wwPDB NMR Structure Validation Summary Report

Jun 7, 2020 – 03:39 am BST

PDB ID	:	6N13
Title	:	UbcH7-Ub Complex with R0RBR Parkin and phosphoubiquitin
Authors	:	Condos, T.E.C.; Dunkerley, K.M.; Freeman, E.A.; Barber, K.R.; Aguirre, J.D.; Chaugule, V.K.; Xiao, Y.; Konermann, L.; Walden, H.; Shaw, G.S.
Deposited on	:	2018-11-08

This is a wwPDB NMR 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/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

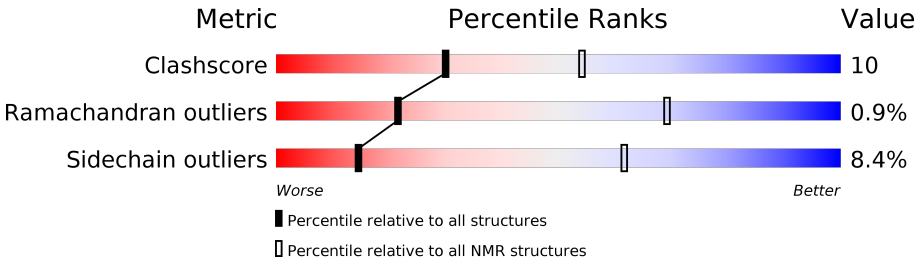
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
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:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 5%.

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)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and 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%

Mol	Chain	Length	Quality of chain
1	B	322	53% 29% • 16%
2	D	76	72% 16% 5% 7%
3	C	156	69% 26% • •
4	A	76	68% 21% • 9%

2 Ensemble composition and analysis

This entry contains 10 models. Model 8 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	B:144-B:337, B:389-B:465, C:501-C:654, A:1-A:64, A:66-A:70 (494)	0.37	8
2	D:701-D:771 (71)	0.21	3

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 3 clusters and 2 single-model clusters were found.

Cluster number	Models
1	2, 4, 5, 6
2	3, 8
3	7, 10
Single-model clusters	1; 9

3 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 6124 atoms, of which 1146 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called E3 ubiquitin-protein ligase parkin.

Mol	Chain	Residues	Atoms						Trace
1	B	322	Total	C	H	N	O	S	0
			3070	1548	567	454	462	39	

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	347	CYS	GLN	engineered mutation	UNP O60260

- Molecule 2 is a protein called ubiquitin.

Mol	Chain	Residues	Atoms						Trace
2	D	76	Total	C	H	N	O	S	0
			746	378	145	105	117	1	

- Molecule 3 is a protein called Ubiquitin-conjugating enzyme E2 L3.

Mol	Chain	Residues	Atoms						Trace
3	C	154	Total	C	H	N	O	S	0
			1551	807	290	218	233	3	

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	499	GLY	-	expression tag	UNP P68036
C	500	HIS	-	expression tag	UNP P68036
C	517	SER	CYS	engineered mutation	UNP P68036
C	586	LYS	CYS	engineered mutation	UNP P68036
C	637	SER	CYS	engineered mutation	UNP P68036

- Molecule 4 is a protein called phosphoubiquitin.

Mol	Chain	Residues	Atoms							Trace
4	A	76	Total	C	H	N	O	P	S	0
			749	378	144	105	120	1	1	

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
5	B	8	8	8

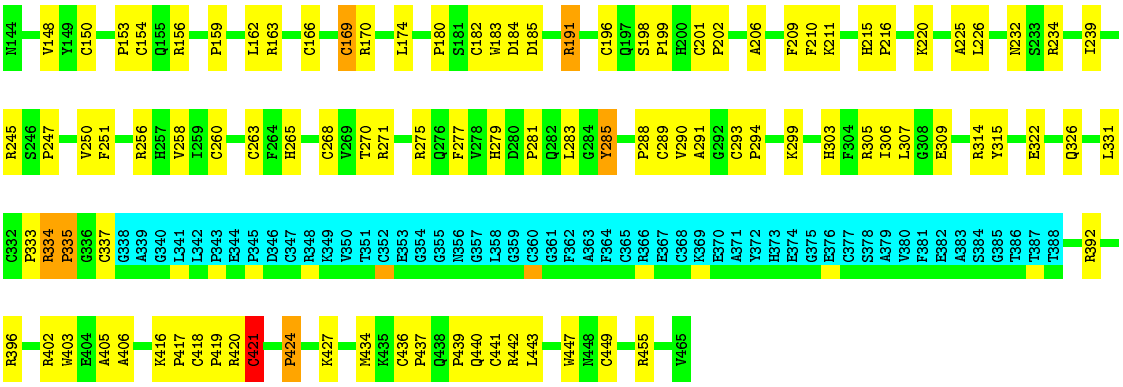
4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-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. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

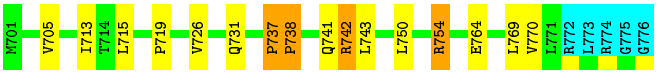
• Molecule 1: E3 ubiquitin-protein ligase parkin

Chain B:



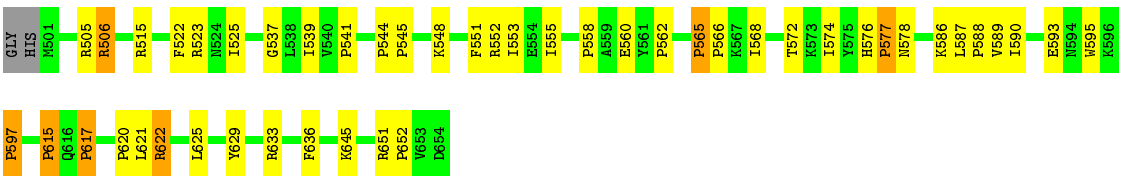
• Molecule 2: ubiquitin

Chain D:



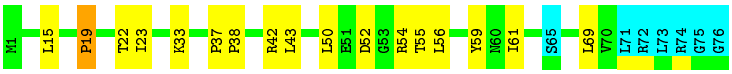
• Molecule 3: Ubiquitin-conjugating enzyme E2 L3

Chain C:



• Molecule 4: phosphoubiquitin

Chain A:



4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 8. Colouring as in section 4.1 above.

- Molecule 1: E3 ubiquitin-protein ligase parkin

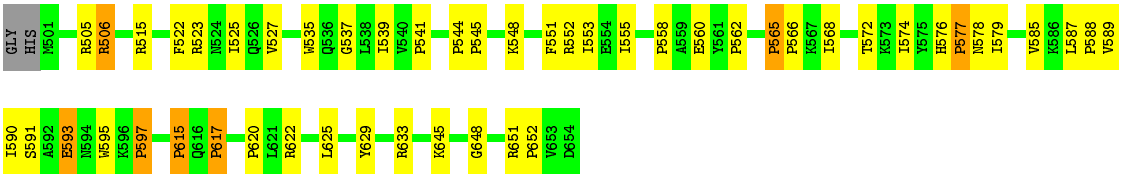
Chain B:

- Molecule 2: ubiquitin

Chain D:

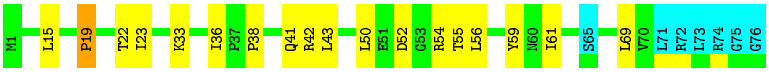
- Molecule 3: Ubiquitin-conjugating enzyme E2 L3

Chain C:



- Molecule 4: phosphoubiquitin

Chain A:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *simulated annealing.*

Of the 1000 calculated structures, 10 were deposited, based on the following criterion: *structures with the lowest energy.*

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
HADDOCK	structure calculation	
PyMOL	geometry optimization	2.0.0

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	398
Number of shifts mapped to atoms	398
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	5%

No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 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 each 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 averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	B	2140	491	2049	54±9
2	D	563	130	583	10±6
3	C	1261	290	1274	22±5
4	A	549	127	570	9±2
5	B	8	0	0	10±2
All	All	45210	10380	44758	900

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 10.

5 of 320 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:201:CYS:HG	5:B:504:ZN:ZN	0.95	0.69	7	9
1:B:250:VAL:HG22	1:B:258:VAL:HG12	0.92	1.40	7	10
1:B:436:CYS:HG	5:B:507:ZN:ZN	0.89	0.75	6	10
3:C:548:LYS:HB3	3:C:645:LYS:HE2	0.87	1.46	8	3
1:B:217:THR:HB	1:B:221:GLU:HG2	0.86	1.42	6	1

5.2 Torsion angles (i)

5.2.1 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 PDB entries followed by that with respect to all NMR entries. 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	B	269/322 (84%)	234±4 (87±2%)	31±4 (11±2%)	5±1 (2±0%)	13	56
2	D	70/76 (92%)	68±2 (97±3%)	2±2 (3±3%)	0±0 (0±0%)	100	100
3	C	152/156 (97%)	140±2 (92±1%)	11±2 (7±1%)	1±0 (0±0%)	38	78
4	A	68/76 (89%)	65±2 (96±2%)	3±2 (4±2%)	0±0 (0±0%)	54	85
All	All	5590/6300 (89%)	5073 (91%)	464 (8%)	53 (1%)	21	69

5 of 17 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	B	406	ALA	10
1	B	421	CYS	10
1	B	424	PRO	7
1	B	328	GLY	4
1	B	335	PRO	4

5.2.2 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 PDB entries followed by that with respect to all NMR

entries. 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	B	241/278 (87%)	220±1 (91±1%)	21±1 (9±1%)	14	61
2	D	65/68 (96%)	61±1 (94±1%)	4±1 (6±1%)	21	69
3	C	138/139 (99%)	123±1 (89±1%)	15±1 (11±1%)	10	55
4	A	63/67 (94%)	60±1 (95±1%)	3±1 (5±1%)	29	78
All	All	5070/5520 (92%)	4643 (92%)	427 (8%)	14	61

5 of 70 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
4	A	19	PRO	10
3	C	541	PRO	10
3	C	597	PRO	10
3	C	566	PRO	10
1	B	417	PRO	10

5.2.3 RNA (i)

There are no RNA molecules in this entry.

MODRES-GEOMETRY INFOmissingINFO

5.3 Carbohydrates (i)

There are no carbohydrates in this entry.

5.4 Ligand geometry (i)

Of 8 ligands modelled in this entry, 8 are monoatomic - leaving 0 for Mogul analysis.

5.5 Other polymers (i)

There are no such molecules in this entry.

5.6 Polymer linkage issues

There are no chain breaks in this entry.

6 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 5% for the well-defined parts and 5% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_0*

6.1.1 Bookkeeping

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	398
Number of shifts mapped to atoms	398
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

6.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	0	—	None (insufficient data)
$^{13}C_{\beta}$	0	—	None (insufficient data)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	205	0.55 ± 0.37	None needed (imprecise)

6.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 5%, i.e. 387 atoms were assigned a chemical shift out of a possible 7126. 0 out of 77 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	387/2749 (14%)	187/1092 (17%)	0/1130 (0%)	200/527 (38%)
Sidechain	0/3872 (0%)	0/2290 (0%)	0/1389 (0%)	0/193 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	0/505 (0%)	0/268 (0%)	0/210 (0%)	0/27 (0%)
Overall	387/7126 (5%)	187/3650 (5%)	0/2729 (0%)	200/747 (27%)

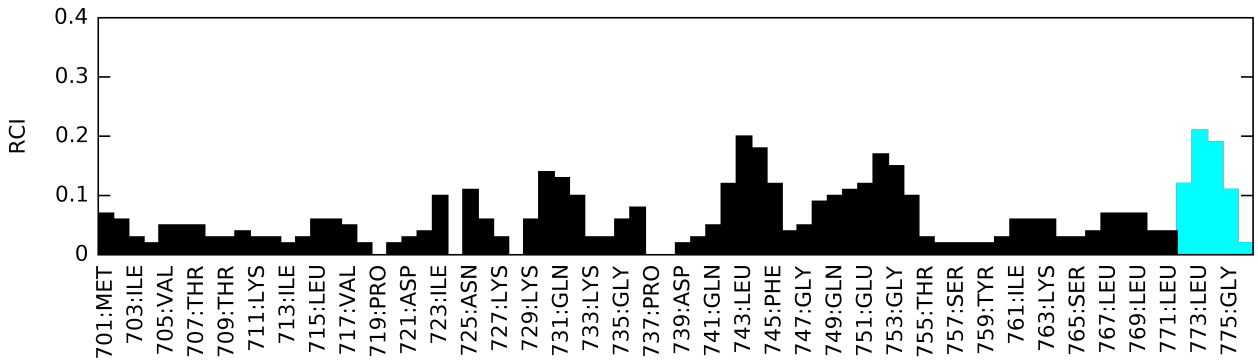
6.1.4 Statistically unusual chemical shifts

There are no statistically unusual chemical shifts.

6.1.5 Random Coil Index (RCI) plots

The images below report *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain D:



Random coil index (RCI) for chain C: