Full wwPDB NMR Structure Validation Report ⓘ

Mar 1, 2022 – 01:16 PM EST

PDB ID	:	2GW9
Title	:	High-resolution solution structure of the mouse defensin Cryptdin4
Authors	:	Rosengren, K.J.; Craik, D.J.; Vogel, H.J.; Daly, N.L.; Ouellette, A.J.
Deposited on	:	2006-05-04

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

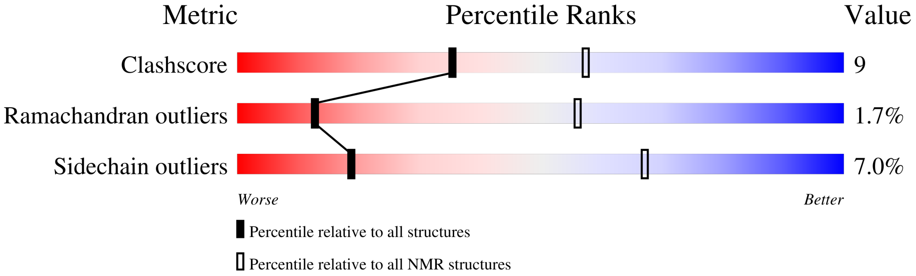
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)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

1 Overall quality at a glance

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	32	

2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 14 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:5-A:30 (26)	0.17	14

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 4 clusters and 3 single-model clusters were found.

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

3 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 525 atoms, of which 266 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Defensin-related cryptdin 4.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	32	525	158	266	57	38	6	0

4 Residue-property plots ⓘ

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: Defensin-related cryptdin 4

4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

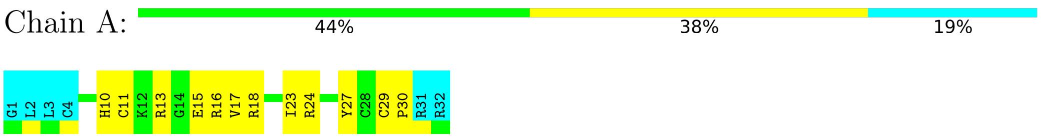
- Molecule 1: Defensin-related cryptdin 4

4.2.2 Score per residue for model 2

- Molecule 1: Defensin-related cryptdin 4

4.2.3 Score per residue for model 3

- Molecule 1: Defensin-related cryptdin 4



4.2.4 Score per residue for model 4

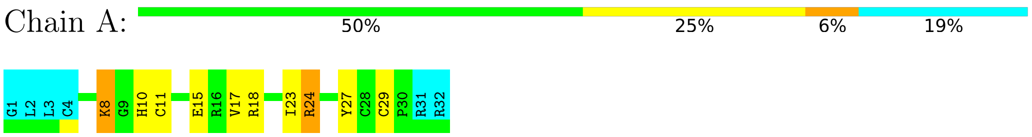
- Molecule 1: Defensin-related cryptdin 4

4.2.5 Score per residue for model 5

- Molecule 1: Defensin-related cryptdin 4

4.2.6 Score per residue for model 6

- Molecule 1: Defensin-related cryptdin 4

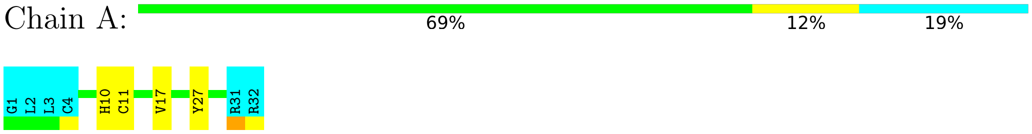


4.2.7 Score per residue for model 7

- Molecule 1: Defensin-related cryptdin 4

4.2.8 Score per residue for model 8

- Molecule 1: Defensin-related cryptdin 4



4.2.9 Score per residue for model 9

- Molecule 1: Defensin-related cryptdin 4

4.2.10 Score per residue for model 10

- Molecule 1: Defensin-related cryptdin 4

4.2.11 Score per residue for model 11

- Molecule 1: Defensin-related cryptdin 4

4.2.12 Score per residue for model 12

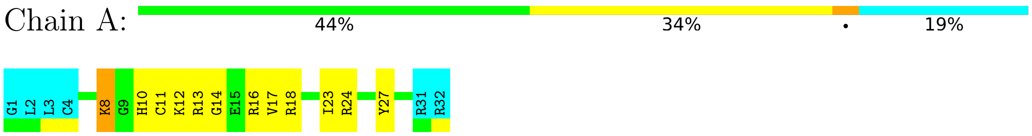
- Molecule 1: Defensin-related cryptdin 4

4.2.13 Score per residue for model 13

- Molecule 1: Defensin-related cryptdin 4

4.2.14 Score per residue for model 14 (medoid)

- Molecule 1: Defensin-related cryptdin 4



4.2.15 Score per residue for model 15

- Molecule 1: Defensin-related cryptdin 4

4.2.16 Score per residue for model 16

- Molecule 1: Defensin-related cryptdin 4

4.2.17 Score per residue for model 17

- Molecule 1: Defensin-related cryptdin 4

4.2.18 Score per residue for model 18

- Molecule 1: Defensin-related cryptdin 4

4.2.19 Score per residue for model 19

- Molecule 1: Defensin-related cryptdin 4

4.2.20 Score per residue for model 20

- Molecule 1: Defensin-related cryptdin 4

5 Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *Simulated annealing using torsion angle dynamics followed by refinement in explicit solvent using Cartesian dynamics within CNS.*.

Of the 50 calculated structures, 20 were deposited, based on the following criterion: *structures with the least restraint violations,structures with the lowest energy*.

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

Software name	**Classification**	**Version**
CYANA	structure solution	2.0
CNS	refinement	1.0

No chemical shift data was provided.

6 Model quality ⓘ

6.1 Standard geometry ⓘ

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 ⓘ

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	210	209	208	4±2
All	All	4200	4180	4160	72

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 9.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:24:ARG:HE	1:A:24:ARG:HA	0.59	1.58	3	1
1:A:11:CYS:SG	1:A:17:VAL:HG22	0.58	2.39	9	14
1:A:8:LYS:HG2	1:A:25:PHE:CE2	0.56	2.36	1	1
1:A:7:ARG:HB2	1:A:10:HIS:O	0.54	2.01	20	1
1:A:12:LYS:HB3	1:A:15:GLU:OE1	0.54	2.03	9	1
1:A:13:ARG:HD2	1:A:13:ARG:H	0.53	1.63	12	1
1:A:18:ARG:HD2	1:A:29:CYS:SG	0.53	2.43	2	4
1:A:8:LYS:HD3	1:A:8:LYS:H	0.53	1.62	19	2
1:A:18:ARG:HA	1:A:18:ARG:NE	0.52	2.19	7	1
1:A:24:ARG:HA	1:A:24:ARG:NE	0.51	2.20	3	2
1:A:18:ARG:HA	1:A:18:ARG:HE	0.50	1.65	7	1
1:A:23:ILE:HD12	1:A:23:ILE:H	0.50	1.67	17	6
1:A:12:LYS:HB3	1:A:15:GLU:CD	0.50	2.26	20	1
1:A:15:GLU:HA	1:A:30:PRO:HA	0.47	1.85	20	6

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:18:ARG:O	1:A:18:ARG:HD2	0.45	2.11	17	1
1:A:8:LYS:HD3	1:A:8:LYS:O	0.45	2.12	14	1
1:A:13:ARG:HA	1:A:13:ARG:NE	0.45	2.26	3	1
1:A:13:ARG:HA	1:A:13:ARG:CZ	0.44	2.41	3	1
1:A:23:ILE:O	1:A:24:ARG:HB2	0.44	2.12	19	10
1:A:11:CYS:HB3	1:A:15:GLU:HB2	0.44	1.89	6	2
1:A:18:ARG:HG3	1:A:29:CYS:SG	0.44	2.52	7	1
1:A:15:GLU:HG2	1:A:30:PRO:HG3	0.44	1.89	13	1
1:A:26:LEU:HD13	1:A:26:LEU:N	0.43	2.28	16	1
1:A:10:HIS:HA	1:A:26:LEU:HD12	0.43	1.89	9	1
1:A:12:LYS:HG2	1:A:15:GLU:OE1	0.43	2.14	4	1
1:A:15:GLU:HG2	1:A:30:PRO:CA	0.43	2.44	20	1
1:A:8:LYS:HA	1:A:24:ARG:O	0.42	2.14	17	2
1:A:7:ARG:HB2	1:A:11:CYS:SG	0.42	2.54	16	1
1:A:12:LYS:O	1:A:14:GLY:N	0.41	2.52	13	2
1:A:7:ARG:HD2	1:A:11:CYS:HA	0.41	1.91	12	1
1:A:24:ARG:NE	1:A:24:ARG:HA	0.40	2.31	11	1
1:A:18:ARG:NE	1:A:18:ARG:HA	0.40	2.31	15	1

6.3 Torsion angles ⓘ

6.3.1 Protein backbone ⓘ

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	26/32 (81%)	23±1 (88±3%)	3±1 (11±3%)	0±0 (2±2%)	13	56
All	All	520/640 (81%)	456 (88%)	55 (11%)	9 (2%)	13	56

All 2 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	13	ARG	5
1	A	14	GLY	4

6.3.2 Protein sidechains ⓘ

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	22/27 (81%)	20±1 (93±5%)	2±1 (7±5%)	19	67
All	All	440/540 (81%)	409 (93%)	31 (7%)	19	67

All 10 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	10	HIS	12
1	A	8	LYS	4
1	A	16	ARG	4
1	A	24	ARG	3
1	A	13	ARG	2
1	A	18	ARG	2
1	A	7	ARG	1
1	A	12	LYS	1
1	A	27	TYR	1
1	A	26	LEU	1

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.6 Ligand geometry ⓘ

There are no ligands in this entry.

6.7 Other polymers ⓘ

There are no such molecules in this entry.

6.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

7 Chemical shift validation ⓘ

No chemical shift data were provided