Full wwPDB NMR Structure Validation Report ⓘ

Oct 7, 2025 – 10:31 AM JST

PDB ID : 9ILL / pdb_00009ill
BMRB ID : 51661
Title : monomeric SarA-E89Q in complex with DNA
Authors : Xia, B.; Fu, D.H.
Deposited on : 2024-06-30

The following versions of software and data (see references ⓘ) were used in the production of this report:

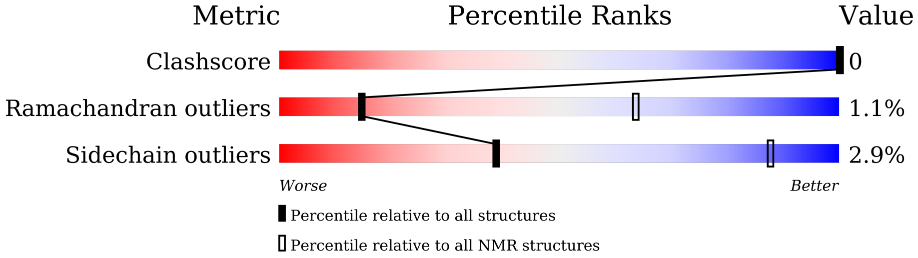
MolProbity	:	4-5-2 with Phenix2.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.46

1 Overall quality at a glance

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

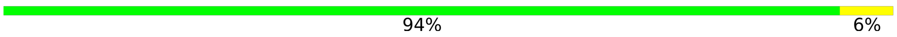
The overall completeness of chemical shifts assignment is 66%.

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	114	
2	B	18	
2	C	18	

2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 8 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:4-A:97 (94)	0.42	8

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 7 single-model clusters were found.

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

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3111 atoms, of which 1407 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Transcriptional regulator SarA.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	114	1970	618	998	174	179	1	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP Q7A732
A	71	GLN	GLU	engineered mutation	UNP Q7A732
A	107	LEU	-	expression tag	UNP Q7A732
A	108	GLU	-	expression tag	UNP Q7A732
A	109	HIS	-	expression tag	UNP Q7A732
A	110	HIS	-	expression tag	UNP Q7A732
A	111	HIS	-	expression tag	UNP Q7A732
A	112	HIS	-	expression tag	UNP Q7A732
A	113	HIS	-	expression tag	UNP Q7A732
A	114	HIS	-	expression tag	UNP Q7A732

- Molecule 2 is a DNA chain called DNA.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	18	570	176	204	67	106	17	0
			Total	C	H	N	O	P	
2	C	18	571	176	205	67	106	17	0

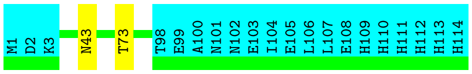
4 Residue-property plots ⓘ

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.

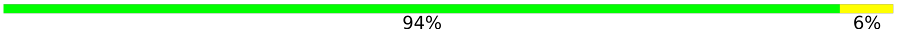
- Molecule 1: Transcriptional regulator SarA

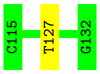
Chain A:



- Molecule 2: DNA

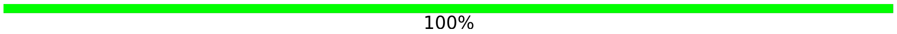
Chain B:





- Molecule 2: DNA

Chain C:



There are no outlier residues in this chain.

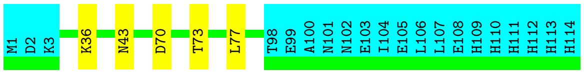
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

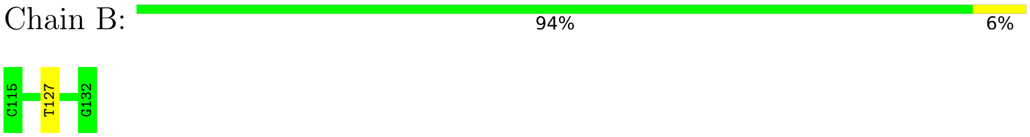
- Molecule 1: Transcriptional regulator SarA

Chain A:



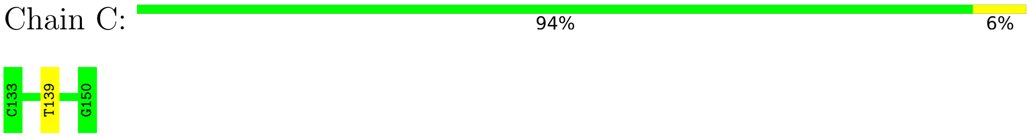
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

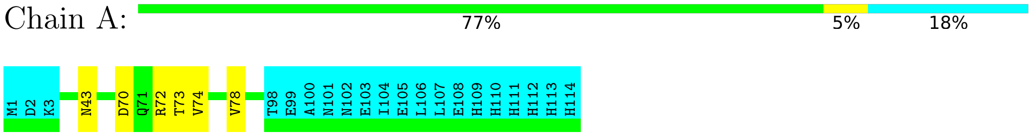
Chain C:



4.2.2 Score per residue for model 2

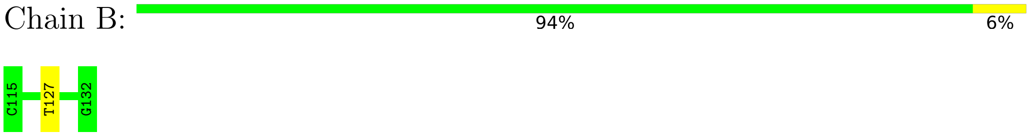
- Molecule 1: Transcriptional regulator SarA

Chain A:



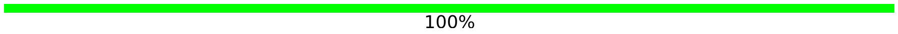
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

Chain C:

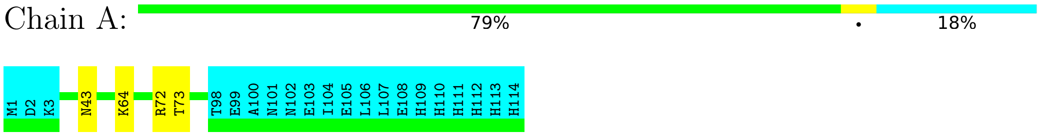


There are no outlier residues in this chain.

4.2.3 Score per residue for model 3

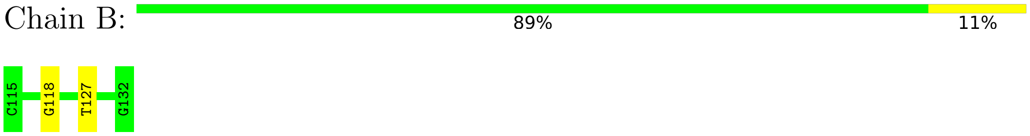
- Molecule 1: Transcriptional regulator SarA

Chain A:



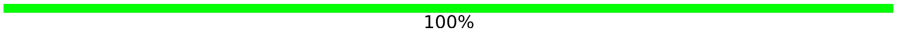
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

Chain C:

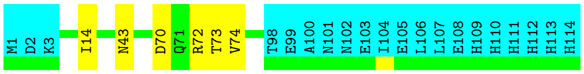


There are no outlier residues in this chain.

4.2.4 Score per residue for model 4

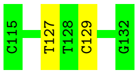
- Molecule 1: Transcriptional regulator SarA

Chain A:



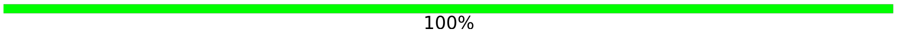
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

Chain C:

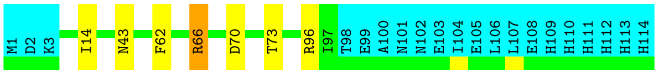


There are no outlier residues in this chain.

4.2.5 Score per residue for model 5

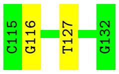
- Molecule 1: Transcriptional regulator SarA

Chain A:



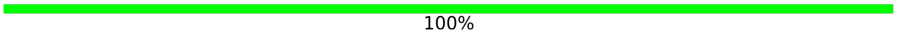
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

Chain C:

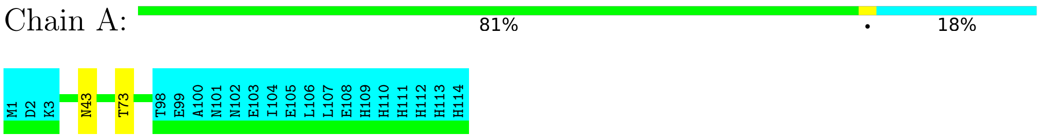


There are no outlier residues in this chain.

4.2.6 Score per residue for model 6

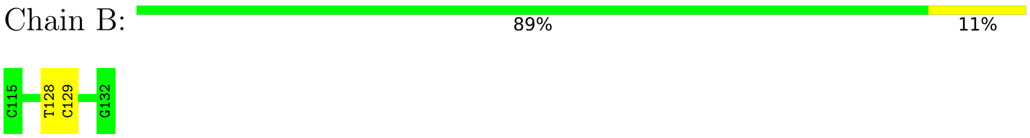
- Molecule 1: Transcriptional regulator SarA

Chain A:



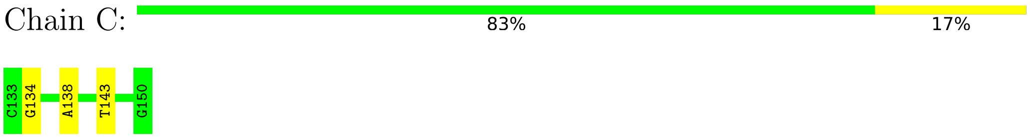
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

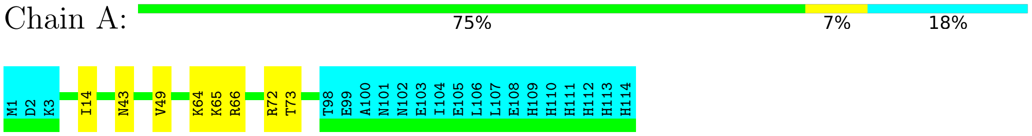
Chain C:



4.2.7 Score per residue for model 7

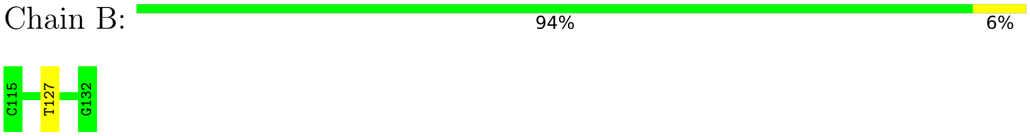
- Molecule 1: Transcriptional regulator SarA

Chain A:



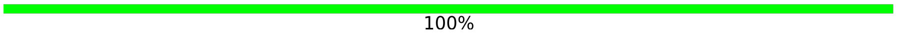
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

Chain C:

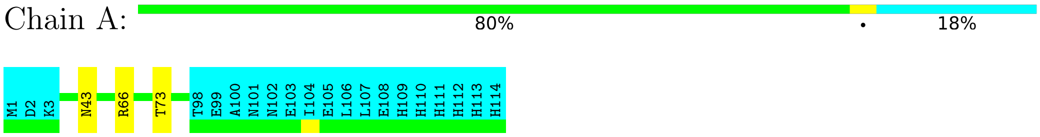


There are no outlier residues in this chain.

4.2.8 Score per residue for model 8 (medoid)

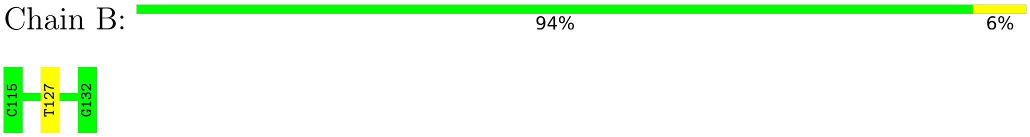
- Molecule 1: Transcriptional regulator SarA

Chain A:



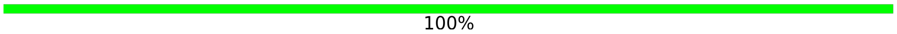
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

Chain C:

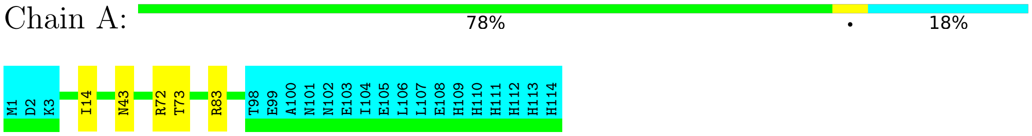


There are no outlier residues in this chain.

4.2.9 Score per residue for model 9

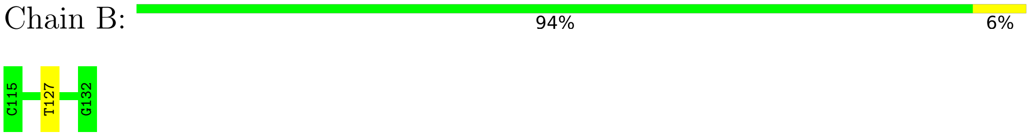
- Molecule 1: Transcriptional regulator SarA

Chain A:



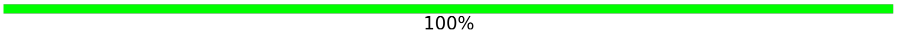
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

Chain C:

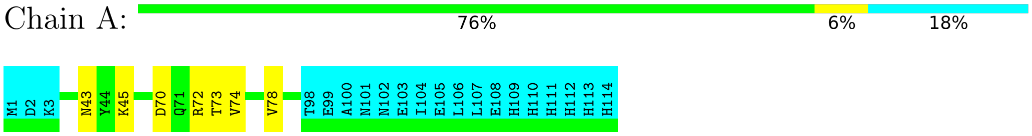


There are no outlier residues in this chain.

4.2.10 Score per residue for model 10

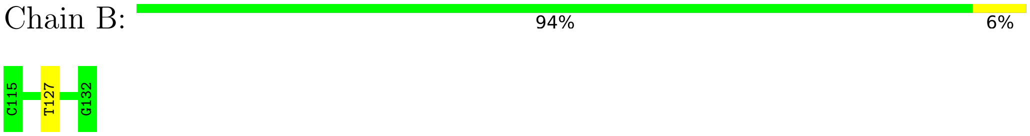
- Molecule 1: Transcriptional regulator SarA

Chain A:



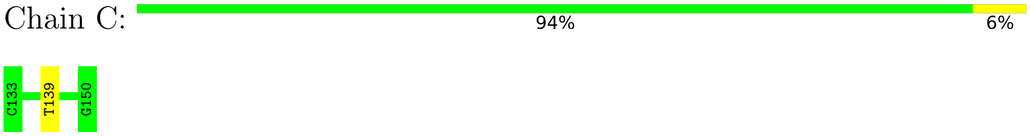
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

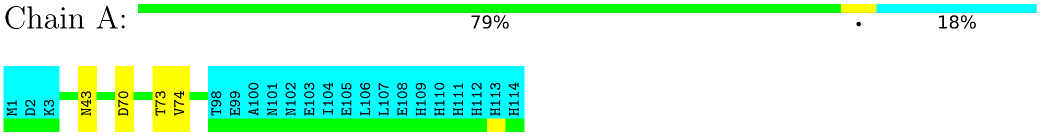
Chain C:



4.2.11 Score per residue for model 11

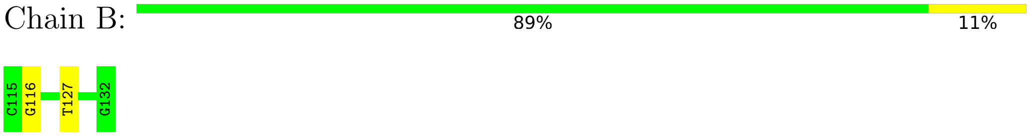
- Molecule 1: Transcriptional regulator SarA

Chain A:



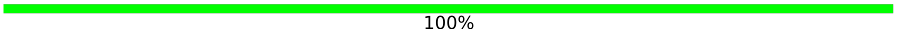
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

Chain C:

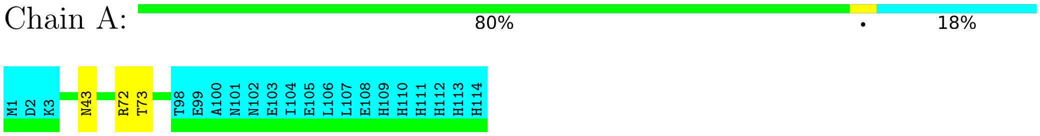


There are no outlier residues in this chain.

4.2.12 Score per residue for model 12

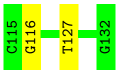
- Molecule 1: Transcriptional regulator SarA

Chain A:



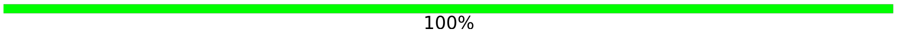
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

Chain C:

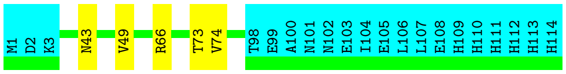


There are no outlier residues in this chain.

4.2.13 Score per residue for model 13

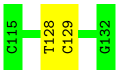
- Molecule 1: Transcriptional regulator SarA

Chain A:



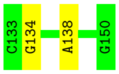
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

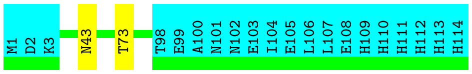
Chain C:



4.2.14 Score per residue for model 14

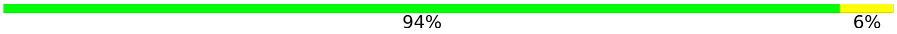
- Molecule 1: Transcriptional regulator SarA

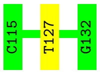
Chain A:



- Molecule 2: DNA

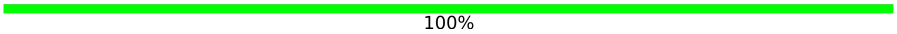
Chain B:





- Molecule 2: DNA

Chain C:

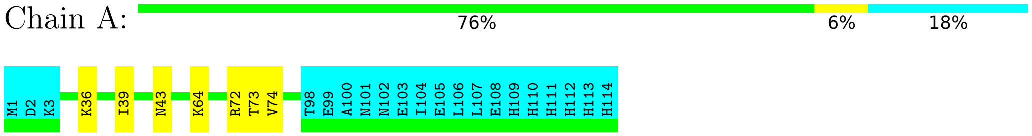


There are no outlier residues in this chain.

4.2.15 Score per residue for model 15

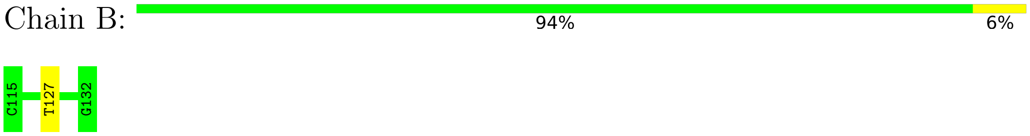
- Molecule 1: Transcriptional regulator SarA

Chain A:



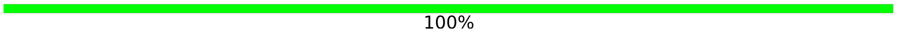
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

Chain C:

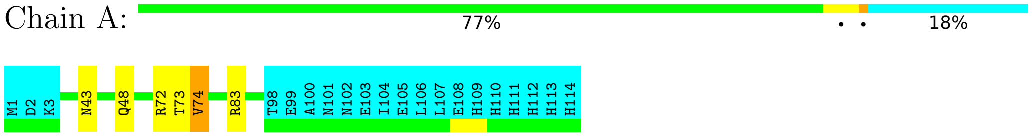


There are no outlier residues in this chain.

4.2.16 Score per residue for model 16

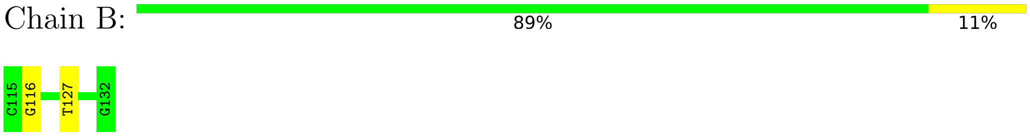
- Molecule 1: Transcriptional regulator SarA

Chain A:



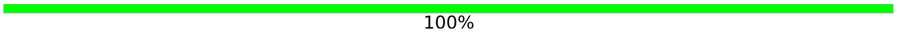
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

Chain C:

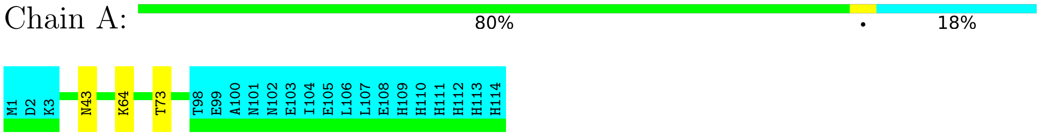


There are no outlier residues in this chain.

4.2.17 Score per residue for model 17

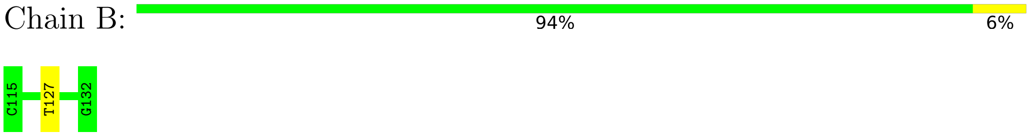
- Molecule 1: Transcriptional regulator SarA

Chain A:



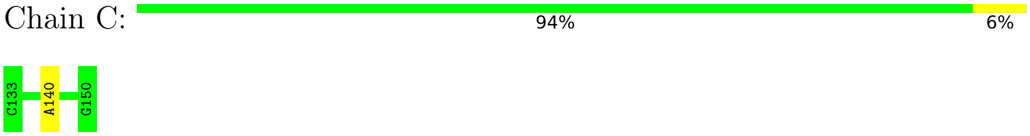
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

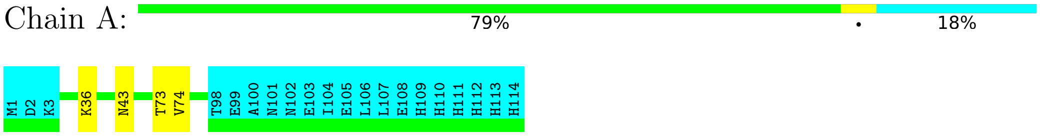
Chain C:



4.2.18 Score per residue for model 18

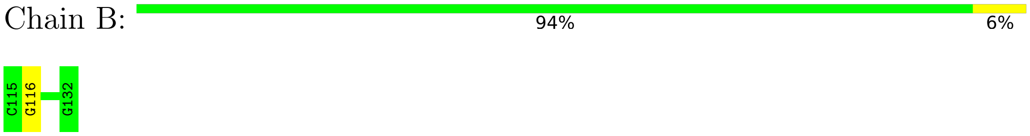
- Molecule 1: Transcriptional regulator SarA

Chain A:



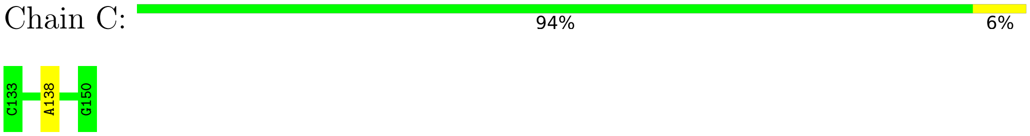
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

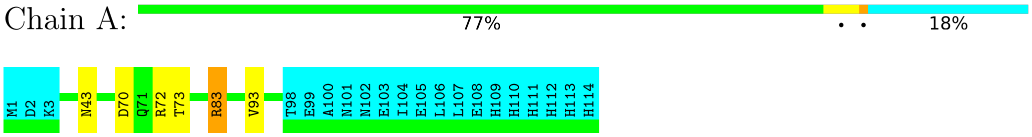
Chain C:



4.2.19 Score per residue for model 19

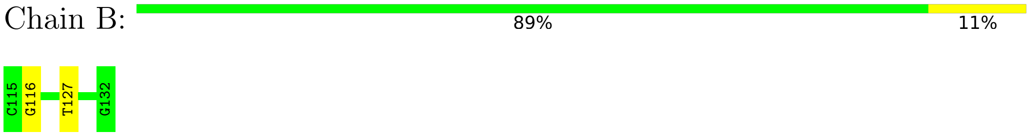
- Molecule 1: Transcriptional regulator SarA

Chain A:



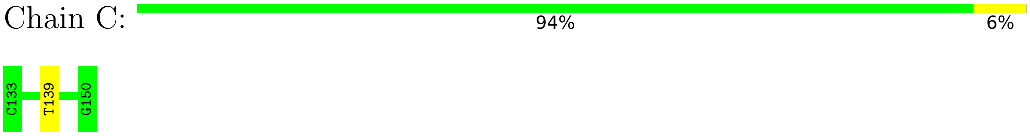
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

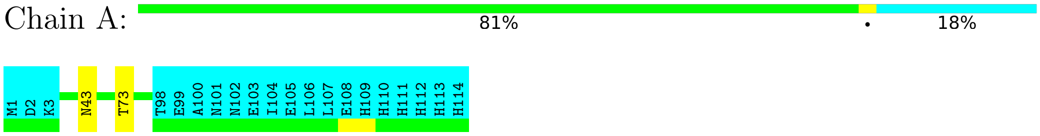
Chain C:



4.2.20 Score per residue for model 20

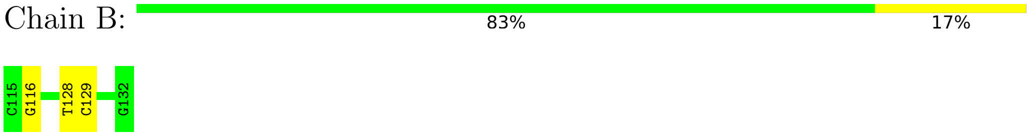
- Molecule 1: Transcriptional regulator SarA

Chain A:



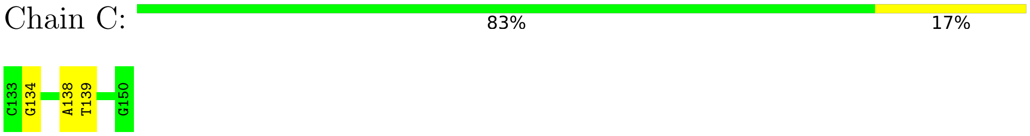
- Molecule 2: DNA

Chain B:



- Molecule 2: DNA

Chain C:



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	structure calculation	
DYANA	structure calculation	
Amber	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	1615
Number of shifts mapped to atoms	1587
Number of unparsed shifts	0
Number of shifts with mapping errors	28
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	66%

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	0.86±0.01	0±0/811 (0.0± 0.0%)	1.40±0.02	1±1/1086 (0.1± 0.1%)
2	B	0.77±0.00	0±0/410 (0.0± 0.0%)	1.24±0.01	0±0/631 (0.0± 0.0%)
2	C	0.76±0.00	0±0/410 (0.0± 0.0%)	1.24±0.01	0±0/631 (0.0± 0.0%)
All	All	0.81	0/32620 (0.0%)	1.31	13/46960 (0.0%)

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	0.6±0.7
2	B	0.0±0.0	1.6±0.6
2	C	0.0±0.0	0.6±1.0
All	All	0	54

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	96	ARG	NE-CZ-NH2	6.03	124.63	119.20	5	1
1	A	66	ARG	NE-CZ-NH2	5.68	124.31	119.20	5	2
1	A	74	VAL	N-CA-C	5.59	115.99	108.17	2	7
1	A	83	ARG	NE-CZ-NH2	5.18	123.87	119.20	19	2
1	A	48	GLN	OE1-CD-NE2	-5.13	117.47	122.60	16	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	B	127	DT	Sidechain	16
1	A	72	ARG	Sidechain	7
2	B	116	DG	Sidechain	7
2	B	129	DC	Sidechain	4
2	C	138	DA	Sidechain	4
1	A	66	ARG	Sidechain	3
2	B	128	DT	Sidechain	3
2	C	134	DG	Sidechain	3
2	C	139	DT	Sidechain	3
2	B	118	DG	Sidechain	1
2	C	143	DT	Sidechain	1
1	A	83	ARG	Sidechain	1
2	C	140	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	799	841	839	0±0
2	C	366	205	205	0±0
All	All	30620	25000	24980	1

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:36:LYS:HE3	2:C:139:DT:H3’	0.41	1.92	1	1

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	94/114 (82%)	92±1 (98±1%)	1±1 (1±1%)	1±0 (1±0%)	15	64
All	All	1880/2280 (82%)	1838 (98%)	22 (1%)	20 (1%)	15	64

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	43	ASN	20

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	91/110 (83%)	88±1 (97±2%)	3±1 (3±2%)	39	88
All	All	1820/2200 (83%)	1767 (97%)	53 (3%)	39	88

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	73	THR	20
1	A	70	ASP	7
1	A	64	LYS	4
1	A	14	ILE	4
1	A	72	ARG	3
1	A	78	VAL	2
1	A	49	VAL	2
1	A	74	VAL	2
1	A	36	LYS	2
1	A	77	LEU	1
1	A	62	PHE	1
1	A	65	LYS	1
1	A	45	LYS	1
1	A	39	ILE	1
1	A	83	ARG	1
1	A	93	VAL	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

There are no oligosaccharides in this entry.

6.6 Ligand geometry

There are no ligands in this entry.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 66% for the well-defined parts and 65% 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	1615
Number of shifts mapped to atoms	1587
Number of unparsed shifts	0
Number of shifts with mapping errors	28
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

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

- No matching atom found in the structure. All 28 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	B	121	DT	H4'1	4.236	.	.
1	B	123	DT	H4'1	4.377	.	.
1	B	125	DT	H4'1	4.244	.	.
1	B	127	DT	H4'1	4.302	.	.
1	B	128	DT	H4'1	4.163	.	.
1	C	139	DT	H4'1	4.236	.	.
1	C	141	DT	H4'1	4.377	.	.
1	C	143	DT	H4'1	4.244	.	.
1	C	145	DT	H4'1	4.302	.	.
1	C	146	DT	H4'1	4.163	.	.
1	B	133	DCZ	H1'	5.774	.	1
1	B	133	DCZ	H2'1	1.934	.	.
1	B	133	DCZ	H2'2	2.392	.	.
1	B	133	DCZ	H3'	4.683	.	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	B	133	DCZ	H4’	4.057	.	1
1	B	133	DCZ	H5	5.908	.	1
1	B	133	DCZ	H5’1	3.726	.	.
1	B	133	DCZ	H5’2	3.701	.	.
1	B	133	DCZ	H6	7.618	.	1
1	C	151	DCZ	H1’	5.774	.	1
1	C	151	DCZ	H2’1	1.934	.	.
1	C	151	DCZ	H2’2	2.384	.	.
1	C	151	DCZ	H3’	4.683	.	1
1	C	151	DCZ	H4’	4.057	.	1
1	C	151	DCZ	H5	5.908	.	1
1	C	151	DCZ	H5’1	3.727	.	.
1	C	151	DCZ	H5’2	3.701	.	.
1	C	151	DCZ	H6	7.618	.	1

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	107	-0.51 ± 0.07	Should be checked
$^{13}C_{\beta}$	106	-0.07 ± 0.07	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	104	0.31 ± 0.36	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 66%, i.e. 1415 atoms were assigned a chemical shift out of a possible 2134. 0 out of 17 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	364/468 (78%)	181/187 (97%)	92/188 (49%)	91/93 (98%)
Sidechain	707/857 (82%)	473/550 (86%)	222/266 (83%)	12/41 (29%)
Aromatic	66/101 (65%)	38/48 (79%)	28/49 (57%)	0/4 (0%)
Sugar	222/432 (51%)	222/252 (88%)	0/180 (0%)	0/0 (—%)
Base	56/276 (20%)	56/168 (33%)	0/64 (0%)	0/44 (0%)
Overall	1415/2134 (66%)	970/1205 (80%)	342/747 (46%)	103/182 (57%)

The following table shows the completeness of the chemical shift assignments for the full structure.

The overall completeness is 65%, i.e. 1576 atoms were assigned a chemical shift out of a possible 2418. 0 out of 19 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Backbone	419/568 (74%)	208/227 (92%)	107/228 (47%)	104/113 (92%)
Sidechain	812/993 (82%)	543/637 (85%)	255/312 (82%)	14/44 (32%)
Aromatic	67/149 (45%)	39/72 (54%)	28/61 (46%)	0/16 (0%)
Sugar	222/432 (51%)	222/252 (88%)	0/180 (0%)	0/0 (—%)
Base	56/276 (20%)	56/168 (33%)	0/64 (0%)	0/44 (0%)
Overall	1576/2418 (65%)	1068/1356 (79%)	390/845 (46%)	118/217 (54%)

7.1.4 Statistically unusual chemical shifts

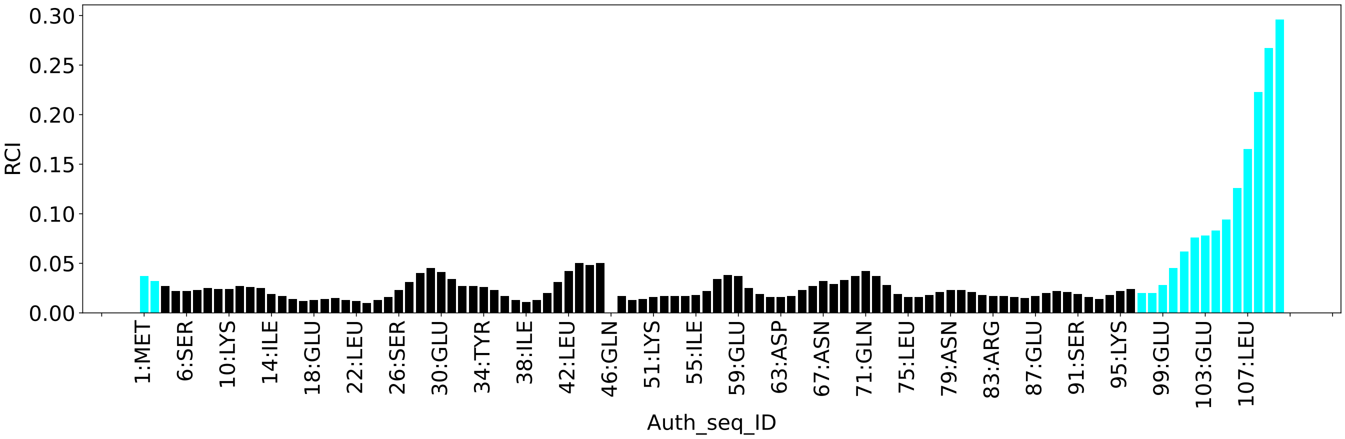
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	**Chain**	**Res**	**Type**	**Atom**	**Shift, *ppm***	**Expected range, *ppm***	**Z-score**
1	A	30	GLU	HB3	0.62	0.95 – 3.05	-6.5

7.1.5 Random Coil Index (RCI) plots

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:



DIST-REST INFOmissingINFO