



# Full wwPDB NMR Structure Validation Report ⓘ

May 28, 2020 – 10:30 pm BST

PDB ID : 2JZ4  
Title : Putative 32 kDa myrosinase binding protein At3g16450.1 from *Arabidopsis thaliana*  
Authors : Takeda, N.; Sugimori, N.; Torizawa, T.; Terauchi, T.; Ono, A.M.; Yagi, H.; Yamaguchi, Y.; Kato, K.; Ikeya, T.; Guntert, P.; Aceti, D.J.; Markley, J.L.; Kainosho, M.; Center for Eukaryotic Structural Genomics (CESG)  
Deposited on : 2007-12-28

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

We welcome your comments at [validation@mail.wwpdb.org](mailto: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 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)  
RCI : v\_1n\_11\_5\_13\_A (Berjanski et al., 2005)  
PANAV : Wang et al. (2010)  
ShiftChecker : 2.11  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11

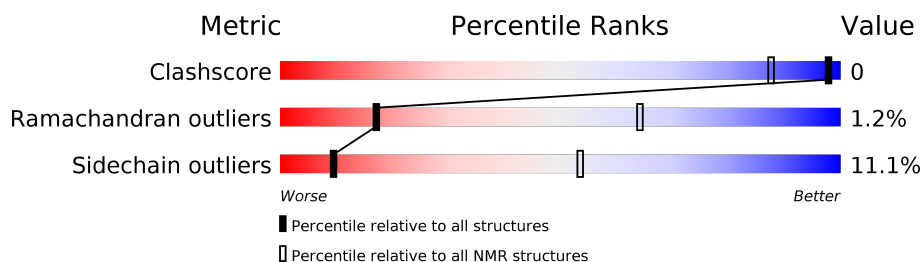
# 1 Overall quality at a glance

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	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  $\geq 3$ , 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	299	 93% • 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:2-A:18, A:22-A:144 (140)	0.98	3
2	A:153-A:297 (145)	0.75	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 6 clusters. No single-model clusters were found.

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

### 3 Entry composition [i](#)

There is only 1 type of molecule in this entry. The entry contains 4427 atoms, of which 2171 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Jasmonate inducible protein isolog.

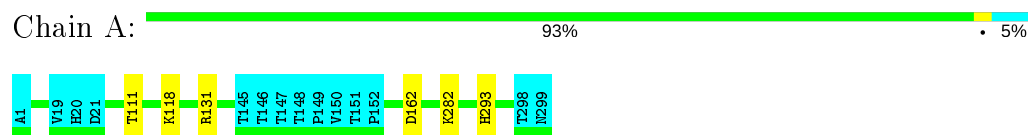
Mol	Chain	Residues	Atoms						Trace
1	A	299	Total	C	H	N	O	S	0
			4427	1433	2171	372	449	2	

## 4 Residue-property plots

### 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: Jasmonate inducible protein isolog

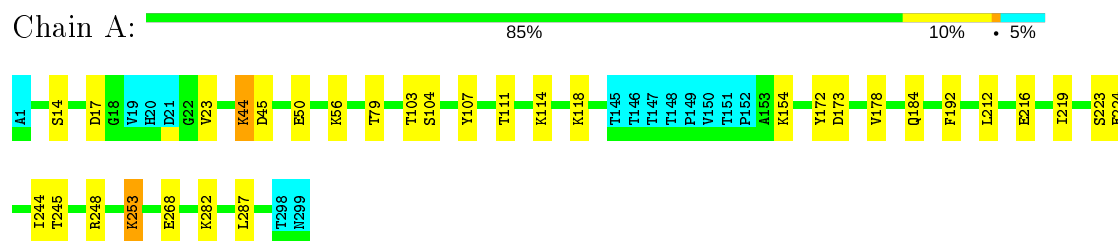


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

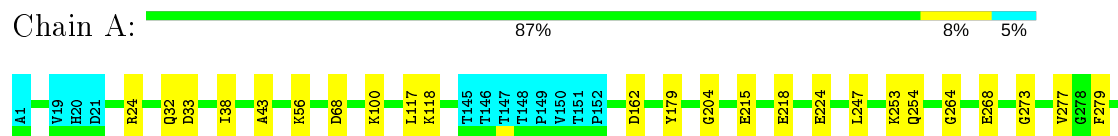
#### 4.2.1 Score per residue for model 1 (medoid)

- Molecule 1: Jasmonate inducible protein isolog



#### 4.2.2 Score per residue for model 2

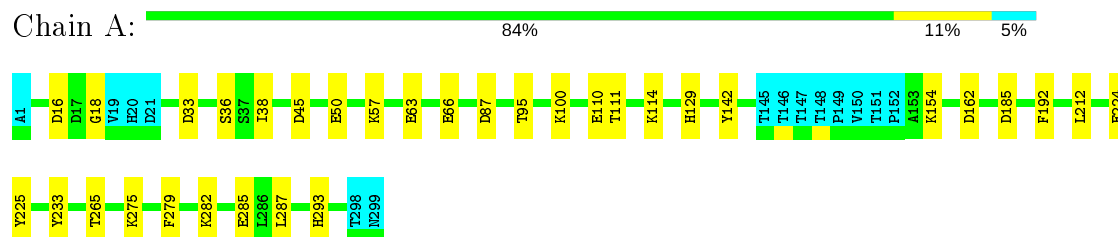
- Molecule 1: Jasmonate inducible protein isolog



T298  
I299

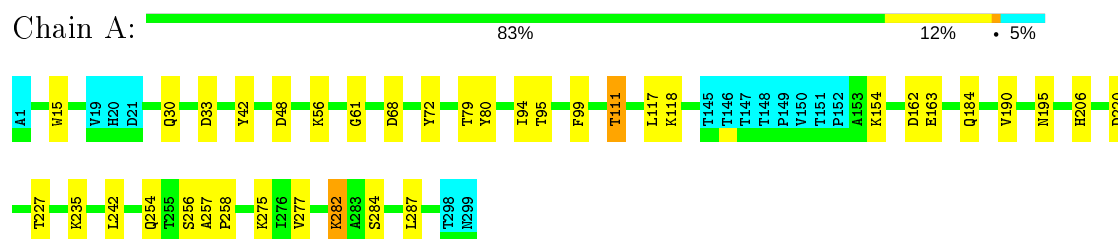
### 4.2.3 Score per residue for model 3

- Molecule 1: Jasmonate inducible protein isolog



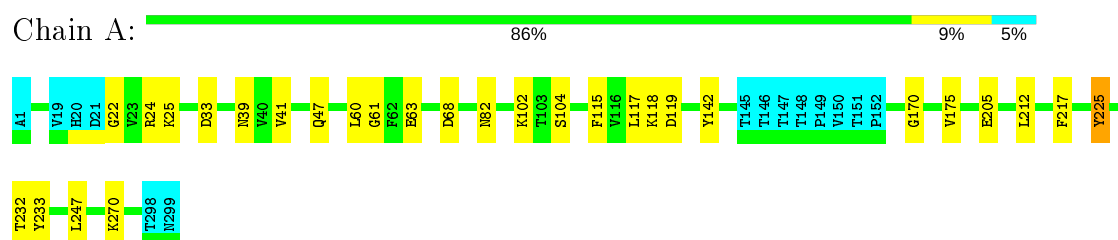
### 4.2.4 Score per residue for model 4

- Molecule 1: Jasmonate inducible protein isolog



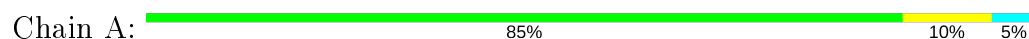
### 4.2.5 Score per residue for model 5

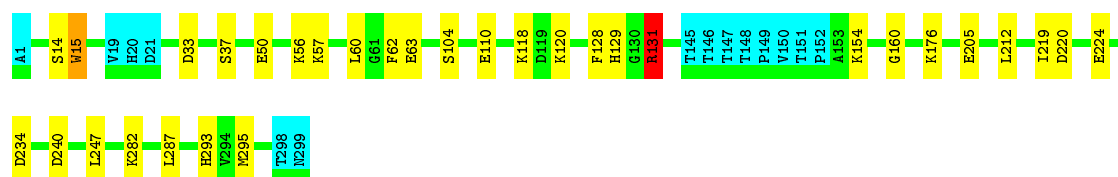
- Molecule 1: Jasmonate inducible protein isolog



### 4.2.6 Score per residue for model 6

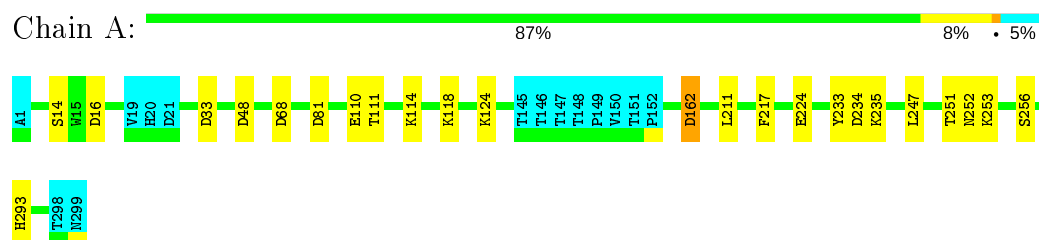
- Molecule 1: Jasmonate inducible protein isolog





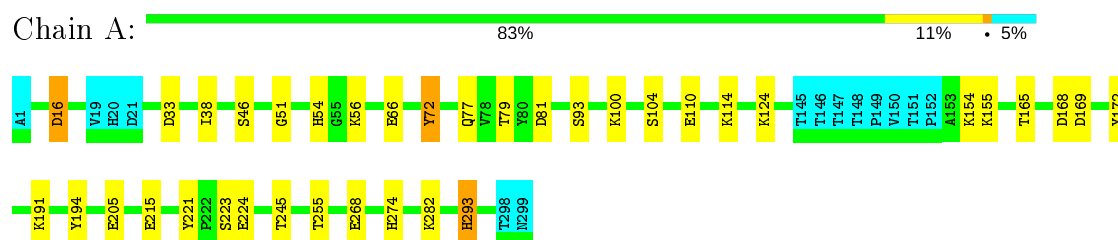
#### 4.2.7 Score per residue for model 7

- Molecule 1: Jasmonate inducible protein isolog



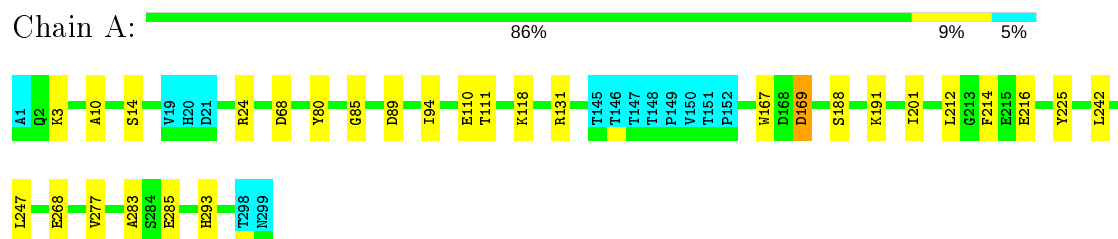
#### 4.2.8 Score per residue for model 8

- Molecule 1: Jasmonate inducible protein isolog



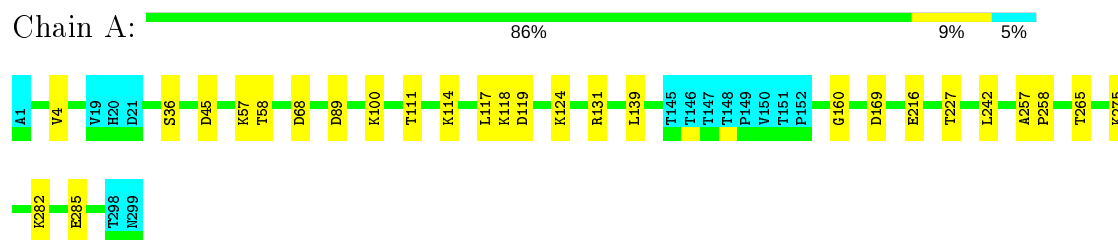
#### 4.2.9 Score per residue for model 9

- Molecule 1: Jasmonate inducible protein isolog



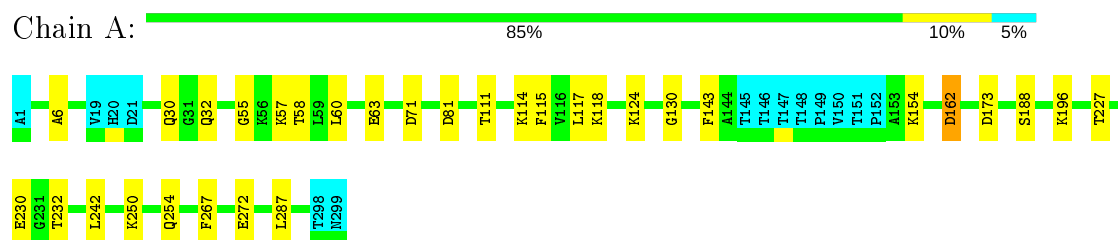
#### 4.2.10 Score per residue for model 10

- Molecule 1: Jasmonate inducible protein isolog



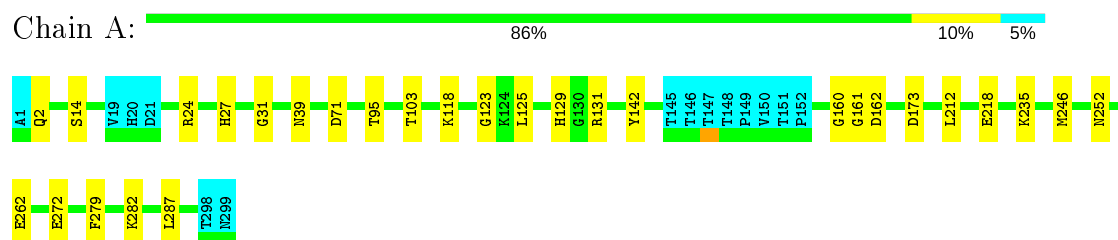
#### 4.2.11 Score per residue for model 11

- Molecule 1: Jasmonate inducible protein isolog



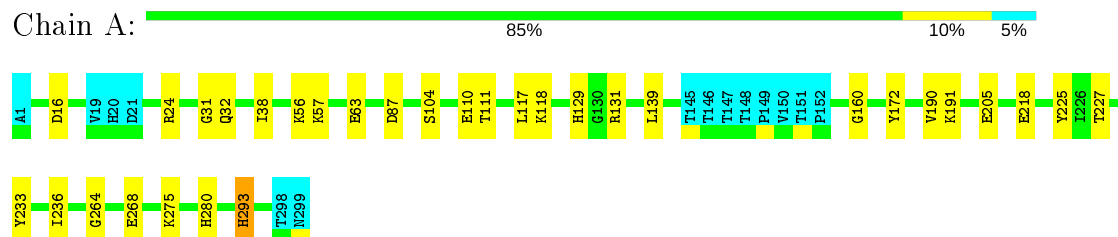
#### 4.2.12 Score per residue for model 12

- Molecule 1: Jasmonate inducible protein isolog



#### 4.2.13 Score per residue for model 13

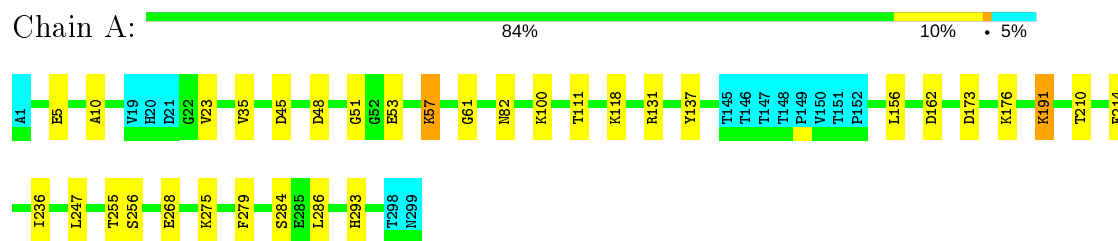
- Molecule 1: Jasmonate inducible protein isolog





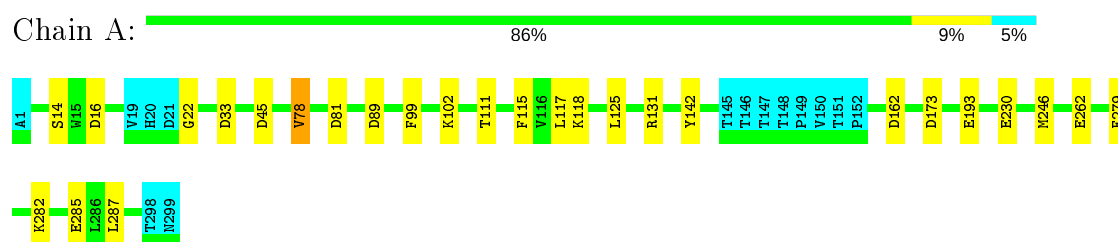
#### 4.2.14 Score per residue for model 14

- Molecule 1: Jasmonate inducible protein isolog



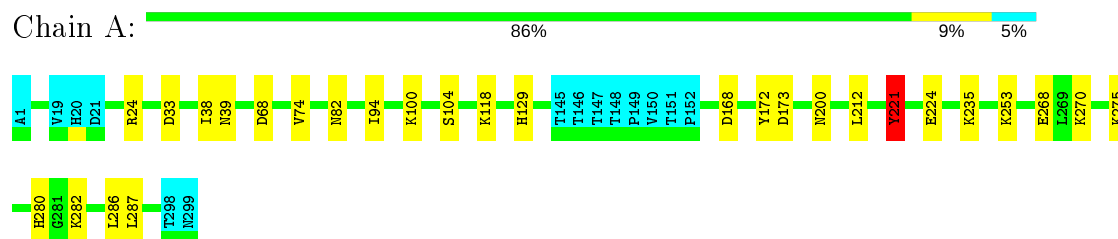
#### 4.2.15 Score per residue for model 15

- Molecule 1: Jasmonate inducible protein isolog



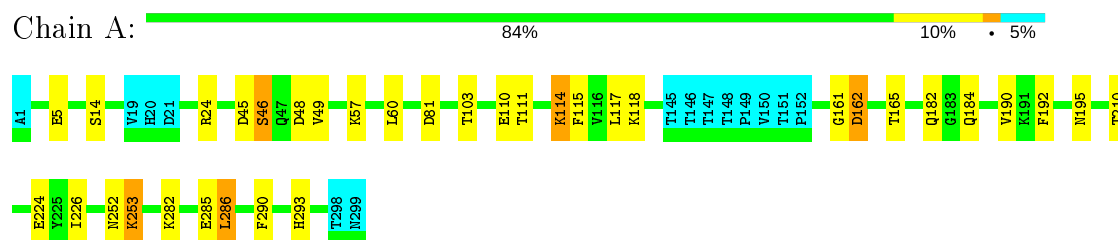
#### 4.2.16 Score per residue for model 16

- Molecule 1: Jasmonate inducible protein isolog



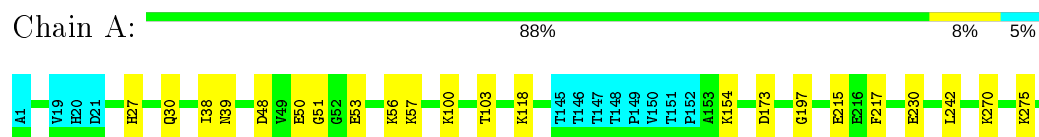
#### 4.2.17 Score per residue for model 17

- Molecule 1: Jasmonate inducible protein isolog



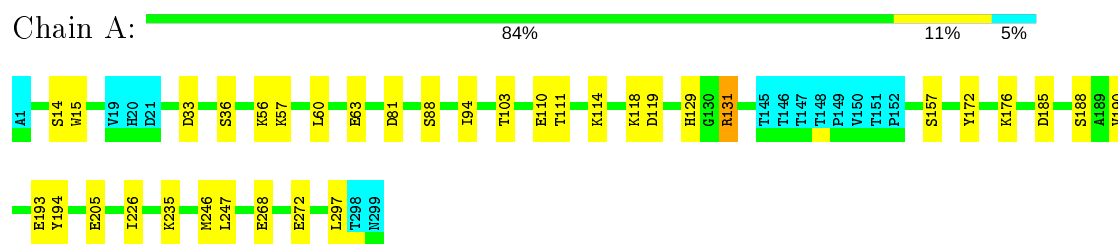
### 4.2.18 Score per residue for model 18

- Molecule 1: Jasmonate inducible protein isolog



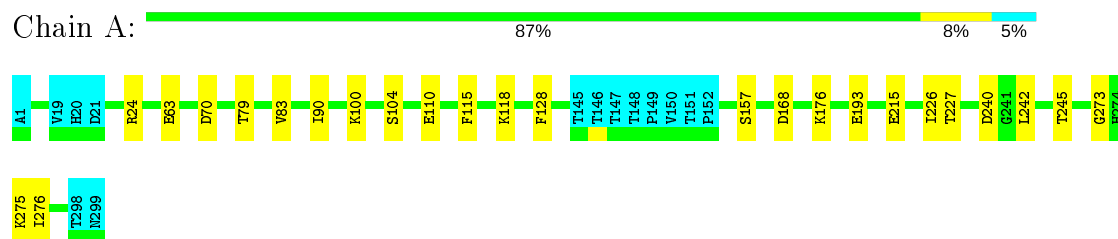
### 4.2.19 Score per residue for model 19

- Molecule 1: Jasmonate inducible protein isolog



### 4.2.20 Score per residue for model 20

- Molecule 1: Jasmonate inducible protein isolog



## 5 Refinement protocol and experimental data overview

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

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 solution	2.2
OPALp	refinement	1.4

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	2463
Number of shifts mapped to atoms	2463
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%

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

COVALENT-GEOMETRY INFOmissingINFO

### 5.1 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	2155	2076	2076	2±1
All	All	43100	41520	41520	34

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:167:TRP:HE1	1:A:283:ALA:HB2	0.61	1.55	9	1
1:A:162:ASP:C	1:A:282:LYS:HE3	0.55	2.22	4	1
1:A:191:LYS:HE2	1:A:214:PHE:CE2	0.54	2.37	14	1
1:A:79:THR:HG21	1:A:110:GLU:HA	0.53	1.80	20	1
1:A:191:LYS:HE2	1:A:214:PHE:CZ	0.50	2.41	14	1
1:A:162:ASP:HA	1:A:282:LYS:CE	0.49	2.38	7	1
1:A:79:THR:HG21	1:A:111:THR:H	0.49	1.66	1	2
1:A:178:VAL:HG13	1:A:192:PHE:CE2	0.48	2.44	1	1
1:A:257:ALA:HB1	1:A:258:PRO:HD2	0.48	1.86	10	2
1:A:191:LYS:HE3	1:A:214:PHE:CE2	0.47	2.45	9	1
1:A:115:PHE:CE2	1:A:117:LEU:HD22	0.46	2.45	5	2
1:A:191:LYS:CE	1:A:214:PHE:CZ	0.46	2.98	14	1
1:A:155:LYS:HE2	1:A:293:HIS:CE1	0.46	2.46	8	1
1:A:72:TYR:CG	1:A:124:LYS:HE3	0.45	2.47	8	1
1:A:78:VAL:HG13	1:A:115:PHE:CE1	0.45	2.47	15	1
1:A:110:GLU:OE1	1:A:114:LYS:NZ	0.43	2.51	19	1
1:A:226:ILE:HG22	1:A:276:ILE:HD12	0.43	1.91	20	1
1:A:57:LYS:HE3	1:A:62:PHE:CZ	0.43	2.49	6	1
1:A:15:TRP:CE3	1:A:129:HIS:CB	0.43	3.02	19	1
1:A:172:TYR:CD1	1:A:194:TYR:CE2	0.43	3.07	19	1
1:A:10:ALA:HB3	1:A:137:TYR:CG	0.42	2.49	14	1
1:A:117:LEU:N	1:A:117:LEU:HD23	0.42	2.30	11	1
1:A:191:LYS:HE3	1:A:214:PHE:CZ	0.41	2.50	9	1
1:A:162:ASP:HA	1:A:282:LYS:HE3	0.41	1.92	3	1
1:A:24:ARG:CG	1:A:43:ALA:HB2	0.41	2.46	2	1
1:A:25:LYS:HE3	1:A:41:VAL:HG21	0.41	1.92	5	1
1:A:15:TRP:CD1	1:A:131:ARG:HG3	0.41	2.51	6	1
1:A:280:HIS:CD2	1:A:293:HIS:HD1	0.41	2.34	13	1
1:A:53:GLU:OE2	1:A:57:LYS:HE2	0.41	2.16	14	1
1:A:27:HIS:CD2	1:A:39:ASN:HD22	0.40	2.35	18	1
1:A:25:LYS:HE3	1:A:41:VAL:CG2	0.40	2.47	5	1

## 5.2 Torsion angles ⓘ

### 5.2.1 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	285/299 (95%)	251±4 (88±1%)	31±3 (11±1%)	3±2 (1±1%)	17	64
All	All	5700/5980 (95%)	5023 (88%)	610 (11%)	67 (1%)	17	64

All 42 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	253	LYS	5
1	A	160	GLY	4
1	A	61	GLY	3
1	A	14	SER	3
1	A	273	GLY	3
1	A	51	GLY	3
1	A	162	ASP	3
1	A	31	GLY	2
1	A	284	SER	2
1	A	114	LYS	2
1	A	285	GLU	2
1	A	264	GLY	2
1	A	46	SER	2
1	A	161	GLY	2
1	A	22	GLY	2
1	A	45	ASP	1
1	A	169	ASP	1
1	A	232	THR	1
1	A	6	ALA	1
1	A	35	VAL	1
1	A	182	GLN	1
1	A	87	ASP	1
1	A	10	ALA	1
1	A	287	LEU	1
1	A	60	LEU	1
1	A	85	GLY	1
1	A	170	GLY	1
1	A	277	VAL	1
1	A	131	ARG	1
1	A	204	GLY	1
1	A	286	LEU	1
1	A	226	ILE	1
1	A	30	GLN	1
1	A	55	GLY	1
1	A	16	ASP	1
1	A	156	LEU	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	130	GLY	1
1	A	221	TYR	1
1	A	227	THR	1
1	A	18	GLY	1
1	A	197	GLY	1
1	A	123	GLY	1

### 5.2.2 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	218/231 (94%)	194±4 (89±2%)	24±4 (11±2%)	9	53
All	All	4360/4620 (94%)	3878 (89%)	482 (11%)	9	53

All 149 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	118	LYS	18
1	A	282	LYS	11
1	A	111	THR	11
1	A	33	ASP	10
1	A	293	HIS	8
1	A	100	LYS	8
1	A	275	LYS	8
1	A	268	GLU	8
1	A	56	LYS	8
1	A	224	GLU	8
1	A	57	LYS	8
1	A	104	SER	7
1	A	110	GLU	7
1	A	212	LEU	7
1	A	63	GLU	7
1	A	173	ASP	7
1	A	68	ASP	7
1	A	154	LYS	7
1	A	247	LEU	7

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	287	LEU	7
1	A	114	LYS	6
1	A	242	LEU	6
1	A	162	ASP	6
1	A	24	ARG	6
1	A	131	ARG	6
1	A	38	ILE	6
1	A	117	LEU	6
1	A	81	ASP	6
1	A	235	LYS	5
1	A	103	THR	5
1	A	279	PHE	5
1	A	45	ASP	5
1	A	205	GLU	5
1	A	14	SER	5
1	A	129	HIS	5
1	A	16	ASP	5
1	A	48	ASP	5
1	A	50	GLU	4
1	A	190	VAL	4
1	A	215	GLU	4
1	A	176	LYS	4
1	A	227	THR	4
1	A	94	ILE	4
1	A	60	LEU	4
1	A	256	SER	3
1	A	252	ASN	3
1	A	169	ASP	3
1	A	272	GLU	3
1	A	246	MET	3
1	A	191	LYS	3
1	A	233	TYR	3
1	A	193	GLU	3
1	A	168	ASP	3
1	A	82	ASN	3
1	A	245	THR	3
1	A	285	GLU	3
1	A	95	THR	3
1	A	39	ASN	3
1	A	218	GLU	3
1	A	124	LYS	3
1	A	216	GLU	3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	230	GLU	3
1	A	286	LEU	3
1	A	254	GLN	3
1	A	119	ASP	3
1	A	225	TYR	3
1	A	188	SER	3
1	A	89	ASP	3
1	A	270	LYS	3
1	A	217	PHE	3
1	A	184	GLN	3
1	A	36	SER	3
1	A	71	ASP	2
1	A	220	ASP	2
1	A	165	THR	2
1	A	262	GLU	2
1	A	58	THR	2
1	A	192	PHE	2
1	A	30	GLN	2
1	A	15	TRP	2
1	A	234	ASP	2
1	A	236	ILE	2
1	A	255	THR	2
1	A	23	VAL	2
1	A	5	GLU	2
1	A	128	PHE	2
1	A	210	THR	2
1	A	195	ASN	2
1	A	219	ILE	2
1	A	99	PHE	2
1	A	66	GLU	2
1	A	240	ASP	2
1	A	265	THR	2
1	A	32	GLN	2
1	A	253	LYS	2
1	A	139	LEU	2
1	A	115	PHE	2
1	A	125	LEU	2
1	A	277	VAL	2
1	A	185	ASP	2
1	A	102	LYS	2
1	A	223	SER	2
1	A	157	SER	2

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	274	HIS	1
1	A	2	GLN	1
1	A	280	HIS	1
1	A	267	PHE	1
1	A	79	THR	1
1	A	93	SER	1
1	A	172	TYR	1
1	A	70	ASP	1
1	A	46	SER	1
1	A	244	ILE	1
1	A	53	GLU	1
1	A	90	ILE	1
1	A	175	VAL	1
1	A	3	LYS	1
1	A	83	VAL	1
1	A	297	LEU	1
1	A	201	ILE	1
1	A	295	MET	1
1	A	194	TYR	1
1	A	196	LYS	1
1	A	290	PHE	1
1	A	4	VAL	1
1	A	120	LYS	1
1	A	232	THR	1
1	A	17	ASP	1
1	A	87	ASP	1
1	A	78	VAL	1
1	A	163	GLU	1
1	A	142	TYR	1
1	A	37	SER	1
1	A	211	LEU	1
1	A	44	LYS	1
1	A	251	THR	1
1	A	47	GLN	1
1	A	77	GLN	1
1	A	200	ASN	1
1	A	143	PHE	1
1	A	49	VAL	1
1	A	88	SER	1
1	A	221	TYR	1
1	A	226	ILE	1
1	A	206	HIS	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	27	HIS	1
1	A	74	VAL	1
1	A	250	LYS	1
1	A	54	HIS	1

### 5.2.3 RNA [i](#)

There are no RNA molecules in this entry.

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

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

### 5.4 Carbohydrates [i](#)

There are no carbohydrates in this entry.

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

There are no ligands in this entry.

### 5.6 Other polymers [i](#)

There are no such molecules in this entry.

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

There are no chain breaks in this entry.

## 6 Chemical shift validation [i](#)

The completeness of assignment taking into account all chemical shift lists is 66% for the well-defined parts and 65% for the entire structure.

### 6.1 Chemical shift list 1

File name: input\_cs.cif

Chemical shift list name: *assigned\_chem\_shift\_list\_1*

#### 6.1.1 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	2463
Number of shifts mapped to atoms	2463
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	11

#### 6.1.2 Chemical shift referencing [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	261	$0.01 \pm 0.10$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}_\beta$	240	$-0.09 \pm 0.08$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	0	—	None (insufficient data)
$^{15}\text{N}$	285	$0.13 \pm 0.42$	None needed ( $< 0.5$ ppm)

#### 6.1.3 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 66%, i.e. 2157 atoms were assigned a chemical shift out of a possible 3264. 0 out of 41 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	1085/1413 (77%)	557/564 (99%)	253/570 (44%)	275/279 (99%)
Sidechain	948/1522 (62%)	499/882 (57%)	435/588 (74%)	14/52 (27%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	124/329 (38%)	65/177 (37%)	58/142 (41%)	1/10 (10%)
Overall	2157/3264 (66%)	1121/1623 (69%)	746/1300 (57%)	290/341 (85%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 65%, i.e. 2230 atoms were assigned a chemical shift out of a possible 3407. 0 out of 43 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	1125/1479 (76%)	579/590 (98%)	261/598 (44%)	285/291 (98%)
Sidechain	981/1592 (62%)	516/921 (56%)	451/618 (73%)	14/53 (26%)
Aromatic	124/336 (37%)	65/181 (36%)	58/144 (40%)	1/11 (9%)
Overall	2230/3407 (65%)	1160/1692 (69%)	770/1360 (57%)	300/355 (85%)

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

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	191	LYS	HE2	0.89	3.87 – 1.97	-10.7
1	A	7	GLY	HA3	1.33	5.80 – 2.00	-6.8
1	A	3	LYS	HE2	1.71	3.87 – 1.97	-6.4
1	A	254	GLN	HE22	9.57	9.27 – 4.77	5.7
1	A	195	ASN	HB2	1.06	4.36 – 1.26	-5.6
1	A	78	VAL	HB	0.21	3.59 – 0.39	-5.6
1	A	3	LYS	HG3	-0.19	2.76 – -0.04	-5.5
1	A	155	LYS	HE2	1.87	3.87 – 1.97	-5.5
1	A	155	LYS	HG3	-0.15	2.76 – -0.04	-5.4
1	A	163	GLU	HB2	0.93	3.08 – 0.98	-5.3
1	A	191	LYS	HD3	0.43	2.75 – 0.45	-5.1

#### 6.1.5 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:

