wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 02:18 PM BST

PDB ID	:	1CX1
Title	:	SECOND N-TERMINAL CELLULOSE-BINDING DOMAIN FROM CELLULOMONAS FIMI BETA-1,4-GLUCANASE C, NMR, 22 STRUCTURES
Authors	:	Brun, E.; Johnson, P.E.; Creagh, L.A.; Haynes, C.A.; Tomme, P.; Webster, P.; Kilburn, D.G; McIntosh, L.P.
Deposited on	:	1999-08-27

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:

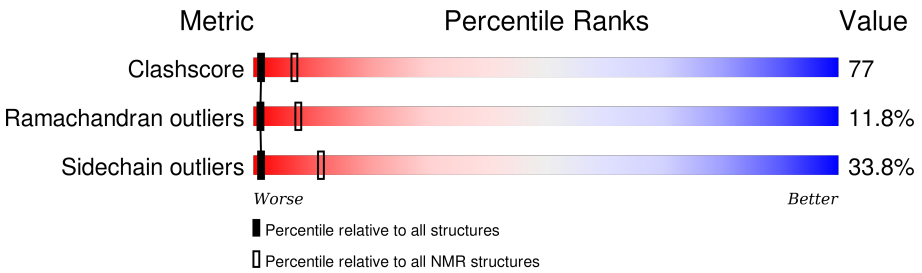
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 85%.

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	153	12% 59% 20% 8%

2 Ensemble composition and analysis (i)

This entry contains 22 models. Model 22 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:7-A:87, A:92-A:150 (140)	0.47	22

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 2 single-model clusters were found.

Cluster number	Models
1	1, 3, 4, 5, 6, 7, 8, 9, 11, 14, 15, 16, 17, 18, 19, 20, 21, 22
2	10, 13
Single-model clusters	2; 12

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2164 atoms, of which 1044 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called ENDOGLUCANASE C.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	153	2164	708	1044	179	229	4	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	ALA	VAL	CLONING ARTIFACT	UNP P14090
A	2	SER	ALA	CLONING ARTIFACT	UNP P14090

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: ENDOGLUCANASE C

Chain A: 12% 59% 20% 8%

A1 S2 L3 D4 S5 E6 V7 E8 L9 L10 P11 H12 T13 S14 F15 A16 E17 S18 L19 W22 S23 L24 Y25 G26 T27 S28 E29 P30 V31 F32 A33 D34 G35 R36 M37 C38 V39 D40 L41 P42 G43 G44 Q45 G46 N47 P48 W49 D50 A51 G52 L53 V54 Y55 N56 G57 V58 P59 V60 G61

E62 G63 E64 S65 Y66 V67 L68 S69 F70 T71 A72 S73 A74 T75 F76 D77 M78 F79 V80 R81 V82 L83 V84 G85 E86 G87 G88 G89 A90 Y91 R92 T93 A94 F95 E96 Q97 G98 S99 A100 P101 L102 T103 G104 E105 P106 A107 T108 R109 E110 Y111 A112 F113 T114 S115 N116 L117 T118 F119 D122

G123 D124 A125 P126 G127 Q128 V129 A130 F131 H132 L133 G134 K135 A136 G137 A138 Y139 E140 F141 C142 I143 S144 Q145 V146 S147 L148 T149 T150 S151 A152 T153

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

The representative model is number 22. Colouring as in section 4.1 above.

- Molecule 1: ENDOGLUCANASE C

Chain A: 29% 41% 19% • 8%

A1 S2 L3 D4 S5 E6 V7 E8 L9 L10 P11 H12 T13 S14 F15 A16 E17 S18 L19 G20 P21 W22 S23 L24 Y25 G26 T27 S28 E29 P30 V31 F32 A33 D34 M37 C38 V39 D40 L41 P42 Q45 G46 N47 P48 W49 D50 A51 G52 L53 V54 Y55 N56 G61 E62 Y66 V67

L68 S69 F70 T71 A72 S73 A74 T75 P76 D77 M78 F79 V80 R81 V82 L83 V84 G85 E86 G87 G88 G89 A90 Y91 R92 T93 A94 F95 E96 Q97 A100 P101 L102 T103 G104 T108 R109 E110 Y111 A112 F113 T114 S115 N116 L117 T118 F119 D122 G123 D124 Q128 V129 A130 F131 H132 L133

G134 K135 A136 G137 A138 Y139 E140 F141 C142 I143 S144 Q145 V146 S147 L148 T149 T150 S151 A152 T153

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *ALL CALCULATIONS WERE PERFORMED USING X-PLOR V3.8 WITH SCRIPTS FROM THE X-PLOR 3.1 MANUAL A. A. BRUNGER (1992), NEW HAVEN: YALE UNIVERSITY PRESS ACCORDING TO THE METHOD DESCRIBED BY M. NILGES, G. M. CLORE & A. M. GRONENBORN.*

Of the 60 calculated structures, 22 were deposited, based on the following criterion: *STRUCTURES WITH ACCEPTABLE COVALENT GEOMETRY,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**
X-PLOR	structure solution	3.8
X-PLOR	refinement	3.8

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 4706
Number of chemical shift lists	1
Total number of shifts	1609
Number of shifts mapped to atoms	1609
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	85%

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	1034	969	969	153±14
All	All	22748	21318	21318	3375

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 77.

5 of 1271 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:10:LEU:HD12	1:A:55:TYR:CZ	0.95	1.95	17	4
1:A:24:LEU:HD12	1:A:24:LEU:O	0.94	1.62	5	3
1:A:53:LEU:N	1:A:53:LEU:HD22	0.90	1.81	17	2
1:A:24:LEU:O	1:A:24:LEU:HD12	0.90	1.67	16	2
1:A:41:LEU:N	1:A:41:LEU:HD22	0.90	1.81	6	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	140/153 (92%)	92±5 (66±4%)	31±5 (22±4%)	16±2 (12±2%)	1 8
All	All	3080/3366 (92%)	2026 (66%)	692 (22%)	362 (12%)	1 8

5 of 57 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	76	PRO	21
1	A	138	ALA	21
1	A	33	ALA	20
1	A	139	TYR	19
1	A	42	PRO	18

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	109/117 (93%)	72±4 (66±4%)	37±4 (34±4%)	1 11
All	All	2398/2574 (93%)	1588 (66%)	810 (34%)	1 11

5 of 82 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	27	THR	22
1	A	143	ILE	22
1	A	114	THR	22
1	A	32	PHE	22
1	A	108	THR	22

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 85% for the well-defined parts and 84% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 4706

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	1609
Number of shifts mapped to atoms	1609
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	7

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	153	0.39 ± 0.17	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	133	-0.30 ± 0.14	None needed (< 0.5 ppm)
$^{13}C'$	0	—	—
^{15}N	138	-1.56 ± 0.28	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 85%, i.e. 1309 atoms were assigned a chemical shift out of a possible 1548. 21 out of 24 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	534/674 (79%)	267/267 (100%)	140/280 (50%)	127/127 (100%)
Sidechain	644/722 (89%)	394/424 (93%)	239/278 (86%)	11/20 (55%)

Continued on next page...

Continued from previous page...

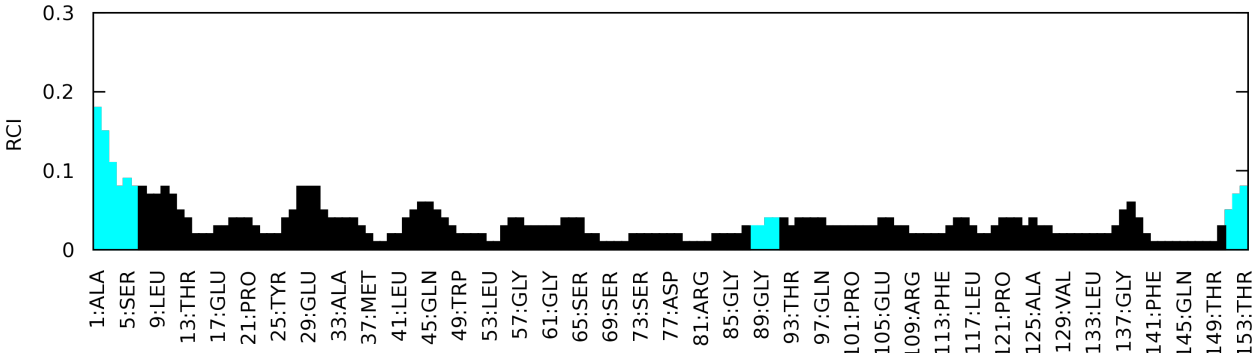
	Total	^{1}H	^{13}C	^{15}N
Aromatic	131/152 (86%)	64/80 (80%)	61/66 (92%)	6/6 (100%)
Overall	1309/1548 (85%)	725/771 (94%)	440/624 (71%)	144/153 (94%)

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	49	TRP	CH2	74.70	133.06 – 114.56	-26.5
1	A	139	TYR	CE1	141.10	124.14 – 111.74	18.7
1	A	139	TYR	CE2	141.10	124.68 – 111.18	17.2
1	A	47	ASN	HB2	-0.27	4.36 – 1.26	-9.9
1	A	55	TYR	HB2	0.30	4.76 – 1.06	-7.0
1	A	49	TRP	CZ2	122.05	121.76 – 106.66	5.2
1	A	66	TYR	HB2	1.03	4.76 – 1.06	-5.1

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: