wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 09:51 PM BST

PDB ID	:	2JMB
Title	:	Solution structure of the protein Atu4866 from Agrobacterium tumefaciens
Authors	:	Ai, X.; Semesi, A.; Yee, A.; Arrowsmith, C.H.; Li, S.S.C.; Choy, W.; Ontario Centre for Structural Proteomics (OCSP)
Deposited on	:	2006-11-01

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:

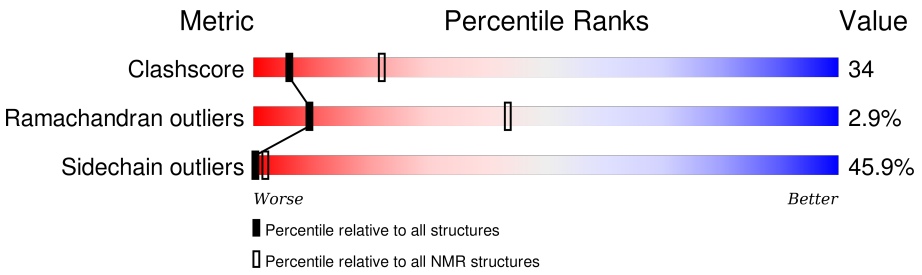
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	79	32% 35% 20% • 11%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 13 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:4-A:30, A:36-A:78 (70)	0.09	13

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 and 3 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1247 atoms, of which 599 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Hypothetical protein Atu4866.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	79	1247	403	599	122	121	2	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: Hypothetical protein Atu4866

Chain A: 32% 35% 20% • 11%

M1 Q2 H3 P4 Y5 V6 G7 I8 W9 V10 T11 A12 D13 G14 R15 I16 R17 L20 L21 R25 Y26 D27 G31 N32 R33 K34 S35 A36 Y37 Q38 G39 R40 Y41 E42 V43 R44 G45 A46 H47 I48 N49 Y50 W51 D52 F56 T57 A58 D59 G60 D61 F62 N66 E67 L68 H69

H70 M73 T74 F75 Y76 R77 E78 K79

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

The representative model is number 13. Colouring as in section 4.1 above.

- Molecule 1: Hypothetical protein Atu4866

Chain A: 33% 39% 15% • 11%

M1 Q2 H3 P4 Y5 V6 G7 I8 W9 V10 T11 A12 D13 G14 R15 I16 R17 L20 L21 R25 Y26 D27 G31 N32 R33 K34 S35 A36 Y37 Q38 G39 R40 Y41 E42 V43 R44 G45 A46 H47 I48 N49 Y50 W51 D52 F56 T57 G60 D61 F62 N66 E67 L68 H69 H70

F75 Y76 R77 E78 K79

5 Refinement protocol and experimental data overview (i)

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

Of the 20 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy.*

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

Software name	Classification	Version
CNS	structure solution	1.1
CNS	refinement	1.1

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	573	519	516	37±3
All	All	11460	10380	10320	732

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:5:TYR:CE1	1:A:43:VAL:HG21	0.92	1.99	6	20
1:A:68:LEU:HD12	1:A:75:PHE:HB2	0.86	1.47	2	20
1:A:68:LEU:HD13	1:A:68:LEU:C	0.83	1.94	11	11
1:A:68:LEU:C	1:A:68:LEU:HD13	0.82	1.94	17	9
1:A:5:TYR:OH	1:A:43:VAL:HG11	0.80	1.77	5	18

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	70/79 (89%)	63±1 (90±1%)	5±1 (7±1%)	2±0 (3±0%)	9 43
All	All	1400/1580 (89%)	1256 (90%)	104 (7%)	40 (3%)	9 43

All 2 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	6	VAL	20
1	A	78	GLU	20

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	55/63 (87%)	30±1 (54±2%)	25±1 (46±2%)	0 2
All	All	1100/1260 (87%)	595 (54%)	505 (46%)	0 2

5 of 36 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	26	TYR	20
1	A	27	ASP	20
1	A	37	TYR	20
1	A	68	LEU	20
1	A	48	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 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