wwPDB NMR Structure Validation Summary Report (i)

Apr 21, 2024 – 04:51 AM EDT

PDB ID	:	2M30
BMRB ID	:	7376
Title	:	Solution NMR refinement of a metal ion bound protein using quantum mechanical/molecular mechanical and molecular dynamics methods
Authors	:	Chakravorty, D.K.; Wang, B.I.; Lee, C.I.; Guerra, A.J.; Giedroc, D.P.; Merz Jr., K.M.; Arunkumar, A.I.; Pennella, M.; Kong, X.
Deposited on	:	2013-01-04

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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

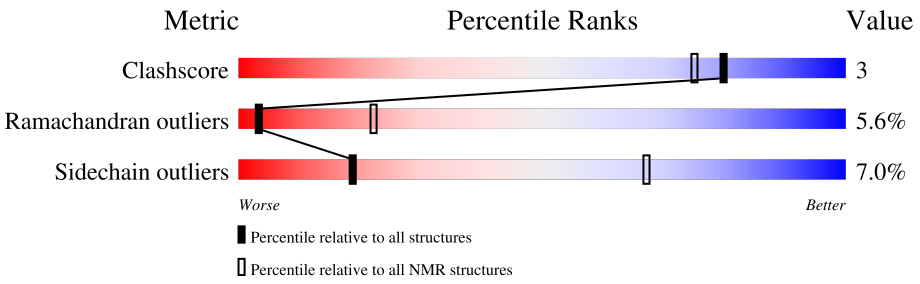
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 26%.

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	A	106	62% 17% • 7% 10%
1	B	106	61% 16% • 8% 10%

2 Ensemble composition and analysis (i)

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: *fewest violations*.

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:10-A:73, A:78-A:101, B:10-B:72, B:78-B:100 (174)	0.52	8

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 1 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 5, 6, 7, 8, 9, 10

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3058 atoms, of which 1546 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Repressor protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	95	1528	471	773	141	140	3	0
1	B	95	1528	471	773	141	140	3	0

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	1	1	1
2	B	1	1	1

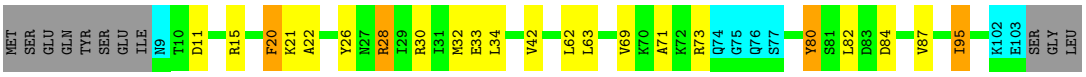
4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide 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.

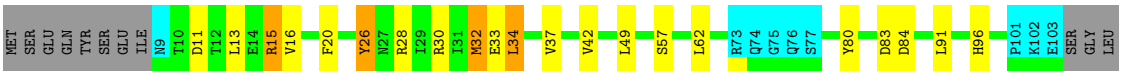
• Molecule 1: Repressor protein

Chain A: 62% 17% • 7% 10%



• Molecule 1: Repressor protein

Chain B: 61% 16% • 8% 10%

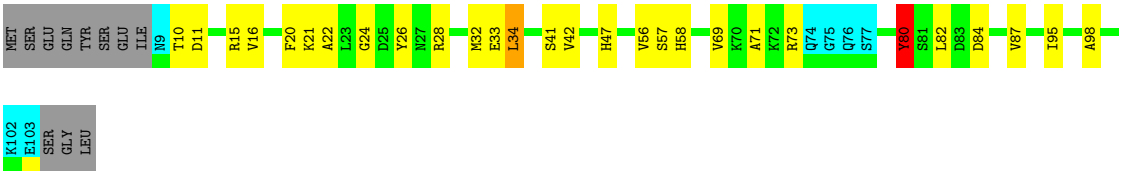


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

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

• Molecule 1: Repressor protein

Chain A: 57% 25% •• 7% 10%



• Molecule 1: Repressor protein

Chain B: 62% 14% 5% 8% 10%

5 Refinement protocol and experimental data overview (i)

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

Of the 132000 calculated structures, 10 were deposited, based on the following criterion: *Structures from 1 ns of independent QM/MM MD sampling*.

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

Software name	**Classification**	**Version**
Amber	structure solution	11
Amber	refinement	11
Amber	geometry optimization	11

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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	730
Number of shifts mapped to atoms	671
Number of unparsed shifts	0
Number of shifts with mapping errors	59
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	26%

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

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	1.46±0.04	2±1/713 (0.3± 0.2%)	1.97±0.08	18±4/963 (1.9± 0.4%)
1	B	1.42±0.05	2±1/694 (0.2± 0.1%)	1.88±0.10	15±3/938 (1.6± 0.4%)
All	All	1.44	40/14070 (0.3%)	1.93	335/19010 (1.8%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	A	0.0±0.0	4.0±2.1
1	B	0.0±0.0	3.6±1.3
All	All	0	76

5 of 37 unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	30	ARG	CZ-NH1	-9.20	1.21	1.33	3	1
1	B	46	SER	CB-OG	-8.41	1.31	1.42	6	1
1	A	46	SER	CB-OG	-8.02	1.31	1.42	7	1
1	A	38	SER	CB-OG	7.44	1.51	1.42	1	1
1	A	15	ARG	CZ-NH1	-7.28	1.23	1.33	5	1

5 of 168 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
1	A	15	ARG	NE-CZ-NH1	18.35	129.48	120.30	9	10

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	B	30	ARG	NE-CZ-NH1	18.30	129.45	120.30	2	7
1	A	30	ARG	NE-CZ-NH1	16.54	128.57	120.30	10	9
1	B	15	ARG	NE-CZ-NH1	16.39	128.50	120.30	2	6
1	B	30	ARG	NE-CZ-NH2	-15.89	112.35	120.30	2	5

There are no chirality outliers.

5 of 30 unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	32	MET	Peptide	9
1	B	32	MET	Peptide	9
1	A	80	TYR	Sidechain	7
1	A	20	PHE	Peptide,Sidechain	6
1	A	30	ARG	Sidechain	3

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	701	724	724	4±2
1	B	683	704	704	5±2
All	All	13860	14280	14280	76

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 3.

5 of 47 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:45:ILE:HG23	1:B:49:LEU:HD13	0.68	1.66	4	2
1:A:71:ALA:HB2	1:A:80:TYR:CE2	0.67	2.25	8	5
1:B:34:LEU:C	1:B:34:LEU:HD13	0.57	2.20	4	4
1:A:13:LEU:O	1:A:17:THR:HG23	0.56	2.01	4	1
1:A:34:LEU:C	1:A:34:LEU:HD13	0.55	2.21	8	2

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	88/106 (83%)	70±2 (80±3%)	11±3 (13±3%)	6±2 (7±2%)	2 16
1	B	86/106 (81%)	74±3 (86±3%)	8±3 (10±3%)	3±1 (4±1%)	5 31
All	All	1740/2120 (82%)	1443 (83%)	199 (11%)	98 (6%)	3 22

5 of 19 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	42	VAL	10
1	B	42	VAL	10
1	A	22	ALA	9
1	A	28	ARG	8
1	A	33	GLU	7

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	80/96 (83%)	74±2 (93±3%)	6±2 (7±3%)	19 68
1	B	78/96 (81%)	72±1 (93±1%)	6±1 (7±1%)	18 67
All	All	1580/1920 (82%)	1470 (93%)	110 (7%)	19 67

5 of 43 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	B	13	LEU	10
1	A	95	ILE	9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	34	LEU	7
1	B	26	TYR	7
1	A	87	VAL	6

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 monosaccharides in this entry.

6.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 2 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 26% for the well-defined parts and 25% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

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	730
Number of shifts mapped to atoms	671
Number of unparsed shifts	0
Number of shifts with mapping errors	59
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. First 5 (of 59) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	2	ALA	CA	51.528	0.400	1
1	A	2	ALA	CB	18.91	0.400	1
1	A	3	GLU	H	8.66	0.020	1
1	A	3	GLU	C	176.132	0.400	1
1	A	3	GLU	CA	59.126	0.400	1
1	A	3	GLU	CB	28.842	0.400	1
1	A	3	GLU	N	120.962	0.400	1
1	A	4	GLN	H	8.442	0.020	1
1	A	4	GLN	HE21	7.414	0.020	2
1	A	4	GLN	HE22	6.798	0.020	2
1	A	4	GLN	C	175.591	0.400	1
1	A	4	GLN	CA	55.755	0.400	1
1	A	4	GLN	CB	29.475	0.400	1
1	A	4	GLN	N	122.004	0.400	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	4	GLN	NE2	111.932	0.400	1
1	A	5	TYR	H	8.305	0.020	1
1	A	5	TYR	CA	57.668	0.400	1
1	A	5	TYR	CB	38.067	0.400	1
1	A	5	TYR	N	121.338	0.400	1
1	A	6	SER	H	7.988	0.020	1
1	A	6	SER	C	173.873	0.400	1
1	A	6	SER	CA	57.59	0.400	1
1	A	6	SER	CB	63.618	0.400	1
1	A	6	SER	N	118.236	0.400	1
1	A	7	GLU	H	8.286	0.020	1
1	A	7	GLU	C	175.802	0.400	1
1	A	7	GLU	CA	56.31	0.400	1
1	A	7	GLU	CB	29.88	0.400	1
1	A	7	GLU	N	123.02	0.400	1
1	A	8	ILE	H	8.101	0.020	1
1	A	8	ILE	HD11	0.772	0.020	1
1	A	8	ILE	HD12	0.772	0.020	1
1	A	8	ILE	HD13	0.772	0.020	1
1	A	8	ILE	C	175.128	0.400	1
1	A	8	ILE	CA	60.246	0.400	1
1	A	8	ILE	CB	37.568	0.400	1
1	A	8	ILE	CD1	12.321	0.400	1
1	A	8	ILE	N	121.13	0.400	1
1	A	104	SER	H	8.367	0.020	1
1	A	104	SER	C	174.942	0.400	1
1	A	104	SER	CA	58.258	0.400	1
1	A	104	SER	CB	63.548	0.400	1
1	A	104	SER	N	116.875	0.400	1
1	A	105	GLY	H	8.381	0.020	1
1	A	105	GLY	C	173.181	0.400	1
1	A	105	GLY	CA	45.047	0.400	1
1	A	105	GLY	N	111.32	0.400	1
1	A	106	LEU	H	7.723	0.020	1
1	A	106	LEU	HD11	0.838	0.020	2
1	A	106	LEU	HD12	0.838	0.020	2
1	A	106	LEU	HD13	0.838	0.020	2
1	A	106	LEU	HD21	0.875	0.020	2
1	A	106	LEU	HD22	0.875	0.020	2
1	A	106	LEU	HD23	0.875	0.020	2
1	A	106	LEU	CA	56.233	0.400	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	106	LEU	CB	43.101	0.400	1
1	A	106	LEU	CD1	23.405	0.400	1
1	A	106	LEU	CD2	25.12	0.400	1
1	A	106	LEU	N	127.393	0.400	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	105	-0.38 ± 0.24	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	96	0.49 ± 0.28	None needed (< 0.5 ppm)
$^{13}C'$	100	-0.59 ± 0.19	Should be applied
^{15}N	103	0.59 ± 0.43	None needed (imprecise)

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 26%, i.e. 631 atoms were assigned a chemical shift out of a possible 2465. 0 out of 38 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

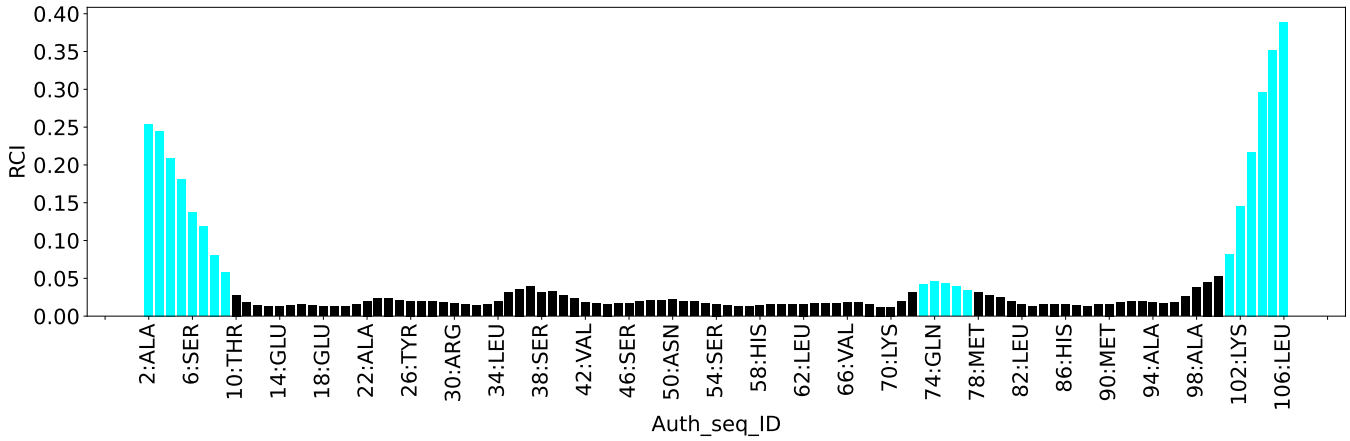
	Total	^{1}H	^{13}C	^{15}N
Backbone	348/872 (40%)	87/351 (25%)	174/348 (50%)	87/173 (50%)
Sidechain	283/1425 (20%)	150/937 (16%)	127/439 (29%)	6/49 (12%)
Aromatic	0/168 (0%)	0/90 (0%)	0/62 (0%)	0/16 (0%)
Overall	631/2465 (26%)	237/1378 (17%)	301/849 (35%)	93/238 (39%)

7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots (i)

The image below reports *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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A: