Full wwPDB X-ray Structure Validation Report (i)

May 29, 2020 – 07:54 am BST

PDB ID	:	5OHP
Title	:	Crystal structure of USP30 (C77A) in complex with Lys6-linked diubiquitin
Authors	:	Gersch, M.; Komander, D.
Deposited on	:	2017-07-17
Resolution	:	2.80 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

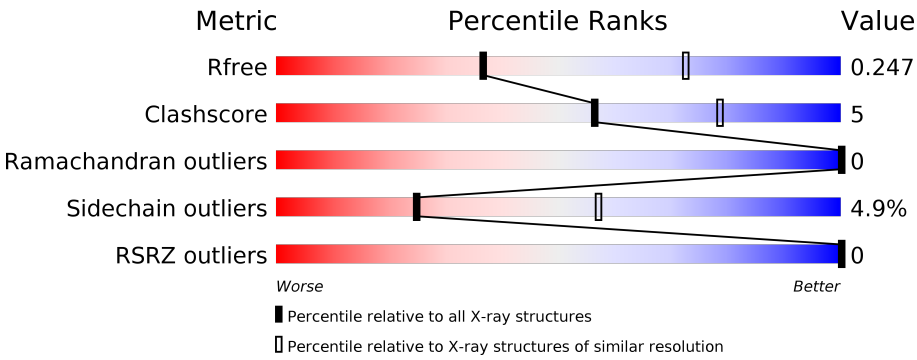
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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}	130704	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (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 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 <=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	336	78% 14% • 6%
2	B	76	89% 8% •
2	C	76	86% 11% • •

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3731 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 30.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	315	2508	1587	442	463	16	0	0	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	62	GLY	-	expression tag	UNP Q70CQ3
A	63	PRO	-	expression tag	UNP Q70CQ3
A	77	ALA	CYS	engineered mutation	UNP Q70CQ3
A	179	GLY	-	linker	UNP Q70CQ3
A	180	SER	-	linker	UNP Q70CQ3
A	181	GLY	-	linker	UNP Q70CQ3
A	182	SER	-	linker	UNP Q70CQ3
A	348	ASP	PHE	engineered mutation	UNP Q70CQ3
A	350	SER	MET	engineered mutation	UNP Q70CQ3
A	353	GLU	ILE	engineered mutation	UNP Q70CQ3
A	358	SER	-	linker	UNP Q70CQ3
A	359	ASN	-	linker	UNP Q70CQ3
A	360	ALA	-	linker	UNP Q70CQ3

- Molecule 2 is a protein called Polyubiquitin-B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	76	594	373	103	117	1	0	0	0
			Total	C	N	O	S			
2	C	74	576	366	100	109	1	0	0	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by author).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
3	A	1	1	1	0	0

- Molecule 4 is water.

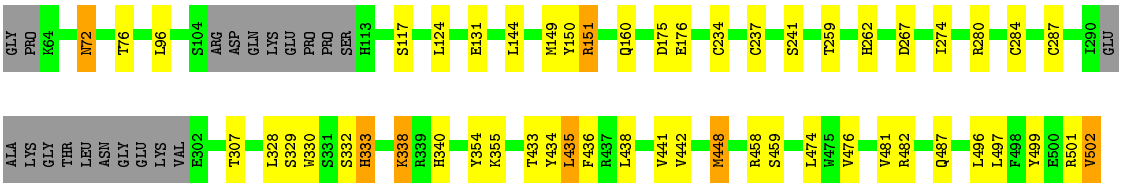
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	32	32	32	0	0
			Total	O		
4	B	9	9	9	0	0
			Total	O		
4	C	11	11	11	0	0

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.

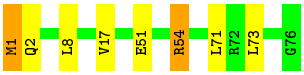
- Molecule 1: Ubiquitin carboxyl-terminal hydrolase 30

Chain A:



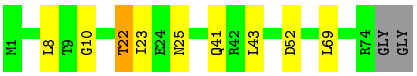
- Molecule 2: Polyubiquitin-B

Chain B:



- Molecule 2: Polyubiquitin-B

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, α, β, γ	117.50Å 117.50Å 138.84Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	82.08 – 2.80 82.08 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.2 (82.08-2.80) 92.3 (82.08-2.80)	Depositor EDS
R_{merge}	0.17	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.70 (at 2.82Å)	Xtriage
Refinement program	PHENIX (1.11rc1_2513)	Depositor
R, R_{free}	0.216 , 0.249 0.216 , 0.247	Depositor DCC
R_{free} test set	729 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	43.7	Xtriage
Anisotropy	0.652	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 29.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	3731	wwPDB-VP
Average B, all atoms (Å^2)	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.58% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: ZN

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	0.30	0/2576	0.46	0/3499
2	B	0.38	0/600	0.52	0/809
2	C	0.58	0/582	0.59	0/785
All	All	0.37	0/3758	0.49	0/5093

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	2508	0	2373	29	1
2	B	594	0	609	6	0
2	C	576	0	599	5	1
3	A	1	0	0	0	0
4	A	32	0	0	1	0
4	B	9	0	0	0	0
4	C	11	0	0	0	0
All	All	3731	0	3581	36	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 5.

All (36) 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:160:GLN:OE1	4:A:901:HOH:O	1.90	0.89
2:B:51:GLU:OE2	2:B:54:ARG:NH1	2.23	0.72
1:A:438:LEU:O	1:A:458:ARG:NH2	2.28	0.66
1:A:442:VAL:HB	1:A:496:LEU:HB2	1.77	0.66
1:A:151:ARG:NH2	1:A:175:ASP:OD2	2.24	0.64
1:A:72:ASN:O	2:C:10:GLY:HA2	1.99	0.62
1:A:328:LEU:HD13	2:B:73:LEU:HD13	1.80	0.62
1:A:284:CYS:HB3	1:A:287:CYS:HB2	1.83	0.59
1:A:96:LEU:CD1	1:A:124:LEU:HD12	2.35	0.56
1:A:476:VAL:HG13	1:A:481:VAL:HG22	1.87	0.56
1:A:72:ASN:ND2	1:A:76:THR:H	2.07	0.52
1:A:438:LEU:HD21	1:A:441:VAL:HG23	1.93	0.50
1:A:328:LEU:HD23	2:B:71:LEU:HB3	1.93	0.50
1:A:280:ARG:NH1	1:A:307:THR:HG21	2.27	0.50
1:A:149:MET:HG3	1:A:150:TYR:CD2	2.47	0.50
1:A:441:VAL:HG22	1:A:497:LEU:HD23	1.94	0.49
1:A:328:LEU:HD11	1:A:448:MET:HB3	1.93	0.49
1:A:330:TRP:HB2	2:B:8:LEU:HD12	1.94	0.48
1:A:259:THR:OG1	1:A:262:HIS:HB2	2.13	0.48
1:A:435:LEU:CD2	1:A:502:VAL:HG21	2.44	0.48
1:A:474:LEU:HD12	1:A:481:VAL:HG12	1.96	0.48
2:B:1:MET:N	2:B:17:VAL:O	2.34	0.47
1:A:436:PHE:HB3	1:A:499:TYR:HB3	1.97	0.47
2:C:23:ILE:HB	2:C:52:ASP:HA	1.97	0.47
1:A:274:ILE:HD13	1:A:354:TYR:HD1	1.81	0.46
1:A:338:LYS:HG2	1:A:340:HIS:CE1	2.51	0.45
2:C:43:LEU:HD23	2:C:43:LEU:HA	1.82	0.45
1:A:355:LYS:HD2	1:A:434:TYR:CE2	2.51	0.45
1:A:433:THR:O	1:A:501:ARG:NH2	2.51	0.44
1:A:435:LEU:HD23	1:A:502:VAL:HG21	2.00	0.44
2:C:22:THR:O	2:C:25:ASN:HB3	2.18	0.43
1:A:117:SER:OG	1:A:176:GLU:OE1	2.25	0.43
1:A:333:HIS:ND1	1:A:333:HIS:N	2.67	0.42
2:B:54:ARG:HD2	2:B:54:ARG:HH21	1.70	0.42
2:C:41:GLN:HB3	2:C:69:LEU:HD11	2.01	0.42
1:A:237:CYS:HB3	1:A:287:CYS:SG	2.61	0.41

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:A:487:GLN:NE2	1:A:487:GLN:NE2[12_564]	1.43	0.77
2:C:8:LEU:CD2	2:C:8:LEU:CD2[7_554]	1.62	0.58

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	309/336 (92%)	297 (96%)	12 (4%)	0	100	100
2	B	74/76 (97%)	72 (97%)	2 (3%)	0	100	100
2	C	72/76 (95%)	72 (100%)	0	0	100	100
All	All	455/488 (93%)	441 (97%)	14 (3%)	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	279/307 (91%)	263 (94%)	16 (6%)	20	50
2	B	66/68 (97%)	63 (96%)	3 (4%)	27	60
2	C	64/68 (94%)	63 (98%)	1 (2%)	62	88
All	All	409/443 (92%)	389 (95%)	20 (5%)	25	57

All (20) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	72	ASN
1	A	131	GLU
1	A	144	LEU
1	A	151	ARG
1	A	234	CYS
1	A	241	SER
1	A	267	ASP
1	A	329	SER
1	A	332	SER
1	A	333	HIS
1	A	338	LYS
1	A	435	LEU
1	A	448	MET
1	A	459	SER
1	A	482	ARG
1	A	502	VAL
2	B	1	MET
2	B	2	GLN
2	B	54	ARG
2	C	22	THR

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

Mol	Chain	Res	Type
1	A	72	ASN

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 carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 1 ligands modelled in this entry, 1 is 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 (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	315/336 (93%)	-0.01	0	100	100	29, 46, 76, 96	0
2	B	76/76 (100%)	-0.16	0	100	100	28, 41, 56, 59	0
2	C	74/76 (97%)	-0.04	0	100	100	31, 51, 62, 77	0
All	All	465/488 (95%)	-0.04	0	100	100	28, 46, 72, 96	0

There are no RSRZ outliers to report.

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)

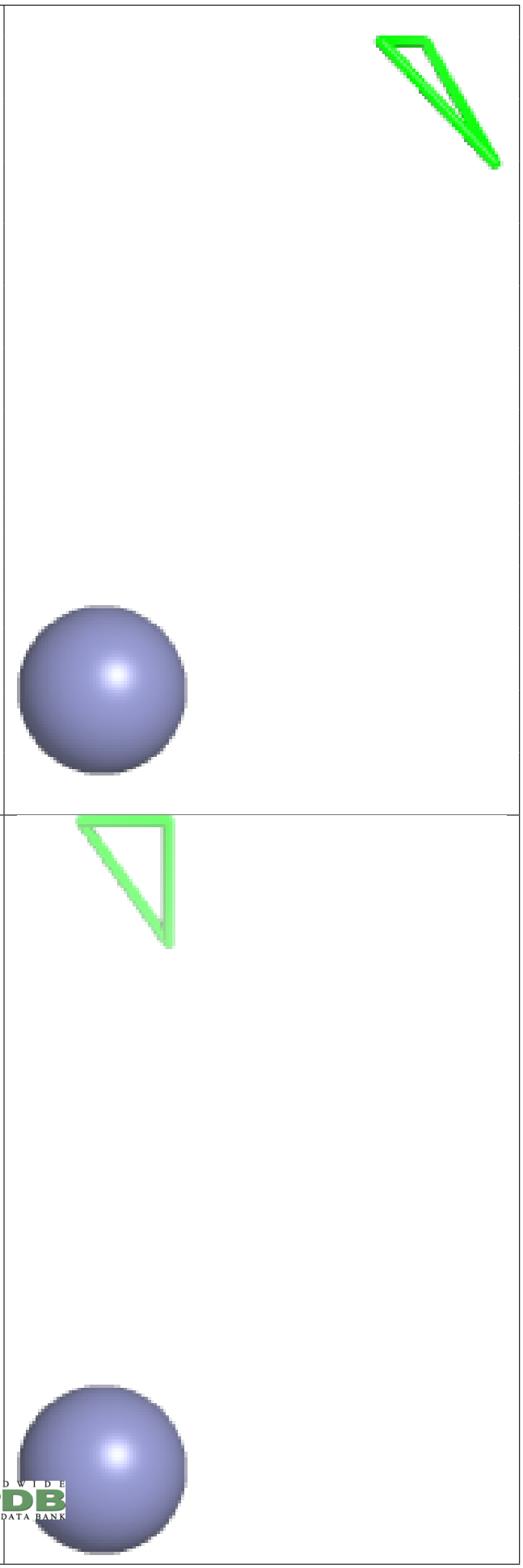
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	ZN	A	800	1/1	0.99	0.12	65,65,65,65	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around ZN A 800:

2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative)
and green (positive)



6.5 Other polymers (i)

There are no such residues in this entry.