wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 05:08 AM BST

PDB ID	:	2RQV
Title	:	Solution structure of SH3CI domain of Bem1p
Authors	:	Takaku, T.; Ogura, K.; Inagaki, F.
Deposited on	:	2009-12-21

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:

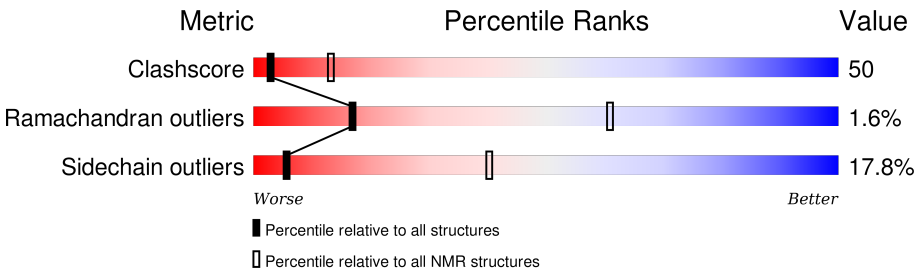
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	A	108	19% 47% • 31%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 17 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:158-A:196, A:203-A:215, A:223-A:245 (75)	0.40	17

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 4 clusters and 6 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1534 atoms, of which 710 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Bud emergence protein 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	108	1534	533	710	133	156	2	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	153	GLY	-	EXPRESSION TAG	UNP P29366
A	154	PRO	-	EXPRESSION TAG	UNP P29366
A	155	LEU	-	EXPRESSION TAG	UNP P29366

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: Bud emergence protein 1

Chain A: 19% 47% • 31%

G153 P154 L155 G156 S157 L158 Y159 A160 I161 V162 L163 Y164 D165 F166 K167 A168 E169 K170 E173 L174 T175 T176 Y177 V178 G179 E180 N181 L182 F183 I184 C185 A186 H187 H188 N189 C190 E191 W192 F193 I194 A195 K196 P197 I198 G199 R200 L201 G202 G203 L206 V207 P208 V209 G210 F211 V212 S213 I214

I215 D216 I217 A218 T219 G220 Y221 A222 N225 D226 V227 I228 E229 D230 I231 V234 N235 L236 V239 Q240 E241 W242 K243 I246 A247 R248 Y249 K250 A251 S252 N253 I254 S255 L256 G257 S258 V259 E260

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

The representative model is number 17. Colouring as in section 4.1 above.

- Molecule 1: Bud emergence protein 1

Chain A: 27% 36% 6% 31%

G153 P154 L155 G156 S157 L158 Y159 A160 I161 V162 L163 F166 K167 A168 E169 K170 A171 D172 E173 L174 T175 T176 Y177 V178 G179 E180 N181 L182 F183 I184 C185 A186 H187 H188 N189 C190 E191 W192 F193 P197 I198 G199 R200 L201 G202 G205 L206 V207 P208 V209 V212 S213 I214 I215 D216 I217

A218 T219 G220 Y221 A222 T223 G224 N225 D226 V227 I228 I231 V234 N235 L236 V239 W242 K243 S244 N245 I246 A247 R248 Y249 K250 A251 S252 N253 I254 S255 L256 G257 S258 V259 E260

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *torsion angle dynamics*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function*.

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

Software name	Classification	Version
CYANA	structure solution	2.1
CYANA	refinement	2.1

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)

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	592	506	581	59±9
All	All	11840	10120	11620	1179

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 50.

5 of 440 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:193:PHE:CE2	1:A:209:VAL:HG22	1.07	1.84	18	3
1:A:158:LEU:HD12	1:A:236:LEU:HD12	0.95	1.38	3	2
1:A:193:PHE:CZ	1:A:227:VAL:HG21	0.93	1.98	18	19
1:A:211:PHE:CD2	1:A:212:VAL:HG13	0.90	2.01	20	1
1:A:191:GLU:O	1:A:209:VAL:HG23	0.90	1.67	10	3

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	75/108 (69%)	65±2 (87±3%)	8±2 (11±3%)	1±1 (2±1%)	17	61
All	All	1500/2160 (69%)	1308 (87%)	168 (11%)	24 (2%)	17	61

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)
1	A	170	LYS	9
1	A	203	GLY	8
1	A	225	ASN	5
1	A	178	VAL	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	
1	A	65/88 (74%)	53±2 (82±4%)	12±2 (18±4%)	6	40
All	All	1300/1760 (74%)	1069 (82%)	231 (18%)	6	40

5 of 36 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	163	LEU	20
1	A	161	ILE	18
1	A	207	VAL	15
1	A	170	LYS	13
1	A	226	ASP	12

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