



Full wwPDB NMR Structure Validation Report ⓘ

May 29, 2020 – 08:00 am BST

PDB ID : 5O7J
Title : Structural insights into the periplasmic sensor domain of the GacS histidine kinase controlling biofilm formation in *Pseudomonas aeruginosa*
Authors : Ali-Ahmad, A.; Bornet, O.; Fadel, F.; Bourne, Y.; Vincent, F.; Guerlesquin, F.; Sebban-Kreuzer, C.
Deposited on : 2017-06-09

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

Cyrange : Kirchner and Güntert (2011)
NmrClust : Kelley et al. (1996)
MolProbity : 4.02b-467
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI : v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV : Wang et al. (2010)
ShiftChecker : 2.11
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.11

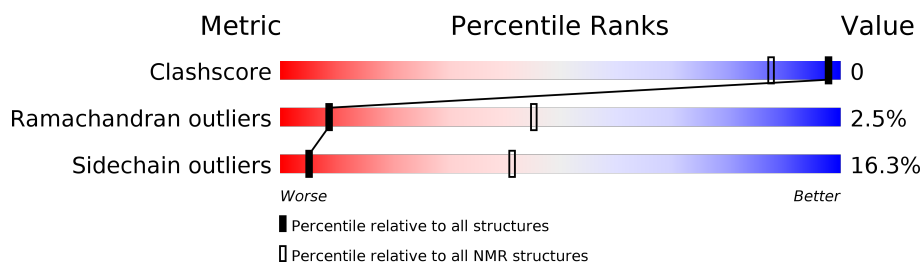
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 87%.

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	A	150	 81% 13% 6%

2 Ensemble composition and analysis

This entry contains 20 models. Model 4 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:24-A:164 (141)	0.81	4

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 2 clusters. No single-model clusters were found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 2271 atoms, of which 1121 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Histidine kinase.

Mol	Chain	Residues	Atoms					Trace	
			Total	C	H	N	O		S
1	A	150	2271	711	1121	217	218	4	0

There are 23 discrepancies between the modelled and reference sequences:

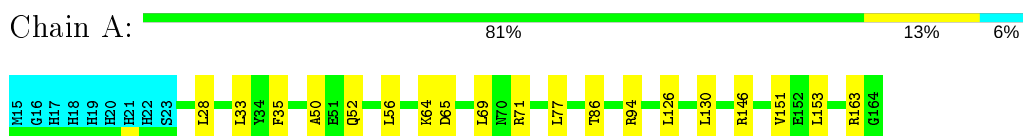
Chain	Residue	Modelled	Actual	Comment	Reference
A	15	MET	-	initiating methionine	UNP A0A0A8RMX6
A	16	GLY	-	expression tag	UNP A0A0A8RMX6
A	17	HIS	-	expression tag	UNP A0A0A8RMX6
A	18	HIS	-	expression tag	UNP A0A0A8RMX6
A	19	HIS	-	expression tag	UNP A0A0A8RMX6
A	20	HIS	-	expression tag	UNP A0A0A8RMX6
A	21	HIS	-	expression tag	UNP A0A0A8RMX6
A	22	HIS	-	expression tag	UNP A0A0A8RMX6
A	23	SER	-	expression tag	UNP A0A0A8RMX6
A	24	SER	-	expression tag	UNP A0A0A8RMX6
A	25	GLY	-	expression tag	UNP A0A0A8RMX6
A	26	VAL	-	expression tag	UNP A0A0A8RMX6
A	27	ASP	-	expression tag	UNP A0A0A8RMX6
A	28	LEU	-	expression tag	UNP A0A0A8RMX6
A	29	GLY	-	expression tag	UNP A0A0A8RMX6
A	30	THR	-	expression tag	UNP A0A0A8RMX6
A	31	GLU	-	expression tag	UNP A0A0A8RMX6
A	32	ASN	-	expression tag	UNP A0A0A8RMX6
A	33	LEU	-	expression tag	UNP A0A0A8RMX6
A	34	TYR	-	expression tag	UNP A0A0A8RMX6
A	35	PHE	-	expression tag	UNP A0A0A8RMX6
A	36	GLN	-	expression tag	UNP A0A0A8RMX6
A	37	SER	-	expression tag	UNP A0A0A8RMX6

4 Residue-property plots [i](#)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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: Histidine kinase

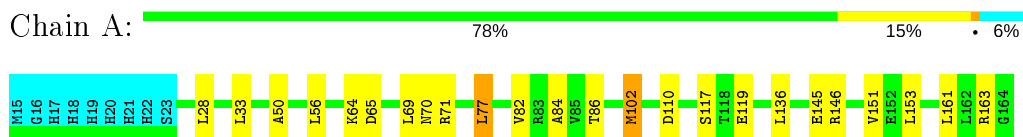


4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

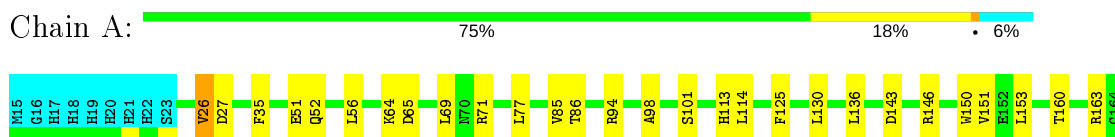
4.2.1 Score per residue for model 1

- Molecule 1: Histidine kinase



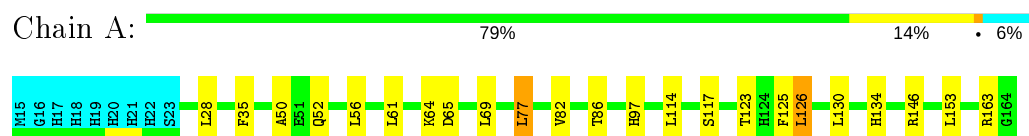
4.2.2 Score per residue for model 2

- Molecule 1: Histidine kinase



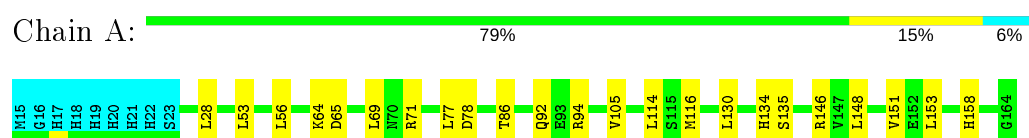
4.2.3 Score per residue for model 3

- Molecule 1: Histidine kinase



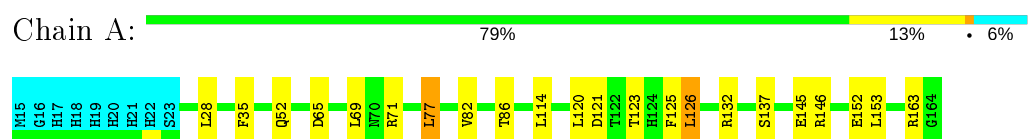
4.2.4 Score per residue for model 4 (medoid)

- Molecule 1: Histidine kinase



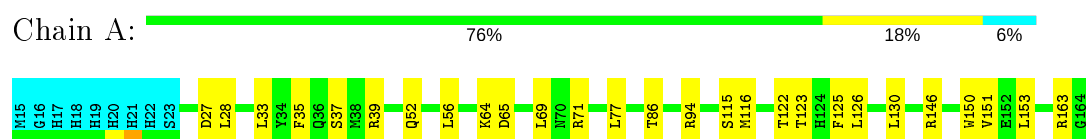
4.2.5 Score per residue for model 5

- Molecule 1: Histidine kinase



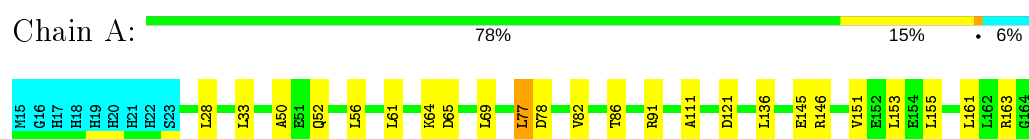
4.2.6 Score per residue for model 6

- Molecule 1: Histidine kinase



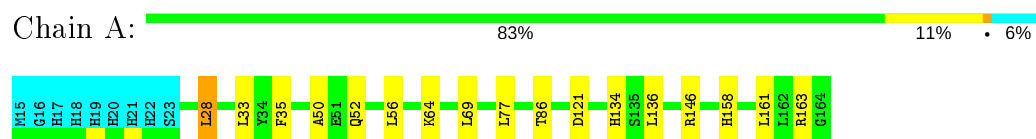
4.2.7 Score per residue for model 7

- Molecule 1: Histidine kinase



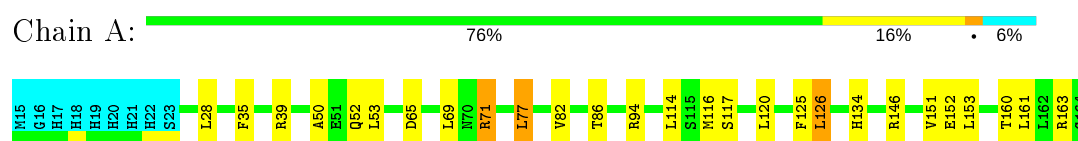
4.2.8 Score per residue for model 8

- Molecule 1: Histidine kinase



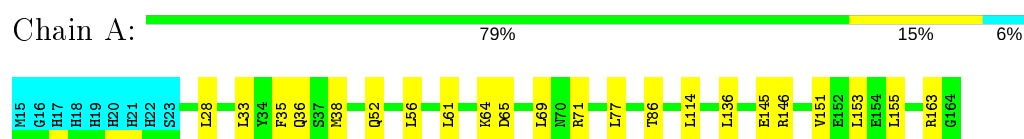
4.2.9 Score per residue for model 9

- Molecule 1: Histidine kinase



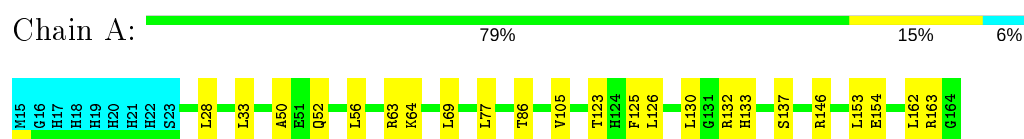
4.2.10 Score per residue for model 10

- Molecule 1: Histidine kinase



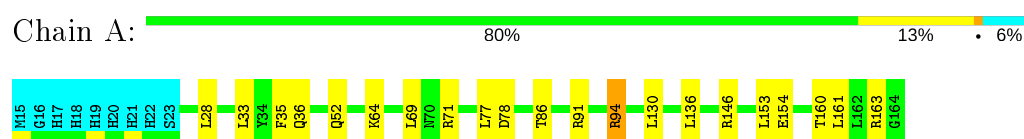
4.2.11 Score per residue for model 11

- Molecule 1: Histidine kinase



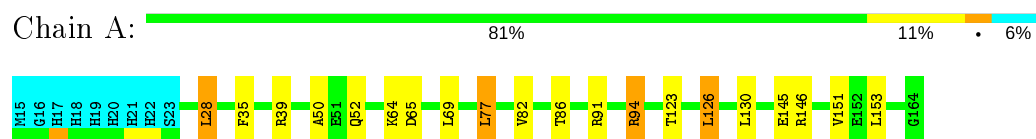
4.2.12 Score per residue for model 12

- Molecule 1: Histidine kinase



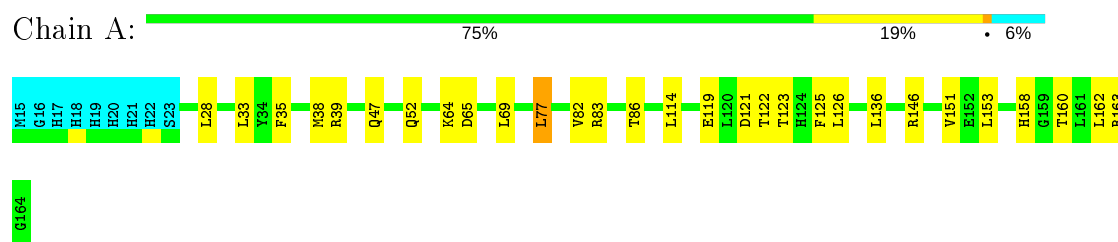
4.2.13 Score per residue for model 13

- Molecule 1: Histidine kinase



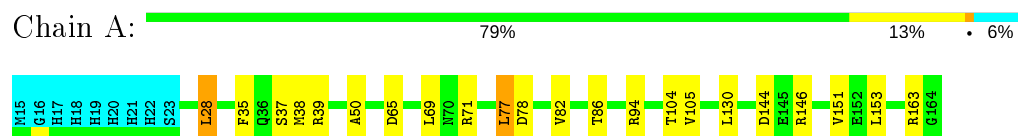
4.2.14 Score per residue for model 14

- Molecule 1: Histidine kinase



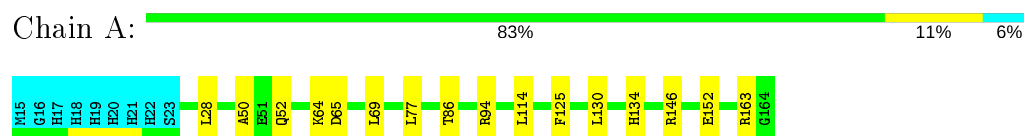
4.2.15 Score per residue for model 15

- Molecule 1: Histidine kinase



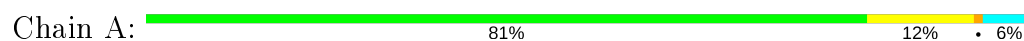
4.2.16 Score per residue for model 16

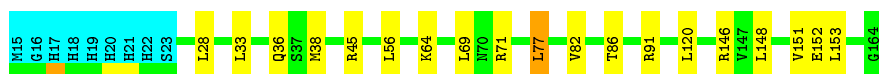
- Molecule 1: Histidine kinase



4.2.17 Score per residue for model 17

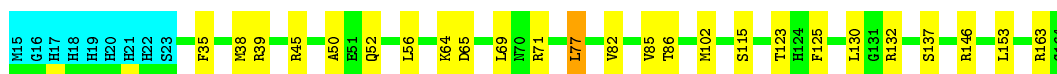
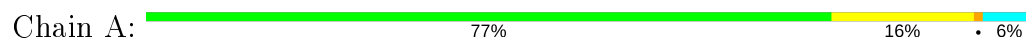
- Molecule 1: Histidine kinase





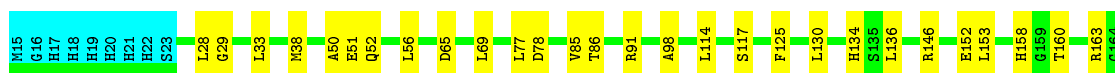
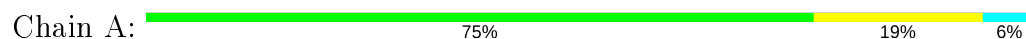
4.2.18 Score per residue for model 18

- Molecule 1: Histidine kinase



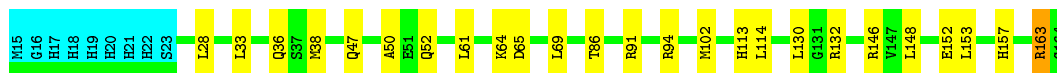
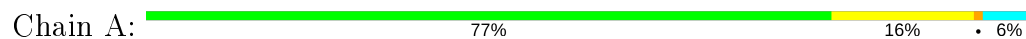
4.2.19 Score per residue for model 19

- Molecule 1: Histidine kinase



4.2.20 Score per residue for model 20

- Molecule 1: Histidine kinase



5 Refinement protocol and experimental data overview (i)

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

Of the 20 calculated structures, 20 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
CYANA	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 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1684
Number of shifts mapped to atoms	1684
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	87%

No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in each chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	1072	1062	1062	1±1
All	All	21440	21240	21240	15

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:130:LEU:HD13	1:A:150:TRP:CZ2	0.48	2.42	6	2
1:A:77:LEU:HA	1:A:82:VAL:HG21	0.45	1.87	7	10
1:A:84:ALA:HB1	1:A:102:MET:SD	0.43	2.53	1	1
1:A:85:VAL:HG12	1:A:98:ALA:HB3	0.41	1.93	19	2

5.2 Torsion angles [i](#)

5.2.1 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	140/150 (93%)	114±3 (81±2%)	23±3 (16±2%)	4±1 (2±1%)	9	45
All	All	2800/3000 (93%)	2275 (81%)	455 (16%)	70 (2%)	9	45

All 22 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	146	ARG	20
1	A	50	ALA	12
1	A	117	SER	4
1	A	126	LEU	4
1	A	105	VAL	3
1	A	137	SER	3
1	A	158	HIS	3
1	A	125	PHE	3
1	A	28	LEU	3
1	A	115	SER	2
1	A	134	HIS	2
1	A	27	ASP	1
1	A	119	GLU	1
1	A	163	ARG	1
1	A	161	LEU	1
1	A	160	THR	1
1	A	136	LEU	1
1	A	29	GLY	1
1	A	51	GLU	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	132	ARG	1
1	A	26	VAL	1
1	A	111	ALA	1

5.2.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the sidechain conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	113/121 (93%)	95±2 (84±2%)	18±2 (16±2%)	5	41
All	All	2260/2420 (93%)	1891 (84%)	369 (16%)	5	41

All 64 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	69	LEU	20
1	A	86	THR	20
1	A	77	LEU	19
1	A	153	LEU	18
1	A	28	LEU	18
1	A	52	GLN	16
1	A	65	ASP	16
1	A	64	LYS	16
1	A	163	ARG	13
1	A	35	PHE	12
1	A	56	LEU	12
1	A	71	ARG	11
1	A	151	VAL	11
1	A	33	LEU	11
1	A	130	LEU	10
1	A	114	LEU	10
1	A	123	THR	7
1	A	136	LEU	7
1	A	126	LEU	7
1	A	38	MET	7
1	A	125	PHE	7
1	A	152	GLU	6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	94	ARG	6
1	A	91	ARG	6
1	A	78	ASP	5
1	A	145	GLU	5
1	A	121	ASP	4
1	A	161	LEU	4
1	A	160	THR	4
1	A	36	GLN	4
1	A	134	HIS	4
1	A	61	LEU	4
1	A	102	MET	3
1	A	148	LEU	3
1	A	120	LEU	3
1	A	116	MET	3
1	A	122	THR	2
1	A	155	LEU	2
1	A	154	GLU	2
1	A	113	HIS	2
1	A	53	LEU	2
1	A	47	GLN	2
1	A	39	ARG	2
1	A	37	SER	2
1	A	162	LEU	2
1	A	83	ARG	1
1	A	27	ASP	1
1	A	143	ASP	1
1	A	97	HIS	1
1	A	144	ASP	1
1	A	133	HIS	1
1	A	70	ASN	1
1	A	92	GLN	1
1	A	110	ASP	1
1	A	158	HIS	1
1	A	135	SER	1
1	A	45	ARG	1
1	A	119	GLU	1
1	A	51	GLU	1
1	A	104	THR	1
1	A	26	VAL	1
1	A	157	HIS	1
1	A	85	VAL	1
1	A	101	SER	1

5.2.3 RNA [i](#)

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.4 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.5 Ligand geometry [i](#)

There are no ligands in this entry.

5.6 Other polymers [i](#)

There are no such molecules in this entry.

5.7 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Chemical shift validation [i](#)

The completeness of assignment taking into account all chemical shift lists is 87% for the well-defined parts and 82% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

6.1.1 Bookkeeping [i](#)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	1684
Number of shifts mapped to atoms	1684
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

6.1.2 Chemical shift referencing [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	139	-0.21 \pm 0.17	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	127	0.17 \pm 0.12	None needed (< 0.5 ppm)
$^{13}\text{C}'$	135	-0.18 \pm 0.14	None needed (< 0.5 ppm)
^{15}N	131	0.21 \pm 0.24	None needed (< 0.5 ppm)

6.1.3 Completeness of resonance assignments [i](#)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 87%, i.e. 1443 atoms were assigned a chemical shift out of a possible 1668. 24 out of 31 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	673/693 (97%)	273/276 (99%)	271/282 (96%)	129/135 (96%)
Sidechain	727/879 (83%)	448/509 (88%)	263/330 (80%)	16/40 (40%)

Continued on next page...

Continued from previous page...

	Total	¹ H	¹³ C	¹⁵ N
Aromatic	43/96 (45%)	23/53 (43%)	19/35 (54%)	1/8 (12%)
Overall	1443/1668 (87%)	744/838 (89%)	553/647 (85%)	146/183 (80%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 82%, i.e. 1455 atoms were assigned a chemical shift out of a possible 1784. 24 out of 31 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	¹ H	¹³ C	¹⁵ N
Backbone	682/738 (92%)	277/294 (94%)	274/300 (91%)	131/144 (91%)
Sidechain	730/908 (80%)	450/528 (85%)	264/340 (78%)	16/40 (40%)
Aromatic	43/138 (31%)	23/77 (30%)	19/47 (40%)	1/14 (7%)
Overall	1455/1784 (82%)	750/899 (83%)	557/687 (81%)	148/198 (75%)

6.1.4 Statistically unusual chemical shifts [i](#)

There are no statistically unusual chemical shifts.

6.1.5 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.

Random coil index (RCI) for chain A:

