wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 03:30 PM BST

PDB ID	:	1L2N
Title	:	Smt3 Solution Structure
Authors	:	Sheng, W.; Liao, X.
Deposited on	:	2002-02-22

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 (i) symbol.

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

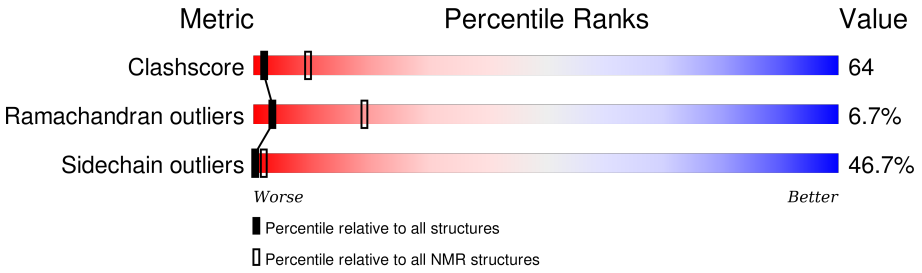
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

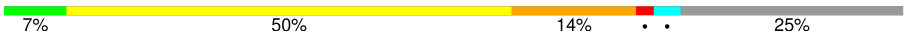
The overall completeness of chemical shifts assignment was not calculated.

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	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 <=5%

Mol	Chain	Length	Quality of chain
1	A	101	

2 Ensemble composition and analysis (i)

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: *minimized target function*.

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:22-A:94 (73)	0.23	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 4 clusters and 2 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1249 atoms, of which 624 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Ubiquitin-like protein SMT3.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	76	1249	389	624	110	123	3	0

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: Ubiquitin-like protein SMT3

Chain A: 7% 50% 14% • • 25%

MET SER ASP SER GLU VAL ASN GLN GLU ALA LYS PRO GLU VAL LYS PRO GLU VAL LYS PRO E21 T22 H23 I24 N25 L26 K27 V28 S29 D30 G31 S32 S33 E34 I35 F36 F37 K38 I39 K40 K41 T42 T43 P44 L45 R46 R47 L48 M49 E50 A51 F52 A53 K54 R55 Q56 G57 K58 E59 M60

D61 S62 L63 R64 F65 L66 Y67 D68 G69 I70 R71 I72 Q73 A74 D75 Q76 T77 P78 E79 D80 L81 D82 M83 E84 D85 N86 D87 I88 I89 E90 A91 H92 R93 E94 Q95 I96 GLY GLY ALA THR TYR

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: Ubiquitin-like protein SMT3

Chain A: 19% 31% 21% • • 25%

MET SER ASP SER GLU VAL ASN GLN GLU ALA LYS PRO GLU VAL LYS PRO GLU VAL LYS PRO E21 T22 H23 I24 N25 L26 S29 D30 G31 S32 S33 E34 I35 I39 T42 T43 P44 L45 R46 R47 L48 A51 F52 A53 K54 R55 K58 E59 M60 D61 S62 L63 R64 F65 L66

Y67 D68 G69 I70 R71 I72 Q73 A74 D75 Q76 T77 P78 E79 D80 L81 D82 M83 E84 D85 N86 D87 I88 I89 E90 A91 H92 R93 E94 Q95 I96 GLY GLY ALA THR TYR

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function*.

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

Software name	Classification	Version
DYANA	refinement	1.5

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 (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	599	598	595	76±7
All	All	11980	11960	11900	1530

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 64.

5 of 372 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:24:ILE:HD11	1:A:87:ASP:OD1	0.95	1.61	12	9
1:A:24:ILE:HD11	1:A:87:ASP:OD2	0.94	1.61	16	5
1:A:52:PHE:CE1	1:A:91:ALA:HB1	0.93	1.99	5	11
1:A:66:LEU:HD11	1:A:69:GLY:HA2	0.90	1.44	10	2
1:A:26:LEU:HD13	1:A:47:ARG:CG	0.89	1.97	5	6

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	73/101 (72%)	50±2 (69±2%)	18±2 (25±3%)	5±1 (7±1%)	3 19
All	All	1460/2020 (72%)	1003 (69%)	359 (25%)	98 (7%)	3 19

5 of 10 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	31	GLY	20
1	A	70	ILE	19
1	A	43	THR	17
1	A	86	ASN	12
1	A	60	MET	10

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	66/90 (73%)	35±3 (53±5%)	31±3 (47±5%)	0 2
All	All	1320/1800 (73%)	704 (53%)	616 (47%)	0 2

5 of 60 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	42	THR	20
1	A	55	ARG	20
1	A	88	ILE	20
1	A	63	LEU	20
1	A	75	ASP	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 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 (i)

No chemical shift data were provided