wwPDB NMR Structure Validation Summary Report (i)

May 7, 2024 – 12:17 pm BST

PDB ID	:	1W9R
Title	:	Solution Structure of Choline Binding Protein A, Domain R2, the Major Adhesin of Streptococcus pneumoniae
Authors	:	Luo, R.; Mann, B.; Lewis, W.S.; Rowe, A.; Heath, R.; Stewart, M.L.; Hamburger, A.E.; Bjorkman, P.J.; Sivakolundu, S.; Lacy, E.R.; Tuomanen, E.; Kriwacki, R.W.
Deposited on	:	2004-10-15

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:

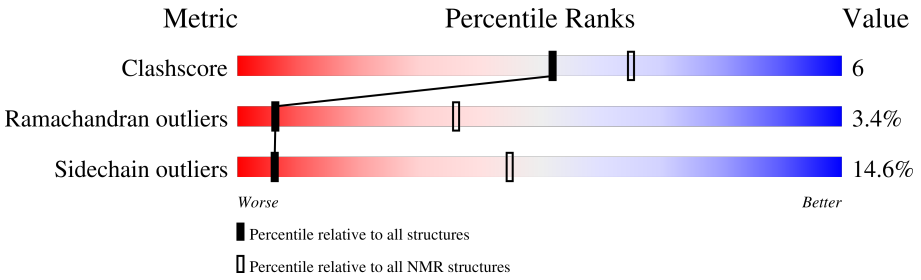
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson 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 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	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	119	73% 10% • 13%

2 Ensemble composition and analysis (i)

This entry contains 19 models. The atoms present in the NMR models are not consistent. Some calculations may have failed as a result. All residues are included in the validation scores. Model 11 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:5-A:100 (96)	1.21	11
2	A:111-A:118 (8)	1.02	2

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 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1965 atoms, of which 998 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called CHOLINE BINDING PROTEIN A.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	119	1965	592	998	171	203	1	0

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: CHOLINE BINDING PROTEIN A

Chain A: 73% 10% • 13%

G1 S2 H3 M4 D30 R31 R32 N33 Y34 P35 T36 N37 T38 Y39 K40 T41 L42 K55 K66 E67 P68 R69 R98 K99 K100 A101 E102 E103 E104 A105 K106 R107 K108 A109 A110 P119

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

The representative model is number 11. Colouring as in section 4.1 above.

• Molecule 1: CHOLINE BINDING PROTEIN A

Chain A: 72% 13% • 13%

G1 S2 H3 M4 A19 K20 Q26 R32 N33 Y34 P35 T36 N37 T38 Y39 K40 V51 K55 A56 E57 E67 P68 R69 R98 K99 K100 A101 E102 E103 E104 A105 K106 R107 K108 A109 A110 P119

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *TAD*.

Of the 200 calculated structures, 19 were deposited, based on the following criterion: *LEAST ENERGY*.

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

Software name	Classification	Version
CNS	refinement	8.0
CNS	structure solution	

No chemical shift data was provided.

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	4.01±14.23	3±12/860 (0.3± 1.4%)	1.71±2.15	11±22/1141 (1.0± 1.9%)
All	All	14.73	55/16333 (0.3%)	2.73	206/21658 (1.0%)

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	A	0.3±1.3	3.7±4.6
All	All	6	70

5 of 55 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	82	GLU	CG-CD	508.70	9.15	1.51	4	1
1	A	87	GLU	CD-OE1	481.28	6.55	1.25	4	1
1	A	87	GLU	CG-CD	450.25	8.27	1.51	4	1
1	A	87	GLU	CD-OE2	435.25	6.04	1.25	4	1
1	A	83	SER	CB-OG	393.32	6.53	1.42	4	1

5 of 123 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(°)	Ideal(°)	Models	
								Worst	Total
1	A	69	ARG	NE-CZ-NH1	-161.14	39.73	120.30	4	5
1	A	87	GLU	OE1-CD-OE2	-86.49	19.51	123.30	4	1
1	A	79	ALA	N-CA-CB	-72.15	9.10	110.10	4	1
1	A	86	ALA	N-CA-CB	-70.77	11.03	110.10	4	1
1	A	77	ALA	N-CA-CB	-65.45	18.47	110.10	4	1

5 of 6 unique chiral outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Models (Total)
1	A	71	GLU	CA	1
1	A	72	GLU	CA	1
1	A	74	VAL	CA	1
1	A	80	GLU	CA	1
1	A	81	VAL	CA	1

5 of 26 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	66	LYS	Peptide	10
1	A	68	PRO	Peptide	7
1	A	37	ASN	Peptide	6
1	A	67	GLU	Peptide	4
1	A	33	ASN	Peptide	3

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	855	888	888	10±36
All	All	16245	16872	16865	188

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 6.

5 of 180 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:50:ASP:HA	1:A:84:LYS:CB	1.62	1.20	4	1
1:A:50:ASP:CA	1:A:84:LYS:HB3	1.59	1.24	4	1
1:A:73:LYS:HG2	1:A:74:VAL:CA	1.56	1.22	4	1
1:A:57:GLU:CD	1:A:81:VAL:HG11	1.52	1.19	4	1
1:A:80:GLU:HG3	1:A:81:VAL:C	1.45	1.25	4	1

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	104/119 (87%)	93±4 (89±4%)	7±3 (7±3%)	3±2 (3±2%)	6 37
All	All	1964/2261 (87%)	1761 (90%)	137 (7%)	66 (3%)	6 36

5 of 16 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	39	TYR	17
1	A	35	PRO	11
1	A	30	ASP	8
1	A	41	THR	6
1	A	68	PRO	5

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	93/103 (90%)	79±3 (85±4%)	14±3 (15±4%)	6 45
All	All	1767/1957 (90%)	1509 (85%)	258 (15%)	6 45

5 of 53 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	36	THR	19
1	A	32	ARG	18
1	A	37	ASN	18
1	A	55	LYS	18
1	A	35	PRO	16

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)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	4-A	7

The worst 5 of 7 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
4	A	86:ALA	C	87:GLU	N	8.48
4	A	87:GLU	C	88:ALA	N	7.99
4	A	80:GLU	C	81:VAL	N	7.48
4	A	83:SER	C	84:LYS	N	6.38
4	A	71:GLU	C	72:GLU	N	5.59

7 Chemical shift validation (i)

No chemical shift data were provided