wwPDB X-ray Structure Validation Summary Report (i)

Jun 23, 2024 – 02:14 AM EDT

PDB ID	:	5IPM
Title	:	SigmaS-transcription initiation complex with 4-nt nascent RNA
Authors	:	Liu, B.; Zuo, Y.; Steitz, T.A.
Deposited on	:	2016-03-09
Resolution	:	4.20 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:

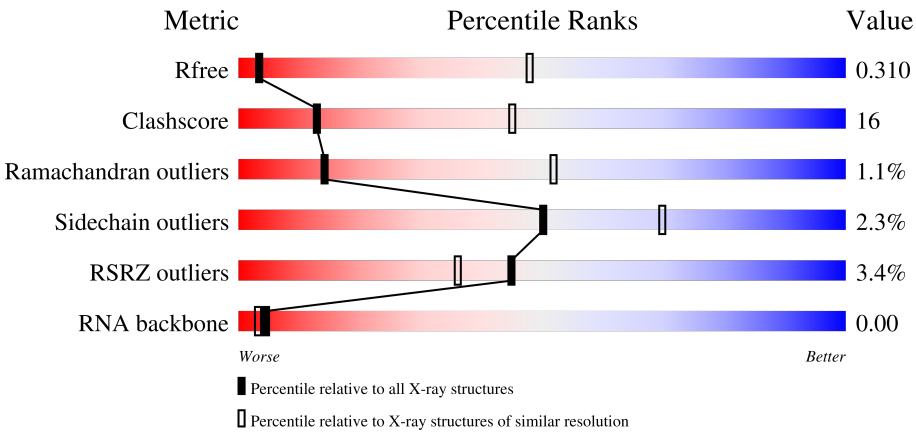
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (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 4.20 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	1005 (4.62-3.78)
Clashscore	141614	1044 (4.60-3.80)
Ramachandran outliers	138981	1000 (4.60-3.80)
Sidechain outliers	138945	1007 (4.62-3.78)
RSRZ outliers	127900	1063 (4.70-3.70)
RNA backbone	3102	1049 (5.04-3.00)

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	242	5% 77% 17% 5%
1	B	242	3% 79% 15% 6%
2	C	1342	2% 75% 24% •
3	D	1407	4% 66% 29% • •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	90	% 76% 12% 12%
5	F	336	5% 58% 23% • 18%
6	1	50	8% 8% 58% 34%
7	2	50	2% 24% 40% • 34%
8	3	4	25% 25% 50%

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
9	ZN	D	1502	-	-	X	-

2 Entry composition (i)

There are 10 unique types of molecules in this entry. The entry contains 29027 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
			Total	C	N	O	S			
1	A	230	1787	1112	317	352	6	0	0	0
1	B	228	1767	1100	312	349	6	0	0	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-6	ALA	-	expression tag	UNP P0A7Z4
A	-5	HIS	-	expression tag	UNP P0A7Z4
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
B	-6	ALA	-	expression tag	UNP P0A7Z4
B	-5	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

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	1340	10570	6631	1841	2055	43	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	D	1362	10568	6633	1887	1998	50	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	E	79	627	382	118	126	1	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	F	277	2253	1411	415	423	4	0	0	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	2	GLY	SER	conflict	UNP P13445
F	33	GLU	GLN	conflict	UNP P13445
F	329	LEU	ARG	conflict	UNP P13445
F	331	HIS	-	expression tag	UNP P13445
F	332	HIS	-	expression tag	UNP P13445
F	333	HIS	-	expression tag	UNP P13445
F	334	HIS	-	expression tag	UNP P13445
F	335	HIS	-	expression tag	UNP P13445
F	336	HIS	-	expression tag	UNP P13445

- Molecule 6 is a DNA chain called synthetic non-template strand DNA (50-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
6	1	33	680	323	124	200	33	0	0	0

- Molecule 7 is a DNA chain called synthetic template strand DNA (50-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
7	2	33	675	322	125	196	32	0	0	0

- Molecule 8 is a RNA chain called nascent RNA 4-mer.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
8	3	4	97	39	17	35	6	0	0	0

- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

- Molecule 10 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

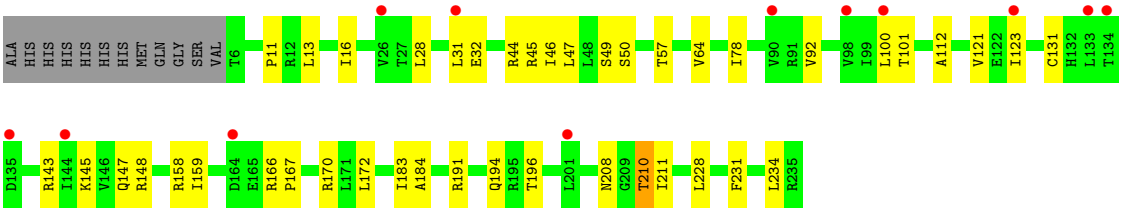
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
10	D	1	1	1	0	0

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.

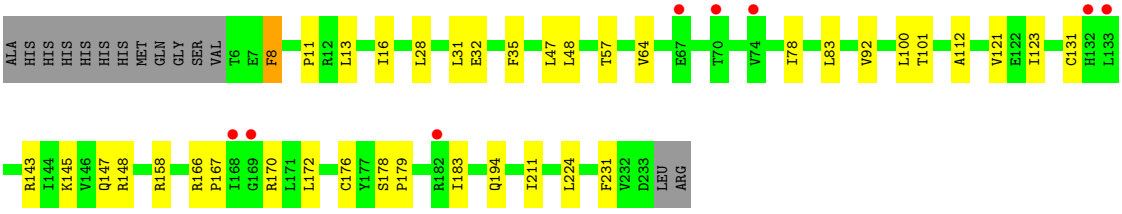
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain A:



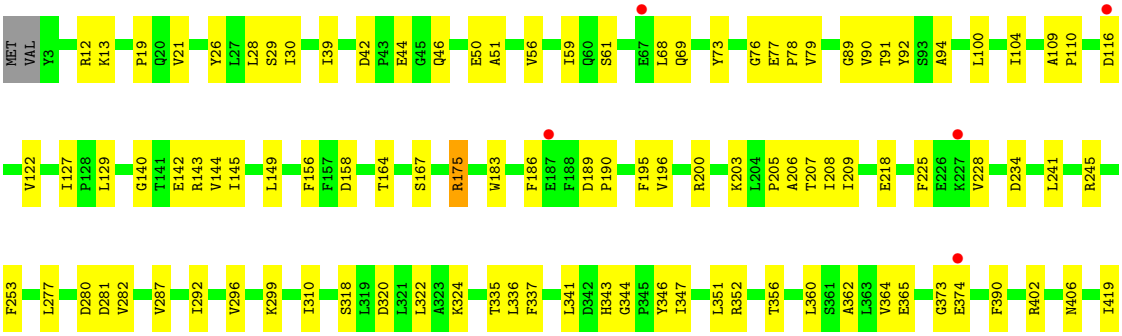
• Molecule 1: DNA-directed RNA polymerase subunit alpha

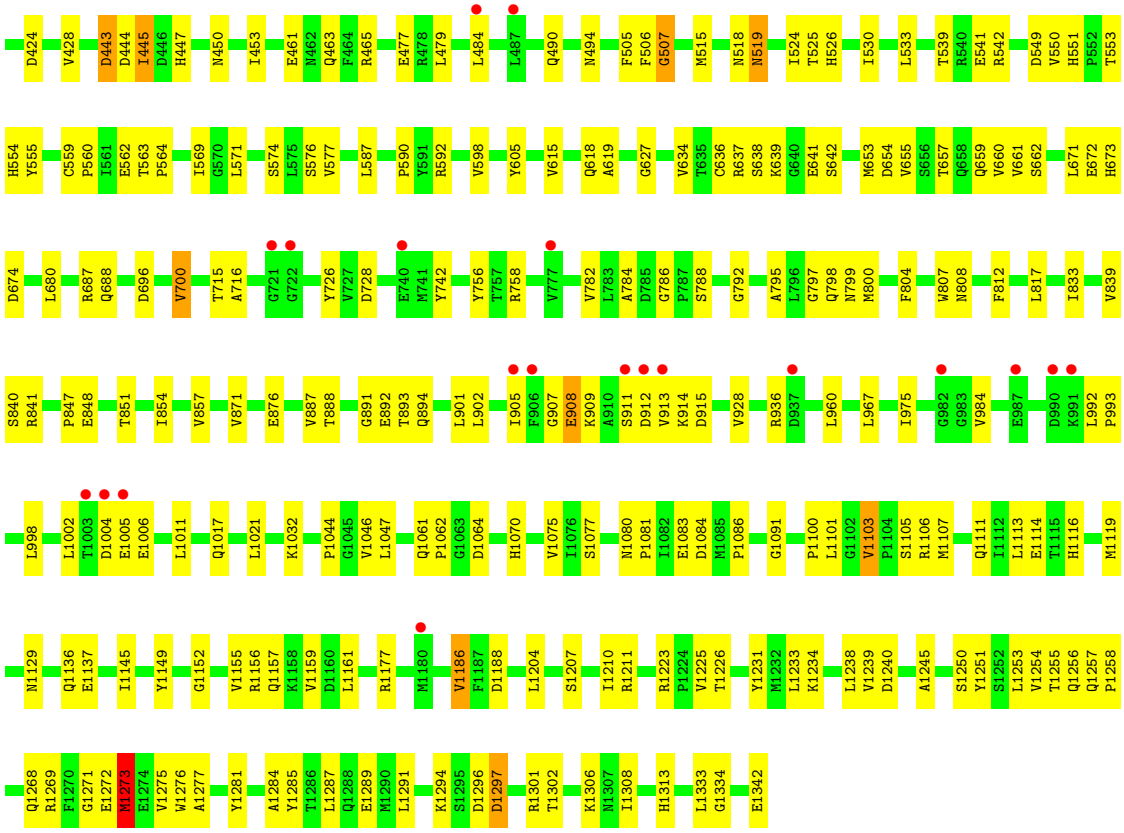
Chain B:



• Molecule 2: DNA-directed RNA polymerase subunit beta

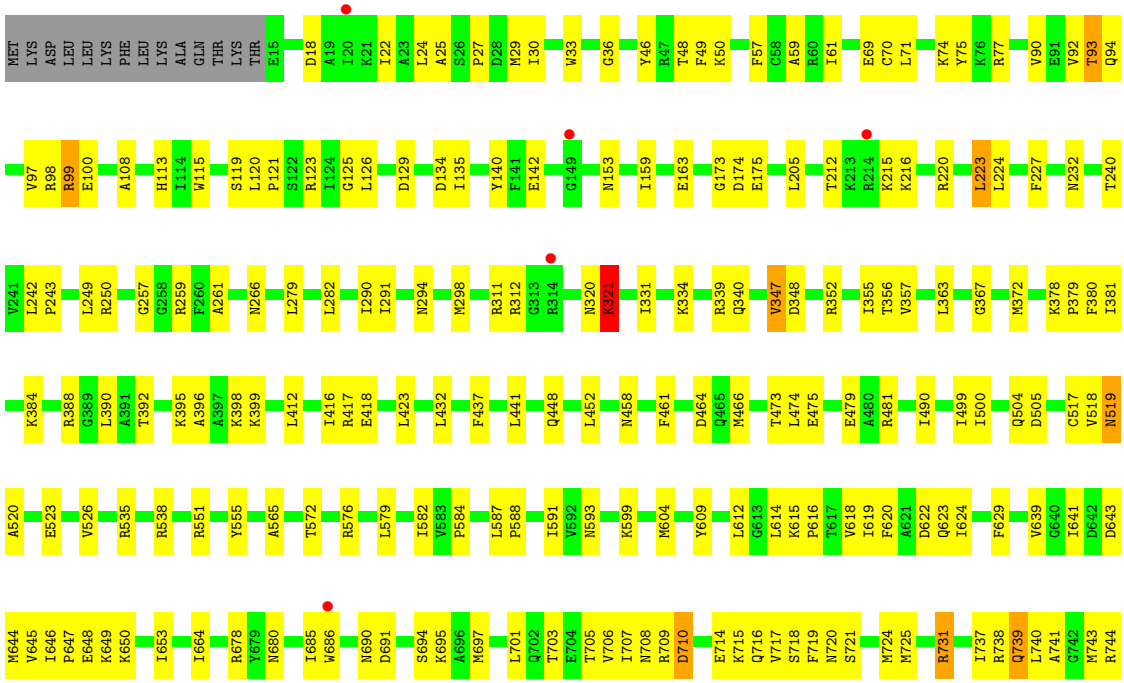
Chain C:

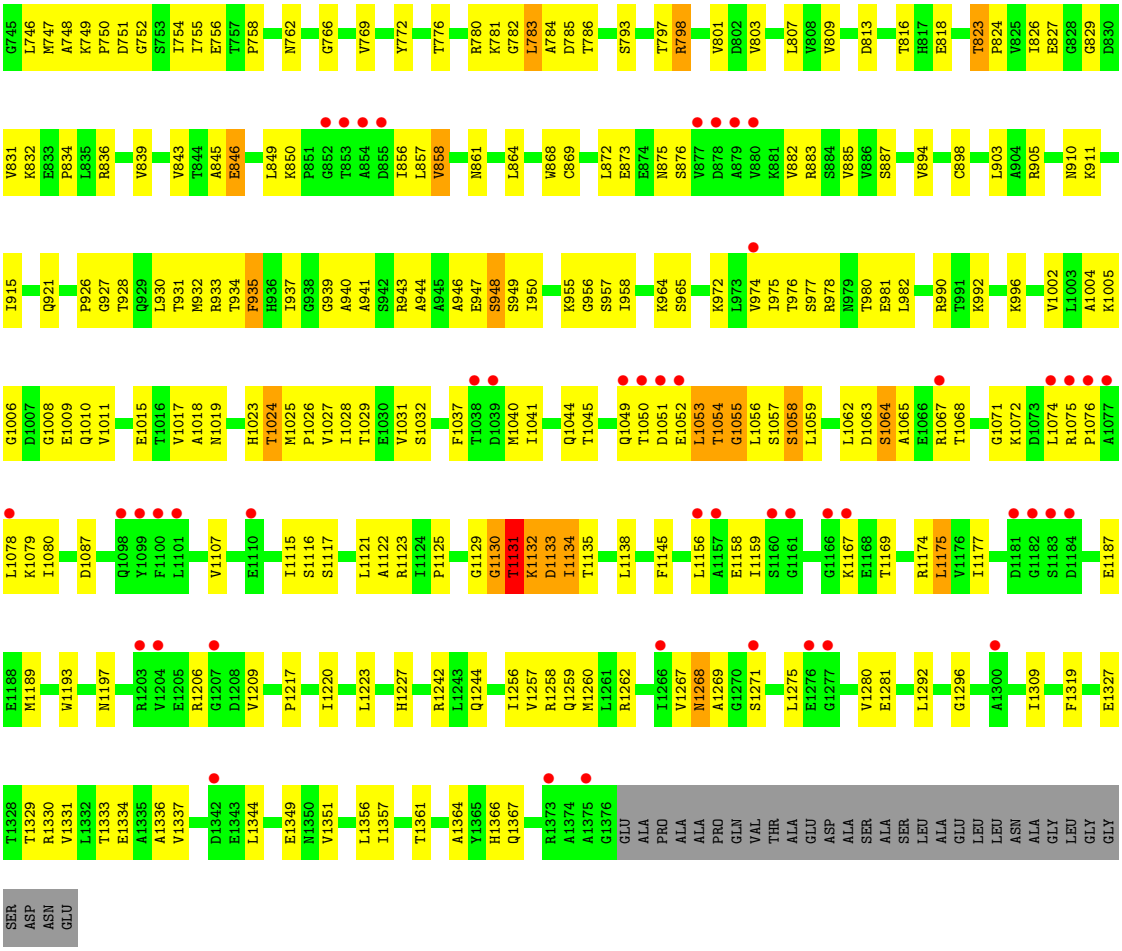




- Molecule 3: DNA-directed RNA polymerase subunit beta'

Chain D:



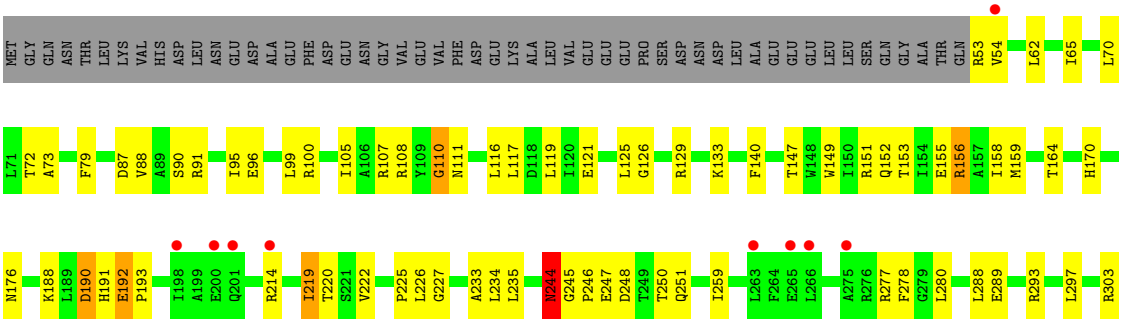


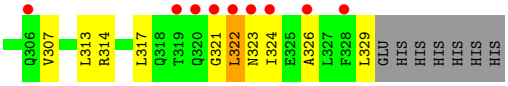
• Molecule 4: DNA-directed RNA polymerase subunit omega

Chain E:

• Molecule 5: RNA polymerase sigma factor RpoS

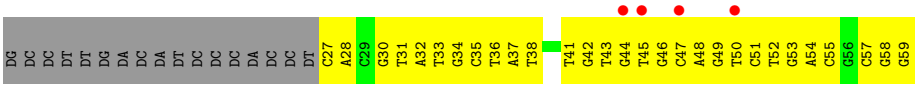
Chain F:





- Molecule 6: synthetic non-template strand DNA (50-MER)

Chain 1:



- Molecule 7: synthetic template strand DNA (50-MER)

Chain 2:

- Molecule 8: nascent RNA 4-mer

Chain 3:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	132.75Å 151.97Å 228.06Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	126.46 – 4.20 49.99 – 4.20	Depositor EDS
% Data completeness (in resolution range)	99.9 (126.46-4.20) 100.0 (49.99-4.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.22 (at 4.14Å)	Xtriage
Refinement program	REFMAC 5.8.0123	Depositor
R, R_{free}	0.270 , 0.330 0.258 , 0.310	Depositor DCC
R_{free} test set	1651 reflections (4.81%)	wwPDB-VP
Wilson B-factor (Å^2)	174.0	Xtriage
Anisotropy	0.560	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 224.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.38$, $< L^2 > = 0.21$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	29027	wwPDB-VP
Average B, all atoms (Å^2)	287.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.64% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, GTP, 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	0.55	0/1809	0.73	1/2450 (0.0%)
1	B	0.53	0/1789	0.71	0/2425
2	C	0.61	1/10736 (0.0%)	0.79	8/14480 (0.1%)
3	D	0.62	2/10729 (0.0%)	0.86	8/14487 (0.1%)
4	E	0.53	0/629	0.73	0/847
5	F	0.64	2/2282 (0.1%)	0.87	1/3076 (0.0%)
6	1	0.39	0/762	0.66	0/1175
7	2	0.52	1/757 (0.1%)	0.68	0/1167
8	3	0.34	0/72	0.62	0/110
All	All	0.60	6/29565 (0.0%)	0.80	18/40217 (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
5	F	0	1

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	1131	THR	C-N	23.67	1.88	1.34
3	D	739	GLN	C-N	14.94	1.68	1.34
5	F	247	GLU	CD-OE1	8.12	1.34	1.25
5	F	247	GLU	CG-CD	7.43	1.63	1.51
2	C	445	ILE	N-CA	-5.76	1.34	1.46

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	D	1131	THR	C-N-CA	-31.47	43.03	121.70
3	D	1131	THR	CA-C-N	-24.32	63.71	117.20
3	D	1131	THR	O-C-N	-20.56	89.81	122.70
3	D	780	ARG	NE-CZ-NH1	7.07	123.83	120.30
2	C	742	TYR	CB-CG-CD1	-6.56	117.06	121.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
5	F	244	ASN	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	1787	0	1813	34	0
1	B	1767	0	1789	30	0
2	C	10570	0	10579	275	0
3	D	10568	0	10778	547	0
4	E	627	0	634	7	0
5	F	2253	0	2298	92	0
6	1	680	0	373	77	0
7	2	675	0	373	37	0
8	3	97	0	44	6	0
9	D	2	0	0	2	0
10	D	1	0	0	0	0
All	All	29027	0	28681	949	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 16.

The worst 5 of 949 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:785:ASP:HB3	3:D:935:PHE:CZ	1.23	1.62

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:D:750:PRO:HA	3:D:781:LYS:CD	1.22	1.60
3:D:739:GLN:C	3:D:740:LEU:N	1.68	1.43
3:D:750:PRO:CA	3:D:781:LYS:HD2	1.51	1.38
3:D:750:PRO:C	3:D:781:LYS:HE3	1.40	1.38

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	228/242 (94%)	216 (95%)	10 (4%)	2 (1%)	17 56
1	B	226/242 (93%)	214 (95%)	11 (5%)	1 (0%)	34 72
2	C	1332/1342 (99%)	1242 (93%)	82 (6%)	8 (1%)	25 64
3	D	1360/1407 (97%)	1258 (92%)	81 (6%)	21 (2%)	10 46
4	E	77/90 (86%)	74 (96%)	3 (4%)	0	100 100
5	F	275/336 (82%)	258 (94%)	12 (4%)	5 (2%)	8 42
All	All	3498/3659 (96%)	3262 (93%)	199 (6%)	37 (1%)	14 52

5 of 37 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	908	GLU
3	D	174	ASP
3	D	519	ASN
3	D	1024	THR
3	D	1053	LEU

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	198/208 (95%)	194 (98%)	4 (2%)	55	73
1	B	196/208 (94%)	191 (97%)	5 (3%)	46	67
2	C	1155/1157 (100%)	1139 (99%)	16 (1%)	67	80
3	D	1135/1168 (97%)	1099 (97%)	36 (3%)	39	62
4	E	67/74 (90%)	65 (97%)	2 (3%)	41	63
5	F	240/292 (82%)	235 (98%)	5 (2%)	53	71
All	All	2991/3107 (96%)	2923 (98%)	68 (2%)	50	70

5 of 68 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	1189	MET
3	D	1356	LEU
5	F	192	GLU
2	C	1273	MET
2	C	1272	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 38 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	1019	ASN
4	E	31	GLN
3	D	1098	GLN
3	D	1218	HIS
5	F	152	GLN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	3	2/4 (50%)	2 (100%)	0

All (2) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
8	3	16	G
8	3	17	U

There are no RNA pucker outliers to report.

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, 3 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)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	C	3
3	D	2

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	C	336:LEU	C	337:PHE	N	2.22
1	C	911:SER	C	912:ASP	N	2.12
1	C	891:GLY	C	892:GLU	N	2.08
1	D	1131:THR	C	1132:LYS	N	1.88
1	D	739:GLN	C	740:LEU	N	1.68

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

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	230/242 (95%)	0.13	12 (5%)	27	23	183, 279, 356, 428	0
1	B	228/242 (94%)	-0.02	8 (3%)	44	35	193, 298, 382, 466	0
2	C	1340/1342 (99%)	-0.16	25 (1%)	66	58	140, 249, 396, 562	0
3	D	1362/1407 (96%)	0.04	52 (3%)	40	32	138, 279, 418, 529	0
4	E	79/90 (87%)	-0.03	1 (1%)	77	68	245, 321, 482, 519	0
5	F	277/336 (82%)	0.25	18 (6%)	18	15	227, 330, 432, 542	0
6	1	33/50 (66%)	0.21	4 (12%)	4	5	272, 339, 468, 483	0
7	2	33/50 (66%)	0.16	1 (3%)	50	39	232, 318, 425, 477	0
8	3	3/4 (75%)	0.40	0	100	100	298, 298, 327, 338	0
All	All	3585/3763 (95%)	-0.02	121 (3%)	45	36	138, 277, 415, 562	0

The worst 5 of 121 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	F	323	ASN	11.1
3	D	1076	PRO	8.9
2	C	1004	ASP	7.1
3	D	1075	ARG	6.2
2	C	982	GLY	5.9

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(Å^2)	Q<0.9
10	MG	D	1503	1/1	0.92	0.07	343,343,343,343	0
9	ZN	D	1502	1/1	0.93	0.13	298,298,298,298	0
9	ZN	D	1501	1/1	0.99	0.08	316,316,316,316	0

6.5 Other polymers (i)

There are no such residues in this entry.