WORLDWIDE wwPDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report ⓘ

May 15, 2023 – 08:11 PM JST

PDB ID	:	8HEP
BMRB ID	:	36515
Title	:	Solution structure of the periplasmic domain of the anti-sigma factor RsgI1 from Clostridium thermocellum
Authors	:	Chen, C.; Feng, Y.
Deposited on	:	2022-11-08

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
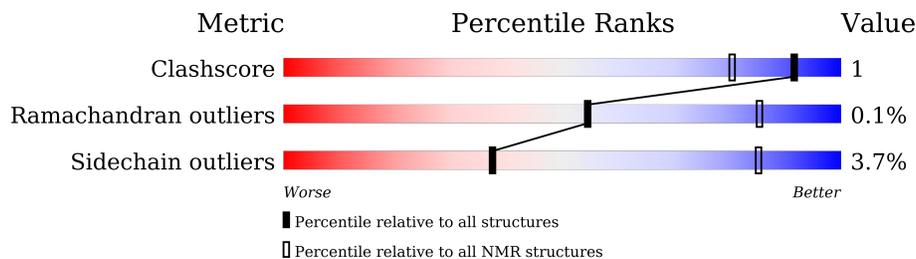http://www.wwpdb.org/validation/2017/FAQs#types.

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 93%.

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	Percentile Ranks	Value
Clashscore		1
Ramachandran outliers		0.1%
Sidechain outliers		3.7%

Worse   *Better*

▌ Percentile relative to all structures
▯ Percentile relative to all NMR structures

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	10	80% 10% 10%
2	B	166	89% · 10%

WORLDWIDE
wwPDB
PROTEIN DATA BANK

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:2-A:10, B:11-B:82, B:93-B:169 (158)	0.36	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 1 single-model cluster was found.

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

3 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 2846 atoms, of which 1429 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Anti-sigma factor.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	10	159	53	78	11	15	2	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP H6SHX8

- Molecule 2 is a protein called Anti-sigma factor.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	166	2687	827	1351	237	267	5	0

There are 6 discrepancies between the modelled and reference sequences:

| Chain	Residue	Modelled	Actual	Comment	Reference
B	171	HIS	-	expression tag	UNP H6SHX8
B	172	HIS	-	expression tag	UNP H6SHX8
B	173	HIS	-	expression tag	UNP H6SHX8
B	174	HIS	-	expression tag	UNP H6SHX8
B	175	HIS	-	expression tag	UNP H6SHX8
B	176	HIS	-	expression tag	UNP H6SHX8

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: Anti-sigma factor
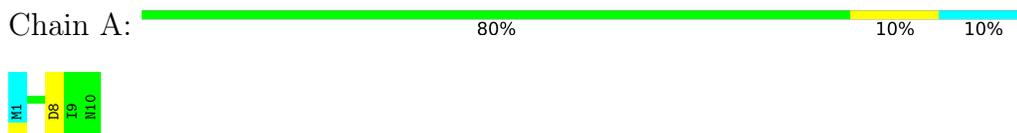
Chain A:

- Molecule 2: Anti-sigma factor

Chain B:
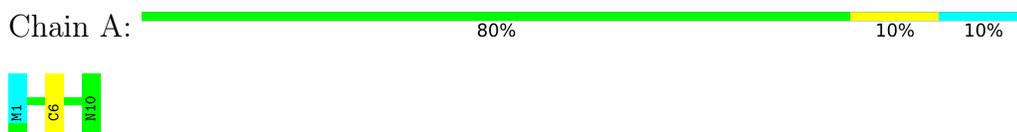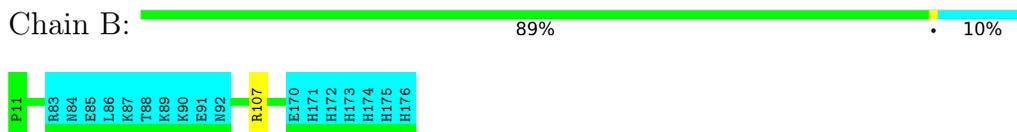
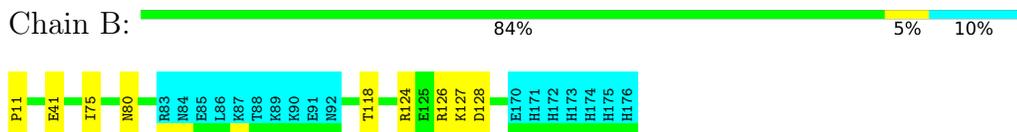## 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

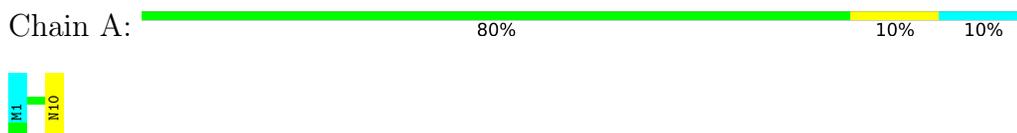### 4.2.1 Score per residue for model 1

- Molecule 1: Anti-sigma factor

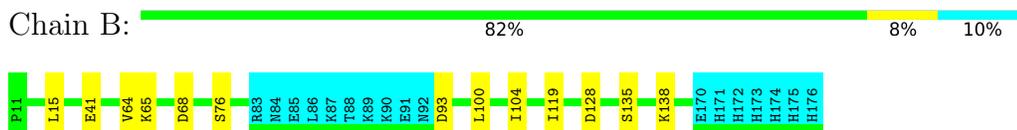Chain A:

- Molecule 2: Anti-sigma factor

Chain B:

4.2.2 Score per residue for model 2

• Molecule 1: Anti-sigma factor

Chain A:

• Molecule 2: Anti-sigma factor

Chain B:

4.2.3 Score per residue for model 3

• Molecule 1: Anti-sigma factor

Chain A:

• Molecule 2: Anti-sigma factor

Chain B:
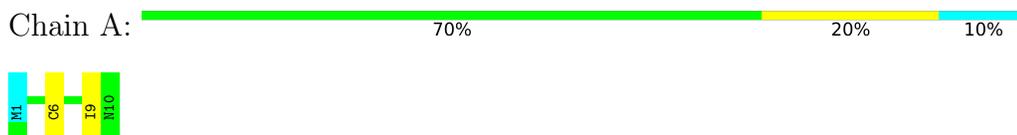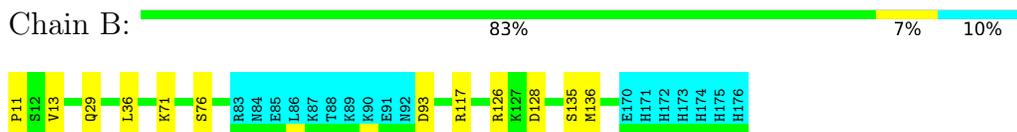
4.2.4 Score per residue for model 4

• Molecule 1: Anti-sigma factor

Chain A:

• Molecule 2: Anti-sigma factor

Chain B:
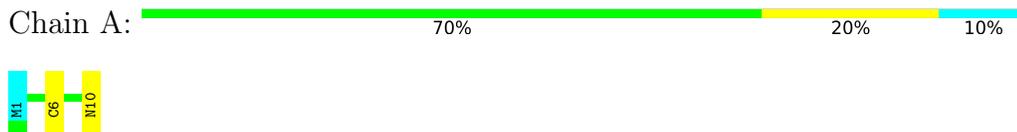
4.2.5 Score per residue for model 5

• Molecule 1: Anti-sigma factor

Chain A:

• Molecule 2: Anti-sigma factor

Chain B:
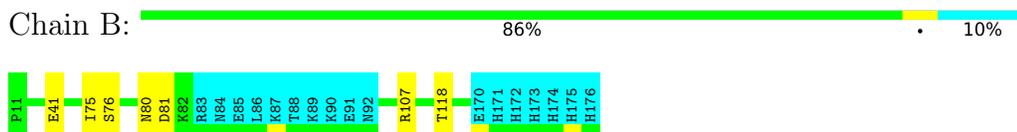
4.2.6 Score per residue for model 6

• Molecule 1: Anti-sigma factor

Chain A:

• Molecule 2: Anti-sigma factor

Chain B:
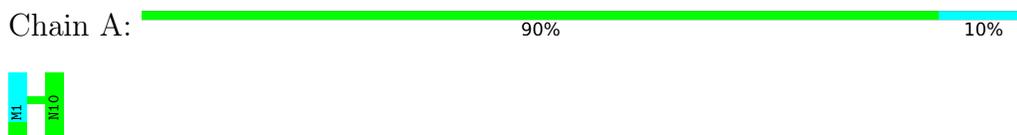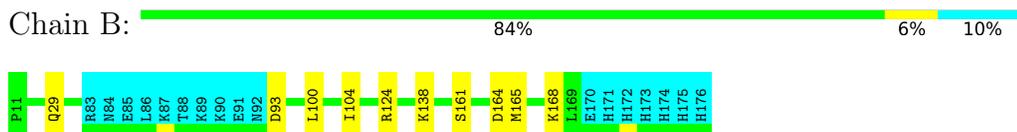
4.2.7 Score per residue for model 7
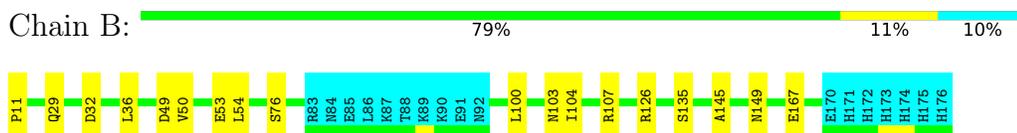
• Molecule 1: Anti-sigma factor

Chain A:

• Molecule 2: Anti-sigma factor

Chain B:
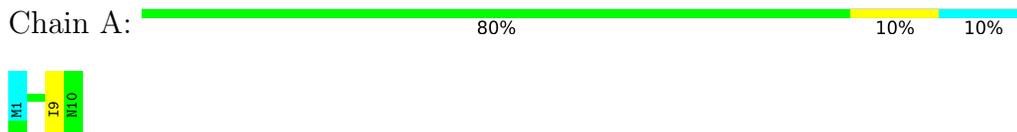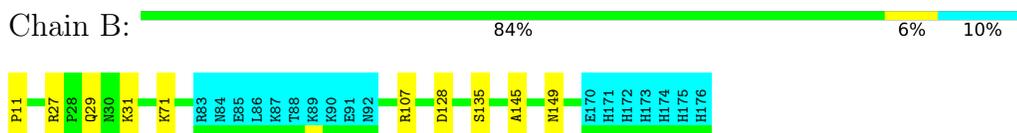
4.2.8 Score per residue for model 8 (medoid)

• Molecule 1: Anti-sigma factor

Chain A:

• Molecule 2: Anti-sigma factor

Chain B:
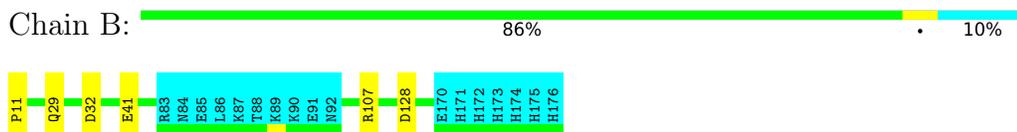
4.2.9 Score per residue for model 9
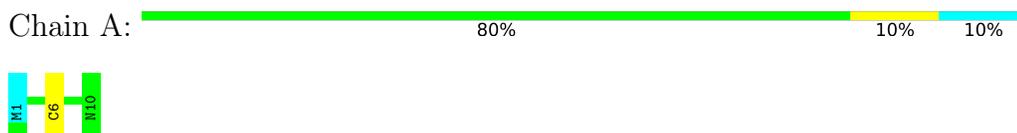
• Molecule 1: Anti-sigma factor

Chain A:

• Molecule 2: Anti-sigma factor

Chain B:
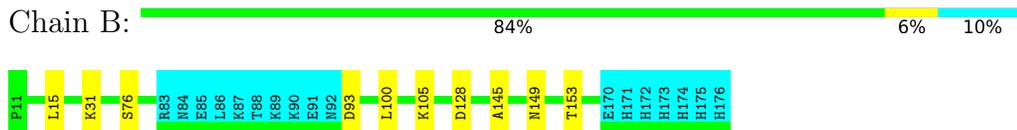
4.2.10 Score per residue for model 10

• Molecule 1: Anti-sigma factor

Chain A:

• Molecule 2: Anti-sigma factor

Chain B:
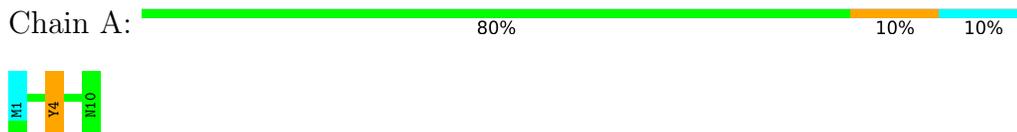
4.2.11   Score per residue for model 11

• Molecule 1: Anti-sigma factor

Chain A:

• Molecule 2: Anti-sigma factor

Chain B:
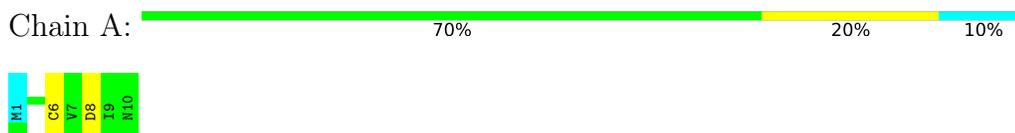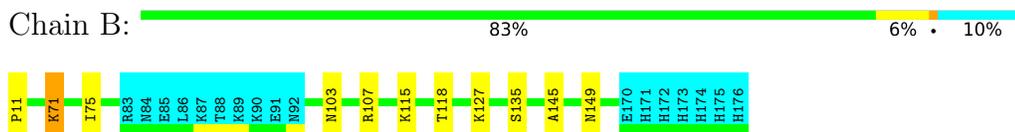
4.2.12   Score per residue for model 12

• Molecule 1: Anti-sigma factor

Chain A:

• Molecule 2: Anti-sigma factor

Chain B:
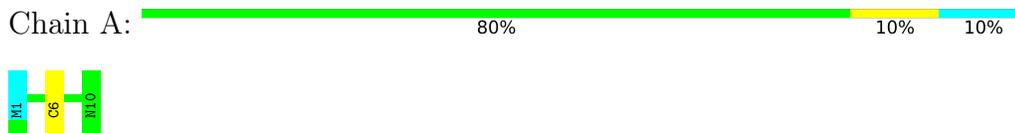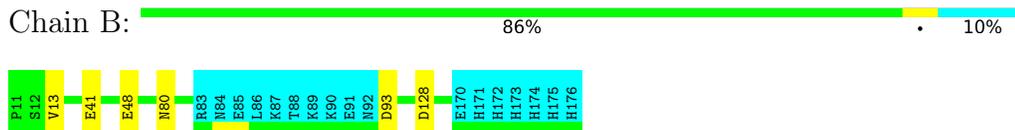
4.2.13   Score per residue for model 13

• Molecule 1: Anti-sigma factor

Chain A:

• Molecule 2: Anti-sigma factor

Chain B:

4.2.14 Score per residue for model 14

• Molecule 1: Anti-sigma factor

Chain A:

• Molecule 2: Anti-sigma factor

Chain B:
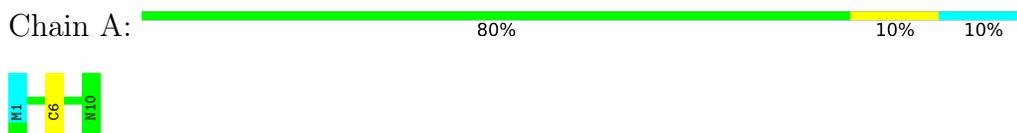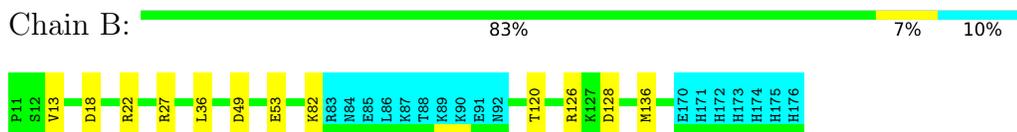
4.2.15 Score per residue for model 15

• Molecule 1: Anti-sigma factor

Chain A:

• Molecule 2: Anti-sigma factor

Chain B:
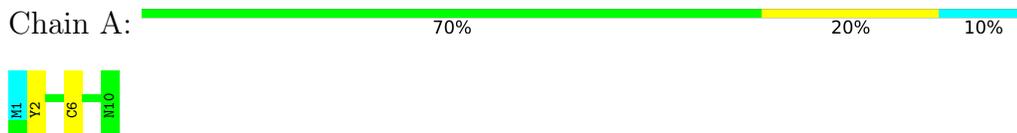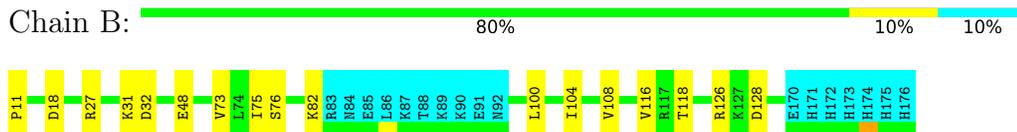
4.2.16 Score per residue for model 16

• Molecule 1: Anti-sigma factor

Chain A:

• Molecule 2: Anti-sigma factor

Chain B:

4.2.17 Score per residue for model 17
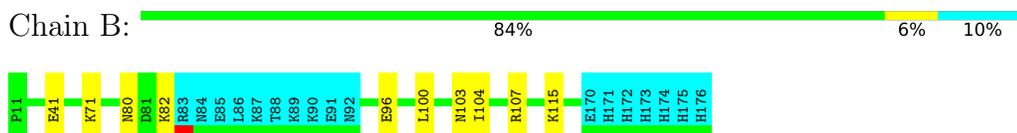
• Molecule 1: Anti-sigma factor

Chain A:
80% 10% 10%

M1 V7 N10

• Molecule 2: Anti-sigma factor

Chain B:
86% . . 10%

P11 S12 R83 N84 E85 L86 K87 T88 K89 K90 E91 N92 L100 I104 R107 R124 K127 K131 E170 H171 H172 H173 H174 H175 H176
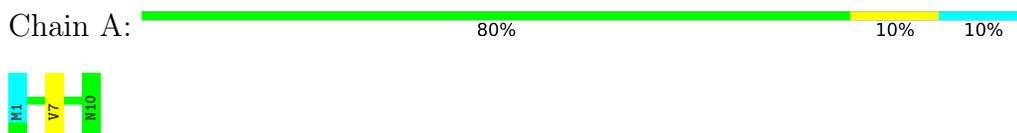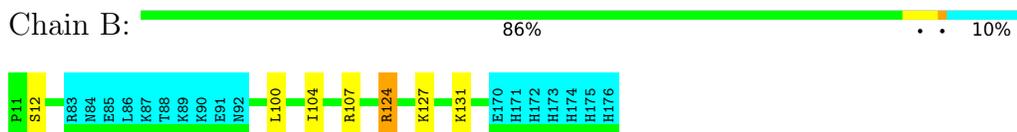
4.2.18 Score per residue for model 18

• Molecule 1: Anti-sigma factor

Chain A:
80% 10% 10%

M1 N10

• Molecule 2: Anti-sigma factor

Chain B:
84% 5% . 10%

P11 D18 L36 N46 E53 R57 A66 K71 R83 N84 E85 L86 K87 T88 K89 K90 E91 N92 E99 R126 M136 E170 H171 H172 H173 H174 H175 H176
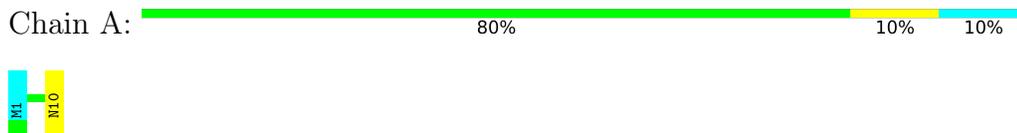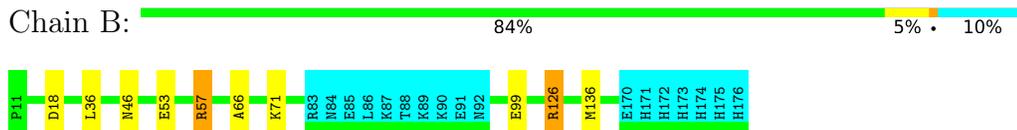
4.2.19 Score per residue for model 19

• Molecule 1: Anti-sigma factor

Chain A:
70% 20% 10%

M1 C6 V7 N10

• Molecule 2: Anti-sigma factor

Chain B:
80% 10% 10%

P11 S12 V13 R27 D32 L36 E48 K65 K71 D81 K82 R83 N84 E85 L86 K87 T88 K89 K90 E91 N92 E96 L100 R107 V108 V116 D128 E170 H171 H172 H173 H174 H175 H176
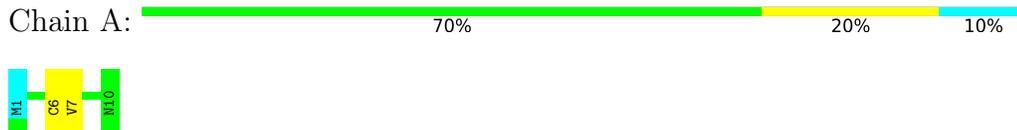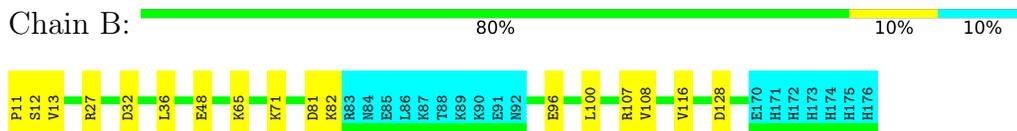
4.2.20 Score per residue for model 20

• Molecule 1: Anti-sigma factor

Chain A:

70%   20%   10%

M1 Y2 C6 N10

• Molecule 2: Anti-sigma factor

Chain B:

87%   •   10%

P11 E25 L54 K65 S76 R83 N84 E85 L86 K87 T88 K89 K90 E91 N92 E170 H171 H172 H173 H174 H175 H176
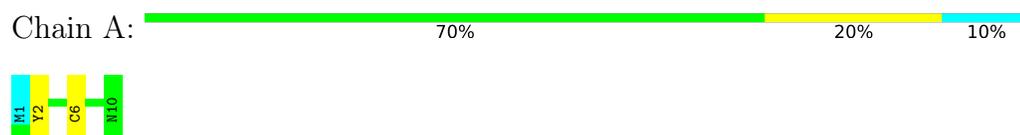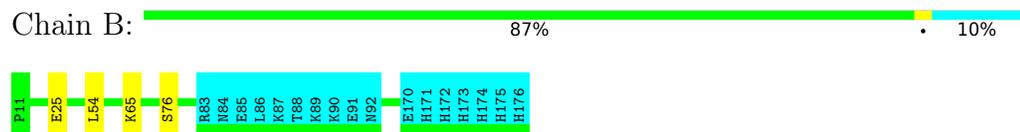
5 Refinement protocol and experimental data overview ⓘ

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
CNS	refinement	
CNS	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	2223
Number of shifts mapped to atoms	2223
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	93%

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	1.16±0.10	0±0/74 (0.0± 0.0%)	0.84±0.06	0±0/100 (0.0± 0.0%)
2	B	0.81±0.02	0±0/1188 (0.0± 0.0%)	0.76±0.02	0±0/1600 (0.0± 0.0%)
All	All	0.83	0/25240 (0.0%)	0.77	1/34000 (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
2	B	0.0±0.0	0.2±0.5
All	All	0	5

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	B	27	ARG	NE-CZ-NH2	-5.08	117.76	120.30	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	B	107	ARG	Sidechain	2
2	B	27	ARG	Sidechain	1
2	B	124	ARG	Sidechain	1
2	B	57	ARG	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	73	67	67	1±1
2	B	1179	1203	1202	4±2
All	All	25040	25400	25380	75

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 1.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
2:B:108:VAL:HG23	2:B:116:VAL:HG21	0.53	1.79	15	2
1:A:6:CYS:SG	2:B:76:SER:HB3	0.51	2.46	4	4
2:B:11:PRO:HA	2:B:32:ASP:OD2	0.50	2.07	19	1
2:B:71:LYS:NZ	2:B:71:LYS:HB2	0.50	2.22	11	1
2:B:145:ALA:O	2:B:149:ASN:HB3	0.49	2.07	4	4
2:B:126:ARG:HG2	2:B:136:MET:SD	0.48	2.48	3	2
2:B:53:GLU:O	2:B:57:ARG:HD3	0.48	2.08	18	1
2:B:65:LYS:HB2	2:B:68:ASP:HB2	0.47	1.85	2	1
2:B:157:VAL:HG13	2:B:165:MET:SD	0.47	2.49	13	1
1:A:6:CYS:HA	2:B:13:VAL:O	0.46	2.10	3	4
1:A:6:CYS:SG	2:B:29:GLN:HG3	0.46	2.51	8	1
2:B:49:ASP:O	2:B:53:GLU:HG2	0.46	2.11	14	2
1:A:7:VAL:O	2:B:12:SER:HA	0.46	2.11	13	3
1:A:8:ASP:OD1	2:B:162:ILE:HB	0.46	2.11	13	1
2:B:164:ASP:O	2:B:168:LYS:HG2	0.46	2.11	6	1
1:A:9:ILE:O	2:B:11:PRO:HD2	0.45	2.12	3	2
2:B:11:PRO:HA	2:B:32:ASP:OD1	0.45	2.10	4	2
2:B:18:ASP:OD2	2:B:22:ARG:HB3	0.45	2.11	14	1
1:A:8:ASP:HA	2:B:11:PRO:O	0.45	2.11	8	3
1:A:4:TYR:CZ	2:B:80:ASN:HB2	0.45	2.47	10	1
2:B:29:GLN:OE1	2:B:135:SER:HA	0.44	2.11	4	3
2:B:49:ASP:HA	2:B:107:ARG:NH1	0.44	2.28	13	1
1:A:6:CYS:SG	2:B:76:SER:HB2	0.44	2.52	15	2
2:B:100:LEU:O	2:B:104:ILE:HG12	0.43	2.12	10	6
2:B:126:ARG:HG3	2:B:136:MET:SD	0.43	2.53	14	1
2:B:75:ILE:O	2:B:118:THR:HA	0.43	2.14	15	4

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
1:A:6:CYS:SG	2:B:136:MET:HG3	0.42	2.54	16	1
2:B:11:PRO:HA	2:B:32:ASP:CG	0.42	2.34	13	3
1:A:6:CYS:SG	2:B:135:SER:HA	0.42	2.55	11	1
2:B:103:ASN:O	2:B:107:ARG:HG2	0.42	2.14	10	2
2:B:76:SER:HA	2:B:119:ILE:O	0.42	2.15	2	1
2:B:73:VAL:O	2:B:116:VAL:HA	0.42	2.13	15	1
2:B:50:VAL:O	2:B:54:LEU:HD22	0.41	2.15	4	1
2:B:127:LYS:O	2:B:131:LYS:HG3	0.41	2.15	17	1
2:B:66:ALA:HA	2:B:71:LYS:HD2	0.41	1.92	18	1
1:A:10:ASN:HB3	2:B:64:VAL:HG11	0.41	1.91	2	1
1:A:4:TYR:CE1	2:B:16:VAL:HG22	0.41	2.50	16	1
2:B:135:SER:OG	2:B:138:LYS:HB2	0.41	2.16	2	1
2:B:17:ILE:HG21	2:B:47:VAL:HB	0.41	1.91	13	1
2:B:138:LYS:HG3	2:B:165:MET:SD	0.41	2.55	6	1
2:B:141:LEU:HB2	2:B:165:MET:SD	0.41	2.55	16	1
2:B:48:GLU:HB3	2:B:107:ARG:HG3	0.40	1.93	19	1
2:B:65:LYS:O	2:B:71:LYS:HD2	0.40	2.16	19	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	8/10 (80%)	7±0 (88±6%)	1±1 (11±7%)	0±0 (2±4%)	11	53
2	B	148/166 (89%)	145±1 (98±1%)	3±1 (2±1%)	0±0 (0±0%)	100	100
All	All	3120/3520 (89%)	3048 (98%)	69 (2%)	3 (0%)	54	85

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	2	TYR	3

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	8/9 (89%)	8±1 (95±7%)	0±1 (5±7%)	28	77
2	B	135/152 (89%)	130±2 (96±1%)	5±2 (4±1%)	38	86
All	All	2860/3220 (89%)	2754 (96%)	106 (4%)	37	85

All 38 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)
2	B	128	ASP	10
2	B	41	GLU	7
2	B	36	LEU	6
1	A	10	ASN	6
2	B	93	ASP	5
2	B	107	ARG	5
2	B	82	LYS	5
2	B	126	ARG	4
2	B	71	LYS	4
2	B	31	LYS	4
2	B	124	ARG	3
2	B	15	LEU	3
2	B	103	ASN	3
2	B	80	ASN	3
2	B	27	ARG	3
2	B	115	LYS	3
2	B	48	GLU	3
2	B	127	LYS	2
2	B	81	ASP	2
2	B	100	LEU	2
2	B	96	GLU	2
2	B	25	GLU	2
2	B	120	THR	2
2	B	18	ASP	2
2	B	65	LYS	2
2	B	117	ARG	1
2	B	167	GLU	1
2	B	29	GLN	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
2	B	161	SER	1
1	A	9	ILE	1
2	B	105	LYS	1
2	B	153	THR	1
1	A	4	TYR	1
2	B	35	GLN	1
2	B	135	SER	1
2	B	46	ASN	1
2	B	99	GLU	1
2	B	54	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 ⓘ

There are no monosaccharides 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 93% for the well-defined parts and 90% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_0*

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	2223
Number of shifts mapped to atoms	2223
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	173	-0.41 ± 0.07	None needed (< 0.5 ppm)
$^{13}C_\beta$	165	0.16 ± 0.05	None needed (< 0.5 ppm)
$^{13}C'$	172	-0.16 ± 0.08	None needed (< 0.5 ppm)
^{15}N	168	0.21 ± 0.28	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 93%, i.e. 2044 atoms were assigned a chemical shift out of a possible 2193. 0 out of 31 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	789/794 (99%)	320/322 (99%)	315/316 (100%)	154/156 (99%)
Sidechain	1187/1317 (90%)	807/852 (95%)	361/415 (87%)	19/50 (38%)

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Aromatic	68/82 (83%)	34/38 (89%)	34/42 (81%)	0/2 (0%)
Overall	2044/2193 (93%)	1161/1212 (96%)	710/773 (92%)	173/208 (83%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 90%, i.e. 2223 atoms were assigned a chemical shift out of a possible 2470. 0 out of 32 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Backbone	862/884 (98%)	349/358 (97%)	345/352 (98%)	168/174 (97%)
Sidechain	1293/1456 (89%)	878/939 (94%)	394/459 (86%)	21/58 (36%)
Aromatic	68/130 (52%)	34/62 (55%)	34/54 (63%)	0/14 (0%)
Overall	2223/2470 (90%)	1261/1359 (93%)	773/865 (89%)	189/246 (77%)

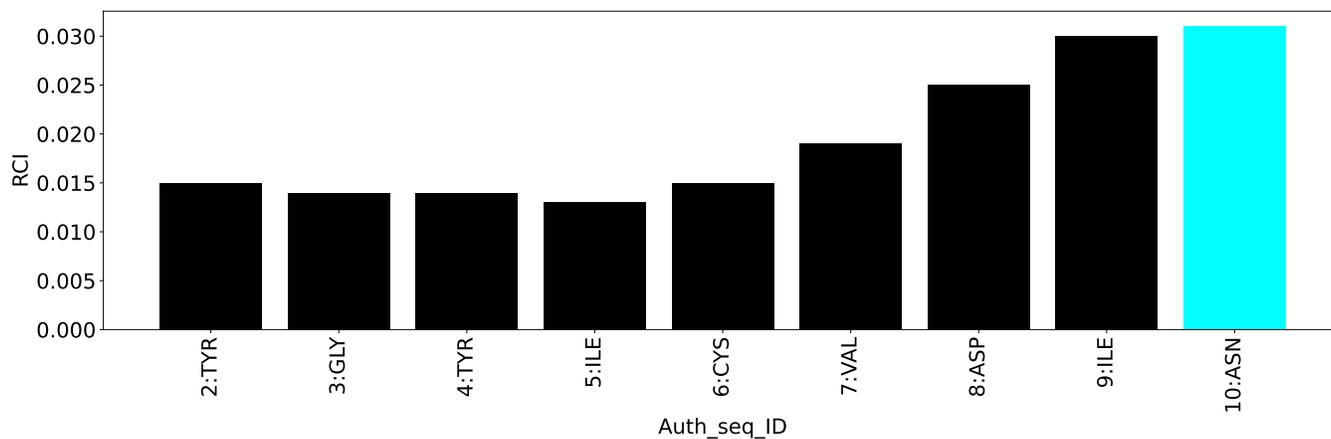### 7.1.4 Statistically unusual chemical shifts ⓘ

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	B	77	GLY	HA3	5.74	2.08 – 5.71	5.1

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:

Random coil index (RCI) for chain B: