



# Full wwPDB NMR Structure Validation Report ⓘ

May 28, 2020 – 09:28 pm BST

PDB ID : 2CUF  
Title : Solution structure of the homeobox domain of the human hypothetical protein FLJ21616  
Authors : Ohnishi, S.; Kigawa, T.; Tomizawa, T.; Koshihara, S.; Inoue, M.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)  
Deposited on : 2005-05-26

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:

Cyrange : Kirchner and Güntert (2011)  
NmrClust : Kelley et al. (1996)  
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 : 2.11  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11

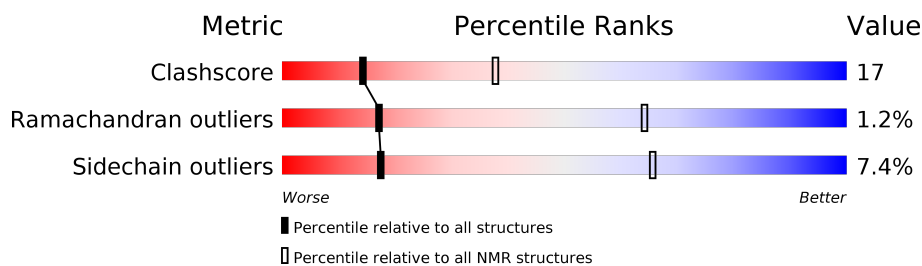
# 1 Overall quality at a glance

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	95	

## 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:14-A:75 (62)	0.15	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 1 single-model cluster was found.

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

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1488 atoms, of which 741 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called FLJ21616 protein.

Mol	Chain	Residues	Atoms					Trace	
			Total	C	H	N	O		S
1	A	95	1488	460	741	141	143	3	0

There are 13 discrepancies between the modelled and reference sequences:

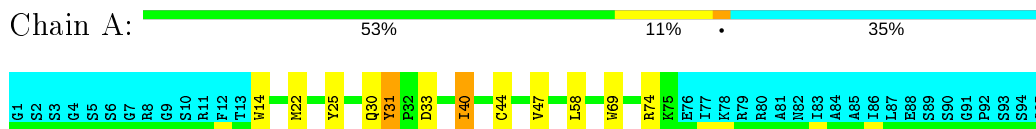
Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	CLONING ARTIFACT	UNP Q9H701
A	2	SER	-	CLONING ARTIFACT	UNP Q9H701
A	3	SER	-	CLONING ARTIFACT	UNP Q9H701
A	4	GLY	-	CLONING ARTIFACT	UNP Q9H701
A	5	SER	-	CLONING ARTIFACT	UNP Q9H701
A	6	SER	-	CLONING ARTIFACT	UNP Q9H701
A	7	GLY	-	CLONING ARTIFACT	UNP Q9H701
A	90	SER	-	CLONING ARTIFACT	UNP Q9H701
A	91	GLY	-	CLONING ARTIFACT	UNP Q9H701
A	92	PRO	-	CLONING ARTIFACT	UNP Q9H701
A	93	SER	-	CLONING ARTIFACT	UNP Q9H701
A	94	SER	-	CLONING ARTIFACT	UNP Q9H701
A	95	GLY	-	CLONING ARTIFACT	UNP Q9H701

## 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: FLJ21616 protein

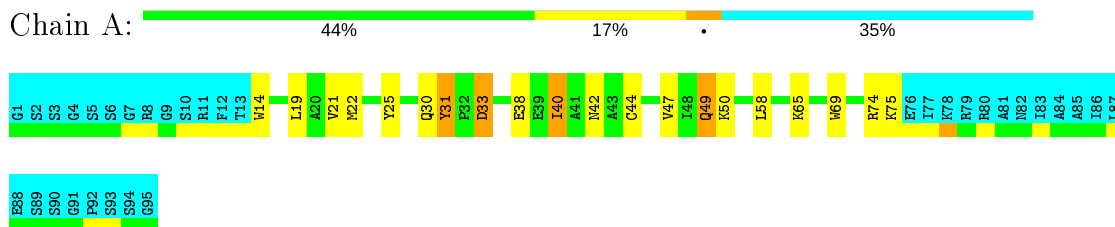


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

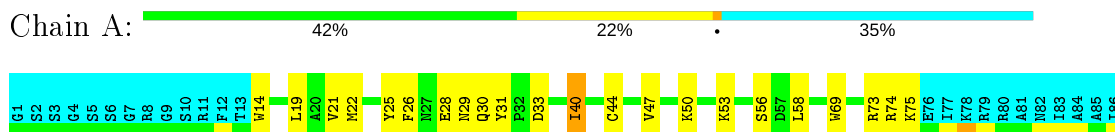
#### 4.2.1 Score per residue for model 1

- Molecule 1: FLJ21616 protein



#### 4.2.2 Score per residue for model 2

- Molecule 1: FLJ21616 protein



L87  
E88  
S89  
S90  
G91  
P92  
S93  
S94  
G95

### 4.2.3 Score per residue for model 3

- Molecule 1: FLJ21616 protein

Chain A:  54% 11% 35%

G1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23 S24 S25 S26 S27 S28 S29 S30 S31 S32 S33 S34 S35 S36 S37 S38 S39 S40 S41 S42 S43 S44 S45 S46 S47 S48 S49 S50 S51 S52 S53 S54 S55 S56 S57 S58 S59 S60 S61 S62 S63 S64 S65 S66 S67 S68 S69 S70 S71 S72 S73 S74 S75 S76 S77 S78 S79 S80 S81 S82 S83 S84 S85 S86 S87 S88 S89 S90 S91 S92 S93 S94 S95

### 4.2.4 Score per residue for model 4

- Molecule 1: FLJ21616 protein

Chain A:  42% 22% 35%

G1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23 S24 S25 S26 S27 S28 S29 S30 S31 S32 S33 S34 S35 S36 S37 S38 S39 S40 S41 S42 S43 S44 S45 S46 S47 S48 S49 S50 S51 S52 S53 S54 S55 S56 S57 S58 S59 S60 S61 S62 S63 S64 S65 S66 S67 S68 S69 S70 S71 S72 S73 S74 S75 S76 S77 S78 S79 S80 S81 S82 S83 S84 S85 S86 S87 S88 S89 S90 S91 S92 S93 S94 S95

S89  
S90  
G91  
P92  
S93  
S94  
G95

### 4.2.5 Score per residue for model 5

- Molecule 1: FLJ21616 protein

Chain A:  43% 20% 35%

G1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23 S24 S25 S26 S27 S28 S29 S30 S31 S32 S33 S34 S35 S36 S37 S38 S39 S40 S41 S42 S43 S44 S45 S46 S47 S48 S49 S50 S51 S52 S53 S54 S55 S56 S57 S58 S59 S60 S61 S62 S63 S64 S65 S66 S67 S68 S69 S70 S71 S72 S73 S74 S75 S76 S77 S78 S79 S80 S81 S82 S83 S84 S85 S86 S87 S88 S89 S90 S91 S92 S93 S94 S95

E88  
S89  
S90  
G91  
P92  
S93  
S94  
G95

### 4.2.6 Score per residue for model 6

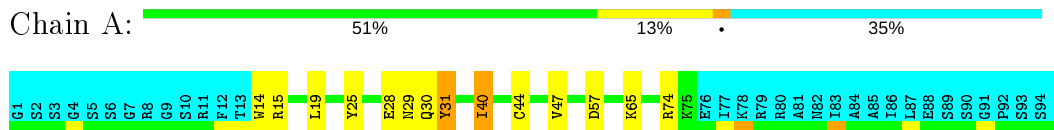
- Molecule 1: FLJ21616 protein

Chain A:  51% 13% 35%

G1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23 S24 S25 S26 S27 S28 S29 S30 S31 S32 S33 S34 S35 S36 S37 S38 S39 S40 S41 S42 S43 S44 S45 S46 S47 S48 S49 S50 S51 S52 S53 S54 S55 S56 S57 S58 S59 S60 S61 S62 S63 S64 S65 S66 S67 S68 S69 S70 S71 S72 S73 S74 S75 S76 S77 S78 S79 S80 S81 S82 S83 S84 S85 S86 S87 S88 S89 S90 S91 S92 S93 S94 S95

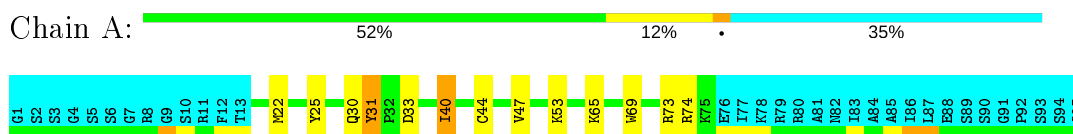
### 4.2.7 Score per residue for model 7

- Molecule 1: FLJ21616 protein



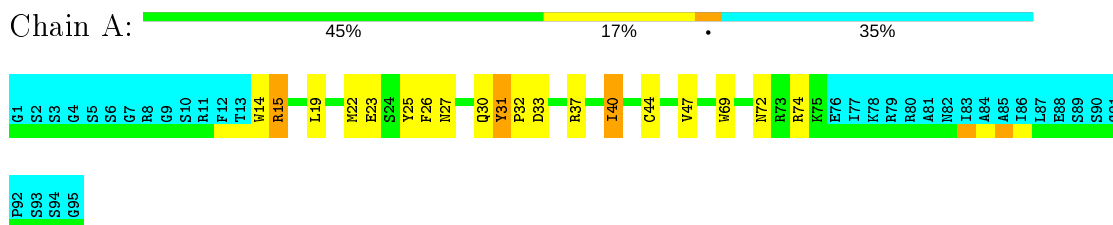
### 4.2.8 Score per residue for model 8

- Molecule 1: FLJ21616 protein



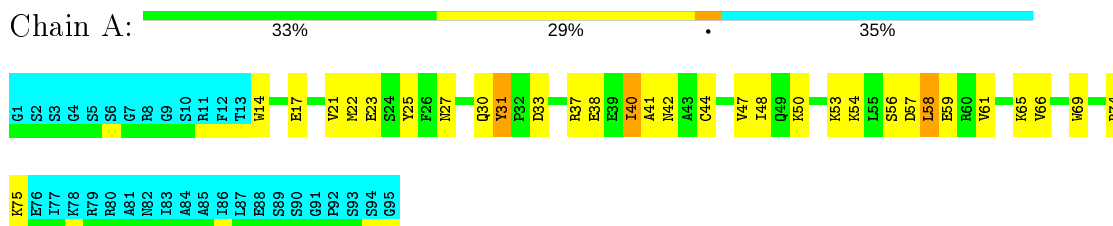
### 4.2.9 Score per residue for model 9

- Molecule 1: FLJ21616 protein



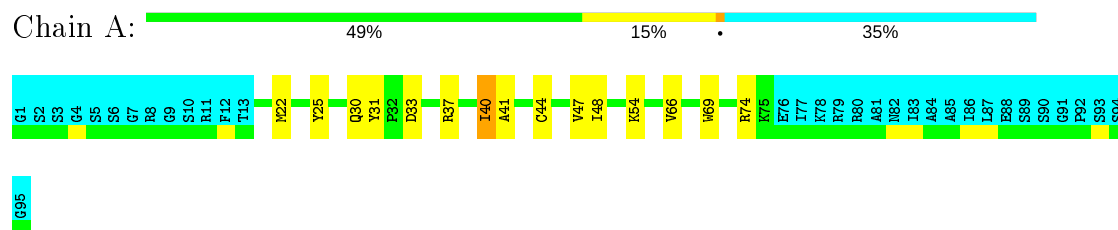
### 4.2.10 Score per residue for model 10

- Molecule 1: FLJ21616 protein



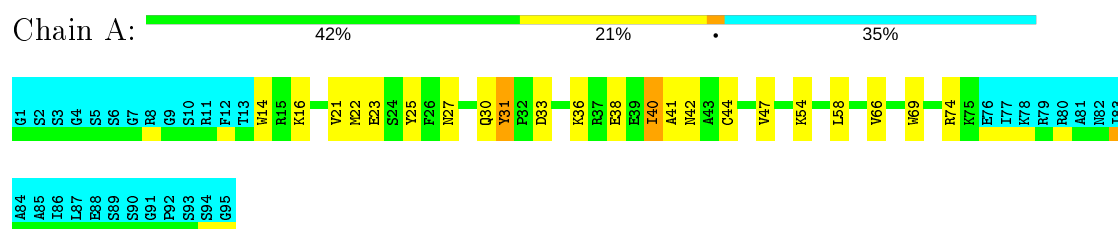
### 4.2.11 Score per residue for model 11

- Molecule 1: FLJ21616 protein



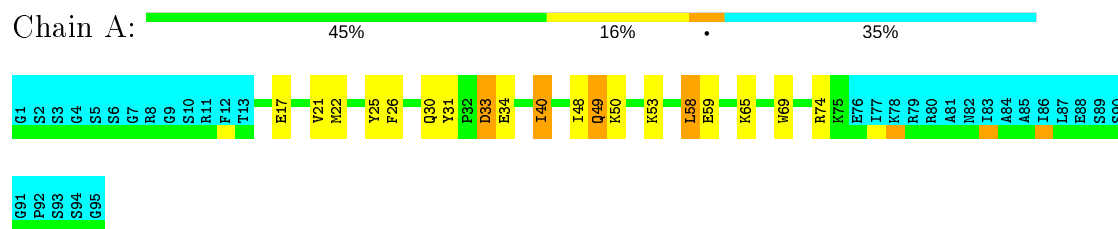
#### 4.2.12 Score per residue for model 12

- Molecule 1: FLJ21616 protein



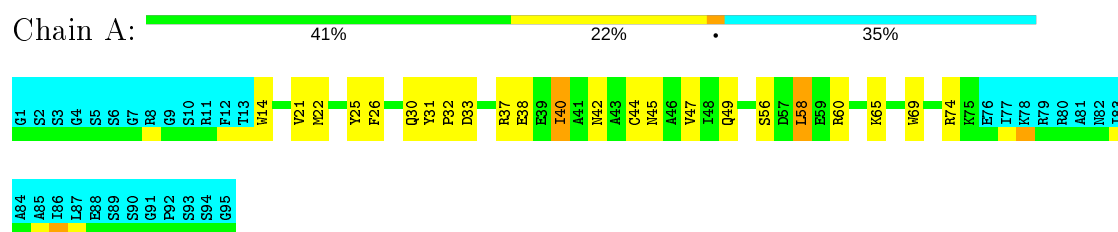
#### 4.2.13 Score per residue for model 13

- Molecule 1: FLJ21616 protein



#### 4.2.14 Score per residue for model 14

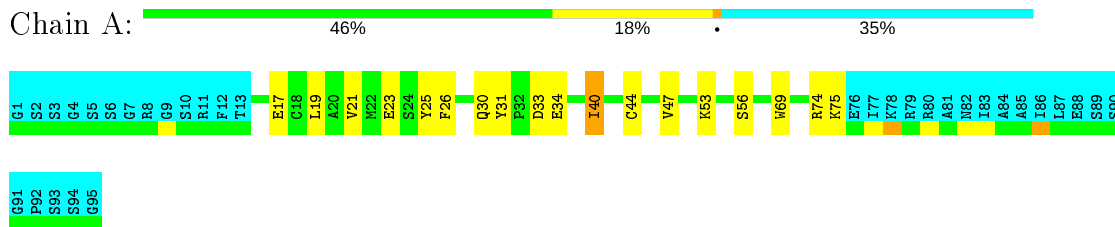
- Molecule 1: FLJ21616 protein





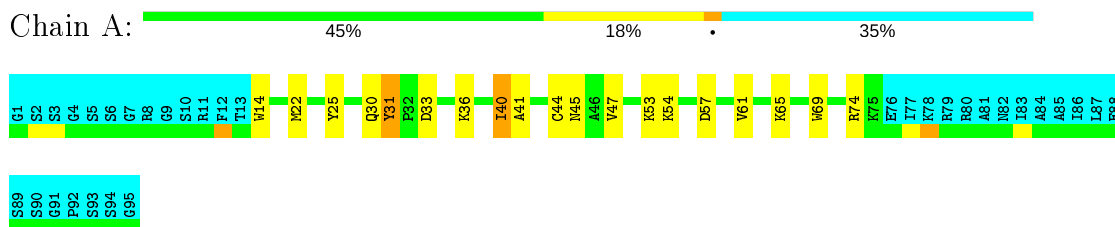
#### 4.2.15 Score per residue for model 15 (medoid)

- Molecule 1: FLJ21616 protein



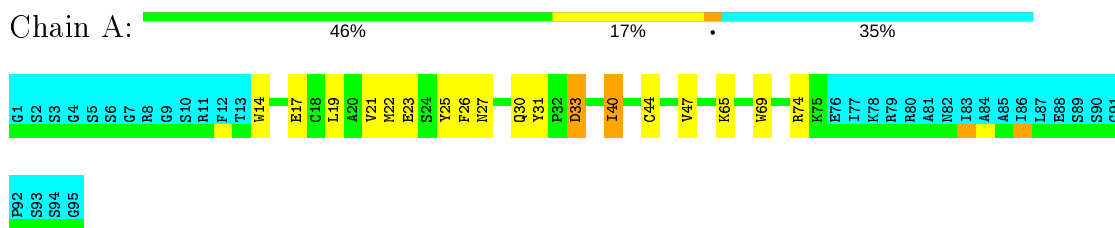
#### 4.2.16 Score per residue for model 16

- Molecule 1: FLJ21616 protein



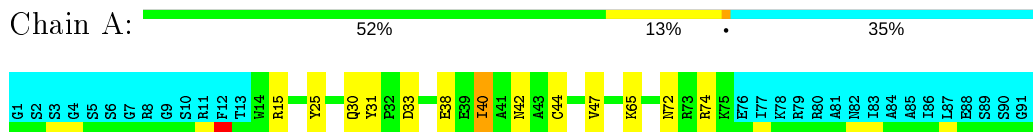
#### 4.2.17 Score per residue for model 17

- Molecule 1: FLJ21616 protein



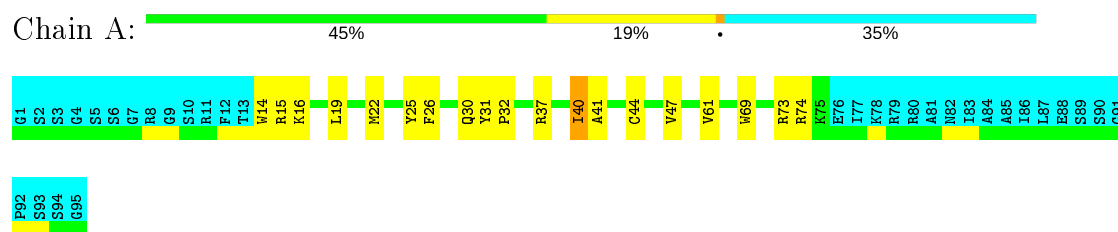
#### 4.2.18 Score per residue for model 18

- Molecule 1: FLJ21616 protein



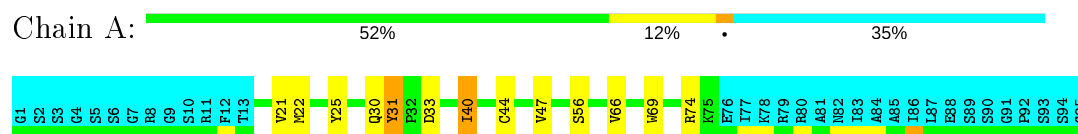
### 4.2.19 Score per residue for model 19

- Molecule 1: FLJ21616 protein



### 4.2.20 Score per residue for model 20

- Molecule 1: FLJ21616 protein



## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *torsion angle dynamics, restrained molecular dynamics*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function, structures with the least restraint violations*.

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

Software name	Classification	Version
CYANA	structure solution	2.0.17
CYANA	refinement	2.0.17

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

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 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	518	513	513	18±4
All	All	10360	10260	10260	357

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 17.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:40:ILE:O	1:A:40:ILE:HD13	0.90	1.66	8	6
1:A:40:ILE:HD13	1:A:40:ILE:O	0.89	1.66	18	6
1:A:31:TYR:CE2	1:A:74:ARG:NE	0.70	2.60	8	20
1:A:26:PHE:CZ	1:A:30:GLN:NE2	0.65	2.65	17	2
1:A:14:TRP:HB3	1:A:19:LEU:HD21	0.63	1.69	7	7
1:A:57:ASP:OD1	1:A:58:LEU:HD23	0.62	1.93	10	1
1:A:14:TRP:CZ3	1:A:22:MET:CE	0.60	2.84	12	3
1:A:26:PHE:CZ	1:A:30:GLN:OE1	0.60	2.55	19	3
1:A:21:VAL:HG12	1:A:40:ILE:HG12	0.60	1.73	12	6
1:A:22:MET:CE	1:A:69:TRP:CG	0.60	2.84	19	14
1:A:26:PHE:CE1	1:A:30:GLN:OE1	0.60	2.55	13	5
1:A:14:TRP:CZ3	1:A:22:MET:SD	0.60	2.95	3	1
1:A:14:TRP:CZ3	1:A:22:MET:HE2	0.59	2.32	10	5
1:A:17:GLU:O	1:A:21:VAL:HG23	0.58	1.98	6	4

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:26:PHE:CE2	1:A:30:GLN:OE1	0.58	2.57	17	1
1:A:25:TYR:CD1	1:A:40:ILE:CG1	0.57	2.87	5	14
1:A:58:LEU:HD23	1:A:58:LEU:N	0.55	2.16	10	1
1:A:22:MET:SD	1:A:69:TRP:CD2	0.55	3.00	19	8
1:A:14:TRP:CE3	1:A:22:MET:CE	0.54	2.91	12	2
1:A:31:TYR:CD2	1:A:74:ARG:CZ	0.54	2.90	4	18
1:A:17:GLU:OE1	1:A:48:ILE:HG22	0.54	2.03	10	3
1:A:31:TYR:CD2	1:A:74:ARG:NE	0.54	2.76	11	15
1:A:40:ILE:HD13	1:A:40:ILE:C	0.54	2.23	8	8
1:A:58:LEU:N	1:A:58:LEU:CD2	0.54	2.70	13	1
1:A:25:TYR:CD1	1:A:40:ILE:HG12	0.53	2.38	5	9
1:A:40:ILE:C	1:A:40:ILE:HD13	0.53	2.24	11	2
1:A:22:MET:CE	1:A:69:TRP:CD1	0.53	2.92	19	5
1:A:41:ALA:HB1	1:A:61:VAL:O	0.53	2.03	10	3
1:A:58:LEU:CD2	1:A:58:LEU:N	0.52	2.72	14	2
1:A:41:ALA:HB2	1:A:66:VAL:HG21	0.52	1.81	12	4
1:A:31:TYR:CD1	1:A:31:TYR:N	0.51	2.77	10	9
1:A:31:TYR:N	1:A:31:TYR:CD1	0.51	2.78	16	8
1:A:25:TYR:CD1	1:A:40:ILE:HG13	0.50	2.41	13	5
1:A:38:GLU:O	1:A:42:ASN:ND2	0.50	2.45	10	5
1:A:37:ARG:O	1:A:40:ILE:HG22	0.50	2.06	11	2
1:A:23:GLU:O	1:A:27:ASN:ND2	0.50	2.45	17	5
1:A:45:ASN:ND2	1:A:61:VAL:O	0.50	2.45	16	1
1:A:30:GLN:HB2	1:A:31:TYR:CD1	0.49	2.43	10	13
1:A:49:GLN:CD	1:A:49:GLN:C	0.49	2.71	13	2
1:A:23:GLU:OE2	1:A:69:TRP:CZ2	0.48	2.66	15	1
1:A:25:TYR:CD2	1:A:40:ILE:HB	0.48	2.44	18	18
1:A:48:ILE:CD1	1:A:59:GLU:O	0.48	2.61	13	1
1:A:22:MET:SD	1:A:69:TRP:CG	0.48	3.07	19	4
1:A:44:CYS:O	1:A:47:VAL:CG2	0.48	2.62	9	18
1:A:48:ILE:HD12	1:A:59:GLU:HB3	0.48	1.84	5	2
1:A:30:GLN:HB2	1:A:31:TYR:CE1	0.47	2.43	20	5
1:A:25:TYR:CG	1:A:40:ILE:HG13	0.47	2.45	10	9
1:A:25:TYR:CE2	1:A:40:ILE:HB	0.47	2.45	5	5
1:A:60:ARG:O	1:A:65:LYS:CE	0.46	2.63	14	1
1:A:44:CYS:O	1:A:47:VAL:HG22	0.46	2.11	12	2
1:A:18:CYS:SG	1:A:48:ILE:HG21	0.46	2.50	4	1
1:A:22:MET:CE	1:A:69:TRP:CD2	0.45	2.99	3	1
1:A:25:TYR:CE1	1:A:40:ILE:HG12	0.45	2.46	5	1
1:A:22:MET:HE1	1:A:69:TRP:CG	0.45	2.46	16	3
1:A:22:MET:HA	1:A:40:ILE:HD11	0.45	1.88	3	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:26:PHE:CE2	1:A:30:GLN:NE2	0.45	2.84	17	1
1:A:30:GLN:O	1:A:74:ARG:NH1	0.45	2.50	6	16
1:A:40:ILE:CG2	1:A:41:ALA:N	0.45	2.79	16	1
1:A:18:CYS:SG	1:A:48:ILE:CG2	0.45	3.04	4	1
1:A:32:PRO:O	1:A:37:ARG:NH1	0.45	2.50	19	4
1:A:19:LEU:O	1:A:23:GLU:CD	0.44	2.55	15	1
1:A:25:TYR:CD2	1:A:36:LYS:HD2	0.43	2.48	16	1
1:A:40:ILE:HD13	1:A:44:CYS:HG	0.43	1.73	8	1
1:A:22:MET:HE1	1:A:69:TRP:CD1	0.43	2.49	1	4
1:A:22:MET:SD	1:A:69:TRP:CE3	0.43	3.12	10	1
1:A:49:GLN:CD	1:A:50:LYS:N	0.42	2.72	1	2
1:A:57:ASP:CG	1:A:58:LEU:HD23	0.42	2.34	10	1
1:A:14:TRP:CE3	1:A:22:MET:HE1	0.42	2.48	12	1
1:A:25:TYR:CE2	1:A:36:LYS:HD2	0.42	2.50	6	2
1:A:15:ARG:N	1:A:15:ARG:HD2	0.42	2.28	7	2
1:A:14:TRP:CD2	1:A:22:MET:HE1	0.42	2.50	12	1
1:A:45:ASN:O	1:A:49:GLN:N	0.42	2.53	14	1
1:A:14:TRP:CE3	1:A:22:MET:HE2	0.41	2.49	14	2
1:A:44:CYS:O	1:A:48:ILE:HG23	0.41	2.16	11	1
1:A:15:ARG:N	1:A:15:ARG:CD	0.41	2.83	19	1
1:A:26:PHE:CE1	1:A:30:GLN:NE2	0.41	2.88	15	1
1:A:31:TYR:CE2	1:A:74:ARG:CD	0.41	3.04	7	2
1:A:40:ILE:CD1	1:A:40:ILE:C	0.40	2.89	8	1
1:A:40:ILE:HG21	1:A:66:VAL:HG11	0.40	1.93	20	1
1:A:28:GLU:HG3	1:A:29:ASN:N	0.40	2.31	5	3
1:A:41:ALA:HB2	1:A:66:VAL:CG2	0.40	2.47	10	1
1:A:40:ILE:C	1:A:40:ILE:CD1	0.40	2.89	9	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	62/95 (65%)	61±1 (98±1%)	0±1 (1±1%)	1±0 (1±1%)	17	64
All	All	1240/1900 (65%)	1216 (98%)	9 (1%)	15 (1%)	17	64

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	33	ASP	15

### 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	55/79 (70%)	51±2 (93±3%)	4±2 (7±3%)	17 65
All	All	1100/1580 (70%)	1019 (93%)	81 (7%)	17 65

All 17 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	40	ILE	20
1	A	31	TYR	10
1	A	53	LYS	7
1	A	56	SER	6
1	A	33	ASP	5
1	A	75	LYS	5
1	A	57	ASP	4
1	A	54	LYS	4
1	A	58	LEU	3
1	A	72	ASN	3
1	A	16	LYS	3
1	A	34	GLU	2
1	A	50	LYS	2
1	A	15	ARG	2
1	A	49	GLN	2
1	A	73	ARG	2
1	A	36	LYS	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