wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 05:15 PM BST

PDB ID	:	1UFZ
Title	:	Solution structure of HBS1-like domain in hypothetical protein BAB28515
Authors	:	He, F.; Muto, Y.; Hayami, N.; Shirouzu, M.; Terada, T.; Kigawa, T.; Inoue, M.; Yabuki, T.; Aoki, M.; Seki, E.; Matsuda, T.; Hirota, H.; Yoshida, M.; Kobayashi, N.; Tanaka, A.; Osanai, T.; Matsuo, Y.; Hayashizaki, Y.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2003-06-11

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:

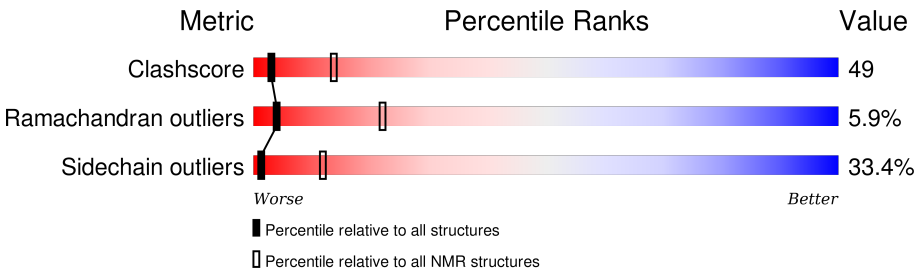
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	83	10% 35% 10% 46%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 4 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:31-A:75 (45)	0.24	4

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 3 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1224 atoms, of which 596 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Hypothetical protein BAB28515.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	83	1224	383	596	107	136	2	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	CLONING ARTIFACT	UNP Q69ZS7
A	2	SER	-	CLONING ARTIFACT	UNP Q69ZS7
A	3	SER	-	CLONING ARTIFACT	UNP Q69ZS7
A	4	GLY	-	CLONING ARTIFACT	UNP Q69ZS7
A	5	SER	-	CLONING ARTIFACT	UNP Q69ZS7
A	6	SER	-	CLONING ARTIFACT	UNP Q69ZS7
A	7	GLY	-	CLONING ARTIFACT	UNP Q69ZS7
A	78	SER	-	CLONING ARTIFACT	UNP Q69ZS7
A	79	GLY	-	CLONING ARTIFACT	UNP Q69ZS7
A	80	PRO	-	CLONING ARTIFACT	UNP Q69ZS7
A	81	SER	-	CLONING ARTIFACT	UNP Q69ZS7
A	82	SER	-	CLONING ARTIFACT	UNP Q69ZS7
A	83	GLY	-	CLONING ARTIFACT	UNP Q69ZS7

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: Hypothetical protein BAB28515

Chain A: 10% 35% 10% 46%

G1 S2 S3 G4 S5 S6 G7 E8 Y9 G10 Y11 E12 D13 L14 R15 E16 S17 S18 N19 S20 L21 L22 N23 H24 Q25 L26 S27 E28 I29 D30 Q31 A32 R33 L34 Y35 S36 C37 L38 D39 H40 M41 R42 E43 V44 L45 G46 D47 A48 V49 P50 D51 D52 I53 L54 T55 E56 A57 I58 L59 K60

H61 K62 F63 D64 V65 Q66 K67 A68 L69 S70 V71 V72 D76 G77 S78 G79 P80 S81 S82 G83

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

The representative model is number 4. Colouring as in section 4.1 above.

- Molecule 1: Hypothetical protein BAB28515

Chain A: 16% 27% 12% 46%

G1 S2 S3 G4 S5 S6 G7 E8 Y9 G10 Y11 E12 D13 L14 R15 E16 S17 S18 N19 S20 L21 L22 N23 H24 Q25 L26 S27 E28 I29 D30 Q31 A32 R33 L34 Y35 S36 C37 L38 D39 R42 E43 V44 L45 G46 D47 A48 V49 P50 D51 D52 I53 L54 T55 I58 L59 K60 H61 K62

F63 D64 V65 Q66 K67 A68 L69 V72 D76 G77 S78 G79 P80 S81 S82 G83

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: *structures with the least restraint violations, structures with the lowest energy, target function*.

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

Software name	Classification	Version
CYANA	structure solution	1.0.7
CYANA	refinement	1.0.7

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	357	364	364	35±8
All	All	7140	7280	7280	708

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 49.

5 of 220 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:49:VAL:HG21	1:A:54:LEU:HD21	0.95	1.38	9	5
1:A:61:HIS:NE2	1:A:71:VAL:HG21	0.90	1.80	15	4
1:A:55:THR:O	1:A:59:LEU:HD22	0.86	1.70	17	13
1:A:65:VAL:HG12	1:A:69:LEU:HD12	0.86	1.47	12	9
1:A:37:CYS:SG	1:A:58:ILE:HD13	0.86	2.11	9	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
1	A	45/83 (54%)	29±2 (64±4%)	13±2 (30±4%)	3±1 (6±2%)	4 22
All	All	900/1660 (54%)	580 (64%)	267 (30%)	53 (6%)	4 22

5 of 9 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	46	GLY	19
1	A	63	PHE	18
1	A	40	HIS	8
1	A	39	ASP	2
1	A	62	LYS	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	40/71 (56%)	27±2 (67±6%)	13±2 (33±6%)	1 12
All	All	800/1420 (56%)	533 (67%)	267 (33%)	1 12

5 of 32 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	62	LYS	20
1	A	38	LEU	19
1	A	60	LYS	17
1	A	59	LEU	16
1	A	66	GLN	16

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