WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Nov 22, 2023 – 03:24 PM JST

PDB ID	:	7YA2
Title	:	Crystal structure of capsular polysaccharide synthesis enzyme CapG from Staphylococcus aureus
Authors	:	Chen, Y.; Wang, Y.C.
Deposited on	:	2022-06-27
Resolution	:	3.20 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

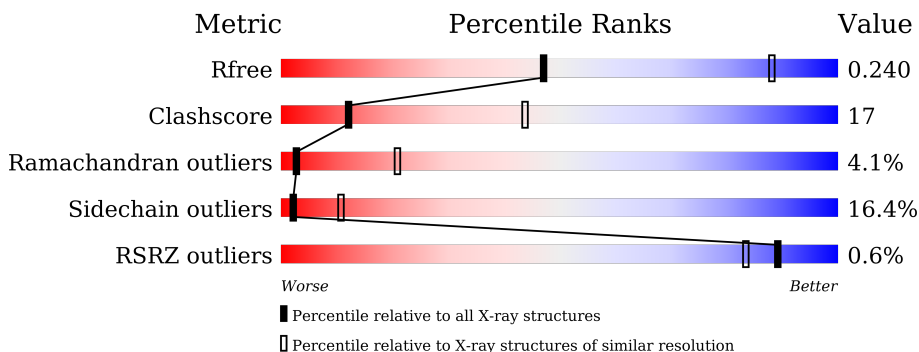|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.20 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	Percentile Ranks	Value
Rfree		0.240
Clashscore		17
Ramachandran outliers		4.1%
Sidechain outliers		16.4%
RSRZ outliers		0.6%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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	374	65% 25% 9% ••
1	B	374	63% 28% 8% •
1	C	374	57% 34% 8% •
1	D	374	58% 31% 10% •
1	E	374	58% 32% 9% •
1	F	374	54% 35% 10% •

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 18050 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 Capsular polysaccharide synthesis enzyme CapG.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	372	Total	C	N	O	S	0	0	0
			2995	1898	512	573	12			
1	B	374	Total	C	N	O	S	0	0	0
			3011	1909	513	577	12			
1	C	374	Total	C	N	O	S	0	0	0
			3011	1909	513	577	12			
1	D	374	Total	C	N	O	S	0	0	0
			3011	1909	513	577	12			
1	E	374	Total	C	N	O	S	0	0	0
			3011	1909	513	577	12			
1	F	374	Total	C	N	O	S	0	0	0
			3011	1909	513	577	12			

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: Capsular polysaccharide synthesis enzyme CapG

Chain A:
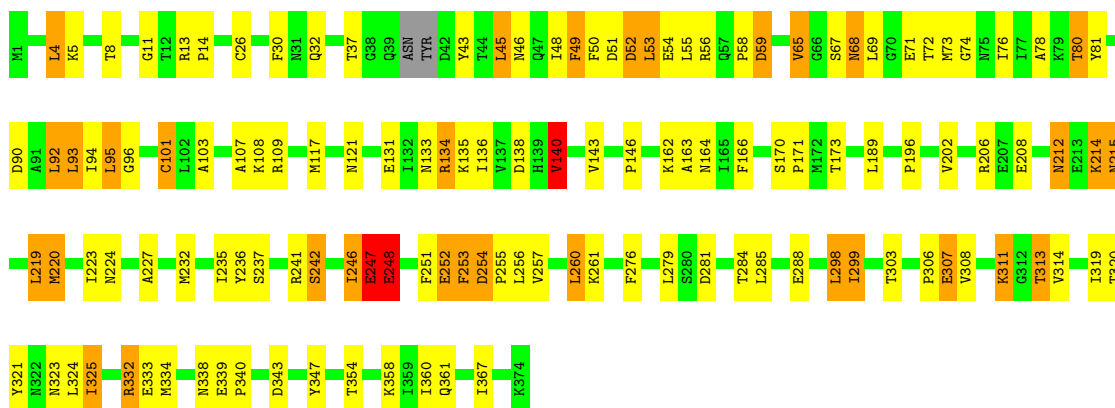
- Molecule 1: Capsular polysaccharide synthesis enzyme CapG

Chain B:
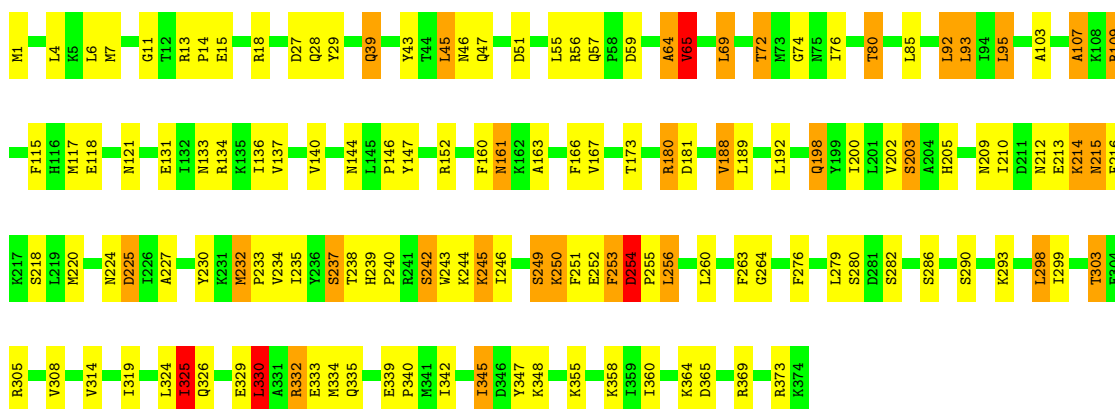
- Molecule 1: Capsular polysaccharide synthesis enzyme CapG

Chain C:

• Molecule 1: Capsular polysaccharide synthesis enzyme CapG

Chain D:

• Molecule 1: Capsular polysaccharide synthesis enzyme CapG

Chain E:

F265 F266 F276 S280 D281 S282 G283 T284 L285 S286 E287 S290 K293 L298 I299 R300 T303 E304 V308 T313 V314 I319 L324 E329 L330 A331 R332 E333 M334 Q335 I345 D346 Y347 S353 T354 K355 K358 I359 I360 Q361 S362 Y363 K364 I367

R373 K374

• Molecule 1: Capsular polysaccharide synthesis enzyme CapG

Chain F:

%

54%   35%   10% •

M1 E2 L6 V10 G11 T12 R13 P14 R18 L19 S20 K24 Q28 T37 G38 Q39 N40 Y41 D42 Y43 T44 L45 N46 Q47 I48 F49 F50 D51 D52 L53 E54 L55 R56 Q57 P58 D59 H60 Y61 L62 E63 A64 V65 L69 T72 M73 G74 N75 I76 T80 V83

L84 L85 R86 L92 L93 I94 L95 G96 D97 S100 A103 A107 K108 R109 H116 M117 E118 A119 C123 V137 V140 S141 D142 V143 N144 L145 P146 Y147 R153 Y154 D157 F160 N161 F166 G169 T173 I176 E177 I183 N184 D187 N190

K191 L192 E195 Q198 Y199 I200 L201 V202 N212 E213 F216 K217 M220 N221 N224 D225 I226 A227 K228 K231 M232 P233 V234 I235 Y236 S237 T238 H239 P240 R241 S242 W243 K244 K245 I246 E247 E248 S249 K250 E252 F251 E253 D254 P255 L256 V257 K258 Q259 L260 F265 F266 D267

Y268 N269 A270 L271 V277 V278 L279 E287 I291 L292 L298 I299 R300 T301 S302 T303 E304 R305 V308 K311 G312 T313 V314 I315 G318 I319 N322 N323 L324 I325 Q326 S327 V328 E329 L330 A331 R332 E333 M334 Q335 E339 A344 Y347 N351 V352 S353 T354 K355

V356 V357 K358 D365 I366 I367 N368 R369 N370 R373 K374

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	302.91Å 84.34Å 145.09Å 90.00° 110.65° 90.00°	Depositor
Resolution (Å)	25.57 – 3.20 25.55 – 3.20	Depositor EDS
% Data completeness (in resolution range)	87.9 (25.57-3.20) 88.1 (25.55-3.20)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.56 (at 3.17Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.183 , 0.241 0.185 , 0.240	Depositor DCC
R_{free} test set	2558 reflections (5.11%)	wwPDB-VP
Wilson B-factor (Å^2)	71.8	Xtriage
Anisotropy	0.047	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 38.9	EDS
L-test for twinning[2]	$< |L| > = 0.48, < L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	18050	wwPDB-VP
Average B, all atoms (Å^2)	74.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.22% 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)

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	0.70	0/3049	0.83	0/4122
1	B	0.71	0/3066	0.82	0/4146
1	C	0.71	0/3066	0.81	0/4146
1	D	0.70	0/3066	0.80	0/4146
1	E	0.70	0/3066	0.79	0/4146
1	F	0.69	0/3066	0.81	0/4146
All	All	0.70	0/18379	0.81	0/24852

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
1	E	0	3
1	F	0	1
All	All	0	6

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	212	ASN	Peptide
1	A	96	GLY	Peptide
1	E	253	PHE	Peptide
1	E	41	TYR	Peptide
1	E	50	PHE	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	2995	0	3006	105	0
1	B	3011	0	3013	99	0
1	C	3011	0	3013	100	0
1	D	3011	0	3013	114	0
1	E	3011	0	3013	110	0
1	F	3011	0	3013	124	0
All	All	18050	0	18071	616	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 17.

The worst 5 of 616 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:C:173:THR:HG21	1:C:347:TYR:O	1.46	1.13
1:B:95:LEU:HD22	1:B:117:MET:HB3	1.45	0.99
1:A:219:LEU:CD2	1:A:299:ILE:HD11	1.95	0.96
1:A:219:LEU:HD21	1:A:299:ILE:HD11	1.48	0.96
1:C:64:ALA:O	1:C:65:VAL:HG12	1.65	0.95

There are no symmetry-related clashes.

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	A	368/374 (98%)	301 (82%)	55 (15%)	12 (3%)	4	25

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	372/374 (100%)	305 (82%)	50 (13%)	17 (5%)	2	18
1	C	372/374 (100%)	311 (84%)	45 (12%)	16 (4%)	2	20
1	D	372/374 (100%)	312 (84%)	46 (12%)	14 (4%)	3	22
1	E	372/374 (100%)	306 (82%)	52 (14%)	14 (4%)	3	22
1	F	372/374 (100%)	299 (80%)	55 (15%)	18 (5%)	2	17
All	All	2228/2244 (99%)	1834 (82%)	303 (14%)	91 (4%)	3	21

5 of 91 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	65	VAL
1	A	134	ARG
1	A	140	VAL
1	A	252	GLU
1	A	254	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	339/341 (99%)	288 (85%)	51 (15%)	3	14
1	B	341/341 (100%)	290 (85%)	51 (15%)	3	14
1	C	341/341 (100%)	280 (82%)	61 (18%)	2	9
1	D	341/341 (100%)	286 (84%)	55 (16%)	2	11
1	E	341/341 (100%)	284 (83%)	57 (17%)	2	10
1	F	341/341 (100%)	281 (82%)	60 (18%)	2	9
All	All	2044/2046 (100%)	1709 (84%)	335 (16%)	2	11

5 of 335 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	94	ILE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	F	92	LEU
1	E	181	ASP
1	E	263	PHE
1	F	198	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 69 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	39	GLN
1	F	116	HIS
1	F	224	ASN
1	C	161	ASN
1	C	75	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 ⓘ

There are no ligands in this entry.

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	372/374 (99%)	-0.58	0	100	100	32, 60, 105, 136	0
1	B	374/374 (100%)	-0.54	0	100	100	34, 65, 122, 161	0
1	C	374/374 (100%)	-0.47	5 (1%)	77	65	35, 69, 115, 144	0
1	D	374/374 (100%)	-0.42	3 (0%)	86	78	35, 66, 124, 189	0
1	E	374/374 (100%)	-0.43	3 (0%)	86	78	42, 69, 129, 179	0
1	F	374/374 (100%)	-0.44	3 (0%)	86	78	41, 74, 125, 185	0
All	All	2242/2244 (99%)	-0.48	14 (0%)	89	83	32, 67, 121, 189	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	41	TYR	3.7
1	D	251	PHE	3.5
1	F	41	TYR	3.2
1	C	1	MET	3.0
1	C	41	TYR	3.0

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 ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.