



# Full wwPDB X-ray Structure Validation Report ⓘ

May 26, 2020 – 09:55 pm BST

PDB ID : 5KYD  
Title : Crystal structure of USP7 catalytic domain [V302K] mutant in complex with ubiquitin  
Authors : Rouge, L.; Ozen, A.  
Deposited on : 2016-07-21  
Resolution : 1.62 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
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)	:	2.11

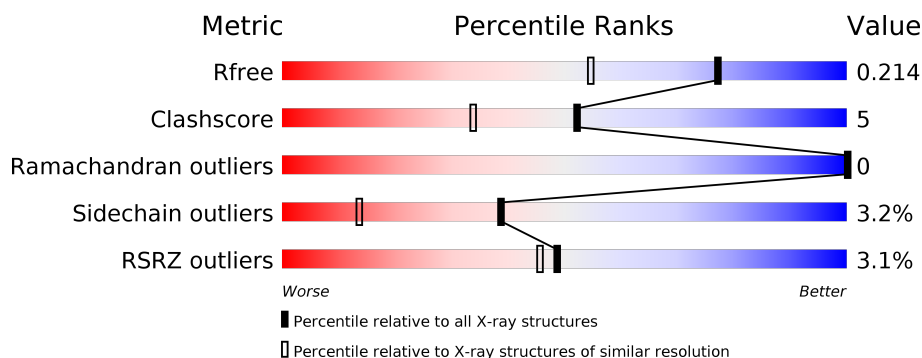
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.62 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	4693 (1.64-1.60)
Clashscore	141614	5002 (1.64-1.60)
Ramachandran outliers	138981	4888 (1.64-1.60)
Sidechain outliers	138945	4887 (1.64-1.60)
RSRZ outliers	127900	4609 (1.64-1.60)

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  $\geq 3$ , 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	D	76	<div> <div>%</div> <div> <div></div> <div>82%</div> <div>17%</div> <div>.</div> </div> </div>
2	A	351	<div> <div>3%</div> <div> <div></div> <div>84%</div> <div>12%</div> <div>..</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3792 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 Polyubiquitin-B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	D	76	Total	C	N	O	S	0	0	0
			600	378	105	116	1			

- Molecule 2 is a protein called Ubiquitin carboxyl-terminal hydrolase 7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	341	Total	C	N	O	S	0	6	0
			2811	1773	480	540	18			

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	204	GLY	-	expression tag	UNP Q93009
A	205	SER	-	expression tag	UNP Q93009
A	206	HIS	-	expression tag	UNP Q93009
A	207	MET	-	expression tag	UNP Q93009
A	302	LYS	VAL	engineered mutation	UNP Q93009

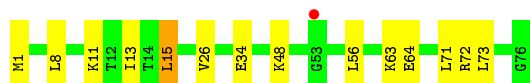
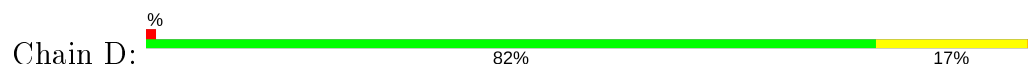
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	D	58	Total	O	0	0
			58	58		
3	A	323	Total	O	0	0
			323	323		

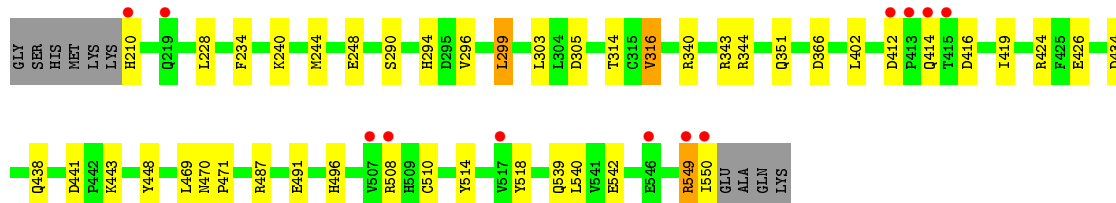
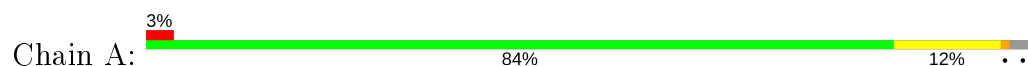
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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: Polyubiquitin-B



#### • Molecule 2: Ubiquitin carboxyl-terminal hydrolase 7



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	62.05Å 80.47Å 86.56Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	27.16 – 1.62 27.16 – 1.62	Depositor EDS
% Data completeness (in resolution range)	99.3 (27.16-1.62) 95.4 (27.16-1.62)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.03 (at 1.62Å)	Xtriage
Refinement program	PHENIX 1.10.1 _2155	Depositor
R, $R_{free}$	0.174 , 0.213 0.176 , 0.214	Depositor DCC
$R_{free}$ test set	2013 reflections (3.61%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	20.8	Xtriage
Anisotropy	0.234	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 43.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	3792	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.87% 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	D	0.34	0/606	0.58	1/815 (0.1%)
2	A	0.37	0/2877	0.58	0/3881
All	All	0.37	0/3483	0.58	1/4696 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	15	LEU	CA-CB-CG	5.49	127.93	115.30

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	D	600	0	629	9	0
2	A	2811	0	2736	29	1
3	A	323	0	0	10	0
3	D	58	0	0	2	0
All	All	3792	0	3365	37	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 5.

All (37) 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)
2:A:424:ARG:NH1	2:A:426:GLU:OE2	2.10	0.84
1:D:11:LYS:NZ	3:D:101:HOH:O	2.16	0.76
2:A:487:ARG:NH1	3:A:603:HOH:O	2.17	0.76
2:A:366:ASP:OD2	3:A:601:HOH:O	2.08	0.72
2:A:228:LEU:HD23	2:A:299:LEU:HD13	1.73	0.69
2:A:414:GLN:HA	2:A:414:GLN:OE1	1.95	0.66
2:A:351:GLN:NE2	3:A:612:HOH:O	2.30	0.65
2:A:340:ARG:NH1	3:A:604:HOH:O	2.21	0.62
2:A:438:GLN:NE2	3:A:613:HOH:O	2.31	0.61
1:D:48:LYS:NZ	2:A:305[B]:ASP:OD1	2.31	0.61
2:A:510[B]:CYS:SG	3:A:628:HOH:O	2.54	0.61
2:A:542:GLU:OE2	3:A:602:HOH:O	2.16	0.60
2:A:290:SER:O	2:A:294:HIS:HD2	1.85	0.60
1:D:26:VAL:HG21	1:D:56:LEU:HD21	1.86	0.58
2:A:508:ARG:HD3	2:A:510[B]:CYS:SG	2.48	0.54
2:A:240:LYS:O	2:A:244:MET:HG2	2.09	0.53
2:A:448:TYR:HB3	2:A:518:TYR:HB3	1.90	0.52
1:D:13:ILE:HD13	1:D:34:GLU:HG2	1.94	0.49
2:A:441:ASP:OD1	2:A:443:LYS:HG2	2.12	0.49
2:A:210:HIS:HB3	3:A:603:HOH:O	2.12	0.49
1:D:8:LEU:CD1	1:D:71:LEU:HD12	2.43	0.49
1:D:8:LEU:HD13	1:D:71:LEU:HD12	1.97	0.47
2:A:549:ARG:HD2	2:A:550:ILE:N	2.29	0.47
2:A:496:HIS:ND1	3:A:614:HOH:O	2.35	0.47
2:A:343:ARG:NH2	3:A:621:HOH:O	2.48	0.47
2:A:234:PHE:CD1	2:A:471:PRO:HB3	2.50	0.46
2:A:549:ARG:HH11	2:A:550:ILE:H	1.63	0.45
1:D:63:LYS:HG2	1:D:64:GLU:HG3	1.98	0.45
2:A:248:GLU:CD	2:A:248:GLU:H	2.21	0.44
2:A:296:VAL:HG11	2:A:514:TYR:HE2	1.82	0.44
3:D:105:HOH:O	2:A:343:ARG:HD2	2.19	0.42
1:D:71:LEU:HD22	1:D:73:LEU:HD23	2.02	0.42
1:D:71:LEU:HD23	1:D:72:ARG:N	2.36	0.41
2:A:491:GLU:CD	2:A:491:GLU:H	2.24	0.41
2:A:412:ASP:HB2	2:A:419:ILE:HD13	2.02	0.41
2:A:314:THR:C	2:A:316:VAL:H	2.25	0.40
2:A:549:ARG:NH1	2:A:550:ILE:HA	2.37	0.40

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)
2:A:344:ARG:NH2	2:A:416:ASP:O[3_545]	1.75	0.45

## 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	D	74/76 (97%)	73 (99%)	1 (1%)	0	100	100
2	A	345/351 (98%)	330 (96%)	15 (4%)	0	100	100
All	All	419/427 (98%)	403 (96%)	16 (4%)	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	D	68/68 (100%)	66 (97%)	2 (3%)	42	16
2	A	315/317 (99%)	305 (97%)	10 (3%)	39	14
All	All	383/385 (100%)	371 (97%)	12 (3%)	39	14

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	1	MET
1	D	15	LEU
2	A	299	LEU

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
2	A	303	LEU
2	A	316	VAL
2	A	402	LEU
2	A	434	ASP
2	A	469	LEU
2	A	470	ASN
2	A	539	GLN
2	A	540	LEU
2	A	549	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	A	294	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

There are no ligands in this entry.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues ⓘ

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	D	76/76 (100%)	-0.12	1 (1%) 77 76	16, 29, 48, 50	0
2	A	341/351 (97%)	-0.06	12 (3%) 44 40	15, 25, 49, 86	0
All	All	417/427 (97%)	-0.07	13 (3%) 49 46	15, 26, 50, 86	0

All (13) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	A	414	GLN	7.0
2	A	550	ILE	6.7
2	A	508	ARG	5.2
2	A	549	ARG	4.7
2	A	413	PRO	4.5
2	A	415	THR	4.4
2	A	210	HIS	3.8
2	A	507	VAL	3.1
2	A	517	VAL	2.7
2	A	412	ASP	2.5
1	D	53	GLY	2.4
2	A	546	GLU	2.1
2	A	219	GLN	2.0

### 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 carbohydrates 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.