



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 19, 2016 – 07:44 PM GMT

PDB ID : 4WPI  
Title : Crystal structure of USP7 ubiquitin-like domains in extended conformation  
Authors : Pföh, R.; Lacdao, I.K.L.; Saridakis, V.  
Deposited on : 2014-10-18  
Resolution : 3.40 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  
<http://wwpdb.org/validation/2016/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 : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026982  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20026982

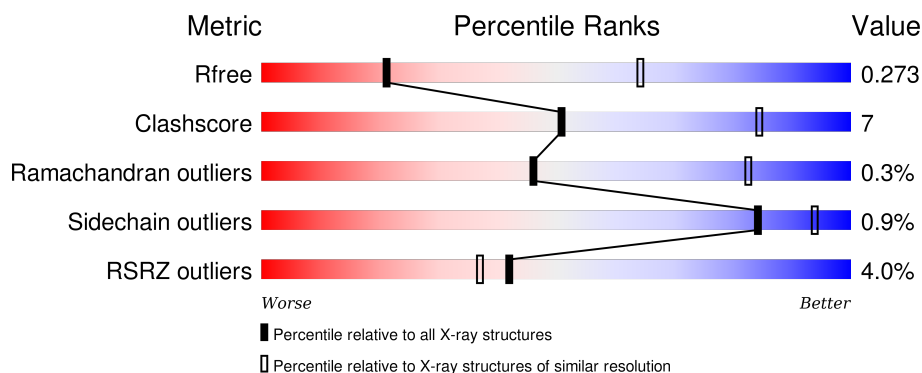
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.40 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))
$R_{free}$	91344	1476 (3.50-3.30)
Clashscore	102246	1611 (3.50-3.30)
Ramachandran outliers	100387	1571 (3.50-3.30)
Sidechain outliers	100360	1571 (3.50-3.30)
RSRZ outliers	91569	1485 (3.50-3.30)

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. 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	375	<div> <div>2%</div> <div>76%</div> <div>15%</div> <div>9%</div> </div>
1	B	375	<div> <div>5%</div> <div>77%</div> <div>14%</div> <div>9%</div> </div>
2	C	11	<div> <div>82%</div> <div>9%</div> <div>9%</div> </div>
2	D	11	<div> <div>73%</div> <div>18%</div> <div>9%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard

residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	CL	A	2002	-	-	X	-
3	CL	B	2001	-	-	X	-

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 5727 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 Ubiquitin carboxyl-terminal hydrolase 7.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	341	Total	C	N	O	S	Se	0	0	0
			2784	1760	474	532	5	13			
1	B	340	Total	C	N	O	S	Se	0	0	0
			2780	1757	477	530	5	11			

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	514	MSE	-	initiating methionine	UNP Q93009
A	515	GLY	-	expression tag	UNP Q93009
A	516	SER	-	expression tag	UNP Q93009
A	517	SER	-	expression tag	UNP Q93009
A	518	HIS	-	expression tag	UNP Q93009
A	519	HIS	-	expression tag	UNP Q93009
A	520	HIS	-	expression tag	UNP Q93009
A	521	HIS	-	expression tag	UNP Q93009
A	522	HIS	-	expression tag	UNP Q93009
A	523	HIS	-	expression tag	UNP Q93009
A	524	SER	-	expression tag	UNP Q93009
A	525	SER	-	expression tag	UNP Q93009
A	526	GLY	-	expression tag	UNP Q93009
A	527	ARG	-	expression tag	UNP Q93009
A	528	GLU	-	expression tag	UNP Q93009
A	529	ASN	-	expression tag	UNP Q93009
A	530	LEU	-	expression tag	UNP Q93009
A	531	TYR	-	expression tag	UNP Q93009
A	532	PHE	-	expression tag	UNP Q93009
A	533	GLN	-	expression tag	UNP Q93009
A	534	GLY	-	expression tag	UNP Q93009
B	514	MSE	-	initiating methionine	UNP Q93009
B	515	GLY	-	expression tag	UNP Q93009
B	516	SER	-	expression tag	UNP Q93009
B	517	SER	-	expression tag	UNP Q93009

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	518	HIS	-	expression tag	UNP Q93009
B	519	HIS	-	expression tag	UNP Q93009
B	520	HIS	-	expression tag	UNP Q93009
B	521	HIS	-	expression tag	UNP Q93009
B	522	HIS	-	expression tag	UNP Q93009
B	523	HIS	-	expression tag	UNP Q93009
B	524	SER	-	expression tag	UNP Q93009
B	525	SER	-	expression tag	UNP Q93009
B	526	GLY	-	expression tag	UNP Q93009
B	527	ARG	-	expression tag	UNP Q93009
B	528	GLU	-	expression tag	UNP Q93009
B	529	ASN	-	expression tag	UNP Q93009
B	530	LEU	-	expression tag	UNP Q93009
B	531	TYR	-	expression tag	UNP Q93009
B	532	PHE	-	expression tag	UNP Q93009
B	533	GLN	-	expression tag	UNP Q93009
B	534	GLY	-	expression tag	UNP Q93009

- Molecule 2 is a protein called ICP0.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	10	Total	C	N	O	S	0	0	0
			80	47	21	11	1			
2	D	10	Total	C	N	O	S	0	0	0
			80	47	21	11	1			

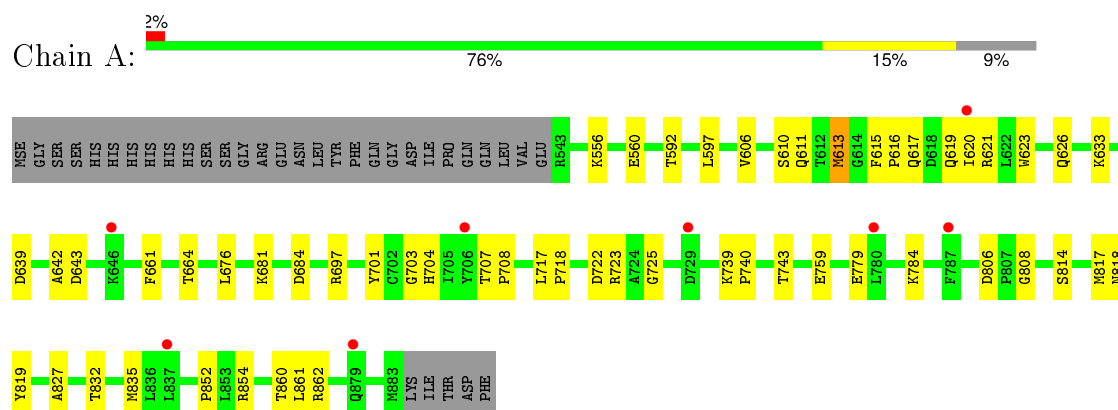
- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Cl	0	0
			1	1		
3	A	2	Total	Cl	0	0
			2	2		

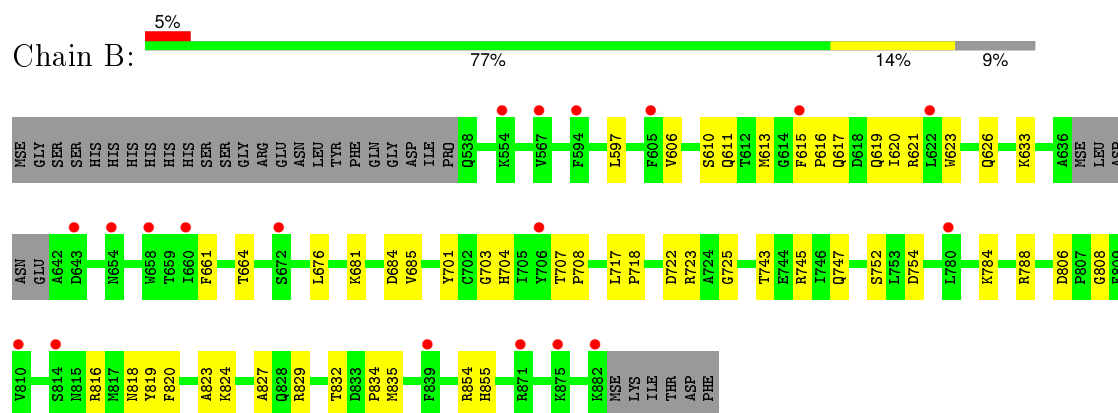
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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 ( $\text{RSRZ} > 2$ ). 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.

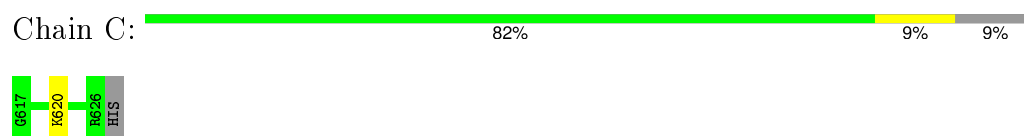
- Molecule 1: Ubiquitin carboxyl-terminal hydrolase 7



- Molecule 1: Ubiquitin carboxyl-terminal hydrolase 7

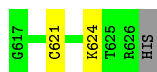


- Molecule 2: ICP0



- Molecule 2: ICP0





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 42 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	164.99Å 164.99Å 110.12Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 3.40 42.26 – 3.40	Depositor EDS
% Data completeness (in resolution range)	99.9 (30.00-3.40) 100.0 (42.26-3.40)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.44 (at 3.40Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, $R_{free}$	0.241 , 0.281 0.235 , 0.273	Depositor DCC
$R_{free}$ test set	1137 reflections (5.60%)	DCC
Wilson B-factor (Å <sup>2</sup> )	135.0	Xtriage
Anisotropy	0.401	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 126.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.28$	Xtriage
Outliers	0 of 21456 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	5727	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	188.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.71% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: 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	0.52	0/2830	0.72	0/3801
1	B	0.49	0/2827	0.71	0/3799
2	C	0.65	0/80	0.86	0/103
2	D	0.78	0/80	0.94	0/103
All	All	0.51	0/5817	0.72	0/7806

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	2784	0	2727	45	0
1	B	2780	0	2725	44	0
2	C	80	0	91	3	0
2	D	80	0	91	2	0
3	A	2	0	0	2	0
3	B	1	0	0	2	0
All	All	5727	0	5634	78	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 7.

The worst 5 of 78 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:A:697:ARG:NE	3:A:2002:CL:CL	2.50	0.82
1:B:788:ARG:NH1	3:B:2001:CL:CL	2.50	0.81
1:A:639:ASP:O	1:A:643:ASP:HB2	1.84	0.76
1:A:835:MSE:HE2	1:B:823:ALA:HB1	1.79	0.65
1:B:819:TYR:CE2	1:B:855:HIS:HB2	2.34	0.63

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	339/375 (90%)	310 (91%)	28 (8%)	1 (0%)	46	82
1	B	336/375 (90%)	307 (91%)	28 (8%)	1 (0%)	46	82
2	C	8/11 (73%)	5 (62%)	3 (38%)	0	100	100
2	D	8/11 (73%)	7 (88%)	1 (12%)	0	100	100
All	All	691/772 (90%)	629 (91%)	60 (9%)	2 (0%)	46	82

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	681	LYS
1	B	681	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	310/329 (94%)	307 (99%)	3 (1%)	82	93
1	B	309/329 (94%)	306 (99%)	3 (1%)	82	93
2	C	8/9 (89%)	8 (100%)	0	100	100
2	D	8/9 (89%)	8 (100%)	0	100	100
All	All	635/676 (94%)	629 (99%)	6 (1%)	84	94

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	613	MSE
1	B	613	MSE
1	B	597	LEU
1	A	611	GLN
1	B	611	GLN

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

### 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 ⓘ

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	A	328/375 (87%)	0.24	8 (2%) 62 57	123, 168, 256, 298	0
1	B	329/375 (87%)	0.33	19 (5%) 26 24	115, 199, 268, 350	0
2	C	10/11 (90%)	-0.32	0 100 100	173, 193, 207, 232	0
2	D	10/11 (90%)	-0.06	0 100 100	146, 159, 209, 213	0
All	All	677/772 (87%)	0.27	27 (3%) 42 37	115, 182, 267, 350	0

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	654	ASN	4.6
1	B	810	VAL	3.4
1	B	605	PHE	3.1
1	A	729	ASP	2.9
1	B	871	ARG	2.8

### 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](#)

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	CL	A	2001	1/1	0.94	0.12	-	136,136,136,136	0
3	CL	A	2002	1/1	0.87	0.20	-	143,143,143,143	0
3	CL	B	2001	1/1	0.91	0.22	-	170,170,170,170	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.