Full wwPDB NMR Structure Validation Report (i)

Apr 20, 2024 – 05:34 PM EDT

PDB ID	:	2KDP
BMRB ID	:	16127
Title	:	Solution Structure of the SAP30 zinc finger motif
Authors	:	He, Y.; Imhoff, R.; Sahu, A.; Radhakrishnan, I.
Deposited on	:	2009-01-14

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:

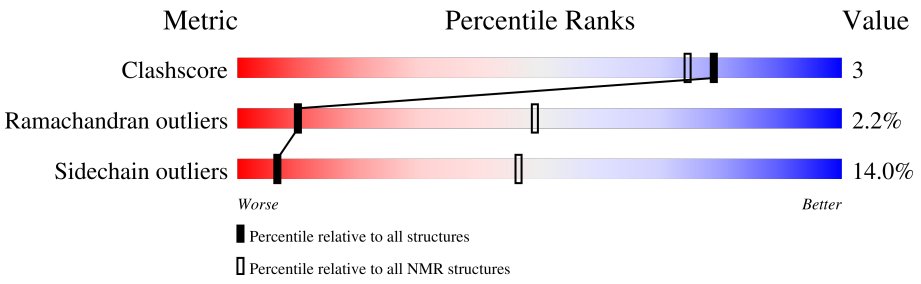
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson 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 85%.

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	71	66% 14% 20%

2 Ensemble composition and analysis (i)

This entry contains 47 models. Model 1 is the overall representative, medoid model (most similar to other models).

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:65-A:121 (57)	0.43	1

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 6 clusters and 12 single-model clusters were found.

Cluster number	Models
1	1, 4, 9, 12, 15, 22, 26, 27, 29, 30, 31, 37, 42, 45
2	10, 24, 32, 33, 36, 40, 41, 44
3	2, 3, 8, 16, 19, 20
4	13, 14, 25
5	5, 6
6	23, 34
Single-model clusters	7; 11; 17; 18; 21; 28; 35; 38; 39; 43; 46; 47

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1156 atoms, of which 589 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Histone deacetylase complex subunit SAP30.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	71	1155	341	589	119	102	4	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	61	SER	-	expression tag	UNP O75446
A	62	ASN	-	expression tag	UNP O75446
A	63	ALA	-	expression tag	UNP O75446

- 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 SAP30

Chain A: 66% 14% 20%

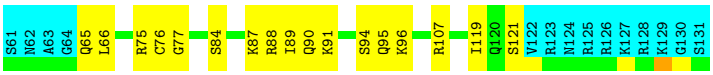
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1 (medoid)

• Molecule 1: Histone deacetylase complex subunit SAP30

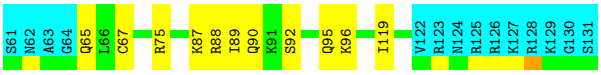
Chain A: 56% 24% 20%



4.2.2 Score per residue for model 2

• Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 65% 15% 20%



4.2.3 Score per residue for model 3

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 70% 10% 20%

4.2.4 Score per residue for model 4

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 59% 21% 20%

4.2.5 Score per residue for model 5

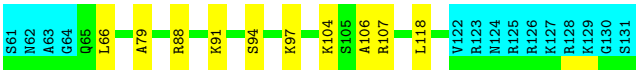
- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 69% 11% 20%

4.2.6 Score per residue for model 6

- Molecule 1: Histone deacetylase complex subunit SAP30

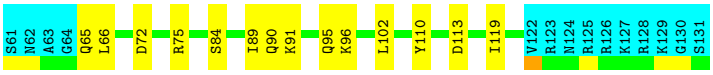
Chain A: 66% 14% 20%



4.2.7 Score per residue for model 7

- Molecule 1: Histone deacetylase complex subunit SAP30

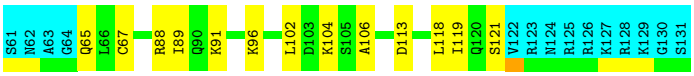
Chain A: 61% 20% 20%



4.2.8 Score per residue for model 8

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 62% 18% 20%



4.2.9 Score per residue for model 9

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 66% 14% 20%

4.2.10 Score per residue for model 10

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 61% 20% 20%

4.2.11 Score per residue for model 11

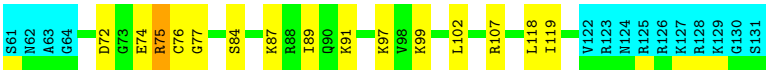
- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 63% 17% 20%

4.2.12 Score per residue for model 12

- Molecule 1: Histone deacetylase complex subunit SAP30

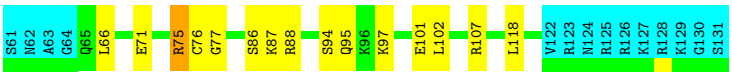
Chain A: 59% 20% • 20%



4.2.13 Score per residue for model 13

- Molecule 1: Histone deacetylase complex subunit SAP30

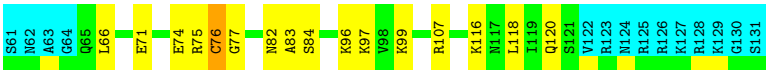
Chain A: 59% 20% • 20%



4.2.14 Score per residue for model 14

- Molecule 1: Histone deacetylase complex subunit SAP30

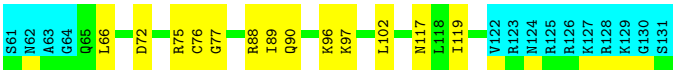
Chain A: 58% 21% • 20%



4.2.15 Score per residue for model 15

- Molecule 1: Histone deacetylase complex subunit SAP30

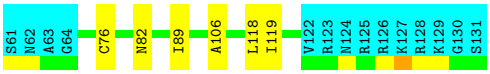
Chain A: 62% 18% 20%



4.2.16 Score per residue for model 16

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 72% 8% 20%



4.2.17 Score per residue for model 17

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 68% 13% 20%

4.2.18 Score per residue for model 18

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 54% 25% • 20%

4.2.19 Score per residue for model 19

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 68% 11% • 20%

4.2.20 Score per residue for model 20

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 68% 13% 20%

4.2.21 Score per residue for model 21

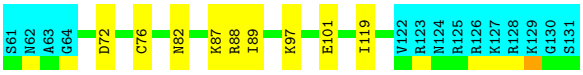
- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 61% 18% • 20%

4.2.22 Score per residue for model 22

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 68% 13% 20%



4.2.23 Score per residue for model 23

• Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 65% 15% 20%

4.2.24 Score per residue for model 24

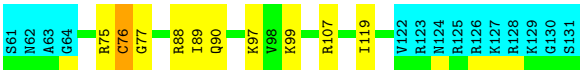
• Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 55% 25% 20%

4.2.25 Score per residue for model 25

• Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 66% 13% • 20%



4.2.26 Score per residue for model 26

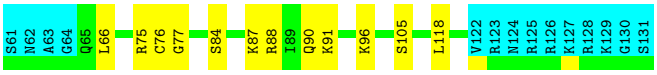
• Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 68% 13% 20%

4.2.27 Score per residue for model 27

• Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 63% 17% 20%



4.2.28 Score per residue for model 28

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 62% 17% • 20%

4.2.29 Score per residue for model 29

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 66% 13% • 20%

4.2.30 Score per residue for model 30

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 63% 17% 20%

4.2.31 Score per residue for model 31

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 68% 13% 20%

4.2.32 Score per residue for model 32

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 61% 20% 20%

4.2.33 Score per residue for model 33

• Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 61% 20% 20%

S61 N62 A63 G64 R75 C76 G77 N82 K87 R88 I89 K97 V98 K99 K104 R107 K116 I119 Q120 S121 V122 R123 N124 R125 R126 K127 R128 K129 G130 S131

4.2.34 Score per residue for model 34

• Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 63% 17% 20%

S61 N62 A63 G64 R75 C76 G77 K87 R88 I89 Q90 L102 D103 K104 H108 K116 N117 L118 S121 V122 R123 N124 R125 R126 K127 R128 K129 G130 S131

4.2.35 Score per residue for model 35

• Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 58% 21% • 20%

S61 N62 A63 G64 C68 R75 C76 G77 K87 R88 I89 Q90 K91 K97 L102 D103 A106 R107 L118 I119 V122 R123 N124 R125 R126 K127 R128 K129 G130 S131

4.2.36 Score per residue for model 36

• Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 68% 13% 20%

S61 N62 A63 G64 Q65 L66 D72 R75 C76 G77 K97 R107 H108 L118 V122 R123 N124 R125 R126 K127 R128 K129 G130 S131

4.2.37 Score per residue for model 37

• Molecule 1: Histone deacetylase complex subunit SAP30

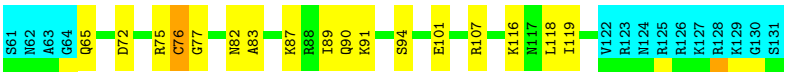
Chain A: 61% 18% • 20%

S61 N62 A63 G64 Q65 L66 C76 N82 A83 R88 I89 Q90 L102 D103 R107 D113 K116 I119 V122 R123 N124 R125 R126 K127 R128 K129 G130 S131

4.2.38 Score per residue for model 38

- Molecule 1: Histone deacetylase complex subunit SAP30

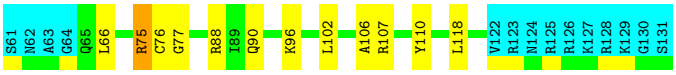
Chain A: 56% 23% • 20%



4.2.39 Score per residue for model 39

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 63% 15% • 20%



4.2.40 Score per residue for model 40

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 63% 15% • 20%

4.2.41 Score per residue for model 41

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 61% 17% • 20%

4.2.42 Score per residue for model 42

- Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 62% 17% • 20%

4.2.43 Score per residue for model 43

• Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 55% 23% • 20%

4.2.44 Score per residue for model 44

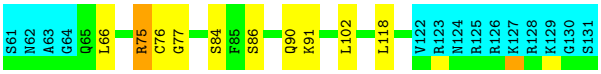
• Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 56% 24% 20%

4.2.45 Score per residue for model 45

• Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 66% 13% • 20%



4.2.46 Score per residue for model 46

• Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 63% 17% 20%

4.2.47 Score per residue for model 47

• Molecule 1: Histone deacetylase complex subunit SAP30

Chain A: 63% 15% • 20%

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *molecular dynamics*.

Of the 3000 calculated structures, 47 were deposited, based on the following criterion: *structures with acceptable covalent geometry and in good agreement with experimental data*.

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

Software name	Classification	Version
TALOS	structure solution	
ARIA	structure solution	
ARIA	refinement	
CNS	structure solution	
CNS	refinement	
Procheck	refinement	
ISD	refinement	1.1

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	818
Number of shifts mapped to atoms	818
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	85%

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

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	455	467	466	3±1
All	All	21432	21949	21902	123

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 3.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:89:ILE:HG23	1:A:119:ILE:HG21	0.79	1.54	41	28
1:A:65:GLN:O	1:A:66:LEU:HD12	0.70	1.87	29	1
1:A:118:LEU:O	1:A:118:LEU:HD23	0.62	1.94	13	24
1:A:83:ALA:HB2	1:A:116:LYS:HD3	0.62	1.72	38	4
1:A:75:ARG:O	1:A:77:GLY:N	0.58	2.36	38	30
1:A:66:LEU:HD12	1:A:79:ALA:HA	0.57	1.77	6	1
1:A:102:LEU:HD12	1:A:103:ASP:H	0.52	1.64	18	2
1:A:102:LEU:HD12	1:A:103:ASP:N	0.47	2.24	35	2
1:A:118:LEU:HD23	1:A:118:LEU:C	0.47	2.29	35	10
1:A:113:ASP:N	1:A:113:ASP:OD1	0.46	2.49	28	1
1:A:96:LYS:NZ	1:A:119:ILE:HG23	0.46	2.26	2	1
1:A:68:CYS:SG	1:A:106:ALA:HB1	0.46	2.50	35	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:96:LYS:HE2	1:A:119:ILE:HG23	0.44	1.88	9	1
1:A:82:ASN:N	1:A:82:ASN:OD1	0.44	2.50	43	1
1:A:87:LYS:N	1:A:87:LYS:HD3	0.44	2.28	40	1
1:A:67:CYS:O	1:A:75:ARG:NH2	0.43	2.51	2	1
1:A:71:GLU:N	1:A:74:GLU:O	0.43	2.52	47	1
1:A:118:LEU:HD23	1:A:118:LEU:O	0.43	2.14	10	1
1:A:117:ASN:ND2	1:A:121:SER:OG	0.42	2.53	32	2
1:A:88:ARG:NE	1:A:89:ILE:HG13	0.42	2.30	19	1
1:A:69:LEU:HD12	1:A:115:HIS:ND1	0.41	2.30	24	1
1:A:85:PHE:CD2	1:A:102:LEU:HD13	0.41	2.49	28	1
1:A:92:SER:OG	1:A:96:LYS:NZ	0.41	2.53	21	2
1:A:82:ASN:ND2	1:A:82:ASN:O	0.41	2.53	22	1
1:A:102:LEU:HG	1:A:103:ASP:N	0.41	2.31	37	1
1:A:103:ASP:OD1	1:A:105:SER:N	0.41	2.54	47	1
1:A:74:GLU:O	1:A:76:CYS:N	0.40	2.52	29	1
1:A:83:ALA:HB2	1:A:116:LYS:CD	0.40	2.47	14	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	57/71 (80%)	48±2 (85±4%)	7±2 (13±4%)	1±1 (2±1%)	10	49
All	All	2679/3337 (80%)	2271 (85%)	348 (13%)	60 (2%)	10	49

All 9 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	76	CYS	36
1	A	106	ALA	11
1	A	110	TYR	4
1	A	65	GLN	2
1	A	75	ARG	2
1	A	107	ARG	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	67	CYS	1
1	A	108	HIS	1
1	A	77	GLY	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	50/61 (82%)	43±3 (86±5%)	7±3 (14±5%)	6 46
All	All	2350/2867 (82%)	2022 (86%)	328 (14%)	6 46

All 32 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	107	ARG	27
1	A	90	GLN	26
1	A	87	LYS	25
1	A	88	ARG	23
1	A	91	LYS	22
1	A	102	LEU	19
1	A	97	LYS	17
1	A	66	LEU	15
1	A	72	ASP	15
1	A	82	ASN	13
1	A	113	ASP	13
1	A	75	ARG	11
1	A	101	GLU	10
1	A	104	LYS	10
1	A	84	SER	9
1	A	65	GLN	8
1	A	99	LYS	8
1	A	95	GLN	7
1	A	121	SER	6
1	A	74	GLU	6
1	A	120	GLN	6
1	A	94	SER	5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	86	SER	5
1	A	76	CYS	5
1	A	117	ASN	4
1	A	116	LYS	4
1	A	71	GLU	2
1	A	92	SER	2
1	A	105	SER	2
1	A	78	ARG	1
1	A	108	HIS	1
1	A	96	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 85% for the well-defined parts and 81% 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	818
Number of shifts mapped to atoms	818
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
$^{13}C_{\alpha}$	70	-0.23 ± 0.16	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	65	-0.09 ± 0.12	None needed (< 0.5 ppm)
$^{13}C'$	65	-0.09 ± 0.13	None needed (< 0.5 ppm)
^{15}N	66	-1.66 ± 0.53	Should be applied

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 85%, i.e. 677 atoms were assigned a chemical shift out of a possible 801. 0 out of 6 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	277/288 (96%)	114/117 (97%)	109/114 (96%)	54/57 (95%)
Sidechain	374/469 (80%)	245/300 (82%)	123/141 (87%)	6/28 (21%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	26/44 (59%)	11/21 (52%)	11/19 (58%)	4/4 (100%)
Overall	677/801 (85%)	370/438 (84%)	243/274 (89%)	64/89 (72%)

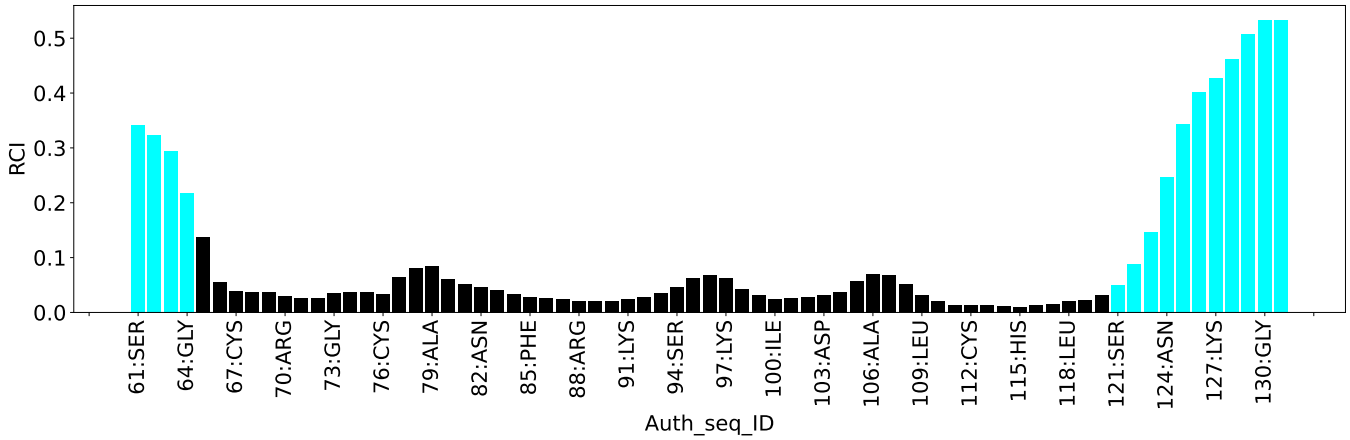
The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 81%, i.e. 817 atoms were assigned a chemical shift out of a possible 1005. 0 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	343/360 (95%)	142/147 (97%)	135/142 (95%)	66/71 (93%)
Sidechain	448/601 (75%)	289/382 (76%)	151/175 (86%)	8/44 (18%)
Aromatic	26/44 (59%)	11/21 (52%)	11/19 (58%)	4/4 (100%)
Overall	817/1005 (81%)	442/550 (80%)	297/336 (88%)	78/119 (66%)

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: