WORLDWIDE wwPDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report ⓘ

Apr 3, 2024 – 01:25 am BST

PDB ID	:	8ONG
BMRB ID	:	18045
Title	:	Structure of the endothelial monocyte activating polypeptide II (EMAP II) in solution
Authors	:	Lozhko, D.; Kolomiiets, L.; Kornelyuk, A.I.; Zhukov, I.
Deposited on	:	2023-04-02

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

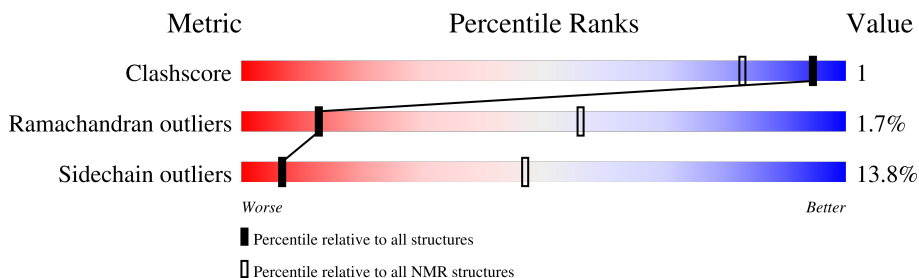|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance ⓘ

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

The overall completeness of chemical shifts assignment is 80%.

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		1.7%
Sidechain outliers		13.8%

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	169	80% 10% · 9%

2 Ensemble composition and analysis ⓘ

This entry contains 19 models. Model 11 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:7-A:113, A:120-A:152, A:156-A:169 (154)	0.65	11

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, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 19
2	3, 14, 18

3 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 2618 atoms, of which 1325 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Aminoacyl tRNA synthase complex-interacting multifunctional protein 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	169	2618	810	1325	226	245	12	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	ALA	-	expression tag	UNP Q12904
A	2	MET	-	expression tag	UNP Q12904

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: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1
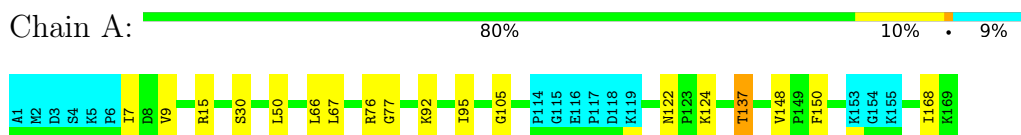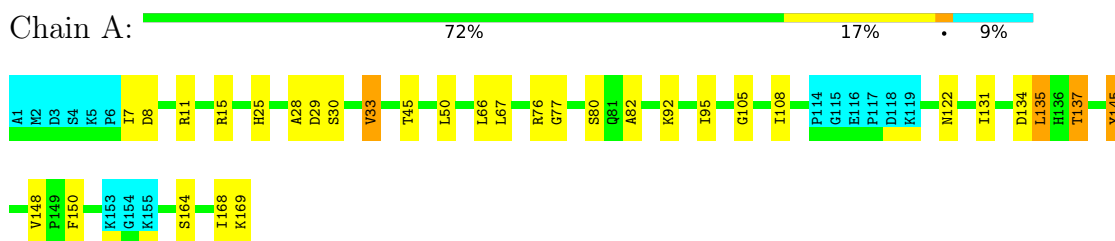

4.2 Scores per residue for each member of the ensemble
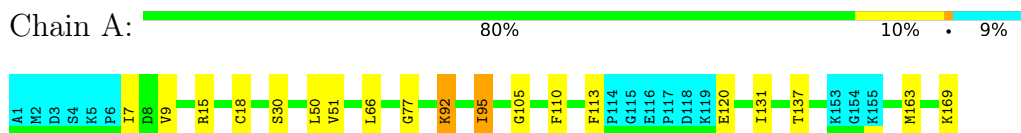
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1
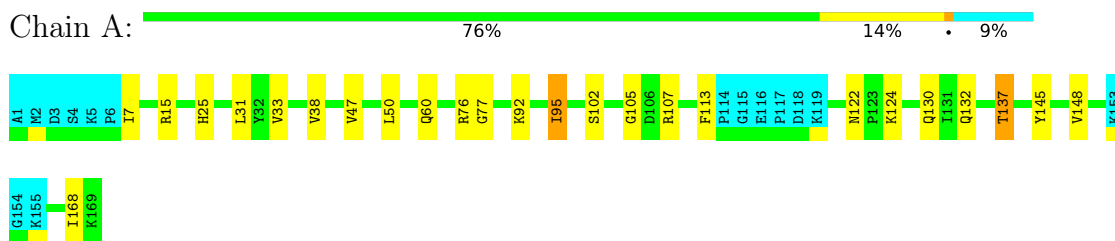

4.2.2 Score per residue for model 2

• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1

4.2.3 Score per residue for model 3

• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1

Chain A:

4.2.4 Score per residue for model 4

• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1

Chain A:

4.2.5 Score per residue for model 5

• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1

Chain A:

4.2.6 Score per residue for model 6

• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1

Chain A:

4.2.7 Score per residue for model 7

• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1

Chain A:

4.2.8 Score per residue for model 8

• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1

Chain A:

4.2.9 Score per residue for model 9

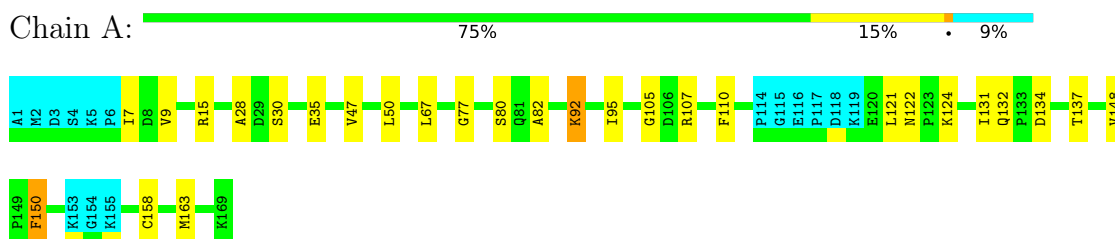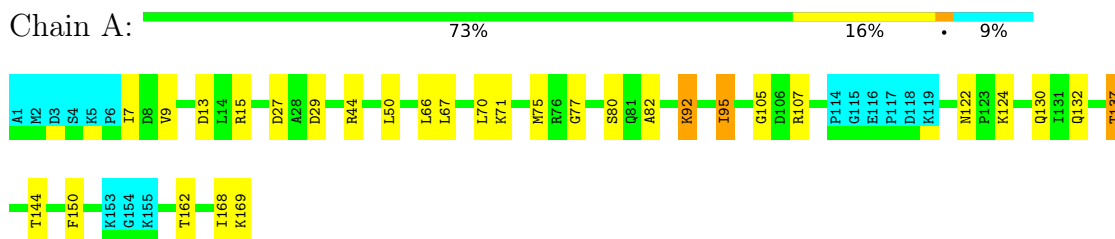• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1

Chain A:

4.2.10 Score per residue for model 10

• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1

Chain A:

4.2.11 Score per residue for model 11 (medoid)

• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1

Chain A:

A1 M2 D3 S4 K5 P6 I7 D8 V9 L12 R15 D29 S30 V47 L50 L66 K71 G77 S80 K92 I95 G105 D106 R107 I108 F113 P114 G115 E116 P117 D118 K119 E120 L121 Q132 T137 F150 K153 G154 K155 K169

4.2.12 Score per residue for model 12

• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1

Chain A: 75% 15% 9%

A1 M2 D3 S4 K5 P6 I7 D13 L14 R15 D27 A28 D29 S30 E34 L50 L66 K71 R76 G77 S80 K92 I95 S102 G105 P114 G115 E116 P117 D118 K119 N122 I131 T137 N138 D139 Y145 V148 P149 F150 K153

G154 K155 C158 N165 I168 K169

4.2.13 Score per residue for model 13

• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1

Chain A: 79% 11% 9%

A1 M2 D3 S4 K5 P6 I7 R15 E35 L50 V51 N52 H53 L66 L67 K71 G77 V78 L79 A82 K92 I93 E94 I95 G105 D106 R107 F110 P114 G115 E116 P117 D118 K119 K124 T137 K153 G154 K155 K169

4.2.14 Score per residue for model 14

• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1

Chain A: 76% 14% 9%

A1 M2 D3 S4 K5 P6 I7 D8 V9 R15 D27 A28 D29 S30 L50 V51 N52 H53 K71 G77 V78 L79 K92 I95 G105 P114 G115 E116 P117 D118 K119 E120 L121 N122 P123 K124 L135 H136 T137 Y145 K146 G147 V148 K153 G154 K155 M163

I168 K169

4.2.15 Score per residue for model 15

• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1

Chain A: 75% 14% 9%

A1 M2 D3 S4 K5 P6 I7 D8 V9 R15 V33 E34 E35 L50 L66 L67 R76 G77 S80 Q81 A82 K92 I95 N100 G105 D106 R107 F110 P114 G115 E116 P117 D118 K119 N122 P123 K124 T137 V148 P149 F150 K153 G154 K155

C158 M163 I168 K169

4.2.16　Score per residue for model 16

• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1
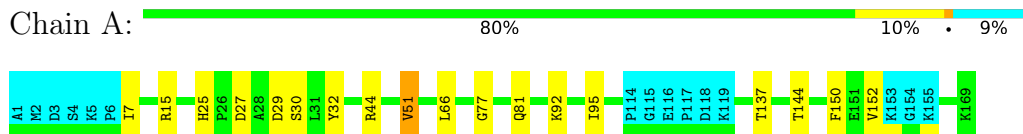
4.2.17　Score per residue for model 17

• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1
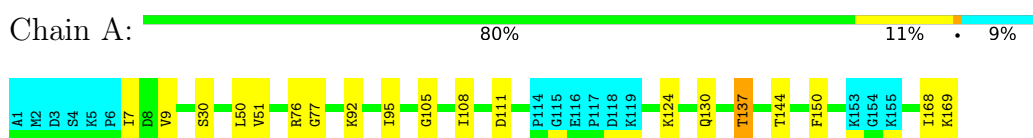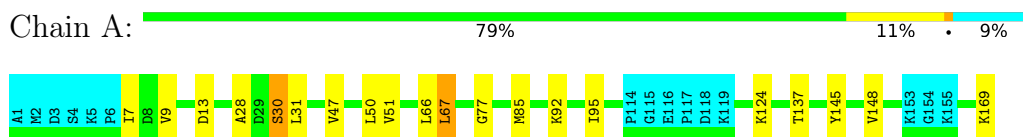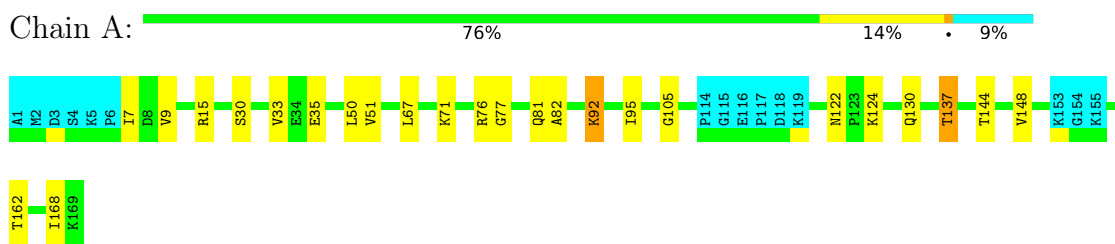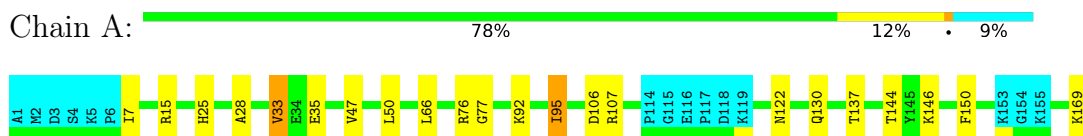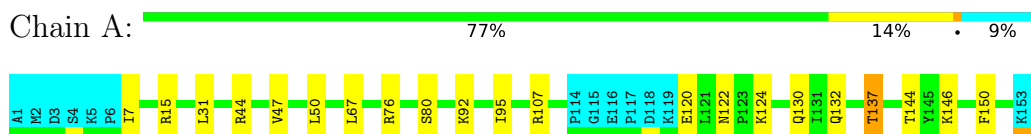

4.2.18　Score per residue for model 18

• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1
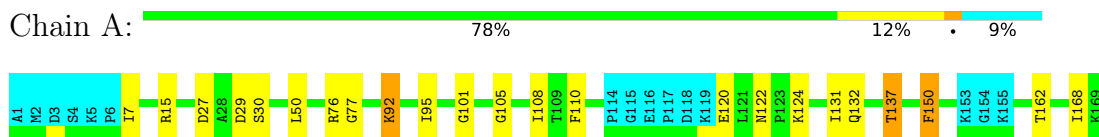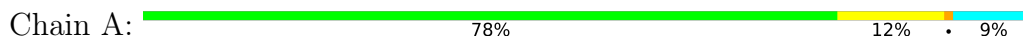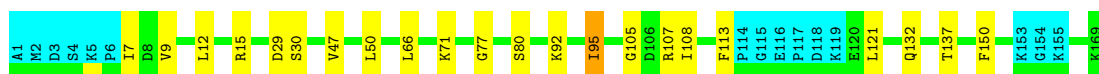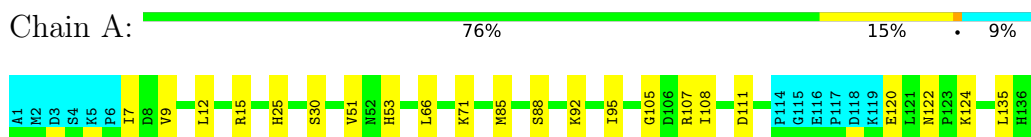

4.2.19　Score per residue for model 19

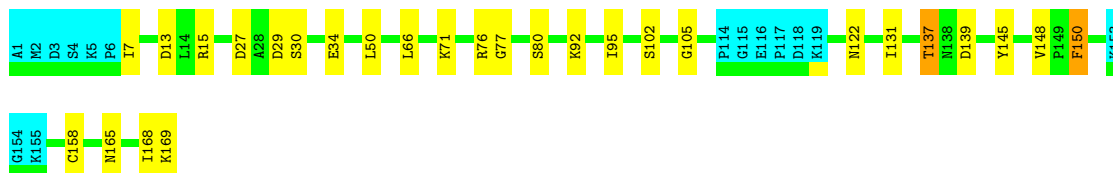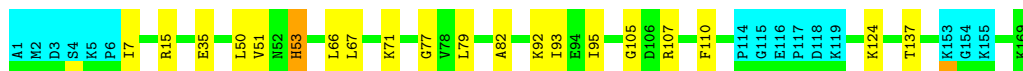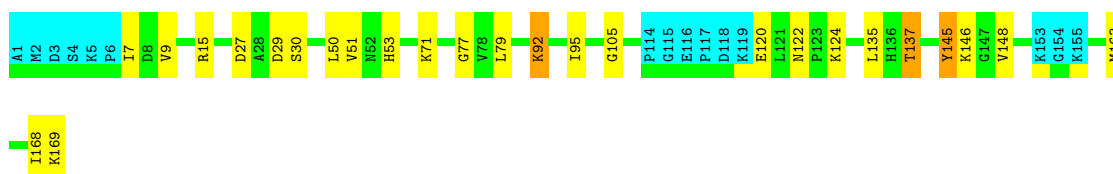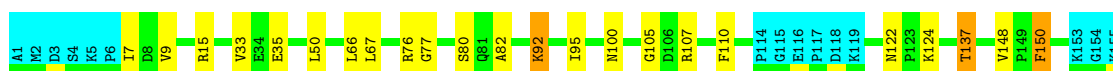• Molecule 1: Aminoacyl tRNA synthase complex-interacting multifunctional protein 1

I168 K169
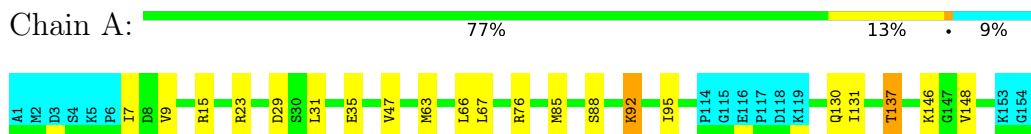
5　Refinement protocol and experimental data overview ⓘ

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

Of the 200 calculated structures, 19 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
CYANA	structure calculation	
YASARA	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	1761
Number of shifts mapped to atoms	1761
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	80%

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	0.59±0.01	0±0/1204 (0.0± 0.0%)	0.67±0.02	0±0/1632 (0.0± 0.0%)
All	All	0.59	0/22876 (0.0%)	0.67	1/31008 (0.0%)

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
1	A	145	TYR	CA-CB-CG	5.11	123.12	113.40	14	1

There are no chirality outliers.

There are no planarity outliers.

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	1184	1211	1211	3±1
All	All	22496	23009	23009	52

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
1:A:137:THR:HG23	1:A:168:ILE:HD11	0.60	1.74	8	12

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
1:A:30:SER:HB3	1:A:51:VAL:HG22	0.57	1.76	7	1
1:A:108:ILE:HG21	1:A:150:PHE:CD1	0.53	2.39	11	3
1:A:95:ILE:HD13	1:A:95:ILE:H	0.52	1.65	5	4
1:A:67:LEU:HD11	1:A:82:ALA:HA	0.50	1.82	13	6
1:A:92:LYS:HE2	1:A:162:THR:HG23	0.48	1.86	8	4
1:A:53:HIS:HB3	1:A:93:ILE:HD13	0.47	1.86	13	1
1:A:67:LEU:O	1:A:67:LEU:HD12	0.46	2.10	19	3
1:A:150:PHE:CE2	1:A:158:CYS:HB3	0.45	2.47	4	3
1:A:92:LYS:HE3	1:A:163:MET:SD	0.45	2.51	4	3
1:A:135:LEU:HD22	1:A:145:TYR:CB	0.44	2.41	1	1
1:A:67:LEU:HD23	1:A:67:LEU:O	0.43	2.13	7	1
1:A:25:HIS:CD2	1:A:33:VAL:HG23	0.43	2.48	9	3
1:A:25:HIS:CD2	1:A:25:HIS:H	0.43	2.31	18	1
1:A:150:PHE:CE1	1:A:158:CYS:HB3	0.43	2.49	17	1
1:A:108:ILE:HG21	1:A:150:PHE:CD2	0.42	2.49	6	2
1:A:67:LEU:HD13	1:A:70:LEU:O	0.41	2.15	5	1
1:A:95:ILE:H	1:A:95:ILE:HD13	0.41	1.75	3	1
1:A:92:LYS:HE2	1:A:163:MET:SD	0.40	2.57	14	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	153/169 (91%)	137±2 (90±1%)	13±2 (8±1%)	3±1 (2±1%)	13	56
All	All	2907/3211 (91%)	2612 (90%)	246 (8%)	49 (2%)	13	56

All 8 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	77	GLY	16
1	A	105	GLY	14
1	A	51	VAL	7
1	A	120	GLU	5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	28	ALA	4
1	A	101	GLY	1
1	A	53	HIS	1
1	A	29	ASP	1

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	135/147 (92%)	116±3 (86±2%)	19±3 (14±2%)	7	47
All	All	2565/2793 (92%)	2210 (86%)	355 (14%)	7	47

All 67 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	7	ILE	19
1	A	92	LYS	19
1	A	95	ILE	19
1	A	137	THR	19
1	A	15	ARG	17
1	A	50	LEU	16
1	A	30	SER	12
1	A	66	LEU	12
1	A	122	ASN	12
1	A	124	LYS	12
1	A	9	VAL	11
1	A	76	ARG	10
1	A	148	VAL	10
1	A	169	LYS	9
1	A	107	ARG	9
1	A	29	ASP	7
1	A	80	SER	7
1	A	47	VAL	7
1	A	130	GLN	7
1	A	150	PHE	7
1	A	71	LYS	7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	131	ILE	6
1	A	132	GLN	6
1	A	35	GLU	6
1	A	144	THR	6
1	A	145	TYR	5
1	A	110	PHE	5
1	A	27	ASP	5
1	A	33	VAL	4
1	A	31	LEU	4
1	A	146	LYS	4
1	A	135	LEU	3
1	A	113	PHE	3
1	A	13	ASP	3
1	A	44	ARG	3
1	A	85	MET	3
1	A	134	ASP	2
1	A	102	SER	2
1	A	121	LEU	2
1	A	111	ASP	2
1	A	67	LEU	2
1	A	81	GLN	2
1	A	12	LEU	2
1	A	79	LEU	2
1	A	53	HIS	2
1	A	88	SER	2
1	A	8	ASP	1
1	A	11	ARG	1
1	A	45	THR	1
1	A	164	SER	1
1	A	18	CYS	1
1	A	38	VAL	1
1	A	60	GLN	1
1	A	75	MET	1
1	A	106	ASP	1
1	A	34	GLU	1
1	A	139	ASP	1
1	A	165	ASN	1
1	A	100	ASN	1
1	A	25	HIS	1
1	A	32	TYR	1
1	A	51	VAL	1
1	A	152	VAL	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	159	ARG	1
1	A	23	ARG	1
1	A	63	MET	1
1	A	158	CYS	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 80% for the well-defined parts and 77% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_ chem_ shift_ list_ 1*

7.1.1 Bookkeeping ⓘ

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

Total number of shifts	1761
Number of shifts mapped to atoms	1761
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	5

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	161	-0.09 ± 0.09	None needed (< 0.5 ppm)
$^{13}C_\beta$	150	0.14 ± 0.09	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	138	0.61 ± 0.31	None needed (imprecise)

7.1.3 Completeness of resonance assignments ⓘ

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 80%, i.e. 1667 atoms were assigned a chemical shift out of a possible 2090. 0 out of 28 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	580/757 (77%)	295/306 (96%)	152/308 (49%)	133/143 (93%)
Sidechain	1083/1252 (87%)	746/818 (91%)	337/388 (87%)	0/46 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	4/81 (5%)	3/41 (7%)	0/36 (0%)	1/4 (25%)
Overall	1667/2090 (80%)	1044/1165 (90%)	489/732 (67%)	134/193 (69%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 77%, i.e. 1758 atoms were assigned a chemical shift out of a possible 2272. 0 out of 28 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	611/828 (74%)	312/335 (93%)	161/338 (48%)	138/155 (89%)
Sidechain	1143/1363 (84%)	786/888 (89%)	357/425 (84%)	0/50 (0%)
Aromatic	4/81 (5%)	3/41 (7%)	0/36 (0%)	1/4 (25%)
Overall	1758/2272 (77%)	1101/1264 (87%)	518/799 (65%)	139/209 (67%)

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	9	VAL	HB	-0.82	0.43 − 3.54	-9.0
1	A	9	VAL	HG21	-0.68	-0.58 − 2.19	-5.4
1	A	9	VAL	HG22	-0.68	-0.58 − 2.19	-5.4
1	A	9	VAL	HG23	-0.68	-0.58 − 2.19	-5.4
1	A	146	LYS	HG3	-0.04	0.04 − 2.67	-5.3

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: