wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 04:19 PM BST

PDB ID	:	1P9K
Title	:	THE SOLUTION STRUCTURE OF YBCJ FROM E. COLI REVEALS A RECENTLY DISCOVERED ALFAL MOTIF INVOLVED IN RNA-BINDING
Authors	:	Volpon, L.; Lievre, C.; Osborne, M.J.; Gandhi, S.; Iannuzzi, P.; Larocque, R.; Matte, A.; Cygler, M.; Gehring, K.; Ekiel, I.; Montreal-Kingston Bacterial Structural Genomics Initiative (BSGI)
Deposited on	:	2003-05-12

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:

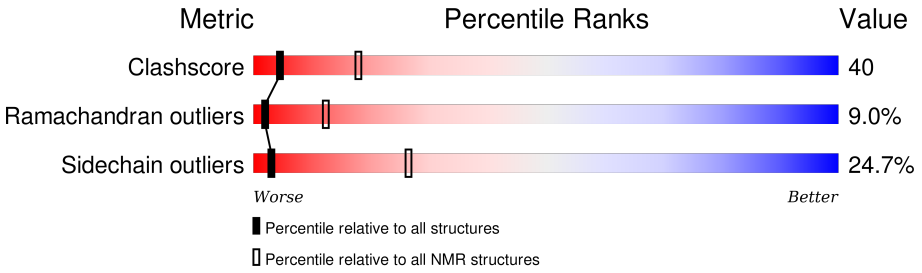
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	79	22% 59% 6% 13%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 19 is the overall representative, medoid model (most similar to other models). The authors have identified model 7 as representative, based on the following criterion: *closest to the average*.

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:11-A:79 (69)	0.34	19

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 2 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1187 atoms, of which 600 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called orf, hypothetical protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	79	1187	365	600	108	109	5	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	CLONING ARTIFACT	UNP P0AAS7
A	2	SER	-	CLONING ARTIFACT	UNP P0AAS7

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: orf, hypothetical protein

Chain A: 22% 59% 6% 13%

G1 S2 M3 I4 H5 R6 M7 S8 N9 M10 A11 T12 F13 S14 L15 G16 K17 H18 P19 H20 V21 E22 L23 C24 D25 L26 L27 K28 L29 E30 G31 W32 S33 E34 S35 Q38 I41 A44 E45 G46 Q47 V48 K49 V50 D51 G52 A53 V54 E55 T56 R57 K58 R59 C60 K61 I62 V63

A64 G65 Q66 T67 V68 S69 F70 A71 G72 H73 S74 V75 Q76 V77 V78 A79

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

The representative model is number 19. Colouring as in section 4.1 above.

- Molecule 1: orf, hypothetical protein

Chain A: 34% 43% 10% 13%

G1 S2 M3 I4 H5 R6 M7 S8 N9 M10 A11 T12 F13 S14 L15 G16 K17 H18 P19 H20 V21 E22 L23 C24 D25 L26 L27 K28 L29 E30 G31 W32 S33 K40 I41 A44 E45 G46 Q47 V48 K49 V50 D51 G52 A53 V54 E55 T56 R57 K58 R59 C60 V63 A64 V68 S69

F70 A71 G72 H73 S74 V77 V78 A79

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *Simulated annealing.*

Of the 60 calculated structures, 20 were deposited, based on the following criterion: *structures with the least restraint violations,structures with the lowest energy.*

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

Software name	Classification	Version
TALOS	refinement	2003.027.13.05
ARIA	refinement	1.1
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)

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	510	521	521	41±7
All	All	10200	10420	10420	821

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 40.

5 of 235 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:27:LEU:HD23	1:A:41:ILE:HG22	0.91	1.43	20	16
1:A:26:LEU:HD11	1:A:77:VAL:HG21	0.90	1.41	6	14
1:A:48:VAL:HG21	1:A:70:PHE:CD1	0.88	2.03	8	20
1:A:11:ALA:HB1	1:A:75:VAL:HG13	0.84	1.48	14	5
1:A:41:ILE:HG23	1:A:70:PHE:CE2	0.83	2.08	5	18

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	68/79 (86%)	47±2 (69±2%)	15±2 (22±3%)	6±1 (9±2%)	2 12
All	All	1360/1580 (86%)	935 (69%)	303 (22%)	122 (9%)	2 12

5 of 11 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	48	VAL	20
1	A	47	GLN	20
1	A	51	ASP	19
1	A	71	ALA	16
1	A	20	HIS	15

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	53/62 (85%)	40±3 (75±5%)	13±3 (25±5%)	3 26
All	All	1060/1240 (85%)	798 (75%)	262 (25%)	3 26

5 of 39 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	12	THR	19
1	A	60	CYS	17
1	A	59	ARG	15
1	A	49	LYS	14
1	A	35	SER	11

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