



Full wwPDB NMR Structure Validation Report ⓘ

Mar 10, 2026 – 09:11 AM UTC

PDB ID : 7PBH / pdb_00007pbh
BMRB ID : 50691
Title : Structural characterization of Thorarchaeota profilin indicates a eukaryotic-like fold but with an extended N-terminus
Authors : Celestine, C.
Deposited on : 2021-08-02

This is a Full wwPDB NMR Structure Validation Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/NMRValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

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

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

MolProbity : 4-5-2 with Phenix2.0
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
wwPDB-RCI : v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV : Wang et al. (2010)
wwPDB-ShiftChecker : v1.2
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

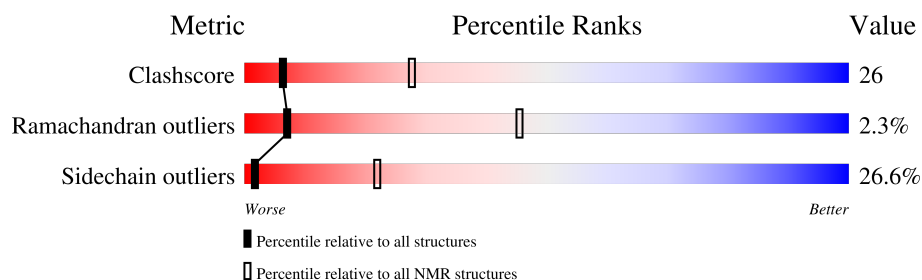
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 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	229148	14424
Ramachandran outliers	224038	12848
Sidechain outliers	223484	12823

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and a grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	145	

2 Ensemble composition and analysis

This entry contains 20 models. Model 9 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *target function*.

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:13-A:145 (133)	0.01	9

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 and 6 single-model clusters were found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 2193 atoms, of which 1085 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Profilin.

Mol	Chain	Residues	Atoms						Trace
1	A	145	Total	C	H	N	O	S	0
			2193	691	1085	191	222	4	

There are 2 discrepancies between the modelled and reference sequences:

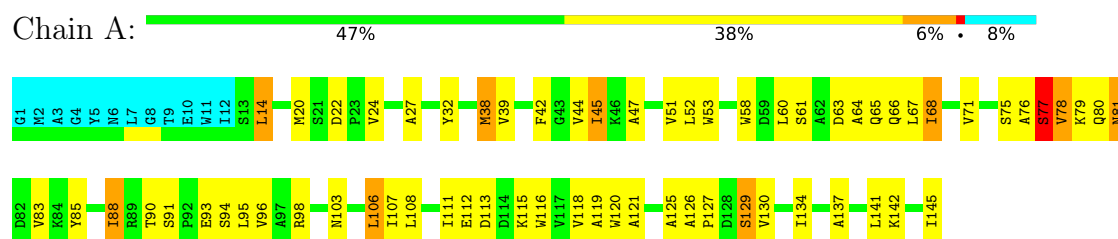
Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP A0A524EIQ6
A	3	ALA	CYS	conflict	UNP A0A524EIQ6

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: Profilin

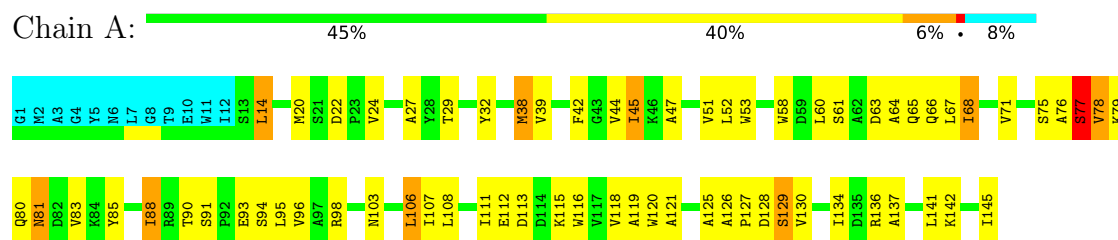


4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

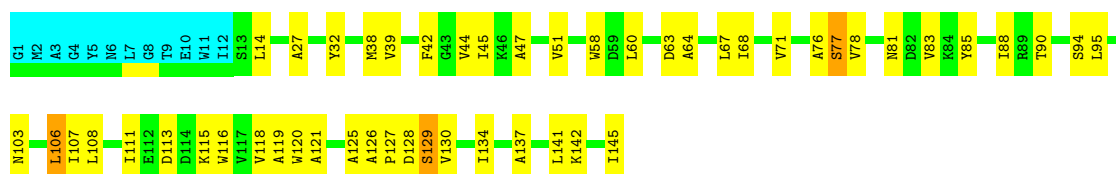
- Molecule 1: Profilin



4.2.2 Score per residue for model 2

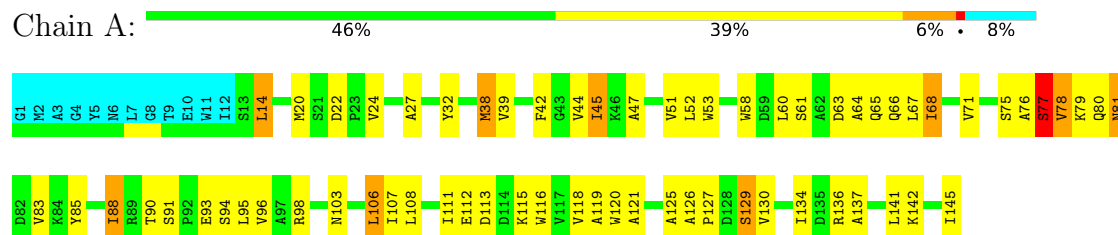
- Molecule 1: Profilin





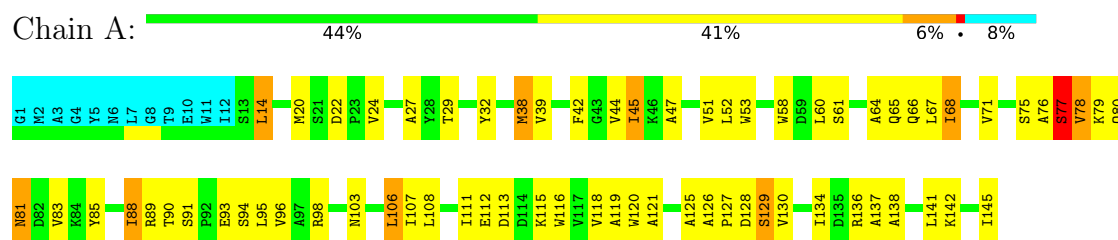
4.2.3 Score per residue for model 3

- Molecule 1: Profilin



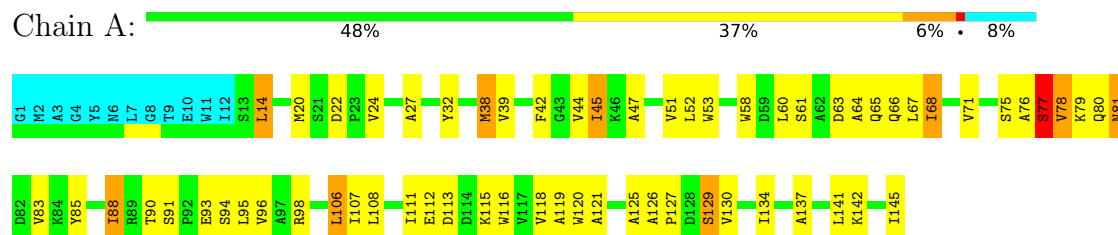
4.2.4 Score per residue for model 4

- Molecule 1: Profilin



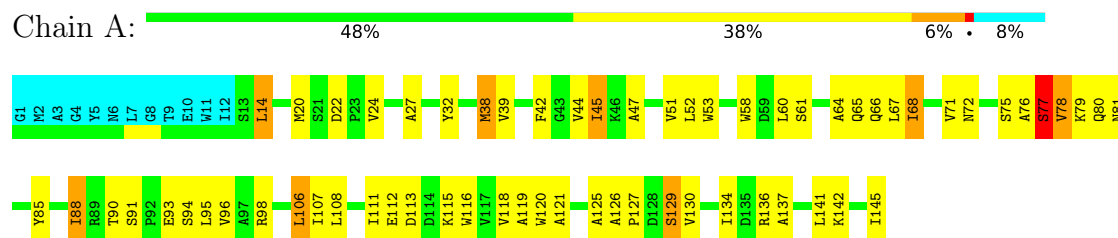
4.2.5 Score per residue for model 5

- Molecule 1: Profilin



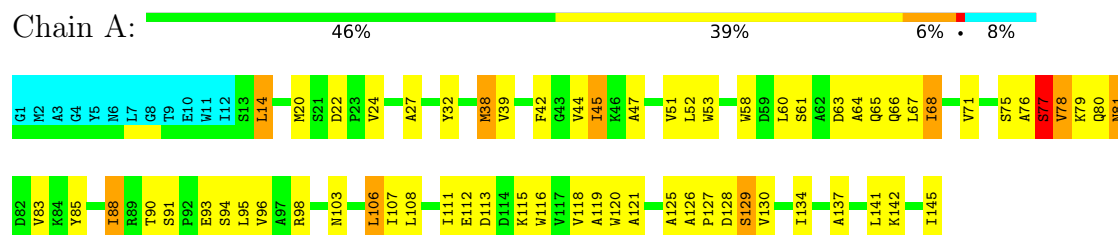
4.2.6 Score per residue for model 6

- Molecule 1: Profilin



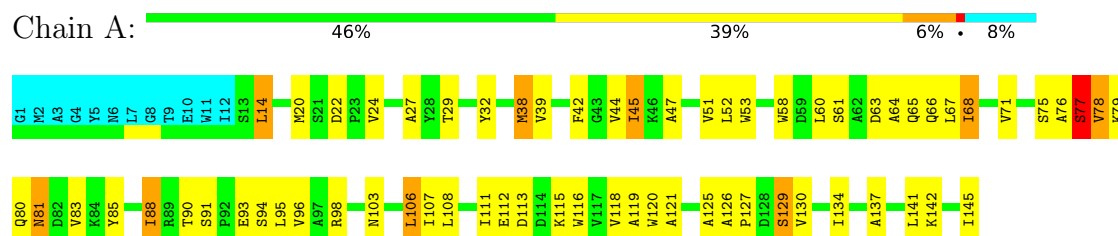
4.2.7 Score per residue for model 7

- Molecule 1: Profilin



4.2.8 Score per residue for model 8

- Molecule 1: Profilin



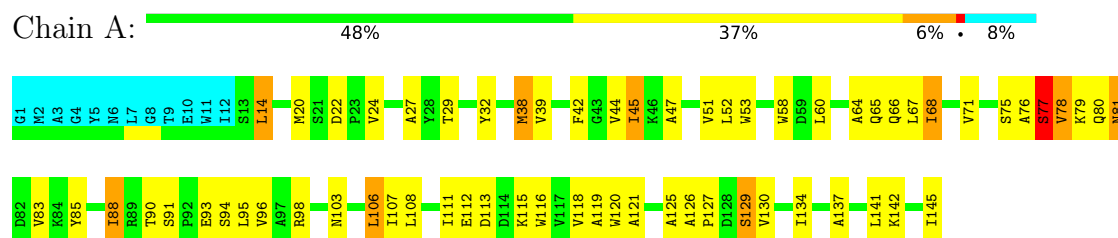
4.2.9 Score per residue for model 9 (medoid)

- Molecule 1: Profilin



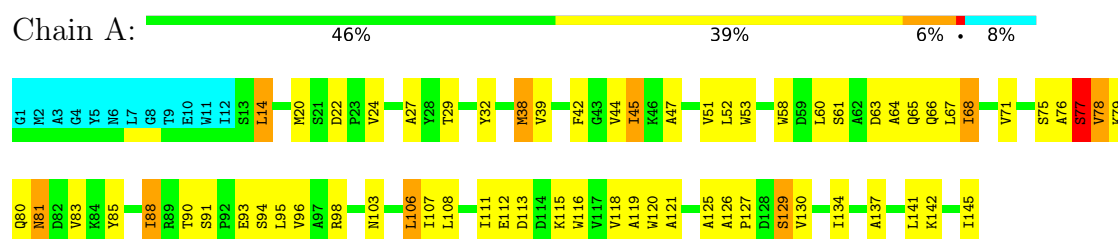
4.2.10 Score per residue for model 10

• Molecule 1: Profilin



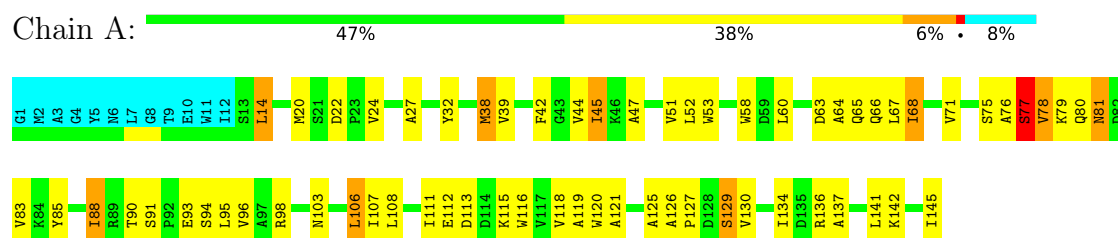
4.2.11 Score per residue for model 11

• Molecule 1: Profilin



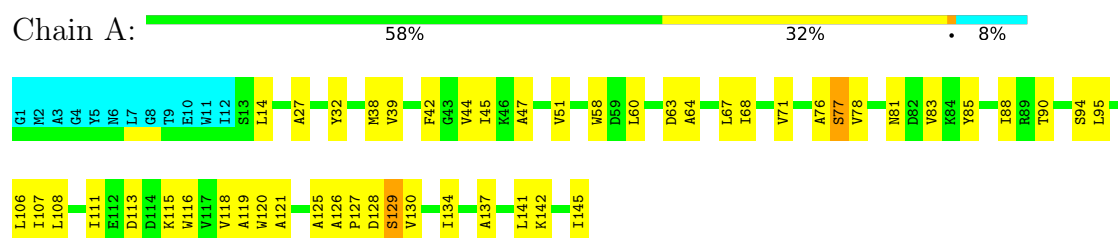
4.2.12 Score per residue for model 12

• Molecule 1: Profilin



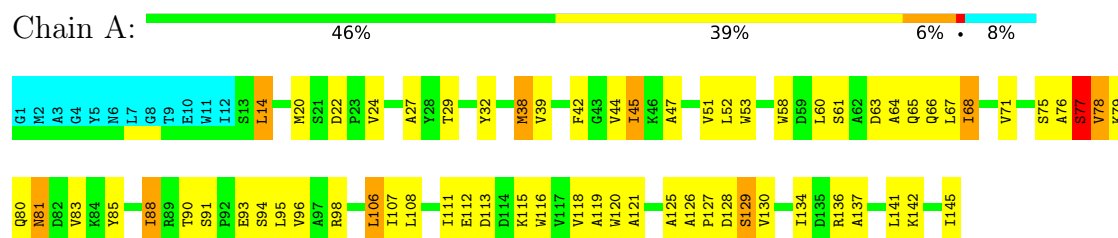
4.2.13 Score per residue for model 13

• Molecule 1: Profilin



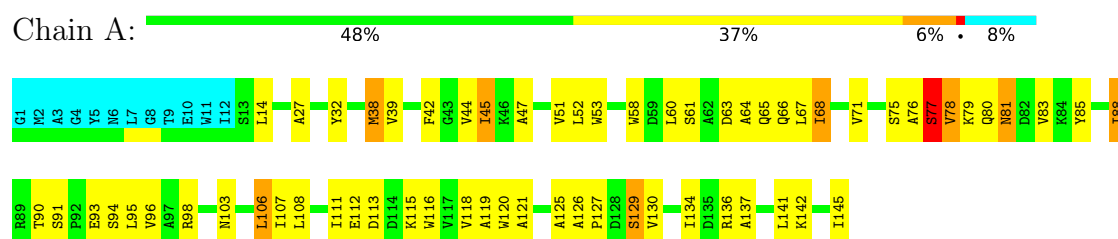
4.2.14 Score per residue for model 14

• Molecule 1: Profilin



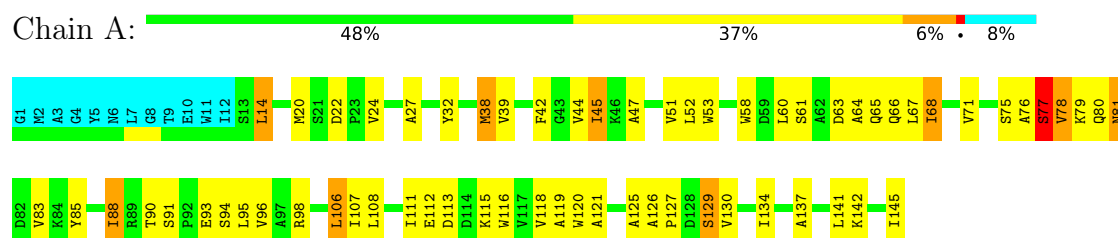
4.2.15 Score per residue for model 15

• Molecule 1: Profilin



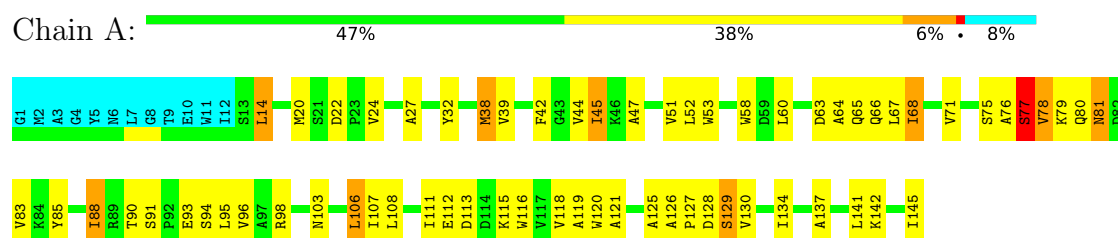
4.2.16 Score per residue for model 16

• Molecule 1: Profilin



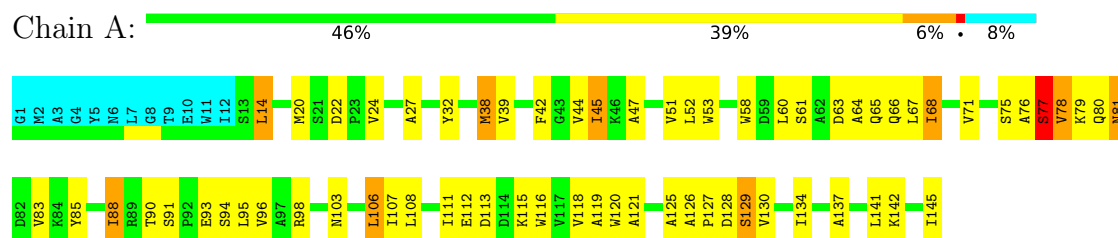
4.2.17 Score per residue for model 17

• Molecule 1: Profilin



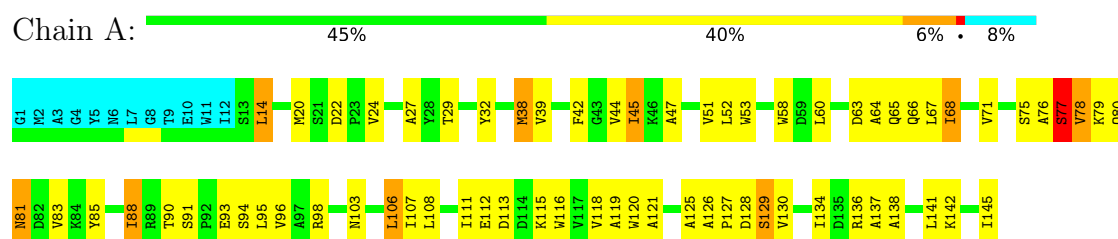
4.2.18 Score per residue for model 18

• Molecule 1: Profilin



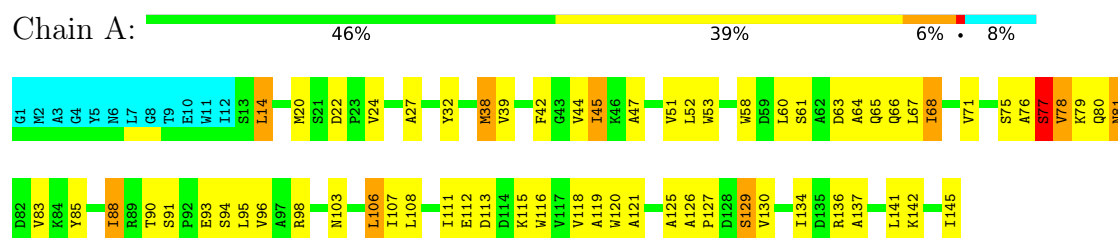
4.2.19 Score per residue for model 19

• Molecule 1: Profilin



4.2.20 Score per residue for model 20

• Molecule 1: Profilin



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: *target function*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
CYANA	structure calculation	
CYANA	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	1205
Number of shifts mapped to atoms	1205
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	66%

6 Model quality [i](#)

6.1 Standard geometry [i](#)

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

6.2 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	1017	1002	1002	52±1
All	All	20340	20040	20040	1030

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 26.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:27:ALA:HB1	1:A:137:ALA:HB1	0.94	1.36	10	20
1:A:111:ILE:HD12	1:A:142:LYS:CB	0.74	2.12	3	20
1:A:58:TRP:CH2	1:A:60:LEU:HD13	0.72	2.19	19	20
1:A:94:SER:O	1:A:95:LEU:HD23	0.72	1.85	5	20
1:A:108:LEU:HA	1:A:118:VAL:HG12	0.67	1.66	6	20
1:A:78:VAL:HG13	1:A:85:TYR:CE2	0.66	2.26	19	20
1:A:107:ILE:HG21	1:A:134:ILE:HG21	0.66	1.66	15	20
1:A:126:ALA:O	1:A:130:VAL:HG12	0.66	1.91	13	20
1:A:27:ALA:HB1	1:A:137:ALA:CB	0.65	2.18	13	20
1:A:106:LEU:HD22	1:A:119:ALA:O	0.62	1.94	8	20
1:A:38:MET:HB2	1:A:125:ALA:HB2	0.62	1.72	15	20
1:A:60:LEU:HD12	1:A:67:LEU:HG	0.61	1.70	2	20
1:A:47:ALA:HA	1:A:68:ILE:HD13	0.60	1.73	15	20
1:A:51:VAL:HG11	1:A:60:LEU:HG	0.60	1.73	15	20

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:106:LEU:HD13	1:A:107:ILE:N	0.59	2.12	15	20
1:A:94:SER:C	1:A:95:LEU:HD23	0.58	2.23	13	20
1:A:90:THR:HA	1:A:95:LEU:HD22	0.58	1.75	13	20
1:A:44:VAL:HG11	1:A:145:ILE:HB	0.58	1.74	20	20
1:A:32:TYR:CD1	1:A:39:VAL:HG12	0.57	2.34	2	20
1:A:107:ILE:CG2	1:A:134:ILE:HG21	0.57	2.30	18	20
1:A:51:VAL:HG13	1:A:64:ALA:CB	0.57	2.30	12	20
1:A:121:ALA:HB2	1:A:130:VAL:HG21	0.57	1.76	17	20
1:A:14:LEU:N	1:A:14:LEU:HD13	0.56	2.15	14	20
1:A:95:LEU:HB2	1:A:108:LEU:HD22	0.53	1.80	6	20
1:A:14:LEU:HD13	1:A:14:LEU:H	0.53	1.64	15	20
1:A:45:ILE:CD1	1:A:118:VAL:HG22	0.53	2.34	8	20
1:A:51:VAL:HG13	1:A:64:ALA:HB1	0.53	1.80	10	20
1:A:111:ILE:HD12	1:A:142:LYS:HB3	0.52	1.78	1	20
1:A:116:TRP:CD1	1:A:116:TRP:N	0.51	2.79	3	20
1:A:42:PHE:CE2	1:A:119:ALA:HB2	0.50	2.41	13	20
1:A:126:ALA:HB1	1:A:127:PRO:HD2	0.49	1.84	13	20
1:A:45:ILE:CG2	1:A:60:LEU:HD11	0.48	2.38	13	20
1:A:115:LYS:CB	1:A:145:ILE:HD11	0.48	2.38	14	20
1:A:126:ALA:HB3	1:A:129:SER:HB2	0.47	1.84	18	20
1:A:60:LEU:HD12	1:A:67:LEU:CG	0.47	2.39	10	20
1:A:121:ALA:HB1	1:A:125:ALA:CB	0.46	2.40	12	20
1:A:115:LYS:HB3	1:A:145:ILE:HD11	0.46	1.87	12	20
1:A:106:LEU:HD11	1:A:118:VAL:HB	0.46	1.86	20	20
1:A:107:ILE:O	1:A:107:ILE:HG23	0.46	2.10	15	20
1:A:45:ILE:N	1:A:45:ILE:HD13	0.46	2.26	17	20
1:A:119:ALA:HB3	1:A:134:ILE:HG22	0.46	1.86	11	20
1:A:107:ILE:HD13	1:A:134:ILE:HG12	0.45	1.88	15	20
1:A:120:TRP:CD1	1:A:120:TRP:C	0.44	2.95	2	20
1:A:63:ASP:O	1:A:67:LEU:HD23	0.43	2.14	12	17
1:A:111:ILE:HD12	1:A:142:LYS:CA	0.43	2.43	15	20
1:A:71:VAL:HG23	1:A:95:LEU:HD11	0.43	1.90	17	20
1:A:76:ALA:O	1:A:77:SER:CB	0.43	2.67	4	20
1:A:45:ILE:HD11	1:A:118:VAL:HG22	0.43	1.90	2	20
1:A:78:VAL:HG13	1:A:85:TYR:CZ	0.42	2.49	6	20
1:A:81:ASN:O	1:A:83:VAL:HG23	0.41	2.14	7	19
1:A:68:ILE:HD12	1:A:116:TRP:CZ2	0.41	2.49	18	16
1:A:45:ILE:HG13	1:A:116:TRP:CE2	0.40	2.51	15	16
1:A:111:ILE:HG21	1:A:138:ALA:O	0.40	2.17	19	2

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	132/145 (91%)	111±0 (84±0%)	18±0 (14±0%)	3±0 (2±0%)	7	45
All	All	2640/2900 (91%)	2220 (84%)	360 (14%)	60 (2%)	7	45

All 3 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	SER	20
1	A	88	ILE	20
1	A	113	ASP	20

6.3.2 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	109/117 (93%)	71±20 (65±19%)	26±8 (24±7%)	2	27
All	All	1932/2340 (83%)	1418 (73%)	514 (27%)	2	22

All 33 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	129	SER	20
1	A	141	LEU	20
1	A	106	LEU	19
1	A	75	SER	18
1	A	77	SER	18
1	A	78	VAL	18
1	A	79	LYS	18

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	80	GLN	18
1	A	81	ASN	18
1	A	88	ILE	18
1	A	91	SER	18
1	A	93	GLU	18
1	A	96	VAL	18
1	A	98	ARG	18
1	A	112	GLU	18
1	A	38	MET	17
1	A	45	ILE	17
1	A	52	LEU	17
1	A	53	TRP	17
1	A	65	GLN	17
1	A	66	GLN	17
1	A	68	ILE	17
1	A	14	LEU	16
1	A	20	MET	16
1	A	22	ASP	16
1	A	24	VAL	16
1	A	103	ASN	15
1	A	61	SER	13
1	A	136	ARG	10
1	A	128	ASP	9
1	A	29	THR	7
1	A	89	ARG	1
1	A	72	ASN	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 64% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *Thorprof.str*

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	1205
Number of shifts mapped to atoms	1205
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	136	2.52 ± 0.18	Should be applied
$^{13}\text{C}_\beta$	125	2.78 ± 0.17	Should be applied
$^{13}\text{C}'$	0	—	None (insufficient data)
^{15}N	133	-0.15 ± 0.33	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. 1155 atoms were assigned a chemical shift out of a possible 1744. 0 out of 22 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	514/664 (77%)	258/268 (96%)	128/266 (48%)	128/130 (98%)
Sidechain	621/986 (63%)	502/642 (78%)	119/304 (39%)	0/40 (0%)

Continued on next page...

Continued from previous page...

	Total	¹H	¹³C	¹⁵N
Aromatic	20/94 (21%)	18/45 (40%)	0/45 (0%)	2/4 (50%)
Overall	1155/1744 (66%)	778/955 (81%)	247/615 (40%)	130/174 (75%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 64%, i.e. 1203 atoms were assigned a chemical shift out of a possible 1894. 0 out of 23 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	¹H	¹³C	¹⁵N
Backbone	542/727 (75%)	273/295 (93%)	136/290 (47%)	133/142 (94%)
Sidechain	637/1052 (61%)	512/686 (75%)	125/325 (38%)	0/41 (0%)
Aromatic	24/115 (21%)	21/55 (38%)	0/55 (0%)	3/5 (60%)
Overall	1203/1894 (64%)	806/1036 (78%)	261/670 (39%)	136/188 (72%)

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	105	THR	HG1	6.14	0.08 – 2.19	23.7
1	A	87	THR	HG1	5.07	0.08 – 2.19	18.6
1	A	9	THR	CB	61.08	61.12 – 78.27	-5.0

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:

