wwPDB NMR Structure Validation Summary Report (i)

Jun 14, 2020 – 05:30 pm BST

PDB ID	:	2RUJ
Title	:	Solution structure of MTSL spin-labeled Schizosaccharomyces pombe Sin1 CRIM domain
Authors	:	Furuita, K.; Kataoka, S.; Sugiki, T.; Kobayashi, N.; Ikegami, T.; Shiozaki, K.; Fujiwara, T.; Kojima, C.
Deposited on	:	2014-07-24

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
https://www.wwpdb.org/validation/2017/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:

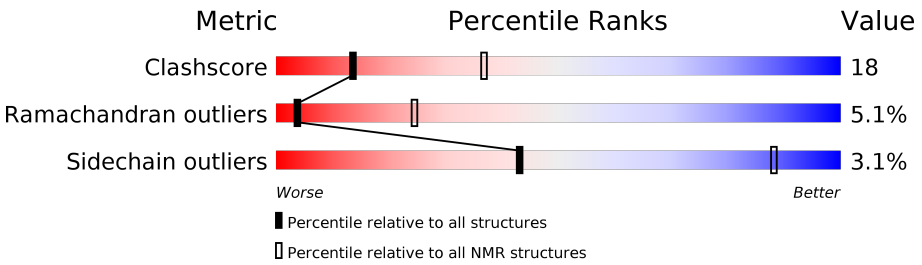
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 77%.

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	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

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	160	43% 28% • 28%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 3 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:271-A:279, A:281-A:281, A:283-A:290, A:292-A:300, A:302-A:311, A:313-A:331, A:333-A:370, A:372-A:383, A:385-A:393 (115)	0.76	3

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. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2656 atoms, of which 1315 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Stress-activated map kinase-interacting protein 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	160	2656	861	1315	215	246	19	0

There are 15 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	241	GLY	-	EXPRESSION TAG	UNP Q9P7Y9
A	242	PRO	-	EXPRESSION TAG	UNP Q9P7Y9
A	243	GLY	-	EXPRESSION TAG	UNP Q9P7Y9
A	244	HIS	-	EXPRESSION TAG	UNP Q9P7Y9
A	245	MET	-	EXPRESSION TAG	UNP Q9P7Y9
A	246	GLY	-	EXPRESSION TAG	UNP Q9P7Y9
A	280	R1A	THR	ENGINEERED MUTATION	UNP Q9P7Y9
A	282	R1A	SER	ENGINEERED MUTATION	UNP Q9P7Y9
A	291	R1A	ARG	ENGINEERED MUTATION	UNP Q9P7Y9
A	301	R1A	SER	ENGINEERED MUTATION	UNP Q9P7Y9
A	312	R1A	LYS	ENGINEERED MUTATION	UNP Q9P7Y9
A	332	R1A	LEU	ENGINEERED MUTATION	UNP Q9P7Y9
A	371	R1A	SER	ENGINEERED MUTATION	UNP Q9P7Y9
A	384	R1A	THR	ENGINEERED MUTATION	UNP Q9P7Y9
A	394	R1A	ALA	ENGINEERED MUTATION	UNP Q9P7Y9

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: Stress-activated map kinase-interacting protein 1

Chain A: 43% 28% • 28%

G241 P242 G243 H244 M245 G246 S247 V248 S249 N250 A251 K252 A253 P254 T255 S256 A257 L258 R259 A260 L261 L262 E263 H264 K265 E266 N267 S268 S269 Q270 N271 A275 F278 A279 C280 F281 C282 A289 L290 C291 L292 F296 S299 E300 C301 P302 S303 L306 E309 L310 R311 C312

V316 S317 E318 A319 I320 G321 Y322 I323 L324 L325 Q326 Q330 Q331 G332 V333 N342 P343 N344 Y345 W346 N347 L348 A363 L364 D365 R366 V367 L370 C371 A377 F378 A379 L380 V381 K382 A383 C384 P385 A386 C394 Y395 P396 F397 K398 S399 K400

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

The representative model is number 3. Colouring as in section 4.1 above.

- Molecule 1: Stress-activated map kinase-interacting protein 1

Chain A: 43% 28% • 28%

G241 P242 G243 H244 M245 G246 S247 V248 S249 N250 A251 K252 A253 P254 T255 S256 A257 L258 R259 A260 L261 L262 E263 H264 K265 E266 N267 S268 S269 Q270 N271 A275 F278 A279 C280 F281 C282 N288 A289 L290 C291 C301 P302 S303 F307 V308 E309 L310 R311 C312 L315 V316 S317

I320 I323 L324 Q330 Q331 G332 V333 D338 E339 A340 Q341 N344 Y345 W346 N347 L348 E352 D358 A363 L364 D365 R366 V367 G368 P369 L370 C371 K372 D376 A377 F378 A379 L380 V381 K382 A383 C384 Q387 I388 C394 Y395 P396 F397 K398 S399 K400

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 10 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
CYANA	structure solution	3.95
X-PLOR NIH	refinement	2.31

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)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1506
Number of shifts mapped to atoms	1506
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	77%

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)

Bond lengths and bond angles in the following residue types are not validated in this section: R1A

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	910	873	872	32±4
All	All	9100	8730	8720	319

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 18.

5 of 202 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:272:GLY:H	1:A:275:ALA:HB3	0.77	1.39	4	4
1:A:309:GLU:C	1:A:310:LEU:HD12	0.70	2.06	4	1
1:A:315:LEU:N	1:A:315:LEU:HD22	0.69	2.02	9	1
1:A:290:LEU:H	1:A:290:LEU:HD22	0.69	1.46	10	1
1:A:310:LEU:HD12	1:A:310:LEU:N	0.65	2.07	9	1

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	115/160 (72%)	100±2 (87±1%)	9±2 (8±1%)	6±1 (5±1%)	4 24
All	All	1150/1600 (72%)	1003 (87%)	88 (8%)	59 (5%)	4 24

5 of 16 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	365	ASP	9
1	A	333	VAL	7
1	A	385	PRO	6
1	A	289	ALA	6
1	A	299	SER	6

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	97/126 (77%)	94±2 (97±2%)	3±2 (3±2%)	43 88
All	All	970/1260 (77%)	940 (97%)	30 (3%)	43 88

5 of 13 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	382	LYS	8
1	A	290	LEU	6
1	A	370	LEU	4
1	A	292	LEU	2
1	A	342	ASN	2

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

9 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
1	R1A	A	301	1	15,18,19	0.55±0.03	0±0 (0±0%)
1	R1A	A	394	1	15,18,19	0.54±0.02	0±0 (0±0%)
1	R1A	A	332	1	15,18,19	0.53±0.02	0±0 (0±0%)
1	R1A	A	282	1	15,18,19	0.54±0.03	0±0 (0±0%)
1	R1A	A	291	1	15,18,19	0.53±0.02	0±0 (0±0%)
1	R1A	A	280	1	15,18,19	0.53±0.01	0±0 (0±0%)
1	R1A	A	312	1	15,18,19	0.54±0.01	0±0 (0±0%)
1	R1A	A	371	1	15,18,19	0.53±0.02	0±0 (0±0%)
1	R1A	A	384	1	15,18,19	0.55±0.03	0±0 (0±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
1	R1A	A	301	1	15,27,29	0.93±0.03	0±0 (0±0%)
1	R1A	A	394	1	15,27,29	0.94±0.03	0±0 (0±0%)
1	R1A	A	332	1	15,27,29	0.91±0.05	0±0 (0±0%)
1	R1A	A	282	1	15,27,29	0.96±0.06	0±0 (0±0%)
1	R1A	A	291	1	15,27,29	0.93±0.05	0±0 (0±0%)
1	R1A	A	280	1	15,27,29	0.90±0.05	0±0 (0±0%)
1	R1A	A	312	1	15,27,29	0.93±0.04	0±0 (0±0%)

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
1	R1A	A	371	1	15,27,29	0.93±0.04	0±0 (0±0%)
1	R1A	A	384	1	15,27,29	0.97±0.07	0±0 (0±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	R1A	A	301	1	-	0±0,5,32,34	0±0,1,1,1
1	R1A	A	394	1	-	0±0,5,32,34	0±0,1,1,1
1	R1A	A	312	1	-	0±0,5,32,34	0±0,1,1,1
1	R1A	A	332	1	-	0±0,5,32,34	0±0,1,1,1
1	R1A	A	384	1	-	0±0,5,32,34	0±0,1,1,1
1	R1A	A	282	1	-	0±0,5,32,34	0±0,1,1,1
1	R1A	A	280	1	-	0±0,5,32,34	0±0,1,1,1
1	R1A	A	291	1	-	0±0,5,32,34	0±0,1,1,1
1	R1A	A	371	1	-	0±0,5,32,34	0±0,1,1,1

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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 77% for the well-defined parts and 74% for the entire structure.

7.1 Chemical shift list 1

File name: input_cs.cif

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	1506
Number of shifts mapped to atoms	1506
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	140	-0.15 ± 0.12	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	132	0.05 ± 0.11	None needed (< 0.5 ppm)
$^{13}C'$	125	-0.02 ± 0.13	None needed (< 0.5 ppm)
^{15}N	129	0.22 ± 0.30	None needed (< 0.5 ppm)

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 77%, i.e. 1090 atoms were assigned a chemical shift out of a possible 1410. 0 out of 21 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	525/555 (95%)	211/220 (96%)	211/230 (92%)	103/105 (98%)
Sidechain	536/731 (73%)	322/428 (75%)	209/274 (76%)	5/29 (17%)

Continued on next page...

Continued from previous page...

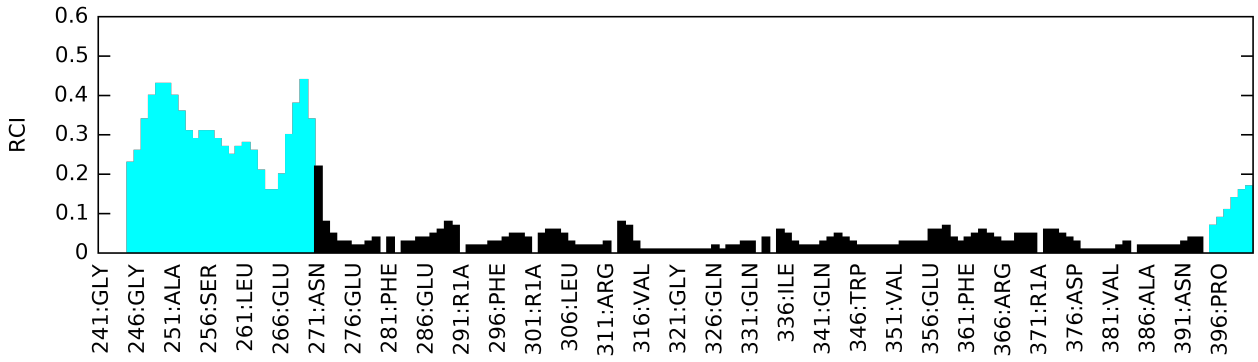
	Total	^{1}H	^{13}C	^{15}N
Aromatic	29/124 (23%)	28/66 (42%)	0/55 (0%)	1/3 (33%)
Overall	1090/1410 (77%)	561/714 (79%)	420/559 (75%)	109/137 (80%)

7.1.4 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	321	GLY	HA2	1.41	5.87 – 2.07	-6.7

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: