wwPDB NMR Structure Validation Summary Report (i)

Mar 6, 2022 – 12:46 PM EST

PDB ID	:	2KA3
Title	:	Structure of EMILIN-1 C1Q-like domain
Authors	:	Verdone, G.; Corazza, A.; Colebrooke, S.A.; Cicero, D.O.; Eliseo, T.; Boyd, J.; Doliana, R.; Fogolari, F.; Viglino, P.; Colombatti, A.; Campbell, I.D.; Esposito, G.
Deposited on	:	2008-10-30

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

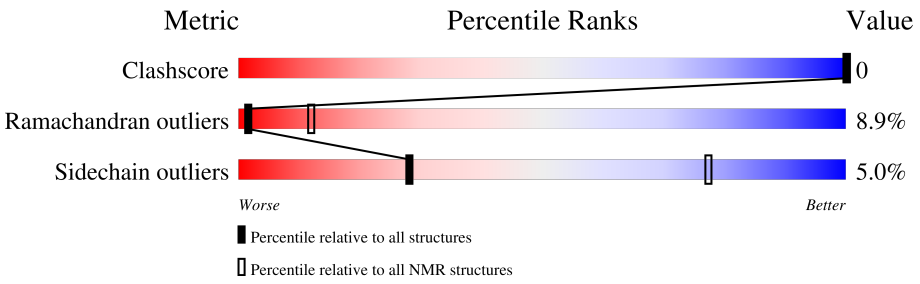
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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	162	71% 9% • 19%
1	B	162	69% 11% • 19%
1	C	162	71% 9% • 19%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 1 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:846-A:926, A:946-A:995, B:846-B:926, B:946-B:995, C:846-C:926, C:946-C:995 (393)	0.17	1

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 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 5805 atoms, of which 2886 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called EMILIN-1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	131	1935	619	962	164	188	2	0
1	B	131	1935	619	962	164	188	2	0
1	C	131	1935	619	962	164	188	2	0

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	834	MET	-	cloning artifact	UNP Q9Y6C2
A	835	ARG	-	cloning artifact	UNP Q9Y6C2
A	836	GLY	-	cloning artifact	UNP Q9Y6C2
A	837	SER	-	cloning artifact	UNP Q9Y6C2
A	838	HIS	-	expression tag	UNP Q9Y6C2
A	839	HIS	-	expression tag	UNP Q9Y6C2
A	840	HIS	-	expression tag	UNP Q9Y6C2
A	841	HIS	-	expression tag	UNP Q9Y6C2
A	842	HIS	-	expression tag	UNP Q9Y6C2
A	843	HIS	-	expression tag	UNP Q9Y6C2
A	844	GLY	-	cloning artifact	UNP Q9Y6C2
A	845	SER	-	cloning artifact	UNP Q9Y6C2
B	834	MET	-	cloning artifact	UNP Q9Y6C2
B	835	ARG	-	cloning artifact	UNP Q9Y6C2
B	836	GLY	-	cloning artifact	UNP Q9Y6C2
B	837	SER	-	cloning artifact	UNP Q9Y6C2
B	838	HIS	-	expression tag	UNP Q9Y6C2
B	839	HIS	-	expression tag	UNP Q9Y6C2
B	840	HIS	-	expression tag	UNP Q9Y6C2
B	841	HIS	-	expression tag	UNP Q9Y6C2
B	842	HIS	-	expression tag	UNP Q9Y6C2
B	843	HIS	-	expression tag	UNP Q9Y6C2
B	844	GLY	-	cloning artifact	UNP Q9Y6C2
B	845	SER	-	cloning artifact	UNP Q9Y6C2
C	834	MET	-	cloning artifact	UNP Q9Y6C2
C	835	ARG	-	cloning artifact	UNP Q9Y6C2
C	836	GLY	-	cloning artifact	UNP Q9Y6C2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	837	SER	-	cloning artifact	UNP Q9Y6C2
C	838	HIS	-	expression tag	UNP Q9Y6C2
C	839	HIS	-	expression tag	UNP Q9Y6C2
C	840	HIS	-	expression tag	UNP Q9Y6C2
C	841	HIS	-	expression tag	UNP Q9Y6C2
C	842	HIS	-	expression tag	UNP Q9Y6C2
C	843	HIS	-	expression tag	UNP Q9Y6C2
C	844	GLY	-	cloning artifact	UNP Q9Y6C2
C	845	SER	-	cloning artifact	UNP Q9Y6C2

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: EMILIN-1

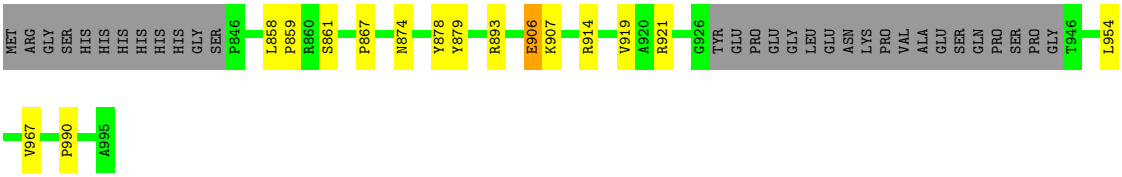
Chain A:

• Molecule 1: EMILIN-1

Chain B:

• Molecule 1: EMILIN-1

Chain C:

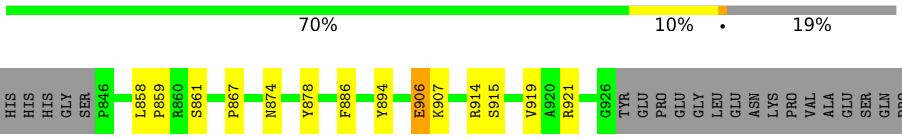


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 1. Colouring as in section 4.1 above.

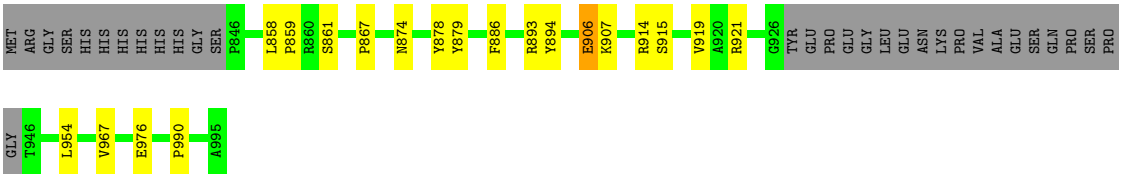
• Molecule 1: EMILIN-1

Chain A:



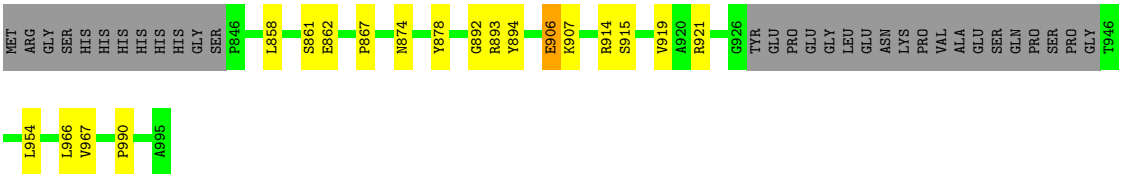
• Molecule 1: EMILIN-1

Chain B:



• Molecule 1: EMILIN-1

Chain C:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing, restrained energy minimization.*

Of the 100 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
X-PLOR	refinement	2.9.9
NAMD	refinement	2.6b2

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	0.93±0.00	0±0/994 (0.0± 0.0%)	1.08±0.01	1±1/1356 (0.1± 0.0%)
1	B	0.93±0.00	0±0/994 (0.0± 0.0%)	1.08±0.01	1±0/1356 (0.1± 0.0%)
1	C	0.93±0.00	0±0/994 (0.0± 0.0%)	1.07±0.01	1±1/1356 (0.1± 0.0%)
All	All	0.93	0/29820 (0.0%)	1.08	39/40680 (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	1.5±0.7
1	B	0.0±0.0	2.5±0.9
1	C	0.0±0.0	2.8±0.7
All	All	0	68

There are no bond-length outliers.

5 of 11 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	B	906	GLU	N-CA-CB	5.95	121.31	110.60	3	10
1	A	874	ASN	N-CA-CB	5.92	121.25	110.60	2	1
1	A	906	GLU	N-CA-CB	5.91	121.23	110.60	4	10
1	C	906	GLU	N-CA-CB	5.88	121.18	110.60	2	8
1	B	874	ASN	N-CA-CB	5.84	121.11	110.60	2	1

There are no chirality outliers.

5 of 14 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	C	893	ARG	Sidechain	10
1	A	921	ARG	Sidechain	9
1	B	921	ARG	Sidechain	9
1	B	893	ARG	Sidechain	8
1	B	879	TYR	Sidechain	6

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	C	973	962	961	0±0
All	All	29190	28860	28830	2

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 0.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:C:954:LEU:HB2	1:C:956:LEU:HD13	0.43	1.91	10	1
1:C:856:LEU:HD11	1:C:858:LEU:HD23	0.41	1.92	7	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	127/162 (78%)	94±1 (74±1%)	22±2 (17±2%)	11±1 (9±1%)	1	12
1	B	127/162 (78%)	95±1 (75±1%)	20±1 (16±1%)	12±1 (9±1%)	1	11
1	C	127/162 (78%)	93±1 (73±1%)	23±1 (18±1%)	11±1 (9±1%)	2	12
All	All	3810/4860 (78%)	2820 (74%)	651 (17%)	339 (9%)	1	12

5 of 43 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	858	LEU	10
1	A	861	SER	10
1	A	867	PRO	10
1	A	874	ASN	10
1	A	906	GLU	10

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	105/131 (80%)	100±1 (95±1%)	6±1 (5±1%)	27	76
1	B	105/131 (80%)	100±1 (95±1%)	6±1 (5±1%)	27	76
1	C	105/131 (80%)	100±1 (95±1%)	5±1 (5±1%)	31	79
All	All	3150/3930 (80%)	2992 (95%)	158 (5%)	28	77

5 of 35 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	914	ARG	10
1	A	954	LEU	10
1	A	976	GLU	10
1	B	914	ARG	10
1	B	954	LEU	10

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)

No chemical shift data were provided