



wwPDB NMR Structure Validation Summary Report ⓘ

Apr 27, 2016 – 02:46 AM BST

PDB ID : 2MBB
Title : Solution Structure of the human Polymerase iota UBM1-Ubiquitin Complex
Authors : Wang, S.; Zhou, P.
Deposited on : 2013-07-29

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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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

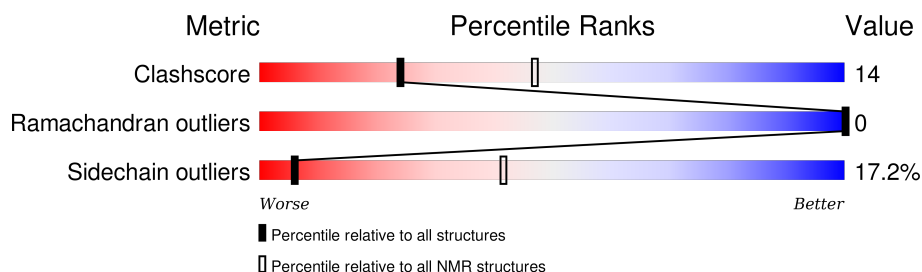
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 84%.

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	106	
2	B	78	

2 Ensemble composition and analysis

This entry contains 20 models. Model 6 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	A:66-A:90, B:201-B:272 (97)	0.12	6

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 3 single-model clusters were found.

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

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1835 atoms, of which 934 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Immunoglobulin G-binding protein G/DNA polymerase iota fusion protein.

Mol	Chain	Residues	Atoms						Trace
1	A	38	Total	C	H	N	O	S	0
			607	191	307	49	59	1	

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	EXPRESSION TAG	UNP P06654
A	2	GLN	-	EXPRESSION TAG	UNP P06654
A	57	GLY	-	LINKER	UNP P06654
A	58	SER	-	LINKER	UNP P06654
A	99	LEU	-	EXPRESSION TAG	UNP Q9UNA4
A	100	GLU	-	EXPRESSION TAG	UNP Q9UNA4
A	101	HIS	-	EXPRESSION TAG	UNP Q9UNA4
A	102	HIS	-	EXPRESSION TAG	UNP Q9UNA4
A	103	HIS	-	EXPRESSION TAG	UNP Q9UNA4
A	104	HIS	-	EXPRESSION TAG	UNP Q9UNA4
A	105	HIS	-	EXPRESSION TAG	UNP Q9UNA4
A	106	HIS	-	EXPRESSION TAG	UNP Q9UNA4

- Molecule 2 is a protein called Polyubiquitin-B.

Mol	Chain	Residues	Atoms						Trace
2	B	76	Total	C	H	N	O	S	0
			1228	378	627	105	117	1	

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	199	SER	-	EXPRESSION TAG	UNP P0CG47
B	200	HIS	-	EXPRESSION TAG	UNP P0CG47

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: Immunoglobulin G-binding protein G/DNA polymerase iota fusion protein

Chain A: 

Sequence: MET GLN TYR LYS LEU ILE LEU ASN GLY LYS THR LYS LEU LYS GLY GLU THR THR GLU ALA VAL ASP ASP ALA ALA THR ALA GLU LYS VAL PHE LYS GLN TYR ALA ASN ASP ASN GLY VAL ASP GLY TRP THR THR ASP ASP ALA THR LYS PHE THR VAL THR GLU GLY SER ASN GLU

Sequence: PHE P62 L63 C64 S65 L66 V70 V74 F75 F76 K77 L78 L82 L87 K90 S91 R92 E93 K94 F95 Q96 Q97 K98 L99 GLU HIS HIS HIS HIS HIS HIS HIS

- Molecule 2: Polyubiquitin-B

Chain B: 

Sequence: SER HIS R201 K206 K207 L208 T209 T212 T213 T214 V217 E218 T222 K225 K226 K227 K233 Q241 R242 L243 L244 F245 K248 Q249 L250 T255 L256 Y259 R260 L261 Q262 K263 K264 S265 R268 L269 V270 L271 L273 L274 G275 G276

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

The representative model is number 6. Colouring as in section 4.1 above.

- Molecule 1: Immunoglobulin G-binding protein G/DNA polymerase iota fusion protein

Chain A: 

Sequence: MET GLN TYR LYS LEU ILE LEU ASN GLY LYS THR LYS LEU LYS GLY GLU THR THR GLU ALA VAL ASP ASP ALA ALA THR ALA GLU LYS VAL PHE LYS GLN TYR ALA ASN ASP ASN GLY VAL ASP GLY TRP THR THR ASP ASP ALA THR LYS PHE THR VAL THR GLU GLY SER ASN GLU

Sequence: PHE P62 L63 C64 S65 L66 V70 D71 V74 F75 F76 K77 L78 L82 Q83 L87 K90 S91 R92 E93 K94 F95 Q96 Q97 K98 L99 GLU HIS HIS HIS HIS HIS HIS HIS

- Molecule 2: Polyubiquitin-B

Chain B: 

SER	HIS	P201	K206	T207	L208	T212	L213	V217	E218	T222	V225	V226	K227	K233	Q241	K242	L243	T244	F245	K248	Q249	L250	E251	K254	T255	L256	Y259	K260	L261	Q262	K263	E264	S265	T266	L267	H268	L269	V270	L271	K272	L273	K274	G275	G276
-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

5 Refinement protocol and experimental data overview

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

Of the 200 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
CYANA	structure solution	
CYANA	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)	2mbb_cs.str
Number of chemical shift lists	1
Total number of shifts	2017
Number of shifts mapped to atoms	2017
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	84%

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](#)

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	197	198	198	8±1
2	B	574	597	596	18±3
All	All	15420	15900	15880	445

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 14.

5 of 87 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:217:VAL:HG11	2:B:226:VAL:HG22	0.94	1.39	16	20
1:A:78:LEU:HD13	1:A:82:ILE:HG21	0.93	1.38	20	18
1:A:82:ILE:HG23	2:B:270:VAL:HG11	0.78	1.56	19	18
1:A:78:LEU:HD23	1:A:82:ILE:HG21	0.77	1.56	9	2
2:B:226:VAL:HG21	2:B:256:LEU:HD21	0.72	1.61	18	19

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	25/106 (24%)	25±0 (99±2%)	0±0 (1±2%)	0±0 (0±0%)	100	100
2	B	71/78 (91%)	69±1 (97±1%)	2±1 (3±1%)	0±0 (0±0%)	100	100
All	All	1920/3680 (52%)	1869 (97%)	51 (3%)	0 (0%)	100	100

There are no Ramachandran outliers.

6.3.2 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 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	23/92 (25%)	20±1 (89±6%)	3±1 (11±6%)	12	55
2	B	66/70 (94%)	53±2 (81±3%)	13±2 (19±3%)	5	37
All	All	1780/3240 (55%)	1473 (83%)	307 (17%)	6	42

5 of 36 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	B	208	LEU	20
2	B	265	SER	20
2	B	212	THR	20
2	B	206	LYS	20
2	B	271	LEU	20

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

The completeness of assignment taking into account all chemical shift lists is 84% for the well-defined parts and 81% for the entire structure.

7.1 Chemical shift list 1

File name: 2mbb_cs.str

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 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	2017
Number of shifts mapped to atoms	2017
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	4

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	175	0.56 ± 0.12	Should be applied
$^{13}\text{C}_\beta$	160	0.63 ± 0.17	Should be applied
$^{13}\text{C}'$	0	—	—
^{15}N	166	0.60 ± 0.34	None needed (imprecise)

7.1.3 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 84%, i.e. 1023 atoms were assigned a chemical shift out of a possible 1224. 18 out of 18 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	375/475 (79%)	188/189 (99%)	97/194 (50%)	90/92 (98%)
Sidechain	616/707 (87%)	384/411 (93%)	221/267 (83%)	11/29 (38%)

Continued on next page...

Continued from previous page...

	Total	¹ H	¹³ C	¹⁵ N
Aromatic	32/42 (76%)	21/23 (91%)	11/18 (61%)	0/1 (0%)
Overall	1023/1224 (84%)	593/623 (95%)	329/479 (69%)	101/122 (83%)

7.1.4 Statistically unusual chemical shifts ⓘ

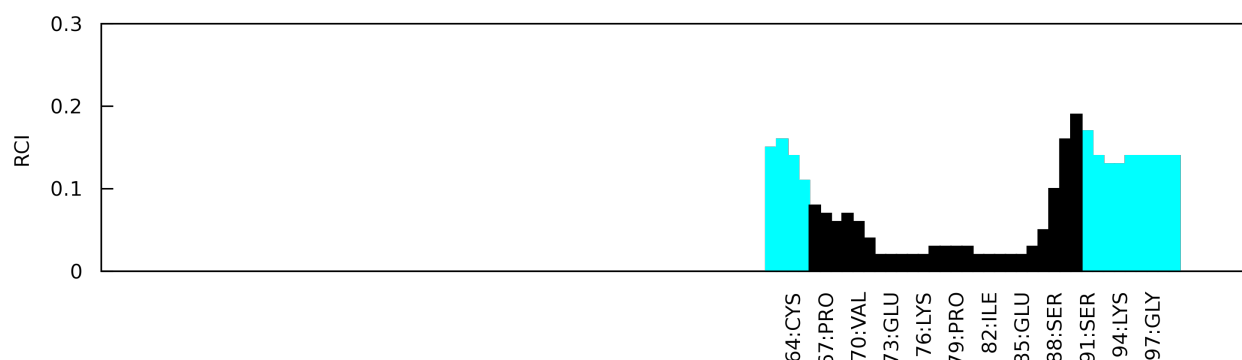
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	5	LEU	HB3	-1.16	3.34 – -0.26	-7.5
1	A	54	VAL	HB	-0.32	3.59 – 0.39	-7.2
1	A	31	LYS	HE2	1.64	3.87 – 1.97	-6.7
1	A	31	LYS	HE3	1.86	3.86 – 1.96	-5.5

7.1.5 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 A:



Random coil index (RCI) for chain B:

