W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report (i)

Oct 5, 2024 – 03:33 pm BST

PDB ID	:	1UND
BMRB ID	:	5966
Title	:	Solution structure of the human advillin C-terminal headpiece subdomain
Authors	:	Vermeulen, W.; Van Troys, M.; Vanhaesebrouck, P.; Verschueren, M.; Fant, F.; Ampe, C.; Martins, J.; Borremans, F.
Deposited on	:	2003-09-09

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

|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
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)	:	2.39

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.
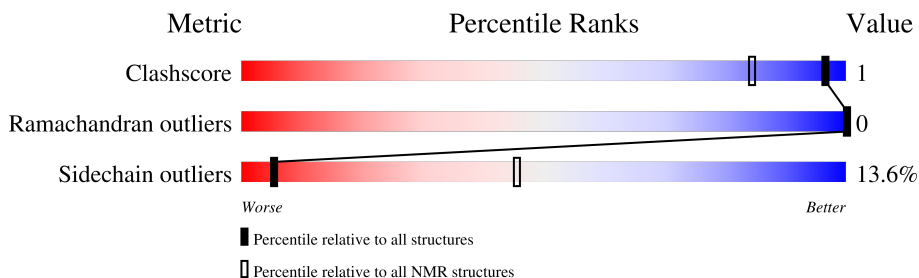
Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	210492	14027
Ramachandran outliers	207382	12486
Sidechain outliers	206894	12463

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	37	62% 14% 24%

2 Ensemble composition and analysis ⓘ

This entry contains 25 models. Model 19 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:32 (28)	0.28	19

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 5 clusters and 1 single-model cluster was found.

Cluster number	Models
1	3, 4, 5, 7, 8, 9, 11, 15, 17, 18, 19, 20, 25
2	1, 6, 22, 24
3	10, 13, 16
4	12, 23
5	14, 21
Single-model clusters	2

3 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 599 atoms, of which 302 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called ADVILLIN.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	37	599	196	302	48	52	1	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.
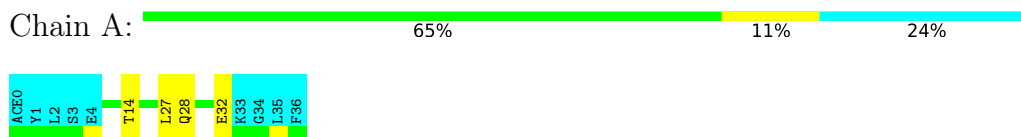
• Molecule 1: ADVILLIN

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: ADVILLIN

Chain A:

4.2.2 Score per residue for model 2
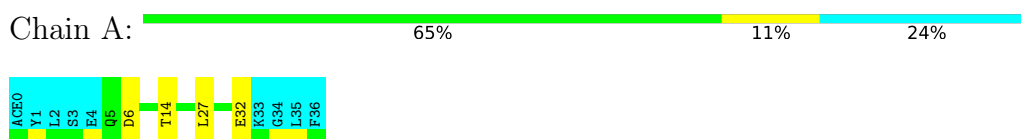
• Molecule 1: ADVILLIN

Chain A:

4.2.3 Score per residue for model 3

• Molecule 1: ADVILLIN

Chain A:

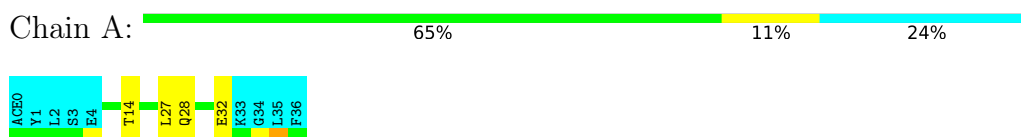#### 4.2.4 Score per residue for model 4

• Molecule 1: ADVILLIN

Chain A:

4.2.5 Score per residue for model 5

• Molecule 1: ADVILLIN

Chain A:

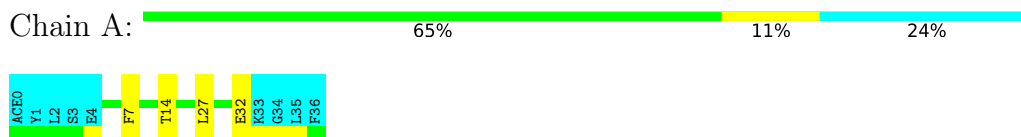#### 4.2.6 Score per residue for model 6

• Molecule 1: ADVILLIN

Chain A:

4.2.7 Score per residue for model 7

• Molecule 1: ADVILLIN

Chain A:

4.2.8 Score per residue for model 8

• Molecule 1: ADVILLIN

Chain A:

4.2.9 Score per residue for model 9

• Molecule 1: ADVILLIN

Chain A:
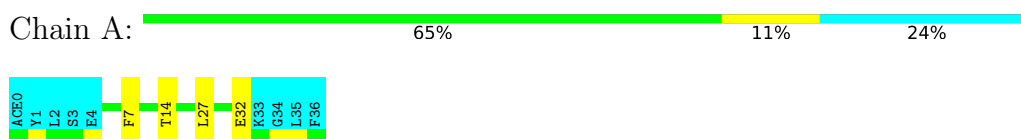
4.2.10 Score per residue for model 10

• Molecule 1: ADVILLIN

Chain A:

4.2.11 Score per residue for model 11

• Molecule 1: ADVILLIN

Chain A:

4.2.12 Score per residue for model 12

• Molecule 1: ADVILLIN

Chain A:

4.2.13 Score per residue for model 13

• Molecule 1: ADVILLIN

Chain A:

4.2.14 Score per residue for model 14

• Molecule 1: ADVILLIN

Chain A:

4.2.15 Score per residue for model 15

• Molecule 1: ADVILLIN

Chain A:

4.2.16 Score per residue for model 16

• Molecule 1: ADVILLIN

Chain A:
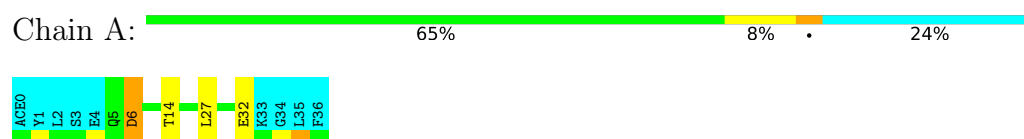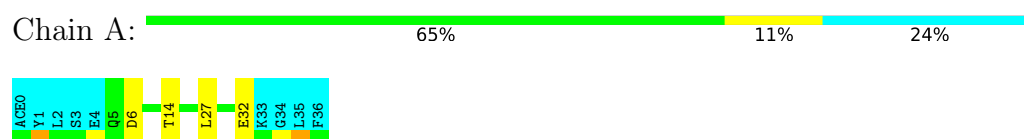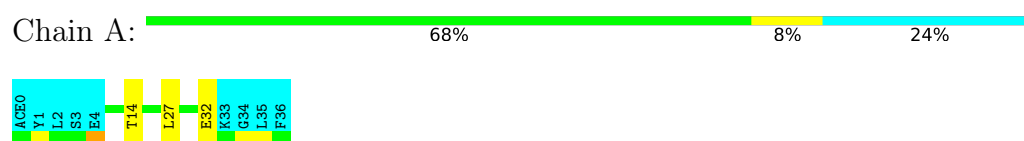
4.2.17 Score per residue for model 17

• Molecule 1: ADVILLIN

Chain A:

4.2.18 Score per residue for model 18

• Molecule 1: ADVILLIN

Chain A: 65%   11%   24%

ACE0 Y1 L2 S3 E4 Q5 D6 T14 L27 E32 K33 G34 L35 F36
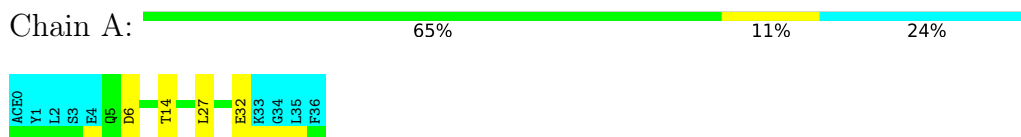
4.2.19 Score per residue for model 19 (medoid)

• Molecule 1: ADVILLIN

Chain A: 62%   14%   24%

ACE0 Y1 L2 S3 E4 Q5 D6 T14 L27 Q28 E32 K33 G34 L35 F36

4.2.20 Score per residue for model 20

• Molecule 1: ADVILLIN

Chain A: 59%   16%   24%

ACE0 Y1 L2 S3 E4 Q5 D6 T14 L27 Q28 K31 E32 K33 G34 L35 F36

4.2.21 Score per residue for model 21

• Molecule 1: ADVILLIN

Chain A: 59%   14%   •   24%

ACE0 Y1 L2 S3 E4 Q5 D6 F7 T14 L27 Q28 E32 K33 G34 L35 F36

4.2.22 Score per residue for model 22

• Molecule 1: ADVILLIN

Chain A: 62%   14%   24%

ACE0 Y1 L2 S3 E4 Q5 D6 T14 L27 Q28 E32 K33 G34 L35 F36

4.2.23 Score per residue for model 23

• Molecule 1: ADVILLIN

Chain A:

4.2.24 Score per residue for model 24

• Molecule 1: ADVILLIN

Chain A:

4.2.25 Score per residue for model 25

• Molecule 1: ADVILLIN

Chain A:

5 Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *torsion angle dynamics*.

Of the 500 calculated structures, 25 were deposited, based on the following criterion: *target function*.

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

Software name	Classification	Version
Insight II	refinement	
DYANA	structure solution	

No chemical shift data was provided.

6   Model quality ⓘ

6.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: ACE

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	0.98±0.01	1±0/232 (0.6± 0.2%)	0.89±0.01	0±0/311 (0.0± 0.0%)
All	All	0.98	33/5800 (0.6%)	0.89	0/7775 (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	0.4±0.5
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	32	GLU	CD-OE2	10.31	1.36	1.25	2	25
1	A	6	ASP	CG-OD2	5.09	1.37	1.25	23	8

There are no bond-angle outliers.

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	7	PHE	Sidechain	10

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	226	231	229	0±1
All	All	5650	5775	5725	8

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 1.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
1:A:7:PHE:CZ	1:A:11:PHE:CD2	0.87	2.61	3	1
1:A:7:PHE:CZ	1:A:11:PHE:CE2	0.81	2.69	3	1
1:A:7:PHE:CE2	1:A:11:PHE:CE2	0.79	2.71	3	1
1:A:7:PHE:CE2	1:A:11:PHE:CZ	0.62	2.88	3	1
1:A:7:PHE:HZ	1:A:18:PHE:CD1	0.54	2.21	10	1
1:A:7:PHE:CZ	1:A:18:PHE:CD1	0.51	2.99	10	1
1:A:7:PHE:CZ	1:A:18:PHE:CG	0.50	3.00	10	1
1:A:7:PHE:CE2	1:A:11:PHE:CD2	0.42	3.08	13	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	28/37 (76%)	28±0 (99±2%)	0±0 (1±2%)	0±0 (0±0%)	100	100
All	All	700/925 (76%)	694 (99%)	6 (1%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	23/30 (77%)	20±1 (86±4%)	3±1 (14±4%)	5	45
All	All	575/750 (77%)	497 (86%)	78 (14%)	5	45

All 7 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	14	THR	25
1	A	27	LEU	25
1	A	6	ASP	13
1	A	28	GLN	12
1	A	15	ARG	1
1	A	31	LYS	1
1	A	32	GLU	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 oligosaccharides 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 (i)

No chemical shift data were provided