wwPDB NMR Structure Validation Summary Report (i)

Oct 23, 2021 – 05:29 PM EDT

PDB ID	:	1FUW
Title	:	SOLUTION STRUCTURE AND BACKBONE DYNAMICS OF A DOUBLE MUTANT SINGLE-CHAIN MONELLIN(SCM) DETERMINED BY NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY
Authors	:	Sung, Y.H.; Shin, J.; Jung, J.; Lee, W.
Deposited on	:	2000-09-18

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:

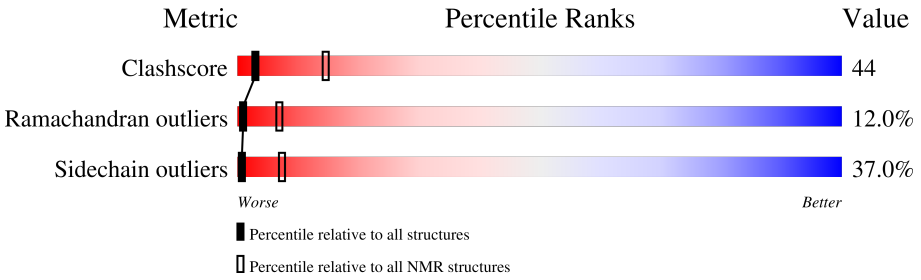
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.23.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 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	91	22% 51% 14% • 12%

2 Ensemble composition and analysis (i)

This entry contains 21 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: *minimized average structure*.

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:76, A:82-A:88 (80)	0.51	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 4 clusters and 2 single-model clusters were found.

Cluster number	Models
1	3, 7, 8, 9, 10, 11, 17, 19
2	1, 2, 6, 14, 16
3	13, 15, 18
4	4, 12, 20
Single-model clusters	5; 21

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1521 atoms, of which 760 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called MONELLIN.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	91	1521	490	760	128	141	2	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	7	GLU	ASP	engineered mutation	UNP P02882
A	39	LYS	ARG	engineered mutation	UNP P02882
A	49	ASN	GLU	SEE REMARK 999	UNP P02882
A	50	GLU	ASN	SEE REMARK 999	UNP P02882

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: MONELLIN

Chain A: 22% 51% 14% • 12%

G1 E2 W3 E4 I5 I6 E7 I8 F11 T12 Q13 N14 L15 G16 K17 F18 A19 E22 E23 N24 K25 I26 Y29 G30 R31 L32 T33 F34 V37 I38 K39 P40 C41 M42 K43 K44 T45 I46 Y47 E48 N49 E52 I53 K54 G55 Y58 Q59 L60 Y61 V62 Y63 A64 S65 D66

K67 L68 F69 R70 A71 D72 I73 S74 E75 D76 Y77 K78 T79 R80 G81 R82 K83 L84 L85 R86 F87 N88 G89 P90 V91

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: MONELLIN

Chain A: 33% 35% 19% • 12%

G1 E2 W3 I6 E7 F11 T12 Q13 N14 L15 G16 K17 F18 A19 E22 E23 I26 R31 L32 T33 V37 I38 K39 P40 C41 M42 K43 K44 T45 I46 Y47 E48 N49 I53 K54 G55 Y58 Q59 L60 Y61 V62 S65 D66 K67 L68 F69 R70 A71 D72 I73 S74

E75 D76 Y77 K78 T79 R80 G81 R82 K83 L84 L85 R86 F87 N88 G89 P90 V91

5 Refinement protocol and experimental data overview (i)

Of the 50 calculated structures, 21 were deposited, based on the following criterion: *structures with the lowest energy*.

The authors did not provide any information on software used for structure solution, optimization or refinement.

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	672	675	675	59±22
All	All	14112	14175	14175	1242

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 44.

5 of 492 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:37:VAL:HG22	1:A:62:VAL:HG23	1.05	1.28	9	8
1:A:37:VAL:HG22	1:A:62:VAL:HG13	1.04	1.28	3	4
1:A:37:VAL:HG13	1:A:62:VAL:HG12	1.01	1.28	18	6
1:A:15:LEU:HD21	1:A:73:ILE:HD13	1.00	1.22	2	2
1:A:73:ILE:HD13	1:A:84:LEU:HD23	0.95	1.38	11	1

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	80/91 (88%)	51±3 (64±4%)	19±3 (24±3%)	10±2 (12±3%)	1 7
All	All	1680/1911 (88%)	1072 (64%)	407 (24%)	201 (12%)	1 7

5 of 35 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	38	ILE	21
1	A	86	ARG	20
1	A	40	PRO	18
1	A	53	ILE	18
1	A	49	ASN	17

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	72/80 (90%)	45±4 (63±6%)	27±4 (37±6%)	1 7
All	All	1512/1680 (90%)	953 (63%)	559 (37%)	1 7

5 of 64 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	12	THR	21
1	A	58	TYR	21
1	A	85	LEU	21
1	A	86	ARG	19
1	A	87	PHE	19

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