wwPDB X-ray Structure Validation Summary Report (i)

Sep 3, 2023 – 08:03 PM EDT

PDB ID	:	3S24
Title	:	Crystal structure of human mRNA guanylyltransferase
Authors	:	Das, K.; Chu, C.; Thyminski, J.R.; Bauman, J.D.; Guan, R.; Qiu, W.; Montelione, G.T.; Arnold, E.; Shatkin, A.J.
Deposited on	:	2011-05-16
Resolution	:	3.01 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:

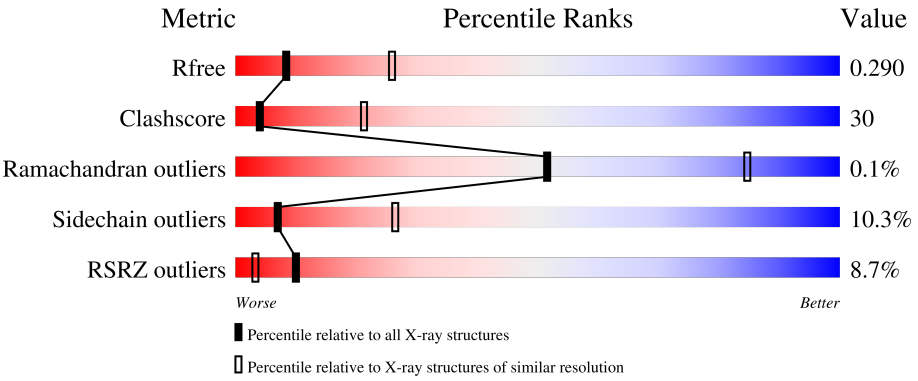
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 3.01 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	2399 (3.04-3.00)
Clashscore	141614	2734 (3.04-3.00)
Ramachandran outliers	138981	2640 (3.04-3.00)
Sidechain outliers	138945	2643 (3.04-3.00)
RSRZ outliers	127900	2287 (3.04-3.00)

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	347	5% 49% 38% 8% 5%
1	B	347	13% 39% 45% 7% 9%
1	C	347	9% 44% 41% 5% 10%
1	D	347	5% 51% 39% 5% 5%
1	E	347	11% 47% 44% • 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	347	
1	G	347	

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
2	SO4	A	5	-	-	X	-

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 18387 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 mRNA-capping enzyme.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	328	Total	C	N	O	S	0	0	0
			2646	1688	457	481	20			
1	B	316	Total	C	N	O	S	0	0	0
			2561	1637	440	465	19			
1	C	314	Total	C	N	O	S	0	0	0
			2547	1628	438	462	19			
1	D	328	Total	C	N	O	S	0	0	0
			2646	1688	457	481	20			
1	E	329	Total	C	N	O	S	0	0	0
			2649	1689	456	483	21			
1	G	329	Total	C	N	O	S	0	0	0
			2651	1691	458	482	20			
1	F	333	Total	C	N	O	S	0	0	0
			2662	1698	460	484	20			

There are 56 discrepancies between the modelled and reference sequences:

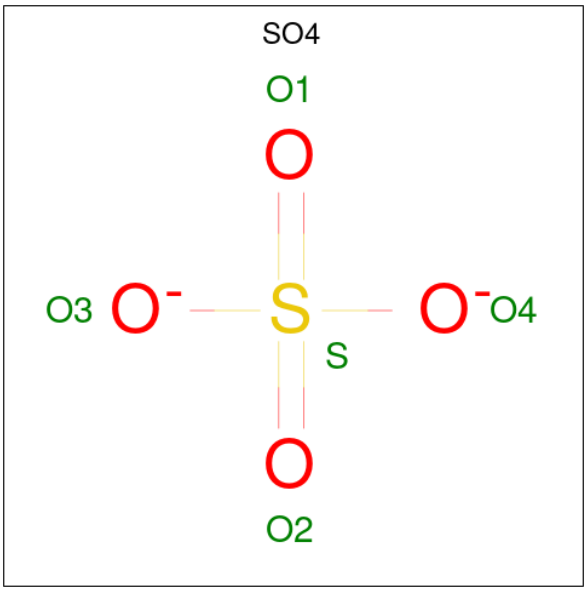
Chain	Residue	Modelled	Actual	Comment	Reference
A	568	LEU	-	expression tag	UNP O60942
A	569	GLU	-	expression tag	UNP O60942
A	570	HIS	-	expression tag	UNP O60942
A	571	HIS	-	expression tag	UNP O60942
A	572	HIS	-	expression tag	UNP O60942
A	573	HIS	-	expression tag	UNP O60942
A	574	HIS	-	expression tag	UNP O60942
A	575	HIS	-	expression tag	UNP O60942
B	568	LEU	-	expression tag	UNP O60942
B	569	GLU	-	expression tag	UNP O60942
B	570	HIS	-	expression tag	UNP O60942
B	571	HIS	-	expression tag	UNP O60942
B	572	HIS	-	expression tag	UNP O60942
B	573	HIS	-	expression tag	UNP O60942
B	574	HIS	-	expression tag	UNP O60942

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	575	HIS	-	expression tag	UNP O60942
C	568	LEU	-	expression tag	UNP O60942
C	569	GLU	-	expression tag	UNP O60942
C	570	HIS	-	expression tag	UNP O60942
C	571	HIS	-	expression tag	UNP O60942
C	572	HIS	-	expression tag	UNP O60942
C	573	HIS	-	expression tag	UNP O60942
C	574	HIS	-	expression tag	UNP O60942
C	575	HIS	-	expression tag	UNP O60942
D	568	LEU	-	expression tag	UNP O60942
D	569	GLU	-	expression tag	UNP O60942
D	570	HIS	-	expression tag	UNP O60942
D	571	HIS	-	expression tag	UNP O60942
D	572	HIS	-	expression tag	UNP O60942
D	573	HIS	-	expression tag	UNP O60942
D	574	HIS	-	expression tag	UNP O60942
D	575	HIS	-	expression tag	UNP O60942
E	568	LEU	-	expression tag	UNP O60942
E	569	GLU	-	expression tag	UNP O60942
E	570	HIS	-	expression tag	UNP O60942
E	571	HIS	-	expression tag	UNP O60942
E	572	HIS	-	expression tag	UNP O60942
E	573	HIS	-	expression tag	UNP O60942
E	574	HIS	-	expression tag	UNP O60942
E	575	HIS	-	expression tag	UNP O60942
G	568	LEU	-	expression tag	UNP O60942
G	569	GLU	-	expression tag	UNP O60942
G	570	HIS	-	expression tag	UNP O60942
G	571	HIS	-	expression tag	UNP O60942
G	572	HIS	-	expression tag	UNP O60942
G	573	HIS	-	expression tag	UNP O60942
G	574	HIS	-	expression tag	UNP O60942
G	575	HIS	-	expression tag	UNP O60942
F	568	LEU	-	expression tag	UNP O60942
F	569	GLU	-	expression tag	UNP O60942
F	570	HIS	-	expression tag	UNP O60942
F	571	HIS	-	expression tag	UNP O60942
F	572	HIS	-	expression tag	UNP O60942
F	573	HIS	-	expression tag	UNP O60942
F	574	HIS	-	expression tag	UNP O60942
F	575	HIS	-	expression tag	UNP O60942

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



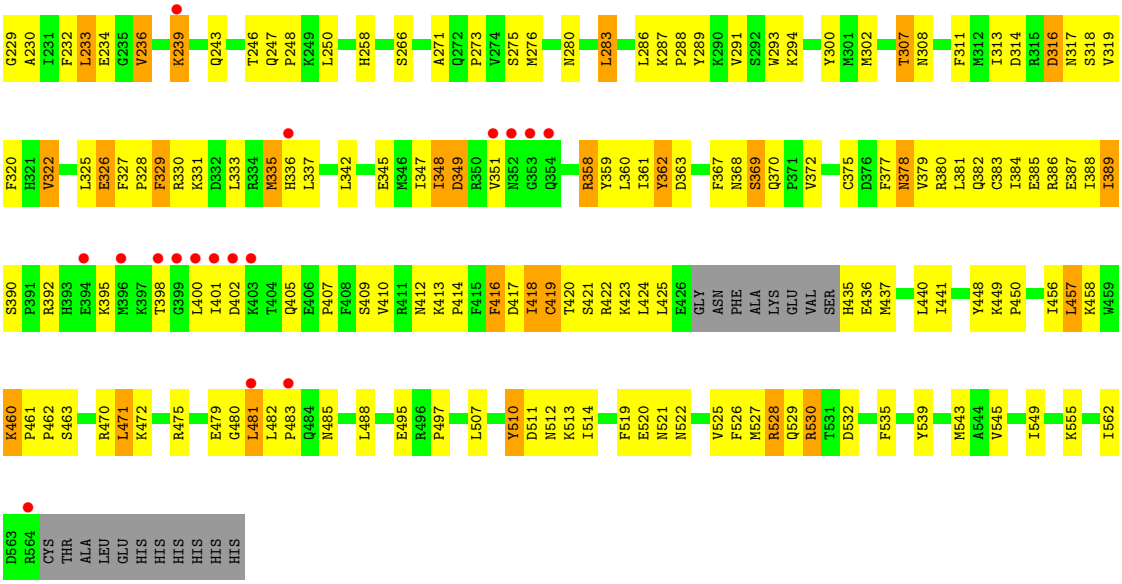
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O S 5 4 1	0	0
2	A	1	Total O S 5 4 1	0	0
2	A	1	Total O S 5 4 1	0	0
2	A	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0

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.

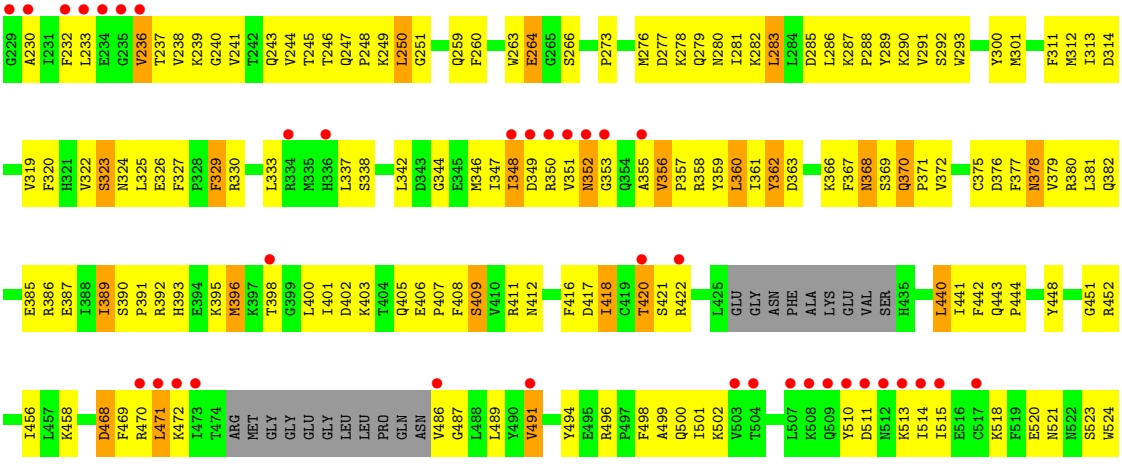
• Molecule 1: mRNA-capping enzyme

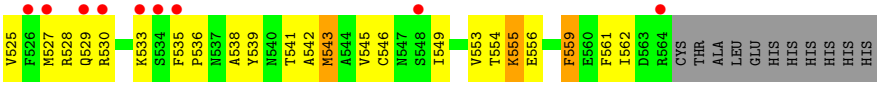
Chain A:



• Molecule 1: mRNA-capping enzyme

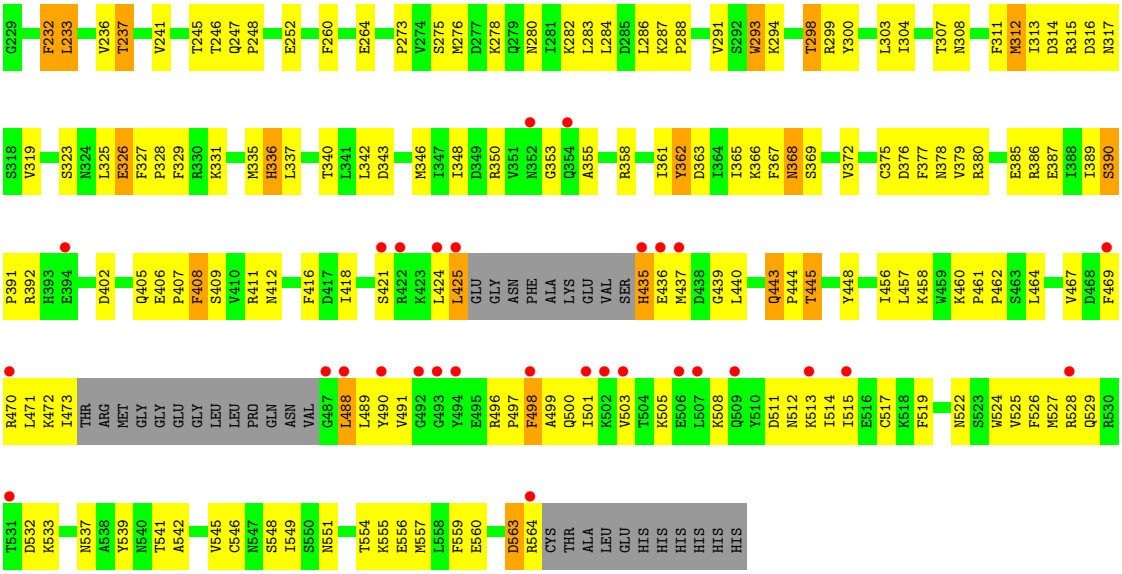
Chain B:





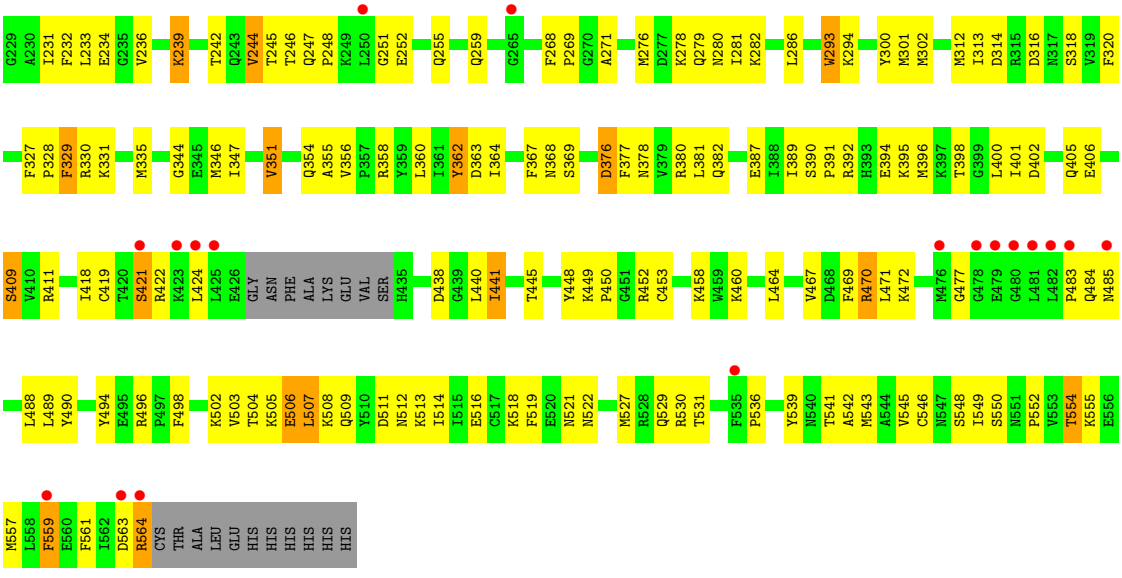
• Molecule 1: mRNA-capping enzyme

Chain C:



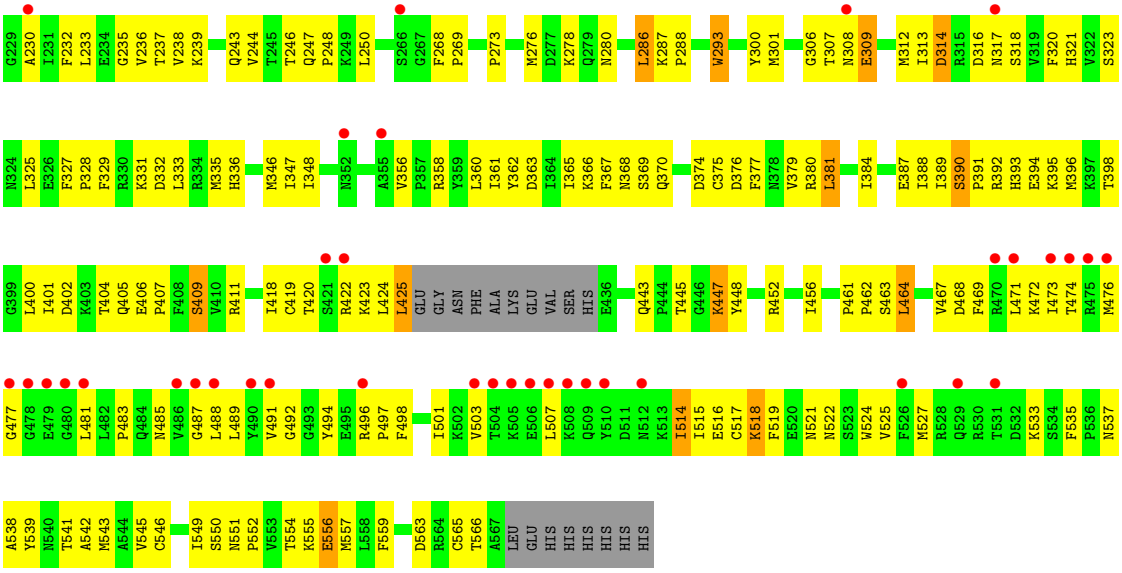
• Molecule 1: mRNA-capping enzyme

Chain D:



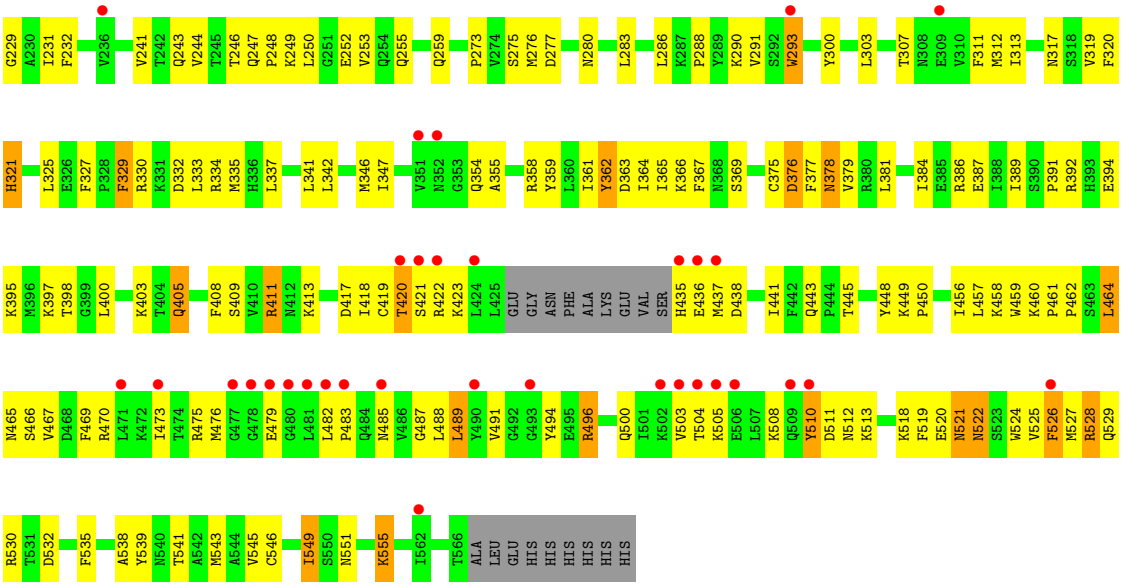
• Molecule 1: mRNA-capping enzyme

Chain E:



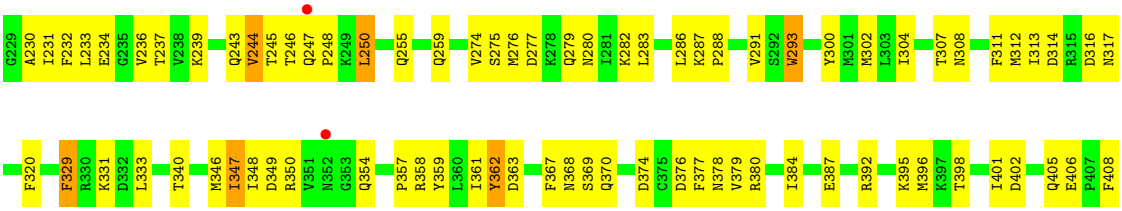
• Molecule 1: mRNA-capping enzyme

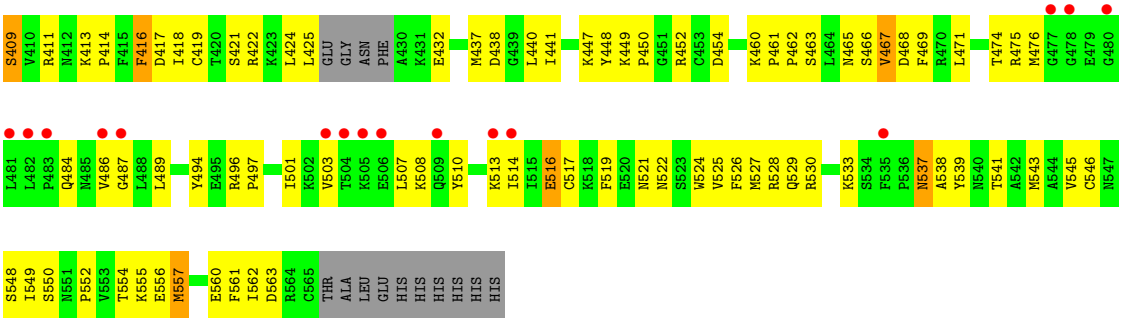
Chain G:



• Molecule 1: mRNA-capping enzyme

Chain F:




S409 V410 R411 N412 K413 P414 F415 F416 D417 I418 C419 T420 S421 R422 K423 L424 L425 GLU GLY ASN PHE A430 K431 E432 M437 D438 G439 L440 I441 K447 Y448 K449 P450 G451 R452 C453 D454 K460 P461 P462 S463 L464 N465 S466 V467 D468 F469 R470 L471 T474 R475 M476 G477 G478 E479 G480
L481 L482 P483 Q484 N485 V486 G487 L488 L489 Y494 E495 R496 P497 I501 K502 V503 T504 K505 E506 L507 K508 Q509 Y510 K513 I514 I515 E516 C517 K518 F519 E520 N521 N522 S523 W524 V525 F526 M527 R528 Q529 R530 K533 S534 F535 P536 N537 A538 Y539 N540 T541 A542 M543 A544 V545 C546 N547
S548 I549 S550 N551 P552 V553 T554 K555 E556 M557 E560 F561 I562 D563 R564 C565 THR ALA LEU GLU HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	92.12Å 104.66Å 149.57Å 90.00° 94.95° 90.00°	Depositor
Resolution (Å)	35.43 – 3.01 35.43 – 3.01	Depositor EDS
% Data completeness (in resolution range)	95.1 (35.43-3.01) 95.0 (35.43-3.01)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.44 (at 3.00Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7_650)	Depositor
R, R_{free}	0.258 , 0.296 0.248 , 0.290	Depositor DCC
R_{free} test set	2724 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å^2)	94.6	Xtriage
Anisotropy	0.139	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 38.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	18387	wwPDB-VP
Average B, all atoms (Å^2)	64.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.25% 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: SO4

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	0.56	0/2704	0.70	1/3643 (0.0%)
1	B	0.48	0/2617	0.65	2/3525 (0.1%)
1	C	0.54	0/2603	0.69	1/3505 (0.0%)
1	D	0.53	0/2704	0.68	1/3643 (0.0%)
1	E	0.55	0/2706	0.73	0/3646
1	F	0.56	0/2719	0.71	2/3665 (0.1%)
1	G	0.49	0/2709	0.66	1/3650 (0.0%)
All	All	0.53	0/18762	0.69	8/25277 (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
1	A	0	2

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	F	362	TYR	CB-CA-C	6.10	122.60	110.40
1	A	362	TYR	CB-CA-C	6.01	122.41	110.40
1	C	362	TYR	CB-CA-C	5.63	121.67	110.40
1	B	362	TYR	CA-CB-CG	5.53	123.90	113.40
1	F	362	TYR	CA-CB-CG	5.30	123.48	113.40

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	351	VAL	Peptide
1	A	480	GLY	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	2646	0	2650	157	0
1	B	2561	0	2567	169	0
1	C	2547	0	2551	150	0
1	D	2646	0	2650	156	0
1	E	2649	0	2658	178	0
1	F	2662	0	2654	161	0
1	G	2651	0	2652	164	0
2	A	20	0	0	2	0
2	B	5	0	0	0	0
All	All	18387	0	18382	1106	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 30.

The worst 5 of 1106 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)
1:B:246:THR:HG22	1:B:248:PRO:HD2	1.22	1.13
1:D:239:LYS:H	1:D:239:LYS:HD3	1.17	1.08
1:C:462:PRO:HD2	1:C:555:LYS:HE3	1.37	1.06
1:A:230:ALA:HA	1:A:243:GLN:HE22	1.21	1.05
1:F:537:ASN:H	1:F:537:ASN:ND2	1.53	1.03

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	324/347 (93%)	313 (97%)	11 (3%)	0	100 100
1	B	310/347 (89%)	291 (94%)	18 (6%)	1 (0%)	41 75
1	C	308/347 (89%)	296 (96%)	11 (4%)	1 (0%)	41 75
1	D	324/347 (93%)	313 (97%)	11 (3%)	0	100 100
1	E	325/347 (94%)	308 (95%)	17 (5%)	0	100 100
1	F	329/347 (95%)	312 (95%)	17 (5%)	0	100 100
1	G	325/347 (94%)	307 (94%)	18 (6%)	0	100 100
All	All	2245/2429 (92%)	2140 (95%)	103 (5%)	2 (0%)	51 85

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	236	VAL
1	C	232	PHE

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	296/313 (95%)	255 (86%)	41 (14%)	3 15
1	B	288/313 (92%)	250 (87%)	38 (13%)	4 17
1	C	286/313 (91%)	258 (90%)	28 (10%)	8 29
1	D	296/313 (95%)	267 (90%)	29 (10%)	8 29

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	E	297/313 (95%)	276 (93%)	21 (7%)	14 44
1	F	294/313 (94%)	273 (93%)	21 (7%)	14 44
1	G	296/313 (95%)	262 (88%)	34 (12%)	5 22
All	All	2053/2191 (94%)	1841 (90%)	212 (10%)	7 26

5 of 212 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	351	VAL
1	E	381	LEU
1	F	378	ASN
1	D	389	ILE
1	D	521	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	280	ASN
1	F	280	ASN
1	F	537	ASN
1	G	529	GLN
1	C	412	ASN

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)

5 ligands are modelled in this entry.

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	$\#\|Z\| > 2$
2	SO4	A	3	-	4,4,4	0.20	0	6,6,6	0.26	0
2	SO4	A	1	-	4,4,4	0.18	0	6,6,6	0.28	0
2	SO4	A	2	-	4,4,4	0.17	0	6,6,6	0.15	0
2	SO4	B	4	-	4,4,4	0.17	0	6,6,6	0.31	0
2	SO4	A	5	-	4,4,4	0.20	0	6,6,6	0.28	0

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.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	5	SO4	2	0

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	328/347 (94%)	0.30	17 (5%) 27 10	28, 51, 102, 136	0
1	B	316/347 (91%)	0.63	46 (14%) 2 1	29, 67, 127, 137	0
1	C	314/347 (90%)	0.46	30 (9%) 8 2	26, 59, 121, 135	0
1	D	328/347 (94%)	0.30	18 (5%) 25 8	28, 55, 112, 143	0
1	E	329/347 (94%)	0.58	37 (11%) 5 1	22, 58, 128, 144	0
1	F	333/347 (95%)	0.29	18 (5%) 25 9	23, 52, 119, 148	0
1	G	329/347 (94%)	0.41	33 (10%) 7 2	30, 61, 127, 149	0
All	All	2277/2429 (93%)	0.43	199 (8%) 10 3	22, 57, 123, 149	0

The worst 5 of 199 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	526	PHE	8.1
1	B	504	THR	7.6
1	E	512	ASN	7.3
1	G	483	PRO	7.1
1	A	564	ARG	7.1

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

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
2	SO4	A	1	5/5	0.84	0.28	68,72,86,94	0
2	SO4	B	4	5/5	0.90	0.19	75,79,82,89	0
2	SO4	A	5	5/5	0.91	0.22	72,75,93,93	0
2	SO4	A	2	5/5	0.96	0.11	47,58,60,63	0
2	SO4	A	3	5/5	0.98	0.28	41,46,53,55	0

6.5 Other polymers

There are no such residues in this entry.