wwPDB X-ray Structure Validation Summary Report (i)

Nov 6, 2023 – 02:59 AM EST

PDB ID	:	5E9T
Title	:	Crystal structure of GtfA/B complex
Authors	:	Chen, Y.; Rapoport, T.A.
Deposited on	:	2015-10-15
Resolution	:	2.92 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 (1)) were used in the production of this report:

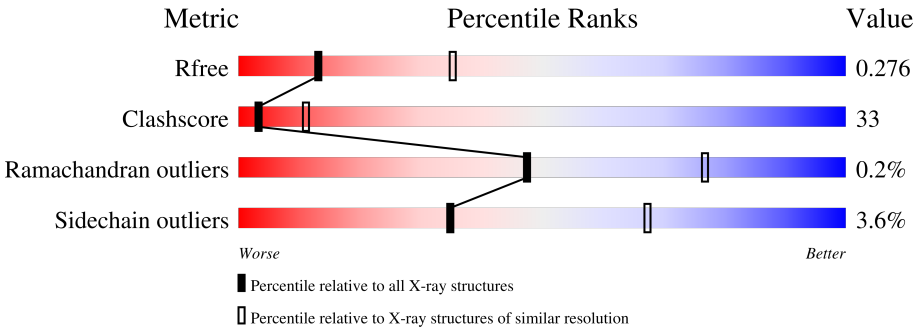
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.92 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	2307 (2.94-2.90)
Clashscore	141614	2531 (2.94-2.90)
Ramachandran outliers	138981	2462 (2.94-2.90)
Sidechain outliers	138945	2464 (2.94-2.90)

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%

Mol	Chain	Length	Quality of chain
1	A	503	60% 36%
1	C	503	55% 40% 5%
2	B	447	53% 44%
2	D	447	50% 45% 5%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 15474 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 Glycosyltransferase Gtf1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	503	Total	C	N	O	S	Se	0	0	0
			4108	2624	687	789	2	6			
1	C	503	Total	C	N	O	S	Se	0	0	0
			4108	2624	687	789	2	6			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	SER	-	expression tag	UNP Q9AET5
A	124	PHE	SER	engineered mutation	UNP Q9AET5
C	1	SER	-	expression tag	UNP Q9AET5
C	124	PHE	SER	engineered mutation	UNP Q9AET5

- Molecule 2 is a protein called Glycosyltransferase-stabilizing protein Gtf2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	447	Total	C	N	O	S	Se	0	0	0
			3615	2302	615	687	3	8			
2	D	445	Total	C	N	O	S	Se	0	0	0
			3607	2298	613	685	3	8			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MSE	-	initiating methionine	UNP Q79T00
D	1	MSE	-	initiating methionine	UNP Q79T00

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Mg	0	0
			1	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total Mg 1 1	0	0

- Molecule 4 is water.

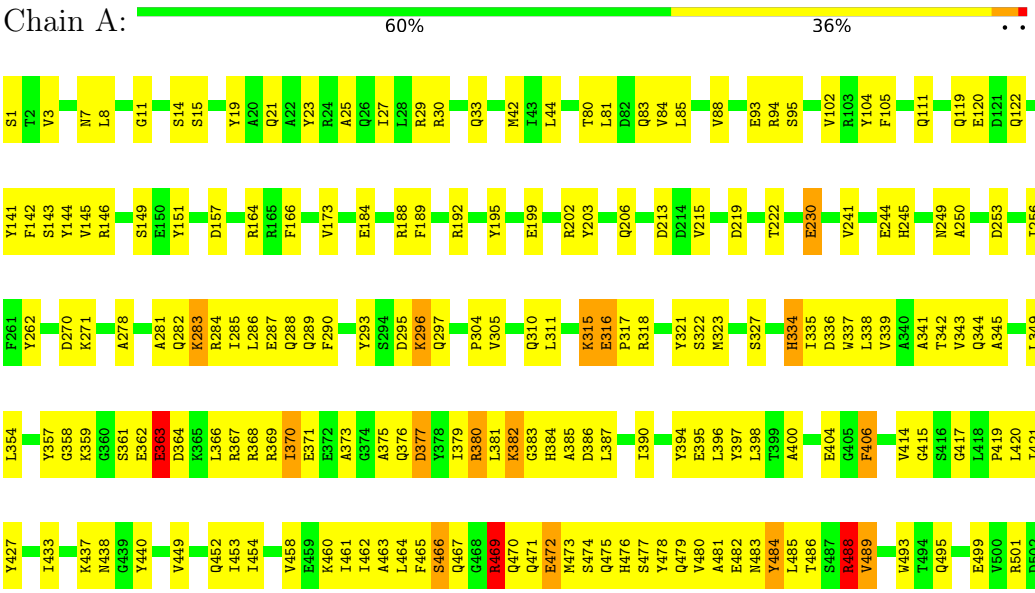
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	10	Total O 10 10	0	0
4	B	9	Total O 9 9	0	0
4	C	8	Total O 8 8	0	0
4	D	7	Total O 7 7	0	0

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. 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. 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.

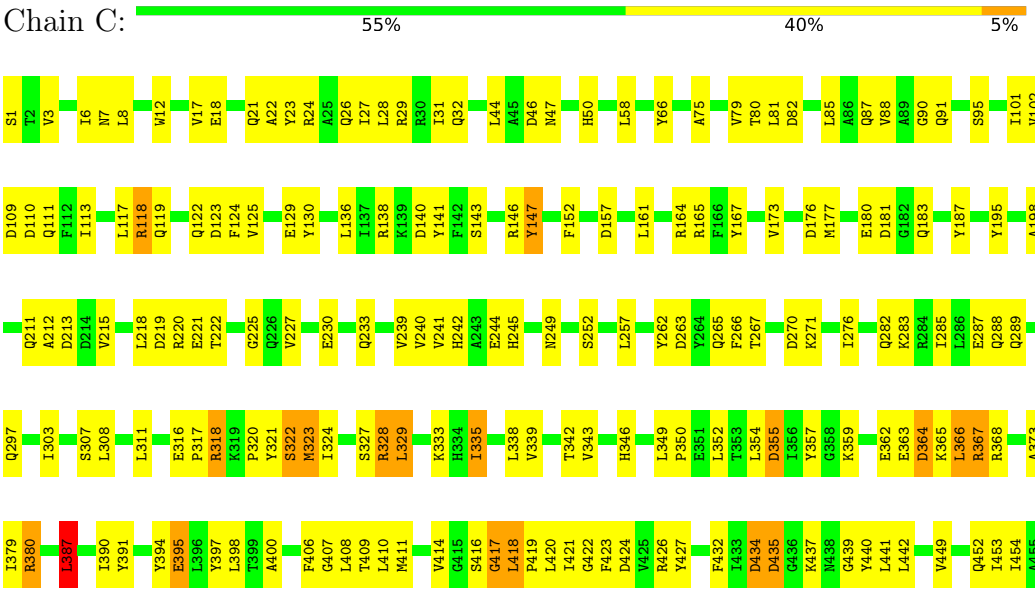
- Molecule 1: Glycosyltransferase Gtf1

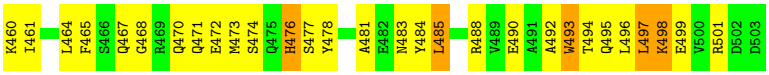
Chain A:



- Molecule 1: Glycosyltransferase Gtf1

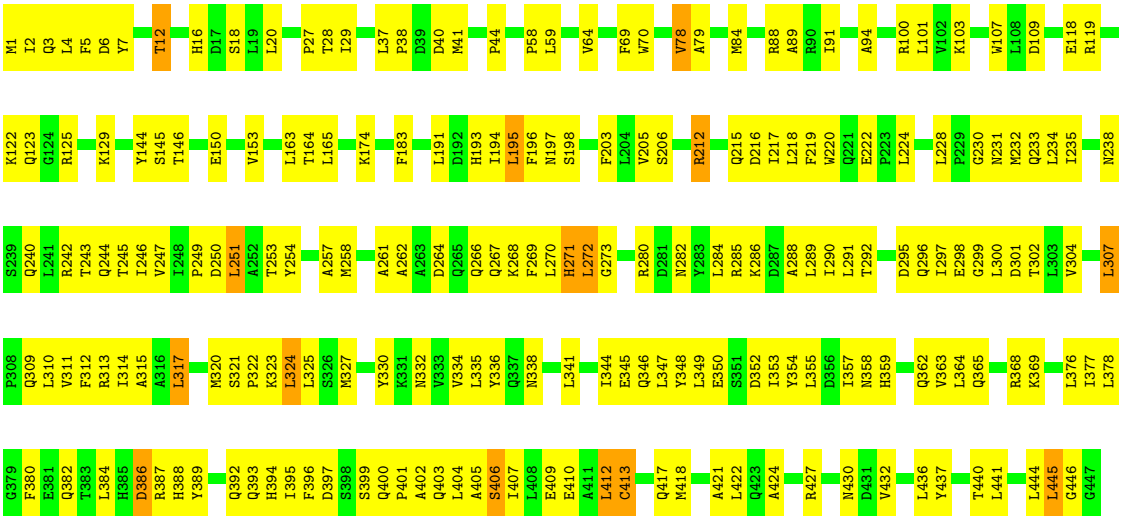
Chain C:





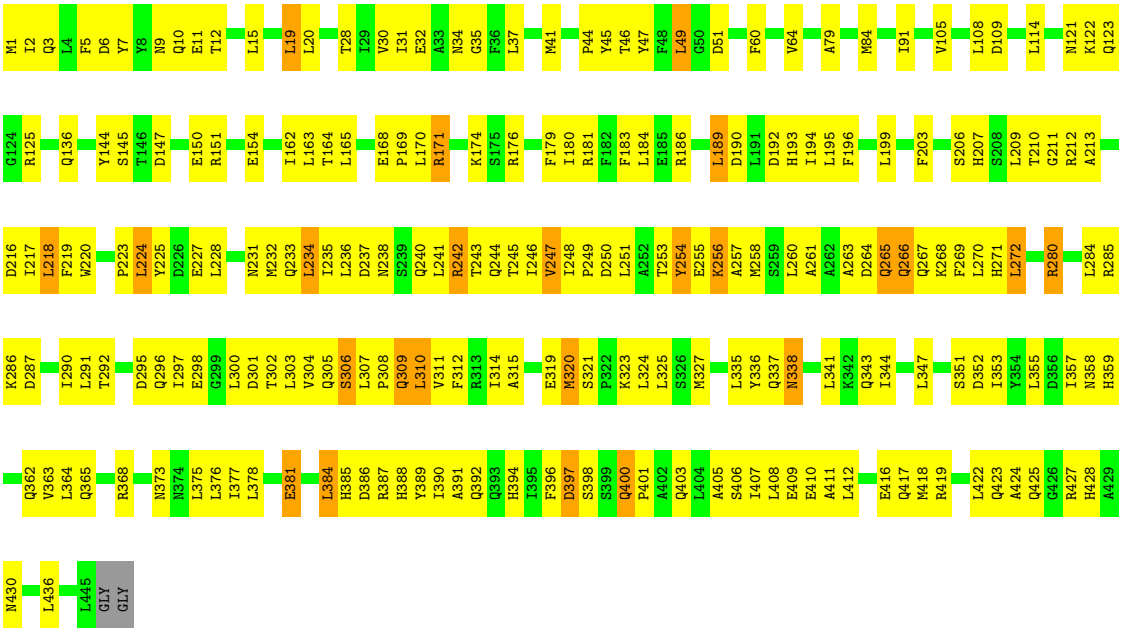
• Molecule 2: Glycosyltransferase-stabilizing protein Gtf2

Chain B:



• Molecule 2: Glycosyltransferase-stabilizing protein Gtf2

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	155.93Å 196.41Å 220.86Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	122.12 – 2.92 122.12 – 2.62	Depositor EDS
% Data completeness (in resolution range)	99.9 (122.12-2.92) 94.8 (122.12-2.62)	Depositor EDS
R_{merge}	0.24	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.95 (at 2.62Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.217 , 0.275 0.218 , 0.276	Depositor DCC
R_{free} test set	4641 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å^2)	30.0	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 59.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.86	EDS
Total number of atoms	15474	wwPDB-VP
Average B, all atoms (Å^2)	68.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.85% 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: 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	0.64	14/4196 (0.3%)	0.81	11/5678 (0.2%)
1	C	0.49	2/4196 (0.0%)	0.78	12/5678 (0.2%)
2	B	0.39	0/3686	0.67	8/4986 (0.2%)
2	D	0.54	5/3678 (0.1%)	0.87	19/4976 (0.4%)
All	All	0.53	21/15756 (0.1%)	0.79	50/21318 (0.2%)

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	C	0	5
2	D	0	3
All	All	0	10

The worst 5 of 21 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	472	GLU	CD-OE1	-10.31	1.14	1.25
1	A	316	GLU	CD-OE2	-9.69	1.15	1.25
1	A	469	ARG	NE-CZ	-9.68	1.20	1.33
1	A	316	GLU	CD-OE1	-9.50	1.15	1.25
2	D	247	VAL	CB-CG1	-7.63	1.36	1.52

The worst 5 of 50 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	488	ARG	NE-CZ-NH2	25.97	133.29	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	329	LEU	CA-CB-CG	16.30	152.78	115.30
1	A	488	ARG	NE-CZ-NH1	-14.02	113.29	120.30
2	D	256	LYS	CD-CE-NZ	-12.70	82.49	111.70
1	C	364	ASP	CB-CG-OD2	-12.15	107.37	118.30

There are no chirality outliers.

5 of 10 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	363	GLU	Peptide
1	A	466	SER	Peptide
1	C	108	GLN	Peptide
1	C	322	SER	Peptide
1	C	328	ARG	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	4108	0	3965	240	0
1	C	4108	0	3965	271	0
2	B	3615	0	3535	210	0
2	D	3607	0	3529	306	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	10	0	0	0	0
4	B	9	0	0	0	0
4	C	8	0	0	0	0
4	D	7	0	0	1	0
All	All	15474	0	14994	1005	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 33.

The worst 5 of 1005 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:D:307:LEU:HA	2:D:309:GLN:NE2	1.64	1.12
1:A:470:GLN:HE21	1:A:471:GLN:HG3	1.06	1.12
2:B:345:GLU:OE2	2:B:369:LYS:NZ	1.83	1.12
2:D:307:LEU:HA	2:D:309:GLN:HE21	1.13	1.09
2:B:272:LEU:HD13	2:B:273:GLY:H	0.92	1.07

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	501/503 (100%)	482 (96%)	16 (3%)	3 (1%)	25 57
1	C	501/503 (100%)	480 (96%)	21 (4%)	0	100 100
2	B	445/447 (100%)	418 (94%)	26 (6%)	1 (0%)	47 77
2	D	443/447 (99%)	413 (93%)	30 (7%)	0	100 100
All	All	1890/1900 (100%)	1793 (95%)	93 (5%)	4 (0%)	47 77

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	316	GLU
1	A	474	SER
1	A	370	ILE
2	B	78	VAL

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	435/429 (101%)	414 (95%)	21 (5%)	25	57
1	C	435/429 (101%)	421 (97%)	14 (3%)	39	71
2	B	392/384 (102%)	381 (97%)	11 (3%)	43	75
2	D	392/384 (102%)	379 (97%)	13 (3%)	38	70
All	All	1654/1626 (102%)	1595 (96%)	59 (4%)	35	67

5 of 59 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	323	LYS
2	D	280	ARG
1	C	157	ASP
2	D	272	LEU
2	D	60	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	163	GLN
2	D	430	ASN
1	C	211	GLN
2	D	309	GLN
1	C	209	GLN

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)

Of 2 ligands modelled in this entry, 2 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)

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.