Full wwPDB NMR Structure Validation Report (i)

Apr 27, 2016 – 05:24 AM BST

PDB ID	:	2RR9
Title	:	The solution structure of the K63-Ub2:tUIMs complex
Authors	:	Sekiyama, N.; Jee, J.; Isogai, S.; Akagi, K.; Huang, T.; Ariyoshi, M.; Tochio, H.; Shirakawa, M.
Deposited on	:	2010-06-16

This is a Full wwPDB NMR Structure Validation 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:

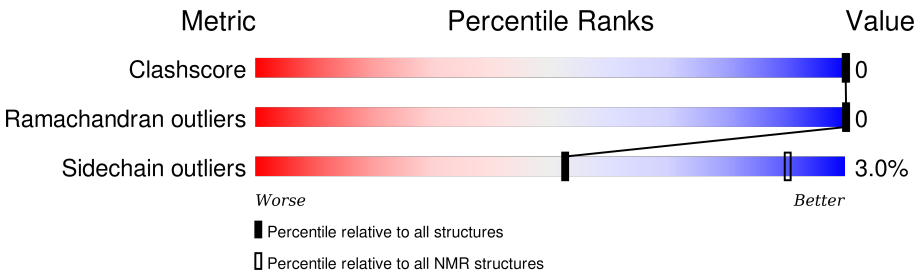
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 was not calculated.

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	76	93% 7%
1	B	76	93% 7%
2	C	46	80% 7% 13%

2 Ensemble composition and analysis (i)

This entry contains 20 models. The atoms present in the NMR models are not consistent. Some calculations may have failed as a result. All residues are included in the validation scores. 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:1-A:71, B:1-B:71, C:80-C:119 (182)	0.28	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 2 clusters and 2 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3171 atoms, of which 1599 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called ubiquitin.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	76	1227	378	626	105	117	1	0
1	B	76	1226	378	624	105	118	1	0

- Molecule 2 is a protein called Putative uncharacterized protein UIMC1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	C	46	718	220	349	62	84	3	0

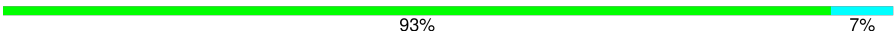
4 Residue-property plots (i)

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.

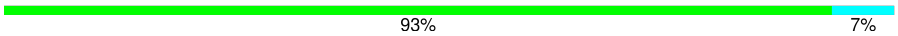
- Molecule 1: ubiquitin

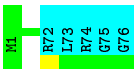
Chain A:



- Molecule 1: ubiquitin

Chain B:





- Molecule 2: Putative uncharacterized protein UIMC1

Chain C:

4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

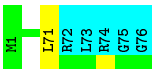
4.2.1 Score per residue for model 1

- Molecule 1: ubiquitin

Chain A:

- Molecule 1: ubiquitin

Chain B:



- Molecule 2: Putative uncharacterized protein UIMC1

Chain C:

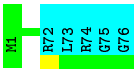
4.2.2 Score per residue for model 2

- Molecule 1: ubiquitin

Chain A:

- Molecule 1: ubiquitin

Chain B:



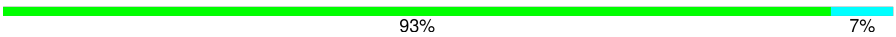
- Molecule 2: Putative uncharacterized protein UIMC1

Chain C:

4.2.3 Score per residue for model 3

- Molecule 1: ubiquitin

Chain A:



- Molecule 1: ubiquitin

Chain B: 93% 7%

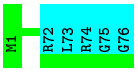
• Molecule 2: Putative uncharacterized protein UIMC1

Chain C: 78% 9% 13%

4.2.4 Score per residue for model 4

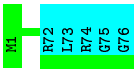
• Molecule 1: ubiquitin

Chain A: 93% 7%



• Molecule 1: ubiquitin

Chain B: 93% 7%



• Molecule 2: Putative uncharacterized protein UIMC1

Chain C: 74% 13% 13%

4.2.5 Score per residue for model 5

• Molecule 1: ubiquitin

Chain A: 93% 7%

• Molecule 1: ubiquitin

Chain B: 92% 7%

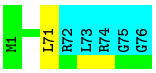
- Molecule 2: Putative uncharacterized protein UIMC1

Chain C:

4.2.6 Score per residue for model 6

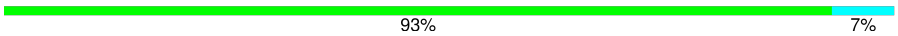
- Molecule 1: ubiquitin

Chain A:



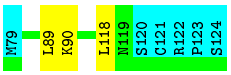
- Molecule 1: ubiquitin

Chain B:



- Molecule 2: Putative uncharacterized protein UIMC1

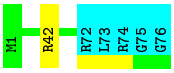
Chain C:



4.2.7 Score per residue for model 7

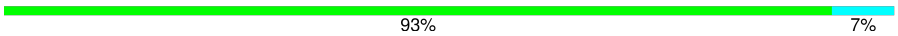
- Molecule 1: ubiquitin

Chain A:



- Molecule 1: ubiquitin

Chain B:



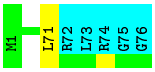
• Molecule 2: Putative uncharacterized protein UIMC1

Chain C:

4.2.8 Score per residue for model 8

• Molecule 1: ubiquitin

Chain A:



• Molecule 1: ubiquitin

Chain B:

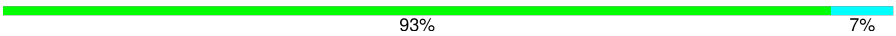
• Molecule 2: Putative uncharacterized protein UIMC1

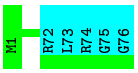
Chain C:

4.2.9 Score per residue for model 9

• Molecule 1: ubiquitin

Chain A:





• Molecule 1: ubiquitin

Chain B:

• Molecule 2: Putative uncharacterized protein UIMC1

Chain C: 80% 7% 13%

M79 K90 M91 S92 R111 S120 C121 R122 P123 S124

4.2.10 Score per residue for model 10

- Molecule 1: ubiquitin

Chain A: 93% 7%

M1 R72 L73 R74 G75 G76

- Molecule 1: ubiquitin

Chain B: 93% 7%

M1 R72 L73 R74 G75 G76

- Molecule 2: Putative uncharacterized protein UIMC1

Chain C: 78% 9% 13%

M79 Q84 K90 M91 S92 R111 S120 C121 R122 P123 S124

4.2.11 Score per residue for model 11

- Molecule 1: ubiquitin

Chain A: 92% • 7%

M1 L71 R72 L73 R74 G75 G76

- Molecule 1: ubiquitin

Chain B: 93% 7%

M1 R72 L73 R74 G75 G76

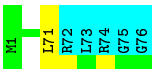
- Molecule 2: Putative uncharacterized protein UIMC1

Chain C: 78% 7% • 13%

4.2.12 Score per residue for model 12 (medoid)

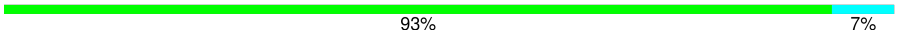
- Molecule 1: ubiquitin

Chain A:



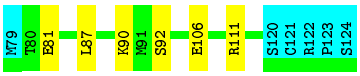
- Molecule 1: ubiquitin

Chain B:



- Molecule 2: Putative uncharacterized protein UIMC1

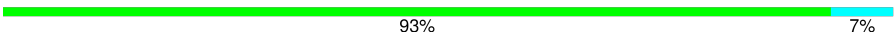
Chain C:



4.2.13 Score per residue for model 13

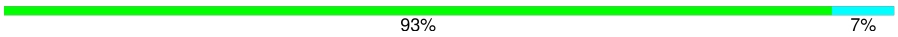
- Molecule 1: ubiquitin

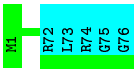
Chain A:



- Molecule 1: ubiquitin

Chain B:





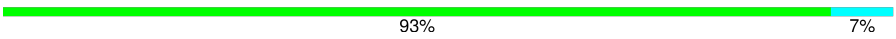
- Molecule 2: Putative uncharacterized protein UIMC1

Chain C:

4.2.14 Score per residue for model 14

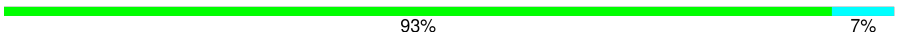
- Molecule 1: ubiquitin

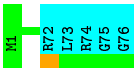
Chain A:



- Molecule 1: ubiquitin

Chain B:





- Molecule 2: Putative uncharacterized protein UIMC1

Chain C:

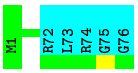
4.2.15 Score per residue for model 15

- Molecule 1: ubiquitin

Chain A:

- Molecule 1: ubiquitin

Chain B:



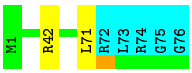
- Molecule 2: Putative uncharacterized protein UIMC1

Chain C:

4.2.16 Score per residue for model 16

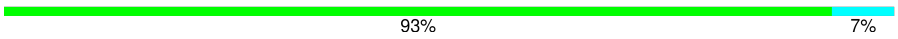
- Molecule 1: ubiquitin

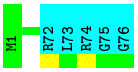
Chain A:



- Molecule 1: ubiquitin

Chain B:





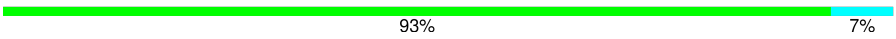
- Molecule 2: Putative uncharacterized protein UIMC1

Chain C:

4.2.17 Score per residue for model 17

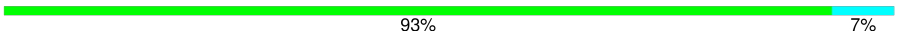
- Molecule 1: ubiquitin

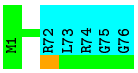
Chain A:



- Molecule 1: ubiquitin

Chain B:





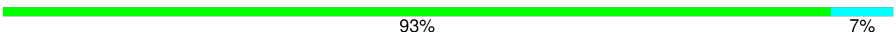
- Molecule 2: Putative uncharacterized protein UIMC1

Chain C:

4.2.18 Score per residue for model 18

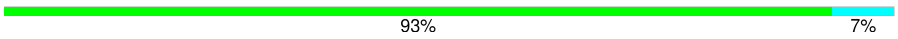
- Molecule 1: ubiquitin

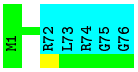
Chain A:



- Molecule 1: ubiquitin

Chain B:





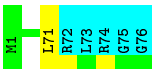
- Molecule 2: Putative uncharacterized protein UIMC1

Chain C:

4.2.19 Score per residue for model 19

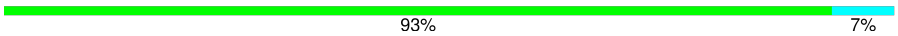
- Molecule 1: ubiquitin

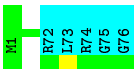
Chain A:



- Molecule 1: ubiquitin

Chain B:





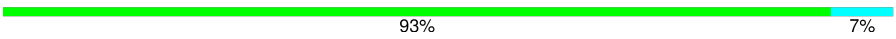
- Molecule 2: Putative uncharacterized protein UIMC1

Chain C:

4.2.20 Score per residue for model 20

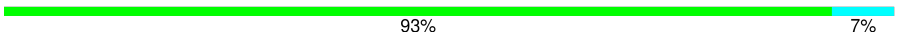
- Molecule 1: ubiquitin

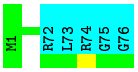
Chain A:



- Molecule 1: ubiquitin

Chain B:





- Molecule 2: Putative uncharacterized protein UIMC1

Chain C:

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *molecular dynamics*.

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
AMBER	refinement	
CYANA	structure solution	
CANDID	structure solution	

No chemical shift data was provided. 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)

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 (average) 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	0.65±0.00	0±0/570 (0.0±0.0%)	0.84±0.01	0±0/770 (0.0±0.0%)
1	B	0.65±0.00	0±0/570 (0.0±0.0%)	0.83±0.01	0±0/770 (0.0±0.0%)
2	C	0.65±0.01	0±0/325 (0.0±0.0%)	0.87±0.04	0±0/435 (0.0±0.1%)
All	All	0.65	0/29300 (0.0%)	0.85	7/39500 (0.0%)

There are no bond-length outliers.

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	C	111	ARG	NE-CZ-NH1	6.52	123.56	120.30	2	3
2	C	80	THR	CA-CB-CG2	-5.25	105.05	112.40	16	1
1	A	42	ARG	NE-CZ-NH1	5.14	122.87	120.30	16	3

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
All	All	29000	29502	29520	-

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 -.

There are no clashes.

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	70/76 (92%)	70±0 (100±0%)	0±0 (0±0%)	0±0 (0±0%)	100	100
1	B	70/76 (92%)	70±0 (100±0%)	0±0 (0±0%)	0±0 (0±0%)	100	100
2	C	40/46 (87%)	40±1 (99±1%)	0±1 (1±1%)	0±0 (0±0%)	100	100
All	All	3600/3960 (91%)	3592 (100%)	8 (0%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	65/68 (96%)	65±0 (99±1%)	0±0 (1±1%)	91	98
1	B	65/68 (96%)	65±0 (100±1%)	0±0 (0±1%)	95	98
2	C	35/41 (85%)	31±1 (88±3%)	4±1 (12±3%)	10	52
All	All	3300/3540 (93%)	3201 (97%)	99 (3%)	52	90

All 19 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)
2	C	90	LYS	18
2	C	111	ARG	16
2	C	92	SER	12
1	A	71	LEU	8
2	C	87	LEU	8
2	C	89	LEU	7
2	C	106	GLU	4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
2	C	118	LEU	4
2	C	110	LEU	3
1	B	71	LEU	3
2	C	84	GLN	3
2	C	93	GLU	3
2	C	85	PHE	2
2	C	98	GLU	2
2	C	116	GLU	2
2	C	81	GLU	1
2	C	80	THR	1
1	B	63	LYS	1
2	C	97	ARG	1

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)

There are no ligands in this entry.

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)

No chemical shift data were provided