wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 02:24 PM BST

PDB ID	:	1EHX
Title	:	NMR SOLUTION STRUCTURE OF THE LAST UNKNOWN MODULE OF THE CELLULOSOMAL SCAFFOLDIN PROTEIN CIPC OF CLOSTRIDUM CELLULOLYTICUM
Authors	:	Mosbah, A.; Belaich, A.; Bornet, O.; Belaich, J.P.; Henrissat, B.; Darbon, H.
Deposited on	:	2000-02-23

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:

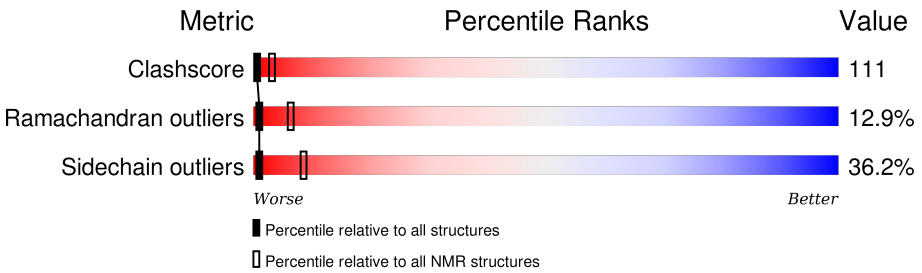
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 is 46%.

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	94	5% 59% 24% • 11%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 14 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:3-A:86 (84)	0.35	14

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1421 atoms, of which 715 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called SCAFFOLDIN PROTEIN.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	94	1421	445	715	112	148	1	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	PRO	CONFLICT	UNP Q45996
A	93	LEU	ASP	CONFLICT	UNP Q45996
A	94	GLU	SER	CONFLICT	UNP Q45996

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: SCAFFOLDIN PROTEIN

Chain A: 5% 59% 24% • 11%

M1 Q2 D3 P4 T5 I6 N7 P8 T9 S10 I11 S12 A13 K14 A15 G16 S17 F18 A19 D20 T21 K22 I23 T24 L25 T26 P27 N28 G29 N30 T31 F32 N33 G34 I35 S36 E37 L38 Q39 S40 S41 Q42 Y43 T44 K45 G46 T47 N48 E49 V50 T51 L52 L53 A54 S55 Y56 L57 N58 T59 L60

P61 E62 N63 T64 T65 K66 T67 L68 T69 F70 D71 F72 G73 V74 G75 T76 K77 N78 P79 K80 L81 T82 I83 T84 V85 L86 P87 K88 D89 I90 P91 G92 L93 E94

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

The representative model is number 14. Colouring as in section 4.1 above.

- Molecule 1: SCAFFOLDIN PROTEIN

Chain A: 10% 49% 29% • 11%

M1 Q2 D3 P4 T5 I6 N7 P8 T9 S10 I11 S12 A13 K14 A15 G16 S17 F18 A19 D20 T21 K22 I23 T24 L25 T26 P27 N28 G29 N30 T31 F32 N33 G34 I35 S36 E37 L38 Q39 S40 S41 Q42 Y43 T44 K45 G46 T47 N48 E49 V50 T51 L52 L53 A54 S55 Y56 L57 N58 T59 L60

P61 E62 N63 T64 T65 K66 T67 L68 T69 F70 D71 F72 G73 V74 G75 T76 K77 N78 P79 K80 L81 T82 I83 T84 V85 L86 P87 K88 D89 I90 P91 G92 L93 E94

5 Refinement protocol and experimental data overview (i)

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

Of the 50 calculated structures, 20 were deposited, based on the following criterion: *back calculated data agree with experimental NOESY spectrum, structures with acceptable covalent geometry, structures with favorable non-bond energy, structures with the least restraint violations, structures with the lowest energy.*

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

Software name	Classification	Version
DIANA	structure solution	2.8
CNS	refinement	1.0

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	BMRB entry 4589
Number of chemical shift lists	1
Total number of shifts	611
Number of shifts mapped to atoms	611
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	46%

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	628	634	634	140±14
All	All	12560	12680	12680	2794

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 111.

5 of 1158 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:35:ILE:HD11	1:A:70:PHE:CE1	1.00	1.91	5	7
1:A:13:ALA:HB3	1:A:18:PHE:CE1	0.99	1.93	11	17
1:A:39:GLN:O	1:A:41:SER:N	0.99	1.96	1	20
1:A:11:ILE:HG23	1:A:83:ILE:HG23	0.99	1.29	12	2
1:A:52:LEU:HD11	1:A:68:LEU:HD11	0.98	1.30	7	2

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	84/94 (89%)	57±3 (67±3%)	17±3 (20±3%)	11±1 (13±1%)	1 6
All	All	1680/1880 (89%)	1132 (67%)	332 (20%)	216 (13%)	1 6

5 of 19 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	40	SER	20
1	A	54	ALA	20
1	A	28	ASN	20
1	A	39	GLN	20
1	A	47	THR	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	74/83 (89%)	47±3 (64±4%)	27±3 (36±4%)	1 8
All	All	1480/1660 (89%)	944 (64%)	536 (36%)	1 8

5 of 62 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	17	SER	20
1	A	30	ASN	20
1	A	22	LYS	19
1	A	65	THR	19
1	A	12	SER	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 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)

The completeness of assignment taking into account all chemical shift lists is 46% for the well-defined parts and 46% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 4589

Chemical shift list name: *assigned_ chem_ shift_ list_ 1*

7.1.1 Bookkeeping (i)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	611
Number of shifts mapped to atoms	611
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	0	—	—
$^{13}C_{\beta}$	0	—	—
$^{13}C'$	0	—	—
^{15}N	86	-1.12 ± 0.51	Should be applied

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 46%, i.e. 438 atoms were assigned a chemical shift out of a possible 953. 0 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	241/410 (59%)	162/163 (99%)	0/168 (0%)	79/79 (100%)
Sidechain	197/491 (40%)	197/283 (70%)	0/192 (0%)	0/16 (0%)

Continued on next page...

Continued from previous page...

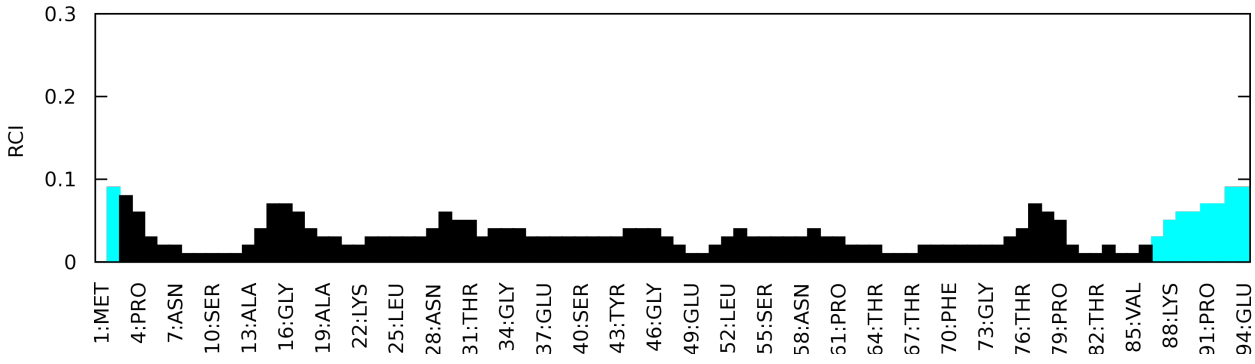
	Total	^{1}H	^{13}C	^{15}N
Aromatic	0/52 (0%)	0/28 (0%)	0/24 (0%)	0/0 (—%)
Overall	438/953 (46%)	359/474 (76%)	0/384 (0%)	79/95 (83%)

7.1.4 Statistically unusual chemical shifts (i)

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	13	ALA	HB3	0.06	2.61 – 0.11	-5.2
1	A	13	ALA	HB2	0.06	2.61 – 0.11	-5.2
1	A	13	ALA	HB1	0.06	2.61 – 0.11	-5.2

7.1.5 Random Coil Index (RCI) plots (i)

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A: