



# Full wwPDB NMR Structure Validation Report ⓘ

Feb 12, 2017 – 11:55 pm GMT

PDB ID : 2LCQ  
Title : Solution structure of the endonuclease Nob1 from P.horikoshii  
Authors : Veith, T.; Martin, R.; Wurm, J.P.; Weis, B.; Duchardt-Ferner, E.; Safferthal, C.; Hennig, R.; Mirus, O.; Bohnsack, M.T.; Woehnert, J.; Schleiff, E.  
Deposited on : 2011-05-05

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

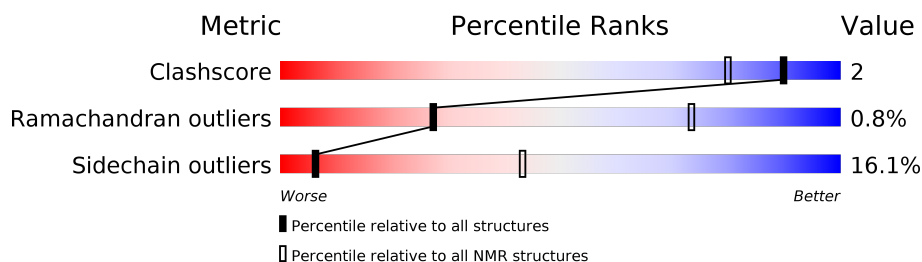
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 89%.

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	165	 75% 11% 12% .

## 2 Ensemble composition and analysis

This entry contains 20 models. Model 8 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:116 (110)	0.25	8
2	A:127-A:157 (31)	0.23	3

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 3 clusters and 2 single-model clusters were found.

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

### 3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2600 atoms, of which 1338 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Putative toxin VapC6.

Mol	Chain	Residues	Atoms						Trace
1	A	161	Total	C	H	N	O	S	0
			2599	806	1338	217	233	5	

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	EXPRESSION TAG	UNP O58440
A	-2	SER	-	EXPRESSION TAG	UNP O58440
A	-1	MET	-	EXPRESSION TAG	UNP O58440
A	0	GLY	-	EXPRESSION TAG	UNP O58440

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

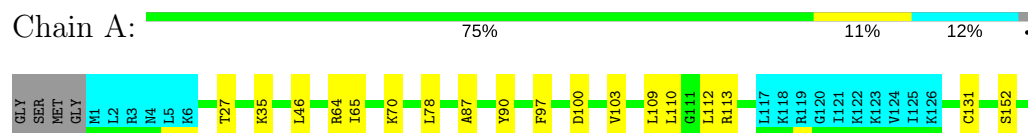
Mol	Chain	Residues	Atoms	
2	A	1	Total	Zn
			1	1

## 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: Putative toxin VapC6

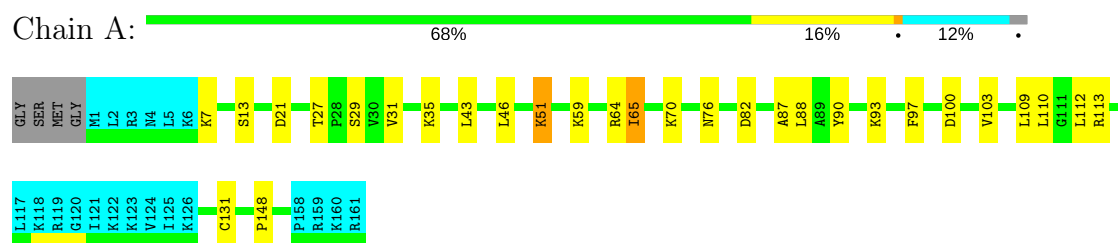


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

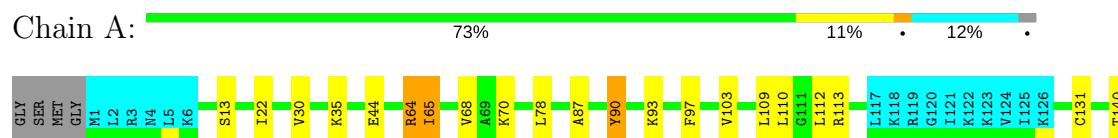
#### 4.2.1 Score per residue for model 1

- Molecule 1: Putative toxin VapC6



#### 4.2.2 Score per residue for model 2

- Molecule 1: Putative toxin VapC6

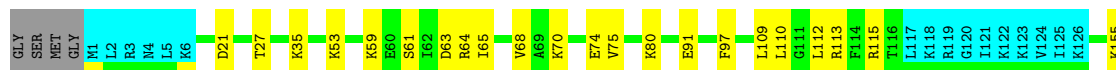




### 4.2.3 Score per residue for model 3

- Molecule 1: Putative toxin VapC6

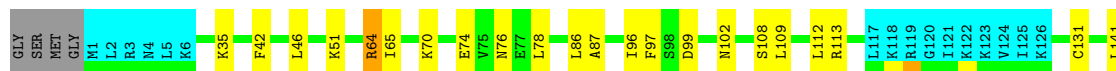
Chain A: 72% 13% 12% •



### 4.2.4 Score per residue for model 4

- Molecule 1: Putative toxin VapC6

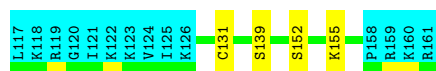
Chain A: 71% 14% • 12% •



### 4.2.5 Score per residue for model 5

- Molecule 1: Putative toxin VapC6

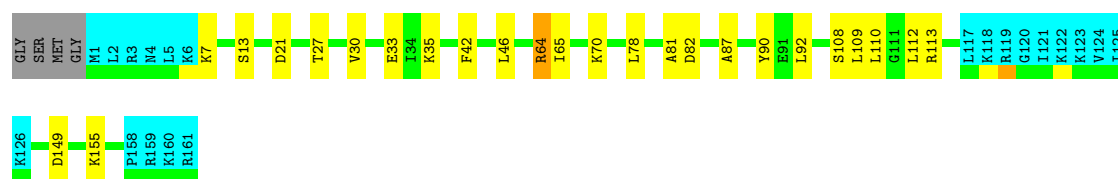
Chain A: 65% 19% • 12% •



### 4.2.6 Score per residue for model 6

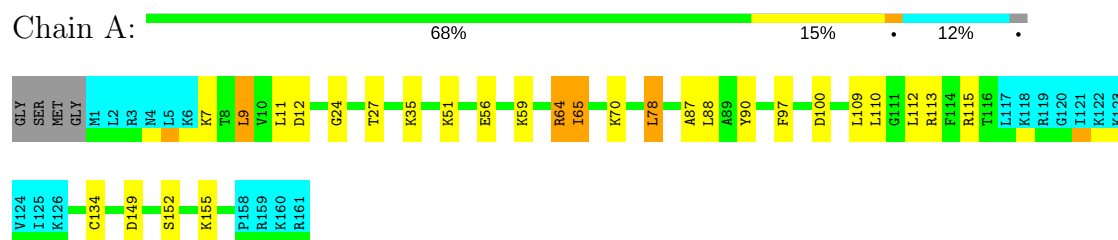
- Molecule 1: Putative toxin VapC6

Chain A: 70% 15% • 12% •



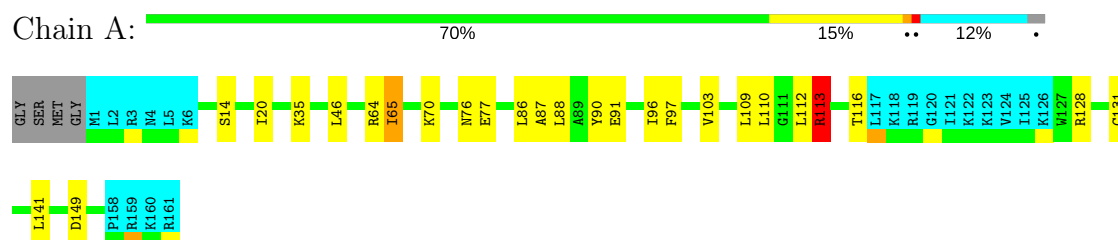
#### 4.2.7 Score per residue for model 7

- Molecule 1: Putative toxin VapC6



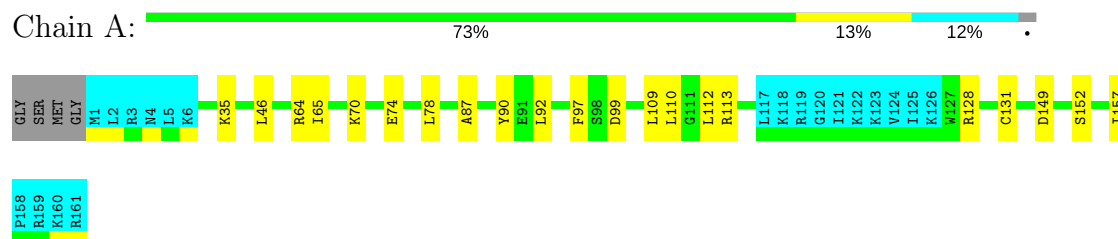
#### 4.2.8 Score per residue for model 8 (medoid)

- Molecule 1: Putative toxin VapC6



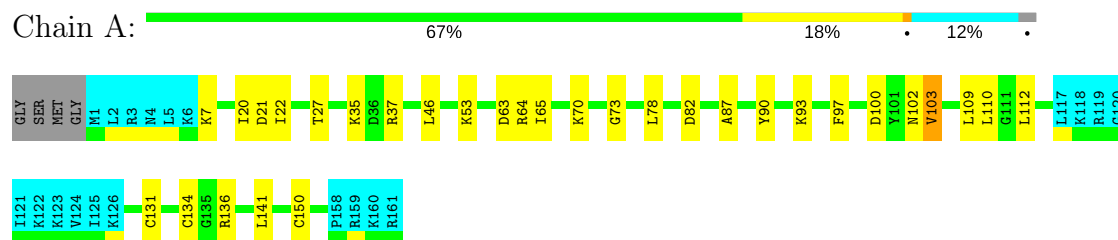
#### 4.2.9 Score per residue for model 9

- Molecule 1: Putative toxin VapC6



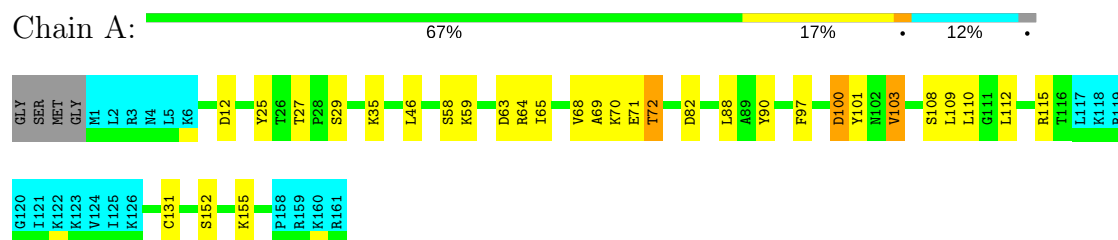
#### 4.2.10 Score per residue for model 10

- Molecule 1: Putative toxin VapC6



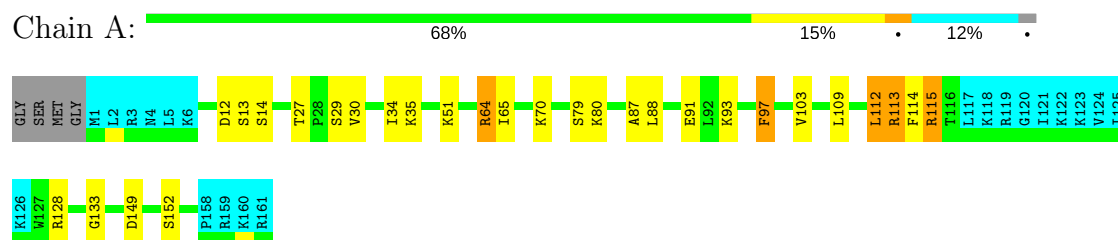
#### 4.2.11 Score per residue for model 11

- Molecule 1: Putative toxin VapC6



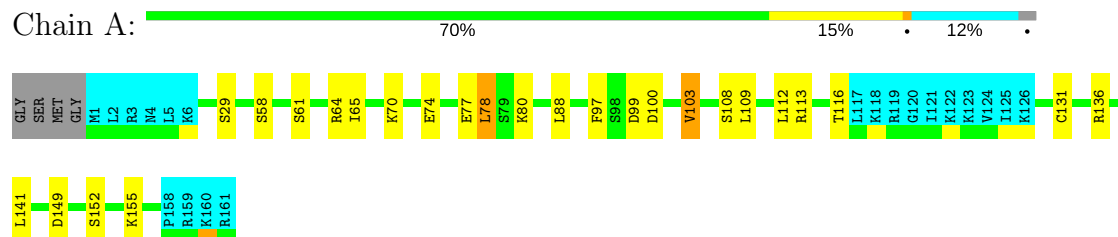
#### 4.2.12 Score per residue for model 12

- Molecule 1: Putative toxin VapC6



#### 4.2.13 Score per residue for model 13

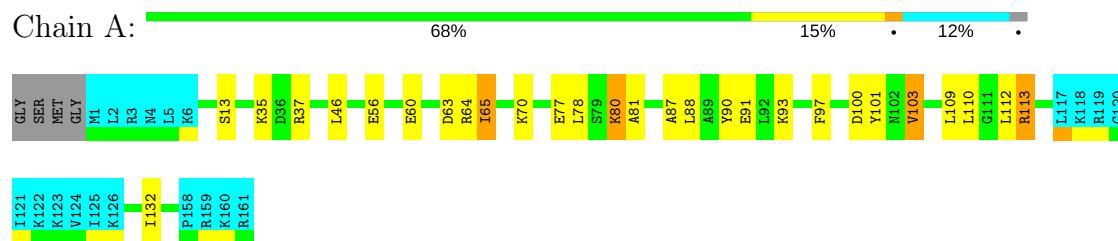
- Molecule 1: Putative toxin VapC6





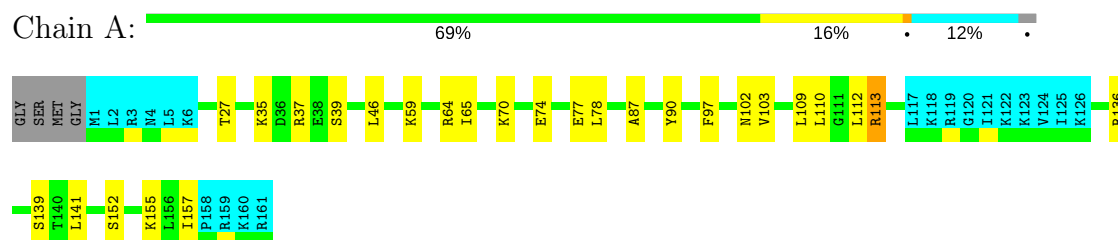
### 4.2.14 Score per residue for model 14

- Molecule 1: Putative toxin VapC6



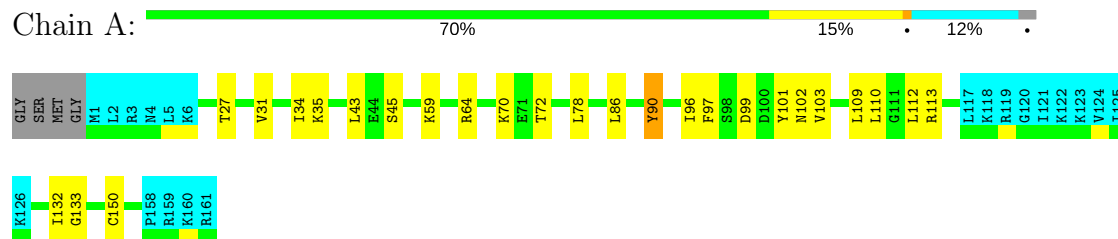
### 4.2.15 Score per residue for model 15

- Molecule 1: Putative toxin VapC6



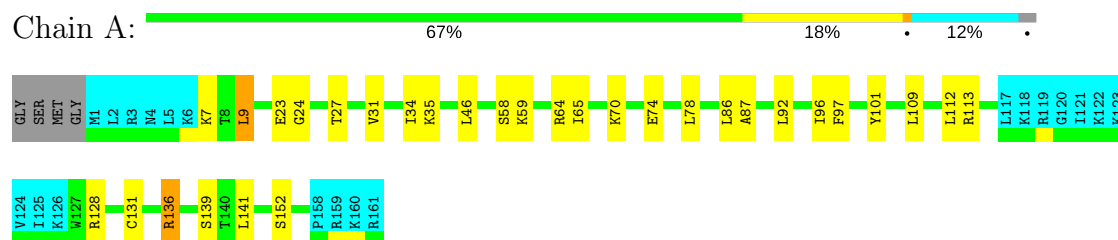
### 4.2.16 Score per residue for model 16

- Molecule 1: Putative toxin VapC6



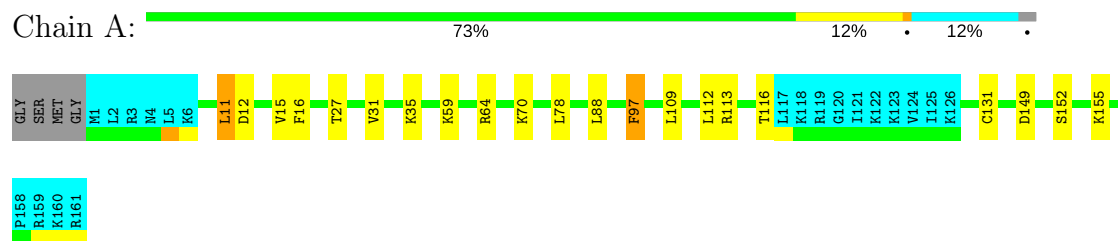
### 4.2.17 Score per residue for model 17

- Molecule 1: Putative toxin VapC6



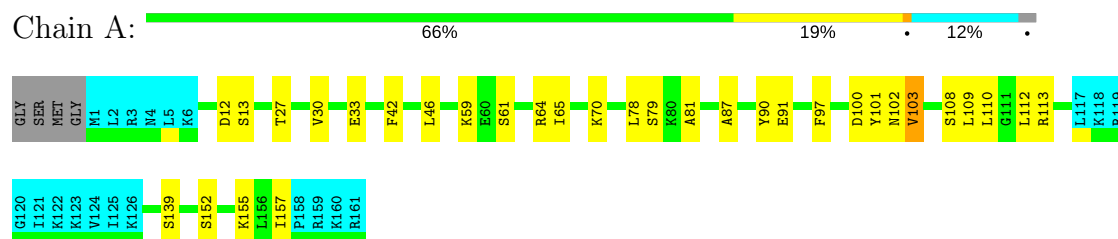
### 4.2.18 Score per residue for model 18

- Molecule 1: Putative toxin VapC6



### 4.2.19 Score per residue for model 19

- Molecule 1: Putative toxin VapC6



### 4.2.20 Score per residue for model 20

- Molecule 1: Putative toxin VapC6



## 5 Refinement protocol and experimental data overview

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

Of the 200 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
OPALp	refinement	
CYANA	structure solution	3.0
TALOS+	structure solution	
CANDID	structure solution	
ATNOS	refinement	
CANDID	refinement	
CYANA	refinement	3.0

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)	2lcq_cs.str
Number of chemical shift lists	1
Total number of shifts	1961
Number of shifts mapped to atoms	1961
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	89%

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

## 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	0.55±0.01	0±0/1105 (0.0±0.0%)	1.02±0.02	1±1/1493 (0.1±0.1%)
All	All	0.55	0/22100 (0.0%)	1.02	23/29860 (0.1%)

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	0.9±0.9
All	All	0	19

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	115	ARG	NE-CZ-NH2	-6.29	117.15	120.30	20	3
1	A	64	ARG	NE-CZ-NH1	6.26	123.43	120.30	6	1
1	A	114	PHE	CB-CG-CD2	-5.95	116.63	120.80	12	1
1	A	113	ARG	NE-CZ-NH2	-5.66	117.47	120.30	8	2
1	A	31	VAL	CA-CB-CG1	5.66	119.39	110.90	18	2
1	A	37	ARG	NE-CZ-NH2	-5.64	117.48	120.30	10	2
1	A	136	ARG	NE-CZ-NH1	5.63	123.11	120.30	17	1
1	A	64	ARG	NE-CZ-NH2	-5.60	117.50	120.30	4	2
1	A	128	ARG	NE-CZ-NH2	-5.59	117.50	120.30	17	1
1	A	97	PHE	CB-CG-CD2	-5.49	116.95	120.80	12	2
1	A	90	TYR	CB-CG-CD2	-5.21	117.87	121.00	16	2
1	A	136	ARG	NE-CZ-NH2	-5.17	117.71	120.30	13	2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	99	ASP	CB-CG-OD1	5.10	122.89	118.30	4	1
1	A	115	ARG	NE-CZ-NH1	5.09	122.85	120.30	3	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	101	TYR	Sidechain	7
1	A	113	ARG	Sidechain	3
1	A	64	ARG	Sidechain	3
1	A	136	ARG	Sidechain	2
1	A	115	ARG	Sidechain	2
1	A	25	TYR	Sidechain	1
1	A	37	ARG	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	1089	1119	1119	4±2
All	All	21800	22380	22380	88

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 2.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:68:VAL:HG11	1:A:110:LEU:HD21	0.82	1.50	11	3
1:A:11:LEU:HB3	1:A:15:VAL:HG11	0.74	1.59	18	1
1:A:65:ILE:HD11	1:A:87:ALA:HA	0.62	1.70	6	15
1:A:78:LEU:H	1:A:78:LEU:HD12	0.53	1.64	7	2
1:A:9:LEU:HD23	1:A:24:GLY:HA3	0.52	1.79	17	3
1:A:12:ASP:O	1:A:15:VAL:HG12	0.51	2.05	18	1
1:A:90:TYR:CE2	1:A:110:LEU:HG	0.50	2.41	8	13

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:42:PHE:CZ	1:A:46:LEU:HD11	0.50	2.42	6	3
1:A:9:LEU:HD22	1:A:23:GLU:O	0.49	2.07	20	1
1:A:9:LEU:CD2	1:A:24:GLY:HA3	0.49	2.38	20	1
1:A:9:LEU:HD21	1:A:11:LEU:HD11	0.47	1.85	7	2
1:A:65:ILE:HD12	1:A:90:TYR:CE2	0.47	2.45	10	7
1:A:112:LEU:HD13	1:A:113:ARG:N	0.46	2.25	12	1
1:A:100:ASP:H	1:A:103:VAL:HG12	0.46	1.71	10	4
1:A:42:PHE:CE2	1:A:46:LEU:HD11	0.46	2.45	6	2
1:A:51:LYS:HA	1:A:51:LYS:HE3	0.45	1.89	1	1
1:A:86:LEU:HD23	1:A:96:ILE:HD13	0.44	1.88	4	4
1:A:31:VAL:HG23	1:A:43:LEU:HD21	0.44	1.88	1	1
1:A:100:ASP:H	1:A:103:VAL:CG1	0.44	2.25	14	3
1:A:65:ILE:HD11	1:A:87:ALA:CA	0.43	2.44	14	3
1:A:13:SER:HB2	1:A:30:VAL:HG21	0.43	1.90	2	1
1:A:31:VAL:HG23	1:A:43:LEU:HD13	0.43	1.89	16	1
1:A:134:CYS:SG	1:A:150:CYS:SG	0.43	3.17	10	1
1:A:86:LEU:CD1	1:A:96:ILE:HD13	0.42	2.45	17	1
1:A:30:VAL:HG22	1:A:81:ALA:CB	0.42	2.45	20	2
1:A:78:LEU:HD12	1:A:78:LEU:H	0.41	1.75	13	2
1:A:15:VAL:HG13	1:A:16:PHE:CD1	0.41	2.50	18	1
1:A:30:VAL:HG22	1:A:81:ALA:HB1	0.41	1.92	6	1
1:A:69:ALA:HA	1:A:72:THR:HG22	0.41	1.91	11	1
1:A:31:VAL:HG23	1:A:43:LEU:CD2	0.41	2.46	1	1
1:A:80:LYS:HE3	1:A:81:ALA:N	0.41	2.31	14	1
1:A:78:LEU:N	1:A:78:LEU:HD12	0.40	2.31	19	1
1:A:127:TRP:CH2	1:A:157:ILE:HD13	0.40	2.52	20	1
1:A:78:LEU:HD12	1:A:78:LEU:N	0.40	2.31	10	1
1:A:15:VAL:HG13	1:A:16:PHE:CE1	0.40	2.52	18	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	141/165 (85%)	129±2 (92±1%)	11±2 (7±1%)	1±1 (1±1%)	27	73

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	2820/3300 (85%)	2587 (92%)	210 (7%)	23 (1%)	27	73

All 10 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	74	GLU	6
1	A	34	ILE	4
1	A	133	GLY	3
1	A	132	ILE	2
1	A	7	LYS	2
1	A	148	PRO	2
1	A	22	ILE	1
1	A	73	GLY	1
1	A	128	ARG	1
1	A	75	VAL	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	123/144 (85%)	103±3 (84±3%)	20±3 (16±3%)	6	43
All	All	2460/2880 (85%)	2063 (84%)	397 (16%)	6	43

All 69 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	109	LEU	20
1	A	112	LEU	20
1	A	70	LYS	20
1	A	64	ARG	20
1	A	97	PHE	19
1	A	35	LYS	18
1	A	113	ARG	16
1	A	27	THR	14
1	A	103	VAL	13

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	131	CYS	12
1	A	78	LEU	11
1	A	152	SER	11
1	A	59	LYS	10
1	A	155	LYS	10
1	A	46	LEU	9
1	A	65	ILE	9
1	A	149	ASP	9
1	A	88	LEU	9
1	A	141	LEU	7
1	A	82	ASP	6
1	A	93	LYS	6
1	A	91	GLU	6
1	A	13	SER	5
1	A	100	ASP	5
1	A	29	SER	5
1	A	63	ASP	5
1	A	108	SER	5
1	A	77	GLU	5
1	A	80	LYS	5
1	A	102	ASN	5
1	A	9	LEU	4
1	A	21	ASP	4
1	A	116	THR	4
1	A	139	SER	4
1	A	58	SER	4
1	A	51	LYS	4
1	A	157	ILE	4
1	A	12	ASP	4
1	A	92	LEU	3
1	A	33	GLU	3
1	A	99	ASP	3
1	A	7	LYS	3
1	A	61	SER	3
1	A	76	ASN	3
1	A	79	SER	2
1	A	14	SER	2
1	A	136	ARG	2
1	A	128	ARG	2
1	A	56	GLU	2
1	A	72	THR	2
1	A	53	LYS	2

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	39	SER	1
1	A	37	ARG	1
1	A	140	THR	1
1	A	132	ILE	1
1	A	74	GLU	1
1	A	71	GLU	1
1	A	23	GLU	1
1	A	30	VAL	1
1	A	45	SER	1
1	A	134	CYS	1
1	A	36	ASP	1
1	A	44	GLU	1
1	A	48	SER	1
1	A	60	GLU	1
1	A	11	LEU	1
1	A	22	ILE	1
1	A	32	GLU	1
1	A	150	CYS	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 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

## 6.7 Other polymers [i](#)

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 89% for the well-defined parts and 85% for the entire structure.

### 7.1 Chemical shift list 1

File name: 2lcq\_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	1961
Number of shifts mapped to atoms	1961
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

#### 7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	159	$-0.12 \pm 0.13$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}_\beta$	146	$0.02 \pm 0.09$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	158	$0.07 \pm 0.09$	None needed ( $< 0.5$ ppm)
$^{15}\text{N}$	152	$0.22 \pm 0.34$	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 89%, i.e. 1527 atoms were assigned a chemical shift out of a possible 1710. 23 out of 25 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	692/695 (100%)	276/277 (100%)	281/282 (100%)	135/136 (99%)
Sidechain	764/926 (83%)	456/539 (85%)	302/351 (86%)	6/36 (17%)

*Continued on next page...*

Continued from previous page...

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Aromatic	71/89 (80%)	47/47 (100%)	23/41 (56%)	1/1 (100%)
Overall	1527/1710 (89%)	779/863 (90%)	606/674 (90%)	142/173 (82%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 85%, i.e. 1736 atoms were assigned a chemical shift out of a possible 2033. 26 out of 29 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	780/793 (98%)	311/316 (98%)	317/322 (98%)	152/155 (98%)
Sidechain	885/1151 (77%)	525/674 (78%)	353/422 (84%)	7/55 (13%)
Aromatic	71/89 (80%)	47/47 (100%)	23/41 (56%)	1/1 (100%)
Overall	1736/2033 (85%)	883/1037 (85%)	693/785 (88%)	160/211 (76%)

#### 7.1.4 Statistically unusual chemical shifts [i](#)

There are no statistically unusual chemical shifts.

#### 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:

