



# Full wwPDB NMR Structure Validation Report ⓘ

Feb 13, 2017 – 12:21 am GMT

PDB ID : 2LRO  
Title : Solution structure, dynamics and binding studies of CtCBM11  
Authors : Viegas, A.; Cabrita, E.J.  
Deposited on : 2012-04-11

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

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/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 : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
RCI : v\_1n\_11\_5\_13\_A (Berjanski et al., 2005)  
PANAV : Wang et al. (2010)  
ShiftChecker : trunk28760  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

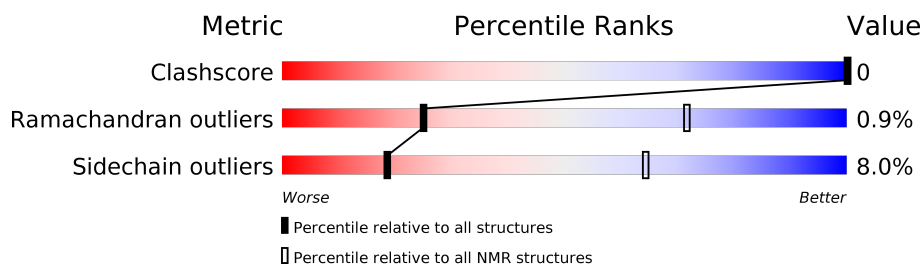
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 90%.

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	125131	11601
Ramachandran outliers	121729	10391
Sidechain outliers	121581	10367

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  $\geq 3$ , 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	172	 88% 8% 5%

## 2 Ensemble composition and analysis

This entry contains 20 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:6-A:169 (164)	0.59	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 1 single-model cluster was found.

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

### 3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2613 atoms, of which 1276 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Endoglucanase H.

Mol	Chain	Residues	Atoms						Trace
1	A	172	Total	C	H	N	O	S	0
			2611	846	1276	216	267	6	

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	EXPRESSION TAG	UNP P16218
A	2	ALA	-	EXPRESSION TAG	UNP P16218
A	3	SER	-	EXPRESSION TAG	UNP P16218
A	171	LEU	-	EXPRESSION TAG	UNP P16218
A	172	GLU	-	EXPRESSION TAG	UNP P16218

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

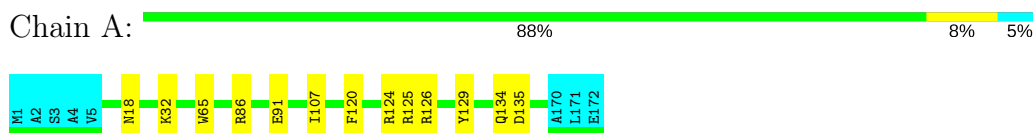
Mol	Chain	Residues	Atoms	
2	A	2	Total	Ca
			2	2

## 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: Endoglucanase H

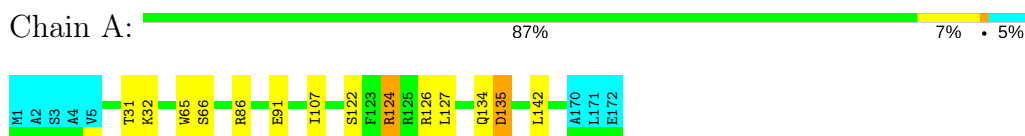


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

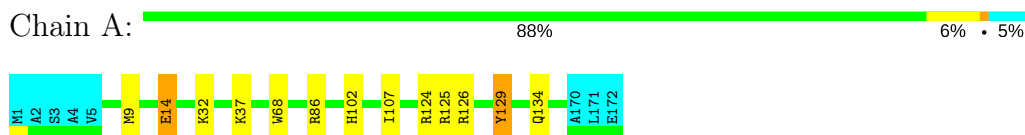
#### 4.2.1 Score per residue for model 1 (medoid)

- Molecule 1: Endoglucanase H



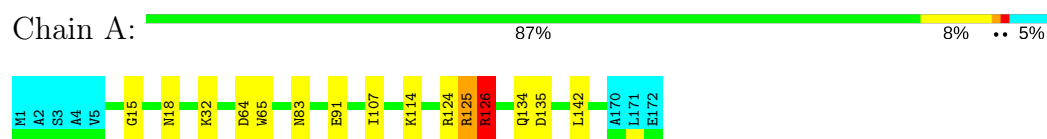
#### 4.2.2 Score per residue for model 2

- Molecule 1: Endoglucanase H



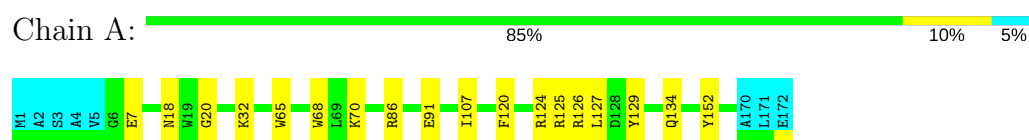
### 4.2.3 Score per residue for model 3

- Molecule 1: Endoglucanase H



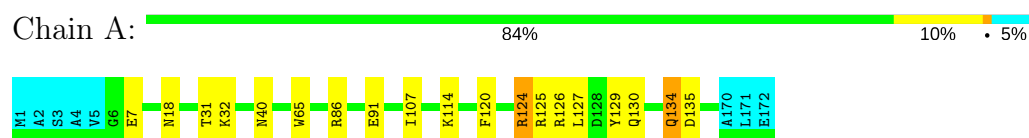
### 4.2.4 Score per residue for model 4

- Molecule 1: Endoglucanase H



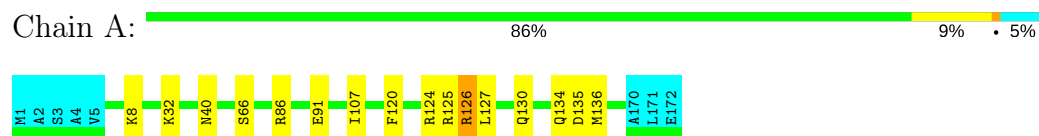
### 4.2.5 Score per residue for model 5

- Molecule 1: Endoglucanase H



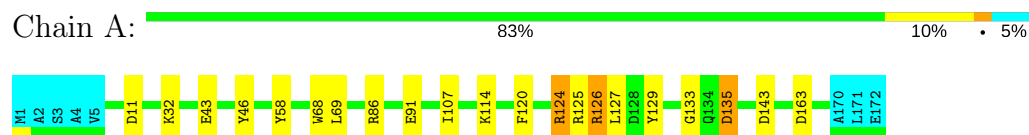
### 4.2.6 Score per residue for model 6

- Molecule 1: Endoglucanase H



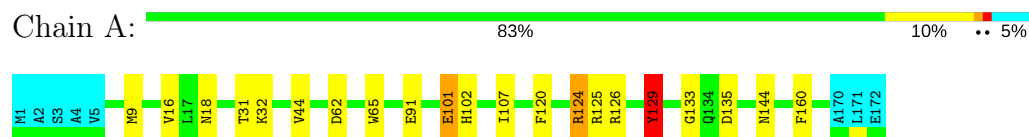
### 4.2.7 Score per residue for model 7

- Molecule 1: Endoglucanase H



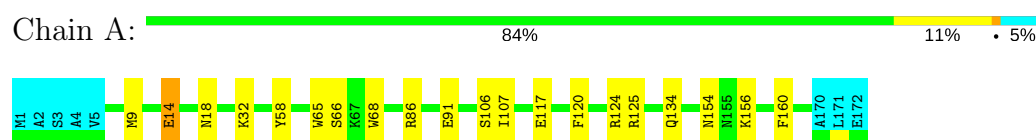
### 4.2.8 Score per residue for model 8

- Molecule 1: Endoglucanase H



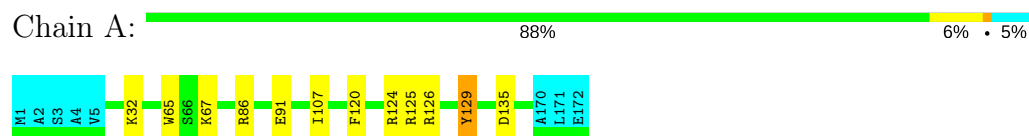
### 4.2.9 Score per residue for model 9

- Molecule 1: Endoglucanase H



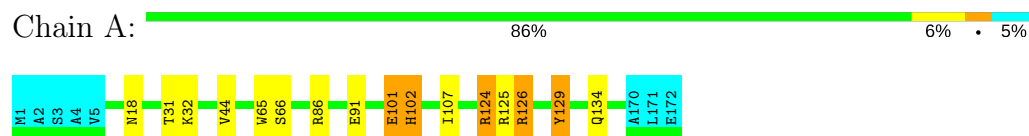
### 4.2.10 Score per residue for model 10

- Molecule 1: Endoglucanase H



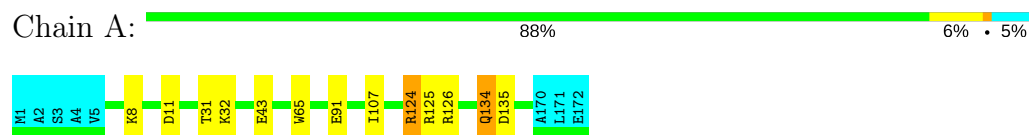
### 4.2.11 Score per residue for model 11

- Molecule 1: Endoglucanase H



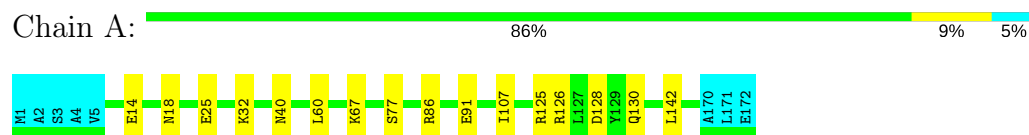
### 4.2.12 Score per residue for model 12

- Molecule 1: Endoglucanase H



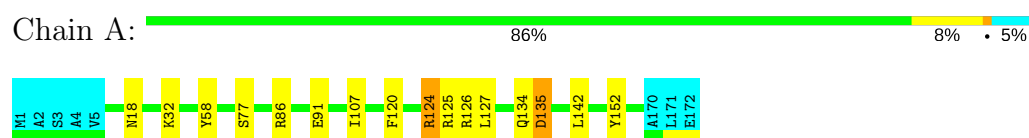
### 4.2.13 Score per residue for model 13

- Molecule 1: Endoglucanase H



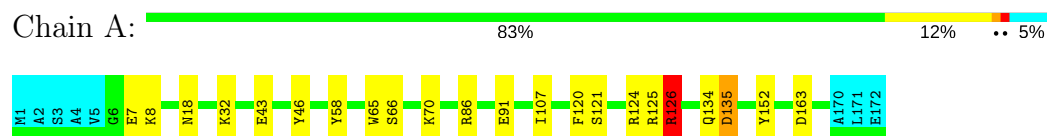
### 4.2.14 Score per residue for model 14

- Molecule 1: Endoglucanase H



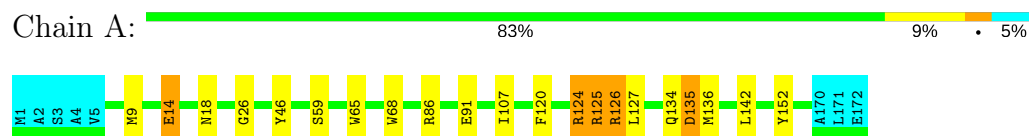
### 4.2.15 Score per residue for model 15

- Molecule 1: Endoglucanase H



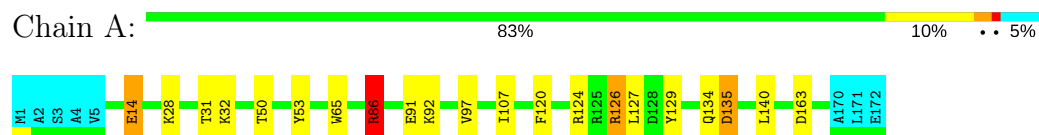
### 4.2.16 Score per residue for model 16

- Molecule 1: Endoglucanase H



### 4.2.17 Score per residue for model 17

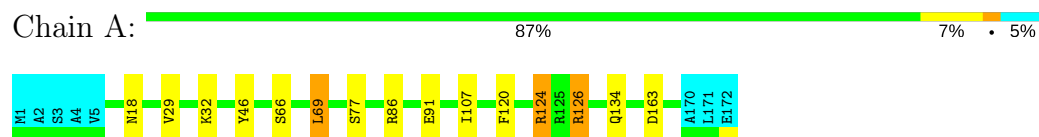
- Molecule 1: Endoglucanase H





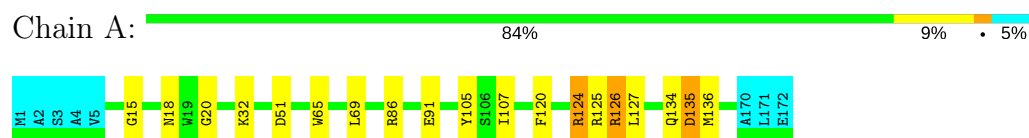
#### 4.2.18 Score per residue for model 18

- Molecule 1: Endoglucanase H



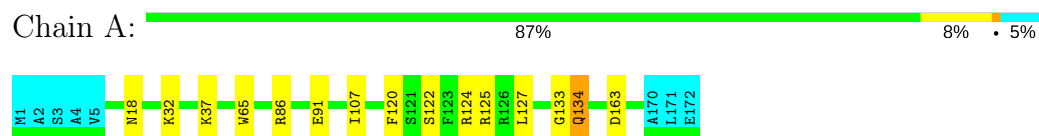
#### 4.2.19 Score per residue for model 19

- Molecule 1: Endoglucanase H



#### 4.2.20 Score per residue for model 20

- Molecule 1: Endoglucanase H



## 5 Refinement protocol and experimental data overview

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

Of the 20 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 solution	2.1
AMBER	refinement	9.0

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)	2lro_cs.str
Number of chemical shift lists	1
Total number of shifts	1959
Number of shifts mapped to atoms	1959
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	90%

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

## 6 Model quality i

### 6.1 Standard geometry i

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

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	0.73±0.01	0±0/1314 (0.0±0.0%)	1.10±0.02	5±1/1778 (0.3±0.1%)
All	All	0.73	0/26280 (0.0%)	1.10	107/35560 (0.3%)

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
1	A	0.0±0.0	1.6±1.5
All	All	0	31

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(°)	Ideal(°)	Models	
								Worst	Total
1	A	86	ARG	NE-CZ-NH1	9.41	125.01	120.30	6	17
1	A	126	ARG	NE-CZ-NH1	9.20	124.90	120.30	18	16
1	A	125	ARG	NE-CZ-NH1	9.02	124.81	120.30	5	16
1	A	124	ARG	NE-CZ-NH1	8.35	124.47	120.30	12	12
1	A	91	GLU	OE1-CD-OE2	-8.23	113.42	123.30	19	19
1	A	14	GLU	OE1-CD-OE2	-7.61	114.17	123.30	9	5
1	A	125	ARG	NE-CZ-NH2	-7.12	116.74	120.30	9	1
1	A	135	ASP	CB-CG-OD1	6.93	124.54	118.30	5	7
1	A	129	TYR	CB-CG-CD2	-6.53	117.08	121.00	11	2
1	A	129	TYR	CB-CG-CD1	5.63	124.38	121.00	11	1
1	A	101	GLU	OE1-CD-OE2	-5.61	116.57	123.30	8	2
1	A	135	ASP	CB-CG-OD2	5.53	123.28	118.30	15	2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	163	ASP	CB-CG-OD2	5.52	123.27	118.30	17	3
1	A	105	TYR	CB-CG-CD2	-5.52	117.69	121.00	19	1
1	A	124	ARG	NE-CZ-NH2	-5.24	117.68	120.30	1	1
1	A	126	ARG	CD-NE-CZ	5.04	130.65	123.60	5	1
1	A	163	ASP	CB-CG-OD1	5.03	122.83	118.30	20	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)
1	A	134	GLN	Peptide	4
1	A	58	TYR	Sidechain	4
1	A	126	ARG	Sidechain	3
1	A	46	TYR	Sidechain	3
1	A	133	GLY	Peptide	3
1	A	129	TYR	Peptide,Sidechain	3
1	A	20	GLY	Peptide	2
1	A	163	ASP	Peptide	1
1	A	43	GLU	Peptide	1
1	A	86	ARG	Peptide	1
1	A	37	LYS	Peptide	1
1	A	69	LEU	Peptide	1
1	A	102	HIS	Sidechain	1
1	A	67	LYS	Peptide	1
1	A	68	TRP	Peptide	1
1	A	152	TYR	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	1281	1219	1217	0±0
All	All	25660	24380	24354	2

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:29:VAL:HG23	1:A:46:TYR:CE1	0.41	2.50	18	1
1:A:102:HIS:CE1	1:A:129:TYR:CE1	0.40	3.09	2	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	164/172 (95%)	150±2 (91±1%)	13±2 (8±1%)	1±1 (1±0%)	25	72
All	All	3280/3440 (95%)	2991 (91%)	261 (8%)	28 (1%)	25	72

All 9 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	120	PHE	14
1	A	142	LEU	3
1	A	102	HIS	2
1	A	68	TRP	2
1	A	66	SER	2
1	A	15	GLY	2
1	A	136	MET	1
1	A	26	GLY	1
1	A	16	VAL	1

### 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	140/145 (97%)	129±3 (92±2%)	11±3 (8±2%)	18	64
All	All	2800/2900 (97%)	2576 (92%)	224 (8%)	18	64

All 57 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	107	ILE	20
1	A	32	LYS	19
1	A	124	ARG	16
1	A	134	GLN	15
1	A	65	TRP	14
1	A	18	ASN	13
1	A	135	ASP	11
1	A	127	LEU	10
1	A	126	ARG	10
1	A	129	TYR	7
1	A	31	THR	6
1	A	9	MET	4
1	A	66	SER	4
1	A	14	GLU	4
1	A	125	ARG	3
1	A	8	LYS	3
1	A	130	GLN	3
1	A	114	LYS	3
1	A	7	GLU	3
1	A	69	LEU	3
1	A	40	ASN	3
1	A	152	TYR	3
1	A	77	SER	3
1	A	122	SER	2
1	A	160	PHE	2
1	A	43	GLU	2
1	A	101	GLU	2
1	A	44	VAL	2
1	A	136	MET	2
1	A	142	LEU	2
1	A	11	ASP	2
1	A	70	LYS	2
1	A	68	TRP	2
1	A	128	ASP	1
1	A	59	SER	1
1	A	50	THR	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	121	SER	1
1	A	53	TYR	1
1	A	117	GLU	1
1	A	25	GLU	1
1	A	86	ARG	1
1	A	97	VAL	1
1	A	64	ASP	1
1	A	28	LYS	1
1	A	156	LYS	1
1	A	67	LYS	1
1	A	62	ASP	1
1	A	143	ASP	1
1	A	140	LEU	1
1	A	83	ASN	1
1	A	51	ASP	1
1	A	92	LYS	1
1	A	144	ASN	1
1	A	60	LEU	1
1	A	106	SER	1
1	A	37	LYS	1
1	A	154	ASN	1

### 6.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

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

### 6.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 6.6 Ligand geometry [i](#)

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

## 6.7 Other polymers [i](#)

There are no such molecules in this entry.

## 6.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.



## 7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 90% for the well-defined parts and 89% for the entire structure.

### 7.1 Chemical shift list 1

File name: 2lro\_cs.str

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	1959
Number of shifts mapped to atoms	1959
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	18

#### 7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	167	$3.02 \pm 0.14$	Should be applied
$^{13}\text{C}_\beta$	147	$2.24 \pm 0.14$	Should be applied
$^{13}\text{C}'$	167	$3.30 \pm 0.15$	Should be applied
$^{15}\text{N}$	161	$0.47 \pm 0.36$	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 90%, i.e. 1738 atoms were assigned a chemical shift out of a possible 1936. 0 out of 18 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	796/810 (98%)	316/323 (98%)	324/328 (99%)	156/159 (98%)
Sidechain	782/921 (85%)	480/540 (89%)	292/347 (84%)	10/34 (29%)

*Continued on next page...*

*Continued from previous page...*

	<b>Total</b>	<b><sup>1</sup>H</b>	<b><sup>13</sup>C</b>	<b><sup>15</sup>N</b>
Aromatic	160/205 (78%)	91/107 (85%)	64/90 (71%)	5/8 (62%)
Overall	1738/1936 (90%)	887/970 (91%)	680/765 (89%)	171/201 (85%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 89%, i.e. 1787 atoms were assigned a chemical shift out of a possible 2015. 0 out of 20 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	<b>Total</b>	<b><sup>1</sup>H</b>	<b><sup>13</sup>C</b>	<b><sup>15</sup>N</b>
Backbone	821/850 (97%)	326/339 (96%)	334/344 (97%)	161/167 (96%)
Sidechain	806/960 (84%)	493/562 (88%)	303/364 (83%)	10/34 (29%)
Aromatic	160/205 (78%)	91/107 (85%)	64/90 (71%)	5/8 (62%)
Overall	1787/2015 (89%)	910/1008 (90%)	701/798 (88%)	176/209 (84%)

#### 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.

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	68	TRP	CG	27.38	118.98 – 101.98	-48.9
1	A	37	LYS	HB2	-0.58	3.03 – 0.53	-9.4
1	A	71	ILE	CG1	13.27	36.54 – 18.94	-8.2
1	A	140	LEU	HB2	-1.12	3.32 – -0.08	-8.1
1	A	58	TYR	HB2	-0.02	4.76 – 1.06	-7.9
1	A	142	LEU	CG	18.27	32.55 – 21.05	-7.4
1	A	17	LEU	CD2	11.49	32.60 – 15.60	-7.4
1	A	76	LYS	HD3	-0.06	2.75 – 0.45	-7.2
1	A	140	LEU	CG	18.85	32.55 – 21.05	-6.9
1	A	71	ILE	CG2	26.50	24.63 – 10.43	6.3
1	A	44	VAL	HB	0.07	3.59 – 0.39	-6.0
1	A	8	LYS	CD	21.93	34.86 – 23.06	-6.0
1	A	76	LYS	HB2	0.39	3.03 – 0.53	-5.5
1	A	46	TYR	HB2	0.88	4.76 – 1.06	-5.5
1	A	167	LEU	HD11	-0.71	2.16 – -0.64	-5.2
1	A	167	LEU	HD12	-0.71	2.16 – -0.64	-5.2
1	A	167	LEU	HD13	-0.71	2.16 – -0.64	-5.2
1	A	46	TYR	HB3	0.88	4.75 – 0.95	-5.2

### 7.1.5 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:

