wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 09:50 PM BST

PDB ID	:	2K13
Title	:	Solution NMR Structure of the Leech Protein Saratin, a Novel Inhibitor of Haemostasis
Authors	:	Gronwald, W.; Bomke, J.; Maurer, T.; Wisotzki, B.; Huber, F.; Schumann, F.; Kremer, W.; Frech, M.; Kalbitzer, H.R.
Deposited on	:	2008-02-20

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:

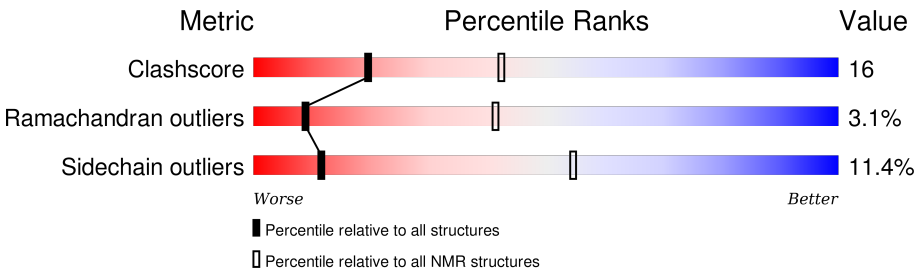
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	X	103	39% 32% • 27%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 5 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest total and noe energy.*

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	X:5-X:79 (75)	0.43	5

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 3 clusters and 2 single-model clusters were found.

Cluster number	Models
1	2, 5, 8, 9
2	3, 4
3	1, 7
Single-model clusters	6; 10

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1582 atoms, of which 736 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Saratin.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	X	103	1582	521	736	127	192	6	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: Saratin

Chain X: 39% 32% • 27%

E1 E2 R3 E4 D5 C6 W7 T8 F9 N12 R13 K14 Y15 T16 F22 K23 K24 D27 L28 D29 E30 C31 K32 K33 T34 C35 T38 C41 Y42 I43 Y44 F45 K51 Y55 N56 V57 V58 D59 Q65 E66 K67 F68 D71 T75 E76 N77 Y78 L79 T80 D81

C82 E83 G84 K85 D86 A87 G88 N89 A90 A91 G92 T93 G94 D95 E96 S97 D98 E99 V100 D101 E102 D103

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

The representative model is number 5. Colouring as in section 4.1 above.

- Molecule 1: Saratin

Chain X: 38% 33% • 27%

E1 E2 R3 E4 D5 C6 W7 T8 F9 Y10 A11 N12 R13 K14 Y15 T16 K20 S21 F22 K23 D27 L28 D29 E30 C31 K32 K33 T34 C35 F36 K37 T38 C41 Y42 I43 K51 Y55 N56 V57 V58 D59 Q65 F68 E72 T75 E76 N77 Y78 L79 T80 D81

C82 E83 G84 K85 D86 A87 G88 N89 A90 A91 G92 T93 G94 D95 E96 S97 D98 E99 V100 D101 E102 D103

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing*.

Of the 1000 calculated structures, 10 were deposited, based on the following criterion: *10 structure of lowest total and NOE energy*.

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

Software name	Classification	Version
CNS	refinement	

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)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	X	1.02±0.02	0±0/658 (0.0±0.0%)	0.97±0.02	0±0/888 (0.0±0.0%)
All	All	1.02	1/6580 (0.0%)	0.97	2/8880 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	X	0.0±0.0	0.2±0.4
All	All	0	2

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	X	14	LYS	CA-CB	-5.42	1.42	1.53	8	1

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	X	9	PHE	CB-CG-CD1	-5.91	116.66	120.80	8	2

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	X	38	THR	Peptide	2

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	X	642	577	577	20±4
All	All	6420	5770	5770	200

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 16.

5 of 89 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:X:28:LEU:O	1:X:32:LYS:HD3	0.66	1.91	6	8
1:X:7:TRP:CH2	1:X:32:LYS:HG3	0.63	2.29	3	1
1:X:6:CYS:O	1:X:79:LEU:HA	0.63	1.94	8	1
1:X:27:ASP:CB	1:X:30:GLU:HB2	0.62	2.24	7	8
1:X:14:LYS:HA	1:X:68:PHE:CD2	0.62	2.30	5	4

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	X	75/103 (73%)	66±2 (88±2%)	6±2 (9±3%)	2±1 (3±1%)	9	41
All	All	750/1030 (73%)	663 (88%)	64 (9%)	23 (3%)	9	41

All 5 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	X	38	THR	10
1	X	41	CYS	4
1	X	16	THR	4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	X	21	SER	3
1	X	42	TYR	2

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	X	73/94 (78%)	65±1 (89±2%)	8±1 (11±2%)	11 54
All	All	730/940 (78%)	647 (89%)	83 (11%)	11 54

5 of 26 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	X	42	TYR	10
1	X	78	TYR	8
1	X	38	THR	7
1	X	12	ASN	6
1	X	14	LYS	6

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