



wwPDB X-ray Structure Validation Summary Report

Jul 29, 2014 – 11:24 PM EDT

PDB ID : 4PYU
Title : The conserved ubiquitin-like protein hub1 plays a critical role in splicing in human cells
Authors : Ammon, T.; Mishra, S.K.; Kowalska, K.; Popowicz, G.M.; Holak, T.A.; Jentsch, S.
Deposited on : 2014-03-28
Resolution : 2.00 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

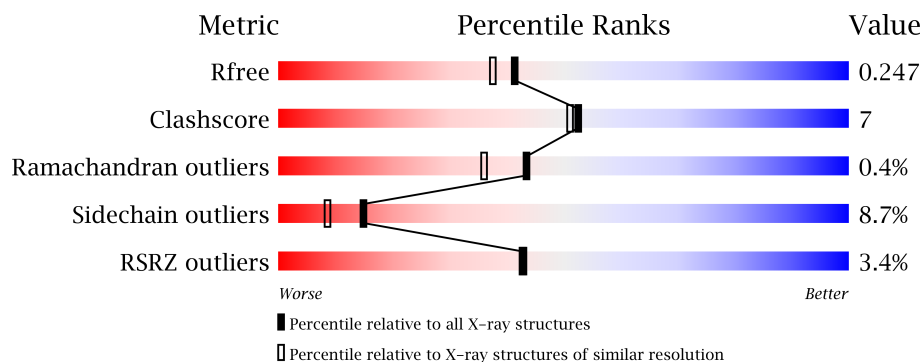
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.16 November 2013
Xtriage (Phenix) : dev-1439
EDS : stable23489
Percentile statistics : 21963
Refmac : 5.8.0049
CCP4 : 6.3.0 (Settle)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable23489

1 Overall quality at a glance

The reported resolution of this entry is 2.00 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}	66092	4888 (2.00-2.00)
Clashscore	79885	6188 (2.00-2.00)
Ramachandran outliers	78287	6102 (2.00-2.00)
Sidechain outliers	78261	6100 (2.00-2.00)
RSRZ outliers	66119	4890 (2.00-2.00)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	76	
1	B	76	
1	G	76	
1	K	76	
1	O	76	
1	S	76	
2	C	19	
2	D	19	
2	H	19	
2	L	19	
2	P	19	
2	T	19	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4741 atoms, of which 0 are hydrogen and 0 are deuterium.

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-like protein 5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	76	Total	C	N	O	S	0	0	0
			600	385	98	113	4			
1	B	75	Total	C	N	O	S	0	0	0
			594	382	98	110	4			
1	G	74	Total	C	N	O	S	0	0	0
			597	382	101	110	4			
1	K	74	Total	C	N	O	S	0	0	0
			588	376	98	110	4			
1	O	75	Total	C	N	O	S	0	0	0
			605	386	103	112	4			
1	S	75	Total	C	N	O	S	0	0	0
			603	386	101	112	4			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	EXPRESSION TAG	UNP Q9BZL1
A	-1	SER	-	EXPRESSION TAG	UNP Q9BZL1
A	0	HIS	-	EXPRESSION TAG	UNP Q9BZL1
B	-2	GLY	-	EXPRESSION TAG	UNP Q9BZL1
B	-1	SER	-	EXPRESSION TAG	UNP Q9BZL1
B	0	HIS	-	EXPRESSION TAG	UNP Q9BZL1
G	-2	GLY	-	EXPRESSION TAG	UNP Q9BZL1
G	-1	SER	-	EXPRESSION TAG	UNP Q9BZL1
G	0	HIS	-	EXPRESSION TAG	UNP Q9BZL1
K	-2	GLY	-	EXPRESSION TAG	UNP Q9BZL1
K	-1	SER	-	EXPRESSION TAG	UNP Q9BZL1
K	0	HIS	-	EXPRESSION TAG	UNP Q9BZL1
O	-2	GLY	-	EXPRESSION TAG	UNP Q9BZL1
O	-1	SER	-	EXPRESSION TAG	UNP Q9BZL1
O	0	HIS	-	EXPRESSION TAG	UNP Q9BZL1
S	-2	GLY	-	EXPRESSION TAG	UNP Q9BZL1
S	-1	SER	-	EXPRESSION TAG	UNP Q9BZL1

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
S	0	HIS	-	EXPRESSION TAG	UNP Q9BZL1

- Molecule 2 is a protein called U4/U6.U5 tri-snRNP-associated protein 1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	19	Total	C	N	O	0	0	0
			146	92	26	28			
2	D	19	Total	C	N	O	0	0	0
			145	92	25	28			
2	H	19	Total	C	N	O	0	0	0
			144	92	24	28			
2	L	19	Total	C	N	O	0	0	0
			146	93	25	28			
2	P	19	Total	C	N	O	0	0	0
			142	92	24	26			
2	T	19	Total	C	N	O	0	0	0
			144	92	24	28			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	44	Total	O	0	0
			44	44		
3	B	24	Total	O	0	0
			24	24		
3	C	14	Total	O	0	0
			14	14		
3	D	9	Total	O	0	0
			9	9		
3	G	34	Total	O	0	0
			34	34		
3	H	19	Total	O	0	0
			19	19		
3	K	27	Total	O	0	0
			27	27		
3	L	15	Total	O	0	0
			15	15		
3	O	41	Total	O	0	0
			41	41		
3	P	10	Total	O	0	0
			10	10		
3	S	41	Total	O	0	0
			41	41		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	T	9	Total	O	0	0
			9	9		

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 ($RSRZ > 2$). 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.

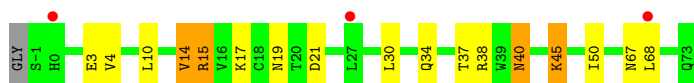
- Molecule 1: Ubiquitin-like protein 5

Chain A: 



- Molecule 1: Ubiquitin-like protein 5

Chain B: 



- Molecule 1: Ubiquitin-like protein 5

Chain G: 



- Molecule 1: Ubiquitin-like protein 5

Chain K: 



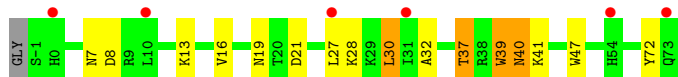
- Molecule 1: Ubiquitin-like protein 5

Chain O: 



- Molecule 1: Ubiquitin-like protein 5

Chain S: 



- Molecule 2: U4/U6.U5 tri-snRNP-associated protein 1

Chain C: 



- Molecule 2: U4/U6.U5 tri-snRNP-associated protein 1

Chain D: 



- Molecule 2: U4/U6.U5 tri-snRNP-associated protein 1

Chain H: 



- Molecule 2: U4/U6.U5 tri-snRNP-associated protein 1

Chain L: 



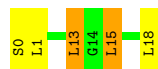
- Molecule 2: U4/U6.U5 tri-snRNP-associated protein 1

Chain P: 



- Molecule 2: U4/U6.U5 tri-snRNP-associated protein 1

Chain T: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	87.51Å 103.63Å 67.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.97 – 2.00 19.97 – 2.00	Depositor EDS
% Data completeness (in resolution range)	87.0 (19.97-2.00) 87.2 (19.97-2.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	5.24 (at 2.01Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, R_{free}	0.186 , 0.246 0.192 , 0.247	Depositor DCC
R_{free} test set	1839 reflections (5.31%)	DCC
Wilson B-factor (Å ²)	29.6	Xtriage
Anisotropy	0.090	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.39 , 50.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	0 of 36473 reflections	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	4741	wwPDB-VP
Average B, all atoms (Å ²)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 39.39 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 3.1981e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

5 Model quality

5.1 Standard geometry

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	0.81	0/611	1.12	4/826 (0.5%)
1	B	0.86	0/605	1.03	3/818 (0.4%)
1	G	0.79	0/608	1.00	2/821 (0.2%)
1	K	0.70	0/599	0.88	1/810 (0.1%)
1	O	0.80	0/616	0.90	1/830 (0.1%)
1	S	0.75	0/614	0.84	0/829
2	C	0.79	0/145	0.99	1/192 (0.5%)
2	D	0.71	0/145	0.99	1/192 (0.5%)
2	H	1.00	0/144	1.13	2/191 (1.0%)
2	L	0.77	0/146	0.94	0/193
2	P	0.78	0/141	0.89	0/187
2	T	0.85	0/144	1.07	0/191
All	All	0.79	0/4518	0.97	15/6080 (0.2%)

There are no bond length outliers.

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	15	ARG	NE-CZ-NH2	-11.41	114.60	120.30
1	A	15	ARG	NE-CZ-NH1	10.29	125.45	120.30
1	A	14	VAL	CB-CA-C	-9.17	93.97	111.40
1	B	15	ARG	NE-CZ-NH1	7.93	124.26	120.30
1	B	15	ARG	NE-CZ-NH2	-7.67	116.47	120.30

There are no chirality outliers.

There are no planarity outliers.

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit,

and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	600	0	579	8	0
1	B	594	0	580	11	0
1	G	597	0	590	9	0
1	K	588	0	571	9	0
1	O	605	0	600	7	0
1	S	603	0	593	11	0
2	C	146	0	161	0	0
2	D	145	0	161	1	0
2	H	144	0	156	2	0
2	L	146	0	163	4	0
2	P	142	0	154	3	0
2	T	144	0	156	1	0
3	A	44	0	0	2	0
3	B	24	0	0	1	0
3	C	14	0	0	0	0
3	D	9	0	0	0	0
3	G	34	0	0	0	0
3	H	19	0	0	2	0
3	K	27	0	0	3	0
3	L	15	0	0	2	0
3	O	41	0	0	1	0
3	P	10	0	0	1	0
3	S	41	0	0	0	0
3	T	9	0	0	0	0
All	All	4741	0	4464	63	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 7.

The worst 5 of 63 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:O:32:ALA:HA	1:O:37:THR:HG22	1.54	0.89
2:L:4:GLU:HG3	3:L:111:HOH:O	1.72	0.89
1:B:45:LYS:HD2	1:B:50:ILE:HG12	1.56	0.87
2:P:4:GLU:CG	3:P:108:HOH:O	2.24	0.85
3:K:119:HOH:O	2:L:0:SER:HA	1.79	0.82

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone ⓘ

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	74/76 (97%)	72 (97%)	2 (3%)	0	100	100
1	B	73/76 (96%)	71 (97%)	2 (3%)	0	100	100
1	G	72/76 (95%)	70 (97%)	2 (3%)	0	100	100
1	K	72/76 (95%)	67 (93%)	4 (6%)	1 (1%)	16	7
1	O	73/76 (96%)	72 (99%)	1 (1%)	0	100	100
1	S	73/76 (96%)	72 (99%)	0	1 (1%)	16	7
2	C	17/19 (90%)	17 (100%)	0	0	100	100
2	D	17/19 (90%)	17 (100%)	0	0	100	100
2	H	17/19 (90%)	17 (100%)	0	0	100	100
2	L	17/19 (90%)	17 (100%)	0	0	100	100
2	P	17/19 (90%)	17 (100%)	0	0	100	100
2	T	17/19 (90%)	17 (100%)	0	0	100	100
All	All	539/570 (95%)	526 (98%)	11 (2%)	2 (0%)	43	36

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	K	39	TRP
1	S	47	TRP

5.3.2 Protein sidechains ⓘ

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	61/68 (90%)	57 (93%)	4 (7%)	24	16
1	B	62/68 (91%)	58 (94%)	4 (6%)	24	17
1	G	63/68 (93%)	56 (89%)	7 (11%)	9	5
1	K	62/68 (91%)	61 (98%)	1 (2%)	75	77
1	O	65/68 (96%)	62 (95%)	3 (5%)	37	30
1	S	64/68 (94%)	58 (91%)	6 (9%)	13	7
2	C	16/17 (94%)	14 (88%)	2 (12%)	7	3
2	D	16/17 (94%)	15 (94%)	1 (6%)	25	18
2	H	15/17 (88%)	11 (73%)	4 (27%)	1	0
2	L	16/17 (94%)	14 (88%)	2 (12%)	7	3
2	P	14/17 (82%)	12 (86%)	2 (14%)	5	2
2	T	15/17 (88%)	10 (67%)	5 (33%)	0	0
All	All	469/510 (92%)	428 (91%)	41 (9%)	15	8

5 of 41 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	H	1	LEU
1	K	27	LEU
2	T	1	LEU
2	H	4	GLU
2	H	13	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	40	ASN
1	B	67	ASN
1	O	73	GLN
1	B	34	GLN
1	K	73	GLN

5.3.3 RNA ⓘ

There are no RNA chains 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 ⓘ

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, 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	76/76 (100%)	-0.06	2 (2%) 53 53	22, 30, 40, 65	0
1	B	75/76 (98%)	-0.01	3 (4%) 36 36	25, 31, 44, 58	0
1	G	74/76 (97%)	-0.06	1 (1%) 72 72	21, 32, 50, 71	0
1	K	74/76 (97%)	0.31	5 (6%) 17 16	28, 37, 55, 68	0
1	O	75/76 (98%)	-0.08	1 (1%) 74 74	22, 30, 45, 68	0
1	S	75/76 (98%)	0.24	6 (8%) 12 12	24, 36, 53, 61	0
2	C	19/19 (100%)	-0.36	1 (5%) 25 25	20, 27, 37, 41	0
2	D	19/19 (100%)	-0.34	0 100 100	26, 34, 46, 49	0
2	H	19/19 (100%)	-0.32	0 100 100	24, 30, 43, 48	0
2	L	19/19 (100%)	-0.26	0 100 100	29, 32, 44, 52	0
2	P	19/19 (100%)	-0.15	0 100 100	27, 36, 46, 57	0
2	T	19/19 (100%)	-0.05	0 100 100	25, 38, 49, 55	0
All	All	563/570 (98%)	-0.01	19 (3%) 43 43	20, 32, 50, 71	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	S	10	LEU	3.2
1	K	39	TRP	2.8
1	S	73	GLN	2.8
1	K	27	LEU	2.8
1	B	27	LEU	2.7

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 carbohydrates in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.