wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 02:15 PM BST

PDB ID	:	1D1N
Title	:	SOLUTION STRUCTURE OF THE FMET-TRNAFMET BINDING DOMAIN OF BECILLUS STEAROTHERMOPHILLUS TRANSLATION INITIATION FACTOR IF2
Authors	:	Meunier, S.; Spurio, S.; Czisch, M.; Wechselberger, R.; Gueunneugues, M.
Deposited on	:	1999-09-20

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:

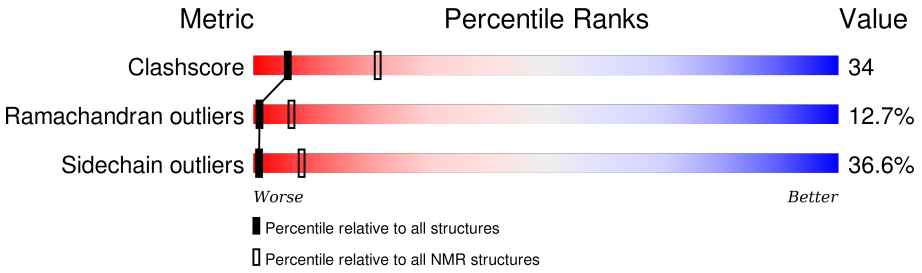
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 is 74%.

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	A	99	27% 42% 15% • 13%

2 Ensemble composition and analysis (i)

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:146-A:200, A:204-A:234 (86)	0.22	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 5 clusters and 3 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1581 atoms, of which 797 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called INITIATION FACTOR 2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	99	1581	492	797	136	153	3	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: INITIATION FACTOR 2

Chain A: 27% 42% 15% • 13%

Y143 E144 E145 K146 V147 I148 G149 Q150 A151 E152 V153 R154 Q155 T156 F157 K161 V162 G163 T164 I165 C168 Y169 V170 G173 K174 I175 T176 R177 D178 S179 K180 V181 R182 L183 I184 R185 Q186 G187 I188 V189 V190 Y191 E192 G193 E194 I195 D196 S197 L198 K199 R200 Y201 K202 D203 R206 E207

V208 A209 Q210 G211 Y212 E213 L216 T217 I218 F221 N222 D223 I224 K225 E226 G227 D228 V229 I230 E231 A232 Y233 V234 M235 Q236 E237 V238 A239 R240 A241

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: INITIATION FACTOR 2

Chain A: 26% 37% 19% • 13%

Y143 E144 E145 K146 V147 I148 G149 Q150 A151 E152 V153 R154 Q155 T156 F157 K158 V162 G163 T164 I165 C168 Y169 V170 G173 K174 I175 T176 R177 D178 S179 K180 V181 R182 L183 I184 R185 Q186 G187 I188 V189 V190 Y191 E192 G193 E194 I195 D196 S197 L198 K199 R200 Y201 K202 D203 E207 V208

A209 Q210 G211 Y212 E213 C214 G215 L216 T217 I218 F221 N222 D223 I224 K225 E226 G227 D228 V229 I230 E231 A232 Y233 V234 M235 Q236 E237 V238 A239 R240 A241

5 Refinement protocol and experimental data overview (i)

Of the 50 calculated structures, 20 were deposited, based on the following criterion: *STRUCTURES WITH ACCEPTABLE COVALENT GEOMETRY,STRUCTURES WITH FAVORABLE NON-BOND ENERGY,STRUCTURES WITH THE LEAST RESTRAINT VIOLATIONS,STRUCTURES WITH THE LOWEST ENERGY.*

The authors did not provide any information on software used for structure solution, optimization or refinement.

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	BMRB entry 4697
Number of chemical shift lists	1
Total number of shifts	1101
Number of shifts mapped to atoms	1028
Number of unparsed shifts	0
Number of shifts with mapping errors	73
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	74%

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)

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	A	672	693	693	47±5
All	All	13440	13860	13860	940

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 34.

5 of 285 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:195:ILE:HD11	1:A:216:LEU:HD13	1.10	1.10	6	3
1:A:148:ILE:HG21	1:A:234:VAL:HG12	1.01	1.29	6	10
1:A:164:THR:HG21	1:A:224:ILE:HG21	0.92	1.38	4	8
1:A:198:LEU:CD1	1:A:216:LEU:HD23	0.80	2.06	18	1
1:A:190:VAL:HG12	1:A:223:ASP:OD2	0.80	1.76	15	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
1	A	86/99 (87%)	58±3 (67±3%)	17±2 (20±3%)	11±1 (13±2%)	1 6
All	All	1720/1980 (87%)	1157 (67%)	344 (20%)	219 (13%)	1 6

5 of 15 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	173	GLY	20
1	A	211	GLY	20
1	A	187	GLY	20
1	A	165	ILE	20
1	A	162	VAL	20

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
1	A	73/84 (87%)	46±3 (63±4%)	27±3 (37±4%)	1 8
All	All	1460/1680 (87%)	926 (63%)	534 (37%)	1 8

5 of 57 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)
1	A	188	ILE	20
1	A	164	THR	20
1	A	155	GLN	20
1	A	216	LEU	20
1	A	223	ASP	20

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)

The completeness of assignment taking into account all chemical shift lists is 74% for the well-defined parts and 74% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 4697

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping (i)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	1101
Number of shifts mapped to atoms	1028
Number of unparsed shifts	0
Number of shifts with mapping errors	73
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	9

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- Residue not found in structure. First 5 (of 73) occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	4	ALA	HB1	1.881	-1.0	1
A	12	PRO	C	176.099	-1.0	1
A	4	ALA	HB2	1.881	-1.0	1
A	9	MET	H	7.969	-1.0	1
A	8	GLY	N	110.33	-1.0	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	109	-0.32 ± 0.10	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	68	-0.52 ± 0.17	Should be applied
$^{13}C'$	109	0.39 ± 0.14	None needed (< 0.5 ppm)

Continued on next page...

Continued from previous page...

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
^{15}N	110	-1.59 ± 0.34	Should be applied

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 74%, i.e. 784 atoms were assigned a chemical shift out of a possible 1060. 0 out of 15 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	424/430 (99%)	168/172 (98%)	170/172 (99%)	86/86 (100%)
Sidechain	335/580 (58%)	264/334 (79%)	61/214 (29%)	10/32 (31%)
Aromatic	25/50 (50%)	25/26 (96%)	0/24 (0%)	0/0 (—%)
Overall	784/1060 (74%)	457/532 (86%)	231/410 (56%)	96/118 (81%)

7.1.4 Statistically unusual chemical shifts (i)

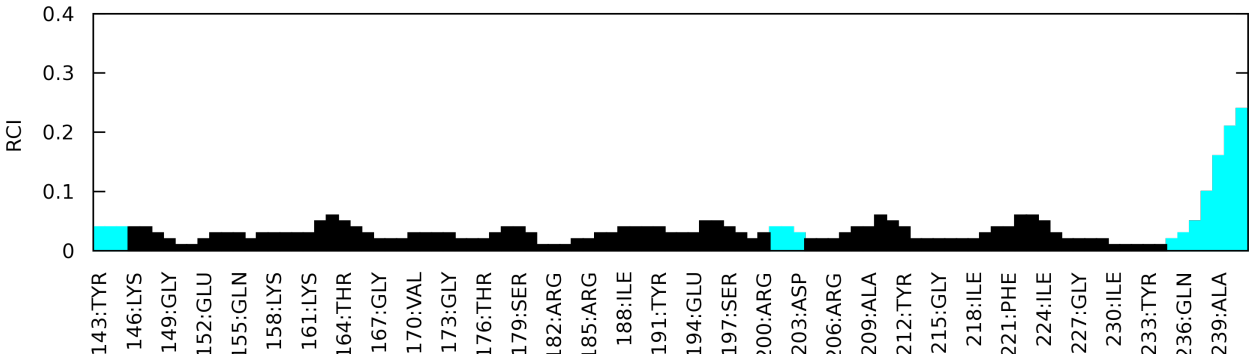
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	206	ARG	NE	136.16	92.63 – 76.73	32.4
1	A	240	ARG	NE	126.09	92.63 – 76.73	26.0
1	A	185	ARG	NE	125.85	92.63 – 76.73	25.9
1	A	177	ARG	NE	125.37	92.63 – 76.73	25.6
1	A	154	ARG	NE	125.12	92.63 – 76.73	25.4
1	A	182	ARG	NE	123.72	92.63 – 76.73	24.6
1	A	4	ALA	CB	32.98	28.03 – 9.93	7.7
1	A	180	LYS	HB3	0.27	3.10 – 0.40	-5.5
1	A	223	ASP	H	5.43	11.17 – 5.47	-5.1

7.1.5 Random Coil Index (RCI) plots (i)

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A:


RCI
0.4
0.3
0.2
0.1
0
143:TYR
146:LYS
149:GLY
152:GLU
155:GLN
158:LYS
161:LYS
164:THR
167:GLY
170:VAL
173:GLY
176:THR
179:SER
182:ARG
185:ARG
188:ILE
191:TYR
194:GLU
197:SER
200:ARG
203:ASP
206:ARG
209:ALA
212:TYR
215:GLY
218:ILE
221:PHE
224:ILE
227:GLY
230:ILE
233:TYR
236:GLN
239:ALA