wwPDB NMR Structure Validation Summary Report (i)

Apr 21, 2024 – 04:11 AM EDT

PDB ID	:	2LKV
BMRB ID	:	18013
Title	:	Staphylococcal Nuclease PHS variant
Authors	:	Matzapetakis, M.; Pais, T.M.; Lamosa, P.; Turner, D.L.; Santos, H.
Deposited on	:	2011-10-21

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
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

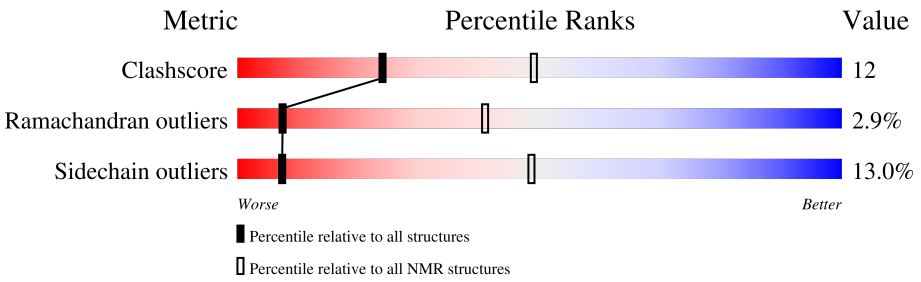
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 91%.

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	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

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	149	

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 3 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:8-A:41, A:55-A:140 (120)	0.41	3

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 1 clusters and 3 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2389 atoms, of which 1212 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Thermonuclease.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	149	2389	744	1212	206	223	4	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	117	GLY	PRO	engineered mutation	UNP Q8NXI6
A	128	ALA	SER	engineered mutation	UNP Q8NXI6

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide 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: Thermonuclease

Chain A: 52% 27% • 19%

A1 T2 S3 T4 K5 K6 L7 H8 A12 T13 T22 M26 F34 R35 L36 L37 P42 E43 T44 K45 H46 P47 K48 K49 G50 V51 E52 K53 Y54 G55 P56 S59 T62 K63 K64 H65 V66 E67 N68 D77 K78 G79 T82 D83 K84 R87 G88 L89 A90 Y91

I92 V99 V104 R105 Q106 G107 L108 V114 Y115 K116 G117 N118 N119 Q123 L124 A128 Q131 K134 L137 N138 I139 W140 S141 E142 D143 N144 A145 D146 S147 G148 Q149

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

The representative model is number 3. Colouring as in section 4.1 above.

- Molecule 1: Thermonuclease

Chain A: 49% 28% • • 19%

A1 T2 S3 T4 K5 K6 L7 H8 K9 E10 P11 A12 T13 A17 T22 M26 T33 F34 R35 L36 L37 P42 E43 T44 K45 H46 P47 K48 K49 G50 V51 E52 K53 Y54 G55 P56 E57 A60 F61 T62 K63 V66 E67 N68 V74 D77 K78 G79 T82 D83

K84 Y85 G86 R87 G88 L89 I92 V99 V104 G107 L108 V114 N118 Q123 L124 L125 R126 K127 A128 Q131 K134 E135 K136 L137 N138 I139 W140 S141 E142 D143 N144 A145 D146 S147 G148 Q149

5 Refinement protocol and experimental data overview (i)

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

Of the 80 calculated structures, 10 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	refinement	1.1
CcpNmr Analysis	structure solution	2.1
TopSpin	structure solution	2.1
UNIO	structure solution	2.0
CYANA	structure solution	2.1
CYANA	refinement	2.1

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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	1847
Number of shifts mapped to atoms	1847
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	91%

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	A	1.15±0.04	2±1/970 (0.2± 0.1%)	0.90±0.02	1±0/1302 (0.1± 0.0%)
All	All	1.15	22/9700 (0.2%)	0.90	7/13020 (0.1%)

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	A	0.0±0.0	0.6±0.5
All	All	0	6

All unique bond outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	91	TYR	CB-CG	-6.16	1.42	1.51	8	2
1	A	89	LEU	N-CA	-6.03	1.34	1.46	9	8
1	A	96	GLY	N-CA	-5.83	1.37	1.46	2	2
1	A	88	GLY	C-N	-5.74	1.20	1.34	10	9
1	A	88	GLY	CA-C	-5.19	1.43	1.51	2	1

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	89	LEU	N-CA-CB	-5.89	98.61	110.40	10	6
1	A	22	THR	N-CA-CB	5.21	120.20	110.30	8	1

There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the

ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	35	ARG	Sidechain	4
1	A	81	ARG	Sidechain	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	A	954	990	989	23±4
All	All	9540	9900	9890	233

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 12.

5 of 91 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:82:THR:HA	1:A:87:ARG:O	0.71	1.85	1	10
1:A:37:LEU:HD21	1:A:89:LEU:HD13	0.71	1.61	10	1
1:A:77:ASP:HB3	1:A:118:ASN:OD1	0.69	1.87	4	9
1:A:77:ASP:OD1	1:A:118:ASN:HA	0.67	1.87	10	6
1:A:10:GLU:HB2	1:A:74:VAL:HG23	0.67	1.66	4	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	A	120/149 (81%)	106±2 (88±2%)	10±2 (9±2%)	4±1 (3±1%)	7 41
All	All	1200/1490 (81%)	1062 (88%)	103 (9%)	35 (3%)	7 41

5 of 6 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	138	ASN	10
1	A	119	ASN	8
1	A	84	LYS	7
1	A	22	THR	6
1	A	136	LYS	3

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	98/123 (80%)	85±2 (87±2%)	13±2 (13±2%)	7 48
All	All	980/1230 (80%)	853 (87%)	127 (13%)	7 48

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	123	GLN	10
1	A	131	GLN	10
1	A	134	LYS	10
1	A	108	LEU	8
1	A	26	MET	8

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 monosaccharides 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 91% for the well-defined parts and 90% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_2*

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	1847
Number of shifts mapped to atoms	1847
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	148	-0.21 ± 0.10	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	137	0.09 ± 0.17	None needed (< 0.5 ppm)
$^{13}C'$	123	-0.00 ± 0.08	None needed (< 0.5 ppm)
^{15}N	138	0.01 ± 0.57	None needed (< 0.5 ppm)

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 91%, i.e. 1542 atoms were assigned a chemical shift out of a possible 1686. 0 out of 19 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	585/603 (97%)	245/246 (100%)	224/240 (93%)	116/117 (99%)
Sidechain	874/971 (90%)	590/628 (94%)	269/301 (89%)	15/42 (36%)

Continued on next page...

Continued from previous page...

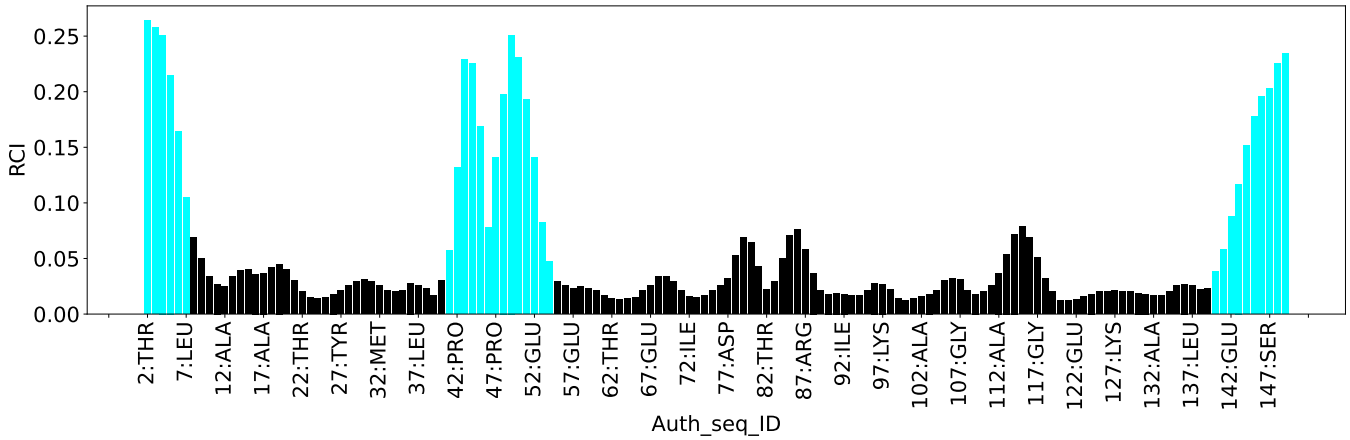
	Total	^{1}H	^{13}C	^{15}N
Aromatic	83/112 (74%)	41/53 (77%)	38/54 (70%)	4/5 (80%)
Overall	1542/1686 (91%)	876/927 (94%)	531/595 (89%)	135/164 (82%)

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.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	9	LYS	HE2	1.82	1.95 – 3.88	-5.7
1	A	100	ASN	HD22	9.62	4.69 – 9.61	5.0

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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A: