



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 21, 2016 – 06:58 AM EDT

PDB ID : 4Y91  
Title : Crystal Structure of a Thermotoga maritima Hfq homolog  
Authors : Randolph, P.S.; Patterson, J.; Mura, C.  
Deposited on : 2015-02-16  
Resolution : 2.66 Å(reported)

This is a wwPDB X-ray 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  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
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  
Mogul : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027107  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0122  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027107

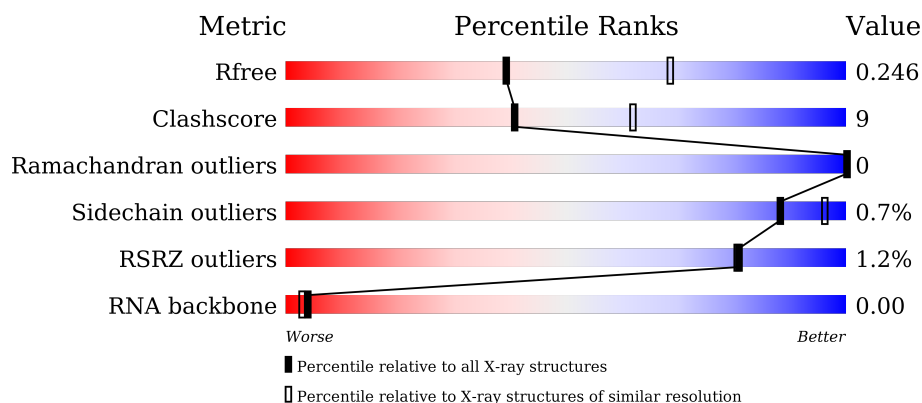
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.66 Å.

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)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	91344	3152 (2.70-2.62)
Clashscore	102246	3524 (2.70-2.62)
Ramachandran outliers	100387	3469 (2.70-2.62)
Sidechain outliers	100360	3469 (2.70-2.62)
RSRZ outliers	91569	3161 (2.70-2.62)
RNA backbone	2183	1001 (3.08-2.24)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	95	<div> <div>53%</div> <div>16%</div> <div>•</div> <div>31%</div> </div>
1	B	95	<div> <div>49%</div> <div>19%</div> <div>•</div> <div>31%</div> </div>
1	C	95	<div> <div>53%</div> <div>16%</div> <div>•</div> <div>32%</div> </div>
1	D	95	<div> <div>55%</div> <div>15%</div> <div>•</div> <div>31%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	E	95	<div><div></div><div>56%14%31%</div></div>
1	F	95	<div><div>%</div><div><div></div><div>60%14%23%</div></div></div>
1	G	95	<div><div></div><div>61%8%31%</div></div>
1	H	95	<div><div>2%</div><div><div></div><div>49%19%31%</div></div></div>
1	I	95	<div><div></div><div>53%16%32%</div></div>
1	J	95	<div><div></div><div>57%13%31%</div></div>
1	K	95	<div><div></div><div>56%14%31%</div></div>
1	L	95	<div><div></div><div>59%9%31%</div></div>
2	N	6	<div><div>50%</div><div><div></div><div>67%33%</div></div></div>
2	O	6	<div><div>67%</div><div><div></div><div>17%33%50%</div></div></div>

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 6777 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called RNA-binding protein Hfq.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	66	Total	C	N	O	S	0	0	0
			543	353	92	97	1			
1	B	66	Total	C	N	O	S	0	0	0
			540	350	92	97	1			
1	C	65	Total	C	N	O	S	0	0	0
			532	344	91	96	1			
1	D	66	Total	C	N	O	S	0	0	0
			540	350	92	97	1			
1	E	66	Total	C	N	O	S	0	0	0
			540	350	92	97	1			
1	F	73	Total	C	N	O	S	4	0	0
			597	387	102	106	2			
1	G	66	Total	C	N	O	S	0	0	0
			540	350	92	97	1			
1	H	66	Total	C	N	O	S	0	0	0
			540	350	92	97	1			
1	I	65	Total	C	N	O	S	0	0	0
			532	344	91	96	1			
1	J	66	Total	C	N	O	S	0	0	0
			540	350	92	97	1			
1	K	66	Total	C	N	O	S	0	0	0
			540	350	92	97	1			
1	L	66	Total	C	N	O	S	0	0	0
			540	350	92	97	1			

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP Q9WYZ6
A	-1	SER	-	expression tag	UNP Q9WYZ6
A	0	HIS	-	expression tag	UNP Q9WYZ6
B	-2	GLY	-	expression tag	UNP Q9WYZ6
B	-1	SER	-	expression tag	UNP Q9WYZ6

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	HIS	-	expression tag	UNP Q9WYZ6
C	-2	GLY	-	expression tag	UNP Q9WYZ6
C	-1	SER	-	expression tag	UNP Q9WYZ6
C	0	HIS	-	expression tag	UNP Q9WYZ6
D	-2	GLY	-	expression tag	UNP Q9WYZ6
D	-1	SER	-	expression tag	UNP Q9WYZ6
D	0	HIS	-	expression tag	UNP Q9WYZ6
E	-2	GLY	-	expression tag	UNP Q9WYZ6
E	-1	SER	-	expression tag	UNP Q9WYZ6
E	0	HIS	-	expression tag	UNP Q9WYZ6
F	-2	GLY	-	expression tag	UNP Q9WYZ6
F	-1	SER	-	expression tag	UNP Q9WYZ6
F	0	HIS	-	expression tag	UNP Q9WYZ6
G	-2	GLY	-	expression tag	UNP Q9WYZ6
G	-1	SER	-	expression tag	UNP Q9WYZ6
G	0	HIS	-	expression tag	UNP Q9WYZ6
H	-2	GLY	-	expression tag	UNP Q9WYZ6
H	-1	SER	-	expression tag	UNP Q9WYZ6
H	0	HIS	-	expression tag	UNP Q9WYZ6
I	-2	GLY	-	expression tag	UNP Q9WYZ6
I	-1	SER	-	expression tag	UNP Q9WYZ6
I	0	HIS	-	expression tag	UNP Q9WYZ6
J	-2	GLY	-	expression tag	UNP Q9WYZ6
J	-1	SER	-	expression tag	UNP Q9WYZ6
J	0	HIS	-	expression tag	UNP Q9WYZ6
K	-2	GLY	-	expression tag	UNP Q9WYZ6
K	-1	SER	-	expression tag	UNP Q9WYZ6
K	0	HIS	-	expression tag	UNP Q9WYZ6
L	-2	GLY	-	expression tag	UNP Q9WYZ6
L	-1	SER	-	expression tag	UNP Q9WYZ6
L	0	HIS	-	expression tag	UNP Q9WYZ6

- Molecule 2 is a RNA chain called RNA (5'-R(P\*UP\*UP\*UP\*UP\*U)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	N	6	Total	C	N	O	P	0	0	0
			120	54	12	48	6			
2	O	6	Total	C	N	O	P	0	0	0
			120	54	12	48	6			

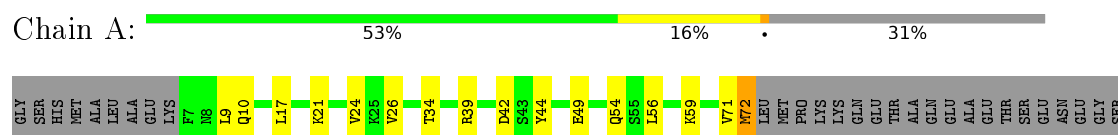
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total 1	O 1	0	0
3	F	1	Total 1	O 1	0	0
3	G	3	Total 3	O 3	0	0
3	H	1	Total 1	O 1	0	0
3	I	1	Total 1	O 1	0	0
3	J	4	Total 4	O 4	0	0
3	N	1	Total 1	O 1	0	0
3	O	1	Total 1	O 1	0	0

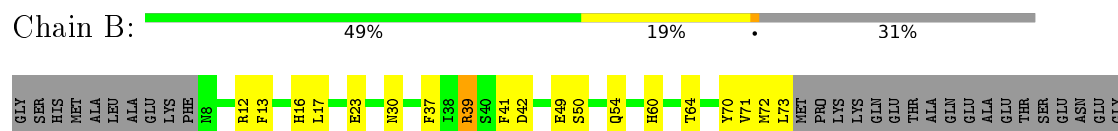
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-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. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

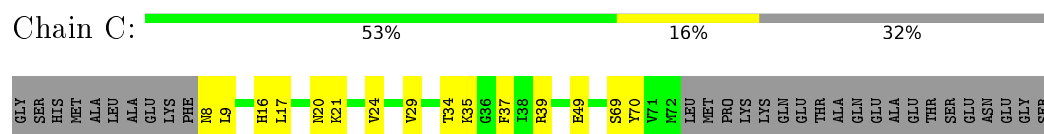
- Molecule 1: RNA-binding protein Hfq



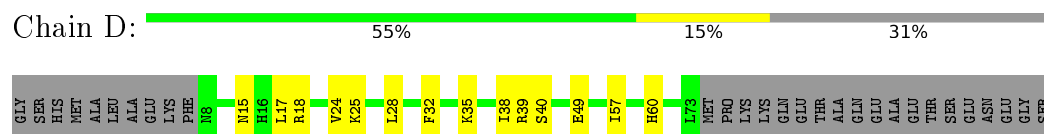
- Molecule 1: RNA-binding protein Hfq



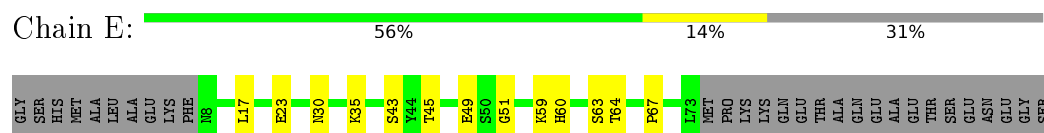
- Molecule 1: RNA-binding protein Hfq



- Molecule 1: RNA-binding protein Hfq



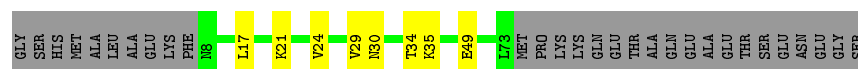
- Molecule 1: RNA-binding protein Hfq



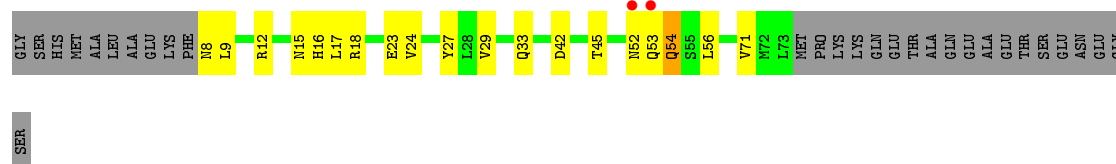
- Molecule 1: RNA-binding protein Hfq



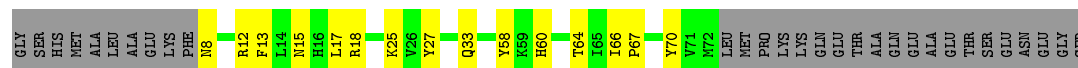
- Molecule 1: RNA-binding protein Hfq



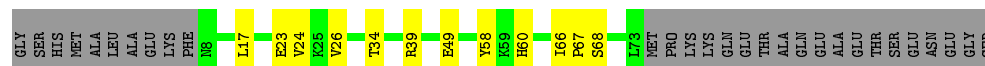
- Molecule 1: RNA-binding protein Hfq



- Molecule 1: RNA-binding protein Hfq



- Molecule 1: RNA-binding protein Hfq



- Molecule 1: RNA-binding protein Hfq

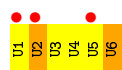


- Molecule 1: RNA-binding protein Hfq

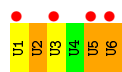


- Molecule 2: RNA (5'-R(P\*UP\*UP\*UP\*UP\*U)-3')





- Molecule 2: RNA (5'-R(P\*UP\*UP\*UP\*UP\*UP\*U)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	39.08Å 133.50Å 206.18Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	56.03 – 2.66 81.59 – 2.66	Depositor EDS
% Data completeness (in resolution range)	84.6 (56.03-2.66) 84.7 (81.59-2.66)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.84 (at 2.65Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.184 , 0.246 0.182 , 0.246	Depositor DCC
$R_{free}$ test set	1372 reflections (5.02%)	DCC
Wilson B-factor (Å <sup>2</sup> )	32.6	Xtriage
Anisotropy	0.152	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 25.1	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	1 of 27360 reflections (0.004%)	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	6777	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.50% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

## 5 Model quality

### 5.1 Standard geometry

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 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	0.49	0/554	0.68	0/747
1	B	0.54	0/550	0.76	1/742 (0.1%)
1	C	0.51	0/542	0.63	0/731
1	D	0.50	0/550	0.65	0/742
1	E	0.52	0/550	0.67	0/742
1	F	0.99	3/609 (0.5%)	0.84	2/820 (0.2%)
1	G	0.57	0/550	0.70	0/742
1	H	0.51	0/550	0.65	0/742
1	I	0.47	0/542	0.64	0/731
1	J	0.51	0/550	0.68	0/742
1	K	0.54	0/550	0.68	0/742
1	L	0.65	0/550	0.83	3/742 (0.4%)
2	N	0.61	0/131	1.08	0/200
2	O	0.54	0/131	0.86	0/200
All	All	0.59	3/6909 (0.0%)	0.72	6/9365 (0.1%)

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 outliers	#Planarity outliers
1	B	0	1
1	H	0	1
All	All	0	2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	2	ALA	C-O	-6.12	1.11	1.23
1	F	1	MET	C-O	-5.17	1.13	1.23
1	F	7	PHE	CB-CG	-5.08	1.42	1.51

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	L	9	LEU	CB-CG-CD2	8.59	125.60	111.00
1	F	8	ASN	N-CA-C	-6.73	92.84	111.00
1	B	39	ARG	NE-CZ-NH2	-5.87	117.37	120.30
1	F	9	LEU	CA-CB-CG	5.78	128.60	115.30
1	L	9	LEU	N-CA-C	5.56	126.02	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	72	MET	Peptide
1	H	52	ASN	Peptide

## 5.2 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 the 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 within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	543	0	550	11	0
1	B	540	0	552	19	0
1	C	532	0	541	12	0
1	D	540	0	552	11	0
1	E	540	0	552	10	0
1	F	597	0	609	10	0
1	G	540	0	552	6	0
1	H	540	0	552	13	0
1	I	532	0	541	14	0
1	J	540	0	552	9	0
1	K	540	0	552	11	0
1	L	540	0	552	9	0
2	N	120	0	61	2	0
2	O	120	0	61	5	0
3	A	1	0	0	1	0
3	F	1	0	0	0	0
3	G	3	0	0	0	0
3	H	1	0	0	1	0
3	I	1	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	J	4	0	0	1	0
3	N	1	0	0	0	0
3	O	1	0	0	0	0
All	All	6777	0	6779	116	0

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 9.

The worst 5 of 116 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:F:5:GLU:HB3	1:F:6:LYS:O	1.62	0.99
1:B:60:HIS:HB3	2:O:5:U:H5''	1.54	0.89
1:F:15:ASN:OD1	1:F:18:ARG:NH1	2.09	0.84
1:E:49:GLU:OE2	1:E:51:GLY:O	1.96	0.84
1:A:21:LYS:NZ	1:G:49:GLU:OE1	2.11	0.83

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 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 X-ray entries followed by that with respect to entries of similar resolution.

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	64/95 (67%)	61 (95%)	3 (5%)	0	100	100
1	B	64/95 (67%)	59 (92%)	5 (8%)	0	100	100
1	C	63/95 (66%)	58 (92%)	5 (8%)	0	100	100
1	D	64/95 (67%)	59 (92%)	5 (8%)	0	100	100
1	E	64/95 (67%)	60 (94%)	4 (6%)	0	100	100
1	F	71/95 (75%)	64 (90%)	7 (10%)	0	100	100
1	G	64/95 (67%)	62 (97%)	2 (3%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	H	64/95 (67%)	60 (94%)	4 (6%)	0	100	100
1	I	63/95 (66%)	61 (97%)	2 (3%)	0	100	100
1	J	64/95 (67%)	61 (95%)	3 (5%)	0	100	100
1	K	64/95 (67%)	60 (94%)	4 (6%)	0	100	100
1	L	64/95 (67%)	62 (97%)	2 (3%)	0	100	100
All	All	773/1140 (68%)	727 (94%)	46 (6%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 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 X-ray entries followed by that with respect to entries of similar resolution.

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	62/85 (73%)	61 (98%)	1 (2%)	70	89
1	B	62/85 (73%)	62 (100%)	0	100	100
1	C	61/85 (72%)	61 (100%)	0	100	100
1	D	62/85 (73%)	62 (100%)	0	100	100
1	E	62/85 (73%)	62 (100%)	0	100	100
1	F	67/85 (79%)	65 (97%)	2 (3%)	48	77
1	G	62/85 (73%)	62 (100%)	0	100	100
1	H	62/85 (73%)	60 (97%)	2 (3%)	46	74
1	I	61/85 (72%)	61 (100%)	0	100	100
1	J	62/85 (73%)	62 (100%)	0	100	100
1	K	62/85 (73%)	62 (100%)	0	100	100
1	L	62/85 (73%)	62 (100%)	0	100	100
All	All	747/1020 (73%)	742 (99%)	5 (1%)	88	96

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	72	MET
1	F	6	LYS
1	F	9	LEU
1	H	23	GLU
1	H	54	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	16	HIS
1	H	10	GLN
1	H	53	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	N	5/6 (83%)	5 (100%)	0
2	O	5/6 (83%)	4 (80%)	0
All	All	10/12 (83%)	9 (90%)	0

5 of 9 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	N	2	U
2	N	3	U
2	N	4	U
2	N	5	U
2	N	6	U

There are no RNA pucker outliers to report.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.



## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	66/95 (69%)	-0.24	0 100 100	10, 27, 46, 60	0
1	B	66/95 (69%)	-0.11	0 100 100	15, 29, 51, 63	0
1	C	65/95 (68%)	-0.27	0 100 100	11, 26, 43, 59	0
1	D	66/95 (69%)	-0.21	0 100 100	12, 27, 42, 49	0
1	E	66/95 (69%)	-0.14	0 100 100	15, 29, 51, 64	0
1	F	73/95 (76%)	-0.04	1 (1%) 78 76	13, 29, 56, 80	5 (6%)
1	G	66/95 (69%)	-0.25	0 100 100	10, 24, 45, 53	0
1	H	66/95 (69%)	-0.04	2 (3%) 54 52	14, 32, 52, 71	0
1	I	65/95 (68%)	-0.12	0 100 100	14, 30, 49, 62	0
1	J	66/95 (69%)	-0.27	0 100 100	11, 27, 45, 51	0
1	K	66/95 (69%)	-0.10	0 100 100	13, 29, 50, 61	0
1	L	66/95 (69%)	-0.22	0 100 100	11, 25, 43, 59	0
2	N	6/6 (100%)	1.89	3 (50%) 0 0	36, 37, 41, 44	6 (100%)
2	O	6/6 (100%)	2.47	4 (66%) 0 0	36, 39, 42, 42	6 (100%)
All	All	809/1152 (70%)	-0.13	10 (1%) 81 80	10, 28, 49, 80	17 (2%)

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	53	GLN	4.0
2	O	1	U	3.6
1	F	5	GLU	3.5
2	N	1	U	3.1
2	O	5	U	2.9

## 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

## 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.