



Full wwPDB NMR Structure Validation Report ⓘ

Mar 8, 2026 – 07:42 AM UTC

PDB ID : 5ZUX / pdb_00005zux
BMRB ID : 36186
Title : Solution Structure of the DNA complex of the C-terminal Domain of Rok
Authors : Xia, B.; Duan, B.
Deposited on : 2018-05-08

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 ⓘ 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 : 4-5-2 with Phenix2.0
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
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) : 2.49

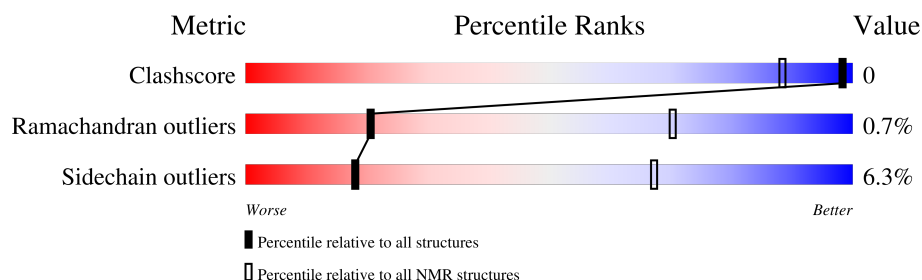
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 70%.

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	229148	14424
Ramachandran outliers	224038	12848
Sidechain outliers	223484	12823

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	A	92	76% 5% 18%
2	B	18	100%
2	C	18	89% 11%

2 Ensemble composition and analysis

This entry contains 20 models. Model 18 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:109-A:183 (75)	0.67	18

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 4 clusters and 2 single-model clusters were found.

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

3 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 2663 atoms, of which 1185 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Rok.

Mol	Chain	Residues	Atoms						Trace
1	A	92	Total	C	H	N	O	S	0
			1517	463	771	145	135	3	

- Molecule 2 is a DNA chain called DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3').

Mol	Chain	Residues	Atoms						Trace
2	B	18	Total	C	H	N	O	P	0
			573	178	207	65	106	17	
2	C	18	Total	C	H	N	O	P	0
			573	178	207	65	106	17	

4 Residue-property plots [i](#)

4.1 Average score per residue in the NMR ensemble

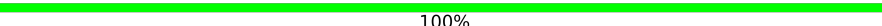
These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for 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: Rok

Chain A: 



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B: 

There are no outlier residues in this chain.

- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C: 



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: Rok

Chain A: 



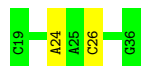
- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  83% 11% 6%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  89% 11%



4.2.2 Score per residue for model 2

- Molecule 1: Rok

Chain A:  75% 18%

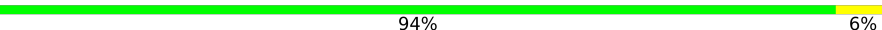


- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  89% 11%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  94% 6%



4.2.3 Score per residue for model 3

- Molecule 1: Rok

Chain A:  77% 18%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

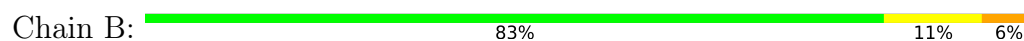


4.2.4 Score per residue for model 4

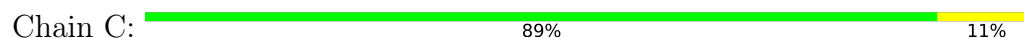
- Molecule 1: Rok



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')



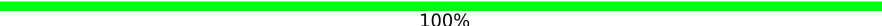
4.2.5 Score per residue for model 5

- Molecule 1: Rok

Chain A:  75% 5% 18%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  100%

There are no outlier residues in this chain.

- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  89% 11%



4.2.6 Score per residue for model 6

- Molecule 1: Rok

Chain A:  71% 9% 18%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  89% 11%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  94% 6%



4.2.7 Score per residue for model 7

- Molecule 1: Rok

Chain A:  74% 18%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  72% 28%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  83% 17%



4.2.8 Score per residue for model 8

- Molecule 1: Rok

Chain A:  74% 8% 18%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  89% 11%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  83% 17%



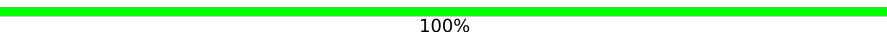
4.2.9 Score per residue for model 9

- Molecule 1: Rok

Chain A:  72% 8% 18%



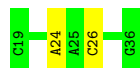
- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  100%

There are no outlier residues in this chain.

- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  89% 11%



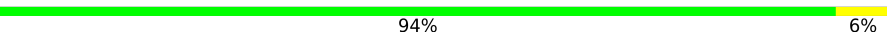
4.2.10 Score per residue for model 10

- Molecule 1: Rok

Chain A:  75% 5% 18%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  94% 6%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  83% 17%



4.2.11 Score per residue for model 11

- Molecule 1: Rok

Chain A:  72% 7% 18%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  89% 11%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  78% 22%



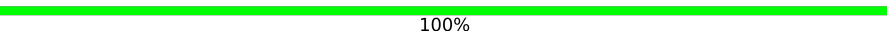
4.2.12 Score per residue for model 12

- Molecule 1: Rok

Chain A:  71% 9% 18%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  100%

There are no outlier residues in this chain.

- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  72% 28%



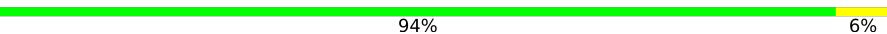
4.2.13 Score per residue for model 13

- Molecule 1: Rok

Chain A:  72% 10% 18%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  94% 6%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  83% 17%



4.2.14 Score per residue for model 14

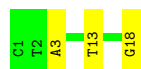
- Molecule 1: Rok

Chain A:  70% 11% 18%

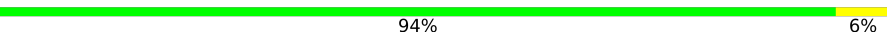


- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  83% 17%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  94% 6%



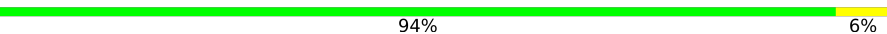
4.2.15 Score per residue for model 15

- Molecule 1: Rok

Chain A:  73% 8% 18%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  94% 6%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  89% 11%



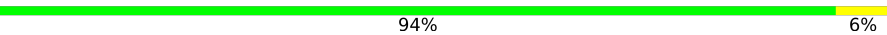
4.2.16 Score per residue for model 16

- Molecule 1: Rok

Chain A:  72% 8% 18%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  94% 6%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  89% 11%



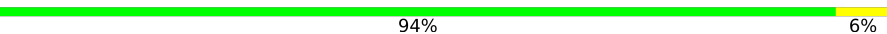
4.2.17 Score per residue for model 17

- Molecule 1: Rok

Chain A:  71% 9% 18%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  94% 6%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  83% 17%



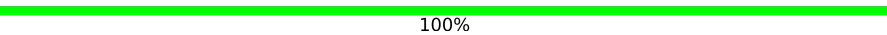
4.2.18 Score per residue for model 18 (medoid)

- Molecule 1: Rok

Chain A:  71% 10% 18%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  100%

There are no outlier residues in this chain.

- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  89% 11%



4.2.19 Score per residue for model 19

- Molecule 1: Rok

Chain A:  72% 9% 18%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  89% 11%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  78% 22%



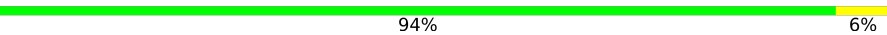
4.2.20 Score per residue for model 20

- Molecule 1: Rok

Chain A:  74% 8% 18%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B:  94% 6%



- Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C:  83% 17%



5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 20 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
Amber	refinement	
CYANA	structure calculation	

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	1409
Number of shifts mapped to atoms	1409
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	70%

6 Model quality (i)

6.1 Standard geometry (i)

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	0.95±0.02	0±0/600 (0.0± 0.0%)	1.55±0.03	4±2/802 (0.5± 0.2%)
2	B	0.78±0.02	0±0/410 (0.0± 0.0%)	1.38±0.02	0±0/631 (0.0± 0.1%)
2	C	0.76±0.01	0±0/410 (0.0± 0.0%)	1.36±0.01	0±0/631 (0.0± 0.0%)
All	All	0.85	0/28400 (0.0%)	1.44	82/41280 (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	0.3±0.6
2	B	0.0±0.0	1.4±1.2
2	C	0.0±0.0	2.5±1.0
All	All	0	85

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(°)	Ideal(°)	Models	
								Worst	Total
1	A	109	GLY	N-CA-C	8.66	121.86	112.33	9	3
1	A	114	GLY	N-CA-C	7.02	120.01	111.93	6	10
1	A	152	ILE	N-CA-C	6.58	114.07	107.55	18	5
1	A	121	ARG	NE-CZ-NH2	6.53	125.08	119.20	15	3
1	A	173	TYR	N-CA-C	6.48	117.68	108.74	14	11
1	A	130	ARG	NE-CZ-NH2	6.23	124.81	119.20	7	6
1	A	113	PRO	CA-C-N	5.88	125.68	120.10	17	4
1	A	113	PRO	C-N-CA	5.88	125.68	120.10	17	4
1	A	136	LYS	CA-C-N	5.80	129.52	120.82	6	3
1	A	136	LYS	C-N-CA	5.80	129.52	120.82	6	3
1	A	131	HIS	CB-CG-CD2	-5.58	123.95	131.20	20	4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	112	ARG	NE-CZ-NH2	5.48	124.13	119.20	2	6
1	A	158	PHE	N-CA-C	5.42	116.88	110.97	8	2
1	A	144	ILE	CA-CB-CG1	5.33	119.45	110.40	7	1
2	B	14	DA	C5'-C4'-C3'	-5.32	106.92	114.90	1	2
2	B	17	DA	N1-C6-N6	-5.30	111.05	119.00	19	3
1	A	166	TYR	CA-C-N	5.24	124.96	119.82	7	2
1	A	166	TYR	C-N-CA	5.24	124.96	119.82	7	2
1	A	157	THR	N-CA-CB	5.22	117.58	110.01	6	1
1	A	116	LYS	CA-C-N	5.09	128.74	120.60	16	3
1	A	116	LYS	C-N-CA	5.09	128.74	120.60	16	3
1	A	111	ALA	N-CA-C	5.02	115.93	108.60	19	1

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)
2	C	26	DC	Sidechain	14
2	C	24	DA	Sidechain	13
2	B	18	DG	Sidechain	10
2	C	23	DT	Sidechain	6
2	B	12	DT	Sidechain	5
2	B	14	DA	Sidechain	5
1	A	174	ARG	Sidechain	5
2	C	20	DT	Sidechain	5
2	C	30	DT	Sidechain	4
2	C	27	DT	Sidechain	3
2	B	16	DT	Sidechain	3
2	C	36	DG	Sidechain	3
2	B	13	DT	Sidechain	2
2	B	17	DA	Sidechain	2
2	C	35	DA	Sidechain	1
1	A	130	ARG	Sidechain	1
1	A	158	PHE	Sidechain	1
2	B	3	DA	Sidechain	1
2	C	28	DA	Sidechain	1

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	591	626	625	1±0
All	All	26460	20800	20780	14

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:173:TYR:CG	1:A:174:ARG:N	0.47	2.83	5	14

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	A	75/92 (82%)	71±1 (95±1%)	3±1 (4±1%)	1±0 (1±1%)	20	70
All	All	1500/1840 (82%)	1423 (95%)	66 (4%)	11 (1%)	20	70

All 2 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	174	ARG	9
1	A	153	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	A	63/79 (80%)	59±2 (94±3%)	4±2 (6±3%)	18 67
All	All	1260/1580 (80%)	1180 (94%)	80 (6%)	18 67

All 16 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	156	THR	20
1	A	112	ARG	11
1	A	125	ILE	11
1	A	120	LEU	5
1	A	141	GLN	5
1	A	143	GLU	5
1	A	121	ARG	5
1	A	152	ILE	4
1	A	122	GLU	4
1	A	157	THR	3
1	A	127	THR	2
1	A	144	ILE	1
1	A	115	SER	1
1	A	116	LYS	1
1	A	132	ASN	1
1	A	150	LEU	1

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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 70% for the well-defined parts and 68% for the entire structure.

7.1 Chemical shift list 1

File name: `working_cs.cif`

Chemical shift list name: *starch_output*

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	1409
Number of shifts mapped to atoms	1409
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

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	88	-0.41 ± 0.11	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	81	-0.03 ± 0.09	None needed (< 0.5 ppm)
$^{13}\text{C}'$	62	-0.22 ± 0.14	None needed (< 0.5 ppm)
^{15}N	78	0.10 ± 0.29	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 70%, i.e. 1227 atoms were assigned a chemical shift out of a possible 1749. 0 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	347/372 (93%)	151/151 (100%)	129/150 (86%)	67/71 (94%)
Sidechain	569/636 (89%)	386/412 (94%)	178/197 (90%)	5/27 (19%)

Continued on next page...

Continued from previous page...

	Total	¹ H	¹³ C	¹⁵ N
Aromatic	11/45 (24%)	11/21 (52%)	0/22 (0%)	0/2 (0%)
Sugar	250/432 (58%)	250/252 (99%)	0/180 (0%)	0/0 (—%)
Base	50/264 (19%)	50/156 (32%)	0/68 (0%)	0/40 (0%)
Overall	1227/1749 (70%)	848/992 (85%)	307/617 (50%)	72/140 (51%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 68%, i.e. 1367 atoms were assigned a chemical shift out of a possible 2015. 0 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	¹ H	¹³ C	¹⁵ N
Backbone	404/458 (88%)	176/186 (95%)	150/184 (82%)	78/88 (89%)
Sidechain	651/768 (85%)	441/495 (89%)	205/234 (88%)	5/39 (13%)
Aromatic	12/93 (13%)	12/45 (27%)	0/34 (0%)	0/14 (0%)
Sugar	250/432 (58%)	250/252 (99%)	0/180 (0%)	0/0 (—%)
Base	50/264 (19%)	50/156 (32%)	0/68 (0%)	0/40 (0%)
Overall	1367/2015 (68%)	929/1134 (82%)	355/700 (51%)	83/181 (46%)

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:

