wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 09:05 PM BST

PDB ID	:	2EZG
Title	:	SOLUTION STRUCTURE OF A COMPLEX OF THE THIRD DNA BINDING DOMAIN OF HUMAN HMG-I(Y) BOUND TO DNA DODECAMER CONTAINING THE PRDII SITE OF THE INTERFERON-BETA PROMOTER, NMR, 35 STRUCTURES
Authors	:	Clore, G.M.; Huth, J.R.; Bewley, C.; Gronenborn, A.M.
Deposited on	:	1997-06-04

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:

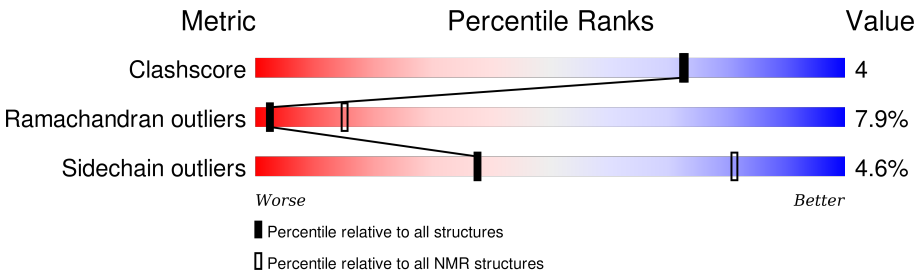
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	12	75% 25%
2	C	12	67% 33%
3	A	10	90% 10%

2 Ensemble composition and analysis (i)

This entry contains 35 models.

Cyrange was unable to find well-defined residues.

Error message: Only domains with < 8 residues could be identified.

NmrClust was unable to cluster the ensemble.

Error message: Wrapper check: not enough residues in core to run NmrClust

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 940 atoms, of which 372 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	B	12	380	117	137	45	70	11	0

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	C	12	380	117	137	45	70	11	0

- Molecule 3 is a protein called HIGH MOBILITY GROUP PROTEIN HMG-I/HMG-Y.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
3	A	10	180	50	98	22	10	0

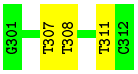
4 Residue-property plots (i)

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.

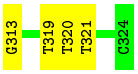
- Molecule 1: DNA (5'-D(*GP*GP*GP*AP*AP*AP*TP*TP*CP*CP*TP*C)-3')

Chain B:



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

Chain C:



- Molecule 3: HIGH MOBILITY GROUP PROTEIN HMG-I/HMG-Y

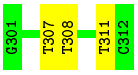
Chain A:

4.2 Residue scores for the first model from the NMR ensemble

No representative models were identified. Colouring as in section 4.1 above.

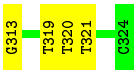
- Molecule 1: DNA (5'-D(*GP*GP*GP*AP*AP*AP*TP*TP*CP*CP*TP*C)-3')

Chain B:



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

Chain C:



- Molecule 3: HIGH MOBILITY GROUP PROTEIN HMG-I/HMG-Y

Chain A:

5 Refinement protocol and experimental data overview (i)

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

Of the 35 calculated structures, 35 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	refinement	3.1
XPLOR MODIFIED	structure solution	MODIFIED

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)

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	1.16±0.04	2±1/272 (0.8±0.3%)	1.65±0.02	5±1/418 (1.1±0.3%)
2	C	1.18±0.04	2±1/272 (0.9±0.3%)	1.64±0.02	4±1/418 (1.1±0.3%)
3	A	1.12±0.03	0±0/83 (0.0±0.0%)	0.88±0.04	0±0/106 (0.0±0.0%)
All	All	1.17	158/21945 (0.7%)	1.58	316/32970 (1.0%)

5 of 6 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	B	307	DT	C5-C7	7.67	1.54	1.50	26	26
2	C	319	DT	C5-C7	7.45	1.54	1.50	25	26
1	B	308	DT	C5-C7	7.16	1.54	1.50	11	26
2	C	321	DT	C5-C7	7.12	1.54	1.50	15	29
2	C	320	DT	C5-C7	6.99	1.54	1.50	15	26

5 of 17 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
1	B	311	DT	C6-C5-C7	-7.07	118.66	122.90	14	26
1	B	307	DT	C6-C5-C7	-6.85	118.79	122.90	10	23
1	B	308	DT	C6-C5-C7	-6.76	118.84	122.90	30	29
2	C	319	DT	C6-C5-C7	-6.75	118.85	122.90	2	29
2	C	321	DT	C6-C5-C7	-6.61	118.93	122.90	34	23

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	243	137	137	1±1
2	C	243	137	137	2±1
3	A	82	98	97	1±1
All	All	19880	13020	12985	137

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 4.

5 of 37 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:301:DG:H8	1:B:301:DG:HO5'	0.94	0.95	33	5
2:C:313:DG:HO5'	2:C:313:DG:H8	0.90	1.04	14	19
2:C:313:DG:H8	2:C:313:DG:HO5'	0.86	1.11	31	12
2:C:313:DG:H8	2:C:313:DG:O5'	0.65	1.75	23	11
2:C:313:DG:O5'	2:C:313:DG:H8	0.63	1.76	24	13

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	8/10 (80%)	7±1 (83±8%)	1±1 (9±8%)	1±1 (8±8%)	2	15
All	All	280/350 (80%)	233 (83%)	25 (9%)	22 (8%)	2	15

All 2 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	35	PRO	14

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
3	A	39	PRO	8

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	8/8 (100%)	8±0 (95±6%)	0±0 (5±6%)	38 81
All	All	280/280 (100%)	267 (95%)	13 (5%)	38 81

All 4 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	33	ARG	4
3	A	34	LYS	4
3	A	40	LYS	4
3	A	41	LYS	1

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)

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