wwPDB NMR Structure Validation Summary Report (i)

May 28, 2020 – 07:36 pm BST

PDB ID	:	1GPX
Title	:	C85S GAPDX, NMR, 20 STRUCTURES
Authors	:	Pochapsky, T.C.; Kuti, M.; Kazanis, S.
Deposited on	:	1998-06-10

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 (1)) were used in the production of this report:

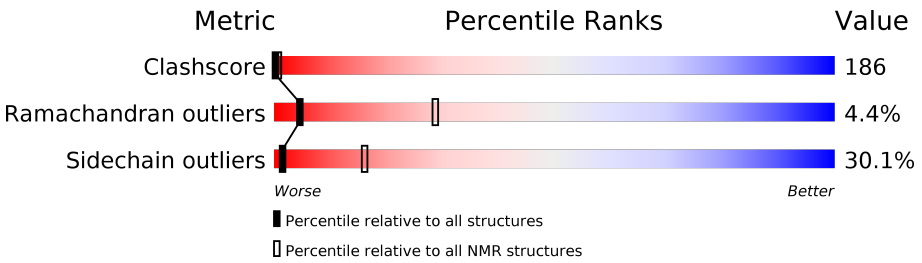
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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	106	8% 59% 22% • 8%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 15 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:1-A:33, A:39-A:102 (97)	0.41	15

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 2 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1565 atoms, of which 770 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called PUTIDAREDOXIN.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	106	1564	488	770	136	162	8	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	85	SER	CYS	ENGINEERED MUTATION	UNP P00259

- Molecule 2 is GALLIUM (III) ION (three-letter code: GA) (formula: Ga).

Mol	Chain	Residues	Atoms	
			Total	Ga
2	A	1	1	1

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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: PUTIDAREDOXIN

Chain A: 8% 59% 22% • 8%

S1 K2 V3 V4 Y5 V6 S7 H8 D9 G10 T11 R12 R13 E14 L15 D16 V17 A18 D19 G20 V21 S22 L23 M24 Q25 A26 A27 V28 S29 N30 G31 I32 Y33 D34 I35 V36 G37 D38 C39 G40 G41 S42 A43 S44 C45 A46 T47 C48 H49 V50 Y51 V52 N53 E54 A55 F56 T57 D58 K59 V60

P61 A62 A63 N64 E65 R66 E67 I68 G69 M70 L71 E72 C73 V74 T75 A76 E77 L78 K79 P80 N81 S82 R83 L84 S85 C86 Q87 I88 I89 M90 T91 P92 E93 L94 D95 G96 I97 V98 V99 D100 V101 P102 D103 R104 Q105 W106

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

The representative model is number 15. Colouring as in section 4.1 above.

- Molecule 1: PUTIDAREDOXIN

Chain A: 8% 57% 25% • 8%

S1 K2 V3 V4 Y5 V6 S7 H8 D9 G10 T11 R12 R13 E14 L15 D16 V17 A18 D19 G20 V21 S22 L23 M24 Q25 A26 A27 V28 S29 N30 G31 I32 Y33 D34 I35 V36 G37 D38 C39 G40 G41 S42 A43 S44 C45 A46 T47 C48 H49 V50 Y51 V52 N53 E54 A55 F56 T57 D58 K59 V60

P61 A62 A63 N64 E65 R66 E67 I68 G69 M70 L71 E72 C73 V74 T75 A76 E77 L78 K79 P80 N81 S82 R83 L84 S85 C86 Q87 I88 I89 M90 E93 L94 D95 G96 I97 V98 V99 D100 V101 P102 D103 R104 Q105 W106

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DG, SIMULATED ANNEALING*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *LEAST RESTRAINT VIOLATION*.

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

Software name	Classification	Version
X-PLOR	refinement	3.85
X-PLOR	structure solution	3.85

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	716	704	704	264±16
All	All	14340	14080	14080	5276

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 186.

5 of 1351 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:24:MET:SD	1:A:86:CYS:SG	1.47	1.47	1	7
1:A:24:MET:SD	1:A:39:CYS:SG	1.30	2.28	7	2
1:A:70:MET:CE	1:A:84:LEU:HD12	1.12	1.75	7	18
1:A:45:CYS:O	1:A:84:LEU:HD11	1.07	1.49	20	19
1:A:71:LEU:HD13	1:A:78:LEU:HD11	1.05	1.20	12	2

5.2 Torsion angles (i)

5.2.1 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	96/106 (91%)	71±2 (74±2%)	21±2 (22±2%)	4±1 (4±1%)	4 29
All	All	1920/2120 (91%)	1421 (74%)	414 (22%)	85 (4%)	4 29

5 of 16 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	75	THR	18
1	A	54	GLU	15
1	A	78	LEU	14
1	A	25	GLN	8
1	A	28	VAL	8

5.2.2 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	81/89 (91%)	57±3 (70±4%)	24±3 (30±4%)	1 16
All	All	1620/1780 (91%)	1132 (70%)	488 (30%)	1 16

5 of 59 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	1	SER	20
1	A	29	SER	20
1	A	71	LEU	20
1	A	22	SER	20
1	A	11	THR	18

5.2.3 RNA (i)

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains (i)

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

5.4 Carbohydrates (i)

There are no carbohydrates in this entry.

5.5 Ligand geometry (i)

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

5.6 Other polymers (i)

There are no such molecules in this entry.

5.7 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Chemical shift validation (i)

No chemical shift data were provided