wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 11:13 PM BST

PDB ID	:	2KIS
Title	:	Solution structure of CA150 FF1 domain and FF1-FF2 interdomain linker
Authors	:	Murphy, J.M.; Hansen, D.; Wiesner, S.; Muhandiram, D.; Borg, M.; Smith, M.J.; Sicheri, F.; Kay, L.E.; Forman-Kay, J.D.; Pawson, T.
Deposited on	:	2009-05-08

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:

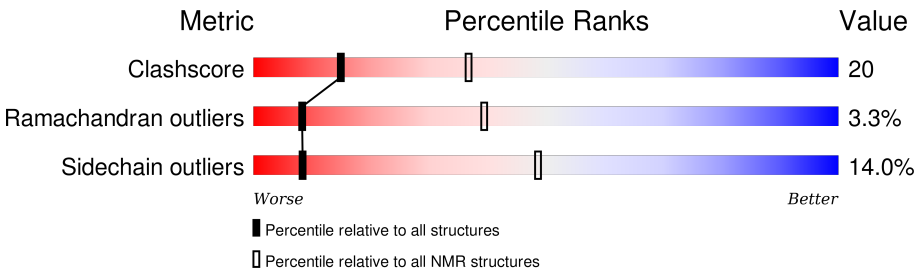
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 84%.

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	71	46% 31% • 20%

2 Ensemble composition and analysis (i)

This entry contains 7 models. Model 2 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:618-A:674 (57)	0.18	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 and 2 single-model clusters were found.

Cluster number	Models
1	1, 2, 7
2	3, 6
Single-model clusters	4; 5

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1221 atoms, of which 619 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Transcription elongation regulator 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	71	1221	381	619	110	108	3	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	613	GLY	-	LINKER	UNP O14776
A	614	ALA	-	LINKER	UNP O14776
A	615	MET	-	LINKER	UNP O14776
A	616	GLY	-	LINKER	UNP O14776
A	617	SER	-	LINKER	UNP O14776

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: Transcription elongation regulator 1

Chain A: 46% 31% • 20%

G613 A614 M615 G616 S617 R621 M622 K623 Q624 R632 G633 V634 S635 T639 W640 E641 K642 E643 L644 H645 V648 F649 D650 P651 R652 L656 N657 P658 K659 E660 R661 K662 Q663 V664 Q667 T671 E675 E676 R677 R678 E679 K680 K681 N682 K683

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: Transcription elongation regulator 1

Chain A: 42% 31% 7% 20%

G613 A614 M615 G616 S617 A620 K623 Q624 F625 L629 L630 E631 R632 G633 V634 S635 T639 W640 E641 K642 E643 L644 H645 K646 I647 V648 F649 D650 L656 N657 E660 R661 K662 Q663 V664 Q667 T671 E675 E676 R677 R678 E679 K680 K681 N682 K683

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing, distance geometry.*

Of the 20 calculated structures, 7 were deposited, based on the following criterion: *structures with the lowest energy.*

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

Software name	Classification	Version
Aria	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 16293
Number of chemical shift lists	2
Total number of shifts	1270
Number of shifts mapped to atoms	1270
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	84%

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	490	503	503	20±4
All	All	3430	3521	3521	137

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 20.

5 of 51 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:634:VAL:HG13	1:A:643:GLU:HB3	0.75	1.55	2	7
1:A:644:LEU:O	1:A:648:VAL:HG13	0.71	1.84	4	5
1:A:660:GLU:O	1:A:664:VAL:HG23	0.67	1.90	7	5
1:A:645:HIS:HA	1:A:648:VAL:HG22	0.62	1.71	4	5
1:A:667:GLN:O	1:A:671:THR:HG22	0.59	1.98	6	7

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	57/71 (80%)	44±1 (77±1%)	11±1 (19±2%)	2±1 (3±1%)	8 39
All	All	399/497 (80%)	309 (77%)	77 (19%)	13 (3%)	8 39

All 5 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	635	SER	7
1	A	633	GLY	3
1	A	658	PRO	1
1	A	644	LEU	1
1	A	640	TRP	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	53/64 (83%)	46±2 (86±4%)	7±2 (14±4%)	8 48
All	All	371/448 (83%)	319 (86%)	52 (14%)	8 48

5 of 22 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	650	ASP	7
1	A	662	LYS	6
1	A	632	ARG	5
1	A	639	THR	5
1	A	623	LYS	4

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)

The completeness of assignment taking into account all chemical shift lists is 84% for the well-defined parts and 81% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 16293

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping (i)

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	736
Number of shifts mapped to atoms	736
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	67	-0.07 ± 0.13	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	68	-0.32 ± 0.14	None needed (< 0.5 ppm)
$^{13}C'$	0	—	—
^{15}N	0	—	—

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 76%, i.e. 608 atoms were assigned a chemical shift out of a possible 798. 0 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	167/281 (59%)	112/112 (100%)	55/114 (48%)	0/55 (0%)
Sidechain	376/446 (84%)	241/264 (91%)	135/156 (87%)	0/26 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	65/71 (92%)	34/38 (89%)	31/31 (100%)	0/2 (0%)
Overall	608/798 (76%)	387/414 (93%)	221/301 (73%)	0/83 (0%)

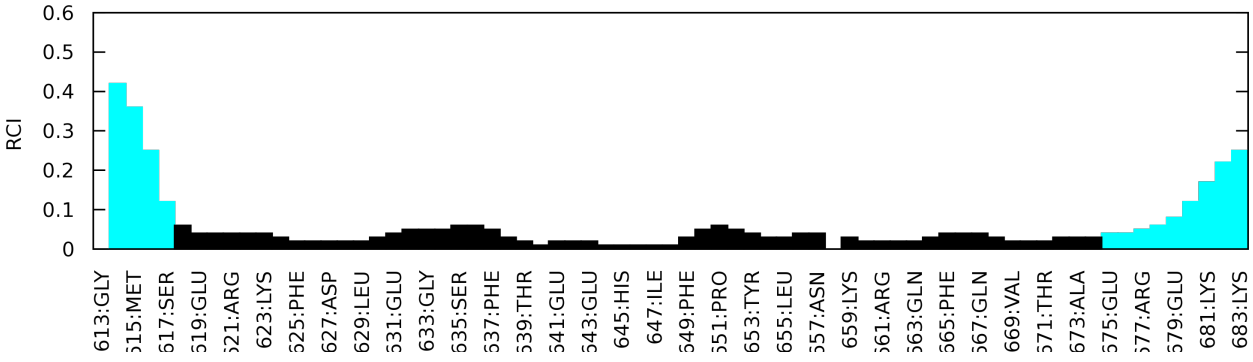
7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots (i)

The image below reports *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:



7.2 Chemical shift list 2

File name: BMRB entry 16293

Chemical shift list name: *assigned_chem_shift_list_2*

7.2.1 Bookkeeping (i)

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	534
Number of shifts mapped to atoms	534
Number of unparsed shifts	0

Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	5

7.2.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	0	—	—
$^{13}C_{\beta}$	0	—	—
$^{13}C'$	0	—	—
^{15}N	67	0.01 ± 0.28	None needed (< 0.5 ppm)

7.2.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 54%, i.e. 431 atoms were assigned a chemical shift out of a possible 798. 0 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

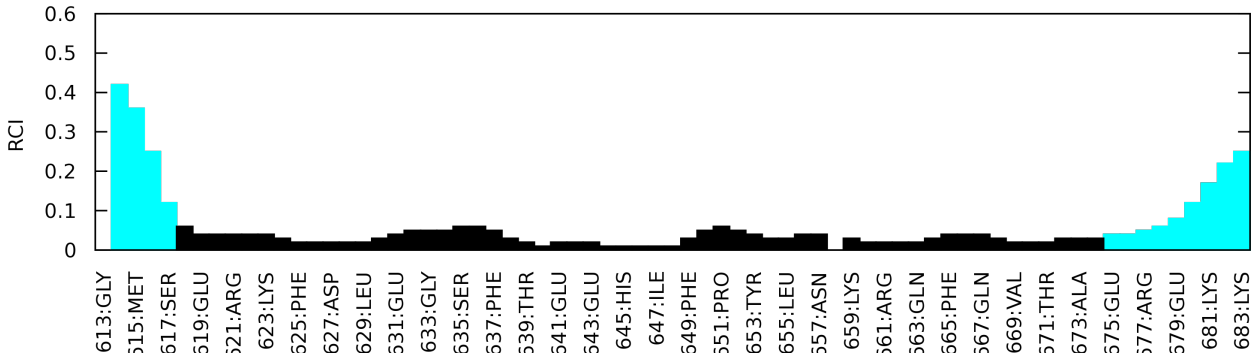
	Total	^{1}H	^{13}C	^{15}N
Backbone	166/281 (59%)	111/112 (99%)	0/114 (0%)	55/55 (100%)
Sidechain	237/446 (53%)	228/264 (86%)	0/156 (0%)	9/26 (35%)
Aromatic	28/71 (39%)	27/38 (71%)	0/31 (0%)	1/2 (50%)
Overall	431/798 (54%)	366/414 (88%)	0/301 (0%)	65/83 (78%)

7.2.4 Statistically unusual chemical shifts (i)

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	661	ARG	NE	111.83	92.63 – 76.73	17.1
1	A	672	ARG	NE	110.81	92.63 – 76.73	16.4
1	A	632	ARG	NE	110.61	92.63 – 76.73	16.3
1	A	652	ARG	NE	110.20	92.63 – 76.73	16.1
1	A	621	ARG	NE	109.80	92.63 – 76.73	15.8

7.2.5 Random Coil Index (RCI) plots (i)

The image below reports *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: