wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 11:14 PM BST

PDB ID	:	2KHJ
Title	:	NMR structure of the domain 6 of the E. coli ribosomal protein S1
Authors	:	Salah, P.; Bisaglia, M.; Aliprandi, P.; Uzan, M.; Sizun, C.; Bontems, F.
Deposited on	:	2009-04-07

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:

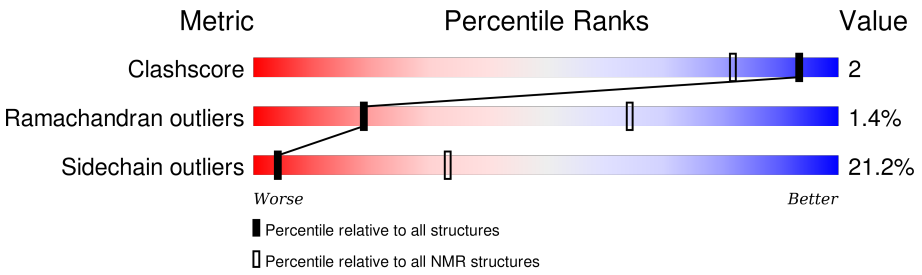
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	109	47% 11% • 21% 18%

2 Ensemble composition and analysis (i)

This entry contains 12 models. Model 2 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:2-A:40, A:57-A:83 (66)	0.62	2

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. No single-model clusters were found.

Cluster number	Models
1	2, 4, 5, 6, 9, 10, 12
2	1, 3, 7, 8, 11

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1343 atoms, of which 674 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called 30S ribosomal protein S1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	89	1343	411	674	120	137	1	0

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	EXPRESSION TAG	UNP P0AG67
A	-18	GLY	-	EXPRESSION TAG	UNP P0AG67
A	-17	SER	-	EXPRESSION TAG	UNP P0AG67
A	-16	SER	-	EXPRESSION TAG	UNP P0AG67
A	-15	HIS	-	EXPRESSION TAG	UNP P0AG67
A	-14	HIS	-	EXPRESSION TAG	UNP P0AG67
A	-13	HIS	-	EXPRESSION TAG	UNP P0AG67
A	-12	HIS	-	EXPRESSION TAG	UNP P0AG67
A	-11	HIS	-	EXPRESSION TAG	UNP P0AG67
A	-10	HIS	-	EXPRESSION TAG	UNP P0AG67
A	-9	SER	-	EXPRESSION TAG	UNP P0AG67
A	-8	SER	-	EXPRESSION TAG	UNP P0AG67
A	-7	GLY	-	EXPRESSION TAG	UNP P0AG67
A	-6	LEU	-	EXPRESSION TAG	UNP P0AG67
A	-5	VAL	-	EXPRESSION TAG	UNP P0AG67
A	-4	PRO	-	EXPRESSION TAG	UNP P0AG67
A	-3	ARG	-	EXPRESSION TAG	UNP P0AG67
A	-2	GLY	-	EXPRESSION TAG	UNP P0AG67
A	-1	SER	-	EXPRESSION TAG	UNP P0AG67
A	0	HIS	-	EXPRESSION TAG	UNP P0AG67
A	1	MET	-	EXPRESSION TAG	UNP P0AG67

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: 30S ribosomal protein S1

Chain A: 47% 11% • 21% 18%

MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS M1 F2 N3 N4 W5 K10 T28 V35 Y38 L39 R40 A41 S42 E43 A44 S45 R46 D47 R48 V49 E50 D51 A52 T53 L54 V55 L56 F66 V69 I76 S77 L78 S79

V80 R81 A82 K83 D84 E85 A86 D87 E88 K89

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

The representative model is number 2. Colouring as in section 4.1 above.

- Molecule 1: 30S ribosomal protein S1

Chain A: 45% 12% • • 21% 18%

MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS M1 F2 N3 N4 W5 V6 N9 K10 L31 V35 Y38 A41 S42 E43 A44 S45 R46 D47 R48 V49 E50 D51 A52 T53 L54 V55 L56 E61 V62 E63 F66 V69

D70 N73 V80 R81 A82 K83 D84 E85 A86 D87 E88 K89

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 12 were deposited, based on the following criterion: *12 structures for lowest energy.*

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

Software name	Classification	Version
INCA	structure solution	
INCA	refinement	

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)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	A	1.00±0.02	0±0/497 (0.0±0.0%)	1.52±0.05	6±2/670 (0.9±0.3%)
All	All	1.00	0/5964 (0.0%)	1.53	70/8040 (0.9%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	A	0.0±0.0	1.8±1.0
All	All	0	22

There are no bond-length outliers.

5 of 22 unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	66	PHE	CB-CG-CD2	-8.86	114.60	120.80	1	8
1	A	66	PHE	CB-CG-CD1	8.63	126.84	120.80	1	8
1	A	27	ALA	CB-CA-C	8.14	122.30	110.10	11	2
1	A	5	TRP	CA-CB-CG	7.16	127.30	113.70	8	4
1	A	5	TRP	CB-CG-CD2	-7.03	117.47	126.60	1	3

There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	2	PHE	Sidechain	8
1	A	80	VAL	Mainchain,Peptide	8
1	A	38	TYR	Sidechain	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
1	A	66	PHE	Sidechain	1

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	492	509	509	2±1
All	All	5904	6108	6108	20

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 2.

5 of 13 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:76:ILE:HD12	1:A:76:ILE:H	0.78	1.38	3	1
1:A:2:PHE:CG	1:A:76:ILE:HG23	0.58	2.34	10	5
1:A:76:ILE:CD1	1:A:76:ILE:H	0.54	2.14	3	1
1:A:27:ALA:HB3	1:A:39:LEU:O	0.53	2.03	11	1
1:A:2:PHE:CZ	1:A:69:VAL:HG11	0.51	2.41	1	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	66/109 (61%)	61±1 (92±1%)	4±1 (7±2%)	1±1 (1±1%)	19	64
All	All	792/1308 (61%)	729 (92%)	52 (7%)	11 (1%)	19	64

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	82	ALA	5
1	A	32	ALA	3
1	A	83	LYS	2
1	A	33	ASP	1

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	50/86 (58%)	39±3 (79±6%)	11±3 (21±6%)	4 33
All	All	600/1032 (58%)	473 (79%)	127 (21%)	4 33

5 of 33 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	69	VAL	11
1	A	66	PHE	10
1	A	5	TRP	8
1	A	2	PHE	7
1	A	28	THR	7

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