wwPDB X-ray Structure Validation Summary Report (i)

Aug 7, 2020 – 10:13 PM BST

PDB ID	:	5G47
Title	:	Structure of Gc glycoprotein from severe fever with thrombocytopenia syndrome virus in the trimeric postfusion conformation
Authors	:	Halldorsson, S.; Behrens, A.J.; Harlos, K.; Huiskonen, J.T.; Elliott, R.M.; Crispin, M.; Brennan, B.; Bowden, T.A.
Deposited on	:	2016-05-05
Resolution	:	2.45 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 following versions of software and data (see references (1)) were used in the production of this report:

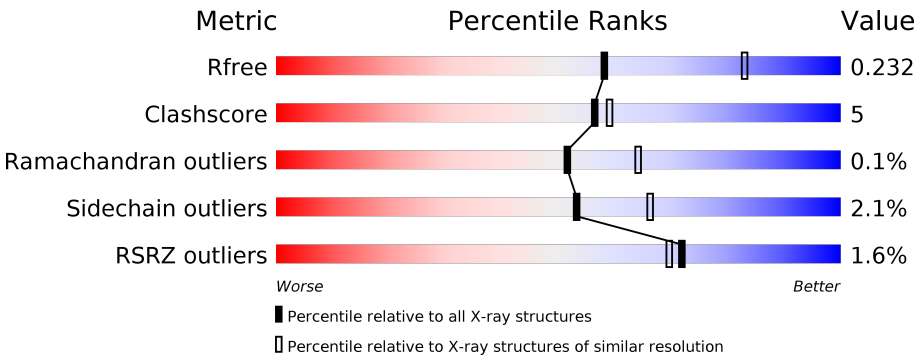
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

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 2.45 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	1544 (2.48-2.44)
Clashscore	141614	1613 (2.48-2.44)
Ramachandran outliers	138981	1598 (2.48-2.44)
Sidechain outliers	138945	1598 (2.48-2.44)
RSRZ outliers	127900	1523 (2.48-2.44)

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 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	445	2% 82% 14% •
1	B	445	2% 78% 18% • •
1	C	445	% 86% 10% •

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 9726 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 SFTSV GC.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	428	3194	1982	562	616	34	0	0	0
1	B	428	3194	1982	562	616	34	0	0	0
1	C	428	3194	1982	562	616	34	0	0	0

There are 33 discrepancies between the modelled and reference sequences:

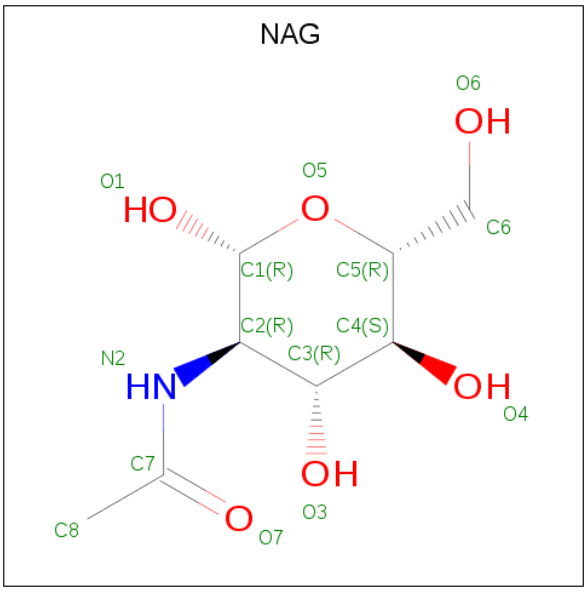
Chain	Residue	Modelled	Actual	Comment	Reference
A	561	THR	-	expression tag	UNP R4V2Q5
A	562	GLY	-	expression tag	UNP R4V2Q5
A	997	GLY	-	expression tag	UNP R4V2Q5
A	998	THR	-	expression tag	UNP R4V2Q5
A	999	LYS	-	expression tag	UNP R4V2Q5
A	1000	HIS	-	expression tag	UNP R4V2Q5
A	1001	HIS	-	expression tag	UNP R4V2Q5
A	1002	HIS	-	expression tag	UNP R4V2Q5
A	1003	HIS	-	expression tag	UNP R4V2Q5
A	1004	HIS	-	expression tag	UNP R4V2Q5
A	1005	HIS	-	expression tag	UNP R4V2Q5
B	561	THR	-	expression tag	UNP R4V2Q5
B	562	GLY	-	expression tag	UNP R4V2Q5
B	997	GLY	-	expression tag	UNP R4V2Q5
B	998	THR	-	expression tag	UNP R4V2Q5
B	999	LYS	-	expression tag	UNP R4V2Q5
B	1000	HIS	-	expression tag	UNP R4V2Q5
B	1001	HIS	-	expression tag	UNP R4V2Q5
B	1002	HIS	-	expression tag	UNP R4V2Q5
B	1003	HIS	-	expression tag	UNP R4V2Q5
B	1004	HIS	-	expression tag	UNP R4V2Q5
B	1005	HIS	-	expression tag	UNP R4V2Q5
C	561	THR	-	expression tag	UNP R4V2Q5

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	562	GLY	-	expression tag	UNP R4V2Q5
C	997	GLY	-	expression tag	UNP R4V2Q5
C	998	THR	-	expression tag	UNP R4V2Q5
C	999	LYS	-	expression tag	UNP R4V2Q5
C	1000	HIS	-	expression tag	UNP R4V2Q5
C	1001	HIS	-	expression tag	UNP R4V2Q5
C	1002	HIS	-	expression tag	UNP R4V2Q5
C	1003	HIS	-	expression tag	UNP R4V2Q5
C	1004	HIS	-	expression tag	UNP R4V2Q5
C	1005	HIS	-	expression tag	UNP R4V2Q5

- Molecule 2 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	14	8	1	5	0	0
			Total	C	N	O		
2	B	1	14	8	1	5	0	0
			Total	C	N	O		
2	C	1	14	8	1	5	0	0

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

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

- Molecule 4 is water.

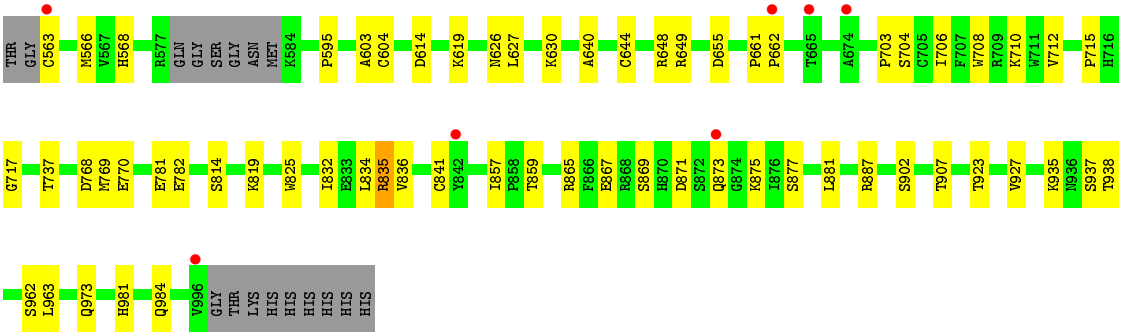
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	42	42	42	0	0
			Total	O		
4	B	32	32	32	0	0
			Total	O		
4	C	27	27	27	0	0

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.

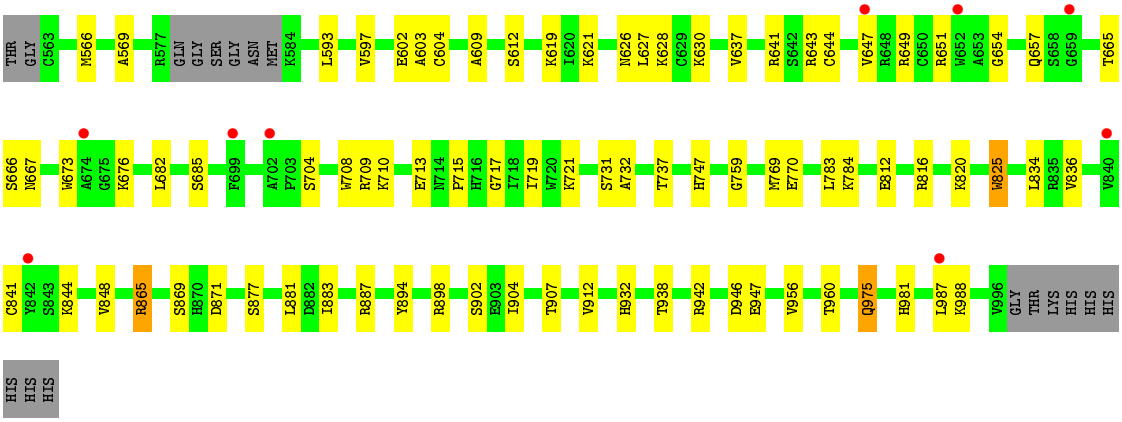
- Molecule 1: SFTSV GC

Chain A:



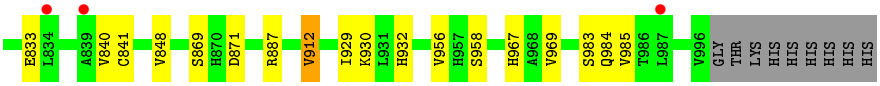
- Molecule 1: SFTSV GC

Chain B:



- Molecule 1: SFTSV GC

Chain C:


E833
L834
A839
V840
C841
V848
S869
H870
D871
R887
V912
I929
K930
L931
H932
V956
H957
S958
H967
A968
V969
S983
Q984
V985
T986
L987
V996
GLY
THR
LYS
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	147.36Å 152.27Å 160.38Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	108.51 – 2.45 108.51 – 2.45	Depositor EDS
% Data completeness (in resolution range)	99.6 (108.51-2.45) 99.6 (108.51-2.45)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.68 (at 2.45Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.194 , 0.232 0.194 , 0.232	Depositor DCC
R_{free} test set	3099 reflections (4.69%)	wwPDB-VP
Wilson B-factor (Å^2)	60.6	Xtriage
Anisotropy	0.361	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 45.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.013 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	9726	wwPDB-VP
Average B, all atoms (Å^2)	81.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.05% 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: NAG, CL

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	0.25	0/3255	0.46	0/4401
1	B	0.24	0/3255	0.45	0/4401
1	C	0.25	0/3255	0.46	0/4401
All	All	0.25	0/9765	0.46	0/13203

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	3194	0	3123	33	0
1	B	3194	0	3123	45	0
1	C	3194	0	3123	22	0
2	A	14	0	13	0	0
2	B	14	0	13	0	0
2	C	14	0	13	0	0
3	B	1	0	0	0	0
4	A	42	0	0	0	0
4	B	32	0	0	0	0
4	C	27	0	0	2	0
All	All	9726	0	9408	96	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 5.

The worst 5 of 96 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:619:LYS:HB2	1:B:737:THR:HB	1.68	0.74
1:B:942:ARG:HB2	1:B:975:GLN:HG3	1.73	0.71
1:C:869:SER:OG	1:C:871:ASP:OD1	2.10	0.69
1:B:628:LYS:NZ	1:B:871:ASP:OD1	2.27	0.68
1:B:628:LYS:NZ	1:B:877:SER:OG	2.25	0.66

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	424/445 (95%)	408 (96%)	16 (4%)	0	100	100
1	B	424/445 (95%)	411 (97%)	13 (3%)	0	100	100
1	C	424/445 (95%)	413 (97%)	10 (2%)	1 (0%)	47	57
All	All	1272/1335 (95%)	1232 (97%)	39 (3%)	1 (0%)	51	64

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	674	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	358/371 (96%)	352 (98%)	6 (2%)	60	73
1	B	358/371 (96%)	350 (98%)	8 (2%)	52	64
1	C	358/371 (96%)	349 (98%)	9 (2%)	47	60
All	All	1074/1113 (96%)	1051 (98%)	23 (2%)	53	66

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	898	ARG
1	B	960	THR
1	C	912	VAL
1	B	904	ILE
1	B	975	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 4 ligands modelled in this entry, 1 is monoatomic - leaving 3 for Mogul analysis.

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	NAG	C	1997	1	14,14,15	0.17	0	17,19,21	0.50	0
2	NAG	B	1997	1	14,14,15	0.19	0	17,19,21	0.56	0
2	NAG	A	1997	1	14,14,15	0.21	0	17,19,21	0.50	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	NAG	C	1997	1	-	4/6/23/26	0/1/1/1
2	NAG	B	1997	1	-	4/6/23/26	0/1/1/1
2	NAG	A	1997	1	-	4/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 12 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1997	NAG	O5-C5-C6-O6
2	A	1997	NAG	C4-C5-C6-O6
2	C	1997	NAG	C8-C7-N2-C2
2	C	1997	NAG	O7-C7-N2-C2
2	B	1997	NAG	C8-C7-N2-C2

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 (i)

6.1 Protein, DNA and RNA chains (i)

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	428/445 (96%)	0.16	7 (1%)	72	69	43, 72, 132, 164	0
1	B	428/445 (96%)	0.16	9 (2%)	63	60	41, 81, 138, 181	0
1	C	428/445 (96%)	0.14	5 (1%)	79	77	42, 74, 124, 153	0
All	All	1284/1335 (96%)	0.15	21 (1%)	72	69	41, 75, 132, 181	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	674	ALA	6.0
1	B	652	TRP	5.3
1	B	702	ALA	4.2
1	B	674	ALA	3.5
1	A	842	TYR	3.3

6.2 Non-standard residues in protein, DNA, RNA chains (i)

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)

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, 95th 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	NAG	C	1997	14/15	0.88	0.20	89,112,136,137	0
2	NAG	A	1997	14/15	0.93	0.13	79,100,109,122	0
2	NAG	B	1997	14/15	0.94	0.12	79,94,105,108	0
3	CL	B	1998	1/1	0.97	0.29	87,87,87,87	0

6.5 Other polymers (i)

There are no such residues in this entry.