wwPDB NMR Structure Validation Summary Report (i)

Feb 22, 2022 – 11:45 AM EST

PDB ID	:	1WSX
Title	:	Solution structure of MCL-1
Authors	:	Day, C.L.; Chen, L.; Richardson, S.J.; Harrison, P.J.; Huang, D.C.; Hinds, M.G.
Deposited on	:	2004-11-12

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

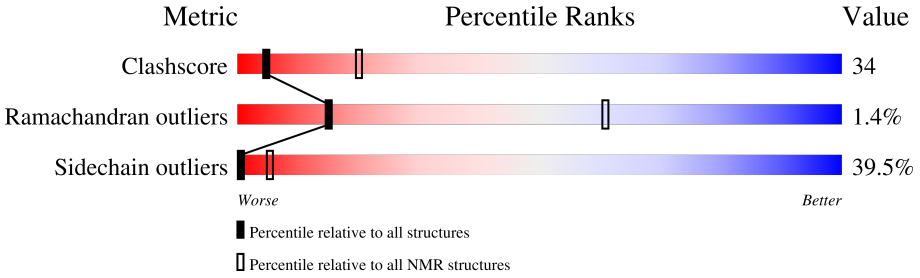
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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	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	162	20% 47% 12% 21%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 18 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:155-A:172, A:185-A:234, A:241-A:300 (128)	0.19	18

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. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2577 atoms, of which 1289 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called myeloid cell leukemia sequence 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	162	2577	810	1289	236	240	2	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	147	GLY	-	cloning artifact	UNP P97287
A	148	PRO	-	cloning artifact	UNP P97287
A	149	LEU	-	cloning artifact	UNP P97287
A	150	GLY	-	cloning artifact	UNP P97287
A	151	SER	-	cloning artifact	UNP P97287

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: myeloid cell leukemia sequence 1

Chain A: 20% 47% 12% 21%

G147 P148 L149 G150 S151 E152 D153 D154 L155 Y156 R157 Q158 S159 L160 E161 I162 I163 S164 R165 Y166 L167 R168 E169 G173 S174 K175 D176 S177 K178 P179 L180 G181 E182 A183 G184 A185 A186 G187 R188 R189 A190 L191 E192 T193 L194 R195 R196 V197 G198 D199 G200 V201 Q202 R203 N204 H205 E206 T207 A208

F209 Q210 G211 M212 L213 R214 K215 L216 K219 N220 D223 V224 K225 S226 F227 S228 R229 V230 M231 V232 H233 V234 F235 K236 D237 G238 V239 T240 N241 W242 I245 V246 T247 L248 I249 S250 F251 G252 A253 F254 V255 A256 K257 H258 L259 K260 S261 V262 N263 Q264 E265 S266 F267 I268 E269 P270 L271

T274 I275 T276 D277 V278 L279 V280 R281 T282 K283 R284 D285 W286 L287 V288 K289 Q290 R291 G292 W293 D294 E298 F299 F300 H301 V302 Q303 D304 L305 E306 G307 G308

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

The representative model is number 18. Colouring as in section 4.1 above.

- Molecule 1: myeloid cell leukemia sequence 1

Chain A: 24% 43% 12% 21%

G147 P148 L149 G150 S151 E152 D153 D154 L155 Y156 R157 Q158 S159 L160 E161 I162 I163 S164 R165 Y166 L167 R168 T172 G173 S174 K175 D176 S177 K178 P179 L180 G181 E182 A183 G184 A185 A186 G187 R188 R189 A190 L191 E192 T193 L194 R195 R196 V197 V201 Q202 R203 N204 H205 E206 T207 A208 F209 Q210

G211 M212 L213 R214 K215 L216 N220 D223 V224 K225 S226 F227 S228 R229 V230 M231 V232 H233 V234 F235 K236 D237 G238 V239 T240 N241 W242 G243 R244 I245 V246 T247 L248 I249 S250 F251 G252 A253 F254 V255 H258 L259 K260 S261 V262 N263 Q264 E265 S266 F267 I268 E269 P270 L271 A272 E273 T274

I275 T276 D277 V278 L279 V280 R281 T282 K283 R284 D285 W286 L287 V288 K289 Q290 R291 G292 W293 D294 E298 H301 V302 Q303 D304 L305 E306 G307 G308

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *torsion angle dynamics, distance geometry, simulated annealing.*

Of the 256 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy, structures with the least restraint violations, structures with acceptable covalent geometry.*

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

Software name	Classification	Version
DYANA	structure solution	1.5
XPLOR-NIH	refinement	2

No chemical shift data was provided.

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	1049	1069	1066	71±5
All	All	20980	21380	21320	1426

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 34.

5 of 290 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:163:ILE:HD13	1:A:276:THR:HG23	0.89	1.45	18	18
1:A:212:MET:O	1:A:216:LEU:HD13	0.87	1.69	18	6
1:A:248:LEU:HD12	1:A:279:LEU:HD23	0.85	1.47	18	10
1:A:163:ILE:HG21	1:A:276:THR:OG1	0.82	1.74	12	19
1:A:252:GLY:HA3	1:A:275:ILE:HD13	0.80	1.52	18	11

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	128/162 (79%)	118±2 (92±1%)	8±2 (6±1%)	2±1 (1±0%)	15 61
All	All	2560/3240 (79%)	2361 (92%)	163 (6%)	36 (1%)	15 61

All 4 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	206	GLU	17
1	A	291	ARG	9
1	A	292	GLY	8
1	A	216	LEU	2

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	111/136 (82%)	67±4 (60±4%)	44±4 (40±4%)	0 5
All	All	2220/2720 (82%)	1343 (60%)	877 (40%)	0 5

5 of 84 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	159	SER	20
1	A	163	ILE	20
1	A	210	GLN	20
1	A	227	PHE	20
1	A	245	ILE	20

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)

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