



wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 01:07 AM GMT

PDB ID : 2JJM
Title : CRYSTAL STRUCTURE OF A FAMILY GT4 GLYCOSYLTRANSFERASE
FROM BACILLUS ANTHRACIS ORF BA1558.
Authors : Ruane, K.M.; Davies, G.J.; Martinez-Fleites, C.
Deposited on : 2008-04-15
Resolution : 3.10 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

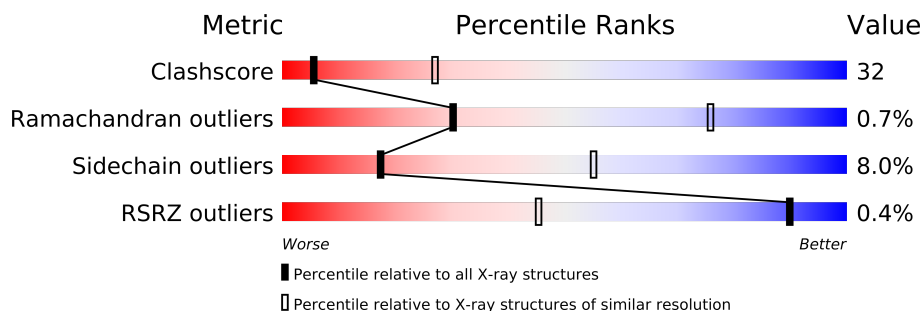
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

1 Overall quality at a glance

The reported resolution of this entry is 3.10 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))
Clashscore	79885	1078 (3.16-3.04)
Ramachandran outliers	78287	1044 (3.16-3.04)
Sidechain outliers	78261	1044 (3.16-3.04)
RSRZ outliers	66119	1008 (3.18-3.02)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	394	
1	B	394	
1	C	394	
1	D	394	
1	E	394	
1	F	394	
1	G	394	
1	H	394	
1	I	394	
1	J	394	
1	K	394	
1	L	394	

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 33924 atoms, of which 0 are hydrogen and 0 are deuterium.

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 GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN.

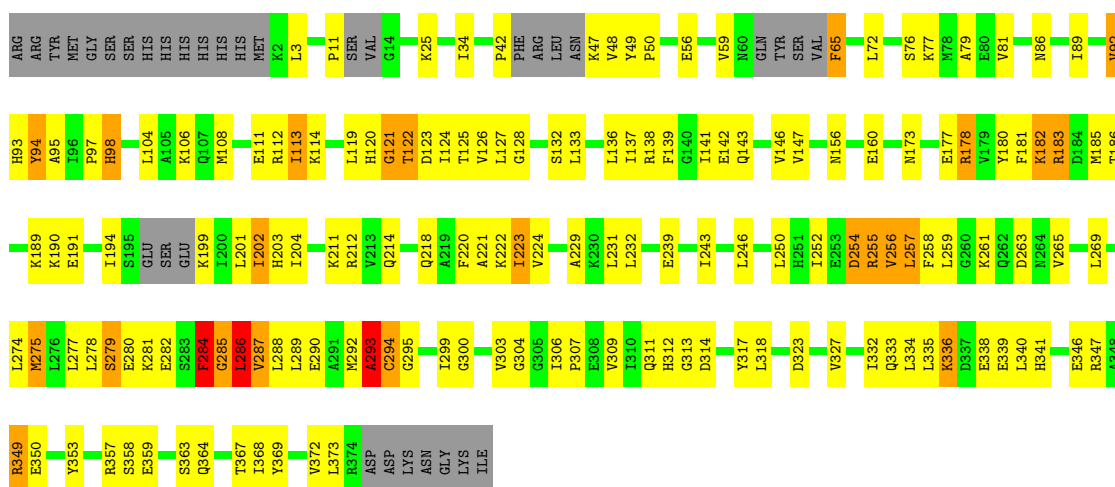
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	360	Total	C	N	O	S	0	0	1
			2827	1806	477	531	13			
1	B	360	Total	C	N	O	S	0	0	1
			2827	1806	477	531	13			
1	C	360	Total	C	N	O	S	0	0	1
			2827	1806	477	531	13			
1	D	360	Total	C	N	O	S	0	0	1
			2827	1806	477	531	13			
1	E	360	Total	C	N	O	S	0	0	1
			2827	1806	477	531	13			
1	F	360	Total	C	N	O	S	0	0	1
			2827	1806	477	531	13			
1	G	360	Total	C	N	O	S	0	0	1
			2827	1806	477	531	13			
1	H	360	Total	C	N	O	S	0	0	1
			2827	1806	477	531	13			
1	I	360	Total	C	N	O	S	0	0	1
			2827	1806	477	531	13			
1	J	360	Total	C	N	O	S	0	0	1
			2827	1806	477	531	13			
1	K	360	Total	C	N	O	S	0	0	1
			2827	1806	477	531	13			
1	L	360	Total	C	N	O	S	0	0	1
			2827	1806	477	531	13			

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

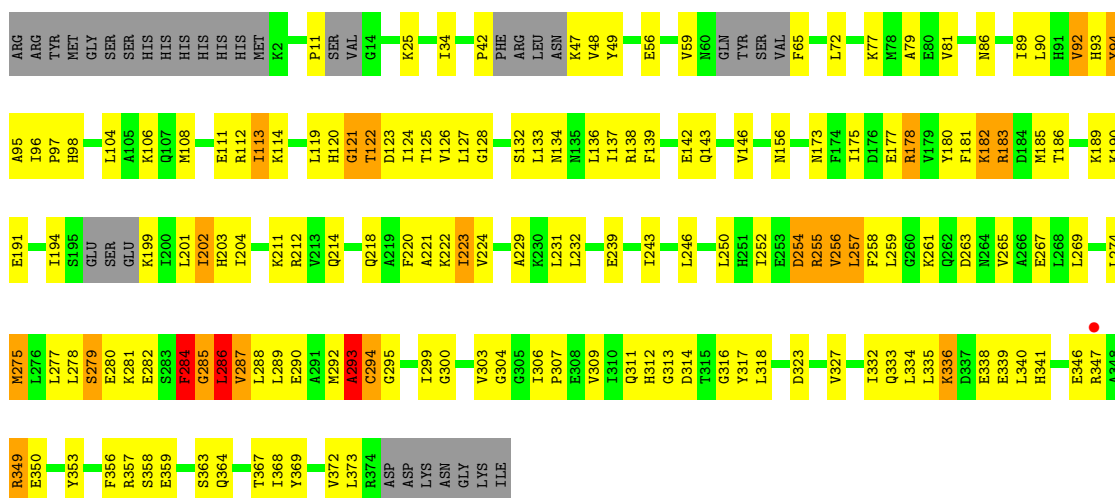
• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

Chain A:



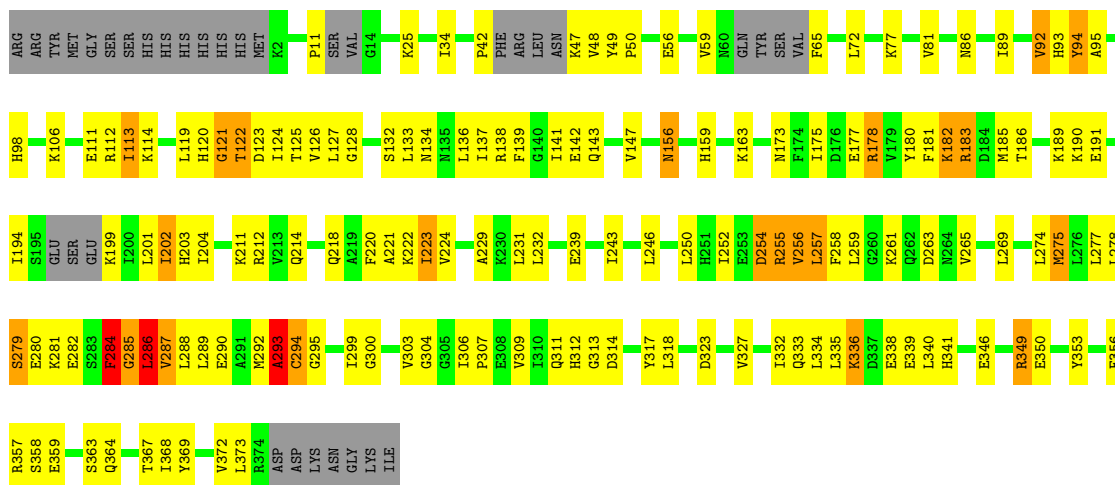
• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

Chain B:



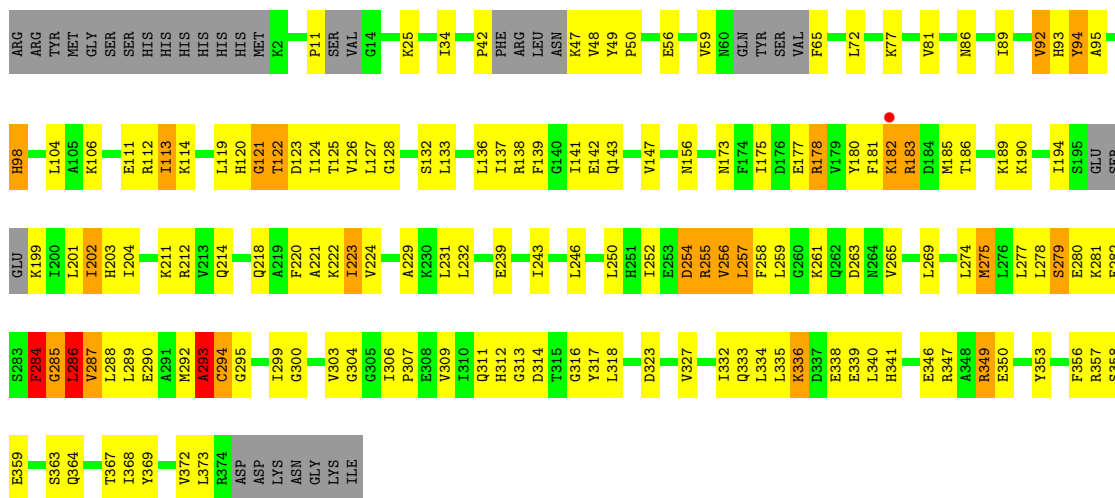
• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

Chain C:



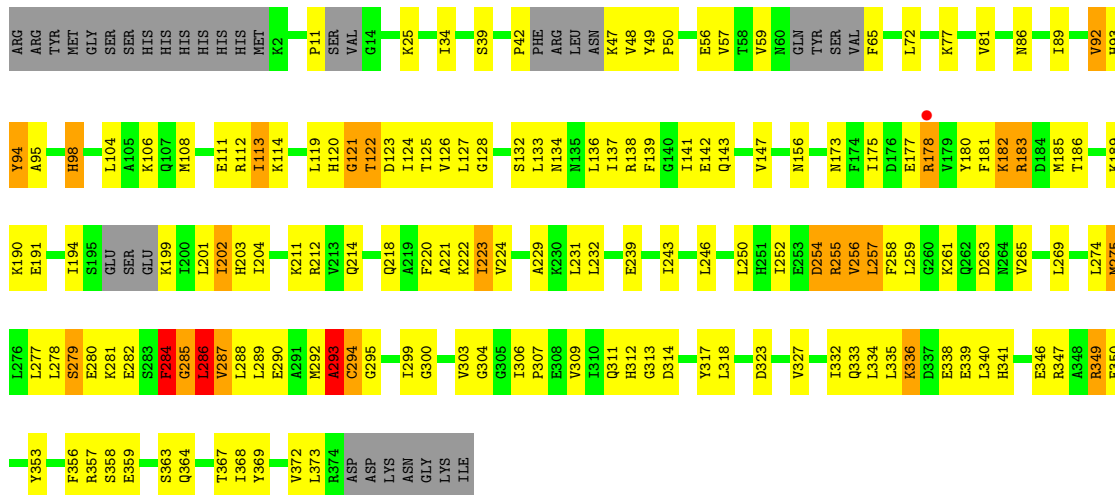
• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

Chain D:



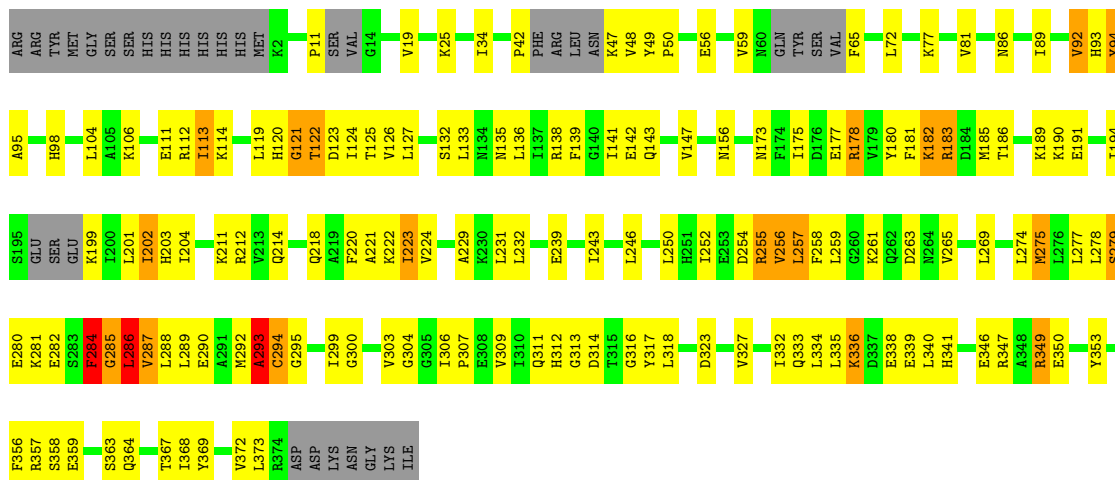
• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

Chain E:



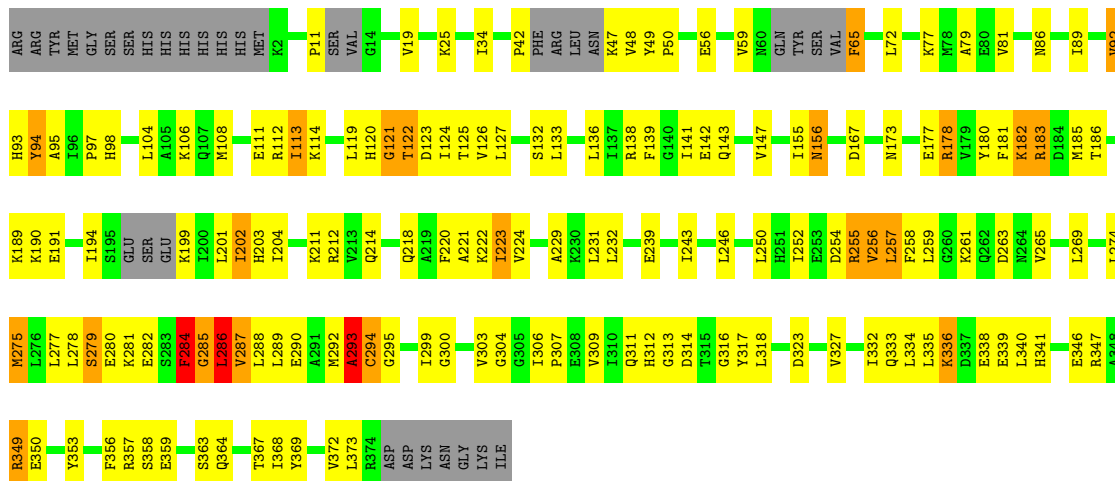
● Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

Chain F:



● Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

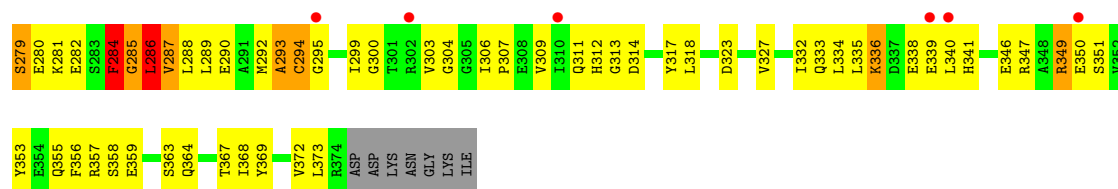
Chain G:



● Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

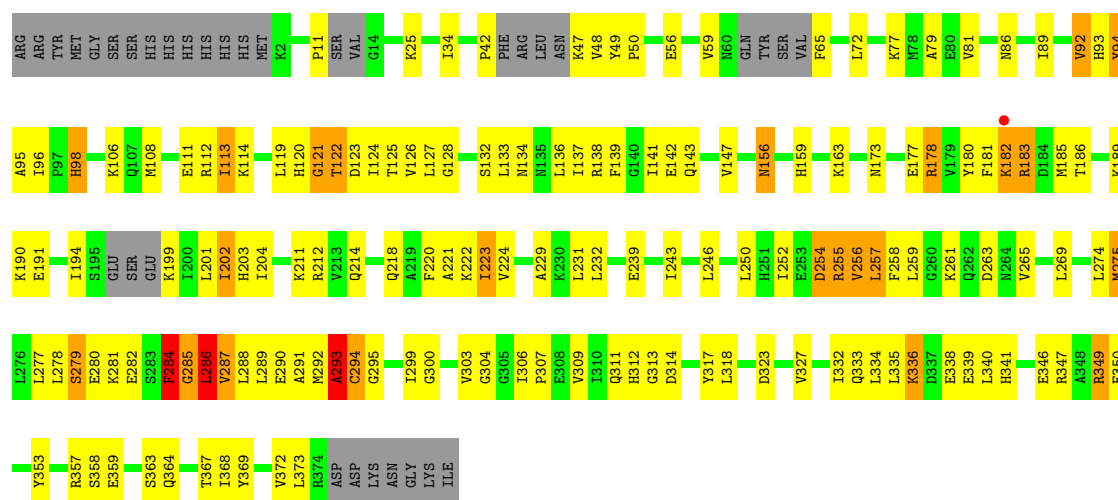
Chain H:





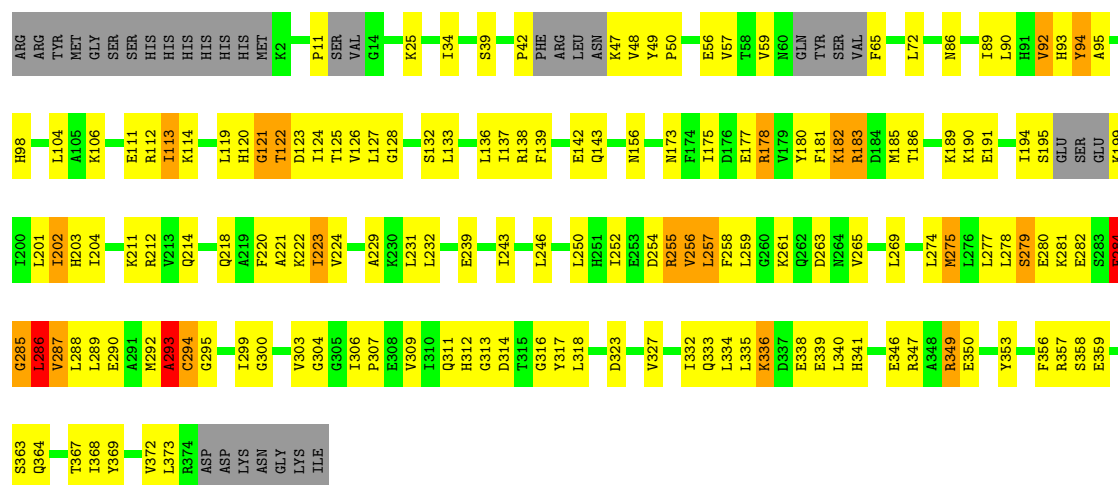
• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

Chain I:



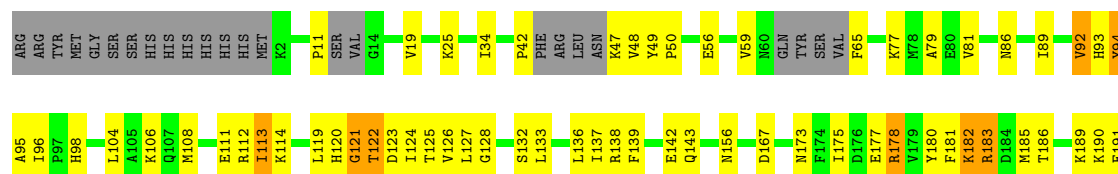
• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

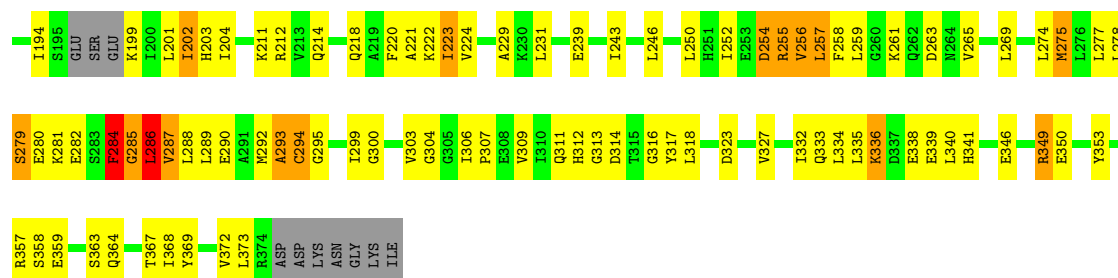
Chain J:



• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

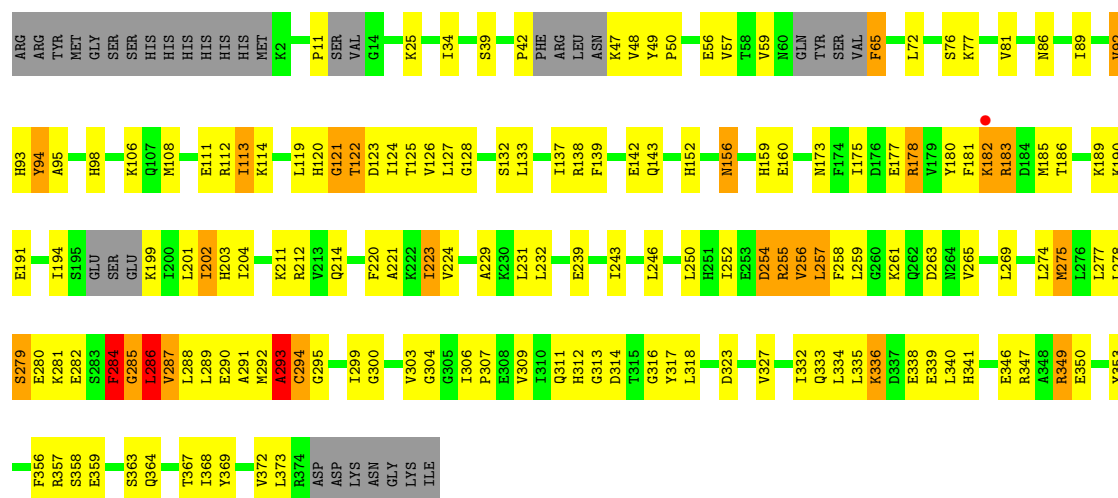
Chain K:





• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

Chain L:



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	134.60Å 204.67Å 135.33Å 90.00° 115.49° 90.00°	Depositor
Resolution (Å)	20.00 – 3.10 78.27 – 2.70	Depositor EDS
% Data completeness (in resolution range)	98.3 (20.00-3.10) 89.5 (78.27-2.70)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.39 (at 2.69Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.254 , 0.262 0.266 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	DCC
Wilson B-factor (Å ²)	88.0	Xtriage
Anisotropy	0.099	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 59.6	EDS
Estimated twinning fraction	0.022 for l,-k,h	Xtriage
L-test for twinning	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 161831 reflections	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	33924	wwPDB-VP
Average B, all atoms (Å ²)	91.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.93% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

5 Model quality

5.1 Standard geometry

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	0.86	8/2873 (0.3%)	0.89	10/3884 (0.3%)
1	B	0.86	7/2873 (0.2%)	0.90	10/3884 (0.3%)
1	C	0.87	8/2873 (0.3%)	0.89	10/3884 (0.3%)
1	D	0.86	8/2873 (0.3%)	0.89	10/3884 (0.3%)
1	E	0.87	8/2873 (0.3%)	0.89	10/3884 (0.3%)
1	F	0.86	8/2873 (0.3%)	0.89	9/3884 (0.2%)
1	G	0.87	8/2873 (0.3%)	0.90	9/3884 (0.2%)
1	H	0.86	6/2873 (0.2%)	0.90	9/3884 (0.2%)
1	I	0.87	8/2873 (0.3%)	0.90	10/3884 (0.3%)
1	J	0.89	8/2873 (0.3%)	0.89	8/3884 (0.2%)
1	K	0.87	7/2873 (0.2%)	0.89	10/3884 (0.3%)
1	L	0.87	7/2873 (0.2%)	0.89	10/3884 (0.3%)
All	All	0.87	91/34476 (0.3%)	0.89	115/46608 (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	B	0	2
1	C	0	2
1	D	0	2
1	E	0	2
1	F	0	2
1	G	0	2
1	H	0	2
1	I	0	3
1	J	0	2
1	K	0	2
1	L	0	2
All	All	0	25

The worst 5 of 91 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	J	287	VAL	CA-CB	-7.51	1.39	1.54
1	F	287	VAL	CA-CB	-7.43	1.39	1.54
1	D	287	VAL	CA-CB	-7.41	1.39	1.54
1	I	287	VAL	CA-CB	-7.41	1.39	1.54
1	G	287	VAL	CA-CB	-7.39	1.39	1.54

The worst 5 of 115 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	L	121	GLY	N-CA-C	9.49	136.84	113.10
1	J	121	GLY	N-CA-C	9.48	136.80	113.10
1	H	121	GLY	N-CA-C	9.46	136.74	113.10
1	B	121	GLY	N-CA-C	9.45	136.72	113.10
1	D	121	GLY	N-CA-C	9.44	136.70	113.10

There are no chirality outliers.

5 of 25 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	284	PHE	Mainchain,Peptide
1	B	284	PHE	Mainchain,Peptide
1	C	284	PHE	Mainchain,Peptide
1	D	284	PHE	Mainchain,Peptide
1	E	284	PHE	Mainchain

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2827	0	2856	182	1
1	B	2827	0	2856	191	0
1	C	2827	0	2856	188	0
1	D	2827	0	2856	182	0
1	E	2827	0	2856	182	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	F	2827	0	2856	185	0
1	G	2827	0	2856	195	0
1	H	2827	0	2856	187	0
1	I	2827	0	2856	189	0
1	J	2827	0	2856	183	0
1	K	2827	0	2856	186	1
1	L	2827	0	2856	190	0
All	All	33924	0	34272	2180	1

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 32.

The worst 5 of 2180 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:G:167:ASP:OD2	1:L:160:GLU:HA	1.15	1.27
1:A:286:LEU:HD12	1:A:286:LEU:N	1.55	1.17
1:B:286:LEU:CD1	1:B:286:LEU:H	1.54	1.16
1:K:280:GLU:O	1:K:281:LYS:HG2	1.46	1.16
1:G:280:GLU:O	1:G:281:LYS:HG2	1.47	1.15

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	Distance(Å)	Clash(Å)
1:A:160:GLU:CA	1:K:167:ASP:OD2[1_455]	2.08	0.12

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	350/394 (89%)	323 (92%)	24 (7%)	3 (1%)	25 71
1	B	350/394 (89%)	322 (92%)	26 (7%)	2 (1%)	33 78
1	C	350/394 (89%)	323 (92%)	25 (7%)	2 (1%)	33 78

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	350/394 (89%)	322 (92%)	25 (7%)	3 (1%)	25	71
1	E	350/394 (89%)	323 (92%)	24 (7%)	3 (1%)	25	71
1	F	350/394 (89%)	322 (92%)	26 (7%)	2 (1%)	33	78
1	G	350/394 (89%)	322 (92%)	26 (7%)	2 (1%)	33	78
1	H	350/394 (89%)	322 (92%)	26 (7%)	2 (1%)	33	78
1	I	350/394 (89%)	322 (92%)	25 (7%)	3 (1%)	25	71
1	J	350/394 (89%)	322 (92%)	26 (7%)	2 (1%)	33	78
1	K	350/394 (89%)	322 (92%)	26 (7%)	2 (1%)	33	78
1	L	350/394 (89%)	323 (92%)	25 (7%)	2 (1%)	33	78
All	All	4200/4728 (89%)	3868 (92%)	304 (7%)	28 (1%)	30	76

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	293	ALA
1	A	304	GLY
1	B	293	ALA
1	B	304	GLY
1	C	293	ALA

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	312/346 (90%)	287 (92%)	25 (8%)	17	55
1	B	312/346 (90%)	287 (92%)	25 (8%)	17	55
1	C	312/346 (90%)	287 (92%)	25 (8%)	17	55
1	D	312/346 (90%)	287 (92%)	25 (8%)	17	55
1	E	312/346 (90%)	287 (92%)	25 (8%)	17	55
1	F	312/346 (90%)	287 (92%)	25 (8%)	17	55
1	G	312/346 (90%)	287 (92%)	25 (8%)	17	55
1	H	312/346 (90%)	287 (92%)	25 (8%)	17	55

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	I	312/346 (90%)	287 (92%)	25 (8%)	17	55
1	J	312/346 (90%)	286 (92%)	26 (8%)	16	53
1	K	312/346 (90%)	287 (92%)	25 (8%)	17	55
1	L	312/346 (90%)	287 (92%)	25 (8%)	17	55
All	All	3744/4152 (90%)	3443 (92%)	301 (8%)	17	55

5 of 301 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	183	ARG
1	G	278	LEU
1	L	65	PHE
1	F	255	ARG
1	G	65	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 111 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	173	ASN
1	G	330	GLN
1	L	134	ASN
1	F	214	GLN
1	G	98	HIS

5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates 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	360/394 (91%)	-0.04	0 100 100	62, 92, 116, 125	0
1	B	360/394 (91%)	0.11	1 (0%) 91 53	63, 94, 116, 126	0
1	C	360/394 (91%)	0.07	0 100 100	57, 93, 116, 125	0
1	D	360/394 (91%)	0.01	1 (0%) 91 53	58, 93, 115, 126	0
1	E	360/394 (91%)	0.04	1 (0%) 91 53	56, 92, 116, 125	0
1	F	360/394 (91%)	0.08	0 100 100	57, 93, 116, 125	0
1	G	360/394 (91%)	-0.00	0 100 100	62, 93, 116, 125	0
1	H	360/394 (91%)	0.33	11 (3%) 47 7	63, 95, 117, 126	0
1	I	360/394 (91%)	0.08	1 (0%) 91 53	57, 92, 116, 125	0
1	J	360/394 (91%)	0.03	0 100 100	38, 87, 116, 124	0
1	K	360/394 (91%)	0.03	0 100 100	55, 90, 116, 124	0
1	L	360/394 (91%)	-0.00	1 (0%) 91 53	61, 92, 115, 124	0
All	All	4320/4728 (91%)	0.06	16 (0%) 90 45	38, 92, 116, 126	0

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	177	GLU	3.3
1	H	190	LYS	3.2
1	I	182	LYS	3.1
1	E	178	ARG	3.1
1	D	182	LYS	2.8

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 carbohydrates 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.