



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 27, 2020 – 02:08 PM BST

PDB ID : 6UTW
Title : E. coli sigma-S transcription initiation complex with a 4-nt RNA ("Fresh" crystal)
Authors : Zuo, Y.; De, S.; Steitz, T.A.
Deposited on : 2019-10-30
Resolution : 3.85 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 following versions of software and data (see [references ⓘ](#)) 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 : 2.13
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) : 2.13

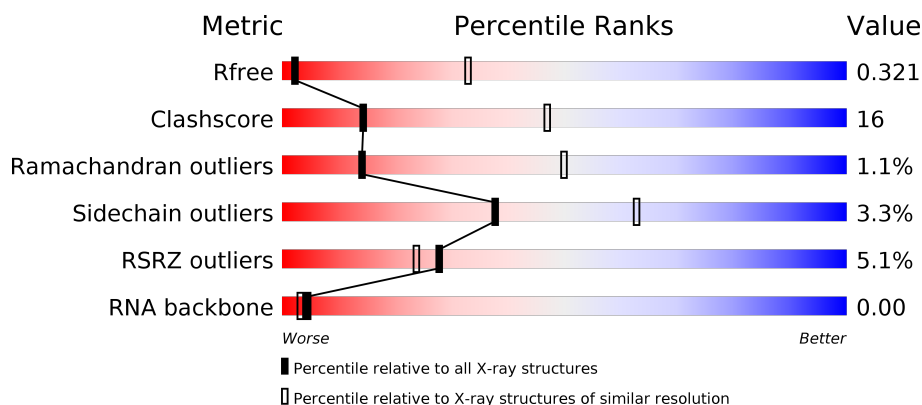
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.85 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	1048 (4.10-3.62)
Clashscore	141614	1015 (4.08-3.64)
Ramachandran outliers	138981	1069 (4.10-3.62)
Sidechain outliers	138945	1062 (4.10-3.62)
RSRZ outliers	127900	1206 (4.12-3.60)
RNA backbone	3102	1039 (4.70-3.00)

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 on 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	AAA	242	<div> <div>3%</div> <div>65% 27% 5%</div> </div>
1	BBB	242	<div> <div>7%</div> <div>69% 25% 6%</div> </div>
2	CCC	1342	<div> <div>2%</div> <div>72% 26%</div> </div>
3	DDD	1407	<div> <div>7%</div> <div>68% 27%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	EEE	90	
5	FFF	336	
6	111	50	
7	222	50	
8	333	4	

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
10	ZN	DDD	1501	-	-	X	-
10	ZN	DDD	1502	-	-	X	-
9	MG	CCC	1401	-	-	-	X

2 Entry composition

There are 11 unique types of molecules in this entry. The entry contains 29078 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	AAA	230	Total	C	N	O	S	0	0	0
			1787	1112	317	352	6			
1	BBB	228	Total	C	N	O	S	0	0	0
			1767	1100	312	349	6			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	-6	ALA	-	expression tag	UNP P0A7Z4
AAA	-5	HIS	-	expression tag	UNP P0A7Z4
AAA	-4	HIS	-	expression tag	UNP P0A7Z4
AAA	-3	HIS	-	expression tag	UNP P0A7Z4
AAA	-2	HIS	-	expression tag	UNP P0A7Z4
AAA	-1	HIS	-	expression tag	UNP P0A7Z4
AAA	0	HIS	-	expression tag	UNP P0A7Z4
BBB	-6	ALA	-	expression tag	UNP P0A7Z4
BBB	-5	HIS	-	expression tag	UNP P0A7Z4
BBB	-4	HIS	-	expression tag	UNP P0A7Z4
BBB	-3	HIS	-	expression tag	UNP P0A7Z4
BBB	-2	HIS	-	expression tag	UNP P0A7Z4
BBB	-1	HIS	-	expression tag	UNP P0A7Z4
BBB	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
2	CCC	1340	Total	C	N	O	S	0	0	0
			10570	6631	1841	2055	43			

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

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

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

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

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

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

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
FFF	2	GLY	SER	conflict	UNP P13445
FFF	33	GLU	GLN	conflict	UNP P13445
FFF	329	LEU	-	expression tag	UNP P13445
FFF	330	GLU	-	expression tag	UNP P13445
FFF	331	HIS	-	expression tag	UNP P13445
FFF	332	HIS	-	expression tag	UNP P13445
FFF	333	HIS	-	expression tag	UNP P13445
FFF	334	HIS	-	expression tag	UNP P13445
FFF	335	HIS	-	expression tag	UNP P13445
FFF	336	HIS	-	expression tag	UNP P13445

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	111	33	Total	C	N	O	P	0	0	0
			680	323	124	200	33			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	222	35	Total	C	N	O	P	0	0	0
			716	342	132	208	34			

- Molecule 8 is a RNA chain called RNA 4-mer (de novo synthesized).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	333	4	Total	C	N	O	P	0	0	0
			97	39	17	35	6			

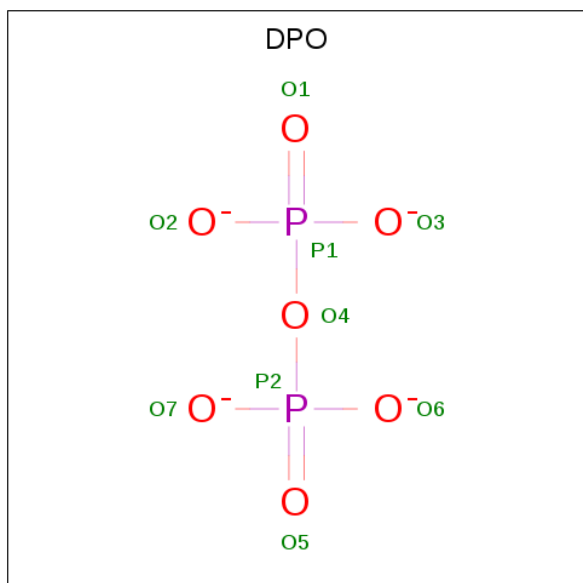
- Molecule 9 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	CCC	1	Total	Mg	0	0
			1	1		
9	DDD	1	Total	Mg	0	0
			1	1		

- Molecule 10 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	DDD	2	Total	Zn	0	0
			2	2		

- Molecule 11 is DIPHOSPHATE (three-letter code: DPO) (formula: O₇P₂).

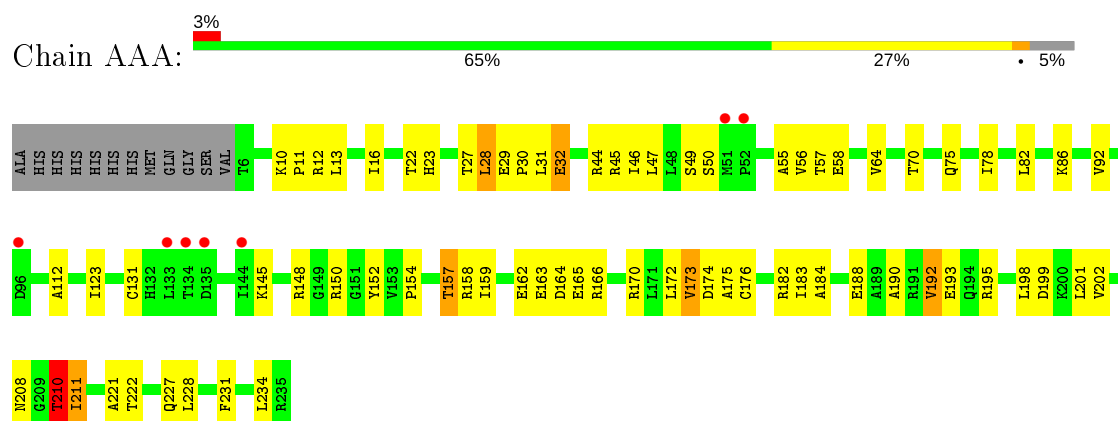


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
11	DDD	1	Total	O	P	0	0
			9	7	2		

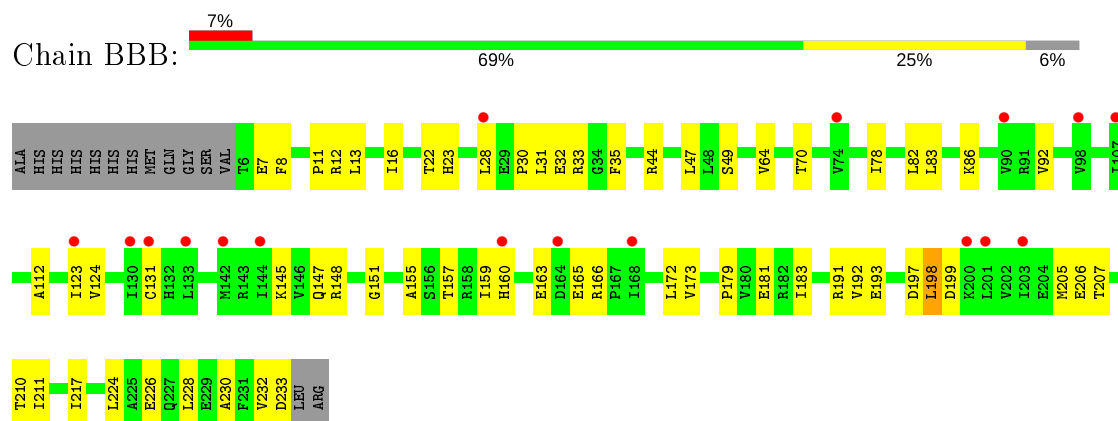
3 Residue-property plots [i](#)

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.

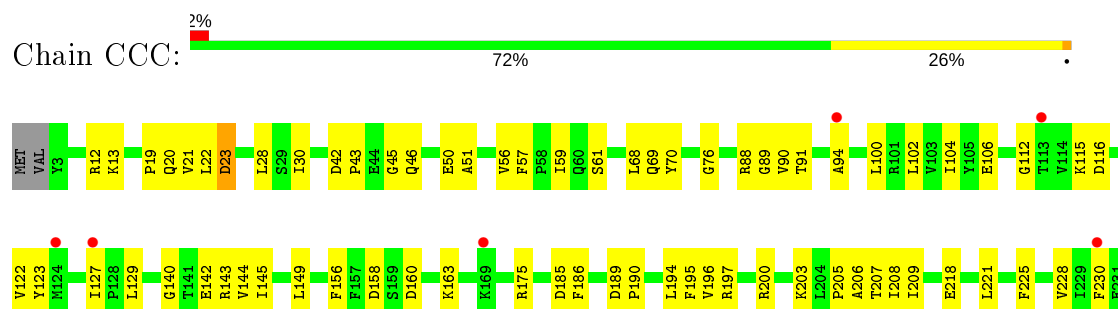
- Molecule 1: DNA-directed RNA polymerase subunit alpha

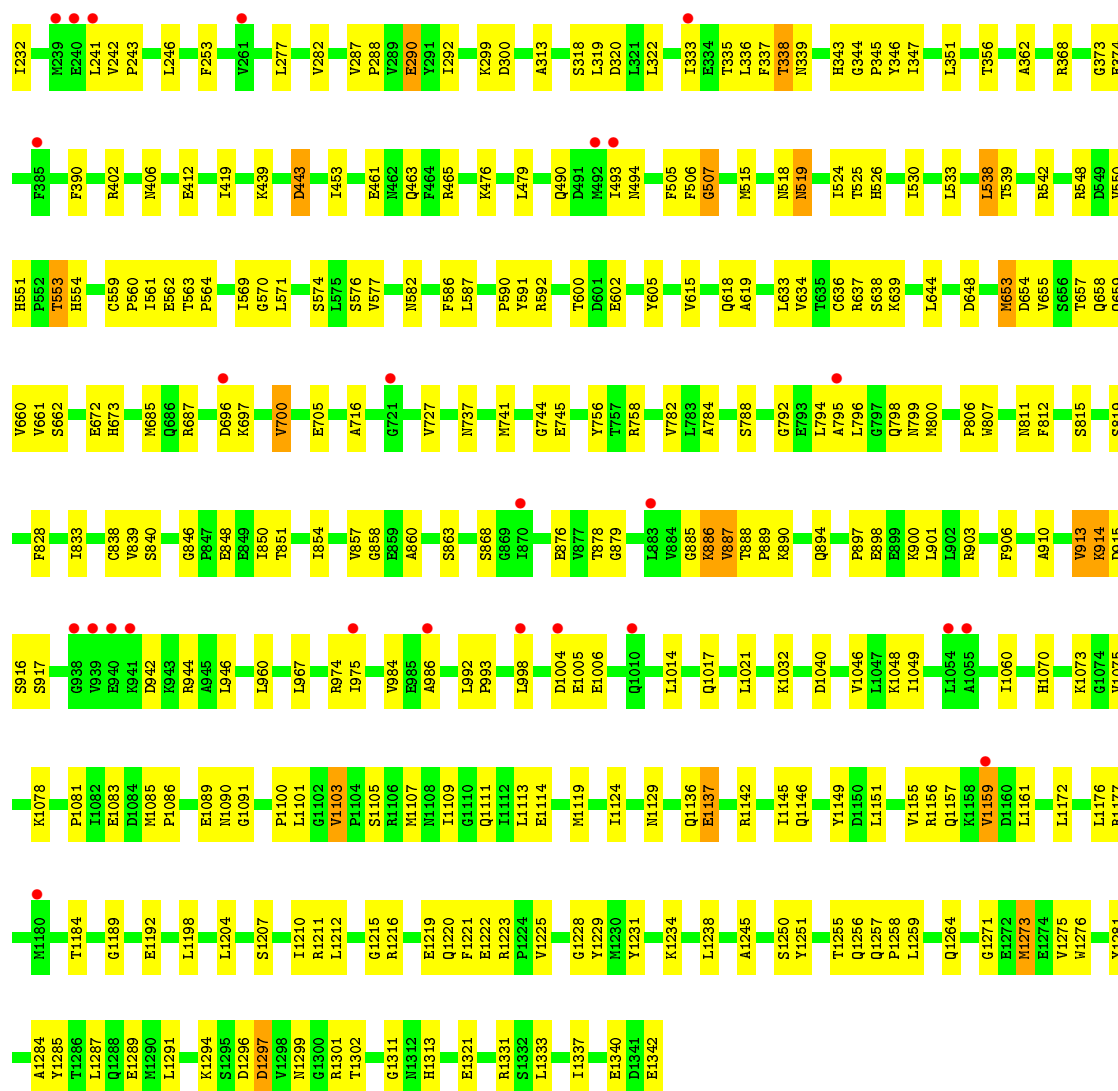


- Molecule 1: DNA-directed RNA polymerase subunit alpha

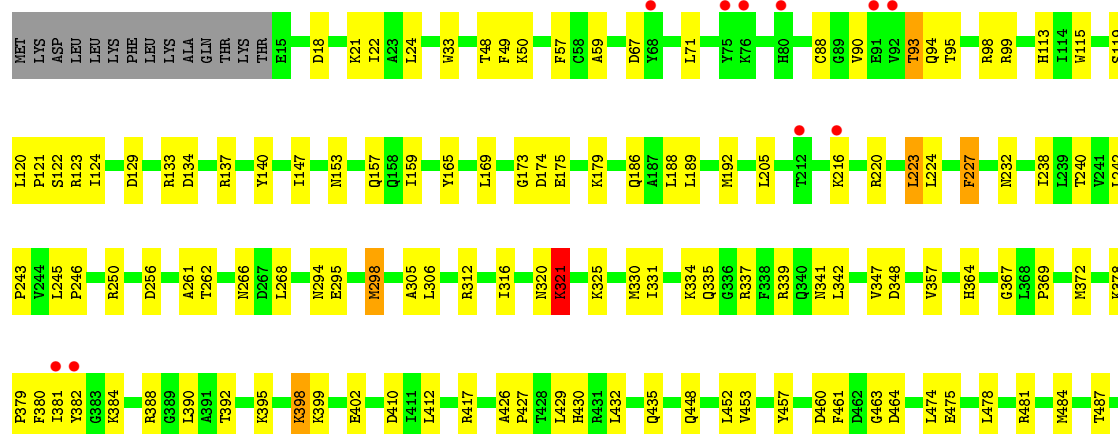
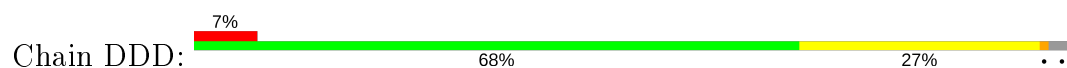


- Molecule 2: DNA-directed RNA polymerase subunit beta

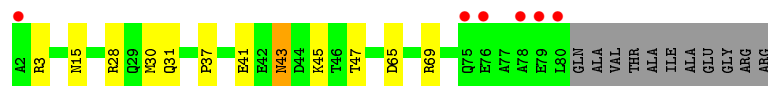
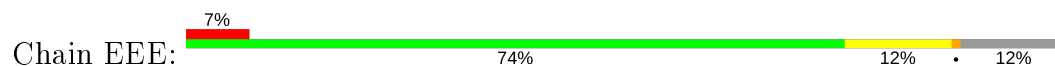




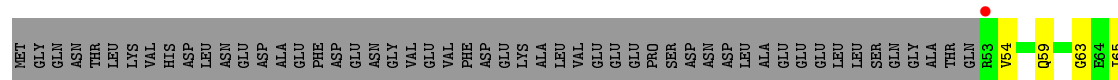
• Molecule 3: DNA-directed RNA polymerase subunit beta'

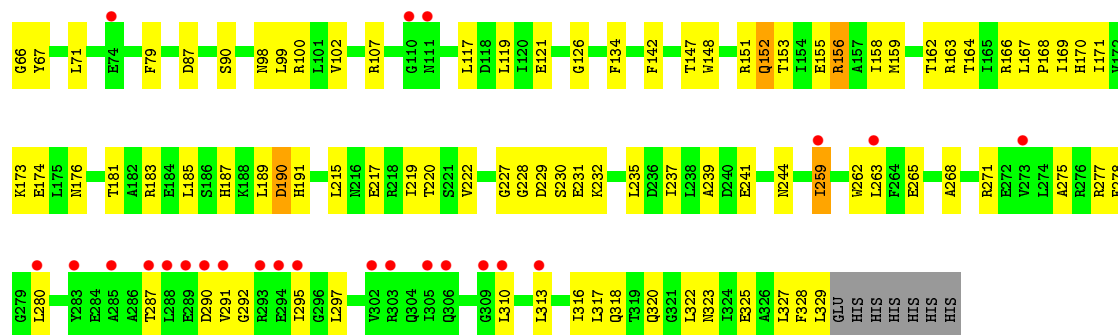


- Molecule 4: DNA-directed RNA polymerase subunit omega

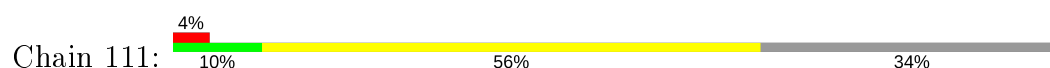


- Molecule 5: RNA polymerase sigma factor RpoS

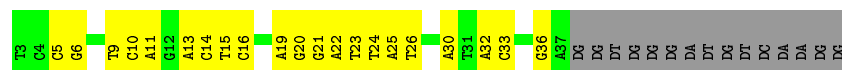




- Molecule 6: Synthetic DNA 50-MER (promoter non-template strand)



- Molecule 7: Synthetic DNA 50-MER (promoter template strand)



- Molecule 8: RNA 4-mer (de novo synthesized)



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	132.99Å 155.45Å 234.74Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.40 – 3.85 49.40 – 3.85	Depositor EDS
% Data completeness (in resolution range)	92.6 (49.40-3.85) 92.7 (49.40-3.85)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.10 (at 3.88Å)	Xtriage
Refinement program	REFMAC 5.8.0257	Depositor
R, R_{free}	0.268 , 0.329 0.263 , 0.321	Depositor DCC
R_{free} test set	2059 reflections (4.76%)	wwPDB-VP
Wilson B-factor (Å ²)	168.6	Xtriage
Anisotropy	0.143	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 176.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.36$, $\langle L^2 \rangle = 0.20$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	29078	wwPDB-VP
Average B, all atoms (Å ²)	257.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.60% 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: GTP, DPO, 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	AAA	0.65	0/1809	0.73	0/2450
1	BBB	0.66	0/1789	0.73	0/2425
2	CCC	0.64	0/10739	0.79	0/14489
3	DDD	0.66	0/10729	0.77	0/14487
4	EEE	0.61	0/629	0.76	0/847
5	FFF	0.67	0/2282	0.71	0/3076
6	111	0.43	0/762	0.66	0/1175
7	222	0.42	0/803	0.69	0/1238
8	333	0.40	0/71	0.73	0/106
All	All	0.64	0/29613	0.76	0/40293

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	AAA	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	AAA	210	THR	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	AAA	1787	0	1813	76	0
1	BBB	1767	0	1789	59	1
2	CCC	10570	0	10582	349	0
3	DDD	10568	0	10783	339	1
4	EEE	627	0	634	16	0
5	FFF	2253	0	2298	110	0
6	111	680	0	373	61	0
7	222	716	0	396	37	0
8	333	97	0	45	4	0
9	CCC	1	0	0	0	0
9	DDD	1	0	0	0	0
10	DDD	2	0	0	4	0
11	DDD	9	0	0	1	0
All	All	29078	0	28713	895	1

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 895 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:CCC:848:GLU:OE2	2:CCC:888:THR:HG22	1.32	1.23
6:111:50:DT:H3'	6:111:51:DC:C5'	1.75	1.16
3:DDD:936:HIS:ND1	8:333:17:U:OP1	1.79	1.14
3:DDD:320:ASN:OD1	7:222:22:DA:N6	1.80	1.14
2:CCC:373:GLY:HA3	5:FFF:54:VAL:HG22	1.18	1.14

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:BBB:166:ARG:NH2	3:DDD:1055:GLY:O[4_445]	1.99	0.21

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	AAA	228/242 (94%)	204 (90%)	20 (9%)	4 (2%)	8	41
1	BBB	226/242 (93%)	211 (93%)	13 (6%)	2 (1%)	17	53
2	CCC	1338/1342 (100%)	1243 (93%)	84 (6%)	11 (1%)	19	56
3	DDD	1360/1407 (97%)	1261 (93%)	81 (6%)	18 (1%)	12	46
4	EEE	77/90 (86%)	74 (96%)	3 (4%)	0	100	100
5	FFF	275/336 (82%)	256 (93%)	16 (6%)	3 (1%)	14	50
All	All	3504/3659 (96%)	3249 (93%)	217 (6%)	38 (1%)	14	50

5 of 38 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	AAA	192	VAL
3	DDD	174	ASP
3	DDD	519	ASN
3	DDD	710	ASP
3	DDD	861	ASN

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	AAA	198/208 (95%)	189 (96%)	9 (4%)	27	55
1	BBB	196/208 (94%)	188 (96%)	8 (4%)	30	57
2	CCC	1155/1157 (100%)	1124 (97%)	31 (3%)	44	67

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	DDD	1135/1168 (97%)	1094 (96%)	41 (4%)	35	61
4	EEE	67/74 (90%)	65 (97%)	2 (3%)	41	64
5	FFF	240/292 (82%)	232 (97%)	8 (3%)	38	63
All	All	2991/3107 (96%)	2892 (97%)	99 (3%)	38	63

5 of 99 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	CCC	1223	ARG
3	DDD	256	ASP
5	FFF	107	ARG
2	CCC	1273	MET
3	DDD	153	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	333	1/4 (25%)	1 (100%)	0

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
8	333	16	G

There are no RNA pucker outliers to report.

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 5 ligands modelled in this entry, 4 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	$\# Z > 2$
11	DPO	DDD	1504	-	6,8,8	0.75	0	13,13,13	1.53	1 (7%)

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
11	DPO	DDD	1504	-	-	0/6/6/6	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
11	DDD	1504	DPO	P2-O4-P1	-4.60	117.04	132.83

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
11	DDD	1504	DPO	1	0

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

The following chains have linkage breaks:

Mol	Chain	Number of breaks
8	333	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	333	16:G	O3'	17:U	P	3.64

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	AAA	230/242 (95%)	0.08	7 (3%) 50 39	179, 261, 332, 374	0
1	BBB	228/242 (94%)	0.09	17 (7%) 14 11	181, 267, 336, 412	0
2	CCC	1340/1342 (99%)	-0.01	32 (2%) 59 49	135, 229, 361, 488	0
3	DDD	1362/1407 (96%)	0.18	94 (6%) 16 12	124, 243, 361, 437	0
4	EEE	79/90 (87%)	-0.07	6 (7%) 13 10	196, 282, 469, 527	0
5	FFF	277/336 (82%)	0.26	25 (9%) 9 7	155, 272, 410, 502	0
6	111	33/50 (66%)	-0.22	2 (6%) 21 16	226, 286, 414, 428	0
7	222	35/50 (70%)	-0.37	0 100 100	175, 280, 384, 460	0
8	333	3/4 (75%)	0.92	0 100 100	176, 176, 176, 210	0
All	All	3587/3763 (95%)	0.09	183 (5%) 28 24	124, 246, 370, 527	0

The worst 5 of 183 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	DDD	1130	GLY	17.7
5	FFF	291	VAL	9.3
5	FFF	310	LEU	7.9
5	FFF	290	ASP	6.2
3	DDD	879	ALA	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(\AA^2)	Q<0.9
9	MG	CCC	1401	1/1	0.27	0.43	138,138,138,138	0
11	DPO	DDD	1504	9/9	0.89	0.44	153,178,187,193	0
9	MG	DDD	1503	1/1	0.90	0.23	105,105,105,105	0
10	ZN	DDD	1501	1/1	0.96	0.04	298,298,298,298	0
10	ZN	DDD	1502	1/1	0.99	0.15	228,228,228,228	0

6.5 Other polymers [i](#)

There are no such residues in this entry.