wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 07:46 PM BST

PDB ID	:	1YUJ
Title	:	SOLUTION NMR STRUCTURE OF THE GAGA FACTOR/DNA COMPLEX, 50 STRUCTURES
Authors	:	Clore, G.M.; Omichinski, J.G.; Gronenborn, A.M.
Deposited on	:	1996-12-31

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
http://wwpdb.org/validation/2016/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:

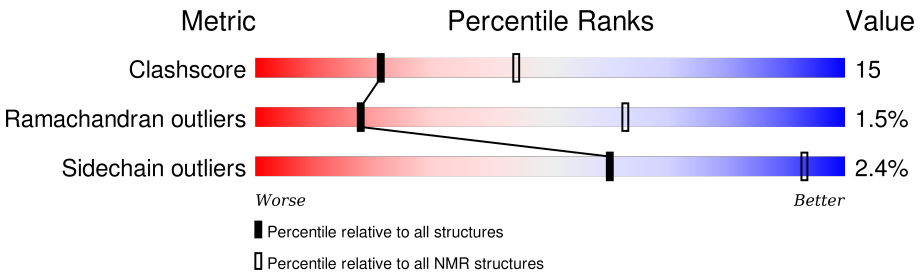
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	B	11	9% 82% 9%
2	C	11	27% 55% 18%
3	A	54	56% 28% 17%

2 Ensemble composition and analysis (i)

This entry contains 50 models. Model 5 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:15-A:59 (45)	0.28	5

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 8 clusters and 5 single-model clusters were found.

Cluster number	Models
1	1, 8, 10, 11, 21, 26, 35, 36, 40
2	4, 5, 17, 22, 27, 44, 48, 49
3	9, 13, 16, 18, 28, 32, 38, 47
4	19, 20, 34, 37, 39, 41, 50
5	2, 7, 23, 42, 45, 46
6	6, 15, 31
7	12, 29
8	3, 24
Single-model clusters	14; 25; 30; 33; 43

3 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 1578 atoms, of which 697 are hydrogens and 0 are deuteriums.

- Molecule 1 is a DNA chain called DNA (5'-D(*GP*CP*CP*GP*AP*GP*AP*GP*TP*AP*C)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	B	11	347	107	122	46	62	10	0

- Molecule 2 is a DNA chain called DNA (5'-D(*GP*TP*AP*CP*TP*CP*TP*CP*GP*GP*C)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	C	11	344	106	124	38	66	10	0

- Molecule 3 is a protein called GAGA-FACTOR.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
3	A	54	886	268	451	93	72	2	0

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

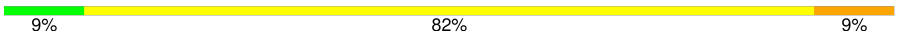
Mol	Chain	Residues	Atoms	
			Total	Zn
4	A	1	1	1

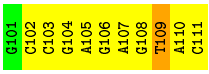
4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

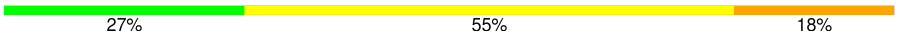
These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain 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: DNA (5'-D(*GP*CP*CP*GP*AP*GP*AP*GP*TP*AP*C)-3')

Chain B: 

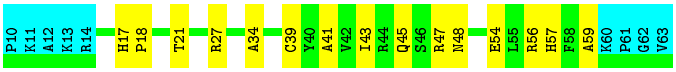


- Molecule 2: DNA (5'-D(*GP*TP*AP*CP*TP*CP*TP*CP*GP*GP*C)-3')

Chain C: 

- Molecule 3: GAGA-FACTOR

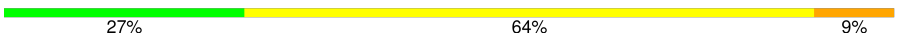
Chain A:

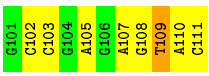


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 5. Colouring as in section 4.1 above.

- Molecule 1: DNA (5'-D(*GP*CP*CP*GP*AP*GP*AP*GP*TP*AP*C)-3')

Chain B: 



- Molecule 2: DNA (5'-D(*GP*TP*AP*CP*TP*CP*TP*CP*GP*GP*C)-3')

Chain C: 45% 45% 9%

G112 T113 T116 C117 T118 C119 G120 G121 C122

- Molecule 3: GAGA-FACTOR

Chain A: 48% 33% • 17%

P10 K11 A12 K13 R14 H17 P18 T21 R27 P33 A34 T35 C36 C39 Y40 A41 V42 I43 R44 Q45 S46 R47 N48 E54 I55 R56 H57 F58 A59 K60 P61 G62 V63

5 Refinement protocol and experimental data overview (i)

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

Of the ? calculated structures, 50 were deposited, based on the following criterion: *?*.

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

Software name	Classification	Version
X-PLOR (SEE ABOVE) ABOVE)	structure solution	ABOVE)
X-PLOR (SEE ABOVE) ABOVE)	refinement	ABOVE)

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

6 Model quality (i)

6.1 Standard geometry (i)

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

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	B	1.07±0.02	1±0/253 (0.2±0.2%)	1.83±0.01	2±1/389 (0.6±0.3%)
2	C	1.19±0.03	3±1/245 (1.0±0.3%)	1.87±0.01	5±1/376 (1.2±0.4%)
3	A	1.09±0.01	0±0/376 (0.0±0.0%)	0.84±0.01	0±0/508 (0.0±0.0%)
All	All	1.11	156/43700 (0.4%)	1.53	344/63650 (0.5%)

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
2	C	118	DT	C5-C7	6.97	1.54	1.50	44	45
2	C	113	DT	C5-C7	6.66	1.54	1.50	49	40
2	C	116	DT	C5-C7	6.51	1.53	1.50	41	43
1	B	109	DT	C5-C7	6.35	1.53	1.50	13	28

5 of 20 unique angle 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(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	C	116	DT	C6-C5-C7	-6.82	118.81	122.90	39	43
2	C	113	DT	C6-C5-C7	-6.73	118.86	122.90	11	37
2	C	118	DT	C6-C5-C7	-6.54	118.97	122.90	35	11
2	C	118	DT	O4'-C1'-N1	6.20	112.34	108.00	39	25
2	C	117	DC	O4'-C1'-N1	6.18	112.33	108.00	41	12

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	B	225	122	124	6±2
2	C	220	124	126	4±1
3	A	366	368	368	15±4
All	All	40600	30700	30900	1106

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 15.

5 of 189 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
3:A:54:GLU:O	3:A:59:ALA:HB2	0.81	1.74	18	50
3:A:21:THR:O	3:A:21:THR:HG23	0.70	1.87	11	19
3:A:21:THR:HG23	3:A:21:THR:O	0.69	1.87	2	17
3:A:45:GLN:HE22	3:A:47:ARG:NH2	0.65	1.89	16	10
3:A:45:GLN:NE2	3:A:47:ARG:NH2	0.65	2.44	16	11

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	
3	A	45/54 (83%)	39±1 (87±3%)	5±1 (12±3%)	1±1 (1±1%)	18	63
All	All	2250/2700 (83%)	1954 (87%)	263 (12%)	33 (1%)	18	63

5 of 7 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
3	A	33	PRO	20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
3	A	40	TYR	5
3	A	19	PRO	3
3	A	59	ALA	2
3	A	37	PRO	1

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
3	A	40/47 (85%)	39±1 (98±3%)	1±1 (2±3%)	60 93
All	All	2000/2350 (85%)	1953 (98%)	47 (2%)	60 93

5 of 11 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)
3	A	21	THR	13
3	A	47	ARG	12
3	A	27	ARG	5
3	A	25	ARG	3
3	A	23	LYS	3

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 carbohydrates in this entry.

6.6 Ligand geometry (i)

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

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