wwPDB X-ray Structure Validation Summary Report (i)

Apr 28, 2024 – 02:40 am BST

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 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:

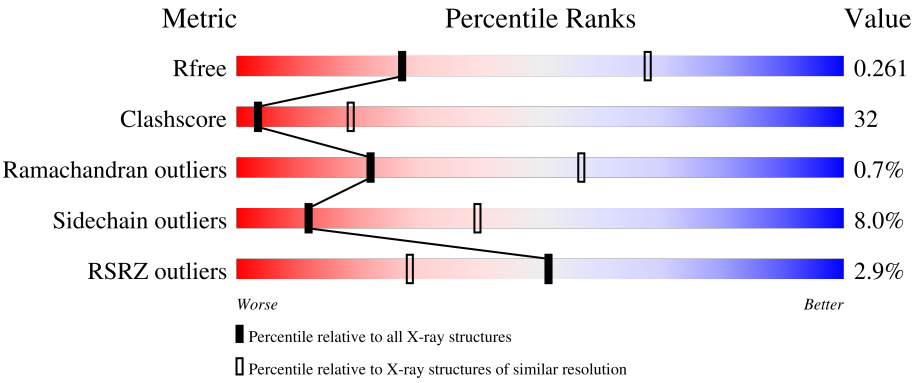
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 3.10 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))
R_{free}	130704	1094 (3.10-3.10)
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)

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	394	51% 34% 6% • 9%
1	B	394	4% 51% 34% 5% • 9%
1	C	394	2% 52% 33% 6% • 9%
1	D	394	53% 32% 6% • 9%
1	E	394	2% 52% 34% 6% • 9%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	394	2% 52% 33% 5% • 9%
1	G	394	% 52% 34% 6% • 9%
1	H	394	13% 51% 34% 5% • 9%
1	I	394	2% 52% 33% 6% • 9%
1	J	394	% 52% 33% 5% • 9%
1	K	394	2% 53% 33% 6% • 9%
1	L	394	% 52% 33% 6% • 9%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 33924 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 GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN.

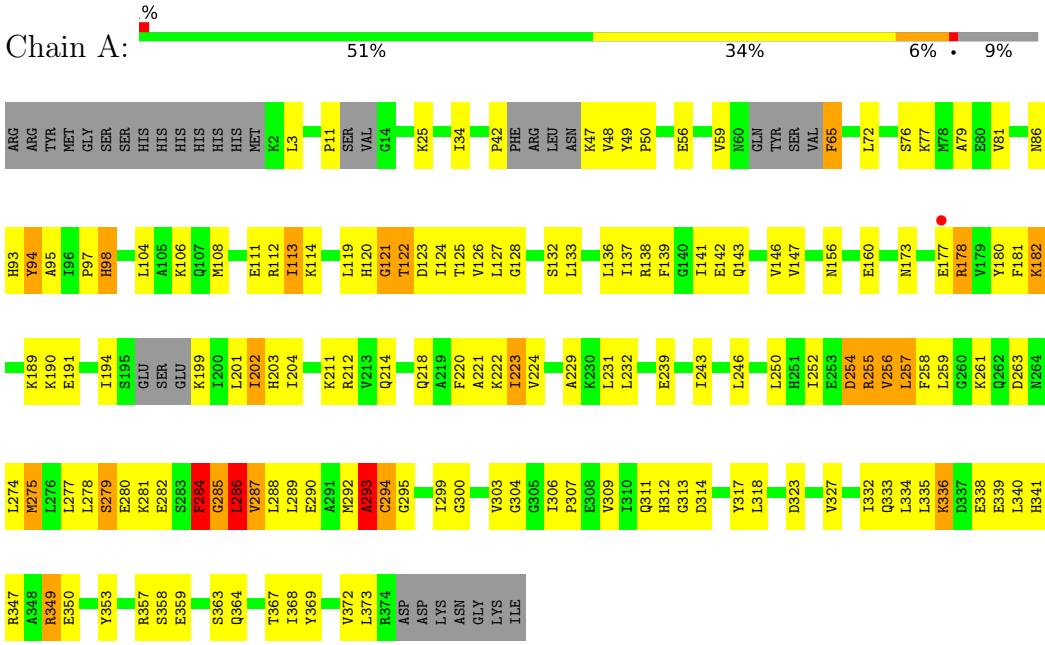
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	360	Total	C	N	O	S	0	0	1
			2827	1806	477	531	13			
1	B	360	Total	C	N	O	S	0	0	1
			2827	1806	477	531	13			
1	C	360	Total	C	N	O	S	0	0	1
			2827	1806	477	531	13			
1	D	360	Total	C	N	O	S	0	0	1
			2827	1806	477	531	13			
1	E	360	Total	C	N	O	S	0	0	1
			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 (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.

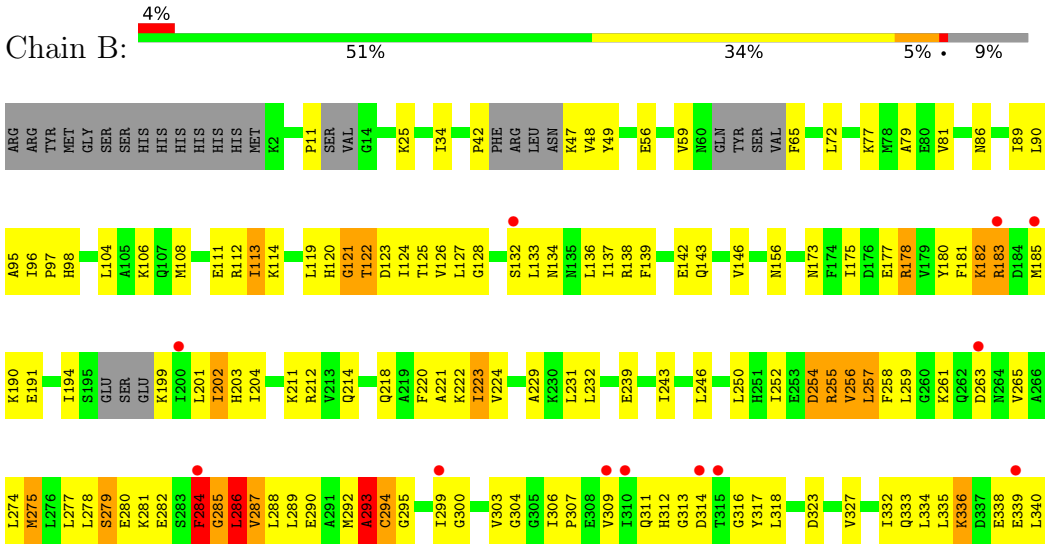
• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

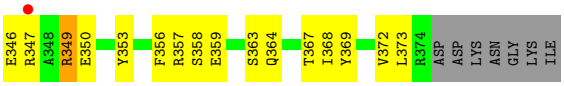
Chain A:



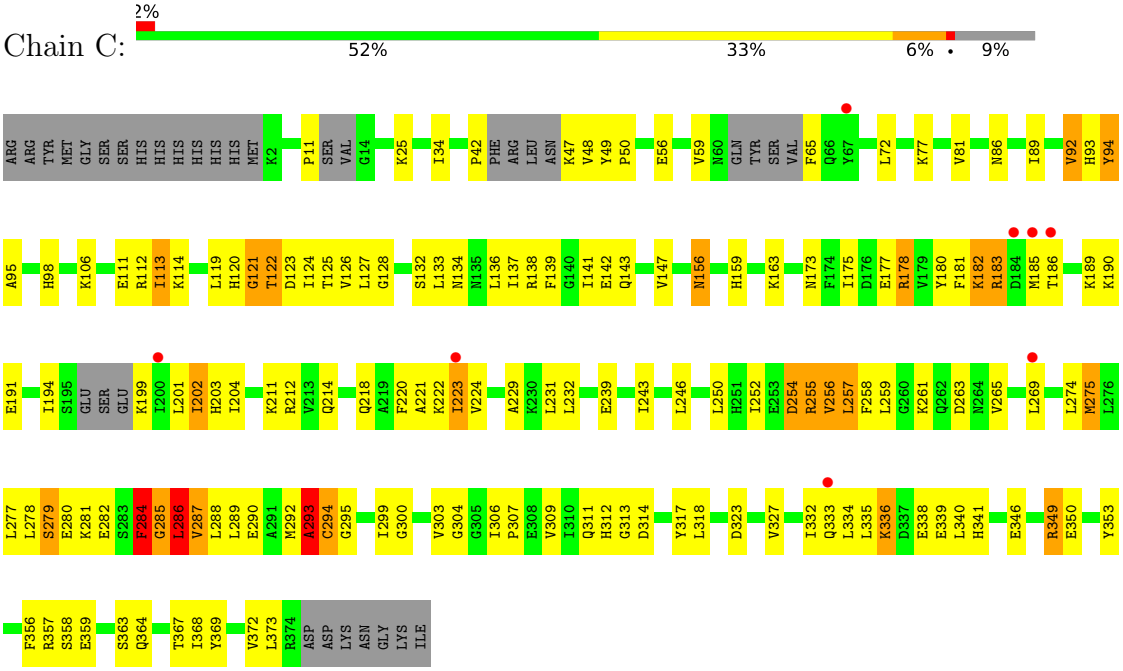
• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

Chain B:

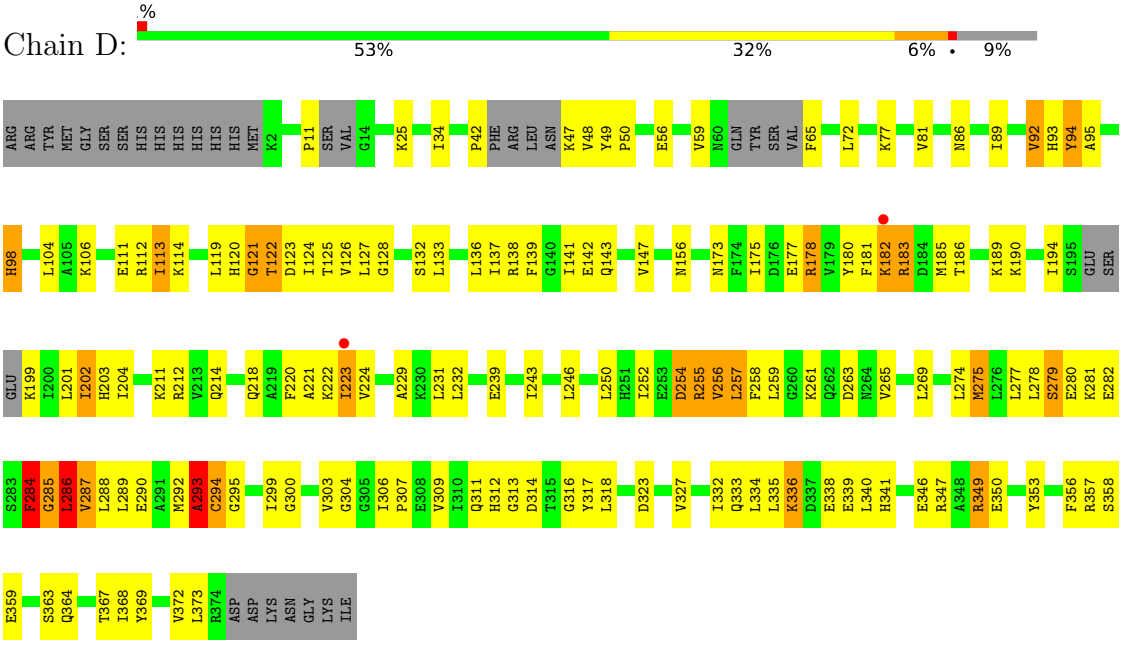




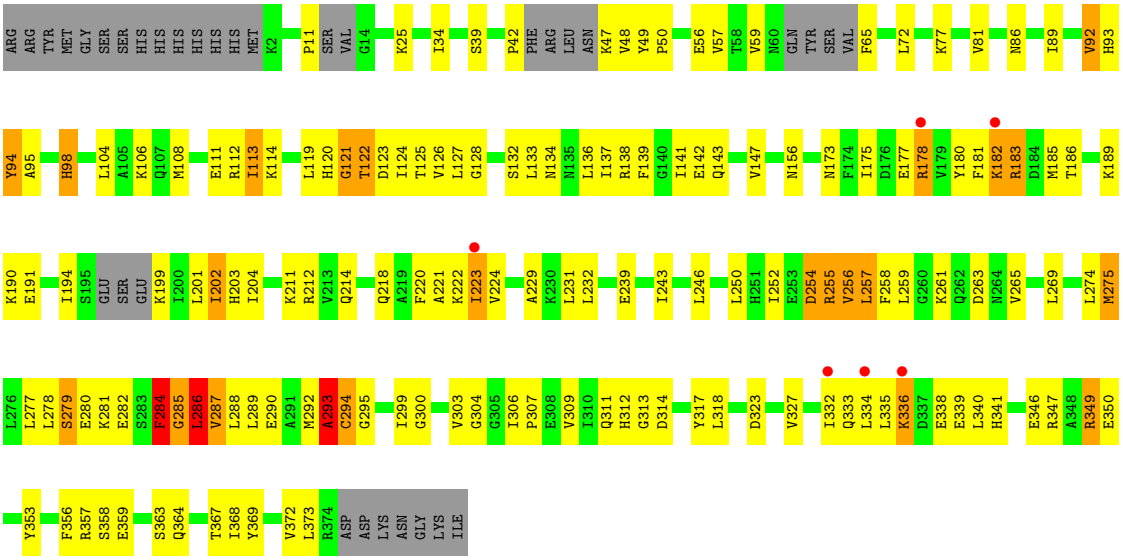
• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN



• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

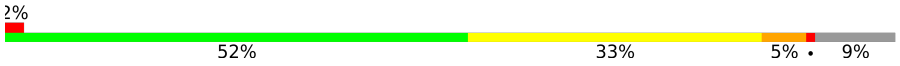


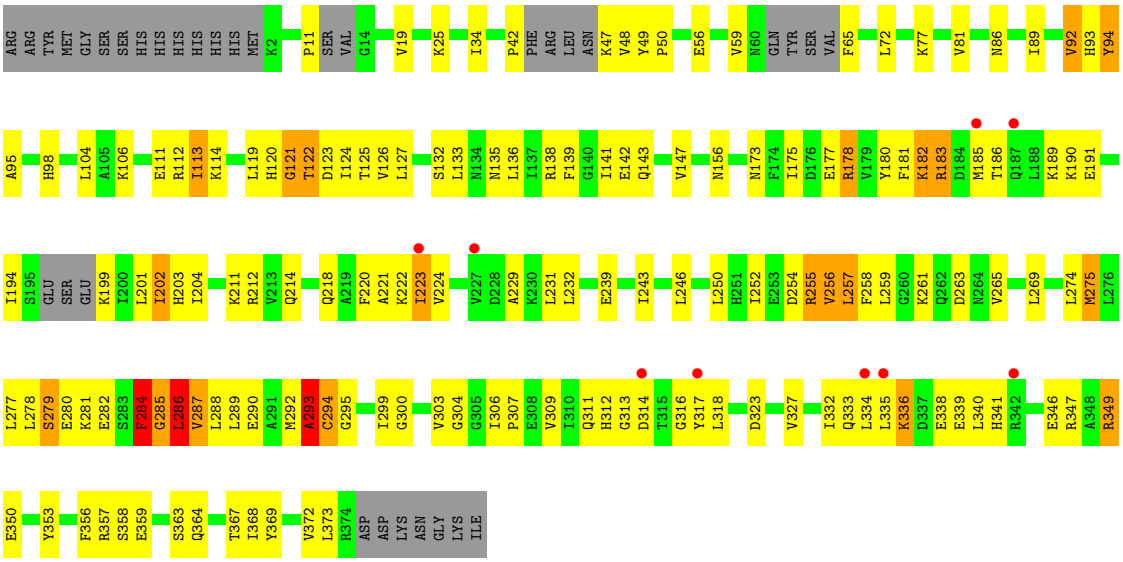
• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN



• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

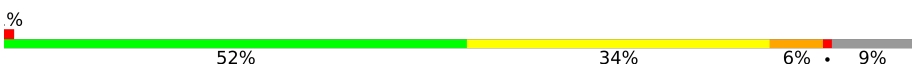
Chain F:

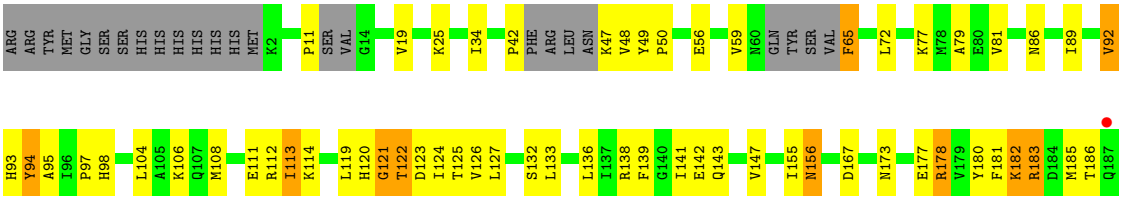


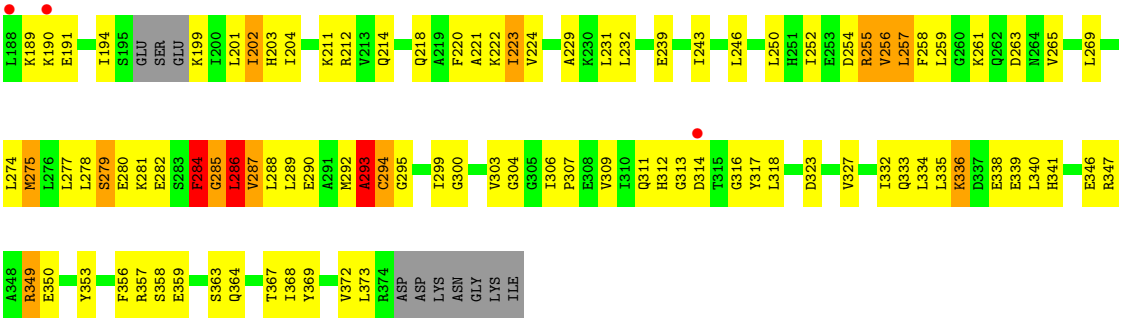


• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

Chain G:

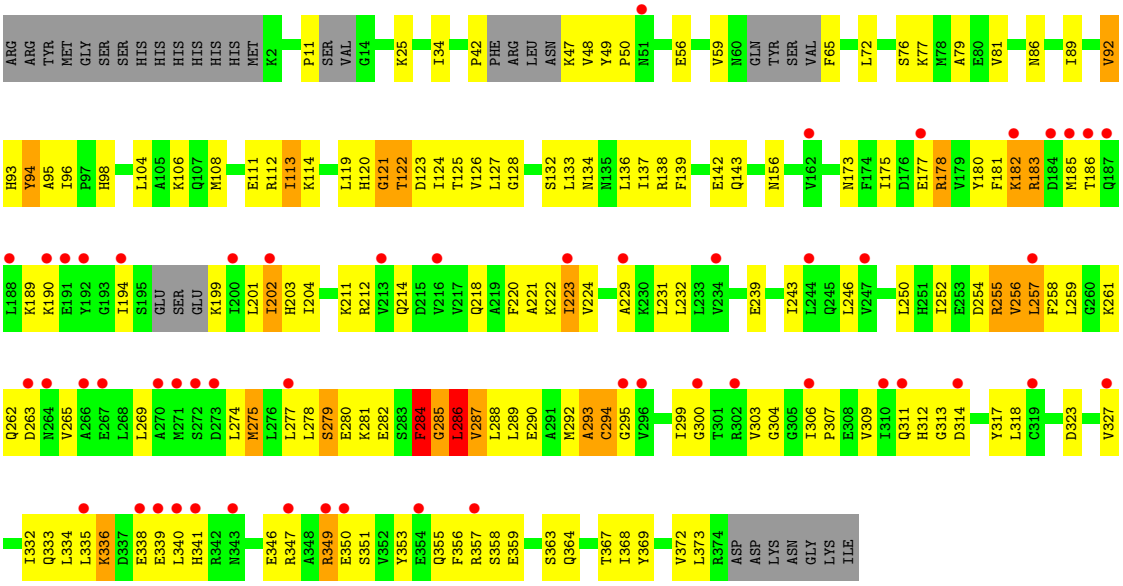






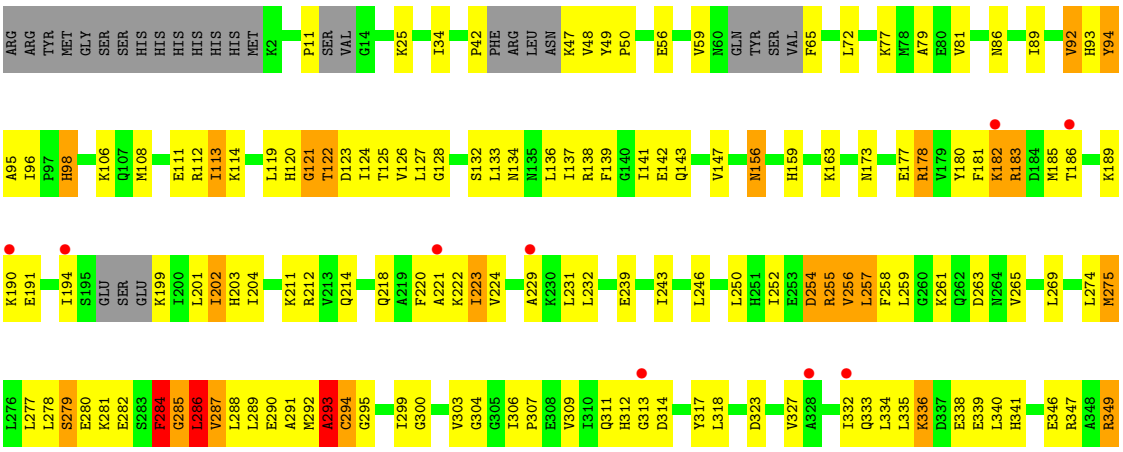
• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

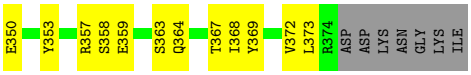
Chain H:



• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

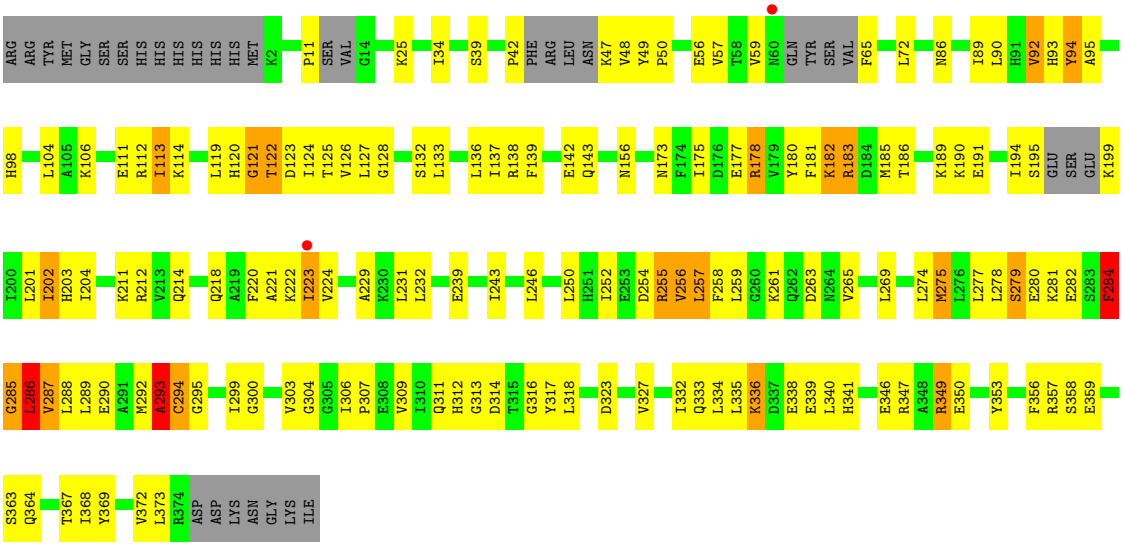
Chain I:





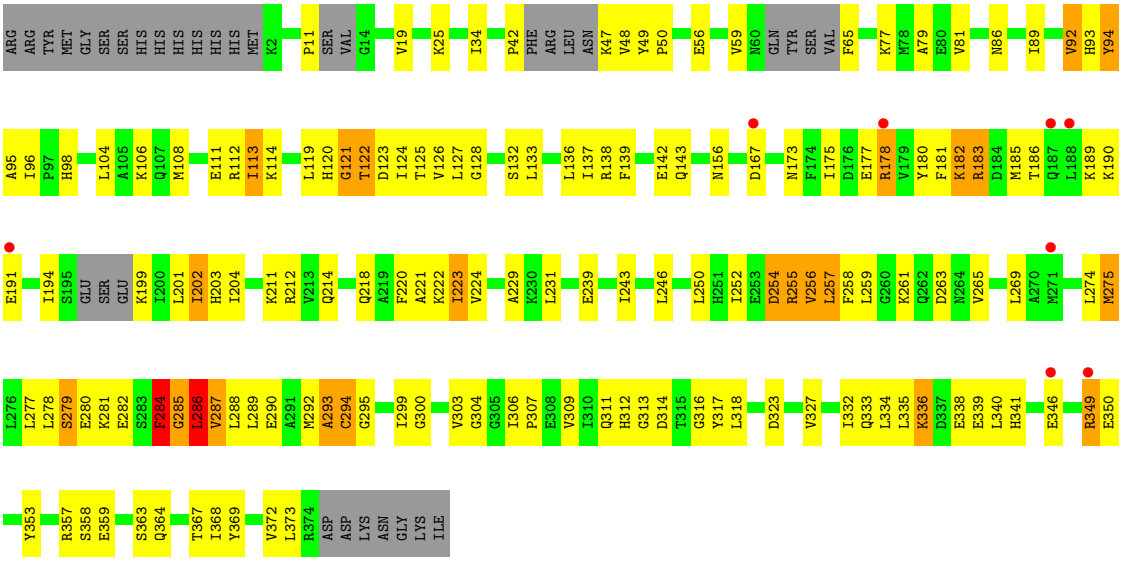
• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

Chain J:



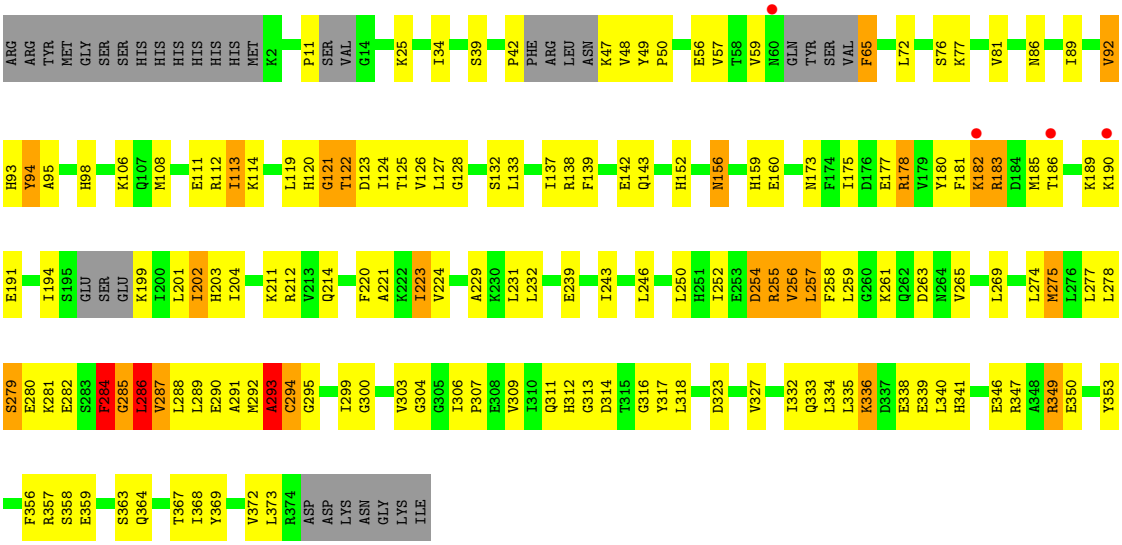
• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

Chain K:



• Molecule 1: GLYCOSYL TRANSFERASE, GROUP 1 FAMILY PROTEIN

Chain L:


ARG ARG TYR MET GLY SER SER HIS HIS HIS HIS HIS MET K2 P11 SER VAL G14 K25 I34 S39 P42 PHE ARG LEU ASN K47 V48 Y49 P50 E56 V57 T58 V59 N60 GLN TYR SER VAL F65 L72 S76 K77 V81 N86 I89 V92
H93 Y94 A95 H98 K106 Q107 M108 E111 R112 I113 K114 L119 H120 G121 T122 D123 I124 T125 V126 L127 G128 S132 L133 I137 R138 F139 E142 Q143 H152 N156 H159 E160 N173 F174 I175 D176 E177 R178 V179 Y180 F181 K182 R183 D184 M185 T186 K189 K190
E191 I194 S195 GLU SER GLU K199 I200 L201 I202 H203 I204 K211 R212 V213 Q214 F220 A221 K222 I223 V224 A229 K230 L231 L232 E239 I243 L246 L250 H251 I252 E253 D254 R255 V256 L257 F258 L259 G260 K261 Q262 D263 N264 V265 L269 L274 M275 L276 L277 L278
S279 E280 K281 E282 S283 F284 G285 L286 V287 L288 L289 E290 A291 M292 A293 C294 G295 I299 G300 V303 G304 G305 I306 P307 E308 V309 I310 Q311 H312 G313 D314 T315 G316 Y317 L318 D323 V327 I332 Q333 L334 L335 K336 D337 E338 E339 L340 H341 E346 R347 A348 R349 E350 Y353
F356 R357 S358 E359 S363 Q364 T367 I368 Y369 V372 L373 R374 ASP ASP LYS ASN GLY LYS ILE

4 Data and refinement statistics (i)

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
$< I/\sigma(I) >$ [1]	1.39 (at 2.69Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.254 , 0.262 0.249 , 0.261	Depositor DCC
R_{free} test set	5915 reflections (3.66%)	wwPDB-VP
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 , 63.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.022 for l,-k,h	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.*

[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	0.86	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(^{o})	Ideal(^{o})
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	Peptide,Mainchain
1	B	284	PHE	Peptide,Mainchain
1	C	284	PHE	Peptide,Mainchain
1	D	284	PHE	Peptide,Mainchain
1	E	284	PHE	Mainchain

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	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
1	F	2827	0	2856	185	0
1	G	2827	0	2856	195	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
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

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 32.

The worst 5 of 2180 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: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:H	1:B:286:LEU:CD1	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	Interatomic distance (Å)	Clash overlap (Å)
1:A:160:GLU:CA	1:K:167:ASP:OD2[1_455]	2.08	0.12

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	350/394 (89%)	323 (92%)	24 (7%)	3 (1%)	17 52
1	B	350/394 (89%)	322 (92%)	26 (7%)	2 (1%)	25 59

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	350/394 (89%)	323 (92%)	25 (7%)	2 (1%)	25	59
1	D	350/394 (89%)	322 (92%)	25 (7%)	3 (1%)	17	52
1	E	350/394 (89%)	323 (92%)	24 (7%)	3 (1%)	17	52
1	F	350/394 (89%)	322 (92%)	26 (7%)	2 (1%)	25	59
1	G	350/394 (89%)	322 (92%)	26 (7%)	2 (1%)	25	59
1	H	350/394 (89%)	322 (92%)	26 (7%)	2 (1%)	25	59
1	I	350/394 (89%)	322 (92%)	25 (7%)	3 (1%)	17	52
1	J	350/394 (89%)	322 (92%)	26 (7%)	2 (1%)	25	59
1	K	350/394 (89%)	322 (92%)	26 (7%)	2 (1%)	25	59
1	L	350/394 (89%)	323 (92%)	25 (7%)	2 (1%)	25	59
All	All	4200/4728 (89%)	3868 (92%)	304 (7%)	28 (1%)	22	57

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 (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	312/346 (90%)	287 (92%)	25 (8%)	12	40
1	B	312/346 (90%)	287 (92%)	25 (8%)	12	40
1	C	312/346 (90%)	287 (92%)	25 (8%)	12	40
1	D	312/346 (90%)	287 (92%)	25 (8%)	12	40
1	E	312/346 (90%)	287 (92%)	25 (8%)	12	40
1	F	312/346 (90%)	287 (92%)	25 (8%)	12	40

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	G	312/346 (90%)	287 (92%)	25 (8%)	12	40
1	H	312/346 (90%)	287 (92%)	25 (8%)	12	40
1	I	312/346 (90%)	287 (92%)	25 (8%)	12	40
1	J	312/346 (90%)	286 (92%)	26 (8%)	11	38
1	K	312/346 (90%)	287 (92%)	25 (8%)	12	40
1	L	312/346 (90%)	287 (92%)	25 (8%)	12	40
All	All	3744/4152 (90%)	3443 (92%)	301 (8%)	12	40

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

Mol	Chain	Res	Type
1	J	122	THR
1	L	182	LYS
1	J	195	SER
1	K	181	PHE
1	L	358	SER

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

Mol	Chain	Res	Type
1	G	134	ASN
1	L	364	GLN
1	H	330	GLN
1	L	333	GLN
1	K	333	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)

There are no ligands in this entry.

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	360/394 (91%)	0.02	3 (0%) 86 72	62, 92, 116, 125	0
1	B	360/394 (91%)	0.22	17 (4%) 31 15	63, 94, 116, 126	0
1	C	360/394 (91%)	0.18	8 (2%) 62 41	57, 93, 116, 125	0
1	D	360/394 (91%)	0.03	2 (0%) 89 78	58, 93, 115, 126	0
1	E	360/394 (91%)	0.12	6 (1%) 70 49	56, 92, 116, 125	0
1	F	360/394 (91%)	0.14	9 (2%) 57 34	57, 93, 116, 125	0
1	G	360/394 (91%)	0.05	4 (1%) 80 64	62, 93, 116, 125	0
1	H	360/394 (91%)	0.66	53 (14%) 2 1	63, 95, 117, 126	0
1	I	360/394 (91%)	0.22	9 (2%) 57 34	57, 92, 116, 125	0
1	J	360/394 (91%)	0.12	2 (0%) 89 78	38, 87, 116, 124	0
1	K	360/394 (91%)	0.16	8 (2%) 62 41	55, 90, 116, 124	0
1	L	360/394 (91%)	0.08	4 (1%) 80 64	61, 92, 115, 124	0
All	All	4320/4728 (91%)	0.17	125 (2%) 51 28	38, 92, 116, 126	0

The worst 5 of 125 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	177	GLU	5.2
1	H	190	LYS	4.4
1	H	257	LEU	4.3
1	H	263	ASP	4.1
1	H	272	SER	4.1

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 (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.