



wwPDB NMR Structure Validation Summary Report ⓘ

Feb 18, 2018 – 02:54 am GMT

PDB ID : 1J5N
Title : Solution Structure of the Non-Sequence-Specific HMGB protein NHP6A in complex with SRY DNA
Authors : Masse, J.E.; Wong, B.; Yen, Y.-M.; Allain, F.H.-T.; Johnson, R.C.; Feigon, J.
Deposited on : 2002-05-15

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 ⓘ symbol.

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

Cyrange : Kirchner and Güntert (2011)
NmrClust : Kelley et al. (1996)
MolProbity : 4.02b-467
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
RCI : v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV : Wang et al. (2010)
ShiftChecker : trunk30686
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk30686

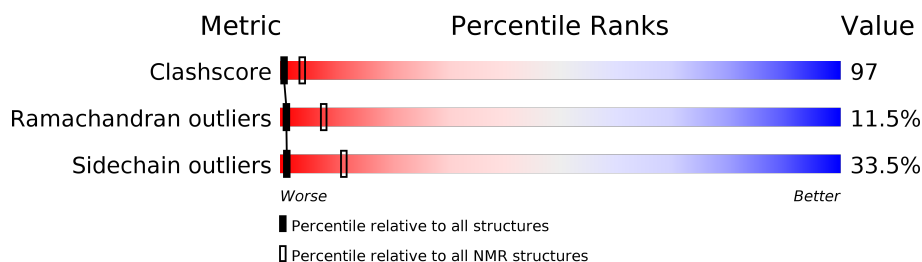
1 Overall quality at a glance

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	136279	12091
Ramachandran outliers	132675	10835
Sidechain outliers	132484	10811

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	B	15	
2	C	15	
3	A	93	

2 Ensemble composition and analysis

This entry contains 20 models. Model 1 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:17-A:88 (72)	0.26	1

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 3 clusters and 2 single-model clusters were found.

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

3 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 2489 atoms, of which 1119 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
1	B	15	Total	C	H	N	O	P	0
			484	149	172	58	91	14	

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

Mol	Chain	Residues	Atoms						Trace
2	C	15	Total	C	H	N	O	P	0
			466	143	169	55	85	14	

- Molecule 3 is a protein called Nonhistone chromosomal protein 6A.

Mol	Chain	Residues	Atoms						Trace
3	A	93	Total	C	H	N	O	S	0
			1539	478	778	138	143	2	

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: 5'-D(*GP*GP*GP*GP*TP*GP*AP*TP*TP*GP*TP*TP*CP*AP*G)-3'

Chain B: 

G101	G102	G103	G104	T105	G106	A107	T108	T109	G110	T111	T112	C113	A114	G115
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

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

Chain C: 

C116	T117	G118	A119	A120	C121	A122	A123	T124	C125	A126	C127	C128	C129	C130
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

- Molecule 3: Nonhistone chromosomal protein 6A

Chain A: 

W1	W2	T3	P4	B5	B6	P7	K8	K9	R10	T11	T12	B13	K14	K15	K16	D17	P18	N19	A20	F21	K22	K23	A24	L25	S26	A27	Y28	M29	F30	F31	A32	N33	E34	N35	R36	D37	T38	V39	R40	S41	E42	N43	P44	D45	L46	T47	F48	C49	O50	N51	G52	K53	K54	L55	G56	E57	K58	N59	K60
----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

A61	L62	T63	P64	E65	E66	K67	Q68	P69	Y70	A71	A72	K73	A74	Q75	A76	D77	K78	K79	R80	R81	E82	S83	E84	K85	E86	L87	Y88	M89	A90	T91	L92	A93
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

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

The representative model is number 1. Colouring as in section 4.1 above.

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

Chain B: 

G101	G102	G103	G104	T105	G106	A107	T108	T109	G110	T111	T112	C113	A114	G115
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

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



C116
T117
G118
A119
A120
C121
A122
T124
C125
A126
C127
C128
C129
C130

● Molecule 3: Nonhistone chromosomal protein 6A



M1	V2	T3	P4	R5	E6	P7	K8	K9	R10	T11	T12	R13	K14	K15	K16	D17	P18	M19	A20	P21	K22	R23	A24	L25	S26	A27	Y28	M29	F30	F31	A32	N33	E34	N35	R36	D37	T38	V39	R40	S41	E42	M43	P44	D45	I46	T47	F48	G49	Q50	V51	G52	K53	K54	L55	G56	E57	K58	W59	K60																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
A61	L62	T63	P64	E65	E66	K67	Q68	P69	Y70	E71	A72	K73	A74	D77	K78	K79	R80	Y81	E82	S83	E84	K85	E86	L87	Y88	N89	A90	T91	L92	A93																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												

5 Refinement protocol and experimental data overview

The models were refined using the following method: *simulated annealing*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *structures with acceptable covalent geometry*.

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

Software name	Classification	Version
X-PLOR	refinement	3.1

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	1.08±0.01	0±0/350 (0.0±0.1%)	2.24±0.02	35±1/541 (6.5±0.2%)
2	C	1.09±0.02	0±0/332 (0.0±0.0%)	2.22±0.02	25±1/508 (5.0±0.1%)
3	A	0.21±0.00	0±0/604 (0.0±0.0%)	0.32±0.00	0±0/812 (0.0±0.0%)
All	All	0.80	2/25720 (0.0%)	1.69	1208/37220 (3.2%)

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	B	109	DT	C5-C7	5.12	1.53	1.50	3	2

5 of 71 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(°)	Ideal(°)	Models	
								Worst	Total
2	C	124	DT	O4'-C1'-N1	10.30	115.21	108.00	16	20
1	B	109	DT	O4'-C1'-N1	10.20	115.14	108.00	15	20
1	B	108	DT	O4'-C1'-N1	10.02	115.02	108.00	13	20
1	B	103	DG	O4'-C1'-N9	9.68	114.78	108.00	20	20
1	B	107	DA	O4'-C1'-N9	9.39	114.57	108.00	16	20

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	312	172	172	28±4
2	C	297	169	169	38±7
3	A	590	579	579	170±9
All	All	23980	18400	18400	4125

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 97.

5 of 788 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
3:A:47:THR:O	3:A:51:VAL:HG12	0.93	1.63	19	11
3:A:59:TRP:O	3:A:62:LEU:HD12	0.93	1.64	5	19
2:C:116:DC:H4'	2:C:117:DT:OP1	0.90	1.66	14	3
3:A:23:ARG:NH2	3:A:24:ALA:HB3	0.90	1.80	8	1
2:C:126:DA:H4'	2:C:127:DC:OP1	0.89	1.67	13	6

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	72/93 (77%)	53±2 (73±3%)	11±2 (15±3%)	8±2 (12±3%)	1	8
All	All	1440/1860 (77%)	1055 (73%)	219 (15%)	166 (12%)	1	8

5 of 17 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	23	ARG	20
3	A	63	THR	20
3	A	17	ASP	18
3	A	62	LEU	18
3	A	36	ARG	16

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	61/80 (76%)	41±2 (66±3%)	20±2 (34±3%)	1	11
All	All	1220/1600 (76%)	811 (66%)	409 (34%)	1	11

5 of 48 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	81	TYR	20
3	A	80	ARG	20
3	A	70	TYR	20
3	A	23	ARG	19
3	A	87	LEU	19

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 ⓘ

There are no chain breaks in this entry.

7 Chemical shift validation

No chemical shift data were provided