



wwPDB NMR Structure Validation Summary Report ⓘ

Jun 15, 2024 – 02:30 PM EDT

PDB ID : 2K7L
BMRB ID : 15919
Title : NMR structure of a complex formed by the C-terminal domain of human RAP74 and a phosphorylated peptide from the central domain of the FCP1
Authors : Yang, A.; Abbott, K.L.; Desjardins, A.; Di Lello, P.; Omichinski, J.G.; Legault, P.
Deposited on : 2008-08-13

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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 2022.3.0, CSD as543be (2022)
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) : 2.37.1

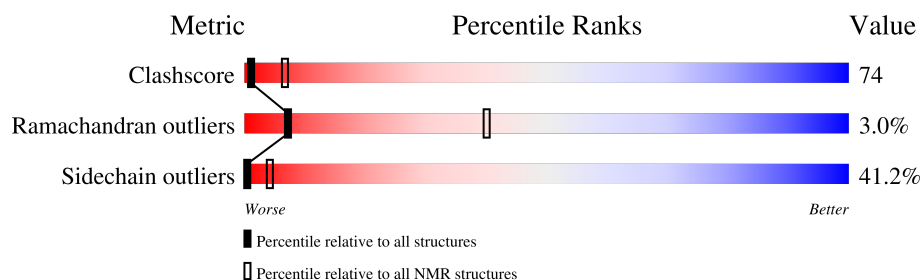
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 72%.

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	B	19	
2	A	67	

2 Ensemble composition and analysis

This entry contains 21 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	B:587-B:598, A:454-A:477, A:484-A:517 (70)	0.33	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 5 clusters and 4 single-model clusters were found.

Cluster number	Models
1	1, 7, 8, 9, 15, 18, 19, 20
2	5, 6, 11
3	12, 17
4	10, 14
5	4, 16
Single-model clusters	2; 3; 13; 21

3 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 1465 atoms, of which 746 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called centFCP1-T584PO4 peptide.

Mol	Chain	Residues	Atoms							Trace
1	B	19	Total	C	H	N	O	P		0
			318	101	151	24	41	1		

- Molecule 2 is a protein called General transcription factor IIF subunit 1.

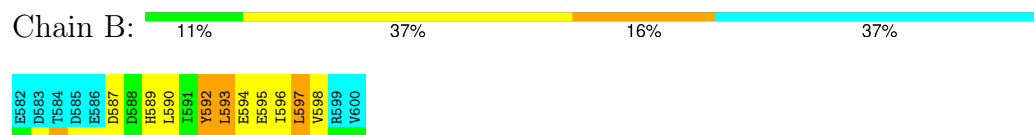
Mol	Chain	Residues	Atoms							Trace
2	A	67	Total	C	H	N	O	S		0
			1147	346	595	101	102	3		

4 Residue-property plots

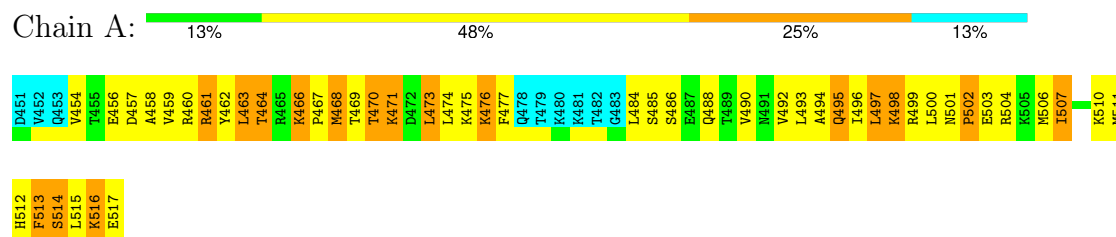
4.1 Average score per residue in the NMR ensemble

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: centFCP1-T584PO4 peptide



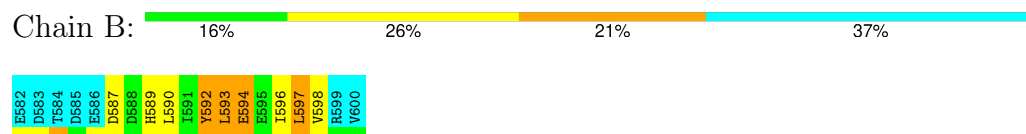
- Molecule 2: General transcription factor IIF subunit 1



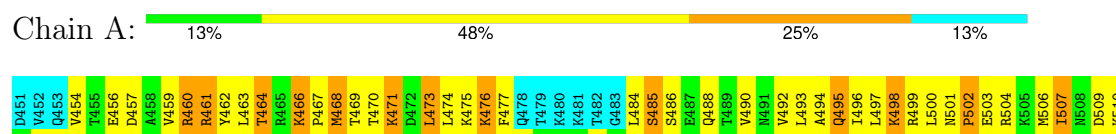
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 1. Colouring as in section 4.1 above.

- Molecule 1: centFCP1-T584PO4 peptide



- Molecule 2: General transcription factor IIF subunit 1



MS11
HS12
FS13
SS14
LS15
KS16
ES17

5 Refinement protocol and experimental data overview

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

Of the 21 calculated structures, 21 were deposited, based on the following criterion: *all calculated structures submitted*.

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

Software name	Classification	Version
CNS	structure solution	1.1
CNS	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	1019
Number of shifts mapped to atoms	867
Number of unparsed shifts	0
Number of shifts with mapping errors	152
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	72%

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: TPO

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	B	103	101	100	33±3
2	A	483	521	520	76±9
All	All	12306	13062	13020	1871

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 74.

5 of 367 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:A:463:LEU:HD11	2:A:500:LEU:HD22	1.06	1.23	17	16
2:A:474:LEU:HD22	2:A:490:VAL:HG22	1.06	1.27	5	11
1:B:593:LEU:HD21	2:A:470:THR:HG21	1.02	1.29	7	4
2:A:484:LEU:HD23	2:A:492:VAL:HG11	1.01	1.29	5	11
2:A:463:LEU:HD11	2:A:500:LEU:HD13	0.99	1.29	4	9

6.3 Torsion angles

6.3.1 Protein backbone

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	B	12/19 (63%)	10±0 (83±0%)	1±0 (11±4%)	1±0 (6±4%)	3	20
2	A	57/67 (85%)	51±2 (89±3%)	5±2 (8±3%)	1±0 (2±1%)	9	46
All	All	1449/1806 (80%)	1279 (88%)	126 (9%)	44 (3%)	7	40

All 4 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	A	502	PRO	21
1	B	597	LEU	15
2	A	485	SER	6
2	A	467	PRO	2

6.3.2 Protein sidechains

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	B	12/18 (67%)	8±1 (70±8%)	4±1 (30±8%)	1	16
2	A	56/64 (88%)	32±2 (56±3%)	24±2 (44±3%)	0	3
All	All	1428/1722 (83%)	839 (59%)	589 (41%)	0	4

5 of 47 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	B	592	TYR	21
1	B	593	LEU	21
2	A	457	ASP	21

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
2	A	461	ARG	21
2	A	468	MET	21

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
1	TPO	B	584	1	8,10,11	1.10±0.02	0±0 (0±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
1	TPO	B	584	1	10,14,16	1.96±0.08	1±0 (10±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	TPO	B	584	1	-	0±0,9,11,13	-

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)	Models	
								Worst	Total
1	B	584	TPO	P-OG1-CB	5.96	107.14	123.33	19	21

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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 72% for the well-defined parts and 69% 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	1019
Number of shifts mapped to atoms	867
Number of unparsed shifts	0
Number of shifts with mapping errors	152
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. First 5 (of 152) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	1	SER	H	8.279	0.02	1
1	A	1	SER	HA	4.483	0.02	1
1	A	1	SER	HB2	3.842	0.02	2
1	A	1	SER	HB3	3.842	0.02	2
1	A	1	SER	C	174.184	0.05	1
1	A	1	SER	CA	58.128	0.05	1
1	A	1	SER	CB	63.892	0.05	1
1	A	1	SER	N	117.058	0.05	1
1	A	2	THR	H	8.152	0.02	1
1	A	2	THR	HA	4.66	0.02	1
1	A	2	THR	HB	4.18	0.02	1
1	A	2	THR	HG21	1.26	0.02	1
1	A	2	THR	HG22	1.26	0.02	1
1	A	2	THR	HG23	1.26	0.02	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	2	THR	CA	59.858	0.05	1
1	A	2	THR	CB	69.849	0.05	1
1	A	2	THR	CG2	21.52	0.05	1
1	A	2	THR	N	118.415	0.05	1
1	A	3	PRO	HA	4.386	0.02	1
1	A	3	PRO	HB2	1.87	0.02	2
1	A	3	PRO	HB3	2.266	0.02	2
1	A	3	PRO	HD2	3.679	0.02	2
1	A	3	PRO	HD3	3.842	0.02	2
1	A	3	PRO	HG2	1.976	0.02	2
1	A	3	PRO	C	176.533	0.05	1
1	A	3	PRO	CA	63.065	0.05	1
1	A	3	PRO	CB	32.236	0.05	1
1	A	3	PRO	CD	51.034	0.05	1
1	A	3	PRO	CG	27.349	0.05	1
1	A	4	GLN	H	8.455	0.02	1
1	A	4	GLN	HA	4.593	0.02	1
1	A	4	GLN	HB2	1.94	0.02	2
1	A	4	GLN	HB3	2.105	0.02	2
1	A	4	GLN	HE21	7.541	0.02	2
1	A	4	GLN	HE22	6.811	0.02	2
1	A	4	GLN	HG2	2.35	0.02	2
1	A	4	GLN	HG3	2.35	0.02	2
1	A	4	GLN	CA	53.517	0.05	1
1	A	4	GLN	CB	29.117	0.05	1
1	A	4	GLN	CG	33.94	0.05	1
1	A	4	GLN	N	122.187	0.05	1
1	A	4	GLN	NE2	112.912	0.05	1
1	A	5	PRO	HA	4.749	0.02	1
1	A	5	PRO	HB2	1.96	0.02	2
1	A	5	PRO	HB3	2.402	0.02	2
1	A	5	PRO	HD2	3.677	0.02	2
1	A	5	PRO	HD3	3.85	0.02	2
1	A	5	PRO	HG2	2.076	0.02	2
1	A	5	PRO	HG3	2.076	0.02	2
1	A	5	PRO	CA	60.84	0.05	1
1	A	5	PRO	CB	30.846	0.05	1
1	A	6	PRO	HA	4.44	0.02	1
1	A	6	PRO	HB2	1.94	0.02	2
1	A	6	PRO	HB3	2.302	0.02	2
1	A	6	PRO	HD2	3.643	0.02	2

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	6	PRO	HD3	3.806	0.02	2
1	A	6	PRO	HG2	2.03	0.02	2
1	A	6	PRO	HG3	2.03	0.02	2
1	A	6	PRO	C	177.09	0.05	1
1	A	6	PRO	CA	63.065	0.05	1
1	A	6	PRO	CB	31.999	0.05	1
1	A	6	PRO	CD	50.47	0.05	1
1	A	6	PRO	CG	27.349	0.05	1
1	A	7	SER	H	8.398	0.02	1
1	A	7	SER	HA	4.422	0.02	1
1	A	7	SER	HB2	3.878	0.02	2
1	A	7	SER	HB3	3.878	0.02	2
1	A	7	SER	C	175.235	0.05	1
1	A	7	SER	CA	58.321	0.05	1
1	A	7	SER	CB	64.085	0.05	1
1	A	7	SER	N	115.996	0.05	1
1	A	8	GLY	H	8.407	0.02	1
1	A	8	GLY	HA2	3.969	0.02	2
1	A	8	GLY	HA3	3.969	0.02	2
1	A	8	GLY	C	173.968	0.05	1
1	A	8	GLY	CA	45.256	0.05	1
1	A	8	GLY	N	110.891	0.05	1
1	A	9	LYS	H	8.189	0.02	1
1	A	9	LYS	HA	4.404	0.02	1
1	A	9	LYS	HB2	1.831	0.02	2
1	A	9	LYS	HB3	1.831	0.02	2
1	A	9	LYS	HD2	1.668	0.02	2
1	A	9	LYS	HD3	1.668	0.02	2
1	A	9	LYS	HE2	2.972	0.02	2
1	A	9	LYS	HE3	2.972	0.02	2
1	A	9	LYS	HG2	1.396	0.02	2
1	A	9	LYS	HG3	1.396	0.02	2
1	A	9	LYS	C	176.811	0.05	1
1	A	9	LYS	CA	56.207	0.05	1
1	A	9	LYS	CB	33.344	0.05	1
1	A	9	LYS	CD	29.041	0.05	1
1	A	9	LYS	CE	42.199	0.05	1
1	A	9	LYS	CG	24.905	0.05	1
1	A	9	LYS	N	120.757	0.05	1
1	A	10	THR	H	8.236	0.02	1
1	A	10	THR	HA	4.386	0.02	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	10	THR	HB	4.186	0.02	1
1	A	10	THR	HG21	1.16	0.02	1
1	A	10	THR	HG22	1.16	0.02	1
1	A	10	THR	HG23	1.16	0.02	1
1	A	10	THR	C	174.463	0.05	1
1	A	10	THR	CA	61.587	0.05	1
1	A	10	THR	CB	70.04	0.05	1
1	A	10	THR	CG2	21.71	0.05	1
1	A	10	THR	N	115.64	0.05	1
1	A	11	THR	H	8.243	0.02	1
1	A	11	THR	HA	4.61	0.02	1
1	A	11	THR	HB	4.13	0.02	1
1	A	11	THR	HG21	1.26	0.02	1
1	A	11	THR	HG22	1.26	0.02	1
1	A	11	THR	HG23	1.26	0.02	1
1	A	11	THR	CA	59.858	0.05	1
1	A	11	THR	CB	69.656	0.05	1
1	A	11	THR	CG2	21.52	0.05	1
1	A	11	THR	N	118.913	0.05	1
1	A	12	PRO	HA	4.404	0.02	1
1	A	12	PRO	HB2	1.903	0.02	2
1	A	12	PRO	HB3	2.284	0.02	2
1	A	12	PRO	HD2	3.715	0.02	2
1	A	12	PRO	HD3	3.86	0.02	2
1	A	12	PRO	HG2	2.012	0.02	2
1	A	12	PRO	C	176.657	0.05	1
1	A	12	PRO	CA	63.316	0.05	1
1	A	12	PRO	CB	32.191	0.05	1
1	A	12	PRO	CD	51.034	0.05	1
1	A	12	PRO	CG	27.537	0.05	1
1	A	13	ASN	H	8.57	0.02	1
1	A	13	ASN	HA	4.712	0.02	1
1	A	13	ASN	HB2	2.768	0.02	2
1	A	13	ASN	HB3	2.851	0.02	2
1	A	13	ASN	HD21	7.66	0.02	2
1	A	13	ASN	HD22	6.934	0.02	2
1	A	13	ASN	C	175.39	0.05	1
1	A	13	ASN	CA	53.325	0.05	1
1	A	13	ASN	CB	38.915	0.05	1
1	A	13	ASN	N	119.127	0.05	1
1	A	13	ASN	ND2	113.375	0.05	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	14	SER	H	8.33	0.02	1
1	A	14	SER	HA	4.404	0.02	1
1	A	14	SER	HB2	3.832	0.02	2
1	A	14	SER	HB3	3.878	0.02	2
1	A	14	SER	C	175.05	0.05	1
1	A	14	SER	CA	58.705	0.05	1
1	A	14	SER	CB	63.7	0.05	1
1	A	14	SER	N	116.494	0.05	1
1	A	15	GLY	H	8.399	0.02	1
1	A	15	GLY	HA2	3.951	0.02	2
1	A	15	GLY	HA3	3.951	0.02	2
1	A	15	GLY	C	173.845	0.05	1
1	A	15	GLY	CA	45.448	0.05	1
1	A	15	GLY	N	110.702	0.05	1
1	A	451	ASP	H	8.159	0.02	1

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	82	-0.47 ± 0.15	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	79	-0.05 ± 0.22	None needed (< 0.5 ppm)
$^{13}\text{C}'$	75	-0.36 ± 0.14	None needed (< 0.5 ppm)
^{15}N	76	0.57 ± 0.28	None needed (imprecise)

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 72%, i.e. 755 atoms were assigned a chemical shift out of a possible 1055. 0 out of 16 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	282/346 (82%)	113/138 (82%)	113/140 (81%)	56/68 (82%)
Sidechain	464/655 (71%)	317/424 (75%)	142/203 (70%)	5/28 (18%)
Aromatic	9/54 (17%)	9/26 (35%)	0/24 (0%)	0/4 (0%)
Overall	755/1055 (72%)	439/588 (75%)	255/367 (69%)	61/100 (61%)

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:

