wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 02:16 PM BST

PDB ID	:	1DAV
Title	:	SOLUTION STRUCTURE OF THE TYPE I DOCKERIN DOMAIN FROM THE CLOSTRIDIUM THERMOCELLUM CELLULOSOME (20 STRUCTURES)
Authors	:	Lytle, B.L.; Volkman, B.F.; Westler, W.M.; Heckman, M.P.; Wu, J.H.D.
Deposited on	:	1999-10-31

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:

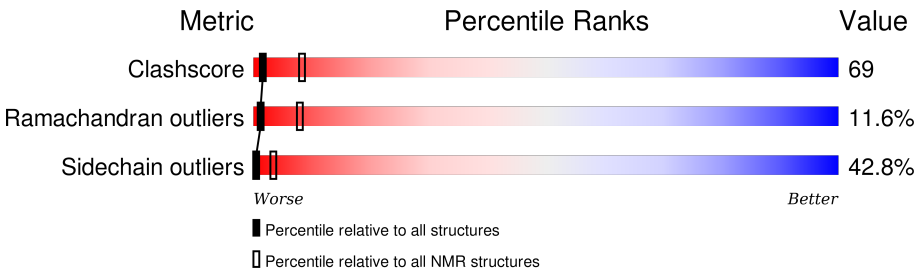
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	71	10% 56% 21% • 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:4-A:66 (63)	0.42	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 4 clusters and 5 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1105 atoms, of which 553 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called ENDOGLUCANASE SS.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	71	1103	339	553	96	114	1	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	PRO	ENGINEERED	UNP P38686
A	71	GLY	-	INSERTION	UNP P38686

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	
			Total	Ca
2	A	2	2	2

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 SS

Chain A: 10% 56% 21% • 11%

M1 S2 T3 K4 L5 Y6 G7 D8 V9 N10 D11 D12 G13 K14 V15 N16 D19 A20 V21 A22 L23 K24 R25 Y26 V27 L28 R29 S30 G31 I32 S33 I34 N35 T36 D37 N38 A39 D40 L41 N42 E43 D44 G45 R46 V47 N48 S49 T50 D51 L52 G53 I54 L55 K56 R57 Y58 I59 L60 K61

E62 I63 D64 T65 L66 P67 Y68 K69 N70 G71

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: ENDOGLUCANASE SS

Chain A: 15% 42% 30% • 11%

M1 S2 T3 K4 L5 Y6 G7 D8 V9 N10 D11 D12 G13 K14 V15 N16 S17 A20 V21 A22 L23 K24 R25 Y26 V27 L28 R29 S30 G31 I32 S33 I34 N35 T36 D37 N38 A39 D40 L41 N42 E43 D44 G45 R46 V47 N48 S49 T50 D51 L52 G53 I54 L55 K56 R57 Y58 I59 L60 K61

E62 I63 D64 T65 L66 P67 Y68 K69 N70 G71

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
DYANA	structure solution	1.5
XEASY	structure solution	1.2

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)

Bond lengths and bond angles in the following residue types are not validated in this section: CA

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	489	494	494	68±6
All	All	9820	9880	9879	1350

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 69.

5 of 508 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:58:TYR:CE2	1:A:66:LEU:HD12	1.09	1.82	9	17
1:A:40:ASP:O	1:A:41:LEU:HD13	0.94	1.62	18	5
1:A:6:TYR:CZ	1:A:41:LEU:HD13	0.93	1.98	12	2
1:A:58:TYR:CE1	1:A:59:ILE:HG22	0.92	1.99	18	14
1:A:47:VAL:HG12	1:A:55:LEU:HD13	0.92	1.39	14	5

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	63/71 (89%)	41±3 (64±6%)	15±3 (24±5%)	7±1 (12±2%)	1 8
All	All	1260/1420 (89%)	811 (64%)	303 (24%)	146 (12%)	1 8

5 of 20 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	13	GLY	20
1	A	64	ASP	19
1	A	49	SER	19
1	A	61	LYS	18
1	A	63	ILE	15

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	55/62 (89%)	31±2 (57±4%)	24±2 (43±4%)	0 3
All	All	1100/1240 (89%)	629 (57%)	471 (43%)	0 3

5 of 52 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	14	LYS	20
1	A	58	TYR	19
1	A	54	ILE	18
1	A	29	ARG	18
1	A	52	LEU	17

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)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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