



# wwPDB NMR Structure Validation Summary Report ⓘ

Apr 26, 2016 – 04:38 PM BST

PDB ID : 1RXL  
Title : Solution structure of the engineered protein Afae-dsc  
Authors : Anderson, K.L.; Billington, J.; Pettigrew, D.; Cota, E; Roversi, P.; Simpson, P.; Chen, H.A.; Urvil, P.; du Merle, L.; Barlow, P.N.; Medof, M.E.; Smith, R.A.; Nowicki, B.; Le Bouguenec, C.; Lea, S.M.; Matthews, S.  
Deposited on : 2003-12-18

This is a wwPDB NMR Structure Validation Summary Report for a publicly released PDB entry.  
We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/NMRValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) 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

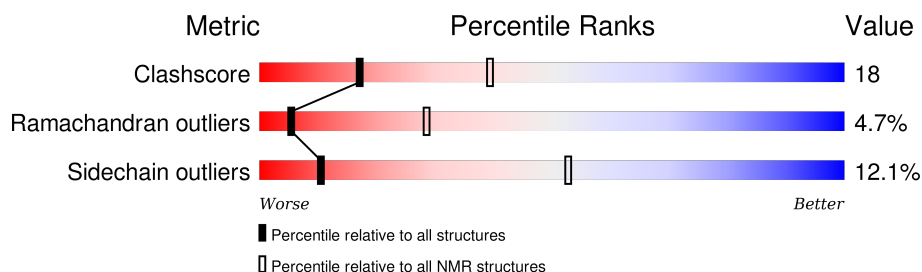
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 84%.

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  $\geq 3$ , 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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	156	

## 2 Ensemble composition and analysis

This entry contains 15 models. Model 15 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:3-A:143 (141)	0.46	15

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 3 single-model clusters were found.

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

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 2137 atoms, of which 1049 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Afimbrial adhesin AFA-III.

Mol	Chain	Residues	Atoms						Trace
1	A	143	Total	C	H	N	O	S	0
			2137	677	1049	188	220	3	

There are 33 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-12	HIS	-	SEE REMARK 999	UNP Q57254
A	-11	HIS	-	SEE REMARK 999	UNP Q57254
A	-10	HIS	-	SEE REMARK 999	UNP Q57254
A	-9	HIS	-	SEE REMARK 999	UNP Q57254
A	-8	HIS	-	SEE REMARK 999	UNP Q57254
A	-7	HIS	-	SEE REMARK 999	UNP Q57254
A	-6	GLY	-	SEE REMARK 999	UNP Q57254
A	-5	LEU	-	SEE REMARK 999	UNP Q57254
A	-4	VAL	-	SEE REMARK 999	UNP Q57254
A	-3	PRO	-	SEE REMARK 999	UNP Q57254
A	-2	ARG	-	SEE REMARK 999	UNP Q57254
A	-1	GLY	-	SEE REMARK 999	UNP Q57254
A	0	SER	-	SEE REMARK 999	UNP Q57254
A	124	ASP	-	SEE REMARK 999	UNP Q57254
A	125	ASN	-	SEE REMARK 999	UNP Q57254
A	126	LYS	-	SEE REMARK 999	UNP Q57254
A	127	GLN	-	SEE REMARK 999	UNP Q57254
A	128	GLY	-	SEE REMARK 999	UNP Q57254
A	129	PHE	-	SEE REMARK 999	UNP Q57254
A	130	THR	-	SEE REMARK 999	UNP Q57254
A	131	PRO	-	SEE REMARK 999	UNP Q57254
A	132	SER	-	SEE REMARK 999	UNP Q57254
A	133	GLY	-	SEE REMARK 999	UNP Q57254
A	134	THR	-	SEE REMARK 999	UNP Q57254
A	135	THR	-	SEE REMARK 999	UNP Q57254
A	136	GLY	-	SEE REMARK 999	UNP Q57254
A	137	THR	-	SEE REMARK 999	UNP Q57254
A	138	THR	-	SEE REMARK 999	UNP Q57254
A	139	LYS	-	SEE REMARK 999	UNP Q57254
A	140	LEU	-	SEE REMARK 999	UNP Q57254
A	141	THR	-	SEE REMARK 999	UNP Q57254

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	142	VAL	-	SEE REMARK 999	UNP Q57254
A	143	THR	-	SEE REMARK 999	UNP Q57254



## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *ARIA-implemented distance geometry/simulated annealing*.

Of the 50 calculated structures, 15 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
ARIA	structure solution	1.2
ARIA	refinement	1.2

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)	BMRB entry 5947
Number of chemical shift lists	1
Total number of shifts	1565
Number of shifts mapped to atoms	1565
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	84%

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	A	0.45±0.04	1±1/1091 (0.1±0.1%)	0.48±0.01	0±0/1486 (0.0±0.0%)
All	All	0.45	10/16365 (0.1%)	0.48	0/22290 (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	A	0.0±0.0	0.7±0.7
All	All	0	10

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	54	PHE	CE1-CZ	7.06	1.50	1.37	13	5
1	A	54	PHE	CE2-CZ	-5.68	1.26	1.37	13	2
1	A	112	TYR	CE2-CZ	5.54	1.45	1.38	11	1
1	A	85	TYR	CE1-CZ	5.19	1.45	1.38	5	1
1	A	112	TYR	CE1-CZ	-5.08	1.31	1.38	10	1

There are no bond-angle outliers.

There are no chirality outliers.

5 of 6 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	49	PHE	Sidechain	4
1	A	85	TYR	Sidechain	2
1	A	112	TYR	Sidechain	1

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Group	Models (Total)
1	A	66	TYR	Sidechain	1
1	A	120	TYR	Sidechain	1

## 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	1070	1037	1037	38±6
All	All	16050	15555	15555	570

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 18.

5 of 305 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:4:GLN:HG3	1:A:133:GLY:HA2	0.92	1.41	8	2
1:A:65:LEU:HB2	1:A:104:GLN:HG3	0.90	1.43	1	2
1:A:123:LYS:HE2	1:A:123:LYS:HA	0.86	1.48	1	2
1:A:41:ALA:HB2	1:A:85:TYR:HB3	0.84	1.48	11	6
1:A:56:LEU:HG	1:A:114:LEU:HD22	0.78	1.55	2	2

## 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	140/156 (90%)	115±3 (82±2%)	19±3 (13±2%)	7±2 (5±1%)	5	28
All	All	2100/2340 (90%)	1721 (82%)	280 (13%)	99 (5%)	5	28

5 of 21 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	108	PRO	15
1	A	93	GLY	13
1	A	109	PRO	11
1	A	21	ASP	9
1	A	52	GLY	9

### 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	115/128 (90%)	101±2 (88±2%)	14±2 (12±2%)	10	53
All	All	1725/1920 (90%)	1516 (88%)	209 (12%)	10	53

5 of 56 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	16	ARG	15
1	A	54	PHE	15
1	A	114	LEU	15
1	A	126	LYS	14
1	A	84	PHE	13

### 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

The completeness of assignment taking into account all chemical shift lists is 84% for the well-defined parts and 84% for the entire structure.

### 7.1 Chemical shift list 1

File name: BMRB entry 5947

Chemical shift list name: *assigned\_chem\_shift\_list\_1*

#### 7.1.1 Bookkeeping

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	1565
Number of shifts mapped to atoms	1565
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	9

#### 7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	141	$-0.70 \pm 0.11$	Should be applied
$^{13}\text{C}_\beta$	124	$-0.82 \pm 0.13$	Should be applied
$^{13}\text{C}'$	137	$0.41 \pm 0.10$	None needed ( $< 0.5$ ppm)
$^{15}\text{N}$	135	$-0.59 \pm 0.39$	None needed (imprecise)

#### 7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 84%, i.e. 1366 atoms were assigned a chemical shift out of a possible 1625. 6 out of 20 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	683/693 (99%)	272/276 (99%)	277/282 (98%)	134/135 (99%)
Sidechain	576/811 (71%)	340/468 (73%)	228/301 (76%)	8/42 (19%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	107/121 (88%)	56/63 (89%)	48/55 (87%)	3/3 (100%)
Overall	1366/1625 (84%)	668/807 (83%)	553/638 (87%)	145/180 (81%)

#### 7.1.4 Statistically unusual chemical shifts ⓘ

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.

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	43	LYS	NZ	117.27	49.86 – 18.16	26.3
1	A	16	ARG	NE	118.29	92.63 – 76.73	21.1
1	A	6	ARG	NE	117.27	92.63 – 76.73	20.5
1	A	86	LYS	HB3	-0.74	3.10 – 0.40	-9.2
1	A	43	LYS	HB3	-0.43	3.10 – 0.40	-8.1
1	A	86	LYS	HG3	-0.58	2.76 – -0.04	-6.9
1	A	78	THR	HB	2.09	5.82 – 2.52	-6.3
1	A	86	LYS	HD3	0.27	2.75 – 0.45	-5.8
1	A	64	LYS	HG3	-0.16	2.76 – -0.04	-5.4

#### 7.1.5 Random Coil Index (RCI) plots ⓘ

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.

Random coil index (RCI) for chain A:

