



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 15, 2025 – 06:02 PM EST

PDB ID : 9CSH  
Title : Turnip Yellow Mosaic Virus (TYMV) protease (PRO) bound to a ubiquitin variant (UbV3)  
Authors : Kim, K.; Mark, B.L.  
Deposited on : 2024-07-23  
Resolution : 2.80 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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

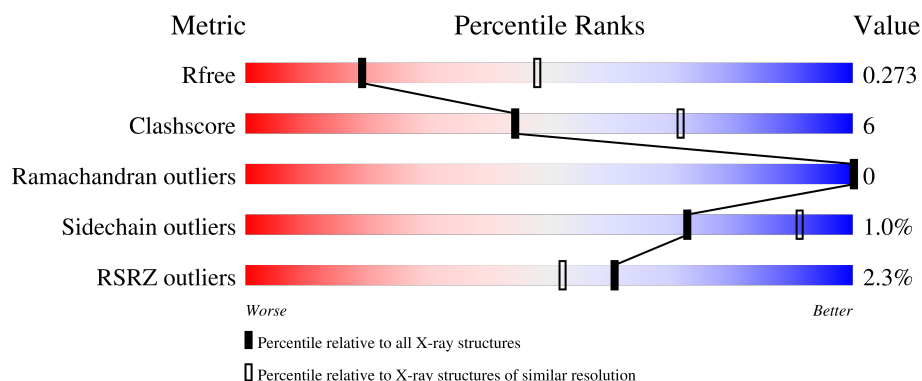
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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))
$R_{free}$	164625	3657 (2.80-2.80)
Clashscore	180529	4123 (2.80-2.80)
Ramachandran outliers	177936	4071 (2.80-2.80)
Sidechain outliers	177891	4073 (2.80-2.80)
RSRZ outliers	164620	3659 (2.80-2.80)

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  $\geq 3$ , 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	159	<div> <div>3%</div> <div> <div></div> <div>76%</div> <div>15%</div> <div>9%</div> </div> </div>
1	C	159	<div> <div>2%</div> <div> <div></div> <div>74%</div> <div>15%</div> <div>11%</div> </div> </div>
2	B	103	<div> <div>2%</div> <div> <div></div> <div>62%</div> <div>11%</div> <div>•</div> <div>26%</div> </div> </div>
2	D	103	<div> <div></div> <div> <div></div> <div>60%</div> <div>13%</div> <div>•</div> <div>26%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3456 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 Methyltransferase/Protease/Ubiquitinyl hydrolase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	145	Total	C	N	O	S	0	0	0
			1133	717	194	221	1			
1	C	141	Total	C	N	O	S	0	0	0
			1101	695	189	216	1			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	721	HIS	-	expression tag	UNP P10358
A	722	HIS	-	expression tag	UNP P10358
A	723	HIS	-	expression tag	UNP P10358
A	724	HIS	-	expression tag	UNP P10358
A	725	HIS	-	expression tag	UNP P10358
A	726	HIS	-	expression tag	UNP P10358
A	727	GLY	-	expression tag	UNP P10358
C	721	HIS	-	expression tag	UNP P10358
C	722	HIS	-	expression tag	UNP P10358
C	723	HIS	-	expression tag	UNP P10358
C	724	HIS	-	expression tag	UNP P10358
C	725	HIS	-	expression tag	UNP P10358
C	726	HIS	-	expression tag	UNP P10358
C	727	GLY	-	expression tag	UNP P10358

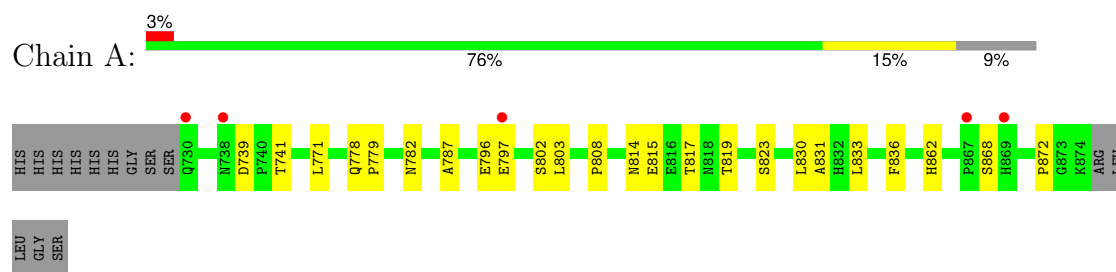
- Molecule 2 is a protein called Ubiquitin Variant UbV3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	76	Total	C	N	O	S	0	0	0
			611	388	107	115	1			
2	D	76	Total	C	N	O	S	0	0	0
			611	388	107	115	1			

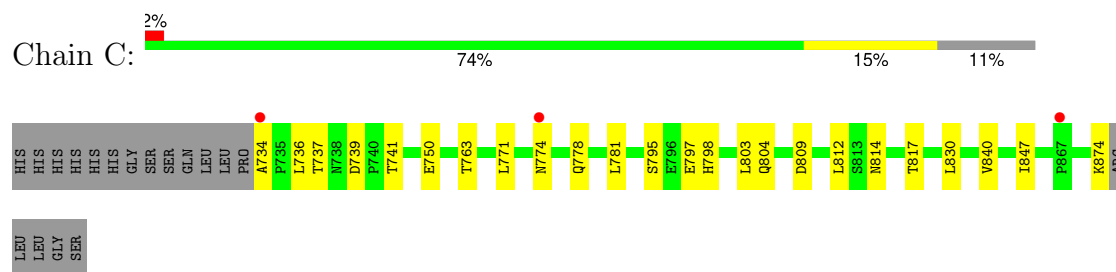
### 3 Residue-property plots [i](#)

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.

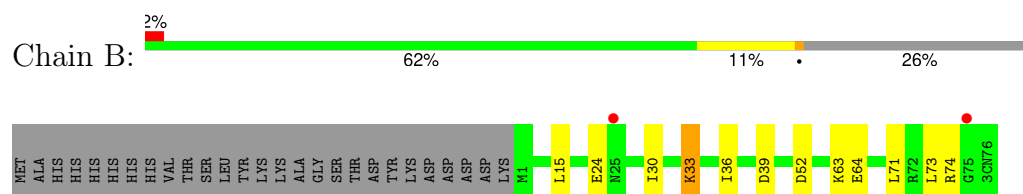
- Molecule 1: Methyltransferase/Protease/Ubiquitinyl hydrolase



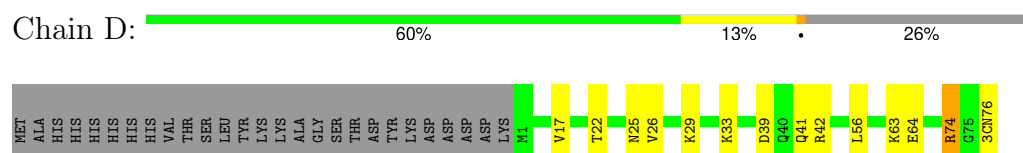
- Molecule 1: Methyltransferase/Protease/Ubiquitinyl hydrolase



- Molecule 2: Ubiquitin Variant UbV3



- Molecule 2: Ubiquitin Variant UbV3



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	72.01Å 89.66Å 137.46Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	35.50 – 2.80 35.50 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.5 (35.50-2.80) 99.4 (35.50-2.80)	Depositor EDS
$R_{merge}$	0.15	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.14 (at 2.81Å)	Xtriage
Refinement program	PHENIX 1.21_5207	Depositor
R, $R_{free}$	0.229 , 0.278 0.225 , 0.273	Depositor DCC
$R_{free}$ test set	10106 reflections (9.99%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	36.2	Xtriage
Anisotropy	0.170	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 18.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	3456	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.37% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 3CN

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	0.24	0/1165	0.47	0/1600
1	C	0.24	0/1132	0.47	0/1554
2	B	0.29	0/615	0.52	0/829
2	D	0.29	0/615	0.52	0/829
All	All	0.26	0/3527	0.49	0/4812

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	1133	0	1112	16	0
1	C	1101	0	1075	12	0
2	B	611	0	637	9	0
2	D	611	0	637	7	0
All	All	3456	0	3461	41	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 6.

The worst 5 of 41 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:C:781:LEU:HB3	2:D:76:3CN:HC2	1.64	0.80
1:A:787:ALA:HA	1:A:872:PRO:HG3	1.75	0.67
2:D:22:THR:HG23	2:D:25:ASN:H	1.64	0.62
2:B:24:GLU:HG2	2:B:52:ASP:HB3	1.83	0.60
1:A:739:ASP:OD2	1:A:808:PRO:HB2	2.02	0.59

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	143/159 (90%)	139 (97%)	4 (3%)	0	100	100
1	C	139/159 (87%)	138 (99%)	1 (1%)	0	100	100
2	B	73/103 (71%)	73 (100%)	0	0	100	100
2	D	73/103 (71%)	72 (99%)	1 (1%)	0	100	100
All	All	428/524 (82%)	422 (99%)	6 (1%)	0	100	100

There are no Ramachandran outliers to report.

### 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	132/144 (92%)	132 (100%)	0	100	100
1	C	128/144 (89%)	127 (99%)	1 (1%)	79	93
2	B	69/93 (74%)	68 (99%)	1 (1%)	62	87
2	D	69/93 (74%)	67 (97%)	2 (3%)	37	71
All	All	398/474 (84%)	394 (99%)	4 (1%)	73	91

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	33	LYS
1	C	874	LYS
2	D	33	LYS
2	D	74	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	778	GLN
1	C	862	HIS

### 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	145/159 (91%)	-0.03	5 (3%) 48 40	19, 30, 54, 71	0
1	C	141/159 (88%)	-0.01	3 (2%) 63 55	18, 31, 52, 75	0
2	B	75/103 (72%)	0.07	2 (2%) 56 47	24, 32, 46, 55	0
2	D	75/103 (72%)	0.23	0 100 100	25, 43, 63, 77	0
All	All	436/524 (83%)	0.04	10 (2%) 61 52	18, 32, 58, 77	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	25	ASN	2.9
1	C	734	ALA	2.8
2	B	75	GLY	2.8
1	A	738	ASN	2.7
1	C	867	PRO	2.4

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.