wwPDB NMR Structure Validation Summary Report

Oct 8, 2024 – 09:22 AM EDT

PDB ID	:	2NBI
BMRB ID	:	4958
Title	:	Structure of the PSCD-region of the cell wall protein pleuralin-1
Authors	:	De Sanctis, S.; Wenzler, M.; Kroeger, N.; Malloni, W.M.; Sumper, M.; Rainer, D.; Zadravec, P.; Brunner, E.; Kremer, W.; Kalbitzer, H.R.
Deposited on	:	2016-02-23

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:

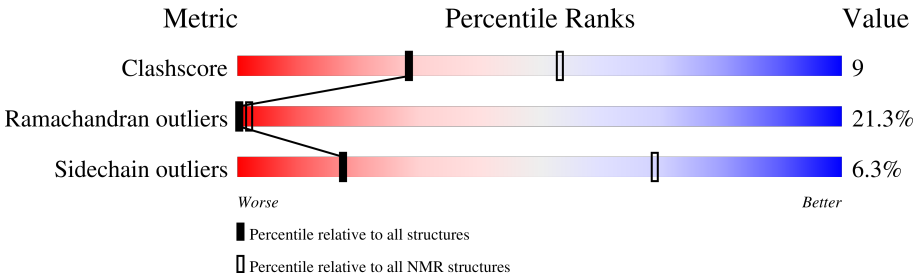
MolProbity	:	4.02b-467
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 18%.

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	210492	14027
Ramachandran outliers	207382	12486
Sidechain outliers	206894	12463

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	503	64% 27% 6%

2 Ensemble composition and analysis (i)

This entry contains 1 models. Identification of well-defined residues and clustering analysis are not possible.

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 6886 atoms, of which 3285 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called HEP200 protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	494	6886	2209	3285	576	754	62	0

There are 9 discrepancies between the modelled and reference sequences:

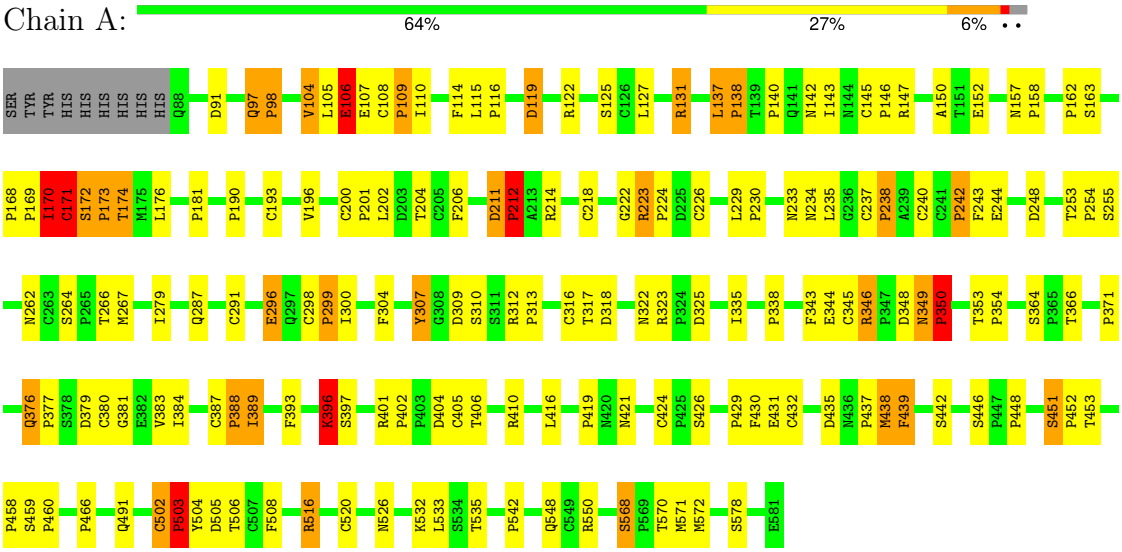
Chain	Residue	Modelled	Actual	Comment	Reference
A	1	SER	-	expression tag	UNP O22015
A	2	TYR	-	expression tag	UNP O22015
A	3	TYR	-	expression tag	UNP O22015
A	4	HIS	-	expression tag	UNP O22015
A	5	HIS	-	expression tag	UNP O22015
A	6	HIS	-	expression tag	UNP O22015
A	7	HIS	-	expression tag	UNP O22015
A	8	HIS	-	expression tag	UNP O22015
A	9	HIS	-	expression tag	UNP O22015

4 Residue-property plots (i)

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: HEP200 protein

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 2000 calculated structures, 1 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
CNS	structure solution	1.21
CNS	refinement	
AUREMOL	refinement	

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	1074
Number of shifts mapped to atoms	1046
Number of unparsed shifts	0
Number of shifts with mapping errors	28
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	18%

6 Model quality

6.1 Standard geometry

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	0.89	4/3735 (0.1%)	0.87	1/5181 (0.0%)
All	All	0.89	4/3735 (0.1%)	0.87	1/5181 (0.0%)

All bond outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	106	GLU	C-O	-6.54	1.10	1.23
1	A	170	ILE	CA-C	6.44	1.69	1.52
1	A	171	CYS	N-CA	6.39	1.59	1.46
1	A	350	PRO	N-CD	-6.11	1.39	1.47

All angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	350	PRO	N-CA-CB	-5.91	96.10	102.60

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	3601	3285	3284	64
All	All	3601	3285	3284	64

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 9.

5 of 64 clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)
1:A:98:PRO:HA	1:A:104:VAL:HG11	0.78	1.54
1:A:107:GLU:HG2	1:A:109:PRO:CD	0.76	2.11
1:A:107:GLU:HG2	1:A:109:PRO:HD2	0.71	1.63
1:A:108:CYS:SG	1:A:109:PRO:HD3	0.67	2.29
1:A:107:GLU:CB	1:A:170:ILE:H	0.66	2.03

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	492/503 (98%)	240 (49%)	147 (30%)	105 (21%)	0	2
All	All	492/503 (98%)	240 (49%)	147 (30%)	105 (21%)	0	2

5 of 105 Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
1	A	91	ASP
1	A	97	GLN
1	A	98	PRO
1	A	104	VAL
1	A	106	GLU

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	460/469 (98%)	431 (94%)	29 (6%)	17	69
All	All	460/469 (98%)	431 (94%)	29 (6%)	17	69

5 of 29 residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
1	A	122	ARG
1	A	127	LEU
1	A	131	ARG
1	A	142	ASN
1	A	143	ILE

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 oligosaccharides 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 18% for the well-defined parts and 18% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

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	1074
Number of shifts mapped to atoms	1046
Number of unparsed shifts	0
Number of shifts with mapping errors	28
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

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

- No matching atom found in the structure. First 5 (of 28) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	2	TYR	C	175.42	0.20	1
1	A	2	TYR	CA	57.75	0.20	1
1	A	2	TYR	CB	40.12	0.20	1
1	A	3	TYR	H	7.64	0.02	1
1	A	3	TYR	HA	4.77	0.05	1
1	A	3	TYR	C	176.1	0.20	1
1	A	3	TYR	CA	59.3	0.20	1
1	A	3	TYR	N	115.11	0.33	1
1	A	5	HIS	HA	3.97	0.05	1
1	A	5	HIS	HB2	3.11	0.05	2
1	A	5	HIS	HB3	3.2	0.05	2
1	A	5	HIS	C	176.39	0.20	1
1	A	5	HIS	CA	56.7	0.20	1
1	A	5	HIS	CB	30.07	0.20	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	8	HIS	HA	4.57	0.05	1
1	A	8	HIS	HB2	2.98	0.05	1
1	A	8	HIS	HB3	2.98	0.05	1
1	A	8	HIS	C	173.92	0.20	1
1	A	8	HIS	CA	55.82	0.20	1
1	A	8	HIS	CB	30.6	0.20	1
1	A	9	HIS	H	8.35	0.02	1
1	A	9	HIS	HA	4.65	0.05	1
1	A	9	HIS	HB2	3.08	0.05	1
1	A	9	HIS	HB3	3.08	0.05	1
1	A	9	HIS	C	175.59	0.20	1
1	A	9	HIS	CA	56.17	0.20	1
1	A	9	HIS	CB	30.74	0.20	1
1	A	9	HIS	N	120.44	0.33	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	105	0.34 ± 0.13	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	93	0.28 ± 0.17	None needed (< 0.5 ppm)
$^{13}C'$	91	-0.01 ± 0.14	None needed (< 0.5 ppm)
^{15}N	78	0.20 ± 0.34	None needed (< 0.5 ppm)

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 18%, i.e. 1042 atoms were assigned a chemical shift out of a possible 5832. 0 out of 40 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	441/2256 (20%)	179/888 (20%)	186/988 (19%)	76/380 (20%)
Sidechain	589/3369 (17%)	390/2192 (18%)	191/1095 (17%)	8/82 (10%)
Aromatic	12/207 (6%)	12/102 (12%)	0/105 (0%)	0/0 (—%)
Overall	1042/5832 (18%)	581/3182 (18%)	377/2188 (17%)	84/462 (18%)

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	401	ARG	NE	127.48	76.53 – 92.65	26.6
1	A	410	ARG	NE	124.58	76.53 – 92.65	24.8

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: