

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

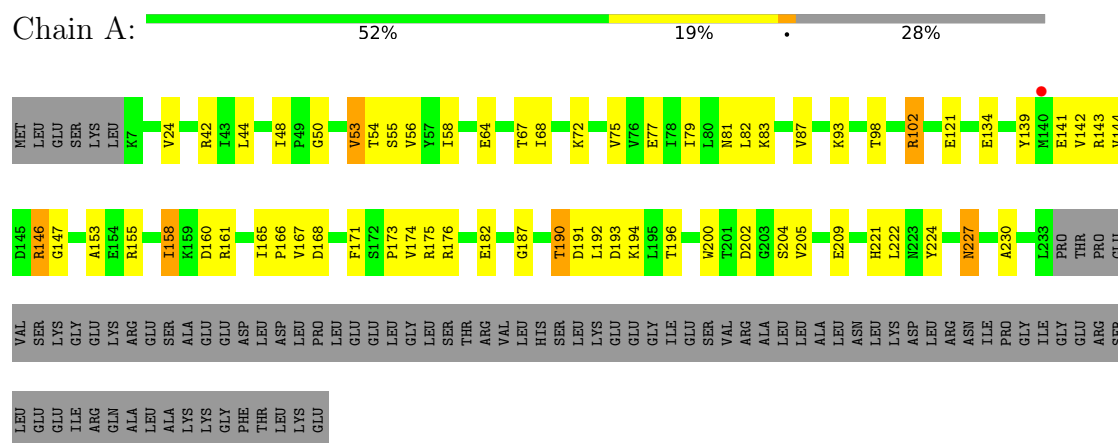
- Molecule 9 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

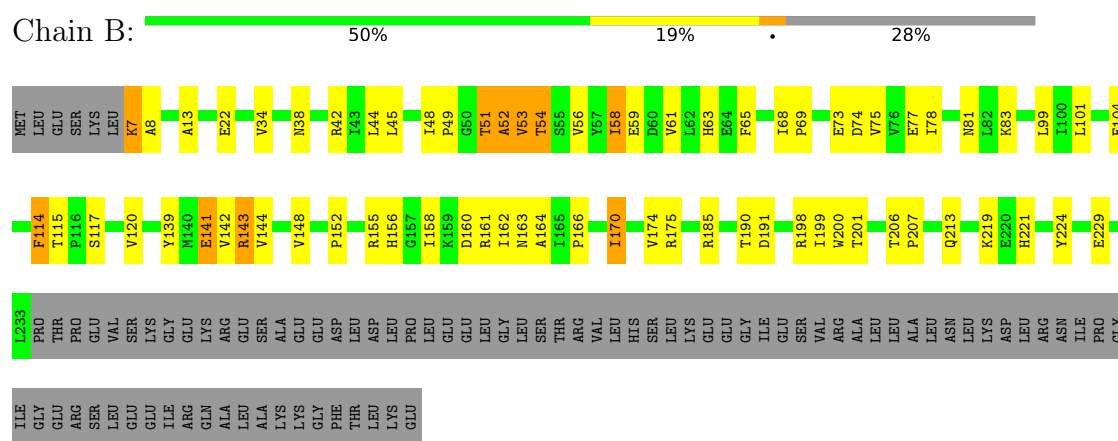
3 Residue-property plots

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.

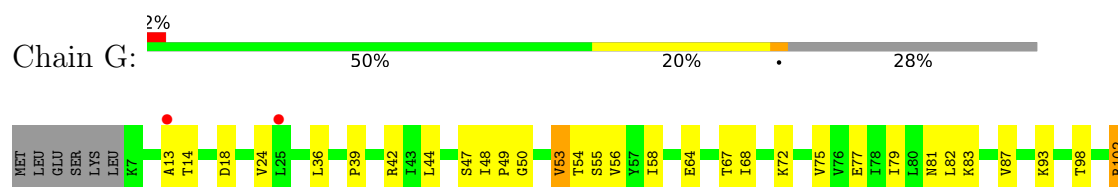
• Molecule 1: DNA-directed RNA polymerase subunit alpha

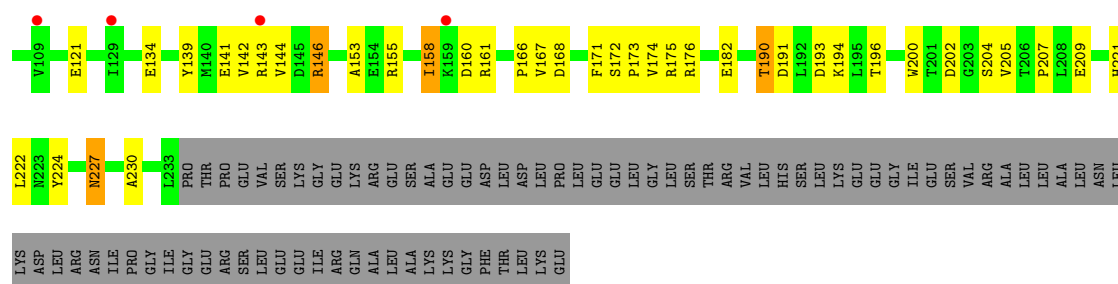


• Molecule 1: DNA-directed RNA polymerase subunit alpha

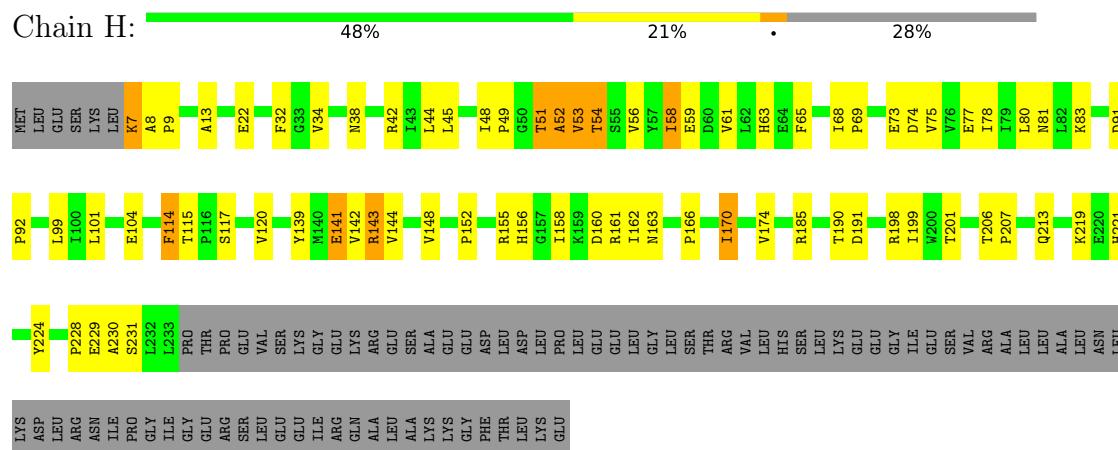


• Molecule 1: DNA-directed RNA polymerase subunit alpha

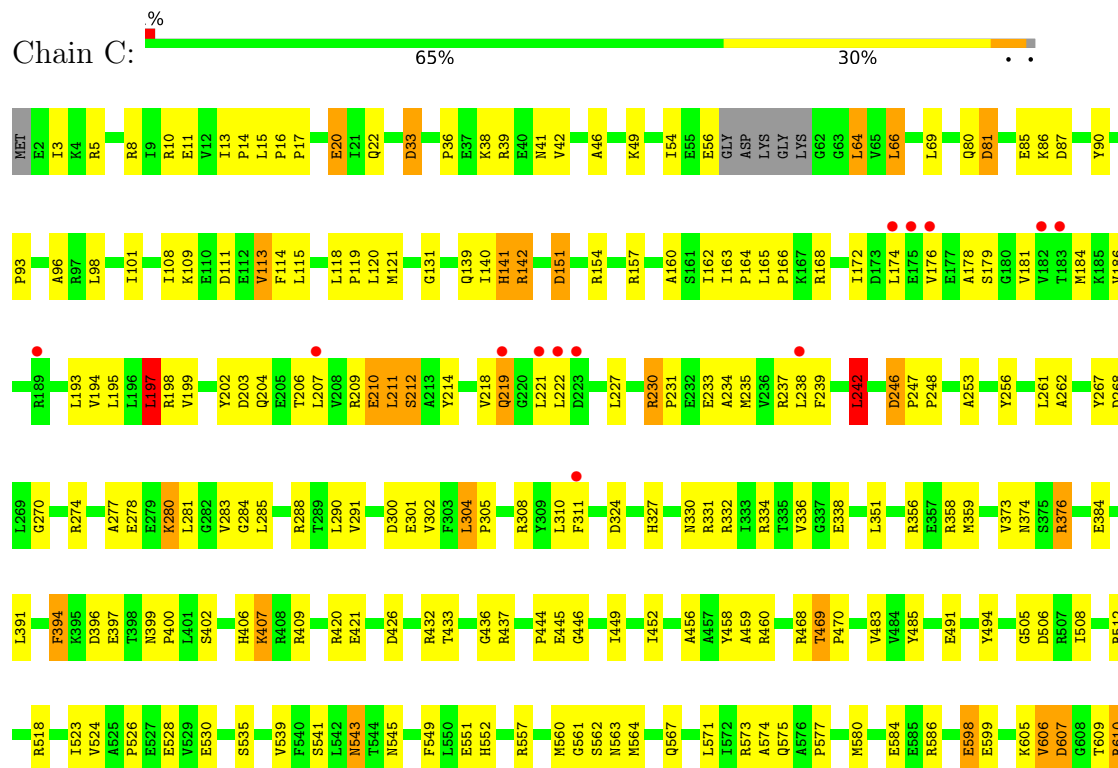


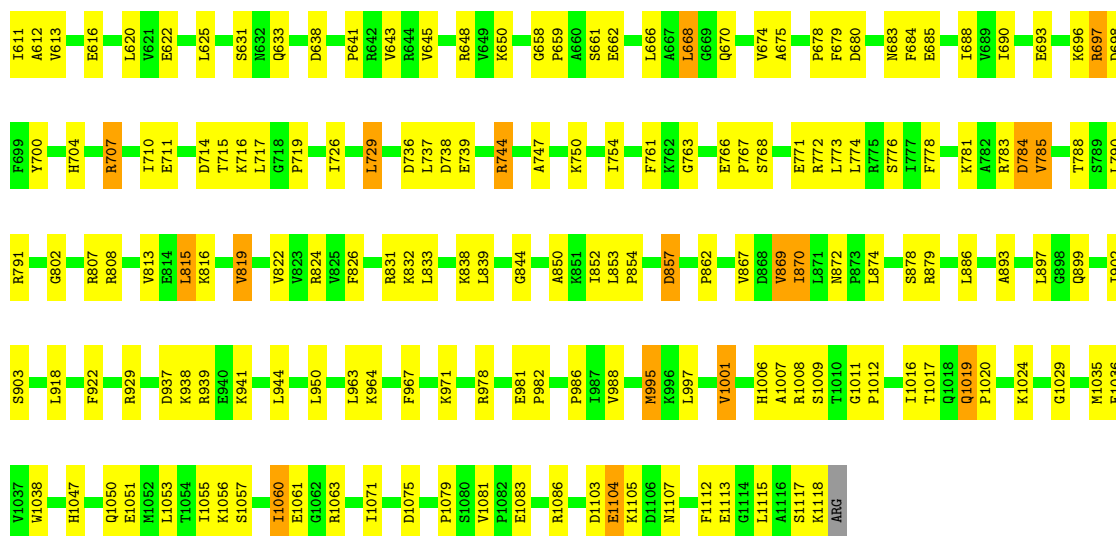


• Molecule 1: DNA-directed RNA polymerase subunit alpha

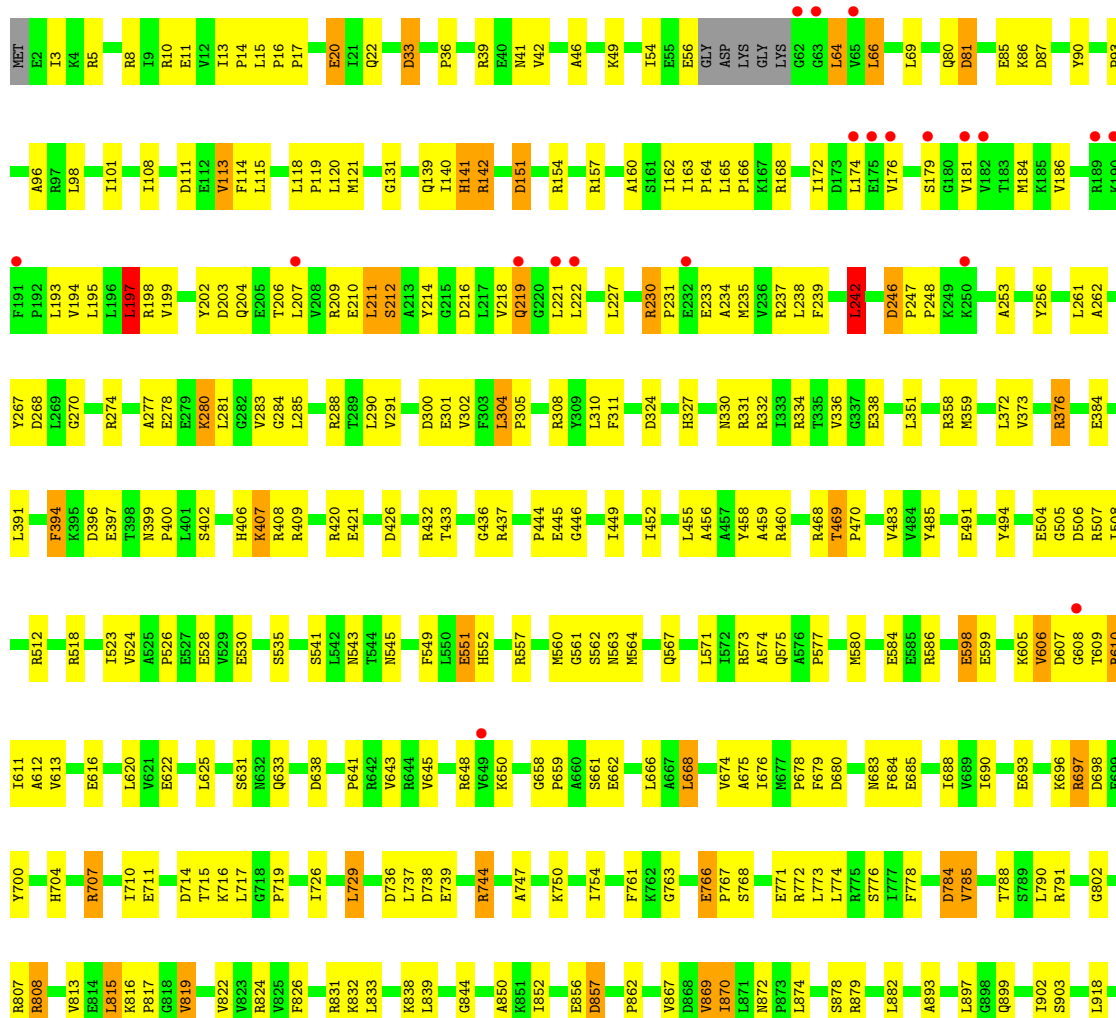


• Molecule 2: DNA-directed RNA polymerase subunit beta

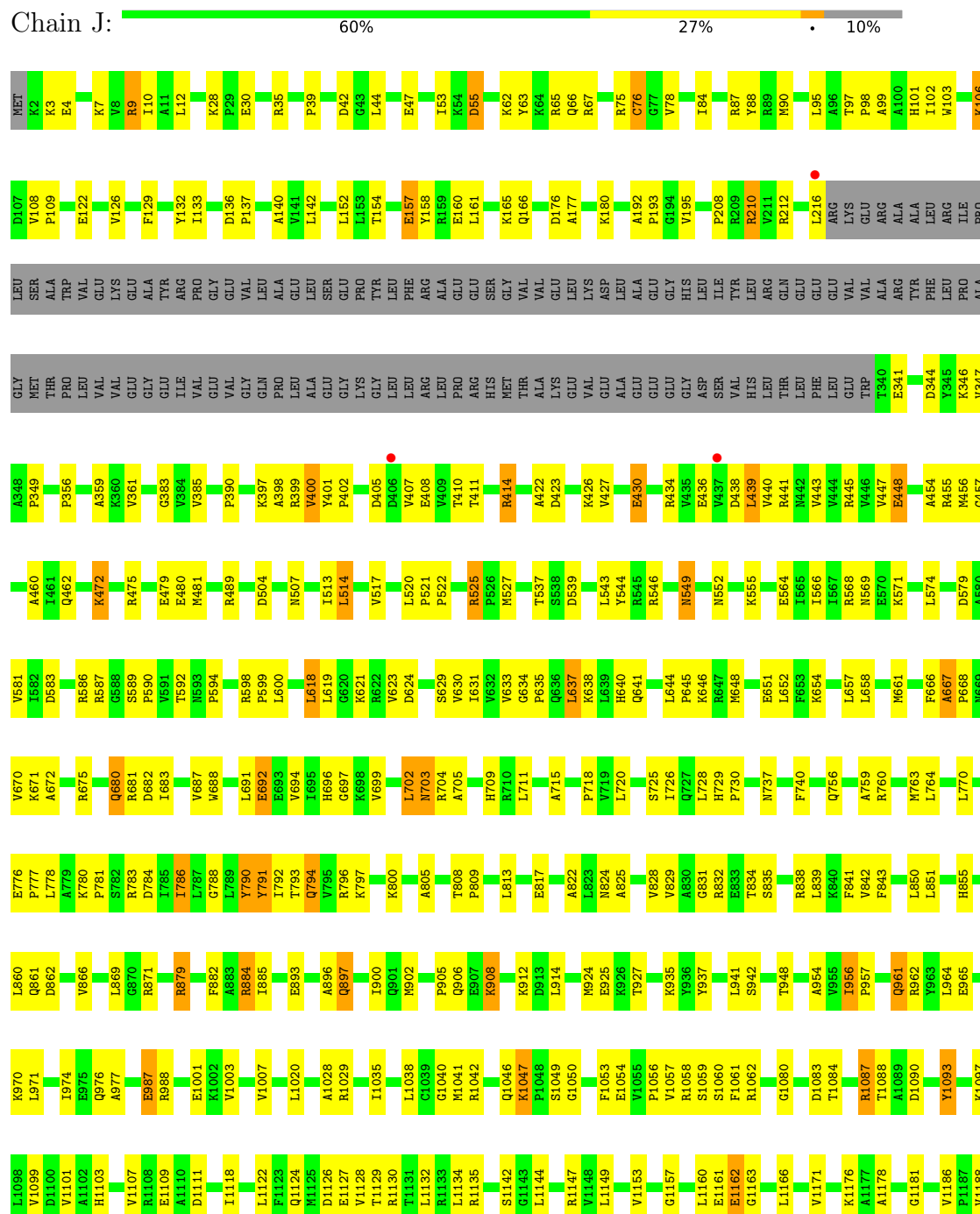


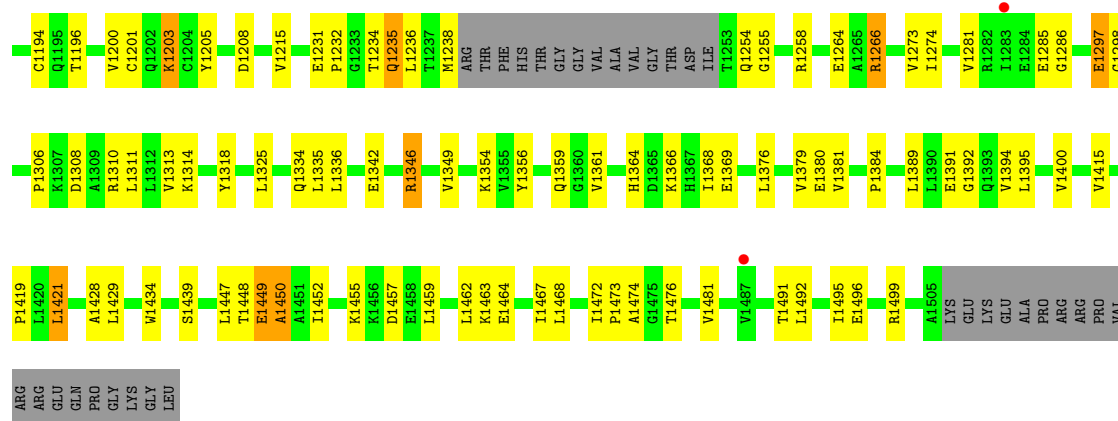


• Molecule 2: DNA-directed RNA polymerase subunit beta









- Molecule 4: DNA-directed RNA polymerase subunit omega

Chain E: 67% 25% 6%



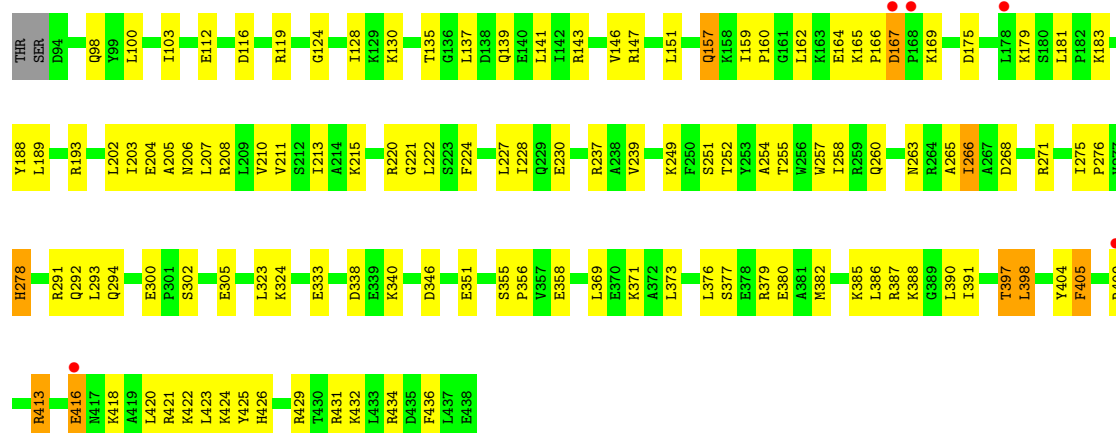
- Molecule 4: DNA-directed RNA polymerase subunit omega

Chain K: 67% 26% 6%



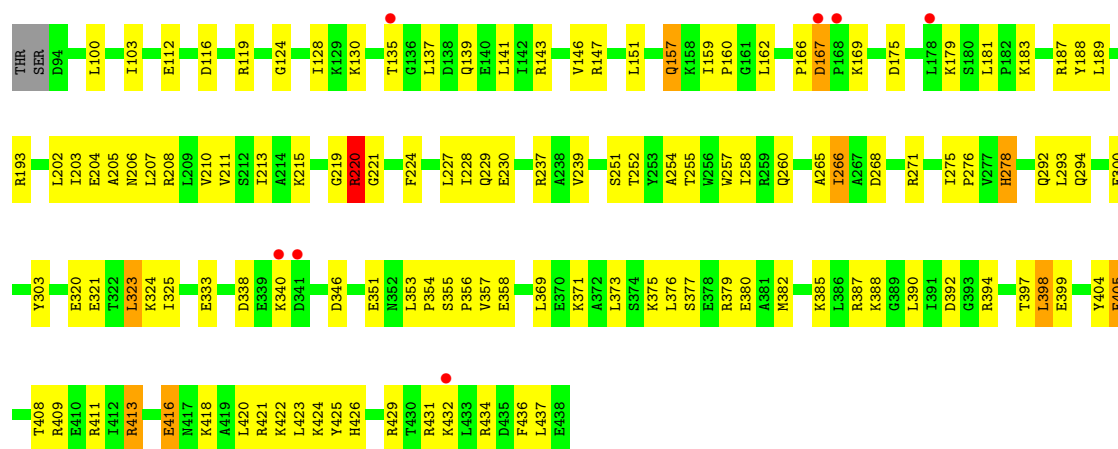
- Molecule 5: RNA polymerase sigma factor SigA

Chain F: 65% 32% 2%

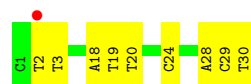


- Molecule 5: RNA polymerase sigma factor SigA

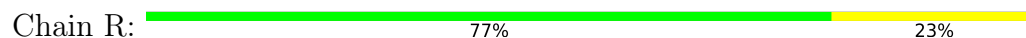
Chain L: 63% 33% 2%



- Molecule 6: DNA (30-MER)



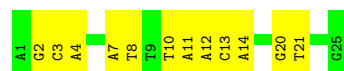
- Molecule 6: DNA (30-MER)



- Molecule 7: DNA (25-MER)



- Molecule 7: DNA (25-MER)



4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α , β , γ	289.87Å 289.87Å 537.36Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.03 – 4.00 50.03 – 3.80	Depositor EDS
% Data completeness (in resolution range)	96.3 (50.03-4.00) 85.0 (50.03-3.80)	Depositor EDS
R_{merge}	0.19	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.92 (at 3.77Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1839)	Depositor
R, R_{free}	0.257 , 0.297 0.257 , 0.297	Depositor DCC
R_{free} test set	9509 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å ²)	135.6	Xtriage
Anisotropy	0.002	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 107.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.36$, $\langle L^2 \rangle = 0.19$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.89	EDS
Total number of atoms	56454	wwPDB-VP
Average B, all atoms (Å ²)	141.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.82% 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

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	0.24	0/1804	0.50	0/2455
1	B	0.24	0/1804	0.50	0/2455
1	G	0.25	0/1804	0.50	0/2455
1	H	0.24	0/1804	0.50	0/2455
2	C	0.24	0/8905	0.49	1/12040 (0.0%)
2	I	0.24	0/8905	0.49	1/12040 (0.0%)
3	D	0.25	0/11963	0.48	1/16165 (0.0%)
3	J	0.25	0/10959	0.48	1/14802 (0.0%)
4	E	0.22	0/783	0.42	0/1054
4	K	0.22	0/783	0.41	0/1054
5	F	0.27	0/2829	0.53	0/3804
5	L	0.28	0/2829	0.54	1/3804 (0.0%)
6	O	0.44	0/687	1.14	0/1059
6	R	0.44	0/687	1.13	0/1059
7	P	0.45	0/568	1.19	0/874
7	S	0.45	0/568	1.19	0/874
All	All	0.26	0/57682	0.54	5/78449 (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	6
2	I	0	5
3	J	0	3
5	F	0	2
5	L	0	3
All	All	0	19

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	J	1286	GLY	N-CA-C	-5.93	98.27	113.10
3	D	1286	GLY	N-CA-C	-5.93	98.28	113.10
5	L	398	LEU	CA-CB-CG	5.44	127.82	115.30
2	C	242	LEU	CA-CB-CG	5.04	126.88	115.30
2	I	242	LEU	CA-CB-CG	5.03	126.87	115.30

There are no chirality outliers.

5 of 19 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	178	ALA	Peptide
2	C	197	LEU	Peptide
2	C	212	SER	Peptide
2	C	230	ARG	Peptide
2	C	737	LEU	Peptide

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	1770	0	1799	36	0
1	B	1770	0	1799	43	0
1	G	1770	0	1799	41	0
1	H	1770	0	1799	47	0
2	C	8739	0	8841	242	0
2	I	8739	0	8841	252	0
3	D	11761	0	11976	310	0
3	J	10779	0	10993	292	0
4	E	768	0	784	18	0
4	K	768	0	784	18	0
5	F	2787	0	2866	75	0
5	L	2787	0	2866	80	0
6	O	613	0	343	10	0
6	R	613	0	343	7	0
7	P	507	0	285	7	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	S	507	0	285	9	0
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	56454	0	56403	1341	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 12.

The worst 5 of 1341 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)
5:L:398:LEU:HG	5:L:409:ARG:HB2	1.45	0.99
5:F:398:LEU:HG	5:F:409:ARG:HB2	1.48	0.91
1:H:53:VAL:HG23	1:H:144:VAL:HG22	1.55	0.88
1:B:53:VAL:HG23	1:B:144:VAL:HG22	1.55	0.86
1:A:42:ARG:HH12	2:C:857:ASP:HB3	1.41	0.86

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%)	209 (93%)	15 (7%)	1 (0%)	34	71
1	B	225/314 (72%)	208 (92%)	16 (7%)	1 (0%)	34	71
1	G	225/314 (72%)	209 (93%)	15 (7%)	1 (0%)	34	71
1	H	225/314 (72%)	208 (92%)	16 (7%)	1 (0%)	34	71
2	C	1108/1119 (99%)	1045 (94%)	58 (5%)	5 (0%)	29	67

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	I	1108/1119 (99%)	1040 (94%)	62 (6%)	6 (0%)	29	67
3	D	1486/1524 (98%)	1395 (94%)	85 (6%)	6 (0%)	34	71
3	J	1361/1524 (89%)	1281 (94%)	74 (5%)	6 (0%)	34	71
4	E	91/99 (92%)	87 (96%)	4 (4%)	0	100	100
4	K	91/99 (92%)	88 (97%)	3 (3%)	0	100	100
5	F	343/347 (99%)	324 (94%)	17 (5%)	2 (1%)	25	63
5	L	343/347 (99%)	326 (95%)	15 (4%)	2 (1%)	25	63
All	All	6831/7434 (92%)	6420 (94%)	380 (6%)	31 (0%)	29	67

5 of 31 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	53	VAL
3	D	1128	VAL
1	G	53	VAL
3	J	1128	VAL
2	C	607	ASP

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%)	178 (92%)	16 (8%)	11	38
1	B	194/270 (72%)	177 (91%)	17 (9%)	10	35
1	G	194/270 (72%)	178 (92%)	16 (8%)	11	38
1	H	194/270 (72%)	177 (91%)	17 (9%)	10	35
2	C	931/936 (100%)	847 (91%)	84 (9%)	9	34
2	I	931/936 (100%)	849 (91%)	82 (9%)	10	35
3	D	1252/1281 (98%)	1151 (92%)	101 (8%)	11	39
3	J	1150/1281 (90%)	1057 (92%)	93 (8%)	11	39
4	E	83/88 (94%)	79 (95%)	4 (5%)	25	53

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	K	83/88 (94%)	79 (95%)	4 (5%)	25	53
5	F	296/299 (99%)	266 (90%)	30 (10%)	7	29
5	L	296/299 (99%)	265 (90%)	31 (10%)	7	27
All	All	5798/6288 (92%)	5303 (92%)	495 (8%)	10	37

5 of 495 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	F	189	LEU
3	J	1093	TYR
2	I	20	GLU
3	J	1053	PHE
5	L	189	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 45 such sidechains are listed below:

Mol	Chain	Res	Type
2	I	204	GLN
3	J	483	HIS
2	I	538	GLN
2	I	829	GLN
3	J	549	ASN

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

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	227/314 (72%)	-0.17	1 (0%) 92 87	87, 140, 176, 201	0
1	B	227/314 (72%)	-0.29	0 100 100	62, 117, 166, 206	0
1	G	227/314 (72%)	0.16	6 (2%) 56 46	124, 168, 194, 220	0
1	H	227/314 (72%)	-0.11	0 100 100	90, 147, 178, 196	0
2	C	1112/1119 (99%)	-0.14	13 (1%) 79 70	59, 139, 200, 249	0
2	I	1112/1119 (99%)	-0.04	20 (1%) 68 59	80, 159, 207, 252	0
3	D	1490/1524 (97%)	-0.25	1 (0%) 95 94	52, 113, 165, 216	0
3	J	1367/1524 (89%)	-0.18	5 (0%) 92 87	66, 134, 184, 230	0
4	E	93/99 (93%)	-0.21	0 100 100	65, 117, 169, 188	0
4	K	93/99 (93%)	-0.15	0 100 100	92, 145, 192, 221	0
5	F	345/347 (99%)	-0.15	5 (1%) 75 65	76, 146, 205, 235	0
5	L	345/347 (99%)	-0.04	7 (2%) 65 56	100, 161, 214, 248	0
6	O	30/30 (100%)	-0.09	1 (3%) 46 37	107, 178, 224, 227	0
6	R	30/30 (100%)	-0.38	0 100 100	134, 169, 201, 211	0
7	P	25/25 (100%)	-0.07	0 100 100	138, 174, 220, 243	0
7	S	25/25 (100%)	-0.48	0 100 100	141, 177, 196, 203	0
All	All	6975/7544 (92%)	-0.15	59 (0%) 86 79	52, 139, 196, 252	0

The worst 5 of 59 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	I	207	LEU	4.6
2	I	221	LEU	4.2
2	C	221	LEU	3.7
2	C	222	LEU	3.6
2	C	176	VAL	3.6

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 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, 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	B-factors(\AA^2)	Q<0.9
9	MG	D	2003	1/1	0.86	0.15	205,205,205,205	0
8	ZN	J	2002	1/1	0.91	0.09	154,154,154,154	0
9	MG	J	2003	1/1	0.93	0.15	255,255,255,255	0
8	ZN	D	2002	1/1	0.97	0.18	118,118,118,118	0
8	ZN	J	2001	1/1	0.98	0.20	151,151,151,151	0
8	ZN	D	2001	1/1	0.99	0.20	107,107,107,107	0

6.5 Other polymers [i](#)

There are no such residues in this entry.