



# wwPDB NMR Structure Validation Summary Report ⓘ

Sep 11, 2025 – 10:15 AM EDT

PDB ID : 9PYS / pdb\_00009pys  
BMRB ID : 31263  
Title : NMR RDC refinement of the helical domain of the SARS-CoV-2 monomeric  
Main Protease (MPROH41Q,10-306)  
Authors : Smith, M.J.; Ying, J.; Shen, Y.; Louis, J.M.; Bax, A.  
Deposited on : 2025-08-08

This is a wwPDB NMR Structure Validation Summary 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4-5-2 with Phenix2.0rc1  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
wwPDB-RCI : v\_1n\_11\_5\_13\_A (Berjanski et al., 2005)  
PANAV : Wang et al. (2010)  
wwPDB-ShiftChecker : v1.2  
BMRB Restraints Analysis : v1.2  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.45.1

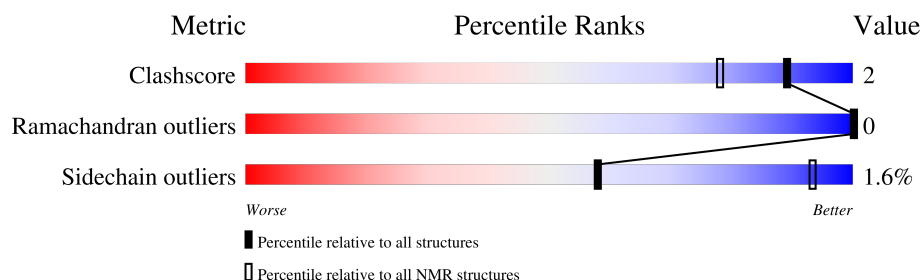
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 23%.

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	210492	14027
Ramachandran outliers	207382	12486
Sidechain outliers	206894	12463

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	306	

## 2 Ensemble composition and analysis

This entry contains 10 models. Model 5 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:200-A:306 (107)	0.14	5

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, 3, 7, 9, 10
2	2, 5, 6
Single-model clusters	4; 8

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1668 atoms, of which 829 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called 3C-like proteinase nsp5.

Mol	Chain	Residues	Atoms						Trace
1	A	107	Total	C	H	N	O	S	0
			1668	537	829	138	159	5	

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	41	GLN	HIS	engineered mutation	UNP P0DTD1



## 5 Refinement protocol and experimental data overview

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

Of the 10 calculated structures, 10 were deposited, based on the following criterion: *all calculated structures submitted*.

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

Software name	Classification	Version
X-PLOR NIH	refinement	

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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	332
Number of shifts mapped to atoms	332
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	23%

## 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	1.24±0.00	0±0/855 ( 0.0± 0.0%)	1.31±0.01	3±1/1164 ( 0.3± 0.1%)
All	All	1.24	0/8550 ( 0.0%)	1.31	32/11640 ( 0.3%)

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	258	GLY	N-CA-C	-5.71	107.36	115.30	10	6
1	A	215	GLY	N-CA-C	-5.69	107.16	115.63	5	10
1	A	214	ASN	N-CA-C	-5.40	106.70	113.28	10	9
1	A	285	ALA	N-CA-C	-5.22	106.60	113.17	8	6
1	A	222	ARG	N-CA-C	-5.11	106.90	112.72	10	1

There are no chirality outliers.

There are no planarity outliers.

### 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	839	829	828	3±1
All	All	8390	8290	8280	28

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.

5 of 11 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:209:TYR:CD2	1:A:264:MET:HE1	0.50	2.42	3	1
1:A:227:LEU:HD11	1:A:242:LEU:O	0.49	2.08	3	1
1:A:239:TYR:CE1	1:A:272:LEU:HD11	0.47	2.45	1	1
1:A:212:VAL:HG21	1:A:264:MET:HE1	0.47	1.86	8	1
1:A:202:VAL:HG21	1:A:246:HIS:CG	0.43	2.49	6	7

## 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	105/306 (34%)	104±0 (99±0%)	1±0 (1±0%)	0±0 (0±0%)	100	100
All	All	1050/3060 (34%)	1040 (99%)	10 (1%)	0 (0%)	100	100

There are no Ramachandran outliers.

### 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	92/263 (35%)	90±1 (98±1%)	2±1 (2±1%)	58	93
All	All	920/2630 (35%)	905 (98%)	15 (2%)	58	93

5 of 7 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	200	ILE	5
1	A	301	SER	4

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	217	ARG	2
1	A	221	ASN	1
1	A	274	ASN	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 oligosaccharides 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 [i](#)

The completeness of assignment taking into account all chemical shift lists is 23% for the well-defined parts and 23% for the entire structure.

### 7.1 Chemical shift list 1

File name: working\_cs.cif

Chemical shift list name: *default\_1*

#### 7.1.1 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	332
Number of shifts mapped to atoms	332
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 [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	86	$0.20 \pm 0.15$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}_\beta$	66	$1.20 \pm 0.16$	Should be checked
$^{13}\text{C}'$	0	—	None (insufficient data)
$^{15}\text{N}$	90	$0.20 \pm 0.42$	None needed ( $< 0.5$ ppm)

#### 7.1.3 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 23%, i.e. 332 atoms were assigned a chemical shift out of a possible 1461. 0 out of 24 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	266/536 (50%)	90/218 (41%)	86/214 (40%)	90/104 (87%)
Sidechain	66/807 (8%)	0/530 (0%)	66/250 (26%)	0/27 (0%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	0/118 (0%)	0/58 (0%)	0/57 (0%)	0/3 (0%)
Overall	332/1461 (23%)	90/806 (11%)	152/521 (29%)	90/134 (67%)

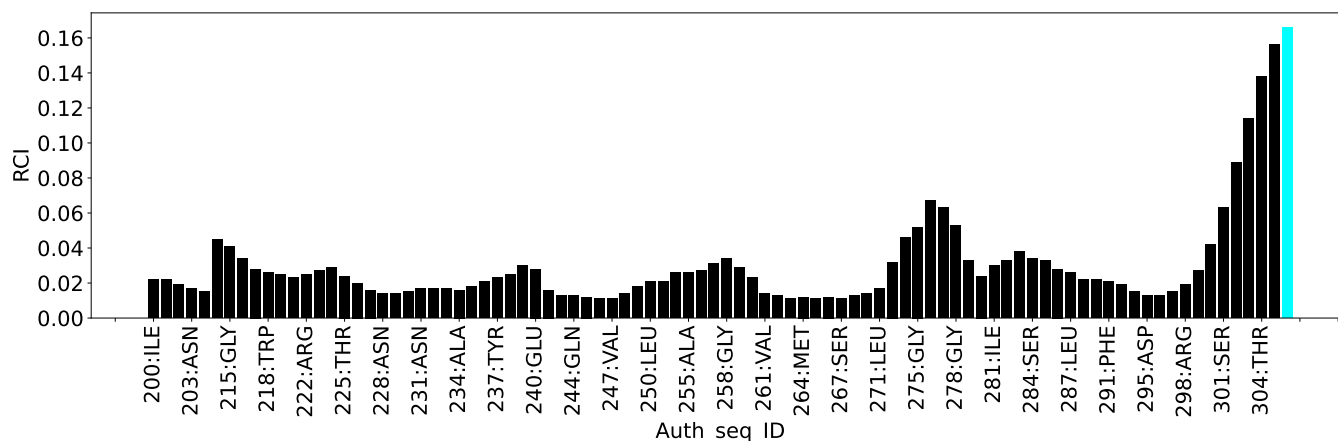
#### 7.1.4 Statistically unusual chemical shifts [i](#)

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. If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:



## 8 NMR restraints analysis

No restraints data found

## 9 Distance violation analysis ⓘ

No distance restraints data found

## 10 Dihedral-angle violation analysis ⓘ

No dihedral-angle restraints found