



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 8, 2020 – 11:29 PM BST

PDB ID : 4GT0
Title : Structure of dengue virus serotype 1 sE containing stem to residue 421
Authors : Klein, D.E.; Choi, J.L.; Harrison, S.C.
Deposited on : 2012-08-28
Resolution : 2.57 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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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

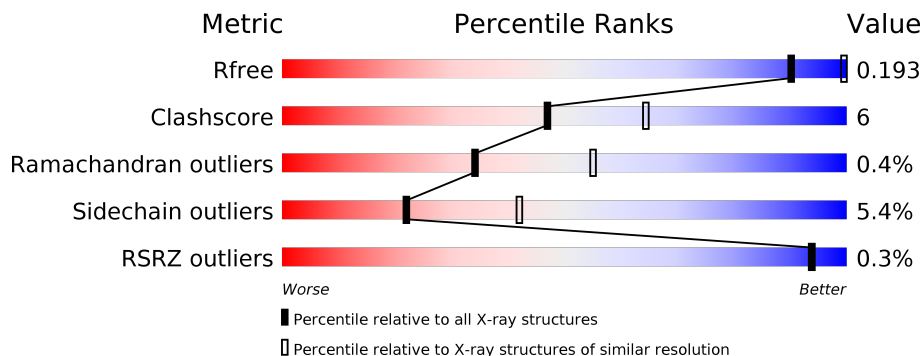
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.57 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	3676 (2.60-2.56)
Clashscore	141614	4049 (2.60-2.56)
Ramachandran outliers	138981	3979 (2.60-2.56)
Sidechain outliers	138945	3979 (2.60-2.56)
RSRZ outliers	127900	3614 (2.60-2.56)

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 $\leq 5\%$. 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	437	
1	B	437	

2 Entry composition i

There are 5 unique types of molecules in this entry. The entry contains 6226 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 Envelope protein E.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	390	3007	1902	507	578	20	0	4	0
1	B	390	2963	1872	498	573	20	0	1	0

There are 40 discrepancies between the modelled and reference sequences:

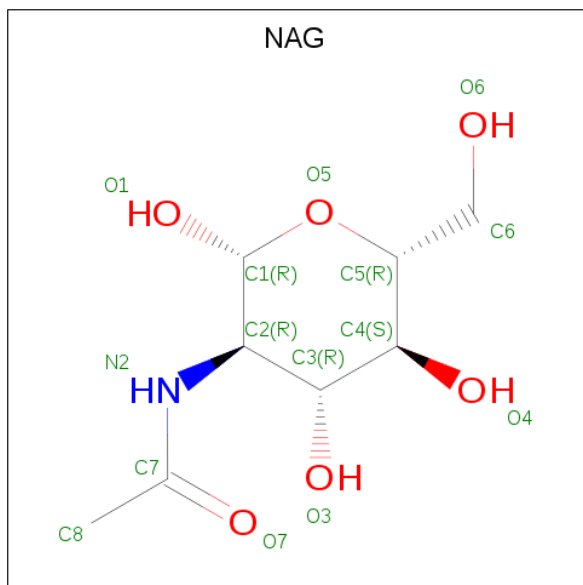
Chain	Residue	Modelled	Actual	Comment	Reference
A	-15	GLY	-	expression tag	UNP P17763
A	-14	HIS	-	expression tag	UNP P17763
A	-13	HIS	-	expression tag	UNP P17763
A	-12	HIS	-	expression tag	UNP P17763
A	-11	HIS	-	expression tag	UNP P17763
A	-10	HIS	-	expression tag	UNP P17763
A	-9	HIS	-	expression tag	UNP P17763
A	-8	HIS	-	expression tag	UNP P17763
A	-7	HIS	-	expression tag	UNP P17763
A	-6	GLY	-	expression tag	UNP P17763
A	-5	SER	-	expression tag	UNP P17763
A	-4	SER	-	expression tag	UNP P17763
A	-3	THR	-	expression tag	UNP P17763
A	-2	SER	-	expression tag	UNP P17763
A	-1	ASN	-	expression tag	UNP P17763
A	0	GLY	-	expression tag	UNP P17763
A	101	HIS	TRP	engineered mutation	UNP P17763
A	161	ILE	THR	conflict	UNP P17763
A	202	LYS	GLU	conflict	UNP P17763
A	203	GLU	LYS	conflict	UNP P17763
B	-15	GLY	-	expression tag	UNP P17763
B	-14	HIS	-	expression tag	UNP P17763
B	-13	HIS	-	expression tag	UNP P17763
B	-12	HIS	-	expression tag	UNP P17763
B	-11	HIS	-	expression tag	UNP P17763

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-10	HIS	-	expression tag	UNP P17763
B	-9	HIS	-	expression tag	UNP P17763
B	-8	HIS	-	expression tag	UNP P17763
B	-7	HIS	-	expression tag	UNP P17763
B	-6	GLY	-	expression tag	UNP P17763
B	-5	SER	-	expression tag	UNP P17763
B	-4	SER	-	expression tag	UNP P17763
B	-3	THR	-	expression tag	UNP P17763
B	-2	SER	-	expression tag	UNP P17763
B	-1	ASN	-	expression tag	UNP P17763
B	0	GLY	-	expression tag	UNP P17763
B	101	HIS	TRP	engineered mutation	UNP P17763
B	161	ILE	THR	conflict	UNP P17763
B	202	LYS	GLU	conflict	UNP P17763
B	203	GLU	LYS	conflict	UNP P17763

- 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
2	A	1	Total	C	N	O	0	0
			14	8	1	5		
2	B	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 3 is CADMIUM ION (three-letter code: CD) (formula: Cd).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	2	Total Cd 2 2	0	0
3	A	2	Total Cd 2 2	0	0

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total Cl 1 1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	118	Total O 118 118	0	0
5	B	105	Total O 105 105	0	0

4 Data and refinement statistics

Property	Value	Source
Space group	P 63	Depositor
Cell constants a, b, c, α , β , γ	77.89Å 77.89Å 292.29Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	28.48 – 2.57 49.57 – 2.57	Depositor EDS
% Data completeness (in resolution range)	99.6 (28.48-2.57) 99.6 (49.57-2.57)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.57 (at 2.58Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, R_{free}	0.177 , 0.210 0.163 , 0.193	Depositor DCC
R_{free} test set	1563 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å ²)	40.3	Xtriage
Anisotropy	0.015	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 7.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.40$, $\langle L^2 \rangle = 0.23$	Xtriage
Estimated twinning fraction	0.499 for h,-h-k,-l	Xtriage
Reported twinning fraction	0.500 for h,-h-k,-l	Depositor
Outliers	0 of 31764 reflections	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	6226	wwPDB-VP
Average B, all atoms (Å ²)	15.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.12% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CL, NAG, CD

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	0.22	0/3071	0.43	0/4156
1	B	0.22	0/3016	0.43	0/4087
All	All	0.22	0/6087	0.43	0/8243

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

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	3007	0	3050	30	0
1	B	2963	0	2965	41	2
2	A	14	0	13	5	0
2	B	14	0	13	3	0
3	A	2	0	0	0	0
3	B	2	0	0	0	0
4	A	1	0	0	1	0
5	A	118	0	0	6	0
5	B	105	0	0	11	1
All	All	6226	0	6041	72	2

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 6.

The worst 5 of 72 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:67:ASN:HD21	2:B:501:NAG:C1	1.25	1.50
1:A:67:ASN:HD21	2:A:501:NAG:C1	1.35	1.36
1:B:399:GLY:N	5:B:653:HOH:O	2.09	0.86
1:B:67:ASN:ND2	2:B:501:NAG:O5	2.10	0.84
1:B:76:THR:OG1	1:B:77:GLN:OE1	1.98	0.80

All (2) 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:B:225:SER:N	5:B:621:HOH:O 3_445	1.99	0.21
1:B:13:GLU:OE2	1:B:16:SER:OG 2_455	2.15	0.05

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	390/437 (89%)	375 (96%)	13 (3%)	2 (0%)	29 50
1	B	385/437 (88%)	366 (95%)	18 (5%)	1 (0%)	41 62
All	All	775/874 (89%)	741 (96%)	31 (4%)	3 (0%)	34 55

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	36	LYS
1	B	291	LYS
1	A	202	LYS

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	338/370 (91%)	318 (94%)	20 (6%)	19	37
1	B	328/370 (89%)	312 (95%)	16 (5%)	25	46
All	All	666/740 (90%)	630 (95%)	36 (5%)	22	42

5 of 36 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	342	GLU
1	B	79	GLU
1	B	298	SER
1	A	387	LEU
1	B	81	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	67	ASN
1	A	355	ASN
1	B	67	ASN
1	B	256	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 7 ligands modelled in this entry, 5 are monoatomic - leaving 2 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	B	501	1	14,14,15	0.57	0	17,19,21	0.81	0
2	NAG	A	501	1	14,14,15	0.53	0	17,19,21	0.87	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	B	501	1	-	3/6/23/26	0/1/1/1
2	NAG	A	501	1	-	3/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	501	NAG	C8-C7-N2-C2
2	B	501	NAG	O7-C7-N2-C2
2	A	501	NAG	C8-C7-N2-C2
2	B	501	NAG	C1-C2-N2-C7
2	A	501	NAG	O7-C7-N2-C2

There are no ring outliers.

2 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	501	NAG	3	0
2	A	501	NAG	5	0

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	390/437 (89%)	-0.48	1 (0%) 94 94	4, 12, 29, 48	0
1	B	390/437 (89%)	-0.41	1 (0%) 94 94	2, 15, 32, 45	0
All	All	780/874 (89%)	-0.44	2 (0%) 94 94	2, 13, 31, 48	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	146	GLY	3.2
1	B	102	GLY	2.5

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	A	501	14/15	0.86	0.11	21,38,46,53	0
2	NAG	B	501	14/15	0.89	0.11	46,53,58,58	0
3	CD	A	502	1/1	0.98	0.06	53,53,53,53	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
3	CD	B	502	1/1	0.98	0.05	65,65,65,65	0
4	CL	A	504	1/1	0.98	0.09	9,9,9,9	0
3	CD	B	503	1/1	0.98	0.04	62,62,62,62	0
3	CD	A	503	1/1	0.99	0.04	52,52,52,52	0

6.5 Other polymers [i](#)

There are no such residues in this entry.