WORLDWIDE wwPDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report (i)

Feb 19, 2022 – 11:37 AM EST

PDB ID	:	1Q1O
Title	:	Solution Structure of the PB1 Domain of Cdc24p (Long Form)
Authors	:	Yoshinaga, S.; Kohjima, M.; Ogura, K.; Yokochi, M.; Takeya, R.; Ito, T.; Sumimoto, H.; Inagaki, F.
Deposited on	:	2003-07-22

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

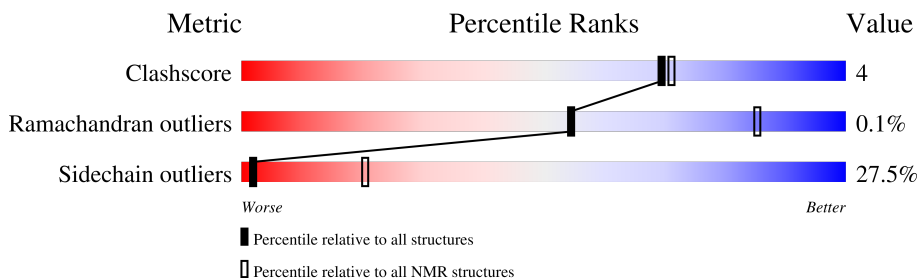|		
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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	Percentile Ranks	Value
Clashscore		4
Ramachandran outliers		0.1%
Sidechain outliers		27.5%

Worse   *Better*

■ Percentile relative to all structures

▯ Percentile relative to all NMR structures

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 >=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 <=5%

Mol	Chain	Length	Quality of chain
1	A	98	48% 23% 29%

2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 15 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:761-A:768, A:779-A:799, A:814-A:854 (70)	0.25	15

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 6 single-model clusters were found.

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

3 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 1555 atoms, of which 765 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Cell division control protein 24.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	98	1555	498	765	130	160	2	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	757	GLY	-	cloning artifact	UNP P11433
A	758	PRO	-	cloning artifact	UNP P11433
A	759	LEU	-	cloning artifact	UNP P11433
A	760	GLY	-	cloning artifact	UNP P11433

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.

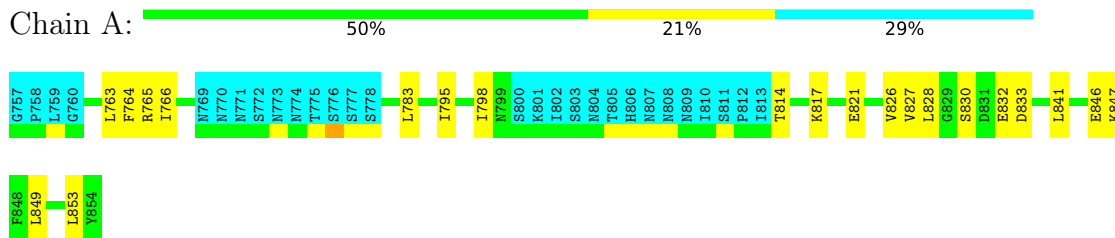• Molecule 1: Cell division control protein 24
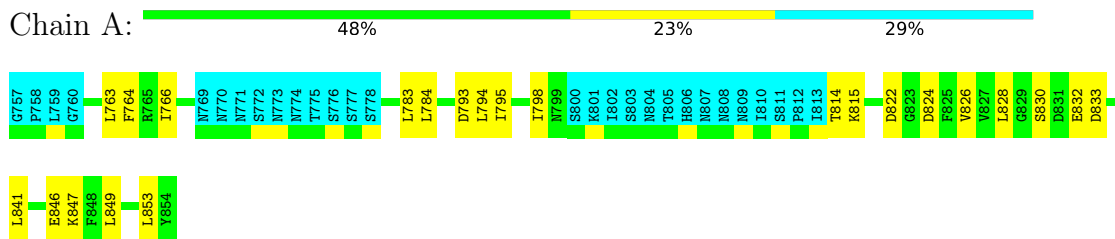
4.2   Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.
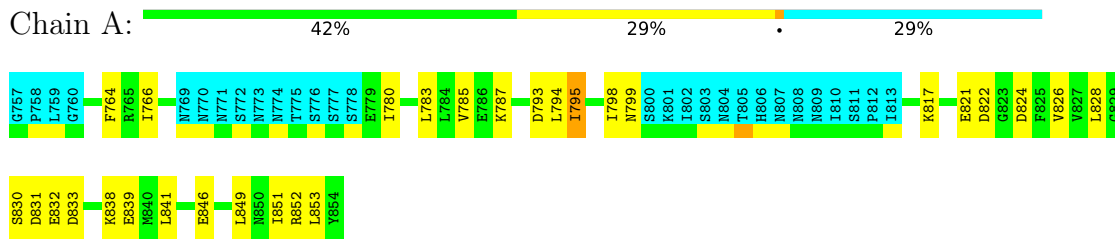
4.2.1   Score per residue for model 1

• Molecule 1: Cell division control protein 24
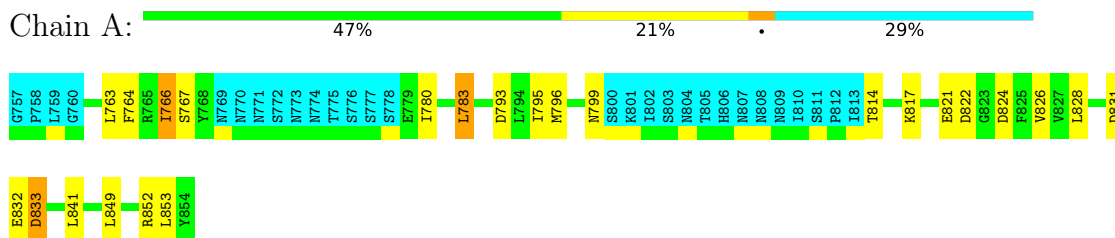

4.2.2   Score per residue for model 2

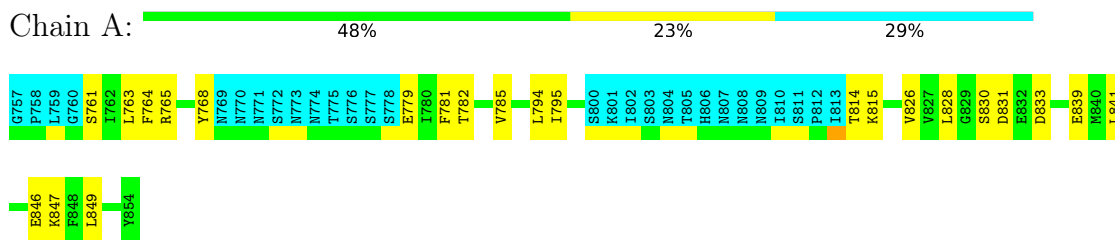• Molecule 1: Cell division control protein 24

4.2.3 Score per residue for model 3

• Molecule 1: Cell division control protein 24

Chain A:

4.2.4 Score per residue for model 4

• Molecule 1: Cell division control protein 24

Chain A:

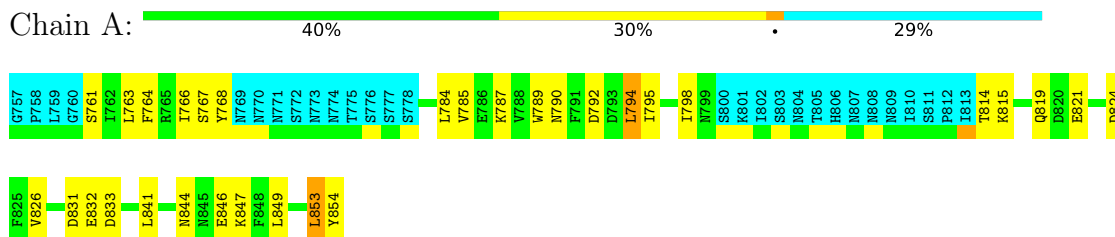### 4.2.5 Score per residue for model 5

• Molecule 1: Cell division control protein 24

Chain A:

4.2.6 Score per residue for model 6

• Molecule 1: Cell division control protein 24

Chain A: 42%   29%   •   29%

4.2.7   Score per residue for model 7

• Molecule 1: Cell division control protein 24

Chain A: 47%   21%   •   29%

4.2.8   Score per residue for model 8

• Molecule 1: Cell division control protein 24

Chain A: 48%   23%   29%

4.2.9   Score per residue for model 9

• Molecule 1: Cell division control protein 24

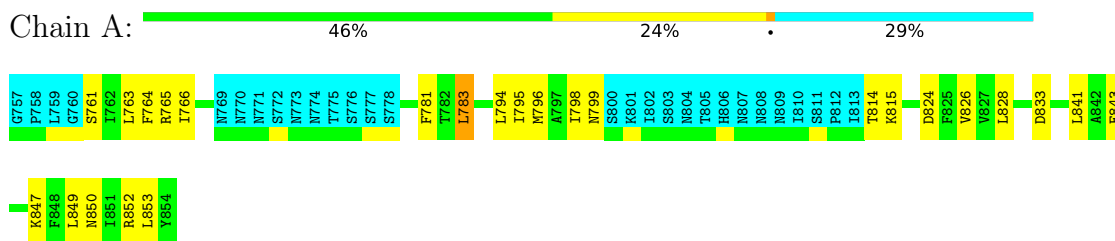Chain A: 40%   30%   •   29%

4.2.10　Score per residue for model 10

• Molecule 1: Cell division control protein 24

Chain A:

4.2.11　Score per residue for model 11

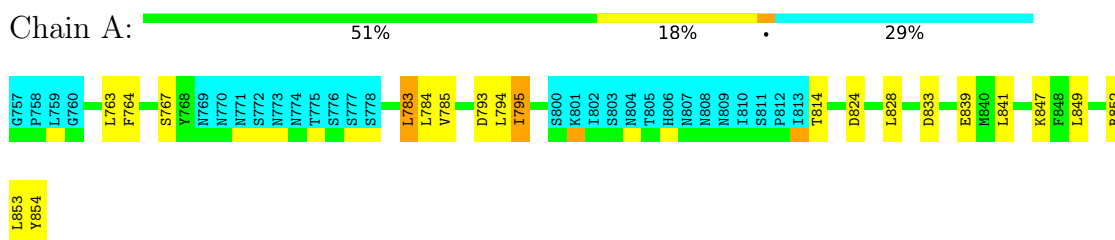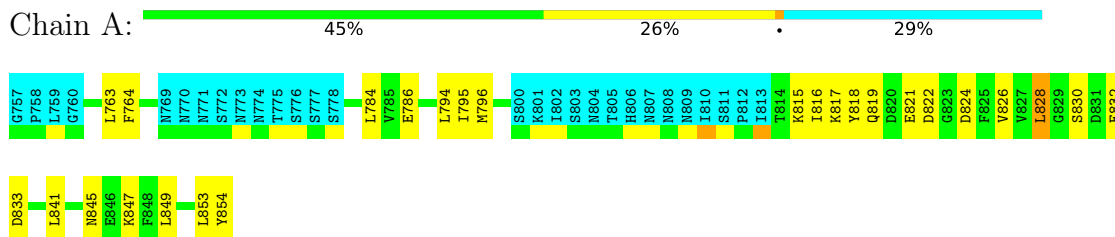• Molecule 1: Cell division control protein 24

Chain A:

4.2.12　Score per residue for model 12

• Molecule 1: Cell division control protein 24

Chain A:

4.2.13　Score per residue for model 13

• Molecule 1: Cell division control protein 24

Chain A:

Full wwPDB NMR Structure Validation Report

4.2.14 Score per residue for model 14

• Molecule 1: Cell division control protein 24

Chain A:

4.2.15 Score per residue for model 15 (medoid)

• Molecule 1: Cell division control protein 24

Chain A:

4.2.16 Score per residue for model 16

• Molecule 1: Cell division control protein 24

Chain A:

4.2.17 Score per residue for model 17

• Molecule 1: Cell division control protein 24

Chain A:

4.2.18 Score per residue for model 18

• Molecule 1: Cell division control protein 24
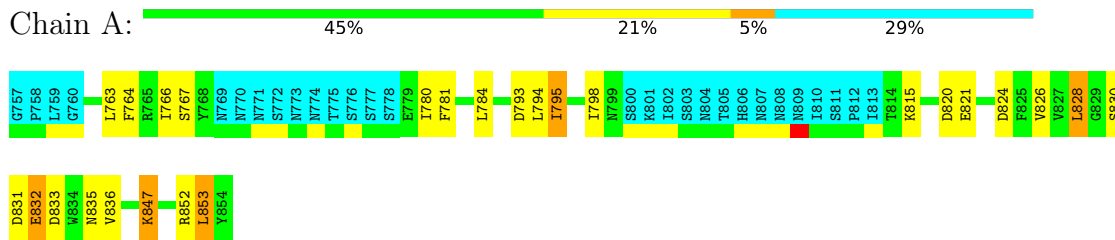
4.2.19 Score per residue for model 19
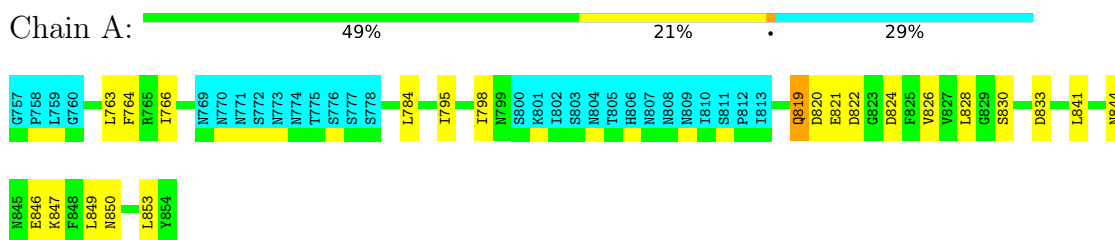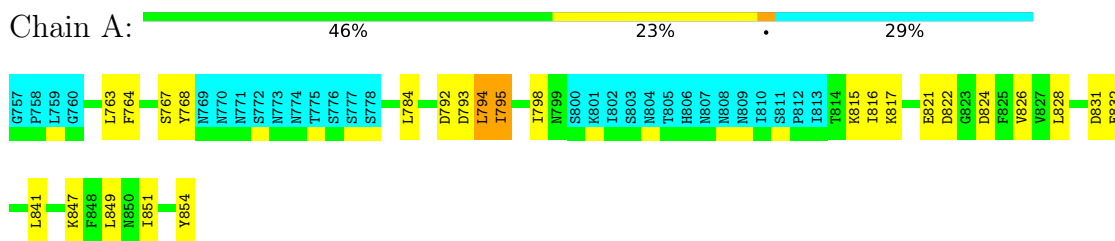
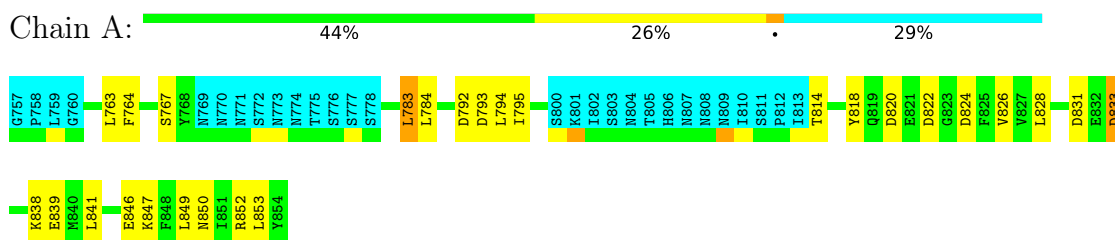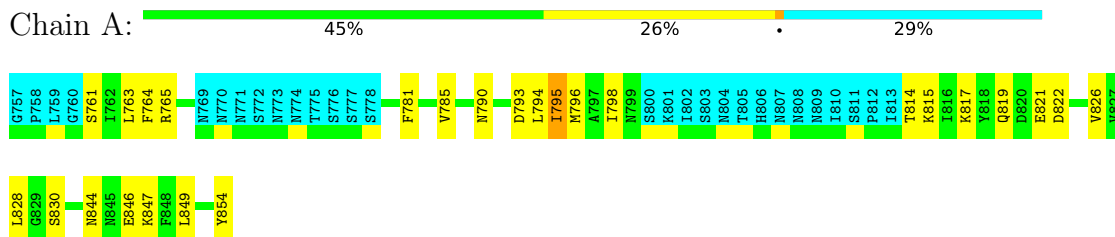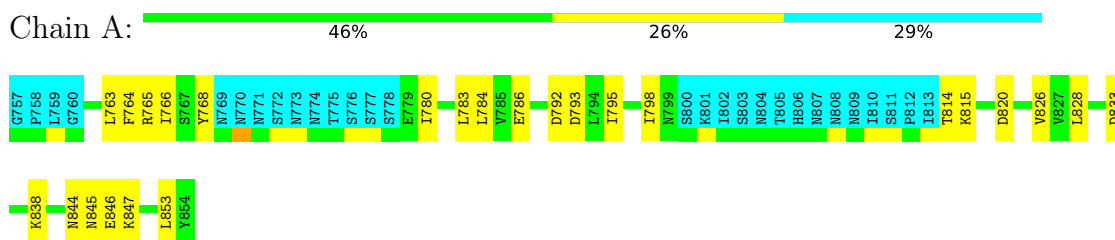• Molecule 1: Cell division control protein 24
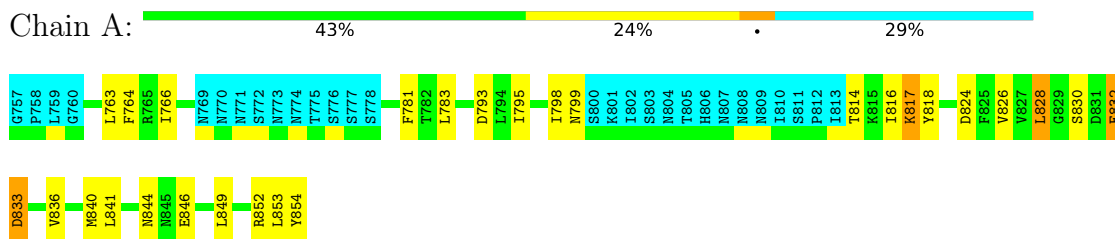
4.2.20 Score per residue for model 20

• Molecule 1: Cell division control protein 24
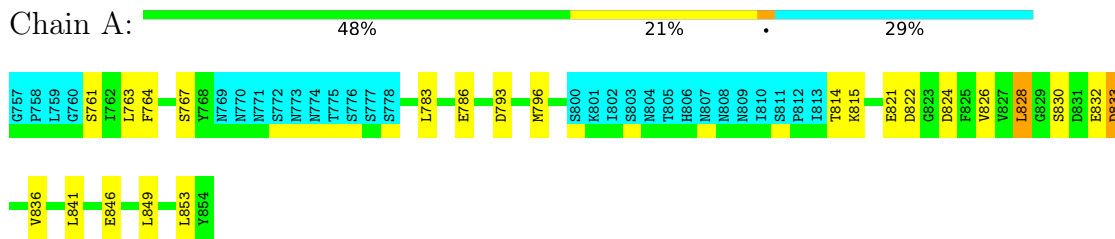
5 Refinement protocol and experimental data overview ⓘ

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

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
ARIA	refinement	
ARIA	structure solution	
CNS	structure solution	

No chemical shift data was provided.

6   Model quality ⓘ

6.1   Standard geometry ⓘ

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

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	589	575	575	4±1
All	All	11780	11500	11500	85

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 4.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
1:A:795:ILE:HA	1:A:798:ILE:HD12	0.67	1.66	18	9
1:A:785:VAL:HG11	1:A:794:LEU:HD12	0.67	1.66	8	5
1:A:841:LEU:HD21	1:A:849:LEU:HD13	0.66	1.66	1	16
1:A:828:LEU:HD12	1:A:833:ASP:HB3	0.65	1.68	11	10
1:A:818:TYR:HB2	1:A:849:LEU:HD21	0.61	1.73	19	2
1:A:766:ILE:HD13	1:A:798:ILE:HG23	0.59	1.75	13	3
1:A:817:LYS:HE3	1:A:827:VAL:HG22	0.53	1.80	1	2
1:A:763:LEU:HD21	1:A:782:THR:HG22	0.52	1.81	8	1
1:A:794:LEU:HD21	1:A:816:ILE:HD11	0.50	1.82	14	1
1:A:766:ILE:HD12	1:A:781:PHE:HB2	0.48	1.85	19	3
1:A:766:ILE:HD11	1:A:783:LEU:HD23	0.48	1.86	7	1
1:A:832:GLU:O	1:A:836:VAL:HG23	0.46	2.11	3	5
1:A:819:GLN:O	1:A:849:LEU:HD11	0.46	2.11	9	4
1:A:817:LYS:HG3	1:A:854:TYR:OH	0.46	2.11	16	4

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:766:ILE:O	1:A:780:ILE:HA	0.45	2.12	6	5
1:A:765:ARG:HA	1:A:781:PHE:O	0.43	2.14	8	3
1:A:787:LYS:HG2	1:A:838:LYS:HD2	0.42	1.90	6	1
1:A:853:LEU:HG	1:A:854:TYR:N	0.42	2.29	11	1
1:A:853:LEU:HD13	1:A:854:TYR:N	0.42	2.30	9	1
1:A:817:LYS:NZ	1:A:827:VAL:HG22	0.42	2.30	5	1
1:A:817:LYS:O	1:A:851:ILE:HA	0.42	2.15	6	2
1:A:785:VAL:HG13	1:A:789:TRP:CE3	0.41	2.50	9	2
1:A:841:LEU:HA	1:A:846:GLU:HB2	0.41	1.91	4	1
1:A:768:TYR:HB3	1:A:779:GLU:HB2	0.41	1.92	8	1
1:A:818:TYR:CD1	1:A:828:LEU:HB2	0.40	2.51	12	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	69/98 (70%)	65±2 (94±2%)	4±2 (6±2%)	0±0 (0±0%)	54	85
All	All	1380/1960 (70%)	1292 (94%)	86 (6%)	2 (0%)	54	85

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	788	VAL	1
1	A	847	LYS	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	65/91 (71%)	47±3 (73±4%)	18±3 (27±4%)	2	20
All	All	1300/1820 (71%)	943 (73%)	357 (27%)	2	20

All 46 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	764	PHE	20
1	A	826	VAL	19
1	A	763	LEU	18
1	A	847	LYS	16
1	A	828	LEU	15
1	A	824	ASP	15
1	A	814	THR	14
1	A	833	ASP	13
1	A	795	ILE	13
1	A	853	LEU	12
1	A	793	ASP	12
1	A	815	LYS	12
1	A	822	ASP	12
1	A	830	SER	11
1	A	784	LEU	11
1	A	783	LEU	10
1	A	821	GLU	10
1	A	846	GLU	10
1	A	832	GLU	9
1	A	831	ASP	9
1	A	852	ARG	8
1	A	794	LEU	8
1	A	761	SER	7
1	A	767	SER	7
1	A	796	MET	6
1	A	820	ASP	6
1	A	844	ASN	6
1	A	799	ASN	5
1	A	845	ASN	5
1	A	790	ASN	4
1	A	786	GLU	4
1	A	839	GLU	4
1	A	792	ASP	4
1	A	850	ASN	3
1	A	765	ARG	2
1	A	766	ILE	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	843	GLU	2
1	A	817	LYS	2
1	A	768	TYR	2
1	A	816	ILE	2
1	A	838	LYS	2
1	A	779	GLU	1
1	A	787	LYS	1
1	A	819	GLN	1
1	A	835	ASN	1
1	A	840	MET	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 monosaccharides 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 ⓘ

No chemical shift data were provided