



# Full wwPDB NMR Structure Validation Report ⓘ

Feb 12, 2017 – 10:37 pm GMT

PDB ID : 2K2I  
Title : NMR Solution structure of the C-terminal domain (T94-Y172) of the human centrin 2 in complex with a repeat sequence of human Sfi1 (R641-T660)  
Authors : Martinez-Sanz, J.; Assairi, L.; Blouquit, Y.; Duchambon, P.; Mouawad, L.; Craescu, C.  
Deposited on : 2008-04-02

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

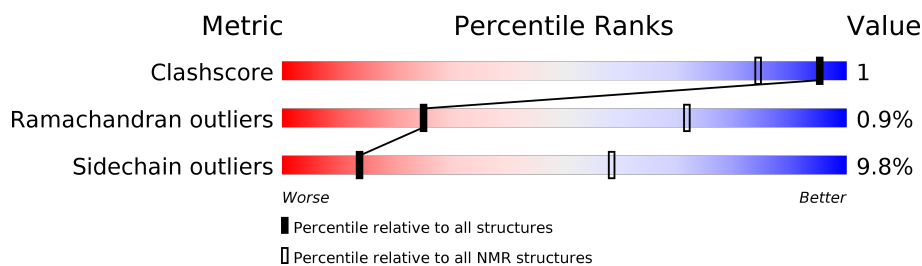
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment was not calculated.

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	79	 75% 6% • 18%
2	B	20	 45% 5% 50%

## 2 Ensemble composition and analysis

This entry contains 15 models. Model 3 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:103-A:167, B:651-B:660 (75)	0.58	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 1 clusters and 1 single-model cluster was found.

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

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

There are 2 unique types of molecules in this entry. The entry contains 1612 atoms, of which 798 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Centrin-2.

Mol	Chain	Residues	Atoms						Trace
1	A	79	Total	C	H	N	O	S	0
			1281	399	636	104	139	3	

- Molecule 2 is a protein called SFI1 peptide.

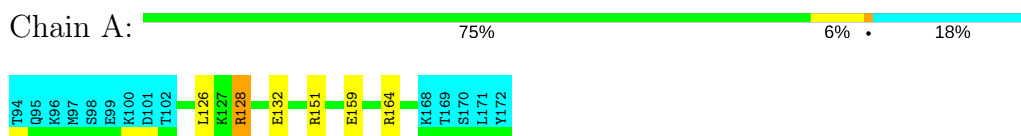
Mol	Chain	Residues	Atoms					Trace
2	B	20	Total	C	H	N	O	0
			331	105	162	37	27	

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

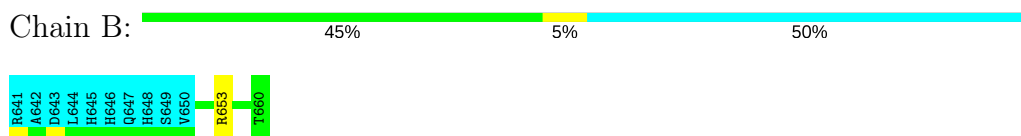
### 4.1 Average score per residue in the NMR ensemble

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: Centrin-2



- Molecule 2: SFI1 peptide

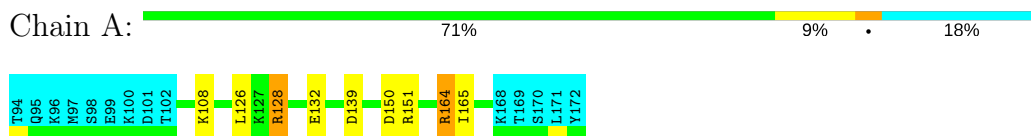


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

#### 4.2.1 Score per residue for model 1

- Molecule 1: Centrin-2

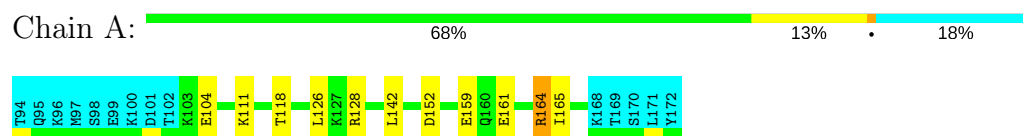


- Molecule 2: SFI1 peptide

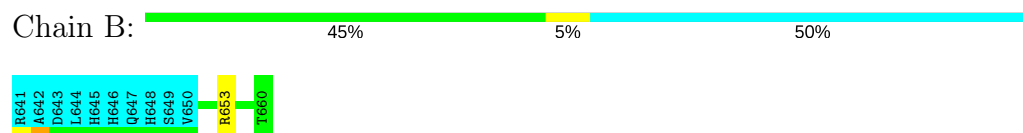


### 4.2.2 Score per residue for model 2

- Molecule 1: Centrin-2

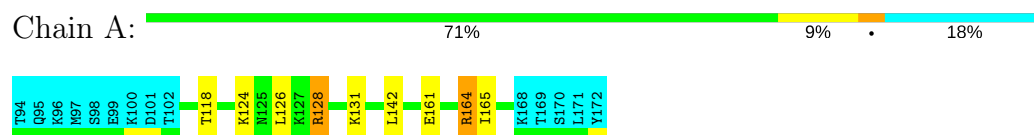


- Molecule 2: SFI1 peptide

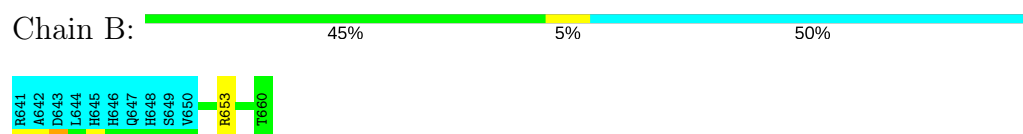


### 4.2.3 Score per residue for model 3 (medoid)

- Molecule 1: Centrin-2

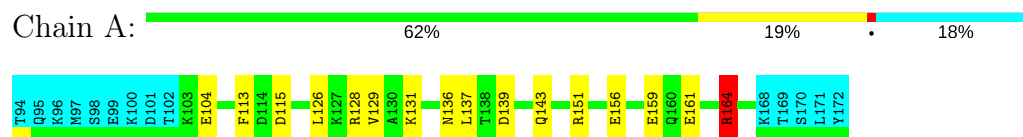


- Molecule 2: SFI1 peptide

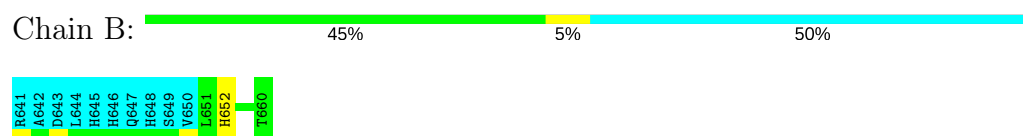


### 4.2.4 Score per residue for model 4

- Molecule 1: Centrin-2

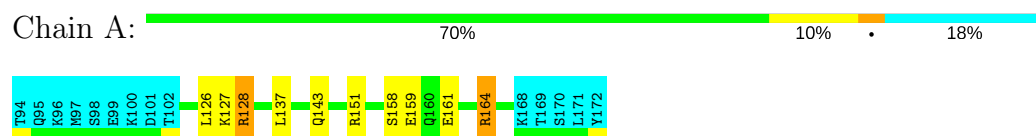


- Molecule 2: SFI1 peptide



### 4.2.5 Score per residue for model 5

- Molecule 1: Centrin-2

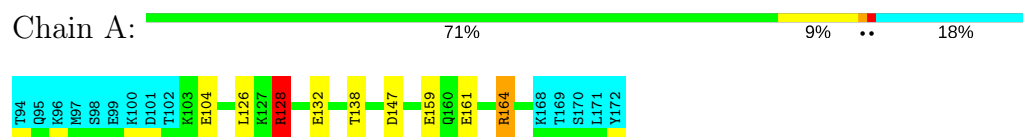


- Molecule 2: SFI1 peptide

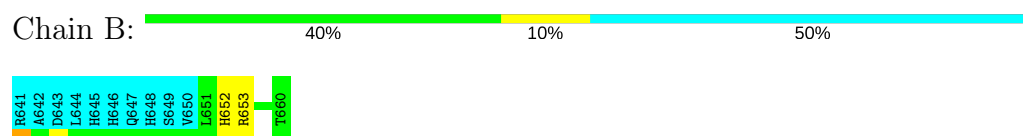


### 4.2.6 Score per residue for model 6

- Molecule 1: Centrin-2

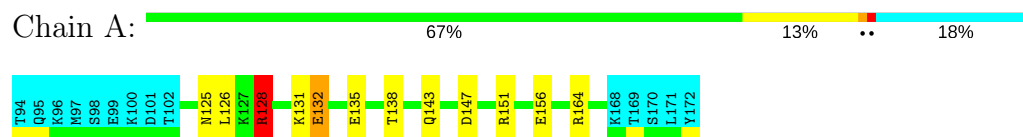


- Molecule 2: SFI1 peptide

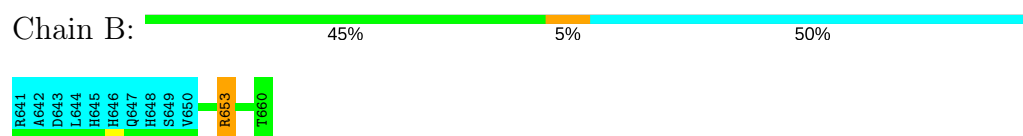


### 4.2.7 Score per residue for model 7

- Molecule 1: Centrin-2

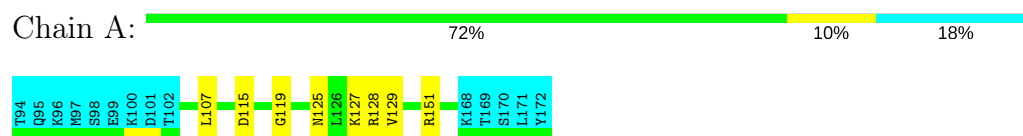


- Molecule 2: SFI1 peptide

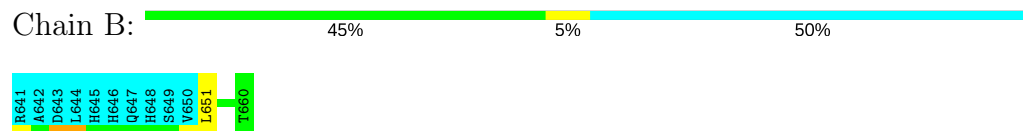


### 4.2.8 Score per residue for model 8

- Molecule 1: Centrin-2

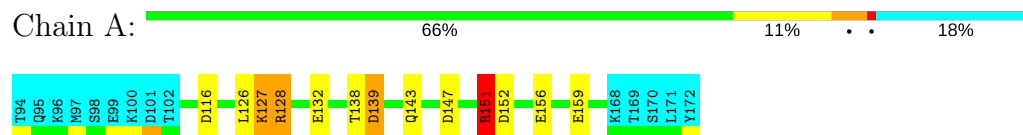


- Molecule 2: SFI1 peptide



### 4.2.9 Score per residue for model 9

- Molecule 1: Centrin-2

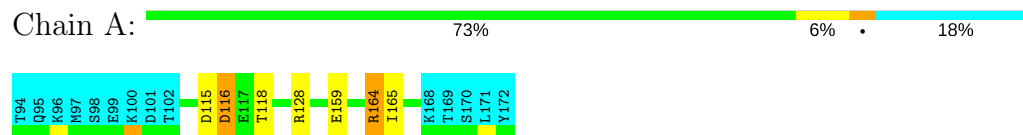


- Molecule 2: SFI1 peptide

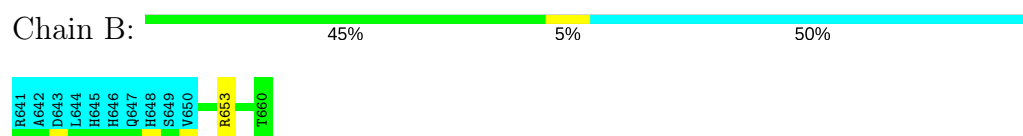


### 4.2.10 Score per residue for model 10

- Molecule 1: Centrin-2



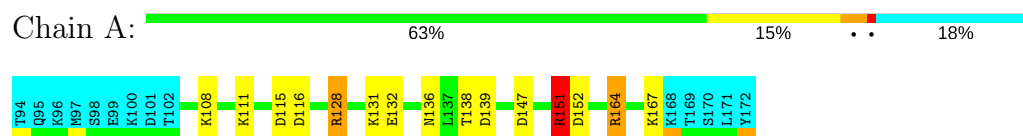
- Molecule 2: SFI1 peptide





### 4.2.11 Score per residue for model 11

- Molecule 1: Centrin-2

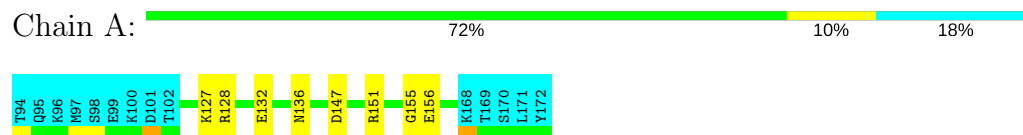


- Molecule 2: SFI1 peptide

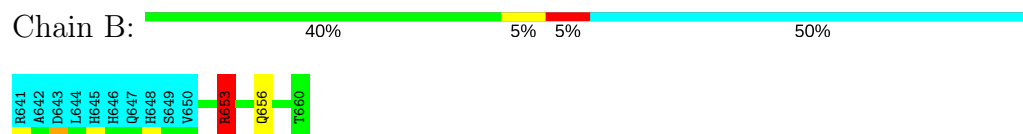


### 4.2.12 Score per residue for model 12

- Molecule 1: Centrin-2

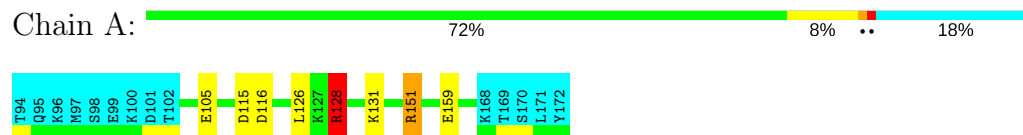


- Molecule 2: SFI1 peptide

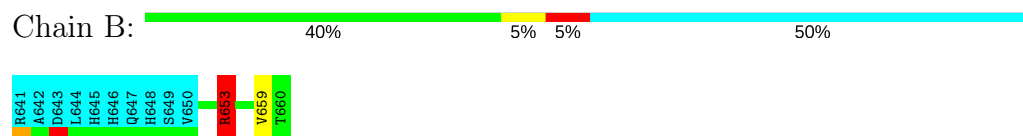


### 4.2.13 Score per residue for model 13

- Molecule 1: Centrin-2

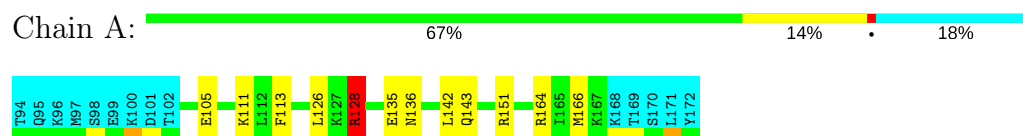


- Molecule 2: SFI1 peptide

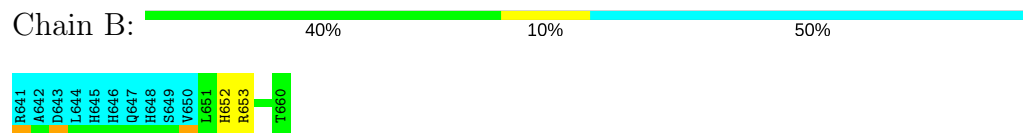


#### 4.2.14 Score per residue for model 14

##### • Molecule 1: Centrin-2

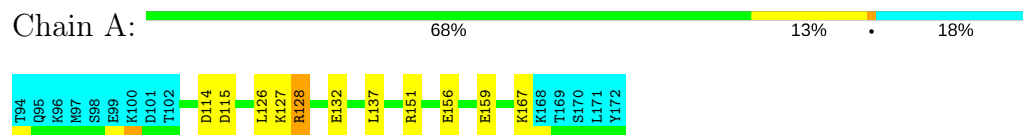


##### • Molecule 2: SFI1 peptide

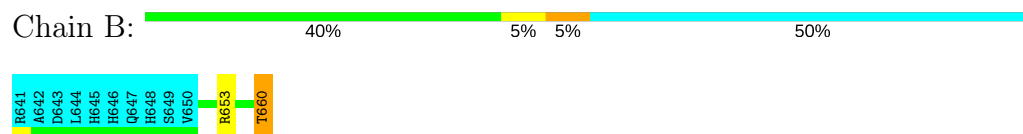


#### 4.2.15 Score per residue for model 15

##### • Molecule 1: Centrin-2



##### • Molecule 2: SFI1 peptide



## 5 Refinement protocol and experimental data overview

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

Of the 500 calculated structures, 15 were deposited, based on the following criterion: ?.

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

Software name	Classification	Version
Discover molecular modeling system	geometry optimization	2.98
Discover molecular modeling system	refinement	2.98

No chemical shift data was provided. 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	0.75±0.02	0±0/534 (0.0±0.0%)	1.11±0.05	2±2/710 (0.3±0.2%)
2	B	0.75±0.02	0±0/87 (0.0±0.0%)	1.17±0.17	0±1/117 (0.4±0.6%)
All	All	0.75	0/9315 (0.0%)	1.12	41/12405 (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±0.8
2	B	0.0±0.0	0.3±0.4
All	All	0	28

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	128	ARG	NE-CZ-NH2	-11.48	114.56	120.30	12	3
1	A	128	ARG	NE-CZ-NH1	11.29	125.94	120.30	9	10
1	A	164	ARG	NE-CZ-NH1	10.67	125.64	120.30	14	9
1	A	151	ARG	NE-CZ-NH2	-10.08	115.26	120.30	1	2
2	B	653	ARG	NE-CZ-NH2	-10.03	115.28	120.30	12	2
1	A	164	ARG	NE-CZ-NH2	-9.85	115.37	120.30	7	1
1	A	151	ARG	NE-CZ-NH1	8.99	124.79	120.30	7	6
2	B	653	ARG	NE-CZ-NH1	7.71	124.15	120.30	12	5
1	A	155	GLY	C-N-CA	5.64	135.80	121.70	12	1
1	A	113	PHE	CB-CG-CD1	5.04	124.33	120.80	14	1
1	A	164	ARG	NH1-CZ-NH2	-5.03	113.87	119.40	1	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	128	ARG	Sidechain	8
1	A	151	ARG	Sidechain,Peptide	6
1	A	164	ARG	Sidechain	5
2	B	653	ARG	Sidechain	4
1	A	156	GLU	Peptide	4
1	A	137	LEU	Peptide	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	530	519	519	1±1
2	B	85	86	86	0±0
All	All	9225	9075	9075	22

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 1.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:128:ARG:NH1	1:A:132:GLU:OE2	0.45	2.49	15	3
1:A:161:GLU:OE2	1:A:164:ARG:NH1	0.44	2.50	3	3
1:A:128:ARG:NH1	1:A:132:GLU:OE1	0.44	2.50	1	1
1:A:115:ASP:OD2	1:A:128:ARG:NH2	0.43	2.52	11	1
1:A:113:PHE:CD1	1:A:129:VAL:HG21	0.43	2.49	4	1
1:A:116:ASP:OD2	1:A:128:ARG:NH2	0.42	2.53	13	1
1:A:161:GLU:OE1	1:A:164:ARG:NH1	0.42	2.53	4	1
1:A:147:ASP:OD1	1:A:151:ARG:NH2	0.42	2.52	12	1
1:A:115:ASP:OD2	1:A:128:ARG:NH1	0.42	2.53	4	2
1:A:105:GLU:OE1	2:B:653:ARG:NH1	0.41	2.54	14	1
1:A:105:GLU:OE1	2:B:653:ARG:NH2	0.41	2.54	13	1
1:A:151:ARG:NH1	1:A:152:ASP:OD2	0.41	2.54	9	2
1:A:127:LYS:NZ	1:A:139:ASP:OD2	0.41	2.53	9	1
1:A:115:ASP:OD1	1:A:128:ARG:NH2	0.40	2.53	15	1
1:A:131:LYS:NZ	1:A:132:GLU:OE1	0.40	2.54	7	1

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:159:GLU:H	1:A:159:GLU:CD	0.40	2.19	9	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	65/79 (82%)	59±2 (91±3%)	6±1 (9±2%)	0±1 (1±1%)	30	75
2	B	9/20 (45%)	8±1 (88±12%)	1±1 (10±10%)	0±0 (2±4%)	12	51
All	All	1110/1485 (75%)	1002 (90%)	98 (9%)	10 (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)
2	B	653	ARG	2
1	A	136	ASN	1
2	B	659	VAL	1
1	A	150	ASP	1
1	A	116	ASP	1
1	A	115	ASP	1
1	A	119	GLY	1
1	A	138	THR	1
1	A	156	GLU	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	58/72 (81%)	52±2 (90±3%)	6±2 (10±3%)	12	56

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	B	8/17 (47%)	8±0 (95±6%)	0±0 (5±6%)	33 78
All	All	990/1335 (74%)	893 (90%)	97 (10%)	14 58

All 38 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	126	LEU	11
1	A	159	GLU	7
1	A	127	LYS	5
1	A	143	GLN	5
1	A	147	ASP	4
1	A	165	ILE	4
1	A	131	LYS	4
1	A	139	ASP	4
1	A	132	GLU	3
1	A	138	THR	3
1	A	111	LYS	3
1	A	118	THR	3
1	A	136	ASN	3
1	A	128	ARG	3
1	A	104	GLU	3
1	A	142	LEU	3
1	A	116	ASP	2
1	A	125	ASN	2
1	A	167	LYS	2
2	B	652	HIS	2
1	A	108	LYS	2
1	A	137	LEU	2
1	A	135	GLU	2
2	B	656	GLN	1
1	A	124	LYS	1
1	A	166	MET	1
1	A	164	ARG	1
1	A	152	ASP	1
1	A	115	ASP	1
2	B	660	THR	1
1	A	114	ASP	1
1	A	161	GLU	1
1	A	107	LEU	1
1	A	129	VAL	1
1	A	158	SER	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
2	B	653	ARG	1
1	A	151	ARG	1
2	B	651	LEU	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](#)

There are no ligands in this entry.

### 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

No chemical shift data were provided