Full wwPDB NMR Structure Validation Report (i)

Apr 20, 2024 – 11:41 AM EDT

PDB ID	:	2N1U
BMRB ID	:	25575
Title	:	Structure of SAP30L corepressor protein
Authors	:	Tossavainen, H.; Permi, P.
Deposited on	:	2015-04-23

This is a Full wwPDB NMR Structure Validation 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:

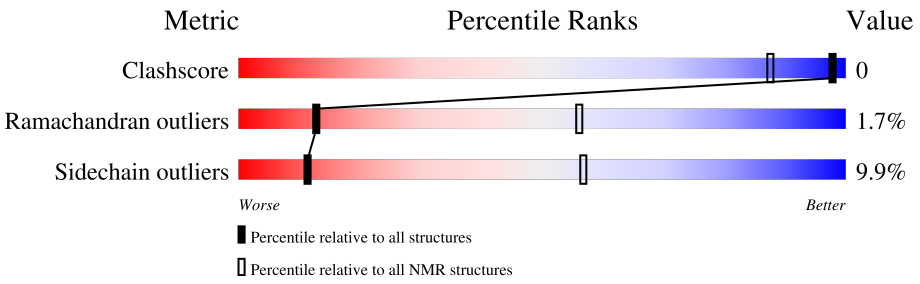
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson 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 81%.

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	70	74% 6% 20%

2 Ensemble composition and analysis (i)

This entry contains 15 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:27-A:82 (56)	0.44	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 3 clusters and 3 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1154 atoms, of which 587 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Histone deacetylase complex subunit SAP30L.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	70	1153	350	587	112	100	4	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	23	GLY	-	expression tag	UNP Q9HAJ7
A	24	SER	-	expression tag	UNP Q9HAJ7

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	1	1	1

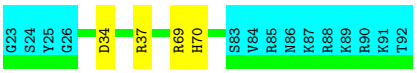
4 Residue-property plots (i)

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: Histone deacetylase complex subunit SAP30L

Chain A: 74% 6% 20%



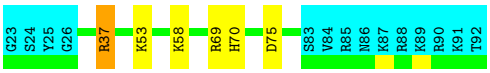
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Histone deacetylase complex subunit SAP30L

Chain A: 71% 7% • 20%



4.2.2 Score per residue for model 2 (medoid)

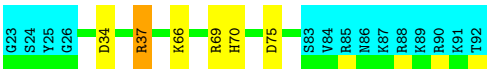
• Molecule 1: Histone deacetylase complex subunit SAP30L

Chain A: 64% 14% • 20%

4.2.3 Score per residue for model 3

- Molecule 1: Histone deacetylase complex subunit SAP30L

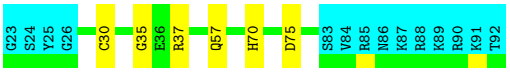
Chain A:



4.2.4 Score per residue for model 4

- Molecule 1: Histone deacetylase complex subunit SAP30L

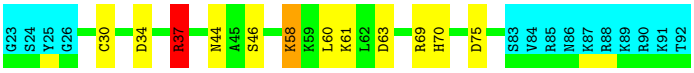
Chain A:



4.2.5 Score per residue for model 5

- Molecule 1: Histone deacetylase complex subunit SAP30L

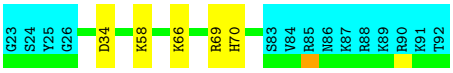
Chain A:



4.2.6 Score per residue for model 6

- Molecule 1: Histone deacetylase complex subunit SAP30L

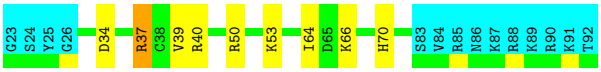
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: Histone deacetylase complex subunit SAP30L

Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: Histone deacetylase complex subunit SAP30L

Chain A: 73% 6% • 20%

G23 S24 Y25 G26 R37 C38 V39 S48 K61 H70 S83 V84 R85 N86 K87 R88 K89 R90 K91 T92

4.2.9 Score per residue for model 9

- Molecule 1: Histone deacetylase complex subunit SAP30L

Chain A: 69% 11% 20%

G23 S24 Y25 G26 Q27 S28 D34 R37 N44 A45 S46 K61 R69 H70 S83 V84 R85 N86 K87 R88 K89 R90 K91 T92

4.2.10 Score per residue for model 10

- Molecule 1: Histone deacetylase complex subunit SAP30L

Chain A: 69% 10% • 20%

G23 S24 Y25 G26 C30 D34 R37 C38 V39 R40 R50 K58 H70 S83 V84 R85 N86 K87 R88 K89 R90 K91 T92

4.2.11 Score per residue for model 11

- Molecule 1: Histone deacetylase complex subunit SAP30L

Chain A: 69% 11% 20%

G23 S24 Y25 G26 Q27 S28 S48 K61 L62 D63 I64 D65 K66 S67 H70 D75 S83 V84 R85 N86 K87 R88 K89 R90 K91 T92

4.2.12 Score per residue for model 12

- Molecule 1: Histone deacetylase complex subunit SAP30L

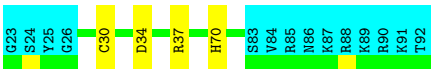
Chain A: 67% 13% 20%

G23 S24 Y25 G26 Q27 S28 C29 D34 G35 E36 R37 C38 K58 R69 H70 S83 V84 R85 N86 K87 R88 K89 R90 K91 T92

4.2.13 Score per residue for model 13

• Molecule 1: Histone deacetylase complex subunit SAP30L

Chain A: 74% 6% 20%



4.2.14 Score per residue for model 14

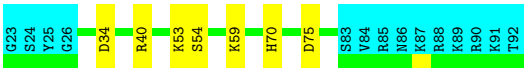
• Molecule 1: Histone deacetylase complex subunit SAP30L

Chain A: 67% 13% 20%

4.2.15 Score per residue for model 15

• Molecule 1: Histone deacetylase complex subunit SAP30L

Chain A: 70% 10% 20%



5 Refinement protocol and experimental data overview (i)

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

Of the 300 calculated structures, 15 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	
CYANA	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	795
Number of shifts mapped to atoms	795
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	81%

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: ZN

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	0.68±0.01	0±0/459 (0.0± 0.0%)	1.06±0.04	1±1/612 (0.2± 0.2%)
All	All	0.68	0/6885 (0.0%)	1.06	19/9180 (0.2%)

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
All	All	0	3

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	A	37	ARG	NE-CZ-NH1	10.00	125.30	120.30	9	10
1	A	40	ARG	NE-CZ-NH1	6.40	123.50	120.30	2	3
1	A	50	ARG	NE-CZ-NH1	5.52	123.06	120.30	10	2
1	A	37	ARG	NE-CZ-NH2	-5.45	117.58	120.30	8	2
1	A	69	ARG	NE-CZ-NH1	5.39	123.00	120.30	1	2

There are no chirality outliers.

All 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	37	ARG	Sidechain	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
1	A	40	ARG	Sidechain	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	452	462	461	0±1
All	All	6795	6930	6915	6

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 0.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:58:LYS:HE3	1:A:60:LEU:HD12	0.57	1.75	5	1
1:A:37:ARG:HH21	1:A:39:VAL:HG12	0.55	1.62	7	1
1:A:47:PHE:CE2	1:A:64:ILE:HD11	0.44	2.48	2	1
1:A:45:ALA:HB2	1:A:78:LYS:CE	0.41	2.44	2	1
1:A:31:LEU:HD21	1:A:47:PHE:CE1	0.40	2.51	14	1
1:A:64:ILE:O	1:A:66:LYS:HE2	0.40	2.15	7	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	56/70 (80%)	47±2 (85±4%)	8±2 (14±4%)	1±1 (2±2%)	13	56
All	All	840/1050 (80%)	710 (85%)	116 (14%)	14 (2%)	13	56

All 7 unique Ramachandran outliers are listed below. They are sorted by the frequency of occur-

rence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	69	ARG	6
1	A	44	ASN	2
1	A	61	LYS	2
1	A	35	GLY	1
1	A	65	ASP	1
1	A	29	CYS	1
1	A	38	CYS	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	52/64 (81%)	47±1 (90±3%)	5±1 (10±3%)	11 57
All	All	780/960 (81%)	703 (90%)	77 (10%)	11 57

All 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	70	HIS	15
1	A	34	ASP	11
1	A	37	ARG	7
1	A	75	ASP	6
1	A	58	LYS	5
1	A	30	CYS	4
1	A	53	LYS	3
1	A	66	LYS	3
1	A	28	SER	3
1	A	50	ARG	2
1	A	46	SER	2
1	A	61	LYS	2
1	A	63	ASP	2
1	A	39	VAL	2
1	A	48	SER	2
1	A	54	SER	2
1	A	69	ARG	1
1	A	57	GLN	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	44	ASN	1
1	A	67	SER	1
1	A	36	GLU	1
1	A	59	LYS	1

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)

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

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 81% for the well-defined parts and 80% 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 (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	795
Number of shifts mapped to atoms	795
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
$^{13}C_{\alpha}$	69	-0.27 ± 0.23	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	65	0.24 ± 0.16	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	62	-0.25 ± 0.59	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 81%, i.e. 645 atoms were assigned a chemical shift out of a possible 793. 0 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	213/280 (76%)	107/113 (95%)	56/112 (50%)	50/55 (91%)
Sidechain	385/459 (84%)	262/294 (89%)	121/140 (86%)	2/25 (8%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	47/54 (87%)	24/27 (89%)	23/24 (96%)	0/3 (0%)
Overall	645/793 (81%)	393/434 (91%)	200/276 (72%)	52/83 (63%)

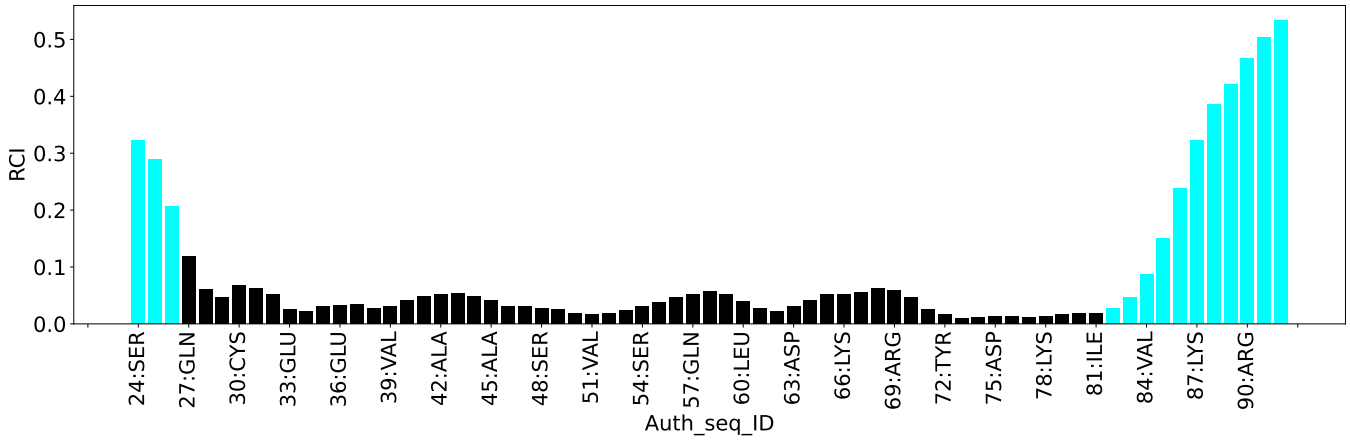
The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 80%, i.e. 795 atoms were assigned a chemical shift out of a possible 999. 0 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	264/352 (75%)	133/143 (93%)	69/140 (49%)	62/69 (90%)
Sidechain	476/584 (82%)	323/372 (87%)	151/174 (87%)	2/38 (5%)
Aromatic	55/63 (87%)	28/31 (90%)	27/29 (93%)	0/3 (0%)
Overall	795/999 (80%)	484/546 (89%)	247/343 (72%)	64/110 (58%)

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.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: