



# Full wwPDB NMR Structure Validation Report ⓘ

Apr 26, 2016 – 08:25 PM BST

PDB ID : 2GFU  
Title : NMR solution structure of the PWWP domain of Mismatch repair protein hMSH6  
Authors : Laguri, C.; Friedrich, N.; Axt, M.; Gilquin, B.; Zinn-Justin, S.; Couprie, J.  
Deposited on : 2006-03-23

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-20027457  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027457

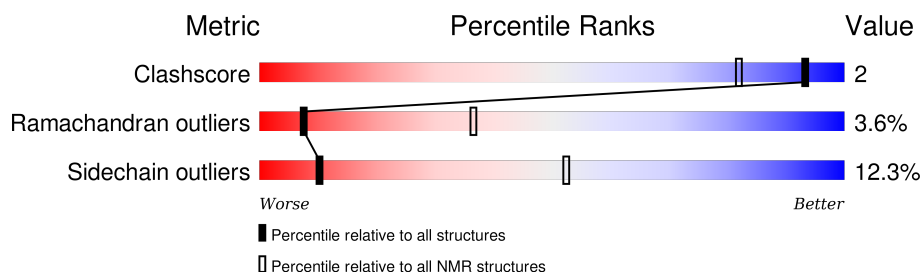
# 1 Overall quality at a glance

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	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	134	 69% 11% • 19%

## 2 Ensemble composition and analysis

This entry contains 20 models. Model 11 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:88-A:195 (108)	0.53	11

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. No single-model clusters were found.

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

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 2126 atoms, of which 1059 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called DNA mismatch repair protein MSH6.

Mol	Chain	Residues	Atoms						Trace
1	A	134	Total	C	H	N	O	S	0
			2126	679	1059	192	194	2	

There are 3 discrepancies between the modelled and reference sequences:

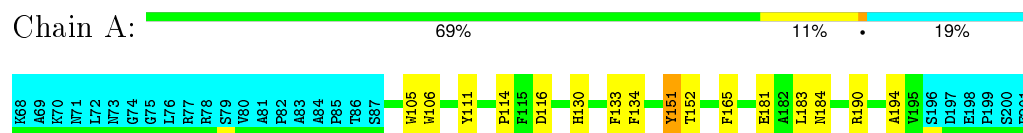
Chain	Residue	Modelled	Actual	Comment	Reference
A	88	SER	CYS	ENGINEERED	UNP P52701
A	108	SER	CYS	ENGINEERED	UNP P52701
A	196	SER	CYS	ENGINEERED	UNP P52701

## 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: DNA mismatch repair protein MSH6

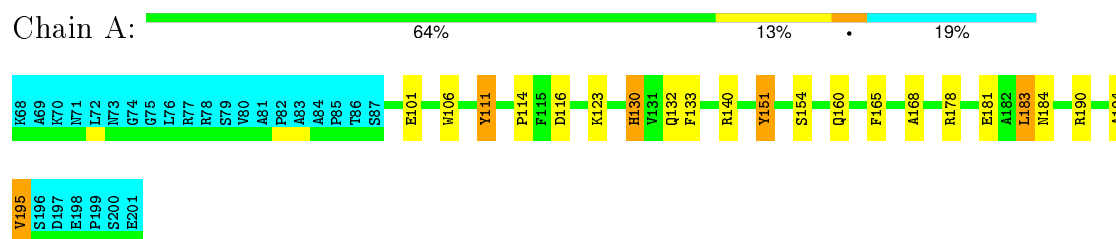


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

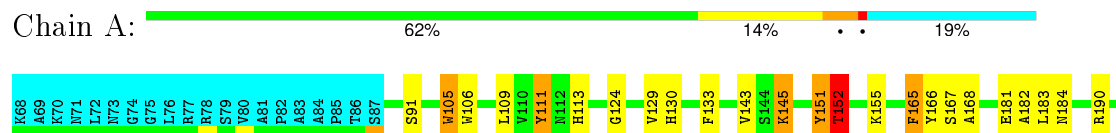
#### 4.2.1 Score per residue for model 1

- Molecule 1: DNA mismatch repair protein MSH6



#### 4.2.2 Score per residue for model 2

- Molecule 1: DNA mismatch repair protein MSH6





### 4.2.3 Score per residue for model 3

- Molecule 1: DNA mismatch repair protein MSH6



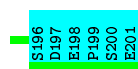
### 4.2.4 Score per residue for model 4

- Molecule 1: DNA mismatch repair protein MSH6



### 4.2.5 Score per residue for model 5

- Molecule 1: DNA mismatch repair protein MSH6



### 4.2.6 Score per residue for model 6

- Molecule 1: DNA mismatch repair protein MSH6

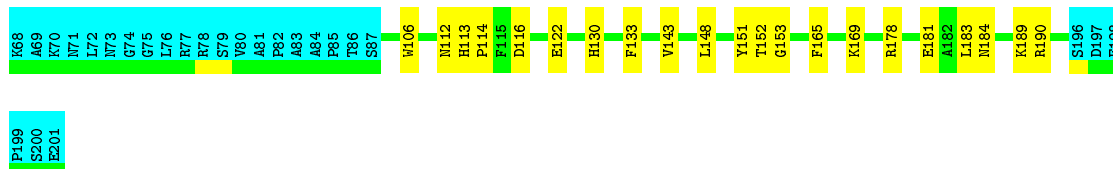




#### 4.2.7 Score per residue for model 7

- Molecule 1: DNA mismatch repair protein MSH6

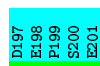
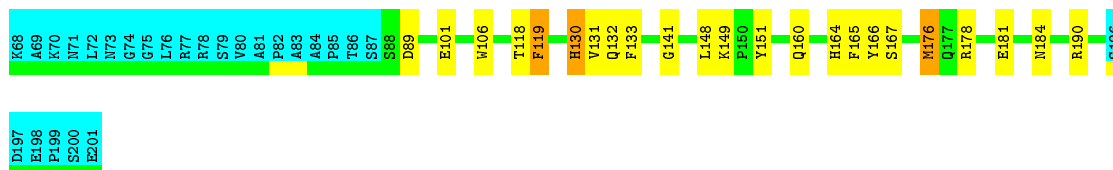
Chain A: 65% 16% 19%



#### 4.2.8 Score per residue for model 8

- Molecule 1: DNA mismatch repair protein MSH6

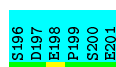
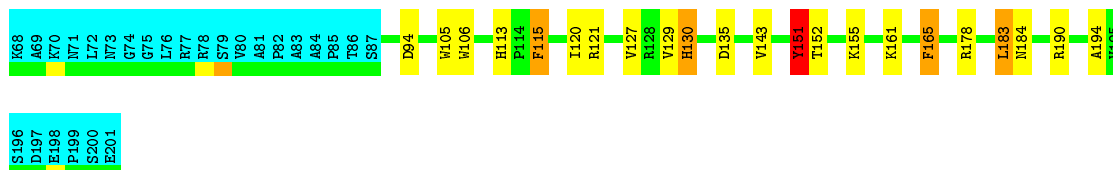
Chain A: 63% 15% 19%



#### 4.2.9 Score per residue for model 9

- Molecule 1: DNA mismatch repair protein MSH6

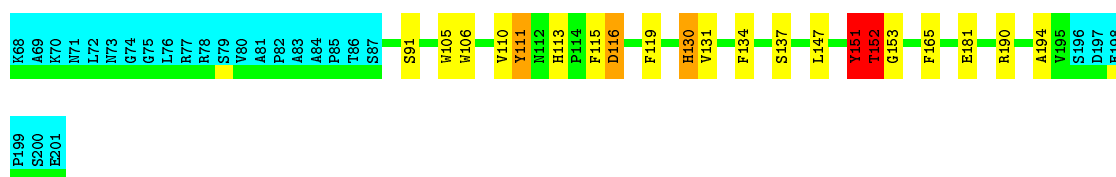
Chain A: 64% 13% 19%



#### 4.2.10 Score per residue for model 10

- Molecule 1: DNA mismatch repair protein MSH6

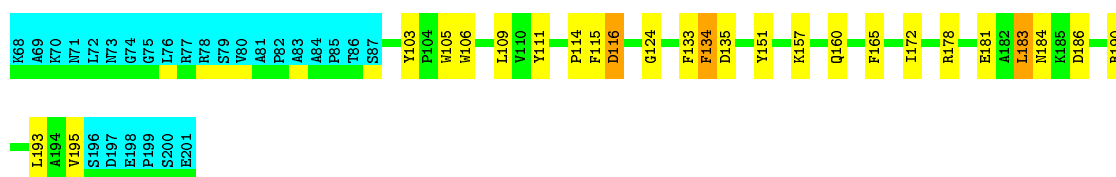
Chain A: 65% 12% 19%



#### 4.2.11 Score per residue for model 11 (medoid)

- Molecule 1: DNA mismatch repair protein MSH6

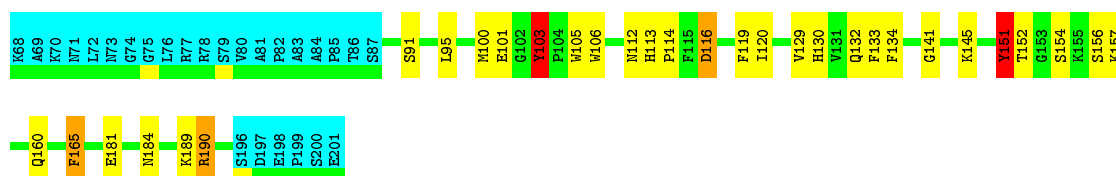
Chain A: 62% 16% 19%



#### 4.2.12 Score per residue for model 12

- Molecule 1: DNA mismatch repair protein MSH6

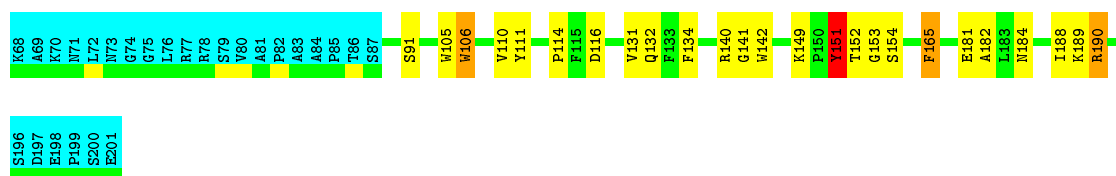
Chain A: 57% 19% 19%



#### 4.2.13 Score per residue for model 13

- Molecule 1: DNA mismatch repair protein MSH6

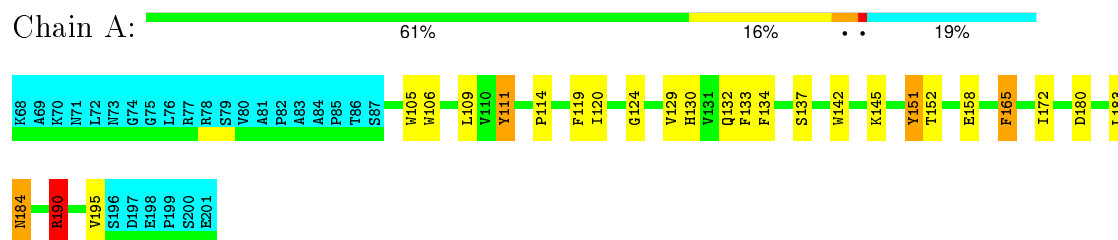
Chain A: 62% 16% 19%



#### 4.2.14 Score per residue for model 14

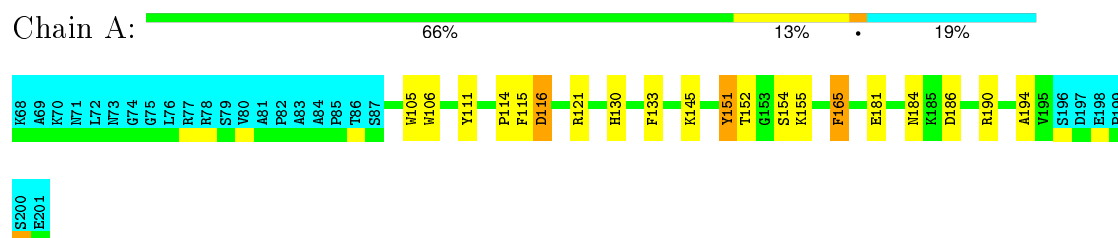
- Molecule 1: DNA mismatch repair protein MSH6





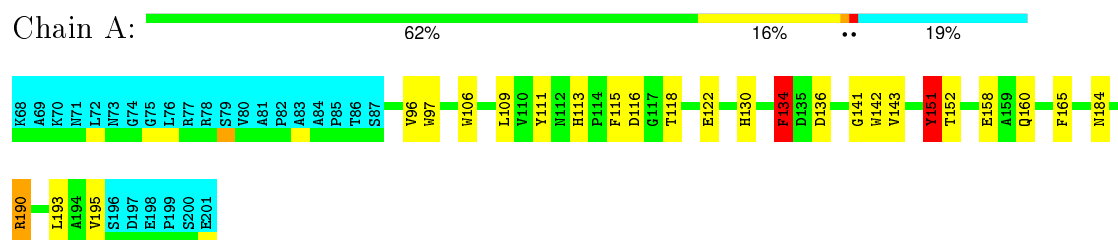
#### 4.2.15 Score per residue for model 15

- Molecule 1: DNA mismatch repair protein MSH6



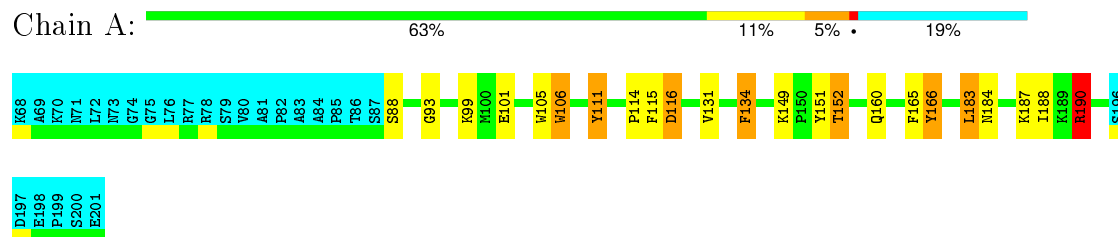
#### 4.2.16 Score per residue for model 16

- Molecule 1: DNA mismatch repair protein MSH6



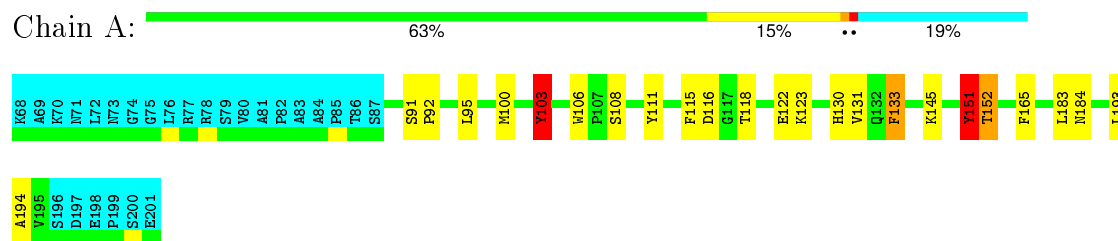
#### 4.2.17 Score per residue for model 17

- Molecule 1: DNA mismatch repair protein MSH6



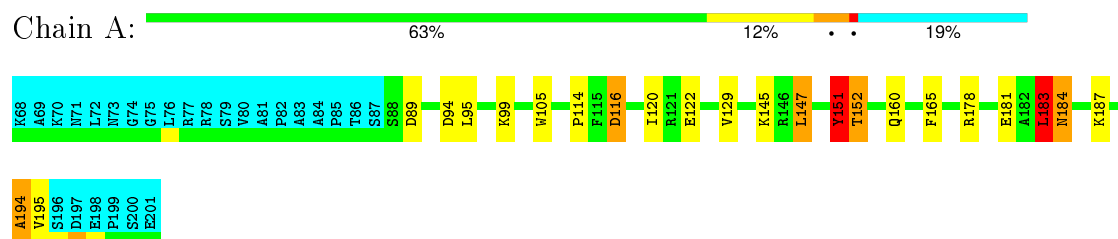
### 4.2.18 Score per residue for model 18

- Molecule 1: DNA mismatch repair protein MSH6



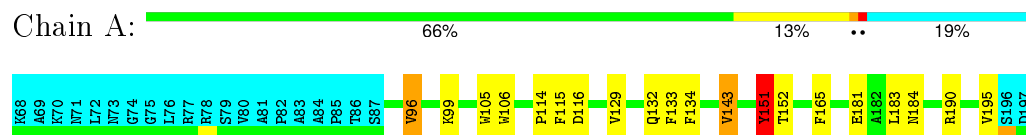
### 4.2.19 Score per residue for model 19

- Molecule 1: DNA mismatch repair protein MSH6



### 4.2.20 Score per residue for model 20

- Molecule 1: DNA mismatch repair protein MSH6



## 5 Refinement protocol and experimental data overview

Of the 1000 calculated structures, 20 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
INCA	structure solution	1.0
INCA	refinement	1.0

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	1.00±0.01	0±0/910 (0.0±0.0%)	1.51±0.04	9±3/1228 (0.8±0.2%)
All	All	1.00	0/18200 (0.0%)	1.51	186/24560 (0.8%)

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	1.5±1.4
All	All	0	30

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	103	TYR	CB-CG-CD1	9.73	126.84	121.00	18	2
1	A	133	PHE	CB-CG-CD1	9.53	127.47	120.80	15	10
1	A	103	TYR	CB-CG-CD2	-9.19	115.48	121.00	18	2
1	A	165	PHE	CB-CG-CD1	8.81	126.97	120.80	11	11
1	A	151	TYR	CB-CG-CD2	-8.36	115.98	121.00	20	10
1	A	106	TRP	CB-CG-CD2	-7.88	116.35	126.60	7	16
1	A	151	TYR	CB-CG-CD1	7.71	125.63	121.00	20	13
1	A	183	LEU	CB-CG-CD2	7.51	123.77	111.00	1	2
1	A	133	PHE	CB-CG-CD2	-7.34	115.66	120.80	1	6
1	A	106	TRP	CB-CG-CD1	7.16	136.31	127.00	7	13
1	A	88	SER	N-CA-CB	6.98	120.97	110.50	6	1
1	A	115	PHE	CB-CG-CD1	6.82	125.58	120.80	15	3
1	A	111	TYR	CB-CG-CD2	-6.76	116.94	121.00	14	2
1	A	151	TYR	C-N-CA	6.74	138.56	121.70	18	5
1	A	96	VAL	CA-CB-CG1	6.61	120.82	110.90	20	2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	190	ARG	CA-CB-CG	6.58	127.87	113.40	4	2
1	A	130	HIS	CA-CB-CG	6.51	124.67	113.60	5	1
1	A	91	SER	CB-CA-C	6.51	122.46	110.10	12	6
1	A	115	PHE	CB-CG-CD2	-6.46	116.28	120.80	15	2
1	A	134	PHE	CB-CG-CD2	-6.41	116.31	120.80	17	5
1	A	165	PHE	CB-CG-CD2	-6.39	116.33	120.80	11	3
1	A	131	VAL	CA-CB-CG1	6.32	120.38	110.90	18	1
1	A	105	TRP	CB-CG-CD2	-6.25	118.47	126.60	4	9
1	A	106	TRP	N-CA-CB	-6.16	99.51	110.60	7	3
1	A	141	GLY	N-CA-C	-6.10	97.85	113.10	13	4
1	A	190	ARG	CB-CA-C	6.06	122.51	110.40	4	1
1	A	176	MET	N-CA-CB	6.04	121.47	110.60	8	1
1	A	183	LEU	CA-CB-CG	6.00	129.10	115.30	9	5
1	A	166	TYR	CB-CG-CD1	5.94	124.56	121.00	17	2
1	A	152	THR	N-CA-C	-5.90	95.07	111.00	10	2
1	A	134	PHE	CB-CG-CD1	5.85	124.90	120.80	17	2
1	A	186	ASP	C-N-CA	5.85	136.32	121.70	11	2
1	A	187	LYS	CB-CA-C	5.81	122.02	110.40	17	1
1	A	106	TRP	CG-CD2-CE3	-5.75	128.72	133.90	7	3
1	A	168	ALA	C-N-CA	5.71	135.99	121.70	1	2
1	A	115	PHE	C-N-CA	5.70	135.94	121.70	10	4
1	A	97	TRP	N-CA-C	-5.63	95.80	111.00	16	1
1	A	103	TYR	CA-CB-CG	5.60	124.03	113.40	18	2
1	A	114	PRO	C-N-CA	5.58	135.64	121.70	6	5
1	A	195	VAL	CB-CA-C	5.57	121.98	111.40	1	1
1	A	166	TYR	CB-CG-CD2	-5.56	117.67	121.00	17	2
1	A	95	LEU	N-CA-CB	5.53	121.46	110.40	18	2
1	A	154	SER	C-N-CA	5.47	135.38	121.70	6	1
1	A	142	TRP	CB-CG-CD1	5.43	134.06	127.00	16	3
1	A	116	ASP	CB-CG-OD1	5.38	123.14	118.30	12	2
1	A	118	THR	N-CA-C	-5.36	96.52	111.00	18	1
1	A	183	LEU	CB-CG-CD1	-5.35	101.91	111.00	20	1
1	A	153	GLY	C-N-CA	5.30	134.95	121.70	13	1
1	A	119	PHE	CB-CG-CD1	-5.29	117.10	120.80	10	1
1	A	194	ALA	C-N-CA	5.27	134.87	121.70	19	1
1	A	156	SER	C-N-CA	5.16	134.59	121.70	12	1
1	A	143	VAL	N-CA-CB	-5.13	100.20	111.50	20	1
1	A	184	ASN	CB-CA-C	5.02	120.43	110.40	19	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	151	TYR	Sidechain	6
1	A	111	TYR	Sidechain	5
1	A	190	ARG	Sidechain	4
1	A	134	PHE	Sidechain	4
1	A	130	HIS	Sidechain	3
1	A	152	THR	Peptide	2
1	A	106	TRP	Peptide	2
1	A	103	TYR	Sidechain	2
1	A	133	PHE	Sidechain	1
1	A	166	TYR	Sidechain	1

## 6.2 Too-close contacts [i](#)

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	882	874	874	3±2
All	All	17640	17480	17480	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:109:LEU:HD11	1:A:190:ARG:HE	0.74	1.42	11	1
1:A:96:VAL:HG21	1:A:148:LEU:HD13	0.65	1.69	3	1
1:A:152:THR:HA	1:A:183:LEU:HD13	0.61	1.70	18	3
1:A:120:ILE:HG22	1:A:129:VAL:HG12	0.60	1.71	19	4
1:A:113:HIS:CD2	1:A:130:HIS:CD2	0.57	2.92	12	5
1:A:105:TRP:CD2	1:A:165:PHE:HB3	0.56	2.35	2	4
1:A:129:VAL:HG22	1:A:145:LYS:HB2	0.55	1.78	2	1
1:A:130:HIS:CD2	1:A:142:TRP:CZ2	0.54	2.96	5	1
1:A:113:HIS:CD2	1:A:130:HIS:CG	0.53	2.96	16	1
1:A:152:THR:HA	1:A:183:LEU:HD11	0.51	1.83	17	1
1:A:109:LEU:CD1	1:A:190:ARG:HE	0.51	2.18	11	1
1:A:149:LYS:HZ2	1:A:149:LYS:HB2	0.50	1.66	13	1
1:A:95:LEU:CD2	1:A:190:ARG:HE	0.49	2.19	12	1
1:A:119:PHE:HA	1:A:130:HIS:CE1	0.48	2.43	8	2

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:113:HIS:CG	1:A:130:HIS:CD2	0.47	3.02	2	2
1:A:109:LEU:HD21	1:A:190:ARG:CZ	0.47	2.38	16	1
1:A:129:VAL:HG11	1:A:145:LYS:HD3	0.46	1.87	12	3
1:A:109:LEU:HD22	1:A:190:ARG:HE	0.46	1.71	14	1
1:A:93:GLY:HA2	1:A:190:ARG:CZ	0.45	2.42	17	1
1:A:151:TYR:CE2	1:A:153:GLY:HA2	0.44	2.47	10	1
1:A:106:TRP:CZ3	1:A:172:ILE:HD13	0.44	2.47	14	2
1:A:97:TRP:CZ3	1:A:107:PRO:HG3	0.44	2.48	6	1
1:A:180:ASP:O	1:A:184:ASN:HB3	0.43	2.13	14	2
1:A:99:LYS:O	1:A:147:LEU:HD13	0.43	2.13	19	1
1:A:95:LEU:HD22	1:A:190:ARG:HE	0.43	1.74	12	1
1:A:99:LYS:O	1:A:147:LEU:HD22	0.43	2.14	19	1
1:A:109:LEU:HD22	1:A:190:ARG:HB2	0.42	1.91	2	2
1:A:119:PHE:HB2	1:A:130:HIS:CE1	0.42	2.49	3	1
1:A:152:THR:HA	1:A:183:LEU:HD21	0.42	1.90	19	1
1:A:106:TRP:CZ2	1:A:135:ASP:HB3	0.41	2.50	11	2
1:A:113:HIS:CD2	1:A:115:PHE:CZ	0.41	3.09	9	1
1:A:111:TYR:O	1:A:130:HIS:CD2	0.41	2.73	1	1
1:A:182:ALA:HB1	1:A:190:ARG:HB3	0.41	1.93	13	1
1:A:100:MET:N	1:A:103:TYR:CE1	0.41	2.88	12	1
1:A:182:ALA:HB1	1:A:190:ARG:HG2	0.41	1.92	2	1
1:A:130:HIS:CG	1:A:140:ARG:HH22	0.41	2.33	5	1
1:A:109:LEU:HB2	1:A:134:PHE:CZ	0.40	2.51	16	1
1:A:129:VAL:O	1:A:143:VAL:HG12	0.40	2.16	20	1
1:A:110:VAL:HG23	1:A:131:VAL:HG12	0.40	1.94	13	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	108/134 (81%)	93±2 (86±2%)	12±3 (11±3%)	4±1 (4±1%)	7	36
All	All	2160/2680 (81%)	1852 (86%)	230 (11%)	78 (4%)	7	36

All 16 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	152	THR	13
1	A	151	TYR	9
1	A	114	PRO	9
1	A	194	ALA	9
1	A	101	GLU	6
1	A	195	VAL	5
1	A	116	ASP	5
1	A	124	GLY	4
1	A	154	SER	4
1	A	134	PHE	3
1	A	157	LYS	2
1	A	153	GLY	2
1	A	88	SER	2
1	A	167	SER	2
1	A	166	TYR	2
1	A	119	PHE	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	93/113 (82%)	82±2 (88±2%)	11±2 (12±2%)	10	52
All	All	1860/2260 (82%)	1632 (88%)	228 (12%)	10	52

All 60 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	184	ASN	19
1	A	151	TYR	16
1	A	190	ARG	15
1	A	116	ASP	14
1	A	181	GLU	12
1	A	165	PHE	11
1	A	160	GLN	9
1	A	132	GLN	8
1	A	183	LEU	8
1	A	111	TYR	7

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	178	ARG	7
1	A	130	HIS	6
1	A	143	VAL	5
1	A	122	GLU	5
1	A	148	LEU	5
1	A	189	LYS	4
1	A	193	LEU	4
1	A	145	LYS	3
1	A	131	VAL	3
1	A	155	LYS	3
1	A	105	TRP	3
1	A	149	LYS	3
1	A	133	PHE	3
1	A	118	THR	2
1	A	94	ASP	2
1	A	158	GLU	2
1	A	135	ASP	2
1	A	112	ASN	2
1	A	103	TYR	2
1	A	147	LEU	2
1	A	140	ARG	2
1	A	137	SER	2
1	A	89	ASP	2
1	A	121	ARG	2
1	A	99	LYS	2
1	A	188	ILE	2
1	A	187	LYS	2
1	A	115	PHE	2
1	A	123	LYS	2
1	A	176	MET	2
1	A	195	VAL	2
1	A	166	TYR	1
1	A	95	LEU	1
1	A	88	SER	1
1	A	161	LYS	1
1	A	174	ARG	1
1	A	136	ASP	1
1	A	152	THR	1
1	A	92	PRO	1
1	A	91	SER	1
1	A	119	PHE	1
1	A	127	VAL	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	108	SER	1
1	A	106	TRP	1
1	A	164	HIS	1
1	A	154	SER	1
1	A	110	VAL	1
1	A	100	MET	1
1	A	169	LYS	1
1	A	96	VAL	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