Full wwPDB NMR Structure Validation Report (i)

Feb 12, 2017 – 05:30 pm GMT

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:

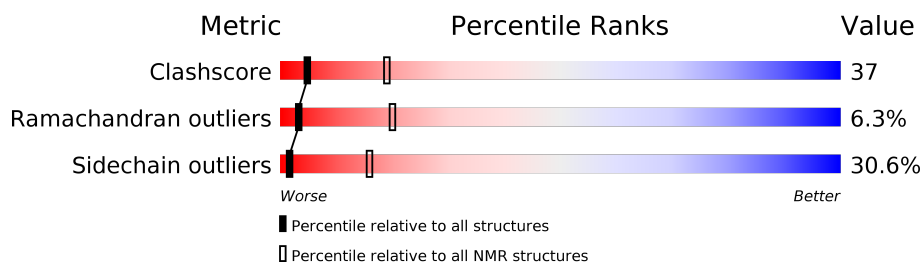|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	trunk28760
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

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 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	Percentile Ranks	Value
Clashscore		37
Ramachandran outliers		6.3%
Sidechain outliers		30.6%

Worse   *Better*

▌ Percentile relative to all structures
⧠ Percentile relative to all NMR structures

Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	125131	11601
Ramachandran outliers	121729	10391
Sidechain outliers	121581	10367

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	53	19% 40% 6% • 9% 25%

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

2 Ensemble composition and analysis ⓘ

This entry contains 10 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 8 is the overall representative, medoid model (most similar to other models).

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:4-A:38 (35)	0.34	8

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

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

NmrClust was unable to cluster the ensemble.

Error message: Inconsistent models in file

3 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 530 atoms, of which 237 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called CU-METALLOTHIONEIN.

Mol	Chain	Residues	Atoms						Trace
1	A	40	Total	C	H	N	O	S	0
			523	156	237	55	65	10	

- Molecule 2 is COPPER (I) ION (three-letter code: CU1) (formula: Cu).

Mol	Chain	Residues	Atoms	
2	A	7	Total	Cu
			7	7

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.

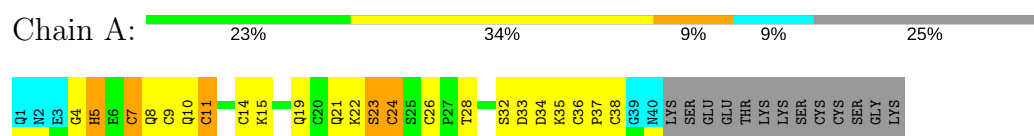• Molecule 1: CU-METALLOTHIONEIN
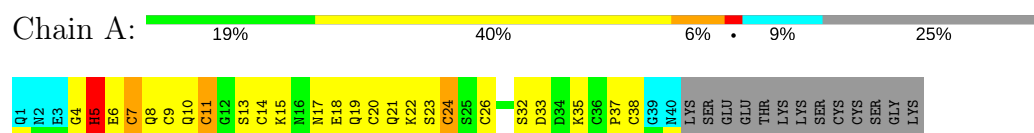
4.2 Scores per residue for each member of the ensemble
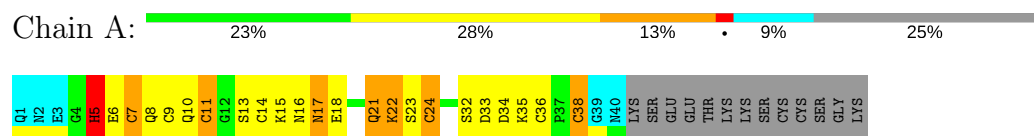
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1
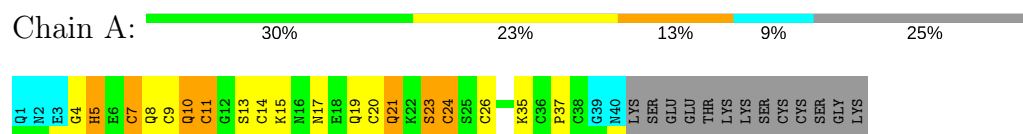
• Molecule 1: CU-METALLOTHIONEIN

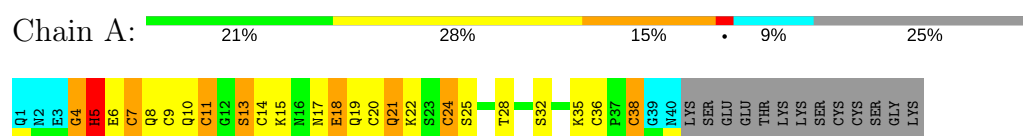### 4.2.2 Score per residue for model 2

• Molecule 1: CU-METALLOTHIONEIN

4.2.3 Score per residue for model 3

• Molecule 1: CU-METALLOTHIONEIN

Chain A:

4.2.4 Score per residue for model 4

• Molecule 1: CU-METALLOTHIONEIN

Chain A:

4.2.5 Score per residue for model 5

• Molecule 1: CU-METALLOTHIONEIN
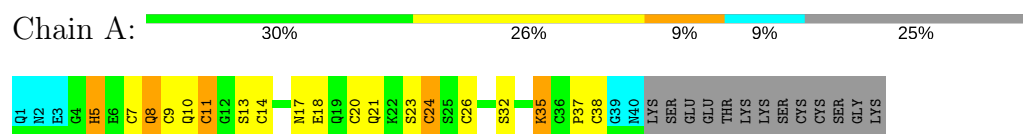
Chain A:

4.2.6 Score per residue for model 6

• Molecule 1: CU-METALLOTHIONEIN

Chain A:

4.2.7 Score per residue for model 7

• Molecule 1: CU-METALLOTHIONEIN

Chain A:

4.2.8 Score per residue for model 8 (medoid)

• Molecule 1: CU-METALLOTHIONEIN
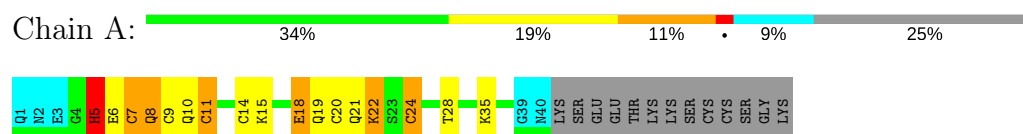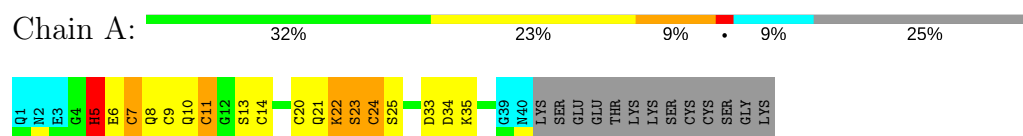
Chain A:
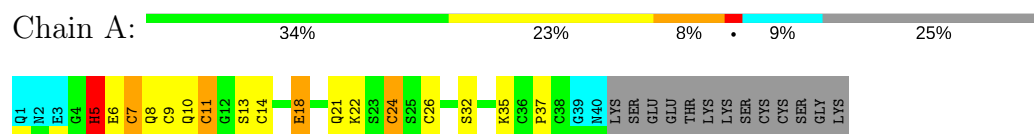
4.2.9 Score per residue for model 9

• Molecule 1: CU-METALLOTHIONEIN

Chain A:
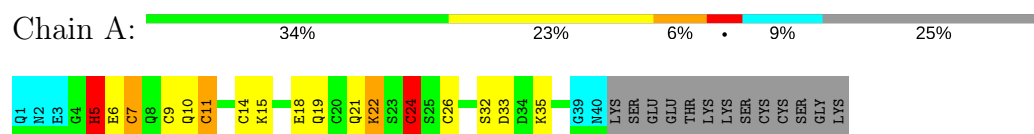
4.2.10 Score per residue for model 10

• Molecule 1: CU-METALLOTHIONEIN

Chain A:
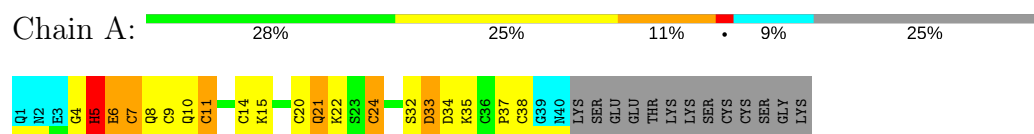
5 Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *DISTANCE GEOMETRY*.

Of the 30 calculated structures, 10 were deposited, based on the following criterion: *LEAST RESTRAINT VIOLATION*.

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

Software name	Classification	Version
X-PLOR	refinement	3.1
X-PLOR	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 ⓘ

6.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: CU1

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	1.01±0.01	0±0/251 (0.0±0.0%)	1.31±0.01	0±0/336 (0.0±0.1%)
All	All	1.01	0/2510 (0.0%)	1.31	1/3360 (0.0%)

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	24	CYS	N-CA-CB	-5.00	101.59	110.60	9	1

There are no chirality outliers.

There are no planarity outliers.

6.2 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	A	248	206	206	17±2
All	All	2550	2060	2060	170

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 37.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
1:A:5:HIS:CE1	1:A:24:CYS:O	0.61	2.54	10	6
1:A:7:CYS:HB2	1:A:24:CYS:HB3	0.61	1.73	9	10
1:A:7:CYS:HB2	1:A:24:CYS:CB	0.60	2.26	9	8
1:A:8:GLN:NE2	1:A:9:CYS:N	0.58	2.51	5	2
1:A:5:HIS:NE2	1:A:24:CYS:O	0.57	2.38	4	6
1:A:7:CYS:SG	1:A:7:CYS:O	0.56	2.64	10	5
1:A:28:THR:HG23	1:A:28:THR:O	0.56	2.00	1	2
1:A:9:CYS:CB	1:A:14:CYS:SG	0.55	2.94	9	10
1:A:5:HIS:CE1	1:A:26:CYS:O	0.55	2.60	9	4
1:A:7:CYS:O	1:A:7:CYS:SG	0.55	2.65	1	5
1:A:5:HIS:CE1	1:A:24:CYS:SG	0.55	2.99	3	3
1:A:7:CYS:SG	1:A:21:GLN:NE2	0.54	2.80	4	3
1:A:7:CYS:HB3	1:A:21:GLN:HA	0.53	1.81	10	9
1:A:10:GLN:O	1:A:11:CYS:O	0.52	2.28	4	10
1:A:6:GLU:CG	1:A:6:GLU:O	0.52	2.58	6	1
1:A:34:ASP:OD1	1:A:35:LYS:N	0.51	2.43	1	1
1:A:28:THR:O	1:A:28:THR:HG23	0.51	2.06	6	1
1:A:18:GLU:HA	1:A:21:GLN:CG	0.50	2.37	2	4
1:A:9:CYS:HB3	1:A:14:CYS:SG	0.49	2.47	1	10
1:A:34:ASP:OD1	1:A:36:CYS:N	0.49	2.46	1	1
1:A:8:GLN:NE2	1:A:8:GLN:HA	0.49	2.23	10	2
1:A:13:SER:O	1:A:17:ASN:CB	0.48	2.60	3	1
1:A:5:HIS:O	1:A:22:LYS:O	0.47	2.32	10	3
1:A:8:GLN:NE2	1:A:8:GLN:C	0.47	2.68	6	2
1:A:33:ASP:N	1:A:33:ASP:OD1	0.47	2.47	10	1
1:A:8:GLN:HA	1:A:8:GLN:NE2	0.47	2.25	1	2
1:A:21:GLN:NE2	1:A:22:LYS:HB2	0.45	2.26	9	1
1:A:9:CYS:SG	1:A:26:CYS:SG	0.45	3.14	3	1
1:A:8:GLN:CD	1:A:8:GLN:C	0.45	2.76	6	1
1:A:7:CYS:SG	1:A:9:CYS:HB2	0.45	2.52	4	8
1:A:36:CYS:SG	1:A:38:CYS:HB2	0.45	2.52	2	2
1:A:4:GLY:C	1:A:5:HIS:CD2	0.45	2.91	4	1
1:A:33:ASP:OD1	1:A:33:ASP:O	0.44	2.35	9	2
1:A:5:HIS:CD2	1:A:5:HIS:N	0.44	2.83	3	2
1:A:24:CYS:HB2	1:A:38:CYS:SG	0.44	2.52	5	1
1:A:5:HIS:O	1:A:22:LYS:C	0.44	2.56	6	4
1:A:7:CYS:C	1:A:9:CYS:H	0.44	2.16	5	5
1:A:4:GLY:CA	1:A:23:SER:HA	0.43	2.43	3	2
1:A:20:CYS:SG	1:A:21:GLN:OE1	0.43	2.76	5	1
1:A:34:ASP:C	1:A:36:CYS:H	0.43	2.17	2	2
1:A:5:HIS:N	1:A:23:SER:HA	0.42	2.29	7	1
1:A:34:ASP:O	1:A:35:LYS:CG	0.42	2.67	10	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
1:A:18:GLU:HA	1:A:21:GLN:HG3	0.42	1.91	8	1
1:A:18:GLU:HA	1:A:21:GLN:HG2	0.42	1.91	5	1
1:A:6:GLU:O	1:A:6:GLU:CG	0.42	2.67	10	1
1:A:20:CYS:O	1:A:24:CYS:HB3	0.42	2.13	4	4
1:A:18:GLU:CG	1:A:21:GLN:HG3	0.42	2.44	8	1
1:A:15:LYS:CG	1:A:16:ASN:N	0.42	2.83	2	1
1:A:5:HIS:O	1:A:22:LYS:HA	0.42	2.15	2	1
1:A:35:LYS:O	1:A:37:PRO:HD3	0.42	2.14	10	5
1:A:7:CYS:HB3	1:A:20:CYS:O	0.41	2.14	6	1
1:A:33:ASP:CG	1:A:33:ASP:O	0.41	2.57	1	1
1:A:18:GLU:HA	1:A:21:GLN:OE1	0.41	2.16	4	1
1:A:21:GLN:HG3	1:A:21:GLN:O	0.41	2.16	10	1
1:A:4:GLY:HA3	1:A:24:CYS:O	0.41	2.16	10	1
1:A:13:SER:O	1:A:17:ASN:HB2	0.41	2.16	4	1
1:A:21:GLN:CD	1:A:21:GLN:C	0.41	2.80	9	1
1:A:6:GLU:HG3	1:A:6:GLU:O	0.40	2.16	6	1

6.3 Torsion angles ⓘ

6.3.1 Protein backbone ⓘ

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	35/53 (66%)	24±1 (67±4%)	9±2 (26±4%)	2±1 (6±2%)	3	20
All	All	350/530 (66%)	236 (67%)	92 (26%)	22 (6%)	3	20

All 5 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	11	CYS	10
1	A	5	HIS	8
1	A	38	CYS	2
1	A	4	GLY	1
1	A	17	ASN	1

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	32/48 (67%)	22±2 (69±7%)	10±2 (31±7%)	2	16
All	All	320/480 (67%)	222 (69%)	98 (31%)	2	16

All 20 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	24	CYS	10
1	A	5	HIS	9
1	A	7	CYS	9
1	A	22	LYS	7
1	A	32	SER	7
1	A	6	GLU	6
1	A	35	LYS	6
1	A	15	LYS	6
1	A	23	SER	5
1	A	19	GLN	5
1	A	8	GLN	5
1	A	13	SER	5
1	A	18	GLU	4
1	A	21	GLN	4
1	A	17	ASN	2
1	A	38	CYS	2
1	A	33	ASP	2
1	A	25	SER	2
1	A	10	GLN	1
1	A	34	ASP	1

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 ⓘ

There are no carbohydrates in this entry.

6.6 Ligand geometry ⓘ

Of 7 ligands modelled in this entry, 7 are monoatomic - leaving 0 for Mogul analysis.

6.7 Other polymers ⓘ

There are no such molecules in this entry.

6.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

7 Chemical shift validation (i)

No chemical shift data were provided