wwPDB NMR Structure Validation Summary Report (i)

Feb 14, 2022 – 09:57 AM EST

PDB ID	:	1J5H
Title	:	Solution Structure of Apo-Neocarzinostatin
Authors	:	Urbaniak, M.D.; Muskett, F.W.; Finucane, M.D.; Caddick, S.; Woolfson, D.N.
Deposited on	:	2002-05-02

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:

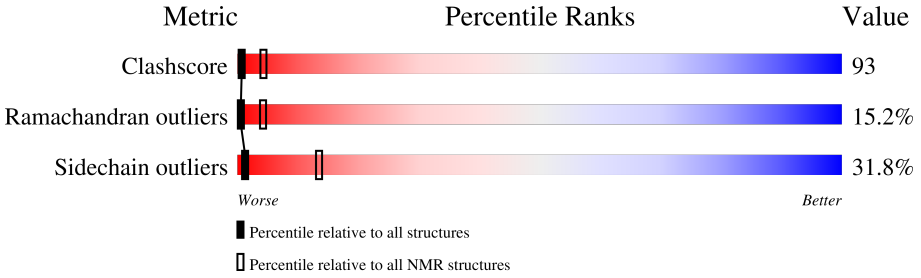
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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	122	12% 52% 20% 7% 7%

2 Ensemble composition and analysis (i)

This entry contains 44 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:4-A:78, A:82-A:99, A:103-A:113 (104)	0.58	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 7 clusters and 6 single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 5, 11, 12, 16, 19, 21, 22, 24, 25, 29, 37, 40
2	8, 14, 18, 23, 33, 36, 39
3	6, 7, 15, 28, 38, 41
4	10, 20, 26
5	13, 17
6	35, 43
7	9, 31
Single-model clusters	27; 30; 32; 34; 42; 44

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1515 atoms, of which 738 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Apo-Neocarzinostatin.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	113	1515	476	738	132	165	4	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-8	HIS	-	expression tag	UNP P0A3R9
A	-7	HIS	-	expression tag	UNP P0A3R9
A	-6	HIS	-	expression tag	UNP P0A3R9
A	-5	HIS	-	expression tag	UNP P0A3R9
A	-4	HIS	-	expression tag	UNP P0A3R9
A	-3	HIS	-	expression tag	UNP P0A3R9
A	-2	LEU	-	expression tag	UNP P0A3R9
A	-1	GLN	-	expression tag	UNP P0A3R9
A	0	GLY	-	expression tag	UNP P0A3R9

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: Apo-Neocarzinostatin

Chain A: 12% 52% 20% 7% 7%

HIS HIS HIS HIS HIS HIS LEU GLN GLY A1 A2 P3 T4 A5 T6 V7 T8 P9 S10 S11 G12 L13 S14 D15 G16 T17 V18 V19 K20 V21 A22 G23 A24 G25 L26 Q27 A28 G29 T30 A31 Y32 D33 V34 G35 Q36 C37 A38 W39 V40 D41 T42 G43 V44 L45 A46 C47 N48 P49 A50 D51

F52 S53 S54 V55 T56 A57 D58 A59 N60 G61 S62 A63 S64 T65 S66 L67 T68 V69 R70 R71 S72 F73 E74 G75 F76 L77 F78 D79 G80 T81 R82 W83 G84 T85 V86 D87 C88 T89 T90 A91 A92 C93 Q94 V95 G96 L97 S98 D99 A100 A101 G102 N103 G104 P105 E106 G107 V108 A109 I110 S111

F112 N113

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

The representative model is number 14. Colouring as in section 4.1 above.

- Molecule 1: Apo-Neocarzinostatin

Chain A: 24% 35% 25% • 7% 7%

HIS HIS HIS HIS HIS HIS LEU GLN GLY A1 A2 P3 T4 A5 T6 V7 T8 P9 S10 S11 G12 L13 S14 D15 G16 T17 V18 V19 K20 V21 A22 G23 A24 G25 L26 Q27 A28 G29 T30 A31 Y32 D33 V34 G35 Q36 W39 V40 D41 T42 L45 N48 P49 A50 D51 F52 S53 S54

V55 T56 A57 D58 A59 N60 G61 S62 A63 S64 T65 S66 L67 T68 V69 R70 R71 S72 F73 E74 L77 F78 D79 G80 T81 R82 W83 G84 T85 V86 D87 C88 T89 T90 A91 A92 C93 Q94 V95 G96 L97 S98 D99 A100 A101 G102 N103 G104 P105 E106 S111 F112 N113

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 44 were deposited, based on the following criterion: *structures with favorable non-bond energy.*

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

Software name	Classification	Version
DYANA	refinement	

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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 (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	727	693	684	131±13
All	All	31988	30492	30100	5747

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 93.

5 of 1377 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:69:VAL:HG12	1:A:112:PHE:CE2	1.08	1.83	44	35
1:A:34:VAL:HG21	1:A:67:LEU:HD21	1.05	1.21	28	3
1:A:83:TRP:CZ3	1:A:86:VAL:HG21	1.04	1.86	42	1
1:A:55:VAL:HG11	1:A:97:LEU:HD22	1.03	1.23	34	1
1:A:32:TYR:CG	1:A:97:LEU:HD21	1.02	1.89	4	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	103/122 (84%)	66±3 (64±3%)	22±3 (21±3%)	16±3 (15±3%)	0 4
All	All	4532/5368 (84%)	2886 (64%)	957 (21%)	689 (15%)	0 4

5 of 53 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	24	ALA	44
1	A	59	ALA	44
1	A	28	ALA	43
1	A	53	SER	38
1	A	103	ASN	33

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	77/88 (88%)	52±3 (68±4%)	25±3 (32±4%)	1 13
All	All	3388/3872 (88%)	2309 (68%)	1079 (32%)	1 13

5 of 62 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	THR	44
1	A	99	ASP	44
1	A	111	SER	37
1	A	4	THR	36
1	A	51	ASP	36

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