



# Full wwPDB NMR Structure Validation Report ⓘ

Jul 11, 2016 – 11:04 AM EDT

PDB ID : 5JYN  
Title : Structure of the transmembrane domain of HIV-1 gp41 in bicelle  
Authors : Dev, J.; Fu, Q.; Park, D.; Chen, B.; Chou, J.J.  
Deposited on : 2016-05-14

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  
Mogul : unknown  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
RCI : v\_1n\_11\_5\_13\_A (Berjanski et al., 2005)  
PANAV : Wang et al. (2010)  
ShiftChecker : rb-20027790  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027790

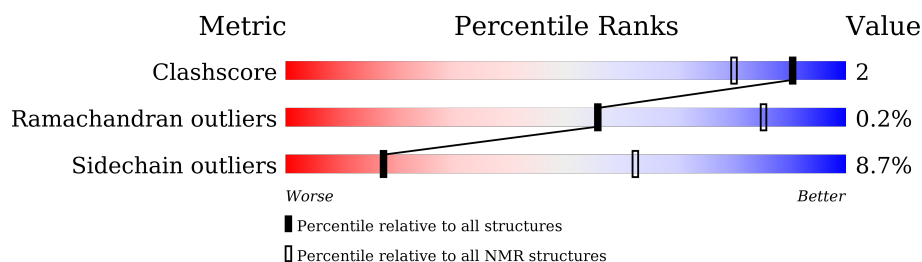
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 10%.

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	40	
1	B	40	
1	C	40	

## 2 Ensemble composition and analysis

This entry contains 15 models. Model 12 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

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:679-A:714, B:679-B:714, C:679-C:715 (109)	0.63	12

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 2 clusters and 2 single-model clusters were found.

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

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 2070 atoms, of which 1077 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Envelope glycoprotein gp160.

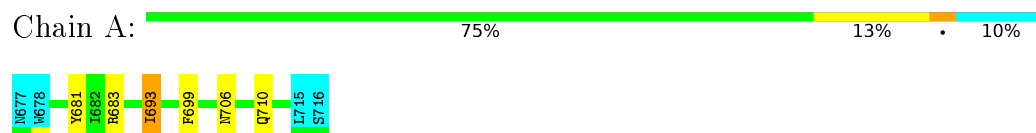
Mol	Chain	Residues	Atoms					Trace
1	A	40	Total	C	H	N	O	0
			690	224	359	57	50	
1	B	40	Total	C	H	N	O	0
			690	224	359	57	50	
1	C	40	Total	C	H	N	O	0
			690	224	359	57	50	

## 4 Residue-property plots [i](#)

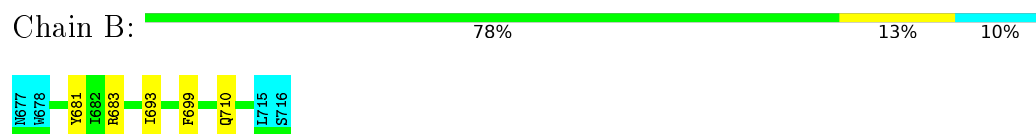
### 4.1 Average score per residue in the NMR ensemble

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.

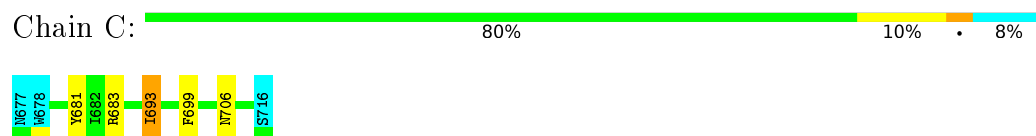
- Molecule 1: Envelope glycoprotein gp160



- Molecule 1: Envelope glycoprotein gp160



- Molecule 1: Envelope glycoprotein gp160

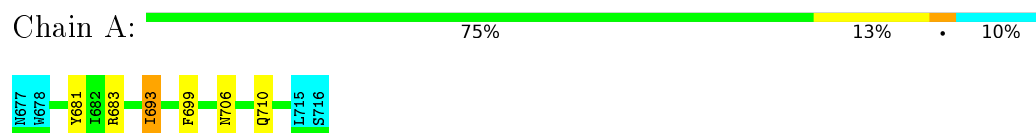


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

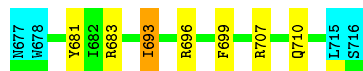
#### 4.2.1 Score per residue for model 1

- Molecule 1: Envelope glycoprotein gp160



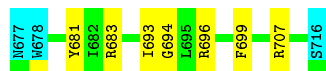
- Molecule 1: Envelope glycoprotein gp160

Chain B:  73% 15% 10%



- Molecule 1: Envelope glycoprotein gp160

Chain C:  75% 18% 8%



#### 4.2.2 Score per residue for model 2

- Molecule 1: Envelope glycoprotein gp160

Chain A:  65% 25% 10%




- Molecule 1: Envelope glycoprotein gp160

Chain B:  73% 18% 10%




- Molecule 1: Envelope glycoprotein gp160

Chain C:  80% 10% 8%



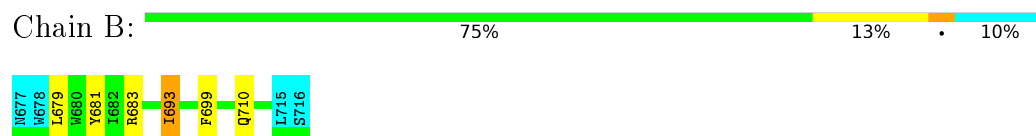
#### 4.2.3 Score per residue for model 3

- Molecule 1: Envelope glycoprotein gp160

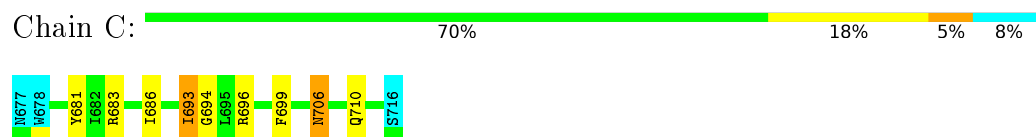
Chain A:  78% 13% 10%



- Molecule 1: Envelope glycoprotein gp160

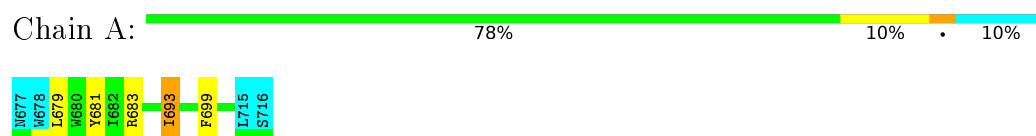


- Molecule 1: Envelope glycoprotein gp160

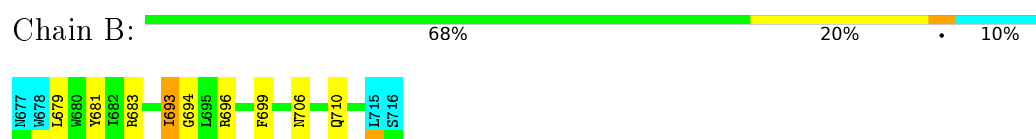


#### 4.2.4 Score per residue for model 4

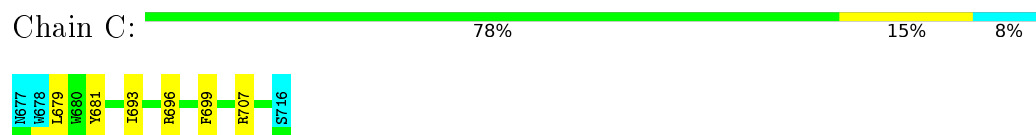
- Molecule 1: Envelope glycoprotein gp160



- Molecule 1: Envelope glycoprotein gp160

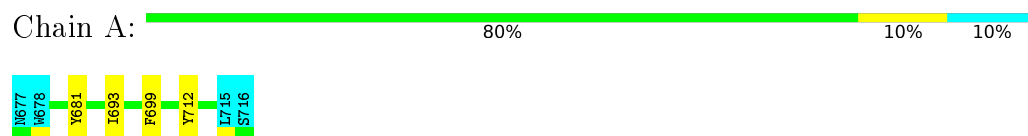


- Molecule 1: Envelope glycoprotein gp160



#### 4.2.5 Score per residue for model 5

- Molecule 1: Envelope glycoprotein gp160



- Molecule 1: Envelope glycoprotein gp160





- Molecule 1: Envelope glycoprotein gp160

Chain C: 78% 13% 8%



#### 4.2.6 Score per residue for model 6

- Molecule 1: Envelope glycoprotein gp160

Chain A: 78% 13% 10%



- Molecule 1: Envelope glycoprotein gp160

Chain B: 78% 13% 10%



- Molecule 1: Envelope glycoprotein gp160

Chain C: 80% 13% 8%



#### 4.2.7 Score per residue for model 7

- Molecule 1: Envelope glycoprotein gp160

Chain A: 78% 10% 10%



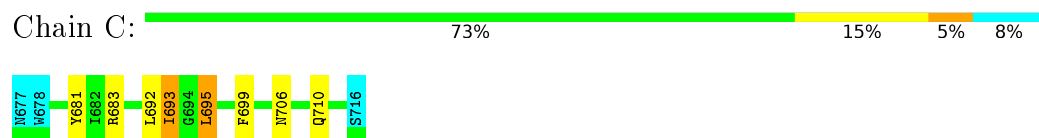
- Molecule 1: Envelope glycoprotein gp160

Chain B: 75% 13% 10%



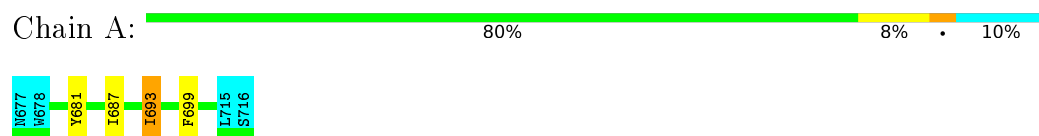


- Molecule 1: Envelope glycoprotein gp160

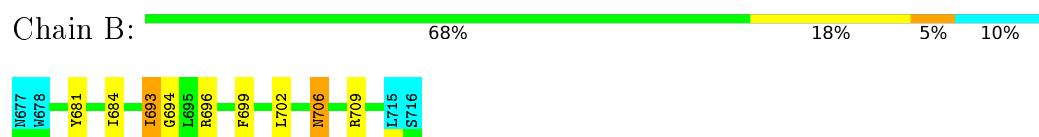


#### 4.2.8 Score per residue for model 8

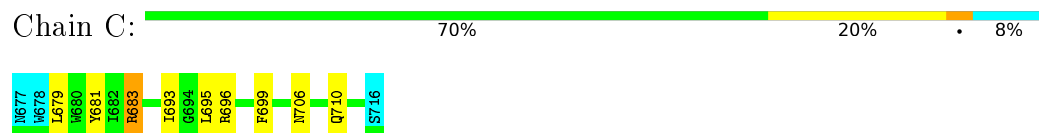
- Molecule 1: Envelope glycoprotein gp160



- Molecule 1: Envelope glycoprotein gp160

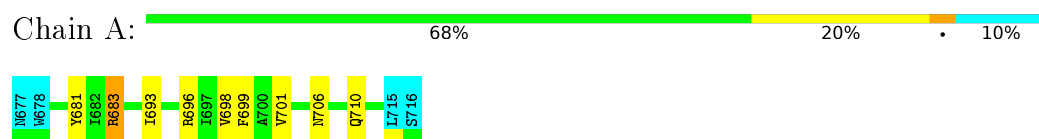


- Molecule 1: Envelope glycoprotein gp160

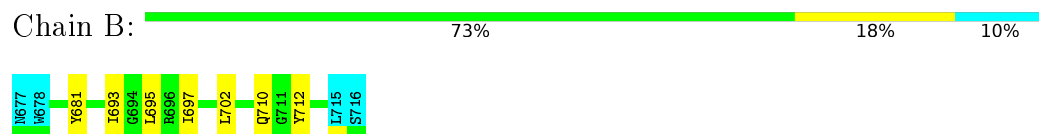


#### 4.2.9 Score per residue for model 9

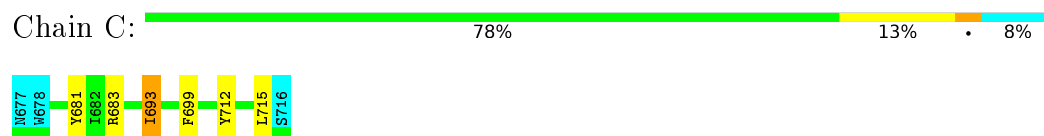
- Molecule 1: Envelope glycoprotein gp160



- Molecule 1: Envelope glycoprotein gp160

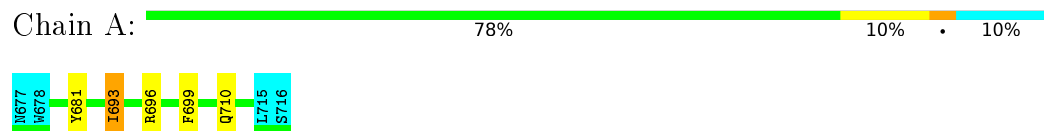


- Molecule 1: Envelope glycoprotein gp160

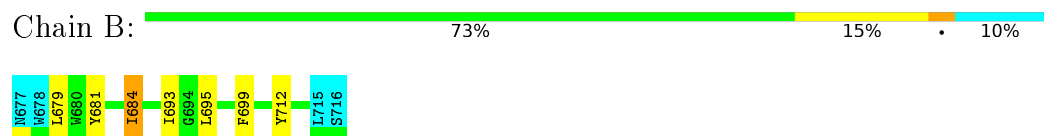


#### 4.2.10 Score per residue for model 10

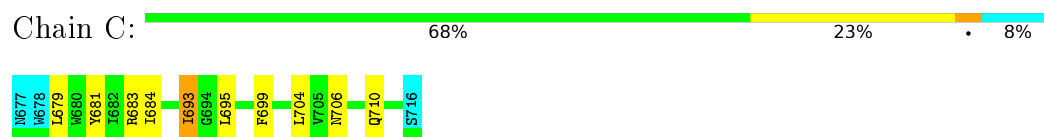
- Molecule 1: Envelope glycoprotein gp160



- Molecule 1: Envelope glycoprotein gp160

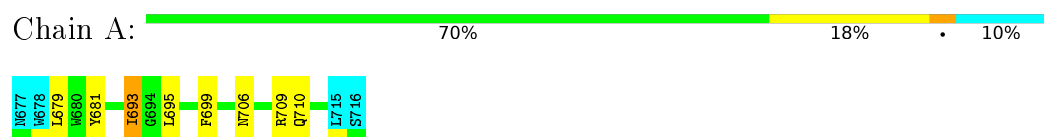


- Molecule 1: Envelope glycoprotein gp160

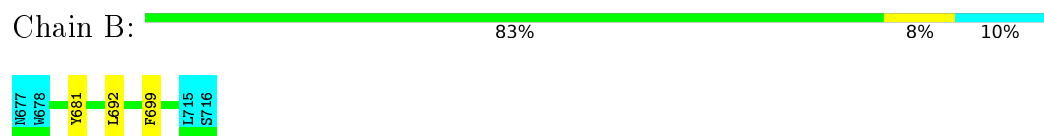


#### 4.2.11 Score per residue for model 11

- Molecule 1: Envelope glycoprotein gp160



- Molecule 1: Envelope glycoprotein gp160



- Molecule 1: Envelope glycoprotein gp160





#### 4.2.12 Score per residue for model 12 (medoid)

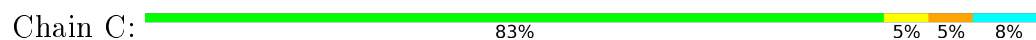
- Molecule 1: Envelope glycoprotein gp160



- Molecule 1: Envelope glycoprotein gp160

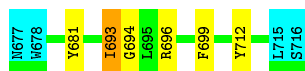


- Molecule 1: Envelope glycoprotein gp160

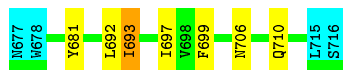


#### 4.2.13 Score per residue for model 13

- Molecule 1: Envelope glycoprotein gp160



- Molecule 1: Envelope glycoprotein gp160

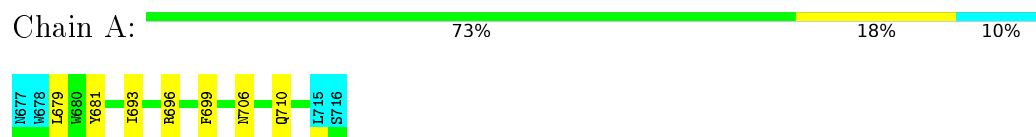


- Molecule 1: Envelope glycoprotein gp160

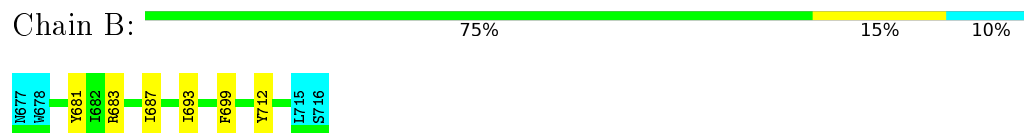


#### 4.2.14 Score per residue for model 14

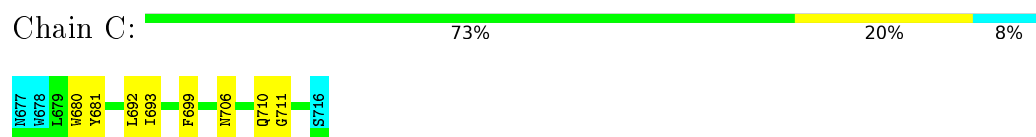
- Molecule 1: Envelope glycoprotein gp160



- Molecule 1: Envelope glycoprotein gp160

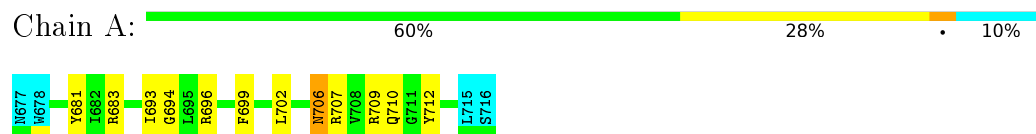


- Molecule 1: Envelope glycoprotein gp160

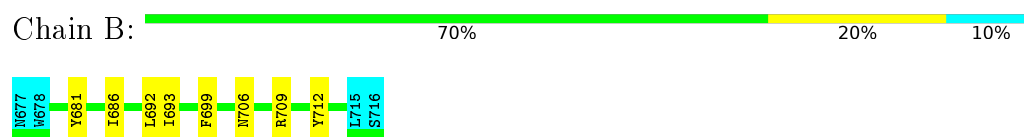


#### 4.2.15 Score per residue for model 15

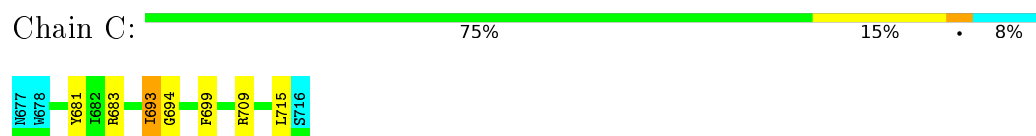
- Molecule 1: Envelope glycoprotein gp160



- Molecule 1: Envelope glycoprotein gp160



- Molecule 1: Envelope glycoprotein gp160



## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *simulated annealing*.

Of the 150 calculated structures, 15 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	Classification	Version
XPLOR-NIH	refinement	2.42

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)	5jyn_cs.cif
Number of chemical shift lists	1
Total number of shifts	149
Number of shifts mapped to atoms	149
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	10%

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

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	2.43±0.01	3±1/301 (0.9±0.3%)	0.92±0.04	0±0/410 (0.0±0.1%)
1	B	2.44±0.02	3±0/301 (0.8±0.2%)	0.89±0.03	0±0/410 (0.0±0.0%)
1	C	2.43±0.02	3±1/309 (0.8±0.2%)	0.91±0.04	0±1/421 (0.1±0.2%)
All	All	2.43	117/13665 (0.9%)	0.91	6/18615 (0.0%)

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	0.1±0.2
All	All	0	1

All unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
1	C	681	TYR	CB-CG	6.26	1.61	1.51	12	15
1	A	699	PHE	CB-CG	6.07	1.61	1.51	10	14
1	A	681	TYR	CB-CG	6.02	1.60	1.51	8	15
1	B	681	TYR	CB-CG	6.00	1.60	1.51	5	15
1	B	699	PHE	CB-CG	5.94	1.61	1.51	1	14
1	C	699	PHE	CB-CG	5.80	1.61	1.51	15	15
1	A	712	TYR	CB-CG	5.74	1.60	1.51	3	3
1	B	712	TYR	CB-CG	5.47	1.59	1.51	10	6
1	A	687	ILE	N-CA	5.39	1.57	1.46	12	1
1	C	712	TYR	CB-CG	5.34	1.59	1.51	2	4
1	A	694	GLY	N-CA	5.24	1.53	1.46	13	2
1	A	711	GLY	N-CA	5.21	1.53	1.46	12	1
1	A	710	GLN	N-CA	5.17	1.56	1.46	11	1
1	A	683	ARG	N-CA	5.16	1.56	1.46	9	2

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	C	693	ILE	N-CA	5.12	1.56	1.46	3	1
1	B	694	GLY	N-CA	5.12	1.53	1.46	4	2
1	C	711	GLY	N-CA	5.12	1.53	1.46	14	1
1	B	710	GLN	N-CA	5.09	1.56	1.46	7	1
1	A	693	ILE	CA-CB	5.07	1.66	1.54	7	1
1	C	694	GLY	N-CA	5.05	1.53	1.46	3	3

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	C	712	TYR	CB-CG-CD2	6.64	124.98	121.00	12	2
1	C	712	TYR	CB-CG-CD1	6.45	124.87	121.00	5	2
1	A	712	TYR	CB-CG-CD2	5.61	124.37	121.00	15	1
1	A	712	TYR	CB-CG-CD1	-5.59	117.65	121.00	15	1

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	696	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	294	325	325	2±1
1	B	294	325	325	2±1
1	C	302	336	336	2±1
All	All	13350	14790	14790	59

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:693:ILE:HD11	1:B:696:ARG:HE	0.61	1.55	8	5
1:A:696:ARG:HE	1:C:693:ILE:HD11	0.61	1.54	2	5
1:B:693:ILE:HD11	1:C:696:ARG:HE	0.57	1.59	4	6
1:A:706:ASN:O	1:A:710:GLN:N	0.56	2.39	12	6
1:C:706:ASN:O	1:C:710:GLN:N	0.53	2.42	10	7
1:B:706:ASN:O	1:B:710:GLN:N	0.52	2.42	4	3
1:B:693:ILE:HD12	1:B:693:ILE:O	0.50	2.07	13	1
1:C:695:LEU:O	1:C:695:LEU:HD22	0.50	2.06	7	1
1:A:706:ASN:HD22	1:A:709:ARG:HE	0.46	1.53	11	1
1:B:682:ILE:HD12	1:B:683:ARG:N	0.45	2.27	7	2
1:B:684:ILE:HD13	1:B:684:ILE:O	0.45	2.12	10	1
1:A:693:ILE:HD12	1:A:693:ILE:O	0.44	2.13	7	2
1:C:693:ILE:O	1:C:693:ILE:HD12	0.44	2.13	13	1
1:B:679:LEU:O	1:B:683:ARG:N	0.44	2.50	3	1
1:C:693:ILE:O	1:C:693:ILE:HD13	0.44	2.13	12	1
1:C:693:ILE:HD13	1:C:693:ILE:O	0.44	2.12	5	2
1:A:693:ILE:HD13	1:A:693:ILE:O	0.43	2.13	11	2
1:B:693:ILE:O	1:B:693:ILE:HD13	0.42	2.14	5	1
1:C:693:ILE:HD12	1:C:693:ILE:O	0.42	2.14	15	2
1:B:706:ASN:O	1:B:709:ARG:N	0.42	2.52	8	2
1:B:686:ILE:HD12	1:B:686:ILE:N	0.42	2.30	15	1
1:A:679:LEU:O	1:A:683:ARG:N	0.41	2.53	2	1
1:C:679:LEU:O	1:C:683:ARG:N	0.41	2.53	13	2
1:A:706:ASN:HD22	1:A:709:ARG:NE	0.41	2.14	11	1
1:B:686:ILE:N	1:B:686:ILE:HD12	0.41	2.31	5	1
1:B:687:ILE:N	1:B:687:ILE:HD12	0.40	2.30	14	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	36/40 (90%)	36±1 (99±2%)	0±1 (1±2%)	0±0 (0±0%)	100	100
1	B	36/40 (90%)	36±0 (100±1%)	0±0 (0±1%)	0±0 (0±0%)	100	100
1	C	37/40 (92%)	36±1 (98±2%)	1±1 (2±2%)	0±0 (1±1%)	38	79

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1635/1800 (91%)	1618 (99%)	14 (1%)	3 (0%)	56 85

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	C	715	LEU	3

### 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	32/36 (89%)	29±1 (92±4%)	3±1 (8±4%)	20 65
1	B	32/36 (89%)	29±1 (91±4%)	3±1 (9±4%)	15 60
1	C	33/36 (92%)	30±1 (91±3%)	3±1 (9±3%)	17 62
All	All	1455/1620 (90%)	1328 (91%)	127 (9%)	17 62

All 43 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	B	693	ILE	14
1	C	693	ILE	14
1	A	693	ILE	13
1	C	683	ARG	9
1	B	683	ARG	8
1	A	683	ARG	6
1	A	679	LEU	4
1	B	710	GLN	4
1	B	684	ILE	3
1	B	679	LEU	3
1	B	692	LEU	3
1	C	695	LEU	3
1	C	679	LEU	3
1	B	697	ILE	2
1	C	680	TRP	2
1	A	702	LEU	2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	C	707	ARG	2
1	B	695	LEU	2
1	A	709	ARG	2
1	B	702	LEU	2
1	C	692	LEU	2
1	A	707	ARG	2
1	C	709	ARG	2
1	A	695	LEU	1
1	C	702	LEU	1
1	B	681	TYR	1
1	A	687	ILE	1
1	A	706	ASN	1
1	C	706	ASN	1
1	A	710	GLN	1
1	C	684	ILE	1
1	C	704	LEU	1
1	B	707	ARG	1
1	A	701	VAL	1
1	B	706	ASN	1
1	A	699	PHE	1
1	A	680	TRP	1
1	A	692	LEU	1
1	A	698	VAL	1
1	C	686	ILE	1
1	C	715	LEU	1
1	C	681	TYR	1
1	B	685	PHE	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 carbohydrates in this entry.

## 6.6 Ligand geometry

There are no ligands in this entry.

## 6.7 Other polymers

There are no such molecules in this entry.

## 6.8 Polymer linkage issues

There are no chain breaks in this entry.

## 7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 10% for the well-defined parts and 9% for the entire structure.

### 7.1 Chemical shift list 1

File name: 5jyn\_cs.cif

Chemical shift list name: *CS.bmrB*

#### 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	149
Number of shifts mapped to atoms	149
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
$^{13}\text{C}_\alpha$	36	$0.28 \pm 0.15$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}_\beta$	0	—	—
$^{13}\text{C}'$	39	$0.31 \pm 0.13$	None needed ( $< 0.5$ ppm)
$^{15}\text{N}$	37	$0.18 \pm 0.14$	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 10%, i.e. 141 atoms were assigned a chemical shift out of a possible 1472. 0 out of 31 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	141/539 (26%)	35/215 (16%)	71/218 (33%)	35/106 (33%)
Sidechain	0/795 (0%)	0/461 (0%)	0/292 (0%)	0/42 (0%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	0/138 (0%)	0/72 (0%)	0/63 (0%)	0/3 (0%)
Overall	141/1472 (10%)	35/748 (5%)	71/573 (12%)	35/151 (23%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 9%, i.e. 149 atoms were assigned a chemical shift out of a possible 1620. 0 out of 33 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	149/594 (25%)	37/237 (16%)	75/240 (31%)	37/117 (32%)
Sidechain	0/852 (0%)	0/495 (0%)	0/312 (0%)	0/45 (0%)
Aromatic	0/174 (0%)	0/90 (0%)	0/78 (0%)	0/6 (0%)
Overall	149/1620 (9%)	37/822 (5%)	75/630 (12%)	37/168 (22%)

#### 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:

