wwPDB NMR Structure Validation Summary Report (i)

Apr 21, 2024 – 07:40 AM EDT

PDB ID	:	2LN7
BMRB ID	:	18152
Title	:	Backbone 1H, 13C, and 15N Chemical Shift Assignments for the catalytic domain of B. anthracis SrtD
Authors	:	Robson, S.A.; Weiner, E.M.; Clubb, R.T.
Deposited on	:	2011-12-19

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:

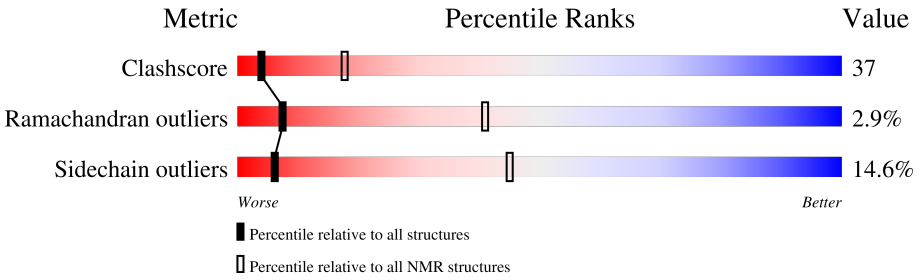
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson 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 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	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	147	31% 39% • • 27%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 19 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 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	A:11-A:34, A:44-A:64, A:74-A:121, A:133-A:147 (108)	0.45	19

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 4 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2325 atoms, of which 1161 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called LPXTG-site transpeptidase family protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	147	2325	742	1161	189	229	4	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP Q81KD1
A	2	SER	-	expression tag	UNP Q81KD1
A	3	HIS	-	expression tag	UNP Q81KD1
A	4	MET	-	expression tag	UNP Q81KD1

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: LPXTG-site transpeptidase family protein

Chain A: 31% 39% • • 27%

G1 S2 H3 M4 S5 S6 Q7 T8 E9 H10 K11 E12 G13 E14 K15 N20 I21 P22 K23 L24 K25 I30 Y31 W32 G33 A34 D35 D36 A37 T38 L39 K40 K41 G42 V43 G44 M45 F46 V47 S48 D49 V50 T51 T52 T53 G56 G57 G58 H59 T60 V61 L62 S63 G64 H65 R66 D67

T68 V69 F70 T71 D72 L73 G74 Q75 L76 K77 D80 T81 L82 V83 L84 E85 Y86 T90 Y91 T92 Y93 E94 I95 W99 I100 T101 H102 D105 R106 T107 V108 I109 I110 K111 K112 E113 E114 L117 T118 L119 T120 T121 C122 Y123 P124 F125 D126 Y127 I128 G129 D130 A131 P132 D133 R134

Y135 I136 I137 L141 T142 G143 S144 Y145 S146 K147

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

The representative model is number 19. Colouring as in section 4.1 above.

- Molecule 1: LPXTG-site transpeptidase family protein

Chain A: 32% 33% 7% • 27%

G1 S2 H3 M4 S5 S6 Q7 T8 E9 H10 K11 E12 G13 E14 K15 P22 K23 L24 K25 K26 I30 Y31 W32 G33 A34 D35 D36 A37 T38 L39 K40 K41 G42 V43 G44 M45 F46 V47 S48 D49 V50 T51 T52 T53 P54 G57 G58 H59 T60 V61 L62 S63 G64 H65 R66 D67 T68

V69 F70 T71 D72 L73 L76 K77 D80 T81 L84 E85 Y86 D87 N88 K89 T90 Y91 T92 Y93 E94 I95 W99 I100 T101 D105 R106 T107 V108 I109 I110 K111 K112 E113 E114 L117 T118 L119 T120 T121 C122 Y123 P124 F125 D126 Y127 I128 G129 D130 A131 P132 D133 R134 Y135 I136

I137 L141 T142 G143 K147

5 Refinement protocol and experimental data overview (i)

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

Of the 200 calculated structures, 20 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
X-PLOR NIH	structure solution	
X-PLOR NIH	refinement	

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	1431
Number of shifts mapped to atoms	1431
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%

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	859	884	884	65±9
All	All	17180	17680	17680	1296

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 37.

5 of 391 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:21:ILE:HD13	1:A:75:GLN:HE22	0.92	1.21	7	2
1:A:23:LYS:HZ2	1:A:77:LYS:HZ1	0.90	1.07	8	1
1:A:119:LEU:HD12	1:A:120:THR:N	0.90	1.80	1	7
1:A:137:ILE:HD12	1:A:137:ILE:N	0.89	1.80	19	1
1:A:106:ARG:O	1:A:110:ILE:HG23	0.88	1.68	4	14

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	107/147 (73%)	96±2 (89±2%)	8±2 (8±2%)	3±1 (3±1%)	7 41
All	All	2140/2940 (73%)	1915 (89%)	163 (8%)	62 (3%)	7 41

5 of 9 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	57	GLY	15
1	A	56	GLY	14
1	A	107	THR	13
1	A	110	ILE	12
1	A	77	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	95/129 (74%)	81±2 (85±3%)	14±2 (15±3%)	6 45
All	All	1900/2580 (74%)	1623 (85%)	277 (15%)	6 45

5 of 47 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	53	THR	20
1	A	92	THR	20
1	A	110	ILE	20
1	A	120	THR	20
1	A	90	THR	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 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 85% for the well-defined parts and 72% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

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	1431
Number of shifts mapped to atoms	1431
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	13

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	125	0.59 ± 0.29	Should be applied
$^{13}C_{\beta}$	114	0.11 ± 0.14	None needed (< 0.5 ppm)
$^{13}C'$	118	0.55 ± 0.14	Should be applied
^{15}N	121	-0.10 ± 0.41	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 85%, i.e. 1272 atoms were assigned a chemical shift out of a possible 1489. 0 out of 15 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	536/543 (99%)	220/222 (99%)	212/216 (98%)	104/105 (99%)
Sidechain	679/834 (81%)	446/543 (82%)	233/267 (87%)	0/24 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	57/112 (51%)	33/54 (61%)	24/54 (44%)	0/4 (0%)
Overall	1272/1489 (85%)	699/819 (85%)	469/537 (87%)	104/133 (78%)

7.1.4 Statistically unusual chemical shifts (i)

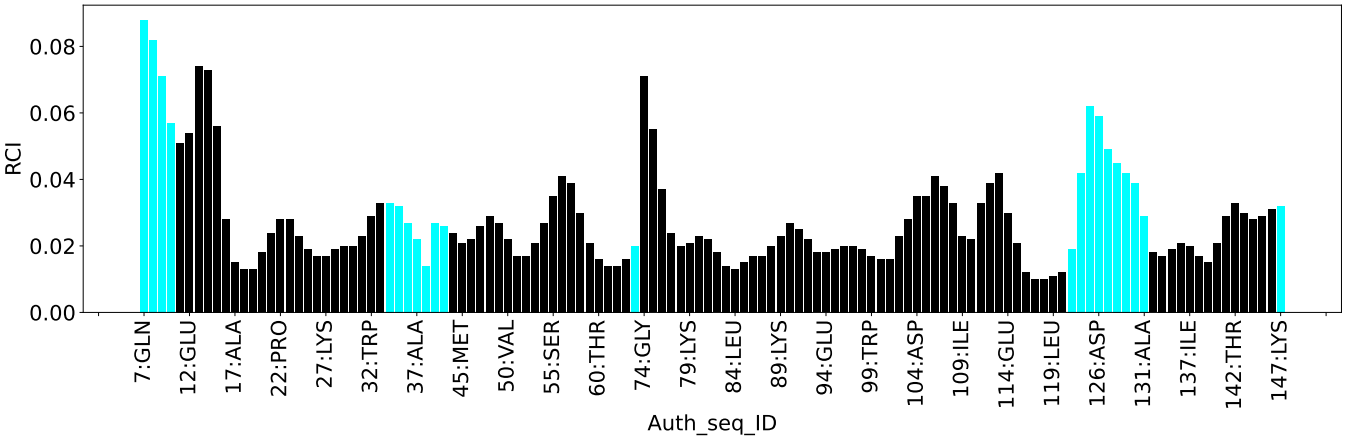
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	93	TYR	CE2	135.82	111.68 – 124.17	14.3
1	A	31	TYR	CE1	131.04	111.24 – 124.66	9.8
1	A	23	LYS	CE	50.00	37.57 – 46.21	9.4
1	A	135	TYR	CE2	106.55	111.68 – 124.17	-9.1
1	A	145	TYR	CE2	107.91	111.68 – 124.17	-8.0
1	A	86	TYR	CE2	108.84	111.68 – 124.17	-7.3
1	A	127	TYR	CE2	109.62	111.68 – 124.17	-6.7
1	A	32	TRP	CZ2	105.28	107.20 – 121.33	-6.4
1	A	10	HIS	HB3	0.87	1.18 – 4.91	-5.8
1	A	47	VAL	HG21	-0.68	-0.58 – 2.19	-5.3
1	A	47	VAL	HG22	-0.68	-0.58 – 2.19	-5.3
1	A	47	VAL	HG23	-0.68	-0.58 – 2.19	-5.3
1	A	136	ILE	CG1	36.30	19.24 – 36.26	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:


RCI
0.00
0.02
0.04
0.06
0.08
7:GLN
12:GLU
17:ALA
22:PRO
27:LYS
32:TRP
37:ALA
45:MET
50:VAL
55:SER
60:THR
74:GLY
79:LYS
84:LEU
89:LYS
94:GLU
99:TRP
104:ASP
109:ILE
114:GLU
119:LEU
126:ASP
131:ALA
137:ILE
142:THR
147:LYS
Auth_seq_ID