



# Full wwPDB NMR Structure Validation Report ⓘ

Apr 26, 2016 – 10:55 PM BST

PDB ID : 2K6G  
Title : Solution structure of the DNA binding BRCT domain from the large subunit of human Replication Factor C  
Authors : Kobayashi, M.; Siegal, G.; Ab, E.; Bonvin, A.M.J.J.  
Deposited on : 2008-07-09

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

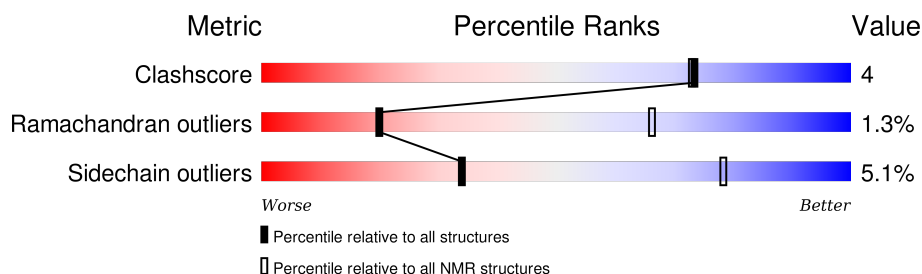
# 1 Overall quality at a glance

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	109	<div> <div style="width: 79%; background-color: green;"></div> <div style="width: 7%; background-color: yellow;"></div> <div style="width: 14%; background-color: cyan;"></div> </div> <div>79% 7% 14%</div>

## 2 Ensemble composition and analysis ⓘ

This entry contains 24 models. Model 7 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 bb conformation z-scores*.

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:380-A:388 (9)	0.40	1
2	A:399-A:483 (85)	0.54	7

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 4 clusters and 2 single-model clusters were found.

Cluster number	Models
1	2, 9, 10, 14, 15, 16, 18, 20, 21, 23
2	4, 6, 7, 13, 19, 24
3	1, 3, 8, 12
4	5, 17
Single-model clusters	11; 22

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1708 atoms, of which 867 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Replication factor C subunit 1.

Mol	Chain	Residues	Atoms						Trace
1	A	109	Total	C	H	N	O	S	0
			1708	522	867	150	167	2	

There are 3 discrepancies between the modelled and reference sequences:

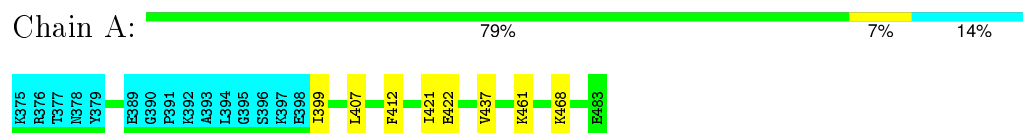
Chain	Residue	Modelled	Actual	Comment	Reference
A	481	ASN	-	EXPRESSION TAG	UNP P35251
A	482	LEU	-	EXPRESSION TAG	UNP P35251
A	483	GLU	-	EXPRESSION TAG	UNP P35251

## 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: Replication factor C subunit 1

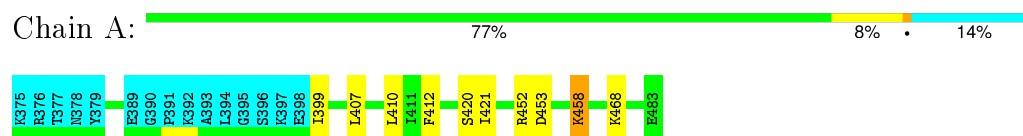


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section [4.1](#) above.

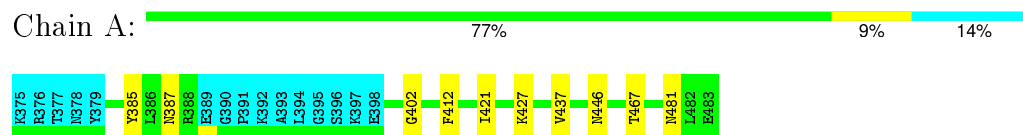
#### 4.2.1 Score per residue for model 1

- Molecule 1: Replication factor C subunit 1



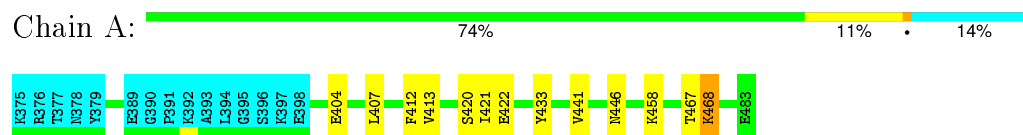
#### 4.2.2 Score per residue for model 2

- Molecule 1: Replication factor C subunit 1



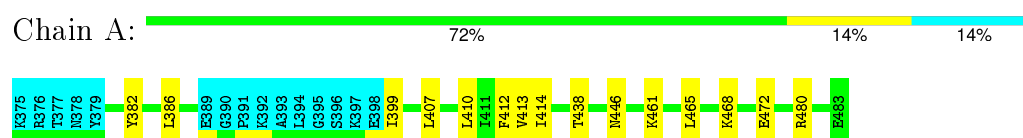
### 4.2.3 Score per residue for model 3

- Molecule 1: Replication factor C subunit 1



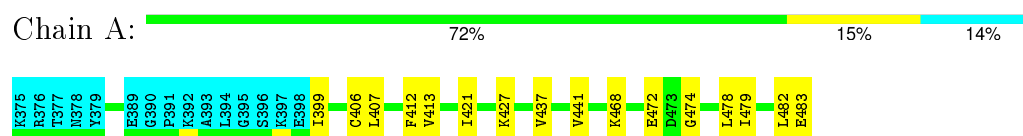
### 4.2.4 Score per residue for model 4

- Molecule 1: Replication factor C subunit 1



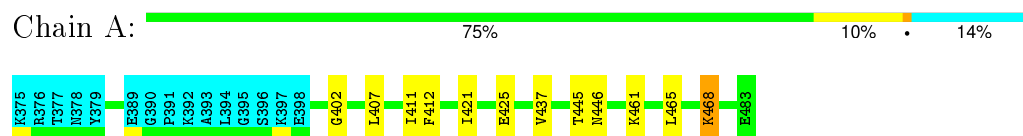
### 4.2.5 Score per residue for model 5

- Molecule 1: Replication factor C subunit 1



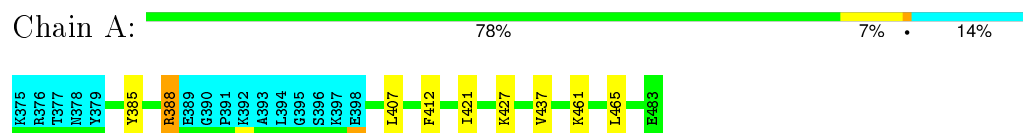
### 4.2.6 Score per residue for model 6

- Molecule 1: Replication factor C subunit 1



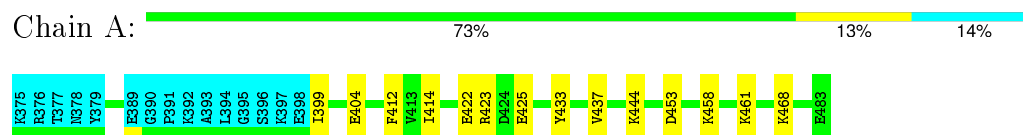
### 4.2.7 Score per residue for model 7 (medoid)

- Molecule 1: Replication factor C subunit 1



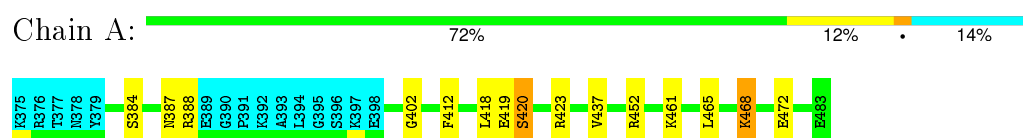
### 4.2.8 Score per residue for model 8

- Molecule 1: Replication factor C subunit 1



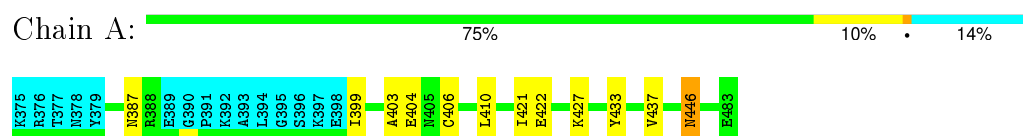
### 4.2.9 Score per residue for model 9

- Molecule 1: Replication factor C subunit 1



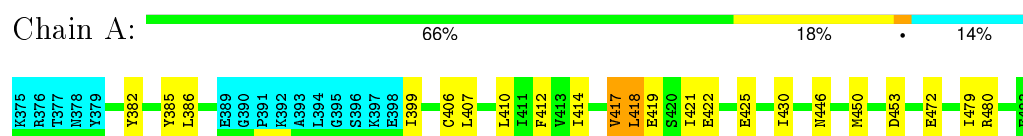
### 4.2.10 Score per residue for model 10

- Molecule 1: Replication factor C subunit 1



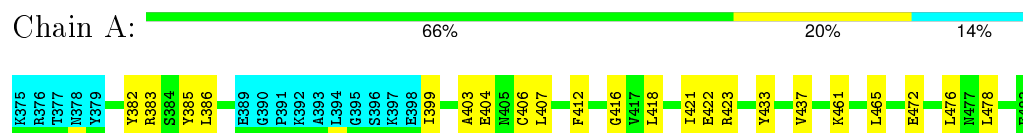
### 4.2.11 Score per residue for model 11

- Molecule 1: Replication factor C subunit 1



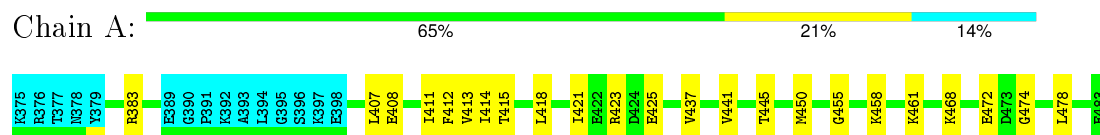
### 4.2.12 Score per residue for model 12

- Molecule 1: Replication factor C subunit 1



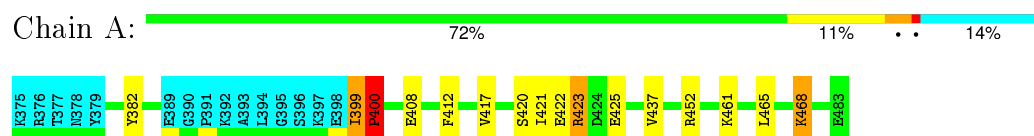
### 4.2.13 Score per residue for model 13

- Molecule 1: Replication factor C subunit 1



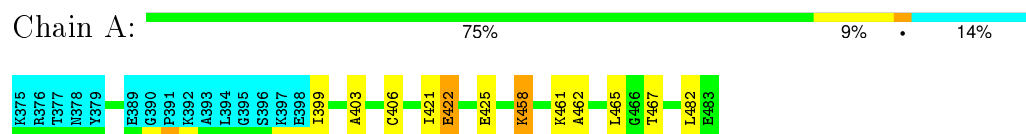
### 4.2.14 Score per residue for model 14

- Molecule 1: Replication factor C subunit 1



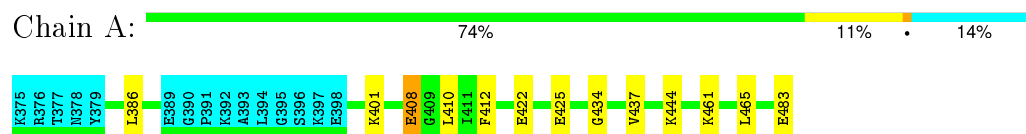
### 4.2.15 Score per residue for model 15

- Molecule 1: Replication factor C subunit 1



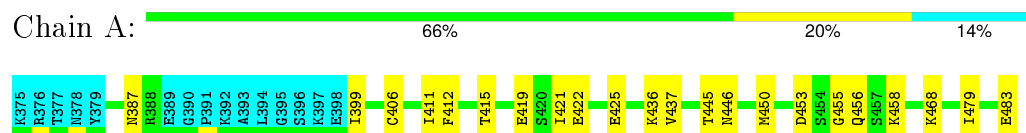
### 4.2.16 Score per residue for model 16

- Molecule 1: Replication factor C subunit 1



### 4.2.17 Score per residue for model 17

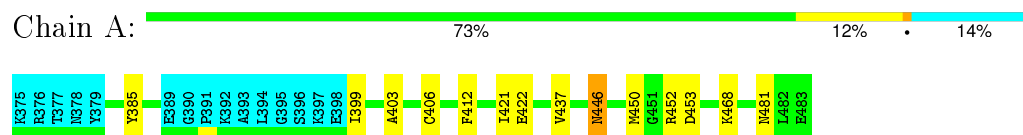
- Molecule 1: Replication factor C subunit 1





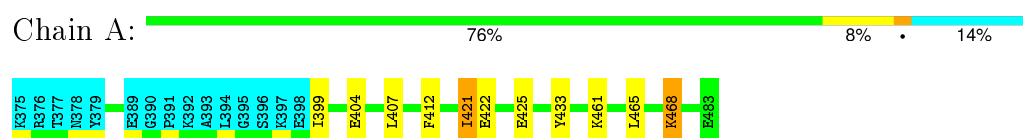
#### 4.2.18 Score per residue for model 18

- Molecule 1: Replication factor C subunit 1



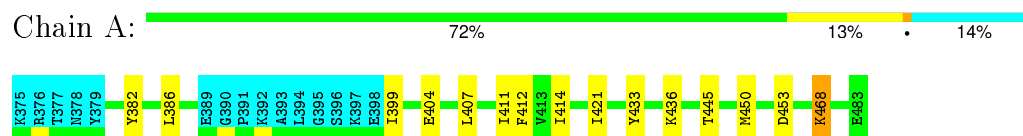
#### 4.2.19 Score per residue for model 19

- Molecule 1: Replication factor C subunit 1



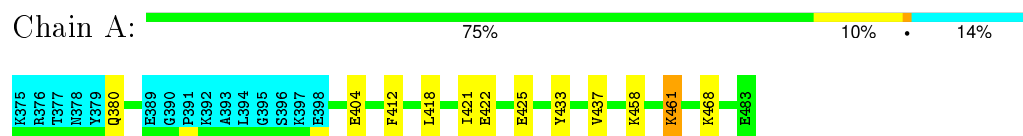
#### 4.2.20 Score per residue for model 20

- Molecule 1: Replication factor C subunit 1



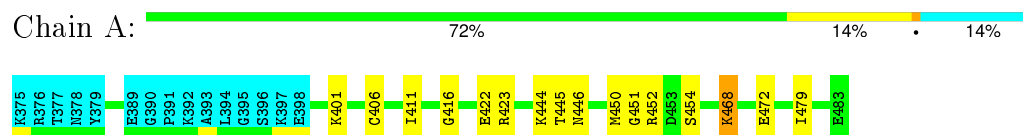
#### 4.2.21 Score per residue for model 21

- Molecule 1: Replication factor C subunit 1



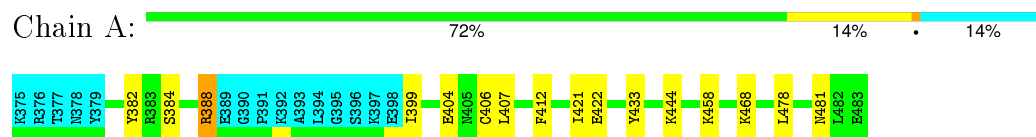
#### 4.2.22 Score per residue for model 22

- Molecule 1: Replication factor C subunit 1



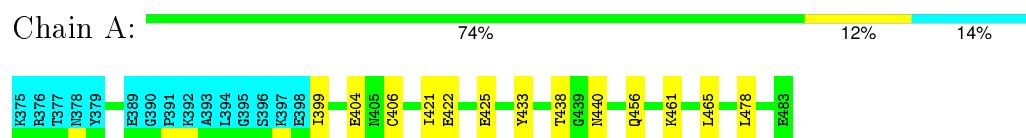
### 4.2.23 Score per residue for model 23

- Molecule 1: Replication factor C subunit 1



### 4.2.24 Score per residue for model 24

- Molecule 1: Replication factor C subunit 1



## 5 Refinement protocol and experimental data overview ⓘ

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

Of the 180 calculated structures, 24 were deposited, based on the following criterion: *bb conformation Z-score (whatif)*.

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

Software name	Classification	Version
CNS	refinement	
CYANA	structure solution	2.1

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	0.80±0.03	0±0/729 (0.0±0.0%)	0.68±0.04	0±0/975 (0.0±0.0%)
All	All	0.80	1/17496 (0.0%)	0.68	0/23400 (0.0%)

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±0.3
All	All	0	3

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	420	SER	N-CA	-5.29	1.35	1.46	9	1

There are no bond-angle outliers.

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	388	ARG	Sidechain	1
1	A	383	ARG	Sidechain	1
1	A	423	ARG	Sidechain	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	724	745	745	6±2
All	All	17376	17880	17880	145

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:406:CYS:SG	1:A:479:ILE:HA	0.70	2.26	17	3
1:A:404:GLU:HA	1:A:433:TYR:CD2	0.62	2.29	20	9
1:A:422:GLU:HA	1:A:423:ARG:NH2	0.59	2.10	22	1
1:A:421:ILE:HG23	1:A:425:GLU:HB2	0.59	1.74	6	3
1:A:399:ILE:HB	1:A:400:PRO:HD3	0.59	1.73	14	1
1:A:384:SER:O	1:A:388:ARG:HB3	0.58	1.99	9	1
1:A:399:ILE:HB	1:A:400:PRO:CD	0.57	2.28	14	1
1:A:416:GLY:HA3	1:A:451:GLY:O	0.57	2.00	22	1
1:A:461:LYS:O	1:A:465:LEU:HG	0.56	2.01	9	10
1:A:452:ARG:HA	1:A:452:ARG:CZ	0.55	2.31	9	1
1:A:468:LYS:N	1:A:468:LYS:HD3	0.53	2.18	20	2
1:A:411:ILE:HD12	1:A:436:LYS:HD3	0.53	1.81	17	1
1:A:450:MET:SD	1:A:454:SER:HA	0.53	2.44	22	1
1:A:446:ASN:O	1:A:468:LYS:HD2	0.52	2.03	22	2
1:A:461:LYS:HD2	1:A:461:LYS:N	0.52	2.20	24	1
1:A:413:VAL:HG21	1:A:441:VAL:HA	0.51	1.82	3	3
1:A:412:PHE:O	1:A:437:VAL:HA	0.51	2.05	18	13
1:A:450:MET:SD	1:A:453:ASP:HA	0.51	2.46	11	3
1:A:482:LEU:O	1:A:483:GLU:HG2	0.51	2.06	5	1
1:A:418:LEU:HD13	1:A:472:GLU:HB3	0.51	1.83	13	1
1:A:474:GLY:O	1:A:478:LEU:HG	0.50	2.06	13	2
1:A:382:TYR:O	1:A:386:LEU:HG	0.49	2.06	11	3
1:A:422:GLU:HB3	1:A:425:GLU:HG2	0.49	1.83	24	2
1:A:382:TYR:HA	1:A:385:TYR:CZ	0.49	2.42	11	1
1:A:418:LEU:HD12	1:A:421:ILE:O	0.49	2.07	13	1
1:A:403:ALA:HB3	1:A:406:CYS:SG	0.49	2.48	10	3
1:A:452:ARG:HA	1:A:452:ARG:NE	0.49	2.22	9	1

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:427:LYS:HG3	1:A:437:VAL:HG11	0.48	1.85	7	4
1:A:415:THR:O	1:A:450:MET:HA	0.48	2.09	17	2
1:A:384:SER:O	1:A:388:ARG:HB2	0.47	2.10	23	1
1:A:406:CYS:SG	1:A:479:ILE:HG12	0.47	2.49	5	1
1:A:422:GLU:HB2	1:A:425:GLU:HG2	0.47	1.86	19	5
1:A:468:LYS:HD3	1:A:468:LYS:N	0.47	2.24	19	3
1:A:407:LEU:HD22	1:A:412:PHE:CE1	0.47	2.45	13	6
1:A:416:GLY:H	1:A:423:ARG:NH2	0.46	2.08	12	1
1:A:411:ILE:O	1:A:445:THR:HA	0.46	2.11	13	1
1:A:414:ILE:HG23	1:A:423:ARG:NH2	0.46	2.26	8	1
1:A:406:CYS:SG	1:A:478:LEU:HB3	0.45	2.52	24	2
1:A:406:CYS:SG	1:A:482:LEU:HB2	0.45	2.51	15	1
1:A:387:ASN:HA	1:A:419:GLU:OE1	0.45	2.11	9	1
1:A:399:ILE:CB	1:A:400:PRO:HD3	0.45	2.42	14	1
1:A:384:SER:HB3	1:A:388:ARG:NH2	0.45	2.27	23	1
1:A:462:ALA:O	1:A:467:THR:HB	0.45	2.12	15	1
1:A:458:LYS:H	1:A:458:LYS:HD2	0.44	1.72	15	2
1:A:455:GLY:O	1:A:458:LYS:HG2	0.44	2.12	17	2
1:A:410:LEU:HD22	1:A:412:PHE:CE1	0.44	2.48	1	2
1:A:406:CYS:SG	1:A:478:LEU:HG	0.43	2.53	23	1
1:A:413:VAL:HG22	1:A:438:THR:HG23	0.43	1.90	4	1
1:A:461:LYS:HE3	1:A:461:LYS:N	0.43	2.28	21	1
1:A:407:LEU:HD22	1:A:412:PHE:CZ	0.43	2.48	4	4
1:A:417:VAL:O	1:A:452:ARG:HG3	0.42	2.14	14	1
1:A:418:LEU:CB	1:A:420:SER:HB3	0.42	2.43	9	1
1:A:407:LEU:HD22	1:A:412:PHE:CE2	0.42	2.49	11	2
1:A:472:GLU:O	1:A:476:LEU:HG	0.42	2.15	12	1
1:A:468:LYS:HD2	1:A:468:LYS:H	0.42	1.75	6	1
1:A:411:ILE:HD12	1:A:436:LYS:HB3	0.42	1.90	20	1
1:A:411:ILE:HB	1:A:445:THR:HA	0.42	1.92	20	4
1:A:427:LYS:HG3	1:A:437:VAL:CG1	0.41	2.45	5	1
1:A:461:LYS:CA	1:A:461:LYS:HE2	0.41	2.45	13	1
1:A:407:LEU:HD12	1:A:430:ILE:HA	0.41	1.92	11	1
1:A:461:LYS:N	1:A:461:LYS:HE2	0.41	2.29	13	1
1:A:417:VAL:C	1:A:419:GLU:H	0.41	2.19	11	1
1:A:382:TYR:O	1:A:386:LEU:HD23	0.41	2.15	4	1
1:A:399:ILE:CB	1:A:400:PRO:CD	0.41	2.97	14	1
1:A:408:GLU:HA	1:A:434:GLY:O	0.41	2.15	16	1
1:A:446:ASN:C	1:A:467:THR:HG23	0.41	2.36	2	2
1:A:388:ARG:HE	1:A:388:ARG:HA	0.41	1.74	7	1
1:A:410:LEU:HG	1:A:446:ASN:HD22	0.40	1.76	10	2

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:461:LYS:H	1:A:461:LYS:HD2	0.40	1.75	24	1
1:A:438:THR:HG23	1:A:440:ASN:O	0.40	2.16	24	1
1:A:417:VAL:HG22	1:A:419:GLU:H	0.40	1.76	11	1
1:A:382:TYR:HA	1:A:385:TYR:CE2	0.40	2.51	12	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	93/109 (85%)	86±2 (93±3%)	6±2 (6±2%)	1±1 (1±1%)	20	66
All	All	2232/2616 (85%)	2069 (93%)	133 (6%)	30 (1%)	20	66

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)
1	A	399	ILE	15
1	A	446	ASN	4
1	A	402	GLY	3
1	A	418	LEU	3
1	A	453	ASP	1
1	A	403	ALA	1
1	A	452	ARG	1
1	A	419	GLU	1
1	A	400	PRO	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	78/90 (87%)	74±2 (95±2%)	4±2 (5±2%)	34 78
All	All	1872/2160 (87%)	1776 (95%)	96 (5%)	34 78

All 28 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	468	LYS	16
1	A	421	ILE	15
1	A	422	GLU	8
1	A	458	LYS	6
1	A	472	GLU	5
1	A	444	LYS	4
1	A	414	ILE	4
1	A	387	ASN	3
1	A	408	GLU	3
1	A	481	ASN	3
1	A	423	ARG	3
1	A	452	ARG	2
1	A	456	GLN	2
1	A	483	GLU	2
1	A	453	ASP	2
1	A	461	LYS	2
1	A	480	ARG	2
1	A	382	TYR	2
1	A	385	TYR	2
1	A	401	LYS	2
1	A	400	PRO	1
1	A	386	LEU	1
1	A	380	GLN	1
1	A	425	GLU	1
1	A	410	LEU	1
1	A	388	ARG	1
1	A	417	VAL	1
1	A	446	ASN	1

### 6.3.3 RNA ⓘ

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