wwPDB NMR Structure Validation Summary Report

Jun 13, 2024 – 06:11 AM EDT

PDB ID	:	4CPG
BMRB ID	:	19779
Title	:	Solution structure of the SGTA N-terminal domain
Authors	:	Darby, J.F.; Krysztofinska, E.M.; Simpson, P.J.; Isaacson, R.L.
Deposited on	:	2014-02-06

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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

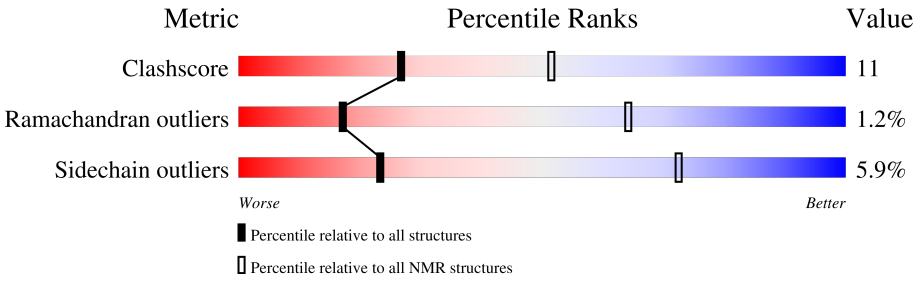
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 47%.

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	83	69% 13% • 17%
1	B	83	69% 14% 17%

2 Ensemble composition and analysis

This entry contains 20 models. Model 9 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:16-A:83, B:15-B:83 (137)	1.07	9

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 3 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2104 atoms, of which 1048 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called SMALL GLUTAMINE-RICH TETRATRICOPEPTIDE REPEAT-CONTAINING PROTEIN ALPHA.

Mol	Chain	Residues	Atoms						Trace
1	A	69	Total	C	H	N	O	S	0
			1052	331	524	88	107	2	
1	B	69	Total	C	H	N	O	S	0
			1052	331	524	88	107	2	

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	expression tag	UNP O43765
A	2	ALA	-	expression tag	UNP O43765
A	3	HIS	-	expression tag	UNP O43765
A	4	HIS	-	expression tag	UNP O43765
A	5	HIS	-	expression tag	UNP O43765
A	6	HIS	-	expression tag	UNP O43765
A	7	HIS	-	expression tag	UNP O43765
A	8	HIS	-	expression tag	UNP O43765
A	9	VAL	-	expression tag	UNP O43765
A	10	ASP	-	expression tag	UNP O43765
A	11	ASP	-	expression tag	UNP O43765
A	12	ASP	-	expression tag	UNP O43765
A	13	ASP	-	expression tag	UNP O43765
A	14	LYS	-	expression tag	UNP O43765
B	1	MET	-	expression tag	UNP O43765
B	2	ALA	-	expression tag	UNP O43765
B	3	HIS	-	expression tag	UNP O43765
B	4	HIS	-	expression tag	UNP O43765
B	5	HIS	-	expression tag	UNP O43765
B	6	HIS	-	expression tag	UNP O43765
B	7	HIS	-	expression tag	UNP O43765
B	8	HIS	-	expression tag	UNP O43765
B	9	VAL	-	expression tag	UNP O43765
B	10	ASP	-	expression tag	UNP O43765
B	11	ASP	-	expression tag	UNP O43765
B	12	ASP	-	expression tag	UNP O43765
B	13	ASP	-	expression tag	UNP O43765
B	14	LYS	-	expression tag	UNP O43765

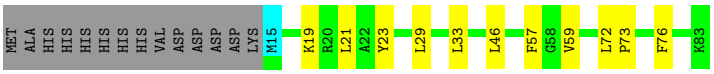
4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide 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.

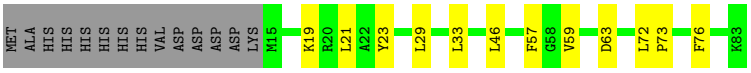
• Molecule 1: SMALL GLUTAMINE-RICH TETRATRICOPEPTIDE REPEAT-CONTAINING PROTEIN ALPHA

Chain A:



• Molecule 1: SMALL GLUTAMINE-RICH TETRATRICOPEPTIDE REPEAT-CONTAINING PROTEIN ALPHA

Chain B:



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

The representative model is number 9. Colouring as in section 4.1 above.

• Molecule 1: SMALL GLUTAMINE-RICH TETRATRICOPEPTIDE REPEAT-CONTAINING PROTEIN ALPHA

Chain A:

• Molecule 1: SMALL GLUTAMINE-RICH TETRATRICOPEPTIDE REPEAT-CONTAINING PROTEIN ALPHA

Chain B:


MET
ALA
HIS
HIS
HIS
HIS
HIS
HIS
VAL
ASP
ASP
ASP
ASP
LYS
M15
K19
R20
L21
A22
Y23
L29
Q32
L33
L46
T71
L72
F76
K83

5 Refinement protocol and experimental data overview

The models were refined using the following method: *ARIA*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *LOWEST ENERGY*.

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

Software name	Classification	Version
CNS	refinement	
NMRView	structure solution	
TALOS	structure solution	
ARIA	structure solution	

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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	861
Number of shifts mapped to atoms	860
Number of unparsed shifts	0
Number of shifts with mapping errors	1
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	47%

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	0.30±0.03	0±0/527 (0.0± 0.0%)	0.42±0.05	0±0/712 (0.0± 0.0%)
1	B	0.31±0.04	0±0/535 (0.0± 0.0%)	0.43±0.07	0±0/722 (0.0± 0.0%)
All	All	0.31	0/21240 (0.0%)	0.43	2/28680 (0.0%)

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	0.2±0.4
1	B	0.0±0.0	0.4±0.7
All	All	0	13

There are no bond-length outliers.

All 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	B	58	GLY	CA-C-N	-5.09	106.01	117.20	16	1
1	A	76	PHE	CB-CG-CD1	5.00	124.30	120.80	13	1

There are no chirality outliers.

5 of 11 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	62	GLU	Peptide	2
1	B	62	GLU	Peptide	2
1	A	60	THR	Peptide	1
1	B	60	THR	Peptide	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
1	B	56	ALA	Peptide	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	520	513	511	12±6
1	B	528	524	519	13±7
All	All	20960	20740	20600	456

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 11.

5 of 266 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:23:TYR:CE1	1:A:62:GLU:HB2	0.86	2.05	17	1
1:B:23:TYR:CE1	1:B:62:GLU:HB2	0.86	2.05	17	1
1:B:68:LEU:HD13	1:B:69:PRO:HD2	0.83	1.51	19	1
1:A:54:GLU:HA	1:A:57:PHE:O	0.82	1.75	20	1
1:A:68:LEU:HD13	1:A:69:PRO:HD2	0.81	1.53	19	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	67/83 (81%)	62±2 (93±3%)	4±2 (6±2%)	1±1 (1±2%)	16	63
1	B	67/83 (81%)	62±2 (93±3%)	4±2 (6±2%)	1±1 (1±2%)	17	64
All	All	2680/3320 (81%)	2486 (93%)	161 (6%)	33 (1%)	17	64

5 of 22 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	B	67	ALA	3
1	A	59	VAL	3
1	A	82	GLY	2
1	B	82	GLY	2
1	A	67	ALA	2

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	55/69 (80%)	52±2 (94±3%)	3±2 (6±3%)	24	73
1	B	56/69 (81%)	53±2 (94±3%)	3±2 (6±3%)	23	72
All	All	2220/2760 (80%)	2090 (94%)	130 (6%)	23	72

5 of 48 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	33	LEU	14
1	B	33	LEU	14
1	A	32	GLN	5
1	B	32	GLN	5
1	A	76	PHE	5

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 monosaccharides 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

The completeness of assignment taking into account all chemical shift lists is 47% for the well-defined parts and 47% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *bmrb_construct.csh*

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	861
Number of shifts mapped to atoms	860
Number of unparsed shifts	0
Number of shifts with mapping errors	1
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

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

- No matching atom found in the structure. All 1 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	15	MET	H	7.973	0.000	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	69	-0.62 ± 0.12	Should be checked
$^{13}C_{\beta}$	65	0.07 ± 0.08	None needed (< 0.5 ppm)
$^{13}C'$	69	-0.28 ± 0.06	None needed (< 0.5 ppm)
^{15}N	67	0.72 ± 0.35	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 47%, i.e. 850 atoms were assigned a chemical shift out of a possible 1813. 0 out of 24 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	340/685 (50%)	138/278 (50%)	136/274 (50%)	66/133 (50%)
Sidechain	466/1018 (46%)	319/665 (48%)	141/323 (44%)	6/30 (20%)
Aromatic	44/110 (40%)	22/54 (41%)	22/48 (46%)	0/8 (0%)
Overall	850/1813 (47%)	479/997 (48%)	299/645 (46%)	72/171 (42%)

7.1.4 Statistically unusual chemical shifts

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots

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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A: