wwPDB NMR Structure Validation Summary Report (i)

May 7, 2024 – 03:58 pm BST

PDB ID	:	4CH1
BMRB ID	:	19653
Title	:	RRM domain from C. elegans SUP-12 bound to GGTGTGC DNA
Authors	:	Amrane, S.; Mackereth, C.D.
Deposited on	:	2013-11-28

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	:	FAILED
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 86%.

There are no overall percentile quality scores available for this entry.

The sequence quality summary graphics cannot be shown.

2 Ensemble composition and analysis (i)

This entry contains 15 models. Model 12 is the overall representative, medoid model (most similar to other models). The authors have identified model 6 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:32-A:114 (83)	0.34	12

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, 5, 6, 7, 8, 9, 12, 13, 14, 15
2	3, 11
Single-model clusters	4; 10

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1734 atoms, of which 830 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called PROTEIN SUP-12, ISOFORM B.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	97	1509	473	749	137	146	4	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	25	GLY	-	expression tag	UNP H2L051
A	26	ALA	-	expression tag	UNP H2L051
A	27	MET	-	expression tag	UNP H2L051

- Molecule 2 is a DNA chain called GGTGTGC.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	7	225	69	81	27	42	6	0

4 Residue-property plots (i)

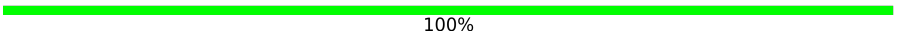
4.1 Average score per residue in the NMR ensemble

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: PROTEIN SUP-12, ISOFORM B

Chain A:

• Molecule 2: GGTGTGC

Chain B: 

There are no outlier residues in this chain.

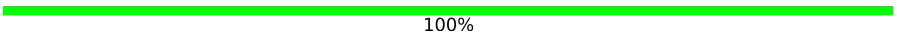
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 12. Colouring as in section 4.1 above.

• Molecule 1: PROTEIN SUP-12, ISOFORM B

Chain A:

• Molecule 2: GGTGTGC

Chain B: 

There are no outlier residues in this chain.

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *ARIA1.2*.

Of the 100 calculated structures, 15 were deposited, based on the following criterion: *LOWEST ENERGY AND NO VIOLATIONS*.

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

Software name	Classification	Version
CNS	refinement	
NMRPipe	structure solution	
Sparky	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	1249
Number of shifts mapped to atoms	1249
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	86%

6 Model quality (i)

6.1 Standard geometry (i)

MolProbity failed to run properly - this section will have to be empty.

6.2 Too-close contacts (i)

MolProbity failed to run properly - this section will have to be empty.

6.3 Torsion angles (i)

6.3.1 Protein backbone (i)

MolProbity failed to run properly - this section will have to be empty.

6.3.2 Protein sidechains (i)

MolProbity failed to run properly - this section will have to be empty.

6.3.3 RNA (i)

MolProbity failed to run properly - this section will have to be empty.

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

MolProbity failed to run properly - this section will have to be empty.

6.5 Carbohydrates (i)

MolProbity failed to run properly - this section will have to be empty.

6.6 Ligand geometry (i)

MolProbity failed to run properly - this section will have to be empty.

6.7 Other polymers (i)

MolProbity failed to run properly - this section will have to be empty.

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 86% for the well-defined parts and 86% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *SUP12_ggtgtgc_for_deposition.str.csh*

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	1249
Number of shifts mapped to atoms	1249
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	96	-0.27 ± 0.14	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	88	0.06 ± 0.09	None needed (< 0.5 ppm)
$^{13}C'$	90	0.26 ± 0.16	None needed (< 0.5 ppm)
^{15}N	90	0.07 ± 0.48	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 86%, i.e. 1090 atoms were assigned a chemical shift out of a possible 1262. 0 out of 9 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	409/416 (98%)	169/170 (99%)	161/166 (97%)	79/80 (99%)
Sidechain	537/604 (89%)	365/390 (94%)	165/187 (88%)	7/27 (26%)

Continued on next page...

Continued from previous page...

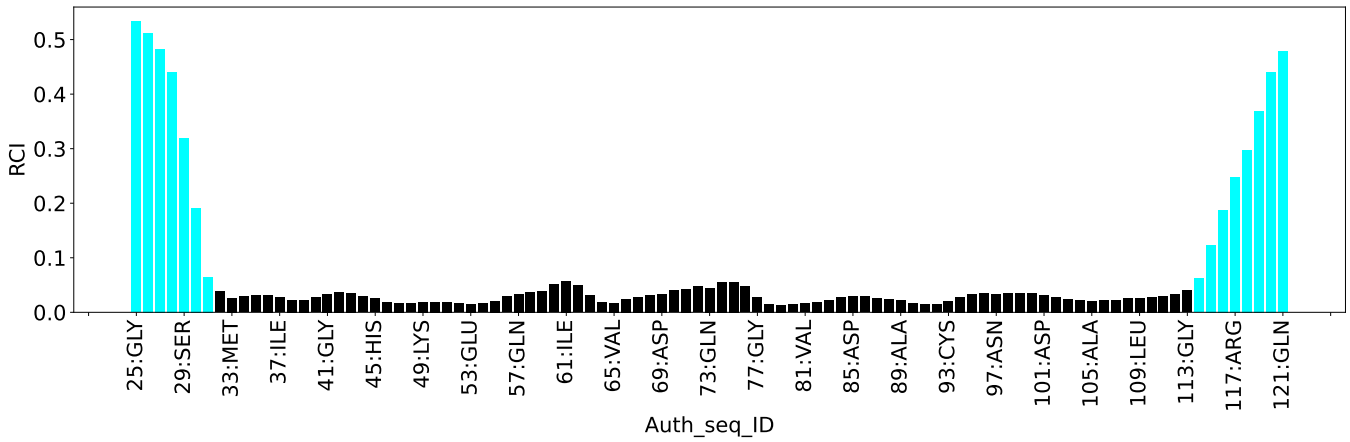
	Total	^{1}H	^{13}C	^{15}N
Aromatic	87/102 (85%)	45/49 (92%)	42/49 (86%)	0/4 (0%)
Sugar	47/84 (56%)	47/49 (96%)	0/35 (0%)	0/0 (—%)
Base	10/56 (18%)	10/35 (29%)	0/10 (0%)	0/11 (0%)
Overall	1090/1262 (86%)	636/693 (92%)	368/447 (82%)	86/122 (70%)

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.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	57	GLN	HG2	0.09	1.01 – 3.62	-8.5
1	A	92	ALA	HA	1.94	2.13 – 6.34	-5.5

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.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: