WORLD WIDE wwPDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report (i)

Sep 7, 2023 – 04:29 pm BST

PDB ID	:	6F3K
EMDB ID	:	EMD-4179
BMRB ID	:	34205
Title	:	Combined solid-state NMR, solution-state NMR and EM data for structure determination of the tetrahedral aminopeptidase TET2 from P. horikoshii
Authors	:	Gauto, D.F.; Estrozi, L.F.; Schwieters, C.D.; Effantin, G.; Macek, P.; Sounier, R.; Kerfah, R.; Sivertsen, A.C.; Colletier, J.P.; Boisbouvier, J.; Schoehn, G.; Favier, A.; Schanda, P.
Deposited on	:	2017-11-28

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 (i) 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 (i)) were used in the production of this report:

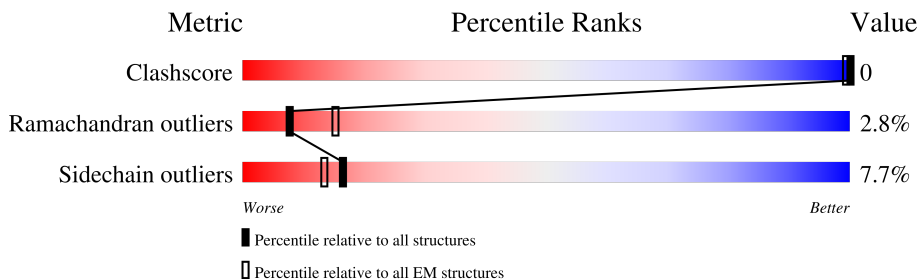|		
EMDB validation analysis	:	NOT EXECUTED
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	NOT EXECUTED
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY, SOLUTION NMR, SOLID-STATE NMR

The reported resolution of this entry is 4.10 Å.

The overall completeness of chemical shifts assignment is 32%.

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		0
Ramachandran outliers		2.8%
Sidechain outliers		7.7%

Worse   *Better*

▮ Percentile relative to all structures

▯ Percentile relative to all EM structures

Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

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 seg-
ment 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	353	87% 5% 9%

2　Ensemble composition and analysis ⓘ

This entry contains 10 models. Model 1 is the overall representative, medoid model (most similar to other models).

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:8-A:95,　　A:102-A:182, A:186-A:247,　A:252-A:318, A:328-A:351 (322)	0.50	1

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 1 clusters and 3 single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 6, 7, 10
Single-model clusters	5; 8; 9

3 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 5544 atoms, of which 2793 are hydrogens and 0 are deuteriums.

In the tables below, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Tetrahedral aminopeptidase.

Mol	Chain	Residues	Atoms						AltConf	Trace
1	A	353	Total 5542	C 1759	H 2793	N 469	O 512	S 9	0	

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
2	A	2	Total 2	Zn 2	

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.
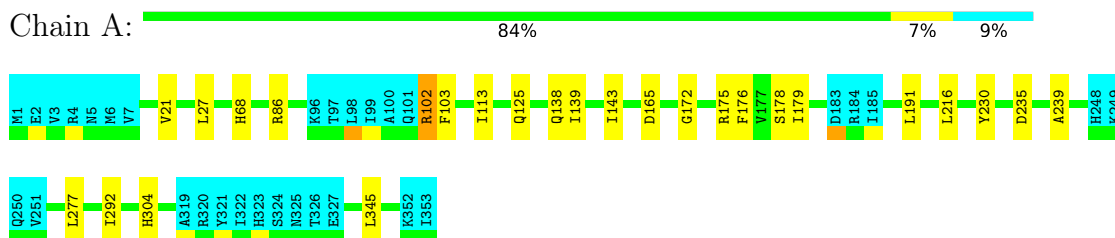
• Molecule 1: Tetrahedral aminopeptidase

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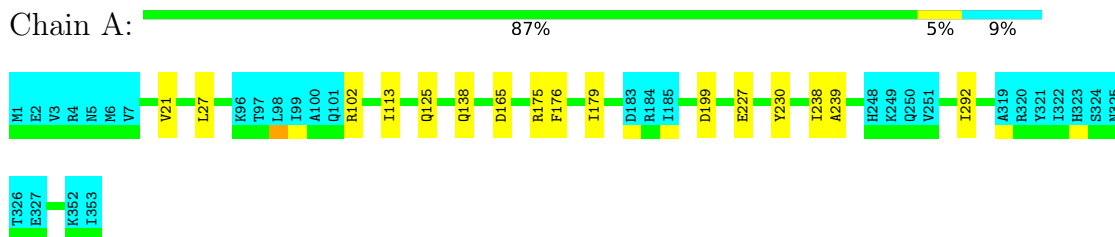
4.2 Scores per residue for each member of the ensemble
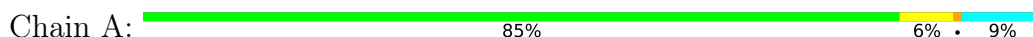
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1 (medoid)

• Molecule 1: Tetrahedral aminopeptidase

Chain A:

4.2.2 Score per residue for model 2
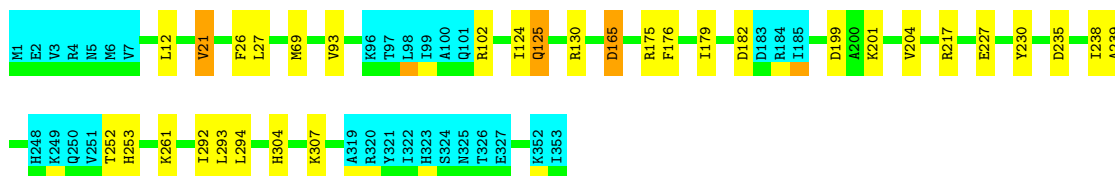
• Molecule 1: Tetrahedral aminopeptidase

Chain A:

M1 E2 V3 R4 N5 M6 V7 V21 L27 H68 L74 K96 T97 L98 I99 A100 Q101 R102 V105 K109 I124 Q125 I139 D165 R175 S178 D183 R184 I185 D199 E213 Y230 I238 A239 H248 K249 Q250 V251 K256 I292

L293 L294 H304 A319 R320 Y321 I322 H323 S324 N325 T326 E327 V328 K352 I353

4.2.3 Score per residue for model 3

• Molecule 1: Tetrahedral aminopeptidase

Chain A:　82%　8% · 9%

M1 E2 V3 R4 N5 M6 V7 L12 V21 F26 L27 M69 V93 K96 T97 L98 I99 A100 Q101 R102 I124 Q125 R130 D165 R175 F176 I179 D182 D183 R184 I185 D199 A200 K201 V204 R217 E227 Y230 D235 I238 A239

H248 K249 Q250 V251 T252 H253 K261 I292 L293 L294 H304 K307 A319 R320 Y321 I322 H323 S324 N325 T326 E327 K352 I353

4.2.4 Score per residue for model 4

• Molecule 1: Tetrahedral aminopeptidase

Chain A:　84%　7%　9%

M1 E2 V3 R4 N5 M6 V7 V21 L27 V32 L49 M64 M69 K96 T97 L98 I99 A100 Q101 K104 I113 Q125 I139 D165 G172 R175 F176 I179 A180 F181 D182 D183 R184 I185 L216 R217 Y230 A239 H248 K249 Q250

V251 T252 H253 L254 I292 A319 R320 Y321 I322 H323 S324 N325 T326 E327 H349 E350 L351 K352 I353

4.2.5 Score per residue for model 5

• Molecule 1: Tetrahedral aminopeptidase

Chain A:　82%　9% · 9%

M1 E2 V3 R4 N5 M6 V7 V21 F26 L27 K56 E59 V76 L85 R86 K96 T97 L98 I99 A100 Q101 R102 I113 I124 Q125 K132 I139 I143 I162 D165 G172 R175 F176 I179 A180 F181 D182 D183 R184 I185 K198

L216 E227 Y230 V236 T237 I238 A239 H248 K249 Q250 V251 W276 I292 L293 L294 K307 A319 R320 Y321 I322 H323 S324 N325 T326 E327 I348 K352 I353

4.2.6 Score per residue for model 6

• Molecule 1: Tetrahedral aminopeptidase

Chain A:

4.2.7 Score per residue for model 7
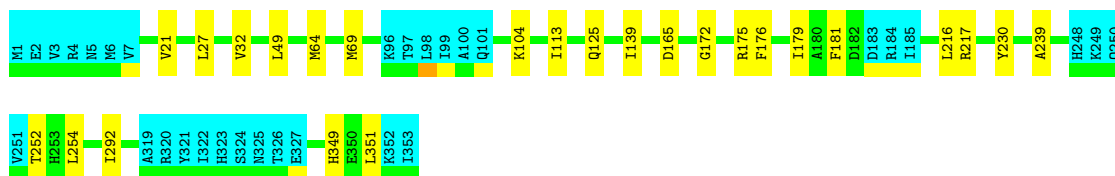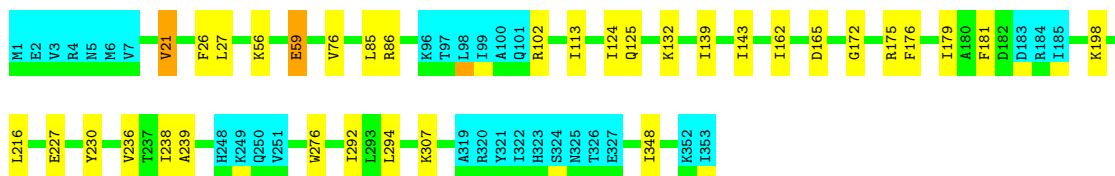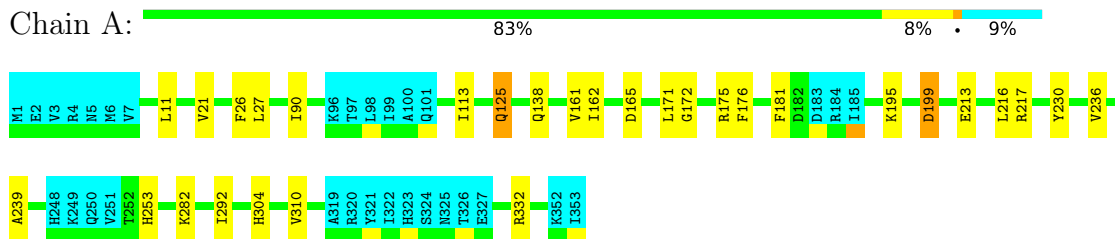
● Molecule 1: Tetrahedral aminopeptidase

Chain A:

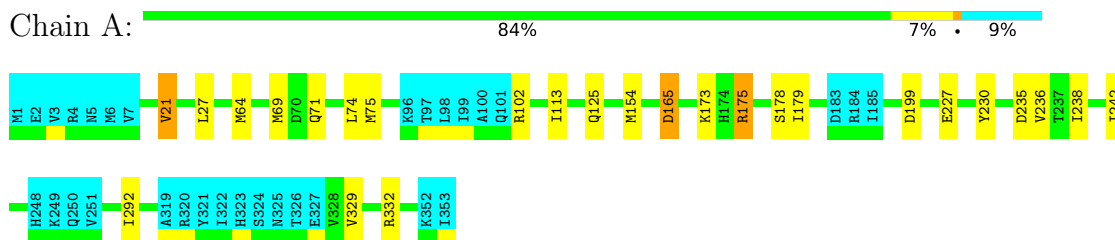### 4.2.8 Score per residue for model 8

● Molecule 1: Tetrahedral aminopeptidase

Chain A:

4.2.9 Score per residue for model 9

● Molecule 1: Tetrahedral aminopeptidase

Chain A:

4.2.10 Score per residue for model 10
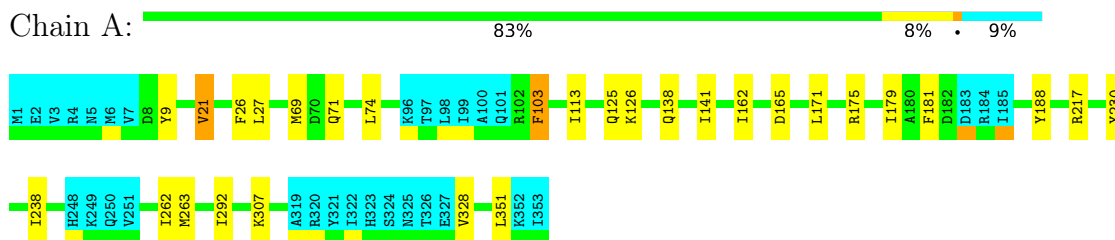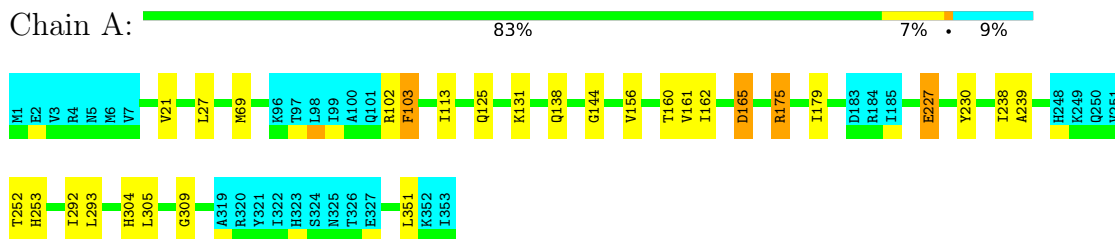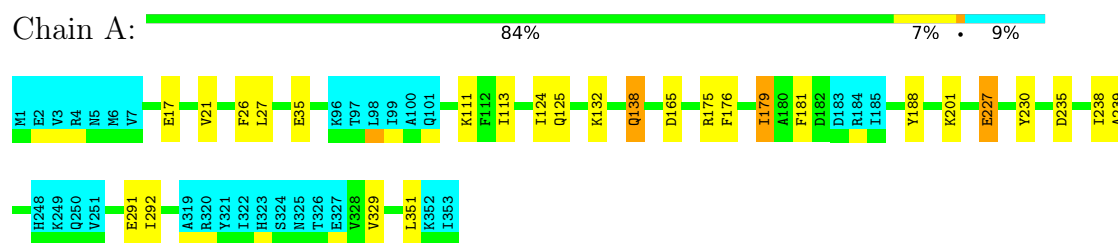
• Molecule 1: Tetrahedral aminopeptidase

5 Refinement protocol and experimental data overview (i)

Of the ? calculated structures, 10 were deposited, based on the following criterion: ?.

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

Software name	Classification	Version
CcpNmr Analysis	refinement	
CYANA	structure calculation	
Xplor-NIH	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	2
Total number of shifts	3036
Number of shifts mapped to atoms	3036
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	32%

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

6 Model quality ⓘ

6.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.96±0.00	0±0/2540 (0.0± 0.0%)	1.01±0.00	1±1/3442 (0.0± 0.0%)
All	All	0.96	0/25400 (0.0%)	1.01	11/34420 (0.0%)

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(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	239	ALA	N-CA-C	-5.85	95.20	111.00	3	8
1	A	235	ASP	N-CA-CB	5.23	120.02	110.60	3	3

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
All	All	24910	25190	25290	-

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 -.

There are no clashes.

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

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	322/353 (91%)	299±3 (93±1%)	14±3 (4±1%)	9±2 (3±1%)	8	42
All	All	3220/3530 (91%)	2991 (93%)	139 (4%)	90 (3%)	8	42

All 21 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	21	VAL	10
1	A	125	GLN	10
1	A	230	TYR	10
1	A	175	ARG	9
1	A	179	ILE	7
1	A	176	PHE	6
1	A	172	GLY	4
1	A	165	ASP	4
1	A	199	ASP	4
1	A	227	GLU	4
1	A	138	GLN	4
1	A	102	ARG	3
1	A	252	THR	3
1	A	178	SER	2
1	A	328	VAL	2
1	A	69	MET	2
1	A	103	PHE	2
1	A	59	GLU	1
1	A	131	LYS	1
1	A	144	GLY	1
1	A	309	GLY	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

WORLDWIDE
wwPDB
PROTEIN DATA BANK

was analysed and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	263/292 (90%)	243±3 (92±1%)	20±3 (8±1%)	16	64
All	All	2630/2920 (90%)	2428 (92%)	202 (8%)	16	64

All 91 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	27	LEU	10
1	A	292	ILE	10
1	A	165	ASP	9
1	A	113	ILE	8
1	A	238	ILE	7
1	A	304	HIS	5
1	A	26	PHE	5
1	A	181	PHE	5
1	A	102	ARG	4
1	A	139	ILE	4
1	A	216	LEU	4
1	A	124	ILE	4
1	A	21	VAL	4
1	A	217	ARG	4
1	A	351	LEU	4
1	A	162	ILE	4
1	A	103	PHE	3
1	A	74	LEU	3
1	A	175	ARG	3
1	A	294	LEU	3
1	A	253	HIS	3
1	A	307	LYS	3
1	A	227	GLU	3
1	A	236	VAL	3
1	A	69	MET	3
1	A	68	HIS	2
1	A	86	ARG	2
1	A	138	GLN	2
1	A	143	ILE	2
1	A	199	ASP	2
1	A	213	GLU	2
1	A	125	GLN	2
1	A	201	LYS	2
1	A	293	LEU	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	64	MET	2
1	A	132	LYS	2
1	A	161	VAL	2
1	A	171	LEU	2
1	A	332	ARG	2
1	A	71	GLN	2
1	A	179	ILE	2
1	A	329	VAL	2
1	A	188	TYR	2
1	A	191	LEU	1
1	A	235	ASP	1
1	A	277	LEU	1
1	A	345	LEU	1
1	A	105	VAL	1
1	A	109	LYS	1
1	A	256	LYS	1
1	A	328	VAL	1
1	A	12	LEU	1
1	A	93	VAL	1
1	A	130	ARG	1
1	A	182	ASP	1
1	A	204	VAL	1
1	A	261	LYS	1
1	A	32	VAL	1
1	A	49	LEU	1
1	A	104	LYS	1
1	A	254	LEU	1
1	A	349	HIS	1
1	A	56	LYS	1
1	A	59	GLU	1
1	A	76	VAL	1
1	A	85	LEU	1
1	A	198	LYS	1
1	A	276	TRP	1
1	A	348	ILE	1
1	A	11	LEU	1
1	A	90	ILE	1
1	A	195	LYS	1
1	A	282	LYS	1
1	A	310	VAL	1
1	A	75	MET	1
1	A	154	MET	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	173	LYS	1
1	A	178	SER	1
1	A	242	ILE	1
1	A	9	TYR	1
1	A	126	LYS	1
1	A	141	ILE	1
1	A	262	ILE	1
1	A	263	MET	1
1	A	156	VAL	1
1	A	160	THR	1
1	A	305	LEU	1
1	A	17	GLU	1
1	A	35	GLU	1
1	A	111	LYS	1
1	A	291	GLU	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 ⓘ

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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 (i)

The completeness of assignment taking into account all chemical shift lists is 32% for the well-defined parts and 31% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list*

7.1.1 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	1518
Number of shifts mapped to atoms	1518
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 (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	283	0.12 ± 0.15	None needed (< 0.5 ppm)
$^{13}C_\beta$	241	0.07 ± 0.21	None needed (< 0.5 ppm)
$^{13}C'$	282	0.17 ± 0.09	None needed (< 0.5 ppm)
^{15}N	289	1.25 ± 0.13	Should be applied

7.1.3 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 32%, i.e. 1404 atoms were assigned a chemical shift out of a possible 4367. 0 out of 56 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	915/1611 (57%)	130/660 (20%)	528/644 (82%)	257/307 (84%)
Sidechain	489/2476 (20%)	9/1620 (1%)	480/784 (61%)	0/72 (0%)

Continued on next page...

Continued from previous page...

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Aromatic	0/280 (0%)	0/146 (0%)	0/130 (0%)	0/4 (0%)
Overall	1404/4367 (32%)	139/2426 (6%)	1008/1558 (65%)	257/383 (67%)

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 31%, i.e. 1506 atoms were assigned a chemical shift out of a possible 4835. 0 out of 60 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Backbone	982/1766 (56%)	140/722 (19%)	565/706 (80%)	277/338 (82%)
Sidechain	524/2768 (19%)	9/1809 (0%)	515/871 (59%)	0/88 (0%)
Aromatic	0/301 (0%)	0/158 (0%)	0/139 (0%)	0/4 (0%)
Overall	1506/4835 (31%)	149/2689 (6%)	1080/1716 (63%)	277/430 (64%)

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

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.
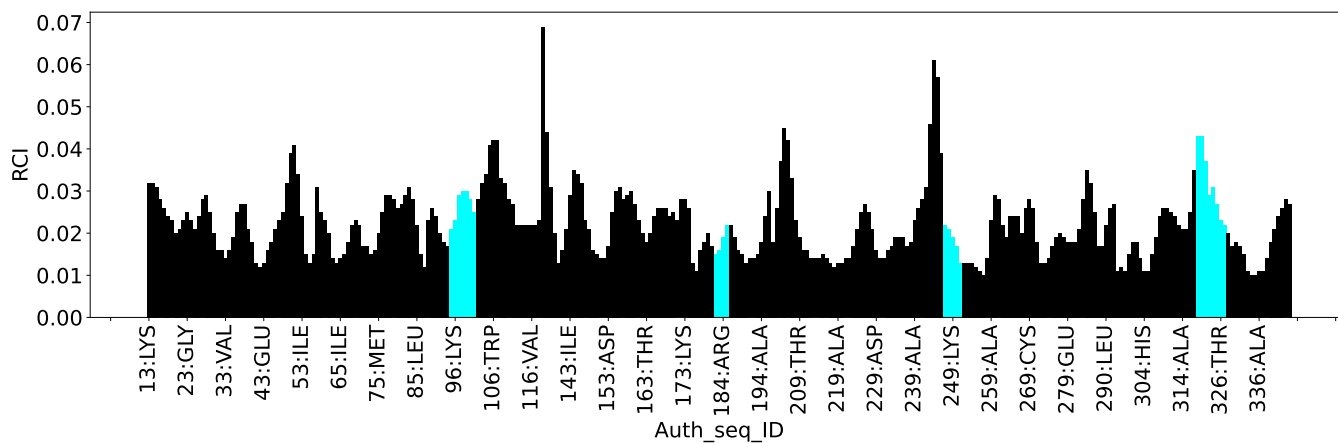
List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	255	GLY	H	12.12	5.23 – 11.42	6.1

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:

7.2   Chemical shift list 2

File name: working_cs.cif

Chemical shift list name: *27211.txt_assigned_chem_shift_list*

7.2.1   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	1518
Number of shifts mapped to atoms	1518
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

7.2.2   Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	283	0.11 ± 0.12	None needed (< 0.5 ppm)
$^{13}C_\beta$	241	0.07 ± 0.14	None needed (< 0.5 ppm)
$^{13}C'$	282	0.16 ± 0.09	None needed (< 0.5 ppm)
^{15}N	289	1.25 ± 0.17	Should be applied

7.2.3 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 32%, i.e. 1404 atoms were assigned a chemical shift out of a possible 4367. 0 out of 56 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	915/1611 (57%)	130/660 (20%)	528/644 (82%)	257/307 (84%)
Sidechain	489/2476 (20%)	9/1620 (1%)	480/784 (61%)	0/72 (0%)
Aromatic	0/280 (0%)	0/146 (0%)	0/130 (0%)	0/4 (0%)
Overall	1404/4367 (32%)	139/2426 (6%)	1008/1558 (65%)	257/383 (67%)

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 31%, i.e. 1506 atoms were assigned a chemical shift out of a possible 4835. 0 out of 60 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	982/1766 (56%)	140/722 (19%)	565/706 (80%)	277/338 (82%)
Sidechain	524/2768 (19%)	9/1809 (0%)	515/871 (59%)	0/88 (0%)
Aromatic	0/301 (0%)	0/158 (0%)	0/139 (0%)	0/4 (0%)
Overall	1506/4835 (31%)	149/2689 (6%)	1080/1716 (63%)	277/430 (64%)

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

7.2.4 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
2	A	255	GLY	H	12.12	5.23 – 11.42	6.1

7.2.5 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: