wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 08:05 PM BST

PDB ID	:	2CXJ
Title	:	3D Solution Structure of S100A13
Authors	:	Vaithiyalingam, S.; Kumar, T.K.S.; Yu, C.
Deposited on	:	2005-06-30

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:

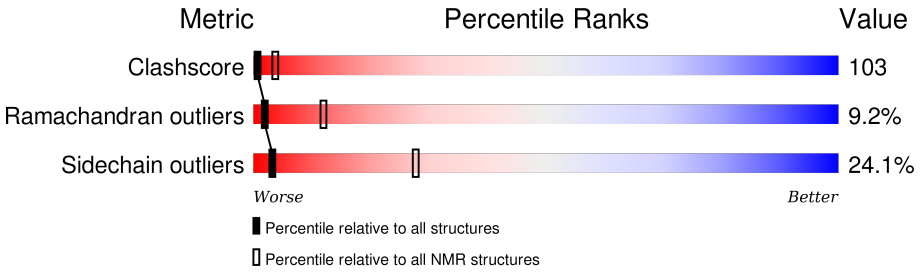
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal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

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	98	14% 64% 21%
1	B	98	13% 65% 21%

2 Ensemble composition and analysis

This entry contains 20 models. Model 7 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:1-A:98, B:1-B:98 (196)	0.46	7

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 3 clusters. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 3184 atoms, of which 1614 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called S100 calcium-binding protein A13.

Mol	Chain	Residues	Atoms						Trace
1	A	98	Total	C	H	N	O	S	0
			1592	496	807	135	152	2	
1	B	98	Total	C	H	N	O	S	0
			1592	496	807	135	152	2	

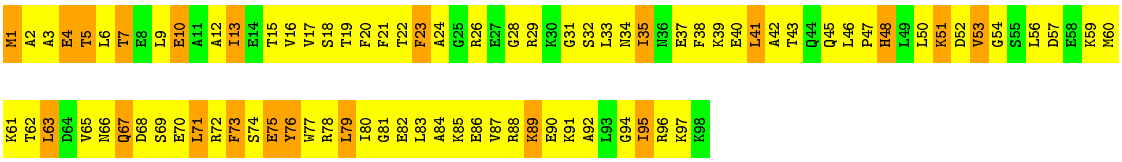
4 Residue-property plots

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.

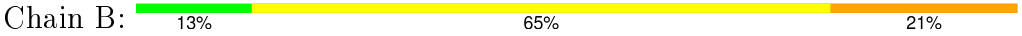
- Molecule 1: S100 calcium-binding protein A13

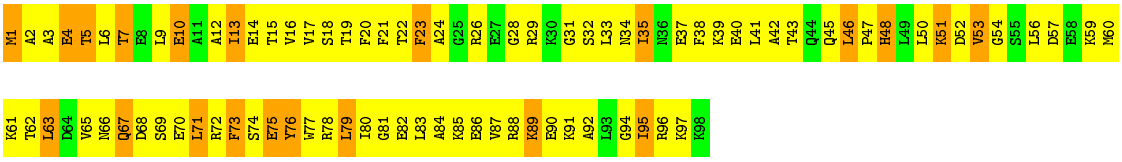
Chain A:



- Molecule 1: S100 calcium-binding protein A13

Chain B:



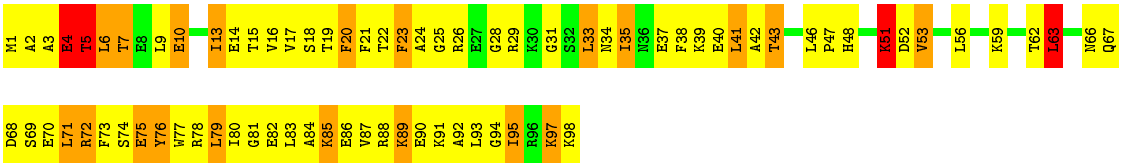


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 7. Colouring as in section 4.1 above.

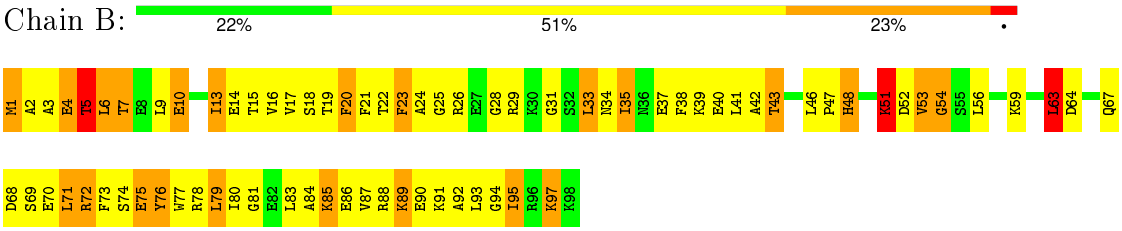
- Molecule 1: S100 calcium-binding protein A13

Chain A:



- Molecule 1: S100 calcium-binding protein A13

Chain B:


22%
51%
23%
M1 A2 A3 E4 T5 L6 T7 E8 L9 E10 I13 E14 T15 V16 W17 S18 T19 F20 F21 T22 F23 A24 G25 R26 E27 G28 R29 K30 G31 S32 I33 N34 I35 N36 E37 F38 K39 E40 L41 A42 T43 L46 P47 H48 K51 D52 V53 G54 S55 L56 K59 L63 D64 Q67
D68 S69 E70 L71 R72 F73 S74 E75 Y76 W77 R78 L79 I80 G81 E82 L83 A84 K85 E86 V87 R88 K89 E90 K91 A92 L93 G94 I95 R96 K97 K98

5 Refinement protocol and experimental data overview

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

Of the 200 calculated structures, 20 were deposited, based on the following criterion: *structures with the least restraint violations*.

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

Software name	Classification	Version
CNS	structure solution	1.1
CNS	refinement	1.1

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 6484
Number of chemical shift lists	1
Total number of shifts	2426
Number of shifts mapped to atoms	2330
Number of unparsed shifts	0
Number of shifts with mapping errors	96
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	83%

No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality

6.1 Standard geometry

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

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	785	807	806	201±13
1	B	785	807	806	199±14
All	All	31400	32280	32240	6543

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 103.

5 of 2053 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:9:LEU:HD22	1:B:16:VAL:HG23	1.10	1.18	8	18
1:A:16:VAL:HG23	1:B:9:LEU:HD22	1.07	1.16	11	18
1:A:76:TYR:CD1	1:A:79:LEU:HD11	1.04	1.87	13	16
1:B:76:TYR:CD1	1:B:79:LEU:HD11	1.03	1.88	13	16
1:A:23:PHE:CD2	1:A:33:LEU:HD11	1.01	1.91	12	13

6.3 Torsion angles

6.3.1 Protein backbone

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	96/98 (98%)	62±2 (64±3%)	25±3 (26±4%)	9±2 (9±2%)	2 11
1	B	96/98 (98%)	62±3 (65±3%)	26±3 (27±3%)	9±2 (9±3%)	2 13
All	All	3840/3920 (98%)	2472 (64%)	1016 (26%)	352 (9%)	2 12

5 of 43 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	95	ILE	20
1	B	95	ILE	20
1	B	53	VAL	19
1	A	41	LEU	19
1	A	53	VAL	19

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	84/84 (100%)	64±3 (76±4%)	20±3 (24±4%)	3 28
1	B	84/84 (100%)	64±3 (76±4%)	20±3 (24±4%)	3 27
All	All	3360/3360 (100%)	2549 (76%)	811 (24%)	3 28

5 of 114 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	B	10	GLU	20
1	A	10	GLU	20
1	A	89	LYS	20
1	B	89	LYS	20
1	A	7	THR	19

6.3.3 RNA

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates

There are no carbohydrates in this entry.

6.6 Ligand geometry

There are no ligands in this entry.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 83% for the well-defined parts and 83% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 6484

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping

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	2426
Number of shifts mapped to atoms	2330
Number of unparsed shifts	0
Number of shifts with mapping errors	96
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	6

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

- Residue not found in structure. First 5 (of 96) occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
B	1	LEU	CB	39.33	0.2	1
A	3	LEU	HD21	0.81	0.01	1
A	1	LEU	CB	39.33	0.2	1
A	1	LEU	HB3	1.52	0.01	1
B	2	VAL	H	8.01	0.01	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	202	2.52 ± 0.09	Should be applied
$^{13}C_{\beta}$	190	2.94 ± 0.07	Should be applied
$^{13}C'$	194	2.36 ± 0.11	Should be applied

Continued on next page...

Continued from previous page...

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
^{15}N	194	0.68 ± 0.20	Should be applied

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 83%, i.e. 2072 atoms were assigned a chemical shift out of a possible 2496. 36 out of 36 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	954/976 (98%)	382/390 (98%)	384/392 (98%)	188/194 (97%)
Sidechain	1060/1374 (77%)	646/800 (81%)	410/506 (81%)	4/68 (6%)
Aromatic	58/146 (40%)	40/78 (51%)	16/62 (26%)	2/6 (33%)
Overall	2072/2496 (83%)	1068/1268 (84%)	810/960 (84%)	194/268 (72%)

7.1.4 Statistically unusual chemical shifts

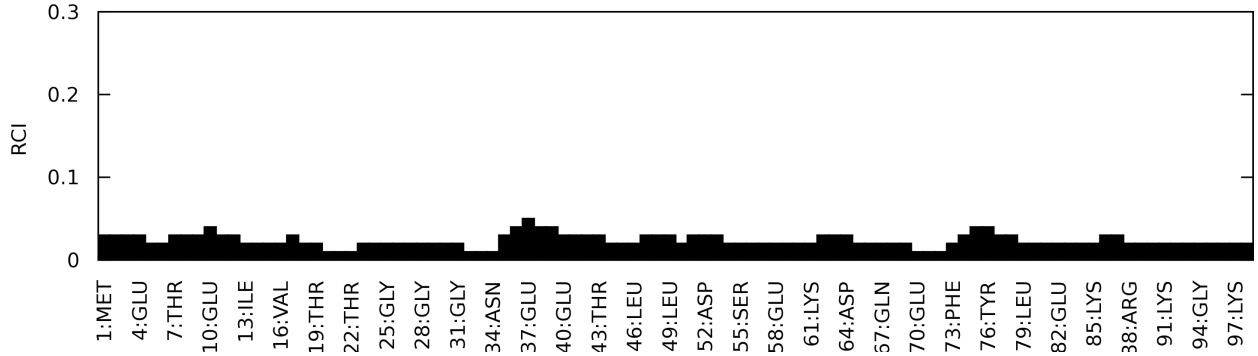
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	98	LYS	CD	40.10	34.86 – 23.06	9.4
1	B	98	LYS	CD	40.10	34.86 – 23.06	9.4
1	A	97	LYS	CD	39.20	34.86 – 23.06	8.7
1	B	97	LYS	CD	39.20	34.86 – 23.06	8.7
1	B	75	GLU	CG	42.60	42.24 – 29.94	5.3
1	A	75	GLU	CG	42.60	42.24 – 29.94	5.3

7.1.5 Random Coil Index (RCI) plots

The images below report *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:



Random coil index (RCI) for chain B: