



wwPDB NMR Structure Validation Summary Report ⓘ

Feb 13, 2017 – 01:56 am GMT

PDB ID : 2OEH
Title : Determination of the Three-dimensional Structure of the Mrf2-DNA Complex
Using Paramagnetic Spin Labeling
Authors : Cai, S.; Zhu, L.; Zhang, Z.
Deposited on : 2006-12-29

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 ⓘ 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	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	trunk28760
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

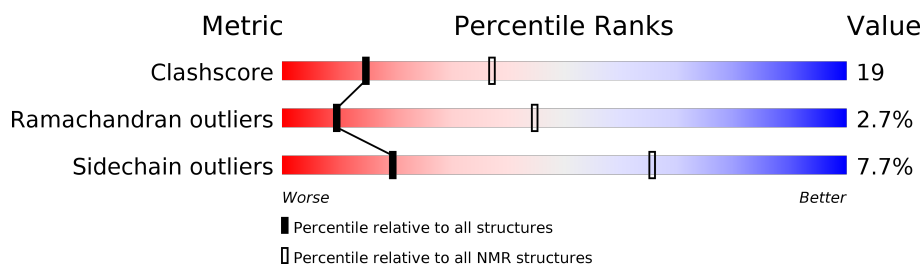
1 Overall quality at a glance

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	125131	11601
Ramachandran outliers	121729	10391
Sidechain outliers	121581	10367

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	93% 7%
2	C	14	93% 7%
3	A	107	59% 11% • 29%

2 Ensemble composition and analysis

This entry contains 17 models. Model 12 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:1-A:28 (28)	0.05	12
2	A:34-A:60 (27)	0.07	12
3	A:77-A:97 (21)	0.06	3

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, 6, 7, 8, 11, 12, 13, 15
2	5, 9, 10, 14, 17
Single-model clusters	16

3 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 2732 atoms, of which 1243 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
1	B	15	Total	C	H	N	O	P	0
			479	147	171	57	89	15	

- Molecule 2 is a DNA chain called 5'-D(P*CP*GP*AP*CP*GP*TP*TP*AP*TP*AP*TP*TP*GP*T)-3'.

Mol	Chain	Residues	Atoms						Trace
2	C	14	Total	C	H	N	O	P	0
			450	138	162	48	88	14	

- Molecule 3 is a protein called AT-rich interactive domain-containing protein 5B.

Mol	Chain	Residues	Atoms						Trace
3	A	107	Total	C	H	N	O	S	0
			1803	575	910	160	155	3	

4 Residue-property plots

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(P*TP*AP*CP*AP*AP*TP*AP*TP*AP*AP*CP*GP*TP*CP*G)-3'

Chain B: 

T1
A2
C3
A4
A5
T6
A7
T8
A9
A10
C11
G12
T13
C14
G15

- Molecule 2: 5'-D(P*CP*GP*AP*CP*GP*TP*TP*AP*TP*AP*TP*TP*GP*T)-3'

Chain C: 

G16
G17
A18
C19
G20
G21
T22
A23
A24
A25
T26
T27
G28
T29

- Molecule 3: AT-rich interactive domain-containing protein 5B

Chain A: 

R1
F7
L8
R18
R24
Y27
L28
G29
F30
K31
Q32
I33
N34
R55
Y61
D62
E63
L64
G65
G66
R67
P68
G69
S70
T71
S72
A73
A74
T75
C76
H80
I85
Y88
D87
K96
P99
L100
P101
P102
I103
K104
P105
R106
K107

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

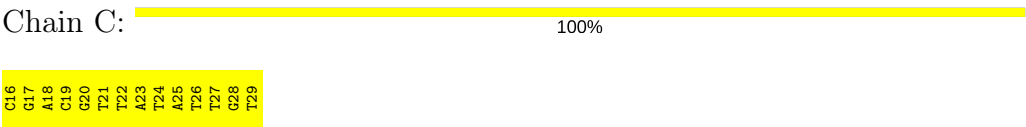
The representative model is number 12. Colouring as in section 4.1 above.

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

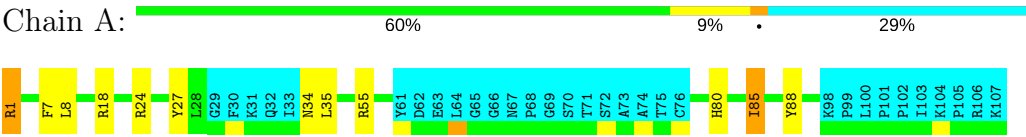
Chain B: 

T1
A2
C3
A4
A5
T6
A7
T8
A9
A10
C11
G12
T13
C14
G15

- Molecule 2: 5'-D(P*CP*GP*AP*CP*GP*TP*TP*AP*TP*AP*TP*TP*GP*T)-3'



● Molecule 3: AT-rich interactive domain-containing protein 5B



5 Refinement protocol and experimental data overview ⓘ

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

Of the 1000 calculated structures, 17 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	Classification	Version
HADDOCK	structure solution	
HADDOCK	refinement	

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

6.1 Standard geometry

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	0.88±0.01	0±0/345 (0.0±0.0%)	1.17±0.02	1±0/528 (0.2±0.1%)
2	C	0.83±0.00	0±0/321 (0.0±0.0%)	1.34±0.00	2±1/492 (0.4±0.2%)
3	A	0.77±0.00	0±0/681 (0.0±0.0%)	1.01±0.00	4±0/917 (0.4±0.1%)
All	All	0.81	0/22899 (0.0%)	1.14	114/32929 (0.3%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	B	0.0±0.0	0.2±0.4
3	A	0.0±0.0	0.1±0.2
All	All	0	4

There are no bond-length outliers.

5 of 15 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
1	B	1	DT	OP1-P-O3'	6.32	119.11	105.20	14	2
3	A	24	ARG	NE-CZ-NH1	5.92	123.26	120.30	14	17
3	A	18	ARG	NE-CZ-NH1	5.56	123.08	120.30	5	17
3	A	1	ARG	NE-CZ-NH1	5.55	123.08	120.30	8	17
3	A	55	ARG	NE-CZ-NH1	5.27	122.93	120.30	15	14

There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	B	1	DT	Sidechain	3
3	A	25	ILE	Mainchain	1

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	308	171	170	22±2
2	C	288	162	161	20±2
3	A	663	673	671	4±2
All	All	21403	17102	17034	730

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 19.

5 of 110 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:6:DT:OP1	3:A:25:ILE:HG23	0.83	1.73	5	1
1:B:1:DT:H1'	1:B:2:DA:H5'	0.70	1.63	2	17
1:B:4:DA:O3'	3:A:26:PRO:HD2	0.62	1.95	15	1
1:B:6:DT:OP1	3:A:80:HIS:CD2	0.60	2.55	14	1
2:C:21:DT:OP1	3:A:55:ARG:HB2	0.59	1.98	3	2

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	75/107 (70%)	68±1 (90±2%)	5±1 (7±2%)	2±1 (3±1%)	10	45
All	All	1275/1819 (70%)	1148 (90%)	93 (7%)	34 (3%)	10	45

All 4 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	85	ILE	17
3	A	34	ASN	12
3	A	26	PRO	3
3	A	35	LEU	2

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	67/92 (73%)	62±1 (92±1%)	5±1 (8±1%)	19	65
All	All	1139/1564 (73%)	1051 (92%)	88 (8%)	19	65

5 of 9 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	8	LEU	17
3	A	88	TYR	17
3	A	7	PHE	17
3	A	1	ARG	17
3	A	27	TYR	7

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

No chemical shift data were provided