Full wwPDB NMR Structure Validation Report

Jun 25, 2024 – 06:51 AM EDT

PDB ID	:	6PX7
BMRB ID	:	30647
Title	:	Dg12a in Weaponisation 'on the fly': Convergent recruitment of knottin and defensin scaffolds as neurotoxins in the venom of assassin fly Dolopus genitalis (Diptera: Asilidae)
Authors	:	Agwa, A.J.; Schroeder, C.
Deposited on	:	2019-07-25

This is a Full wwPDB NMR Structure Validation 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:

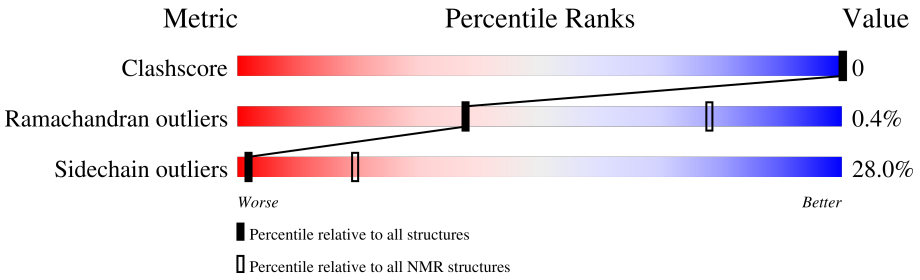
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 82%.

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	34	68% 18% 15%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 17 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:6-A:34 (29)	0.04	17

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. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 473 atoms, of which 205 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Venom polypeptide.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	34	473	163	205	49	50	6	0

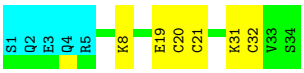
4 Residue-property plots

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: Venom polypeptide

Chain A:



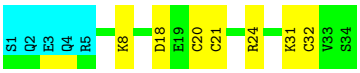
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Venom polypeptide

Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: Venom polypeptide

Chain A:

4.2.3 Score per residue for model 3

• Molecule 1: Venom polypeptide

Chain A:

4.2.4 Score per residue for model 4

• Molecule 1: Venom polypeptide

Chain A:

4.2.5 Score per residue for model 5

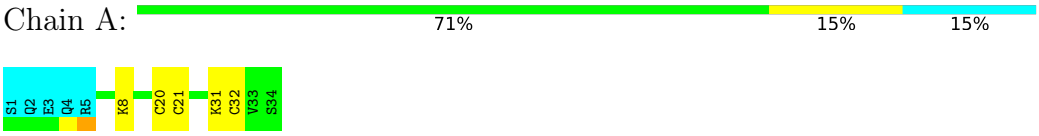
• Molecule 1: Venom polypeptide

Chain A:

4.2.6 Score per residue for model 6

• Molecule 1: Venom polypeptide

Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Venom polypeptide

Chain A:

4.2.8 Score per residue for model 8

• Molecule 1: Venom polypeptide

Chain A:

4.2.9 Score per residue for model 9

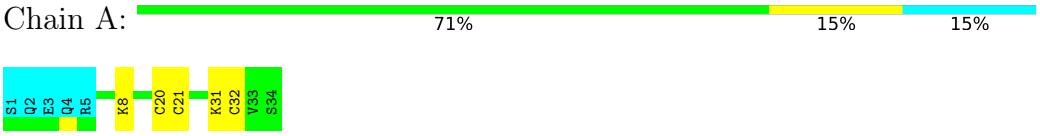
• Molecule 1: Venom polypeptide

Chain A:

4.2.10 Score per residue for model 10

• Molecule 1: Venom polypeptide

Chain A:



4.2.11 Score per residue for model 11

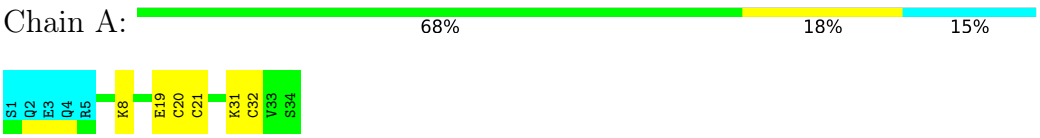
• Molecule 1: Venom polypeptide

Chain A:

4.2.12 Score per residue for model 12

• Molecule 1: Venom polypeptide

Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: Venom polypeptide

Chain A:

4.2.14 Score per residue for model 14

• Molecule 1: Venom polypeptide

Chain A:

4.2.15 Score per residue for model 15

• Molecule 1: Venom polypeptide

Chain A:

4.2.16 Score per residue for model 16

• Molecule 1: Venom polypeptide

Chain A:

4.2.17 Score per residue for model 17 (medoid)

• Molecule 1: Venom polypeptide

Chain A:

4.2.18 Score per residue for model 18

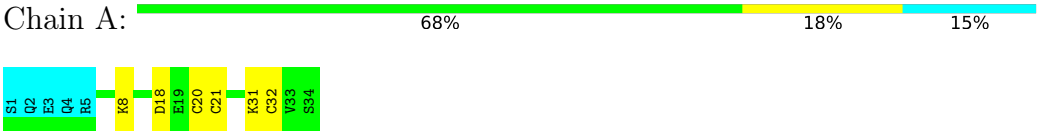
• Molecule 1: Venom polypeptide

Chain A:

4.2.19 Score per residue for model 19

• Molecule 1: Venom polypeptide

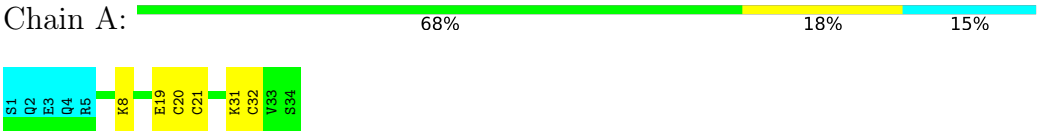
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: Venom polypeptide

Chain A:



5 Refinement protocol and experimental data overview

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

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

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

Software name	Classification	Version
TopSpin	refinement	3.5
CYANA	structure calculation	3.97

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	365
Number of shifts mapped to atoms	315
Number of unparsed shifts	0
Number of shifts with mapping errors	50
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	82%

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
All	All	4480	3500	4300	-

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 -.

There are no clashes.

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	28/34 (82%)	26±0 (92±1%)	2±0 (7±1%)	0±0 (0±1%)	38	78
All	All	560/680 (82%)	517 (92%)	41 (7%)	2 (0%)	38	78

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	6	GLN	2

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	25/30 (83%)	18±1 (72±5%)	7±1 (28±5%)	2 19
All	All	500/600 (83%)	360 (72%)	140 (28%)	2 19

All 12 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	8	LYS	20
1	A	20	CYS	20
1	A	21	CYS	20
1	A	31	LYS	20
1	A	32	CYS	20
1	A	19	GLU	11
1	A	18	ASP	8
1	A	24	ARG	6
1	A	9	LYS	5
1	A	22	SER	4
1	A	23	LYS	4
1	A	6	GLN	2

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

The completeness of assignment taking into account all chemical shift lists is 82% for the well-defined parts and 79% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *starch_ output*

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	365
Number of shifts mapped to atoms	315
Number of unparsed shifts	0
Number of shifts with mapping errors	50
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. All 50 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	1	SER	HB2	4.021	0.001	.
1	A	2	GLN	HB2	2.126	0.001	.
1	A	2	GLN	HG2	2.393	0.001	.
1	A	3	GLU	HB2	2.089	0.002	.
1	A	3	GLU	HG2	2.464	0.002	.
1	A	4	GLN	HB2	2.111	0.001	.
1	A	4	GLN	HG2	2.359	0.001	.
1	A	5	ARG	HB2	1.869	0.003	.
1	A	5	ARG	HG2	1.649	0.004	.
1	A	5	ARG	HD2	3.227	0.002	.
1	A	6	GLN	HB2	2.143	0.006	.
1	A	6	GLN	HG2	2.428	0.012	.
1	A	7	CYS	HB2	3.189	0.003	.
1	A	8	LYS	HB2	1.982	0.001	.

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	8	LYS	HG2	1.636	0.002	.
1	A	8	LYS	HD2	1.736	0.001	.
1	A	8	LYS	HE2	3.138	0.003	.
1	A	9	LYS	HB2	2.029	0.003	.
1	A	9	LYS	HG2	1.529	0.002	.
1	A	9	LYS	HD2	1.62	0.004	.
1	A	9	LYS	HE2	3.068	0.002	.
1	A	10	ILE	HG12	1.588	0.003	.
1	A	12	GLU	HB2	2.254	0.002	.
1	A	12	GLU	HG2	2.395	0.001	.
1	A	13	HIS	HB2	3.316	0.003	.
1	A	14	CYS	HB2	2.927	0.003	.
1	A	15	TYR	HB2	3.197	0.002	.
1	A	18	ASP	HB2	2.863	0.001	.
1	A	19	GLU	HB2	2.084	0.003	.
1	A	19	GLU	HG2	2.596	0.004	.
1	A	20	CYS	HB2	3.488	0.004	.
1	A	21	CYS	HB2	3.268	0.003	.
1	A	22	SER	HB2	4.157	0.011	.
1	A	23	LYS	HB2	2.139	0.004	.
1	A	23	LYS	HG2	1.46	0.003	.
1	A	23	LYS	HD2	1.829	0.003	.
1	A	23	LYS	HE2	3.054	0.004	.
1	A	24	ARG	HB2	1.617	0.004	.
1	A	24	ARG	HG2	1.303	0.002	.
1	A	24	ARG	HD2	3.11	0.003	.
1	A	25	CYS	HB2	2.641	0.001	.
1	A	26	LEU	HB2	1.945	0.002	.
1	A	27	PHE	HB2	3.181	0.002	.
1	A	28	TYR	HB2	3.22	0.002	.
1	A	31	LYS	HB2	1.478	0.015	.
1	A	31	LYS	HG2	1.099	0.003	.
1	A	31	LYS	HD2	1.586	0.002	.
1	A	31	LYS	HE2	2.912	0.004	.
1	A	32	CYS	HB2	3.145	0.002	.
1	A	34	SER	HB2	3.922	0.006	.

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	34	-0.16 ± 0.33	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	33	0.20 ± 0.49	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	33	1.01 ± 1.32	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 82%, i.e. 304 atoms were assigned a chemical shift out of a possible 373. 0 out of 3 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	117/146 (80%)	59/59 (100%)	29/58 (50%)	29/29 (100%)
Sidechain	173/192 (90%)	118/124 (95%)	55/60 (92%)	0/8 (0%)
Aromatic	14/35 (40%)	14/17 (82%)	0/17 (0%)	0/1 (0%)
Overall	304/373 (82%)	191/200 (96%)	84/135 (62%)	29/38 (76%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 79%, i.e. 353 atoms were assigned a chemical shift out of a possible 446. 0 out of 3 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	135/171 (79%)	68/69 (99%)	34/68 (50%)	33/34 (97%)
Sidechain	204/240 (85%)	139/153 (91%)	65/74 (88%)	0/13 (0%)
Aromatic	14/35 (40%)	14/17 (82%)	0/17 (0%)	0/1 (0%)
Overall	353/446 (79%)	221/239 (92%)	99/159 (62%)	33/48 (69%)

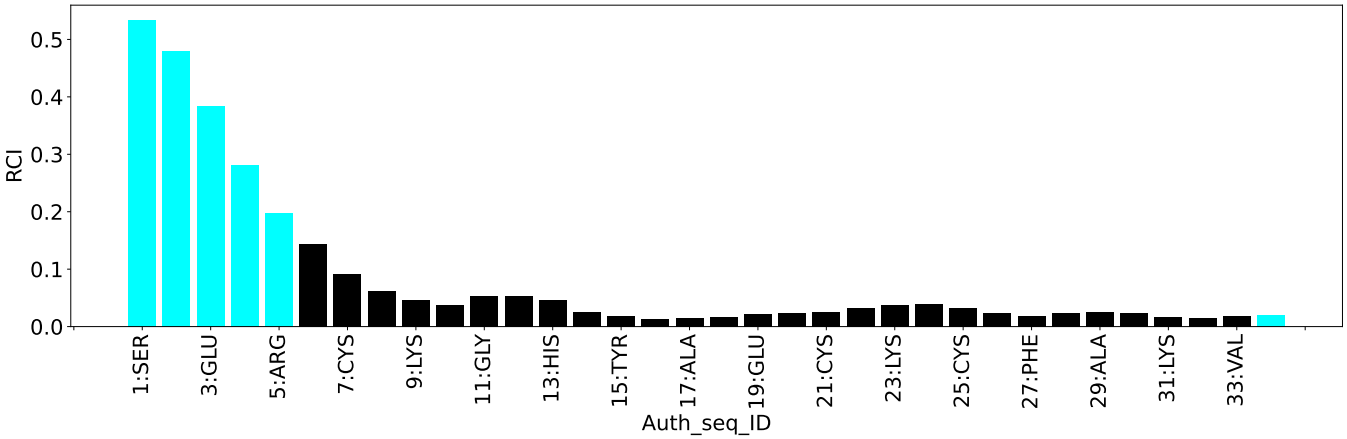
7.1.4 Statistically unusual chemical shifts

There are no statistically unusual chemical shifts.

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.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
0.0
0.1
0.2
0.3
0.4
0.5
1:SER
3:GLU
5:ARG
7:CYS
9:LYS
11:GLY
13:HIS
15:TYR
17:ALA
19:GLU
21:CYS
23:LYS
25:CYS
27:PHE
29:ALA
31:LYS
33:VAL
Auth_seq_ID