



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 01:42 AM GMT

PDB ID : 2DYY  
Title : Crystal structure of putative translation initiation inhibitor PH0854 from *Pyrococcus horikoshii*  
Authors : Ihsanawati; Kishishita, S.; Murayama, K.; Chen, L.; Liu, Z.J.; Wang, B.C.; Shirouzu, M.; Bessho, Y.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)  
Deposited on : 2006-09-19  
Resolution : 2.60 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	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

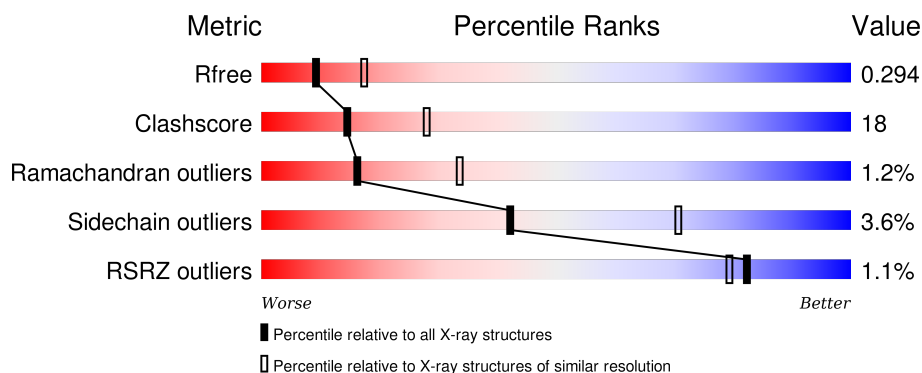
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.60 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	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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	126	<div> <div>62%</div> <div>37%</div> <div>..</div> </div>
1	B	126	<div> <div>62%</div> <div>30%</div> <div>6%</div> </div>
1	C	126	<div> <div>66%</div> <div>29%</div> <div>6%</div> </div>
1	D	126	<div> <div>74%</div> <div>26%</div> </div>
1	E	126	<div> <div>70%</div> <div>23%</div> <div>5%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	126	
1	G	126	
1	H	126	
1	I	126	
1	J	126	
1	K	126	
1	L	126	

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 11461 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 UPF0076 protein PH0854.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	125	Total	C	N	O	S	0	0	0
			965	623	155	185	2			
1	B	119	Total	C	N	O	S	0	0	0
			928	601	149	177	1			
1	C	119	Total	C	N	O	S	0	0	0
			921	596	147	177	1			
1	D	126	Total	C	N	O	S	0	0	0
			970	626	156	186	2			
1	E	120	Total	C	N	O	S	0	0	0
			922	598	146	177	1			
1	F	118	Total	C	N	O		0	0	0
			905	586	145	174				
1	G	125	Total	C	N	O	S	0	0	0
			971	628	155	187	1			
1	H	122	Total	C	N	O	S	0	0	0
			930	599	147	182	2			
1	I	116	Total	C	N	O		0	0	0
			895	580	142	173				
1	J	124	Total	C	N	O	S	0	0	0
			957	619	154	182	2			
1	K	123	Total	C	N	O	S	0	0	0
			944	609	151	183	1			
1	L	119	Total	C	N	O	S	0	0	0
			920	595	147	177	1			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	19	Total	O	0	0
			19	19		
2	B	27	Total	O	0	0
			27	27		

*Continued on next page...*

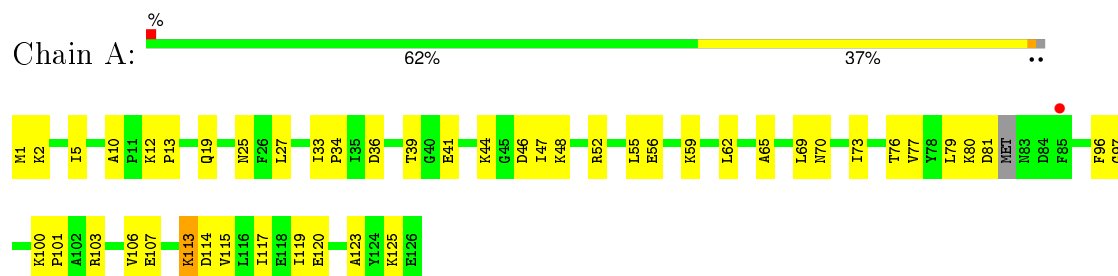
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	C	22	Total 22	O 22	0	0
2	D	22	Total 22	O 22	0	0
2	E	21	Total 21	O 21	0	0
2	F	14	Total 14	O 14	0	0
2	G	29	Total 29	O 29	0	0
2	H	19	Total 19	O 19	0	0
2	I	10	Total 10	O 10	0	0
2	J	19	Total 19	O 19	0	0
2	K	24	Total 24	O 24	0	0
2	L	7	Total 7	O 7	0	0

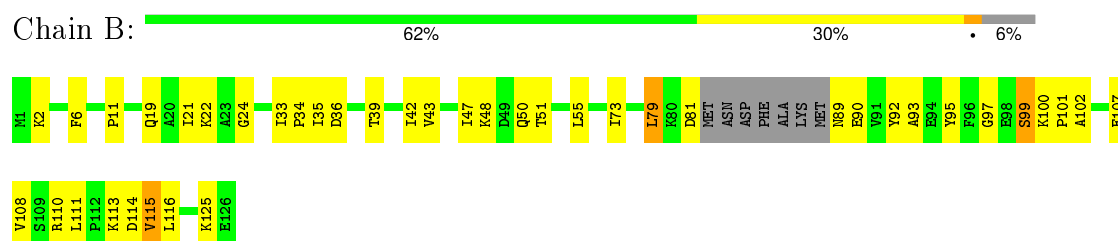
### 3 Residue-property plots

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 ( $\text{RSRZ} > 2$ ). 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.

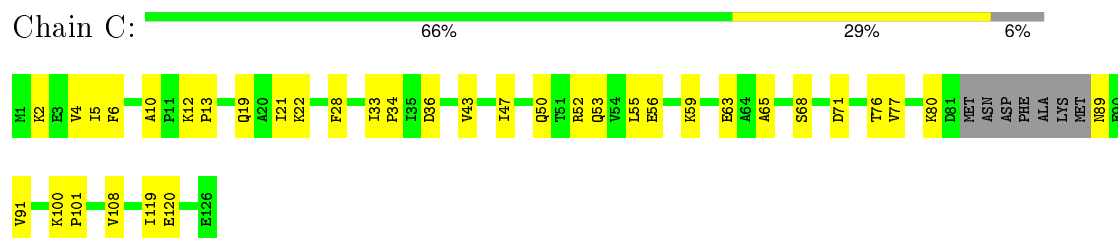
- Molecule 1: UPF0076 protein PH0854



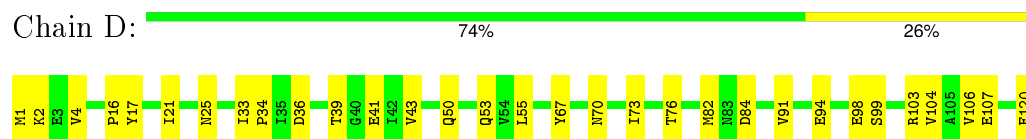
- Molecule 1: UPF0076 protein PH0854




- Molecule 1: UPF0076 protein PH0854

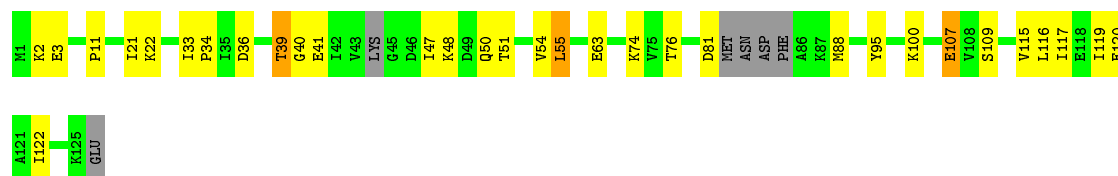


- Molecule 1: UPF0076 protein PH0854



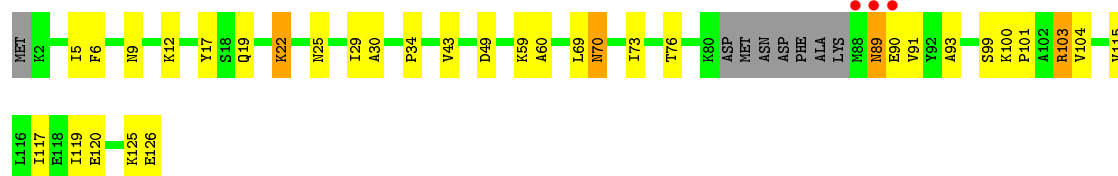
- Molecule 1: UPF0076 protein PH0854

Chain E:  70% 23% 5%



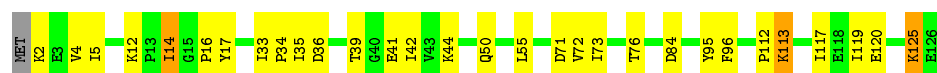
- Molecule 1: UPF0076 protein PH0854

Chain F:  2% 67% 24% 6%



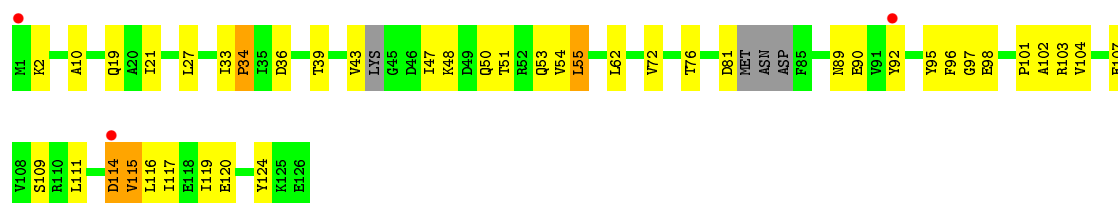
- Molecule 1: UPF0076 protein PH0854

Chain G:  75% 21% 2%



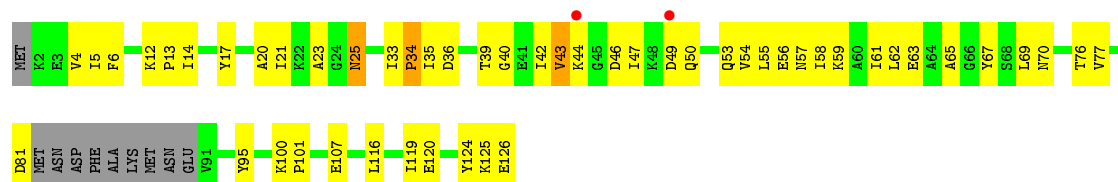
- Molecule 1: UPF0076 protein PH0854

Chain H:  2% 63% 30% 2%



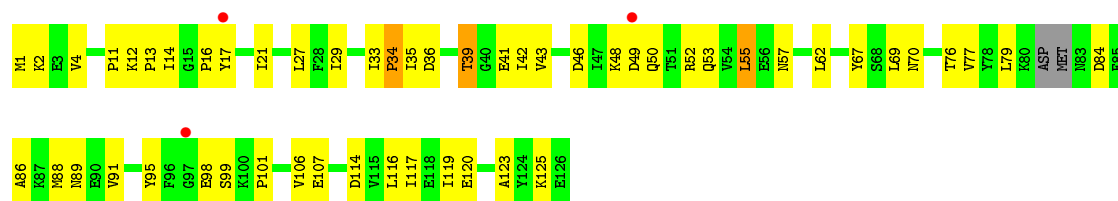
- Molecule 1: UPF0076 protein PH0854

Chain I:  2% 52% 38% 8%

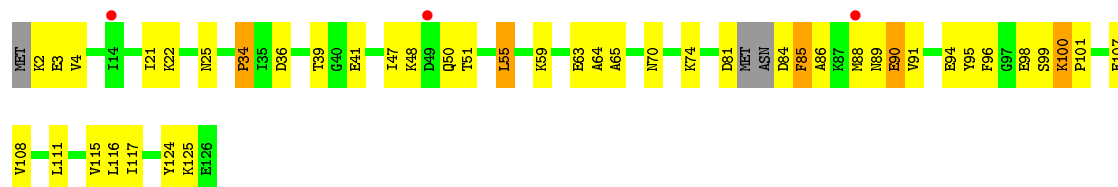


- Molecule 1: UPF0076 protein PH0854

Chain J:  2% 56% 40% 2%



• Molecule 1: UPF0076 protein PH0854



• Molecule 1: UPF0076 protein PH0854





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	92.49 Å 43.09 Å 170.21 Å 90.00° 91.51° 90.00°	Depositor
Resolution (Å)	29.73 – 2.60 48.92 – 2.60	Depositor EDS
% Data completeness (in resolution range)	96.7 (29.73-2.60) 96.8 (48.92-2.60)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	6.47 (at 2.61 Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.209 , 0.295 0.209 , 0.294	Depositor DCC
$R_{free}$ test set	2048 reflections (5.01%)	DCC
Wilson B-factor (Å <sup>2</sup> )	40.0	Xtriage
Anisotropy	0.587	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 64.2	EDS
Estimated twinning fraction	0.017 for h,-k,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	2 of 40909 reflections (0.005%)	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	11461	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 34.61 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 6.6413e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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 [i](#)

### 5.1 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 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	0.38	0/980	0.59	0/1324
1	B	0.43	0/943	0.61	0/1274
1	C	0.41	0/936	0.62	0/1267
1	D	0.39	0/986	0.59	0/1334
1	E	0.40	0/936	0.59	0/1267
1	F	0.39	0/920	0.60	0/1248
1	G	0.44	0/988	0.64	0/1337
1	H	0.38	0/944	0.59	0/1280
1	I	0.35	0/910	0.58	0/1234
1	J	0.38	0/972	0.56	0/1313
1	K	0.36	0/959	0.58	0/1299
1	L	0.38	0/935	0.56	0/1266
All	All	0.39	0/11409	0.59	0/15443

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 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	965	0	983	44	0
1	B	928	0	957	43	0
1	C	921	0	937	27	0
1	D	970	0	986	31	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	922	0	939	27	0
1	F	905	0	910	28	0
1	G	971	0	983	31	0
1	H	930	0	921	37	0
1	I	895	0	904	52	0
1	J	957	0	979	47	0
1	K	944	0	947	46	0
1	L	920	0	935	40	0
2	A	19	0	0	2	0
2	B	27	0	0	0	0
2	C	22	0	0	2	0
2	D	22	0	0	0	0
2	E	21	0	0	3	0
2	F	14	0	0	1	0
2	G	29	0	0	2	0
2	H	19	0	0	0	0
2	I	10	0	0	0	0
2	J	19	0	0	0	0
2	K	24	0	0	1	0
2	L	7	0	0	1	0
All	All	11461	0	11381	405	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 18.

The worst 5 of 405 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:C:89:ASN:HD22	1:C:91:VAL:HG23	1.22	1.02
1:C:12:LYS:HD2	1:C:13:PRO:HD2	1.45	0.97
1:J:106:VAL:HG23	1:K:111:LEU:HD23	1.51	0.92
1:H:76:THR:HB	1:H:120:GLU:HG2	1.50	0.90
1:A:36:ASP:HB3	1:A:39:THR:HG22	1.56	0.87

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	121/126 (96%)	110 (91%)	10 (8%)	1 (1%)	24	46
1	B	115/126 (91%)	113 (98%)	2 (2%)	0	100	100
1	C	115/126 (91%)	111 (96%)	4 (4%)	0	100	100
1	D	124/126 (98%)	115 (93%)	9 (7%)	0	100	100
1	E	114/126 (90%)	110 (96%)	4 (4%)	0	100	100
1	F	114/126 (90%)	101 (89%)	10 (9%)	3 (3%)	7	11
1	G	123/126 (98%)	115 (94%)	7 (6%)	1 (1%)	24	46
1	H	116/126 (92%)	101 (87%)	11 (10%)	4 (3%)	5	7
1	I	112/126 (89%)	98 (88%)	12 (11%)	2 (2%)	11	21
1	J	120/126 (95%)	103 (86%)	14 (12%)	3 (2%)	7	12
1	K	119/126 (94%)	105 (88%)	11 (9%)	3 (2%)	7	12
1	L	115/126 (91%)	104 (90%)	11 (10%)	0	100	100
All	All	1408/1512 (93%)	1286 (91%)	105 (8%)	17 (1%)	16	33

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	44	LYS
1	F	90	GLU
1	H	115	VAL
1	J	39	THR
1	F	43	VAL

### 5.3.2 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 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	101/106 (95%)	98 (97%)	3 (3%)	48	76
1	B	99/106 (93%)	95 (96%)	4 (4%)	38	67
1	C	97/106 (92%)	95 (98%)	2 (2%)	61	85
1	D	101/106 (95%)	99 (98%)	2 (2%)	63	85
1	E	97/106 (92%)	90 (93%)	7 (7%)	18	35
1	F	94/106 (89%)	90 (96%)	4 (4%)	35	64
1	G	102/106 (96%)	97 (95%)	5 (5%)	31	57
1	H	96/106 (91%)	95 (99%)	1 (1%)	82	94
1	I	94/106 (89%)	91 (97%)	3 (3%)	46	74
1	J	100/106 (94%)	94 (94%)	6 (6%)	24	47
1	K	98/106 (92%)	93 (95%)	5 (5%)	29	55
1	L	97/106 (92%)	97 (100%)	0	100	100
All	All	1176/1272 (92%)	1134 (96%)	42 (4%)	42	71

5 of 42 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	22	LYS
1	G	14	ILE
1	K	55	LEU
1	F	49	ASP
1	F	103	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	9	ASN
1	H	70	ASN
1	J	53	GLN
1	F	70	ASN
1	K	9	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates [i](#)

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	125/126 (99%)	-0.25	1 (0%) 87 85	19, 35, 65, 73	0
1	B	119/126 (94%)	-0.31	0 100 100	17, 33, 56, 68	0
1	C	119/126 (94%)	-0.35	0 100 100	20, 33, 57, 65	0
1	D	126/126 (100%)	-0.31	0 100 100	17, 36, 62, 73	0
1	E	120/126 (95%)	-0.29	0 100 100	19, 35, 59, 78	0
1	F	118/126 (93%)	-0.08	3 (2%) 61 54	20, 41, 68, 80	0
1	G	125/126 (99%)	-0.36	0 100 100	18, 31, 53, 57	0
1	H	122/126 (96%)	-0.05	3 (2%) 61 54	16, 41, 80, 89	0
1	I	116/126 (92%)	0.12	2 (1%) 73 68	27, 52, 79, 81	0
1	J	124/126 (98%)	-0.08	3 (2%) 62 56	24, 42, 76, 78	0
1	K	123/126 (97%)	0.06	3 (2%) 62 56	27, 51, 75, 88	0
1	L	119/126 (94%)	-0.04	1 (0%) 87 85	25, 45, 73, 83	0
All	All	1456/1512 (96%)	-0.16	16 (1%) 82 79	16, 39, 71, 89	0

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	114	ASP	4.0
1	K	14	ILE	3.9
1	F	88	MET	2.9
1	I	44	LYS	2.7
1	J	17	TYR	2.7

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

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.