



# wwPDB X-ray Structure Validation Summary Report i

Dec 1, 2025 – 09:05 am GMT

PDB ID : 9T3J / pdb\_00009t3j  
Title : Protease from Norovirus Sydney GII.4 strain with crystallization epitope mutation H50Y  
Authors : Fairhead, M.; MacLean, E.M.; Ni, X.; Wright, N.D.; Koekemoer, L.; von Delft, F.  
Deposited on : 2025-10-27  
Resolution : 2.41 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](mailto: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 : 4-5-2 with Phenix2.0  
Xtriage (Phenix) : 2.0  
EDS : 3.0  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
CCP4 : 9.0.010 (Gargrove)  
Density-Fitness : 1.0.12  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.46

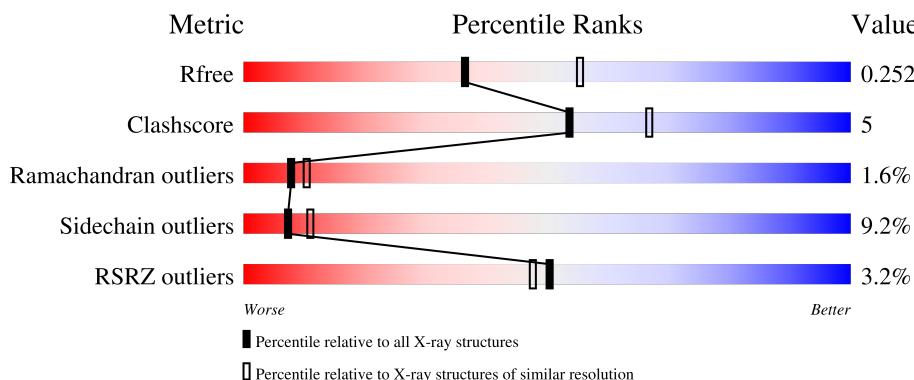
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## X-RAY DIFFRACTION

The reported resolution of this entry is 2.41 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 <sub>free</sub>	164625	5670 (2.44-2.40)
Clashscore	180529	6299 (2.44-2.40)
Ramachandran outliers	177936	6232 (2.44-2.40)
Sidechain outliers	177891	6233 (2.44-2.40)
RSRZ outliers	164620	5670 (2.44-2.40)

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.



## 2 Entry composition [\(i\)](#)

There are 2 unique types of molecules in this entry. The entry contains 10253 atoms, of which 5079 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 Genome polyprotein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	172	Total	C	H	N	O	S	0	4	0
			2608	834	1296	233	234	11			
1	B	172	Total	C	H	N	O	S	0	0	0
			2566	822	1275	226	232	11			
1	C	169	Total	C	H	N	O	S	0	0	0
			2526	809	1254	223	229	11			
1	D	169	Total	C	H	N	O	S	0	0	0
			2526	809	1254	223	229	11			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	50	TYR	HIS	engineered mutation	UNP A0A3Q8EAF6
A	69	SER	THR	conflict	UNP A0A3Q8EAF6
B	50	TYR	HIS	engineered mutation	UNP A0A3Q8EAF6
B	69	SER	THR	conflict	UNP A0A3Q8EAF6
C	50	TYR	HIS	engineered mutation	UNP A0A3Q8EAF6
C	69	SER	THR	conflict	UNP A0A3Q8EAF6
D	50	TYR	HIS	engineered mutation	UNP A0A3Q8EAF6
D	69	SER	THR	conflict	UNP A0A3Q8EAF6

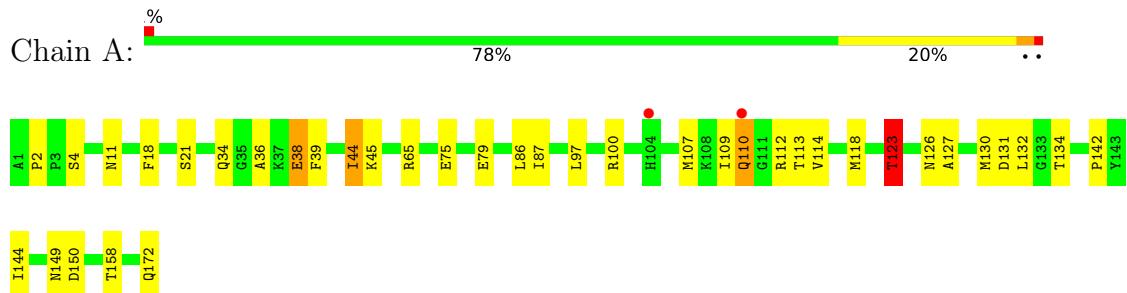
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	13	Total O 13 13	0	0
2	B	10	Total O 10 10	0	0
2	C	3	Total O 3 3	0	0
2	D	1	Total O 1 1	0	0

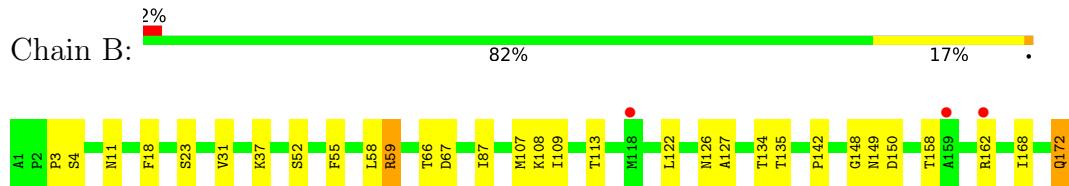
### 3 Residue-property plots [\(i\)](#)

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.

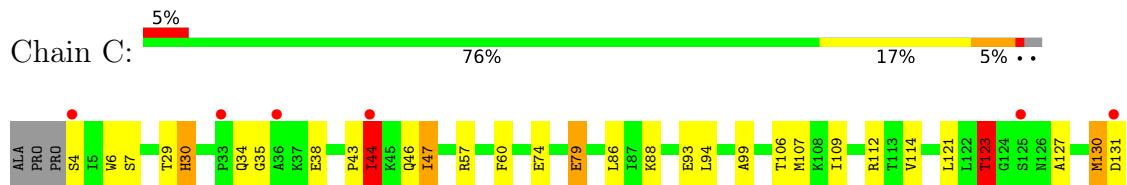
- Molecule 1: Genome polyprotein



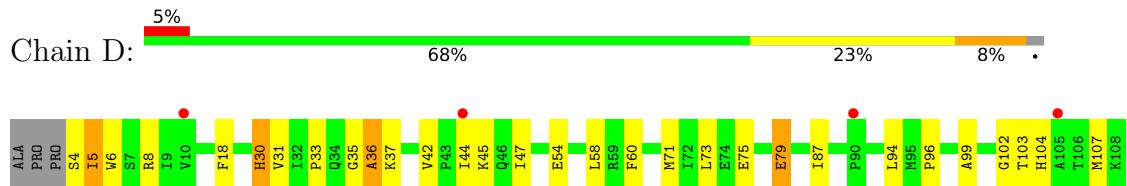
- Molecule 1: Genome polyprotein

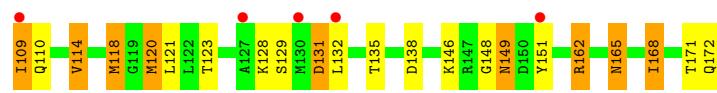


- Molecule 1: Genome polyprotein



- Molecule 1: Genome polyprotein





## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	128.90Å 128.90Å 118.13Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	81.13 – 2.41 81.13 – 2.41	Depositor EDS
% Data completeness (in resolution range)	92.0 (81.13-2.41) 92.0 (81.13-2.41)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.98 (at 2.42Å)	Xtriage
Refinement program	REFMAC 5.8.0431 (refmacat 0.4.105)	Depositor
$R$ , $R_{free}$	0.209 , 0.252 0.209 , 0.252	Depositor DCC
$R_{free}$ test set	1740 reflections (4.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	72.0	Xtriage
Anisotropy	0.303	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.41 , 62.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.039 for h,-h-k,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	10253	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	90.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.75% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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 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	0.89	0/1357	1.63	19/1836 (1.0%)
1	B	0.88	1/1322 (0.1%)	1.46	8/1790 (0.4%)
1	C	0.81	0/1301	1.58	14/1759 (0.8%)
1	D	0.83	0/1301	1.58	16/1759 (0.9%)
All	All	0.85	1/5281 (0.0%)	1.56	57/7144 (0.8%)

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	3
1	B	0	2
1	C	0	4
All	All	0	9

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	172	GLN	C-O	5.42	1.34	1.23

The worst 5 of 57 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	110	GLN	CB-CA-C	13.86	132.04	111.80
1	A	38	GLU	CB-CG-CD	11.71	132.51	112.60
1	D	79	GLU	CB-CG-CD	10.71	130.80	112.60
1	C	79	GLU	CB-CG-CD	9.84	129.32	112.60
1	D	30	HIS	CA-CB-CG	9.46	123.27	113.80

There are no chirality outliers.

5 of 9 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	100[A]	ARG	Sidechain
1	A	112	ARG	Sidechain
1	A	65	ARG	Sidechain
1	B	162	ARG	Sidechain
1	B	59	ARG	Sidechain

## 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	1312	1296	1317	10	0
1	B	1291	1275	1311	9	0
1	C	1272	1254	1289	13	0
1	D	1272	1254	1289	19	0
2	A	13	0	0	0	0
2	B	10	0	0	0	0
2	C	3	0	0	0	0
2	D	1	0	0	0	0
All	All	5174	5079	5206	50	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 50 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:107:MET:HE3	1:B:109:ILE:HD11	1.73	0.70
1:A:109:ILE:HG22	1:A:110:GLN:HG2	1.85	0.57
1:A:123:THR:HG21	1:A:132:LEU:O	2.04	0.57
1:A:86:LEU:HD11	1:A:144:ILE:HG12	1.89	0.55
1:C:123:THR:HG23	1:C:127:ALA:CB	2.38	0.54

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	174/172 (101%)	167 (96%)	7 (4%)	0	100 100
1	B	170/172 (99%)	158 (93%)	10 (6%)	2 (1%)	11 15
1	C	167/172 (97%)	156 (93%)	9 (5%)	2 (1%)	11 15
1	D	167/172 (97%)	151 (90%)	9 (5%)	7 (4%)	2 1
All	All	678/688 (98%)	632 (93%)	35 (5%)	11 (2%)	8 10

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	5	ILE
1	D	36	ALA
1	D	129	SER
1	C	35	GLY
1	D	148	GLY

### 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	140/137 (102%)	129 (92%)	11 (8%)	10 16
1	B	137/137 (100%)	128 (93%)	9 (7%)	14 22
1	C	135/137 (98%)	120 (89%)	15 (11%)	5 6
1	D	135/137 (98%)	120 (89%)	15 (11%)	5 6
All	All	547/548 (100%)	497 (91%)	50 (9%)	7 11

5 of 50 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	109	ILE
1	D	5	ILE
1	D	168	ILE
1	C	112	ARG
1	C	130	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	46	GLN
1	D	104	HIS
1	D	165	ASN
1	B	126	ASN
1	B	172	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 oligosaccharides 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 [\(i\)](#)

### 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	172/172 (100%)	-0.36	2 (1%) 76 74	29, 71, 106, 142	2 (1%)
1	B	172/172 (100%)	-0.19	3 (1%) 69 66	55, 73, 100, 142	0
1	C	169/172 (98%)	0.49	8 (4%) 37 34	64, 99, 153, 182	0
1	D	169/172 (98%)	0.40	9 (5%) 33 30	61, 97, 156, 193	0
All	All	682/688 (99%)	0.08	22 (3%) 50 47	29, 82, 142, 193	2 (0%)

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	104[A]	HIS	3.2
1	B	159	ALA	3.1
1	D	151	TYR	2.9
1	C	4	SER	2.7
1	D	130	MET	2.6

### 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 oligosaccharides 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.