wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 08:38 PM BST

PDB ID	:	2FF0
Title	:	Solution Structure of Steroidogenic Factor 1 DNA Binding Domain Bound to its Target Sequence in the Inhibin alpha-subunit Promoter
Authors	:	Little, T.H.; Zhang, Y.; Matulis, C.K.; Weck, J.; Zhang, Z.; Ramachandran, A.; Mayo, K.E.; Radhakrishnan, I.
Deposited on	:	2005-12-17

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:

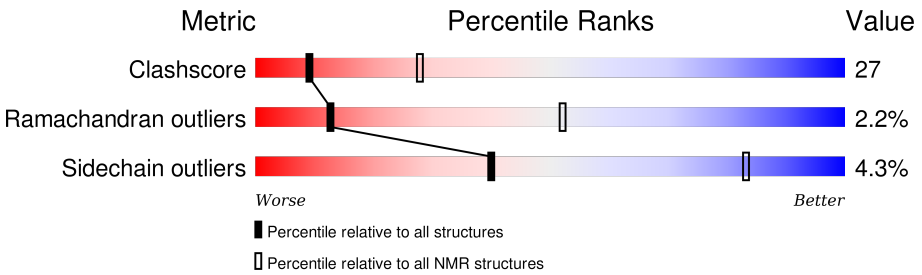
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	15	100%
2	C	15	100%
3	A	102	73% 19% • 8%

2 Ensemble composition and analysis (i)

This entry contains 16 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:10-A:103 (94)	0.71	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 3 single-model clusters were found.

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

3 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2613 atoms, of which 1178 are hydrogens and 0 are deuteriums.

- Molecule 1 is a DNA chain called GGCTCAGGGCCACAG.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	B	15	475	145	168	62	86	14	0

- Molecule 2 is a DNA chain called CTGTGGCCCTGAGCC.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	C	15	472	144	170	54	90	14	0

- Molecule 3 is a protein called Steroidogenic factor 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
3	A	102	1664	507	840	161	144	12	0

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

Mol	Chain	Residues	Atoms	
			Total	Zn
4	A	2	2	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: GGCTCAGGGCCACAG

Chain B: 100%

- Molecule 2: CTGTGGCCCTGAGCC

Chain C: 100%

- Molecule 3: Steroidogenic factor 1

Chain A: 73% 19% • 8%

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: GGCTCAGGGCCACAG

Chain B: 100%

- Molecule 2: CTGTGGCCCTGAGCC

Chain C: 100%

C16 T17 G18 T19 G20 G21 C22 C23 C24 T25 G26 A27 G28 C29 C30

- Molecule 3: Steroidogenic factor 1

Chain A:

D10 V15 C16 G17 D18 H24 T29 C30 E31 S32 K38 R39 T40 V41 S52 K59 K63 R64 C65 C68 K72 R79 A82 V83 R84 A85 D86 R87 M88 R89 G90 G91 R92 N93 K94 F95 G96 P97 M98 Y99 K100 A104 L105 K106 Q107 Q108 K109 K110

A111

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *Simulated annealing and cartesian molecular dynamics.*

Of the 200 calculated structures, 16 were deposited, based on the following criterion: *lowest restraint energies, restraint violations, and RMS deviations from ideal covalent geometry.*

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

Software name	Classification	Version
ARIA	refinement	1.2
CNS	refinement	1.1

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)

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

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

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	307	168	168	25±5
2	C	302	170	170	29±5
3	A	760	764	753	30±6
All	All	21936	17632	17456	1051

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 27.

5 of 549 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:C:18:DG:H1'	2:C:19:DT:H5'	1.10	1.23	7	6
2:C:27:DA:H4'	3:A:91:GLY:H	0.97	1.18	6	3
2:C:20:DG:H1'	2:C:21:DG:H5'	0.96	1.34	9	9
1:B:3:DC:H1'	1:B:4:DT:H5'	0.93	1.36	1	2
2:C:19:DT:H3'	3:A:69:ARG:HH12	0.93	1.24	15	4

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	93/102 (91%)	83±3 (89±3%)	8±3 (9±3%)	2±1 (2±1%)	13 52
All	All	1488/1632 (91%)	1326 (89%)	130 (9%)	32 (2%)	13 52

5 of 12 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	86	ASP	7
3	A	87	ARG	6
3	A	51	GLU	4
3	A	52	SER	3
3	A	50	THR	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	84/90 (93%)	80±1 (96±2%)	4±1 (4±2%)	40 83
All	All	1344/1440 (93%)	1286 (96%)	58 (4%)	40 83

5 of 30 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	44	ASN	6
3	A	102	ASP	4
3	A	71	GLN	4
3	A	43	ASN	4
3	A	72	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 2 ligands modelled in this entry, 2 are 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