wwPDB X-ray Structure Validation Summary Report (i)

Feb 23, 2022 – 12:12 pm GMT

PDB ID	:	7P99
Title	:	Structure of human USPL1 in complex with SUMO2
Authors	:	Reverter, D.; Ying, L.
Deposited on	:	2021-07-26
Resolution	:	1.80 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*
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:

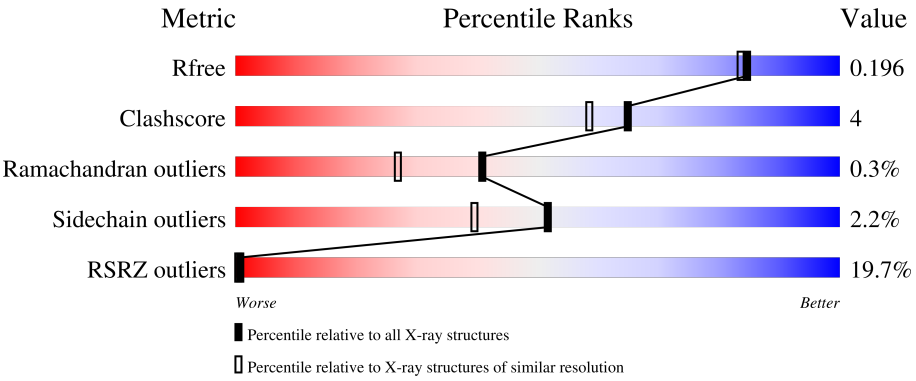
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.26
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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 1.80 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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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 >=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	338	12% 73% 5% 21%
2	C	101	26% 63% 13% · 22%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3000 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 SUMO-specific isopeptidase USPL1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	266	Total	C	N	O	S	0	0	0
			2185	1410	363	395	17			

There are 34 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	179	MET	-	initiating methionine	UNP Q5W0Q7
A	180	GLY	-	expression tag	UNP Q5W0Q7
A	181	SER	-	expression tag	UNP Q5W0Q7
A	182	SER	-	expression tag	UNP Q5W0Q7
A	183	HIS	-	expression tag	UNP Q5W0Q7
A	184	HIS	-	expression tag	UNP Q5W0Q7
A	185	HIS	-	expression tag	UNP Q5W0Q7
A	186	HIS	-	expression tag	UNP Q5W0Q7
A	187	HIS	-	expression tag	UNP Q5W0Q7
A	188	HIS	-	expression tag	UNP Q5W0Q7
A	189	SER	-	expression tag	UNP Q5W0Q7
A	190	SER	-	expression tag	UNP Q5W0Q7
A	191	GLY	-	expression tag	UNP Q5W0Q7
A	192	LEU	-	expression tag	UNP Q5W0Q7
A	193	VAL	-	expression tag	UNP Q5W0Q7
A	194	PRO	-	expression tag	UNP Q5W0Q7
A	195	ARG	-	expression tag	UNP Q5W0Q7
A	196	GLY	-	expression tag	UNP Q5W0Q7
A	197	SER	-	expression tag	UNP Q5W0Q7
A	198	HIS	-	expression tag	UNP Q5W0Q7
A	199	MET	-	expression tag	UNP Q5W0Q7
A	200	ALA	-	expression tag	UNP Q5W0Q7
A	201	SER	-	expression tag	UNP Q5W0Q7
A	202	MET	-	expression tag	UNP Q5W0Q7
A	203	THR	-	expression tag	UNP Q5W0Q7
A	204	GLY	-	expression tag	UNP Q5W0Q7
A	205	GLY	-	expression tag	UNP Q5W0Q7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	206	GLN	-	expression tag	UNP Q5W0Q7
A	207	GLN	-	expression tag	UNP Q5W0Q7
A	208	MET	-	expression tag	UNP Q5W0Q7
A	209	GLY	-	expression tag	UNP Q5W0Q7
A	210	ARG	-	expression tag	UNP Q5W0Q7
A	211	GLY	-	expression tag	UNP Q5W0Q7
A	212	SER	-	expression tag	UNP Q5W0Q7

- Molecule 2 is a protein called Small ubiquitin-related modifier.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	79	Total	C	N	O	S	0	0	0
			638	397	114	124	3			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-5	MET	-	initiating methionine	UNP A0A6J0CIQ7
C	-4	GLY	-	expression tag	UNP A0A6J0CIQ7
C	-3	SER	-	expression tag	UNP A0A6J0CIQ7
C	-2	SER	-	expression tag	UNP A0A6J0CIQ7
C	-1	HIS	-	expression tag	UNP A0A6J0CIQ7
C	0	HIS	-	expression tag	UNP A0A6J0CIQ7
C	1	HIS	-	expression tag	UNP A0A6J0CIQ7
C	2	HIS	-	expression tag	UNP A0A6J0CIQ7
C	3	HIS	-	expression tag	UNP A0A6J0CIQ7
C	4	HIS	-	expression tag	UNP A0A6J0CIQ7
C	5	SER	-	expression tag	UNP A0A6J0CIQ7
C	6	SER	-	expression tag	UNP A0A6J0CIQ7
C	7	GLY	-	expression tag	UNP A0A6J0CIQ7
C	8	LEU	-	expression tag	UNP A0A6J0CIQ7
C	9	VAL	-	expression tag	UNP A0A6J0CIQ7
C	10	PRO	-	expression tag	UNP A0A6J0CIQ7
C	11	ARG	-	expression tag	UNP A0A6J0CIQ7
C	12	GLY	-	expression tag	UNP A0A6J0CIQ7
C	13	SER	-	expression tag	UNP A0A6J0CIQ7
C	14	HIS	-	expression tag	UNP A0A6J0CIQ7
C	48	SER	CYS	engineered mutation	UNP A0A6J0CIQ7
C	93	ALA	GLY	engineered mutation	UNP A0A6J0CIQ7

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

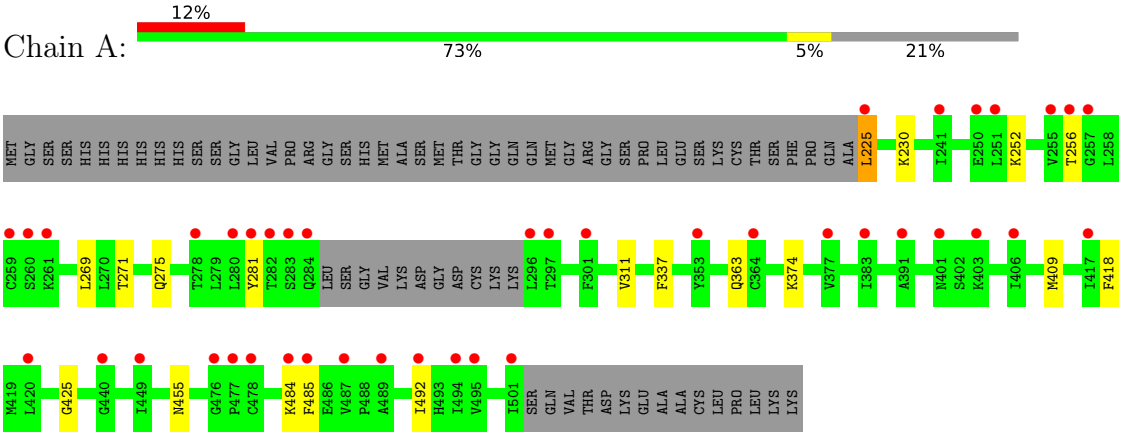
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	140	Total O 140 140	0	0
4	C	36	Total O 36 36	0	0

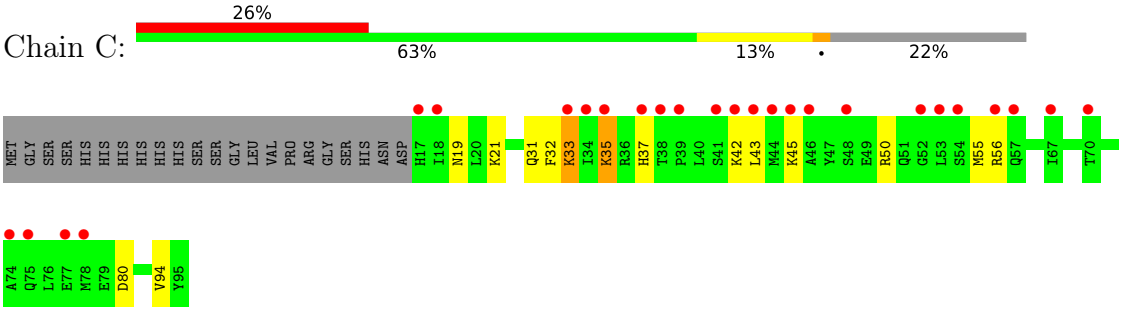
3 Residue-property plots (i)

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: SUMO-specific isopeptidase USPL1



• Molecule 2: Small ubiquitin-related modifier



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	50.71Å 69.88Å 53.64Å 90.00° 90.66° 90.00°	Depositor
Resolution (Å)	41.04 – 1.80 41.04 – 1.80	Depositor EDS
% Data completeness (in resolution range)	96.4 (41.04-1.80) 96.5 (41.04-1.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.01 (at 1.79Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.174 , 0.197 0.175 , 0.196	Depositor DCC
R_{free} test set	1626 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å^2)	25.4	Xtriage
Anisotropy	0.465	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.032 for h,-k,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	3000	wwPDB-VP
Average B, all atoms (Å^2)	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.49% 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

5.1 Standard geometry

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	0.38	0/2245	0.53	0/3047
2	C	0.41	0/648	0.55	0/869
All	All	0.39	0/2893	0.54	0/3916

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

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	2185	0	2123	15	0
2	C	638	0	628	9	0
3	A	1	0	0	0	0
4	A	140	0	0	6	0
4	C	36	0	0	2	0
All	All	3000	0	2751	23	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 4.

The worst 5 of 23 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:275:GLN:NE2	4:A:701:HOH:O	2.14	0.81
2:C:19:ASN:N	2:C:80:ASP:OD1	2.12	0.80
1:A:269:LEU:HA	1:A:311:VAL:HG21	1.74	0.70
2:C:35:LYS:HD3	2:C:37:HIS:H	1.58	0.69
2:C:33:LYS:H	2:C:33:LYS:HD3	1.73	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	262/338 (78%)	256 (98%)	5 (2%)	1 (0%)	34 21
2	C	77/101 (76%)	76 (99%)	1 (1%)	0	100 100
All	All	339/439 (77%)	332 (98%)	6 (2%)	1 (0%)	41 27

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	485	PHE

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	246/305 (81%)	243 (99%)	3 (1%)	71 65

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	C	70/89 (79%)	66 (94%)	4 (6%)	20 8
All	All	316/394 (80%)	309 (98%)	7 (2%)	52 39

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	33	LYS
2	C	35	LYS
2	C	56	ARG
2	C	42	LYS
1	A	337	PHE

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

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 monosaccharides 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

6.1 Protein, DNA and RNA chains

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^{th} 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	266/338 (78%)	1.32	42 (15%)	2	1	17, 28, 52, 69	0
2	C	79/101 (78%)	1.76	26 (32%)	0	0	19, 42, 61, 70	0
All	All	345/439 (78%)	1.42	68 (19%)	1	0	17, 30, 59, 70	0

The worst 5 of 68 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	281	TYR	9.2
2	C	34	ILE	6.8
1	A	225	LEU	6.8
1	A	501	ILE	6.3
1	A	283	SER	6.2

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

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, 95^{th} 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	601	1/1	0.99	0.09	24,24,24,24	0

6.5 Other polymers (i)

There are no such residues in this entry.