wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 01:04 AM BST

PDB ID	:	2LGG
Title	:	Structure of PHD domain of UHRF1 in complex with H3 peptide
Authors	:	Wang, C.; Shen, J.; Yang, Z.; Chen, P.; Zhao, B.; Hu, W.; Lan, W.; Tong, X.; Wu, H.; Li, G.; Cao, C.
Deposited on	:	2011-07-26

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
http://wwpdb.org/validation/2016/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:

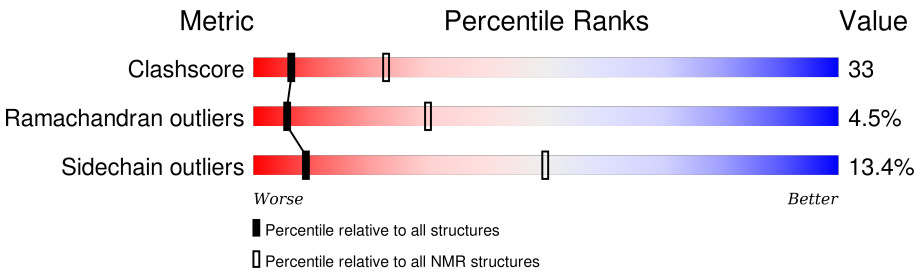
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 76%.

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	A	69	42% 42% 10% 6%
2	B	12	17% 67% 17%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 12 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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	A:315-A:379, B:383-B:392 (75)	1.00	12

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 4 clusters and 1 single-model cluster was found.

Cluster number	Models
1	4, 5, 6, 8, 11, 12, 13, 14, 15, 16
2	1, 2, 7, 9, 20
3	3, 19
4	10, 18
Single-model clusters	17

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1215 atoms, of which 583 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called E3 ubiquitin-protein ligase UHRF1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	69	1020	323	482	95	107	13	0

- Molecule 2 is a protein called histone H3 peptide.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
2	B	12	192	52	101	21	18	0

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

Mol	Chain	Residues	Atoms	
			Total	Zn
3	A	3	3	3

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 UHRF1

Chain A: 42% 42% 10% 6%

S311 G312 P313 S314 C315 C318 K319 D320 D321 V322 L325 C326 C329 A330 C331 H332 L333 C334 Q338 D339 P340 Q343 L344 M345 C346 C349 F353 H354 I355 Y356 C357 L358 D359 L362 S363 S364 D369 E370 W371 Y372 C373 P374 E375 C376 D379

- Molecule 2: histone H3 peptide

Chain B: 17% 67% 17%

A383 R384 T385 K386 Q387 T388 A389 S392 T393 G394

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

The representative model is number 12. Colouring as in section 4.1 above.

- Molecule 1: E3 ubiquitin-protein ligase UHRF1

Chain A: 43% 42% 9% 6%

S311 G312 P313 S314 C315 K316 H317 C318 K319 D320 D321 R324 L325 C326 R327 V328 C329 A330 C331 H332 L333 G336 R337 Q338 D339 P340 M345 C346 C349 F353 H354 I355 Y356 C357 L358 D359 P360 P361 L362 S363 D369 E370 W371 Y372 C373 P374 E375 C376 D379

- Molecule 2: histone H3 peptide

Chain B: 50% 25% 8% 17%

A383 R384 T385 K386 A389 S392 T393 G394

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 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
X-PLOR	structure solution	
X-PLOR	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	2lgg_cs.str
Number of chemical shift lists	1
Total number of shifts	801
Number of shifts mapped to atoms	801
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	76%

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

6 Model quality (i)

6.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

6.2 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 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	A	515	460	458	35±8
2	B	79	91	88	7±3
All	All	11940	11020	10920	745

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 33.

5 of 349 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:345:MET:SD	1:A:352:ALA:HB2	0.83	2.13	8	1
1:A:344:LEU:HD21	2:B:383:ALA:N	0.83	1.88	6	1
1:A:330:ALA:CA	2:B:386:LYS:HZ3	0.80	1.89	15	1
1:A:355:ILE:HD12	1:A:355:ILE:C	0.72	2.05	12	3
1:A:344:LEU:HD11	1:A:366:PRO:CD	0.71	2.15	1	1

6.3 Torsion angles (i)

6.3.1 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 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	A	64/69 (93%)	52±2 (82±3%)	9±2 (14±4%)	3±1 (4±2%)	6 31
2	B	9/12 (75%)	8±1 (84±8%)	1±1 (11±9%)	1±1 (6±7%)	4 23
All	All	1460/1620 (90%)	1195 (82%)	200 (14%)	65 (4%)	6 30

5 of 22 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	320	ASP	11
1	A	363	SER	10
1	A	361	PRO	6
1	A	339	ASP	5
1	A	370	GLU	4

6.3.2 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 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	A	61/64 (95%)	53±1 (86±2%)	9±1 (14±2%)	8 49
2	B	8/9 (89%)	7±1 (91±12%)	1±1 (9±12%)	15 60
All	All	1380/1460 (95%)	1195 (87%)	185 (13%)	9 50

5 of 30 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)
1	A	371	TRP	19
1	A	354	HIS	18

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	331	CYS	17
1	A	318	CYS	17
1	A	326	CYS	14

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

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

6.5 Carbohydrates (i)

There are no carbohydrates in this entry.

6.6 Ligand geometry (i)

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

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 76% for the well-defined parts and 76% for the entire structure.

7.1 Chemical shift list 1

File name: 2lgg_cs.str

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping (i)

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	801
Number of shifts mapped to atoms	801
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	5

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	67	-0.68 ± 0.27	Should be applied
$^{13}C_{\beta}$	65	0.36 ± 0.22	None needed (< 0.5 ppm)
$^{13}C'$	59	-0.24 ± 0.26	None needed (< 0.5 ppm)
^{15}N	62	-0.34 ± 0.39	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 76%, i.e. 703 atoms were assigned a chemical shift out of a possible 921. 7 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	321/365 (88%)	143/145 (99%)	119/150 (79%)	59/70 (84%)
Sidechain	353/496 (71%)	246/297 (83%)	103/171 (60%)	4/28 (14%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	29/60 (48%)	18/31 (58%)	11/23 (48%)	0/6 (0%)
Overall	703/921 (76%)	407/473 (86%)	233/344 (68%)	63/104 (61%)

7.1.4 Statistically unusual chemical shifts (i)

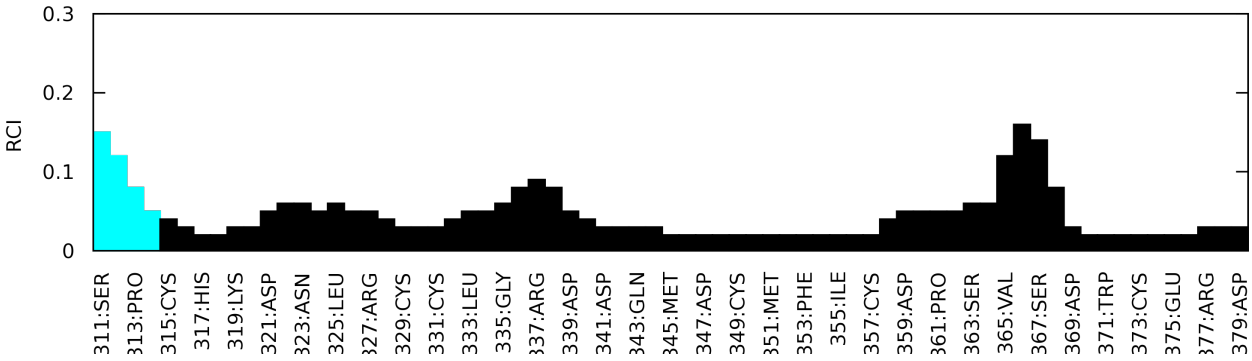
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	377	ARG	NH2	114.33	91.01 – 55.81	11.6
1	A	377	ARG	HH22	9.18	8.76 – 4.76	6.1
1	A	377	ARG	HH21	9.18	9.06 – 4.56	5.3
1	A	377	ARG	HH12	9.04	8.94 – 4.74	5.2
1	A	354	HIS	CD2	137.81	137.40 – 103.40	5.1

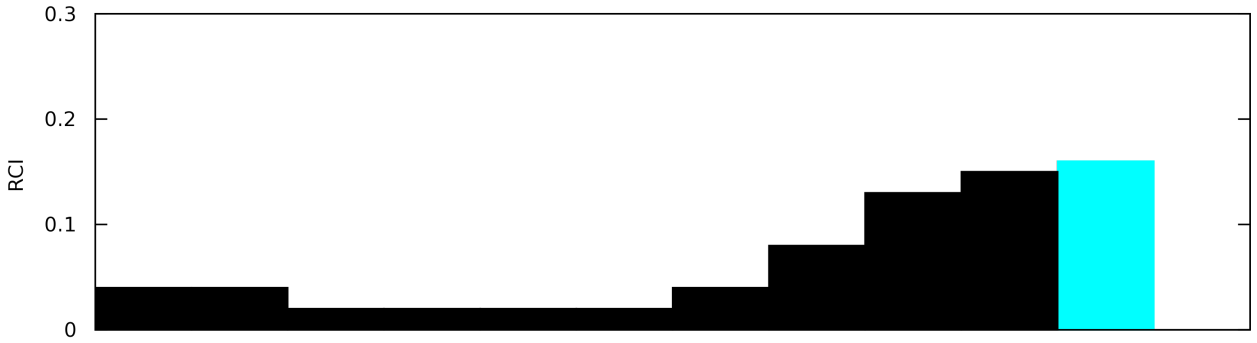
7.1.5 Random Coil Index (RCI) plots (i)

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 A:



Random coil index (RCI) for chain B:


0.3
0.2
0.1
0
RCI