



# Full wwPDB NMR Structure Validation Report ⓘ

Oct 11, 2021 – 02:43 AM EDT

PDB ID : 2JUF  
Title : NMR solution structure of PARC CPH Domain. NESG Target HR3443B/SGC-Toronto  
Authors : Kaustov, L.; Liao, J.C.C.; Lemak, S.; Duan, S.; Muhandiram, R.; Karra, M.; Srisailam, S.H.; Sundstrom, M.; Weigelt, J.; Edwards, A.M.; Dhe-Paganon, S.; Arrowsmith, C.H.; Structural Genomics Consortium (SGC); Northeast Structural Genomics Consortium (NESG)  
Deposited on : 2007-08-23

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

<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:

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 : 2.23.2  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.23.2

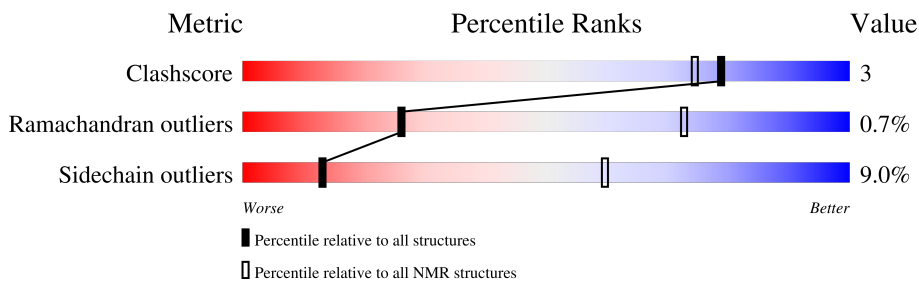
# 1 Overall quality at a glance

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	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	105	50% 6% • 5% 39%

## 2 Ensemble composition and analysis

This entry contains 20 models. Model 10 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:14-A:49, A:53-A:75 (59)	0.25	10

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 3 clusters and 1 single-model cluster was found.

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

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1024 atoms, of which 492 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called p53-associated parkin-like cytoplasmic protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	64	1024	338	492	91	100	3	0

There are 5 discrepancies between the modelled and reference sequences:

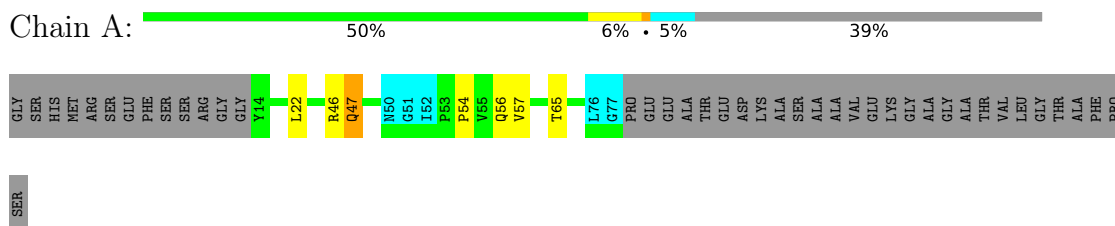
Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP Q8IWT3
A	2	SER	-	expression tag	UNP Q8IWT3
A	3	HIS	-	expression tag	UNP Q8IWT3
A	4	MET	-	expression tag	UNP Q8IWT3
A	12	GLY	SER	engineered mutation	UNP Q8IWT3

## 4 Residue-property plots

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

These plots are provided for all protein, RNA, DNA and oligosaccharide 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: p53-associated parkin-like cytoplasmic protein

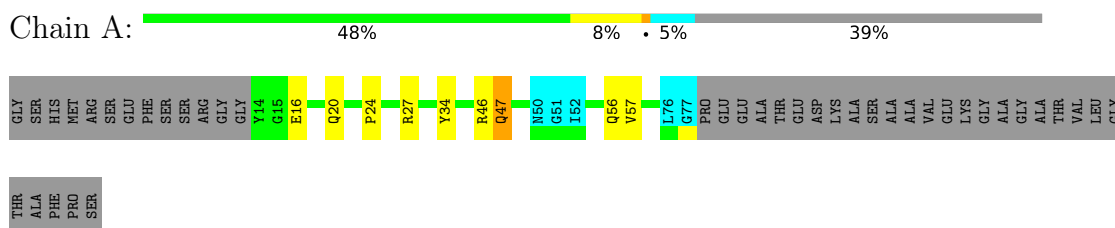


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

#### 4.2.1 Score per residue for model 1

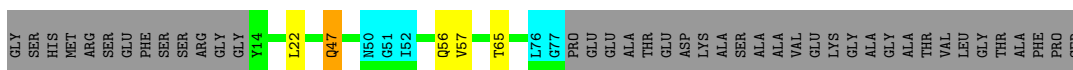
- Molecule 1: p53-associated parkin-like cytoplasmic protein



#### 4.2.2 Score per residue for model 2

- Molecule 1: p53-associated parkin-like cytoplasmic protein





### 4.2.3 Score per residue for model 3

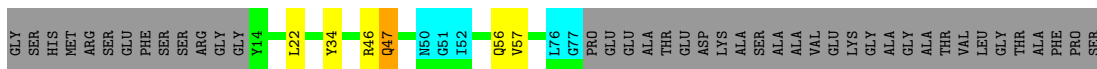
- Molecule 1: p53-associated parkin-like cytoplasmic protein



PHE  
PRO  
SER

### 4.2.4 Score per residue for model 4

- Molecule 1: p53-associated parkin-like cytoplasmic protein



### 4.2.5 Score per residue for model 5

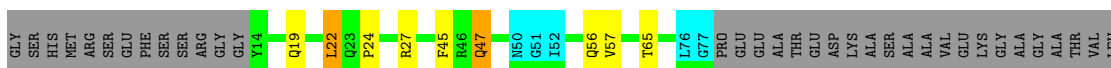
- Molecule 1: p53-associated parkin-like cytoplasmic protein



ALA  
PHE  
PRO  
SER

### 4.2.6 Score per residue for model 6

- Molecule 1: p53-associated parkin-like cytoplasmic protein



GLY  
THR  
ALA  
PHE  
PRO  
SER

#### 4.2.7 Score per residue for model 7

- Molecule 1: p53-associated parkin-like cytoplasmic protein

Chain A: 45% 10% 5% 39%

GLY SER HIS MET ARG SER GLU PHE SER SER ARG ARG GLY GLY Y14 Q19 M26 R29 R30 R46 Q47 N50 G51 I52 P53 P54 V57 V67 V68 H69 M72 I75 L76 G77 PRO GLU GLU ALA THR GLU ASP LYS IYS ALA SER ALA VAL GLY LYS GLY ALA GLY

ALA  
THR  
VAL  
LEU  
GLY  
THR  
ALA  
PHE  
PRO  
SER

#### 4.2.8 Score per residue for model 8

- Molecule 1: p53-associated parkin-like cytoplasmic protein

Chain A: 47% 10% 5% 39%

GLY SER HIS MET ARG SER GLU PHE SER SER ARG ARG GLY GLY Y14 L22 Q23 P24 R27 F45 R46 Q47 S48 M49 N50 G51 I52 P53 P54 V57 T65 L76 G77 PRO GLU GLU ALA THR GLU ASP LYS ALA SER ALA VAL GLU LYS GLY GLY ALA GLY ALA THR VAL LEU

GLY  
THR  
ALA  
PHE  
PRO  
SER

#### 4.2.9 Score per residue for model 9

- Molecule 1: p53-associated parkin-like cytoplasmic protein

Chain A: 48% 8% 5% 39%

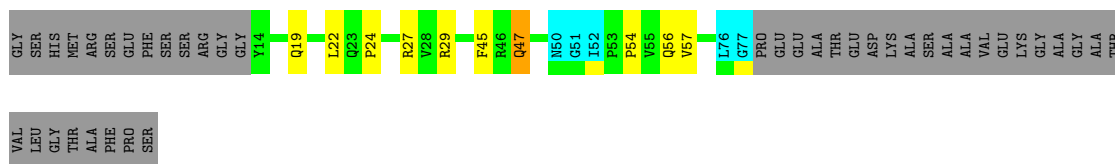
GLY SER HIS MET ARG SER GLU PHE SER SER ARG ARG GLY GLY Y14 L22 Q23 P24 R46 Q47 N50 G51 I52 P53 P54 V57 V68 L73 L76 G77 PRO GLU GLU ALA THR GLU ASP LYS ALA SER ALA VAL GLU LYS GLY GLY ALA GLY ALA THR VAL LEU GLY

THR  
ALA  
PHE  
PRO  
SER

#### 4.2.10 Score per residue for model 10 (medoid)

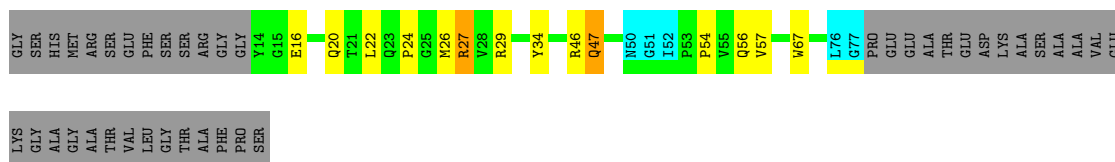
- Molecule 1: p53-associated parkin-like cytoplasmic protein

Chain A: 47% 9% 5% 39%



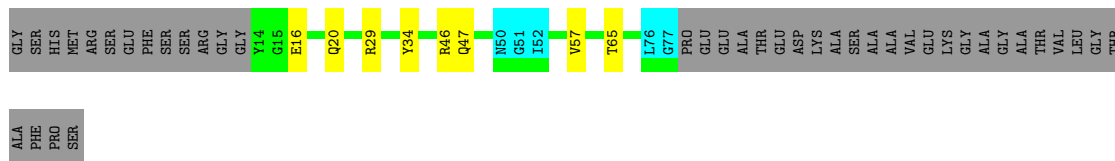
#### 4.2.11 Score per residue for model 11

- Molecule 1: p53-associated parkin-like cytoplasmic protein



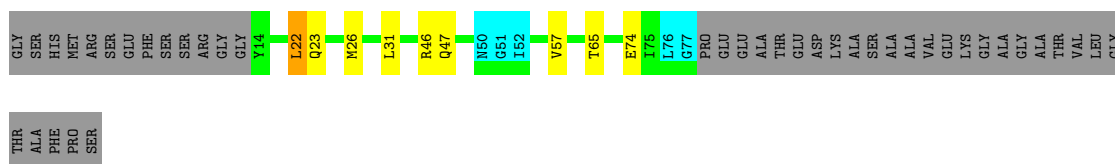
#### 4.2.12 Score per residue for model 12

- Molecule 1: p53-associated parkin-like cytoplasmic protein



#### 4.2.13 Score per residue for model 13

- Molecule 1: p53-associated parkin-like cytoplasmic protein

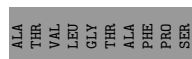


#### 4.2.14 Score per residue for model 14

- Molecule 1: p53-associated parkin-like cytoplasmic protein



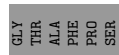
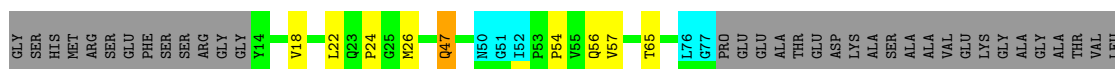
Chain A: 45% 10% 5% 39%



#### 4.2.15 Score per residue for model 15

- Molecule 1: p53-associated parkin-like cytoplasmic protein

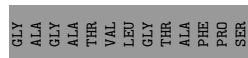
Chain A: 48% 8% 5% 39%



#### 4.2.16 Score per residue for model 16

- Molecule 1: p53-associated parkin-like cytoplasmic protein

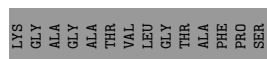
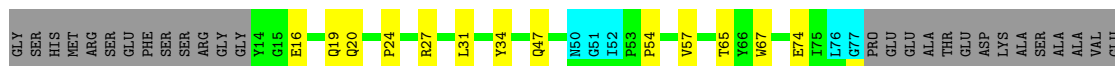
Chain A: 44% 11% 5% 39%



#### 4.2.17 Score per residue for model 17

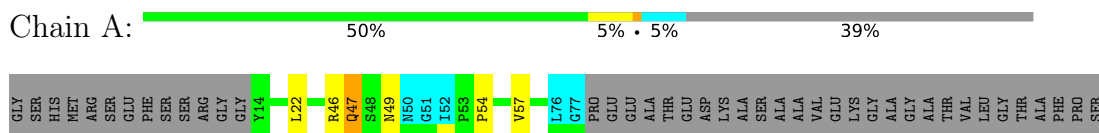
- Molecule 1: p53-associated parkin-like cytoplasmic protein

Chain A: 44% 12% 5% 39%



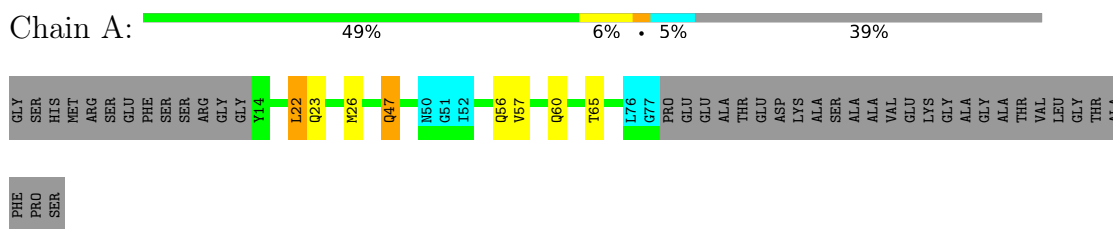
#### 4.2.18 Score per residue for model 18

- Molecule 1: p53-associated parkin-like cytoplasmic protein



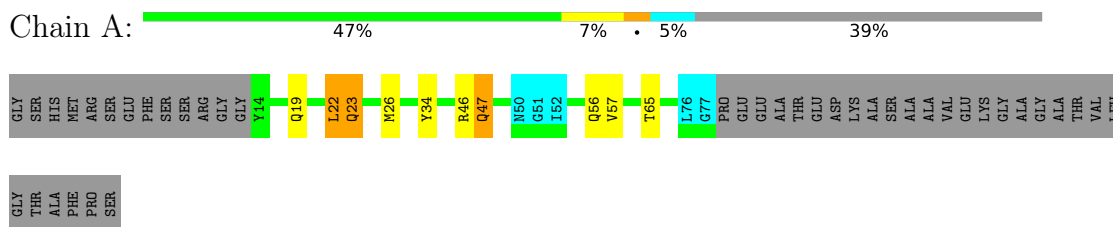
#### 4.2.19 Score per residue for model 19

- Molecule 1: p53-associated parkin-like cytoplasmic protein



#### 4.2.20 Score per residue for model 20

- Molecule 1: p53-associated parkin-like cytoplasmic protein



## 5 Refinement protocol and experimental data overview

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

Of the 100 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
CNS	refinement	1.1
CYANA	structure solution	2.1

No chemical shift data was provided.

## 6 Model quality [i](#)

### 6.1 Standard geometry [i](#)

There are no covalent bond-length or bond-angle outliers.

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

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	46	ARG	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	500	458	455	3±1
All	All	10000	9160	9100	53

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 3.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:22:LEU:HD13	1:A:26:MET:SD	0.61	2.34	19	3
1:A:26:MET:SD	1:A:75:ILE:HG21	0.58	2.38	7	1

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:47:GLN:NE2	1:A:56:GLN:HB2	0.54	2.18	10	12
1:A:22:LEU:HD12	1:A:26:MET:SD	0.52	2.45	11	2
1:A:54:PRO:HB2	1:A:67:TRP:HB3	0.51	1.82	14	6
1:A:23:GLN:HB2	1:A:26:MET:HB3	0.50	1.82	13	2
1:A:68:VAL:HG21	1:A:73:LEU:HD13	0.48	1.85	9	1
1:A:45:PHE:CZ	1:A:48:SER:HB2	0.46	2.45	8	1
1:A:18:VAL:O	1:A:22:LEU:HB2	0.46	2.10	15	1
1:A:16:GLU:O	1:A:20:GLN:HG2	0.45	2.12	11	4
1:A:46:ARG:HB2	1:A:47:GLN:NE2	0.44	2.27	14	9
1:A:23:GLN:NE2	1:A:26:MET:SD	0.44	2.90	20	1
1:A:49:ASN:HD21	1:A:55:VAL:HA	0.44	1.73	14	1
1:A:69:HIS:HB3	1:A:72:MET:SD	0.44	2.53	14	2
1:A:24:PRO:HA	1:A:45:PHE:CD1	0.43	2.49	10	1
1:A:31:LEU:HD23	1:A:74:GLU:HB2	0.43	1.91	13	2
1:A:22:LEU:HD23	1:A:26:MET:SD	0.42	2.53	16	1
1:A:31:LEU:HD11	1:A:72:MET:HA	0.42	1.92	16	1
1:A:22:LEU:HD12	1:A:45:PHE:HB2	0.41	1.93	6	1
1:A:56:GLN:HB3	1:A:65:THR:HG22	0.41	1.91	16	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	58/105 (55%)	56±1 (96±2%)	2±1 (4±2%)	0±1 (1±1%)	26	73
All	All	1160/2100 (55%)	1110 (96%)	42 (4%)	8 (1%)	26	73

All 2 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	24	PRO	7
1	A	27	ARG	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	53/84 (63%)	48±1 (91±2%)	5±1 (9±2%)	13 60
All	All	1060/1680 (63%)	965 (91%)	95 (9%)	13 60

All 14 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	47	GLN	20
1	A	57	VAL	20
1	A	22	LEU	13
1	A	65	THR	10
1	A	34	TYR	8
1	A	19	GLN	8
1	A	29	ARG	4
1	A	46	ARG	3
1	A	60	GLN	2
1	A	27	ARG	2
1	A	49	ASN	2
1	A	64	ARG	1
1	A	30	MET	1
1	A	23	GLN	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 monosaccharides 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