



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:34 PM GMT

PDB ID : 4P6Y  
Title : Crystal structure of the M42 aminopeptidase TmPep1050 from *Thermotoga maritima*  
Authors : Dutoit, R.; Demarez, M.; Van Elder, D.; Bauvois, C.  
Deposited on : 2014-03-25  
Resolution : 2.20 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

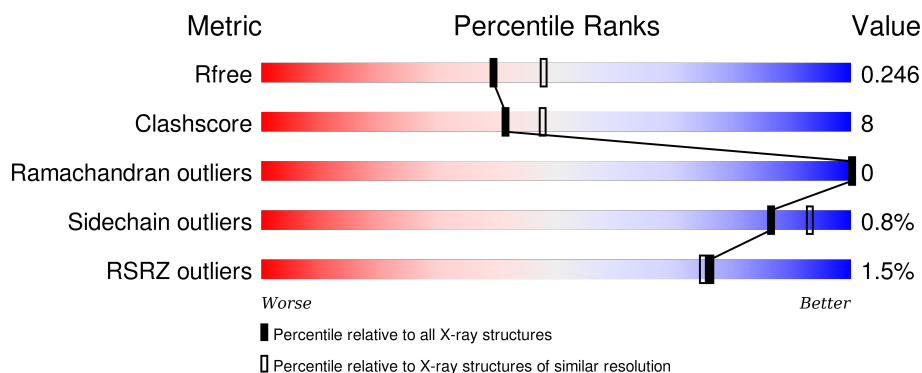
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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	3774 (2.20-2.20)
Clashscore	102246	4477 (2.20-2.20)
Ramachandran outliers	100387	4404 (2.20-2.20)
Sidechain outliers	100360	4405 (2.20-2.20)
RSRZ outliers	91569	3781 (2.20-2.20)

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	331	<div> <div></div> <div>85%14%</div> </div>
1	B	331	<div> <div>2%</div> <div>82%17%</div> </div>
1	C	331	<div> <div></div> <div>86%14%</div> </div>
1	D	331	<div> <div>2%</div> <div>82%17%</div> </div>
1	E	331	<div> <div></div> <div>84%15%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	331	<div><div></div><div>%</div><div><div></div><div>84%</div><div>15%</div></div><div></div></div>
1	G	331	<div><div></div><div>%</div><div><div></div><div>84%</div><div>15%</div></div><div></div></div>
1	H	331	<div><div></div><div>%</div><div><div></div><div>83%</div><div>17%</div></div><div></div></div>
1	I	331	<div><div></div><div>2%</div><div><div></div><div>85%</div><div>15%</div></div><div></div></div>
1	J	331	<div><div></div><div>2%</div><div><div></div><div>79%</div><div>21%</div></div><div></div></div>
1	K	331	<div><div></div><div>%</div><div><div></div><div>82%</div><div>16%</div></div><div></div></div>
1	L	331	<div><div></div><div>5%</div><div><div></div><div>85%</div><div>14%</div></div><div></div></div>

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 32192 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 Aminopeptidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	331	Total	C	N	O	S	0	0	0
			2513	1588	425	489	11			
1	B	330	Total	C	N	O	S	0	0	0
			2491	1579	419	482	11			
1	C	331	Total	C	N	O	S	0	0	0
			2501	1584	420	486	11			
1	D	331	Total	C	N	O	S	0	0	0
			2496	1584	417	484	11			
1	E	330	Total	C	N	O	S	0	0	0
			2513	1589	426	487	11			
1	F	331	Total	C	N	O	S	0	0	0
			2522	1593	429	489	11			
1	G	330	Total	C	N	O	S	0	0	0
			2492	1579	423	479	11			
1	H	330	Total	C	N	O	S	0	0	0
			2480	1572	420	477	11			
1	I	330	Total	C	N	O	S	0	0	0
			2494	1582	420	481	11			
1	J	330	Total	C	N	O	S	0	0	0
			2493	1581	420	481	11			
1	K	330	Total	C	N	O	S	0	0	0
			2482	1573	421	477	11			
1	L	330	Total	C	N	O	S	0	0	0
			2492	1578	422	481	11			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	221	Total	O	0	0
			221	221		
2	B	168	Total	O	0	0
			168	168		

*Continued on next page...*

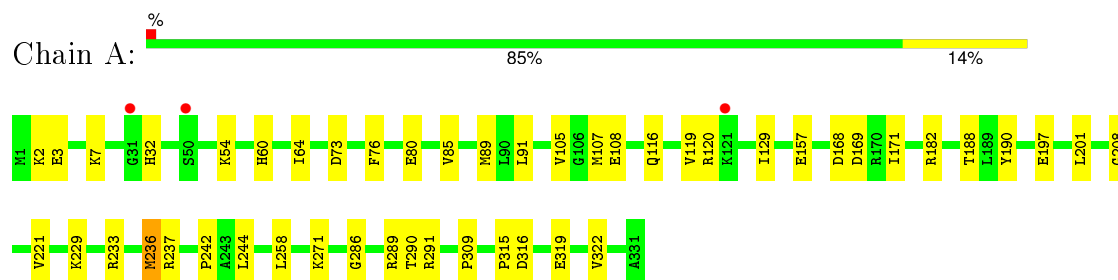
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	C	192	Total 192	O 192	0	0
2	D	206	Total 206	O 206	0	0
2	E	204	Total 204	O 204	0	0
2	F	176	Total 176	O 176	0	0
2	G	195	Total 195	O 195	0	0
2	H	165	Total 165	O 165	0	0
2	I	199	Total 199	O 199	0	0
2	J	147	Total 147	O 147	0	0
2	K	176	Total 176	O 176	0	0
2	L	174	Total 174	O 174	0	0

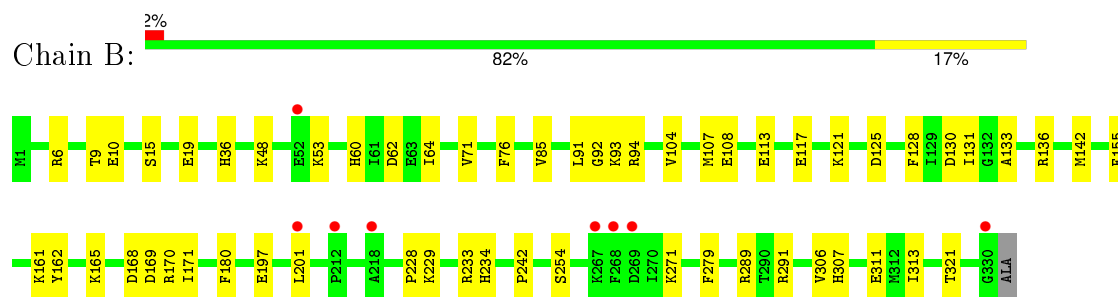
### 3 Residue-property plots

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.

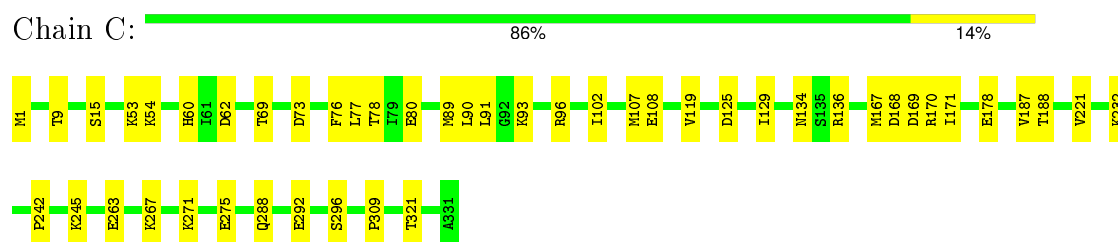
#### • Molecule 1: Aminopeptidase



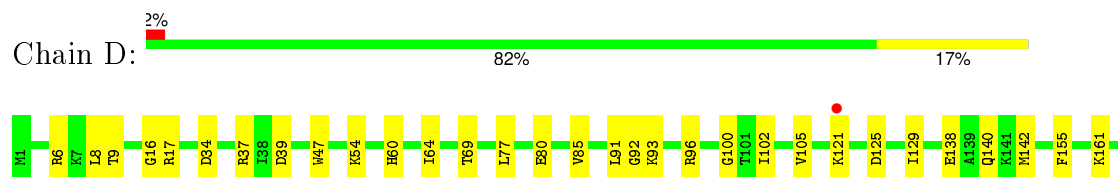
#### • Molecule 1: Aminopeptidase



#### • Molecule 1: Aminopeptidase

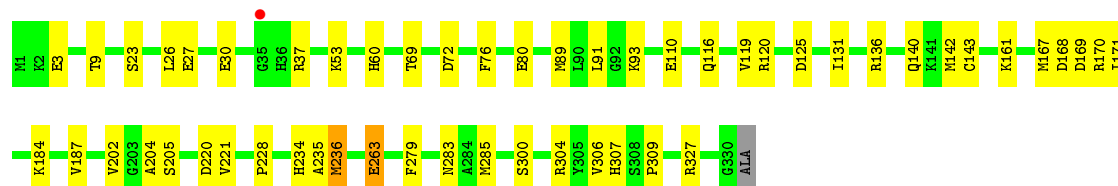
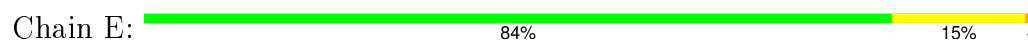


#### • Molecule 1: Aminopeptidase

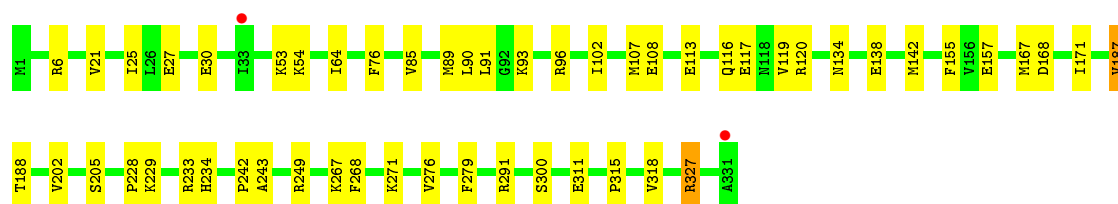
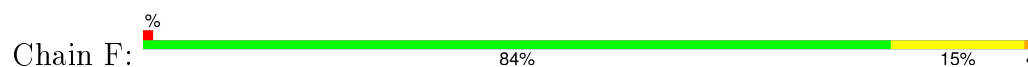




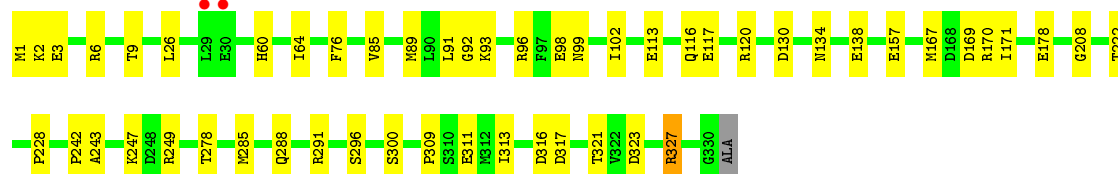
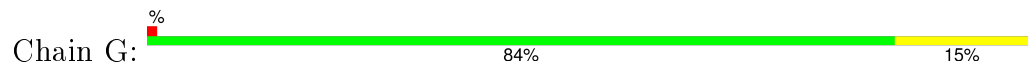
• Molecule 1: Aminopeptidase



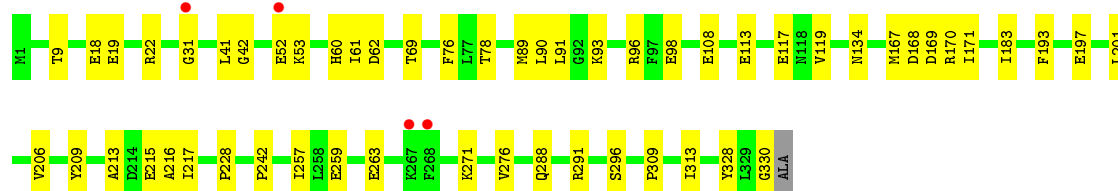
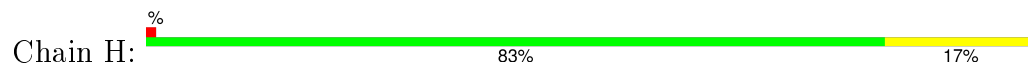
• Molecule 1: Aminopeptidase



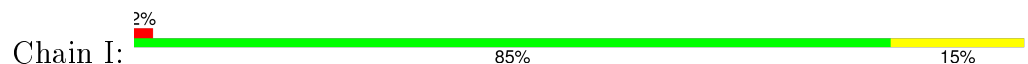
• Molecule 1: Aminopeptidase

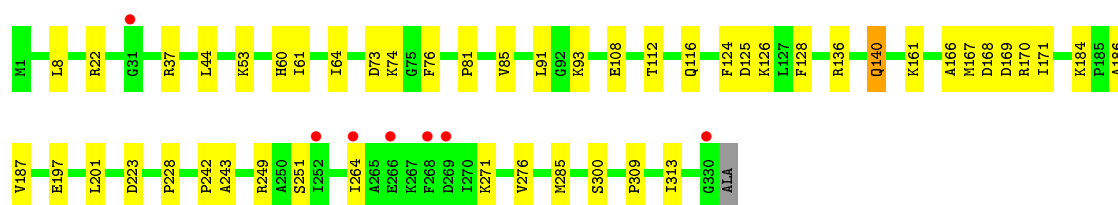


• Molecule 1: Aminopeptidase

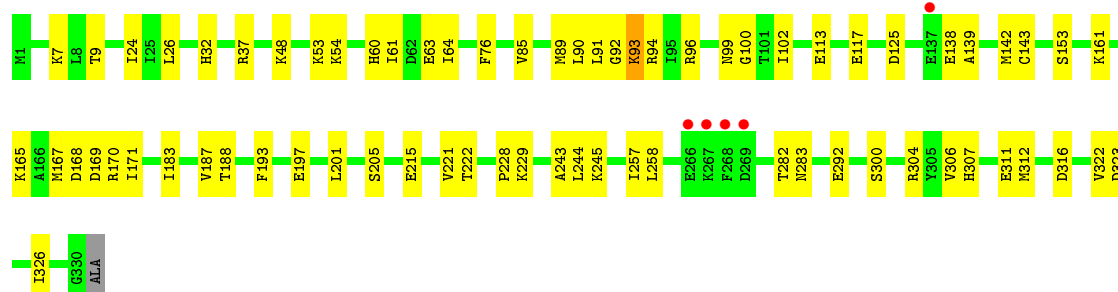
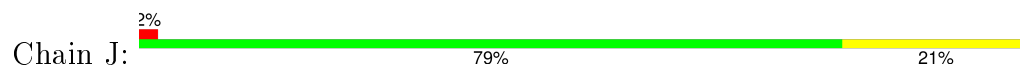


• Molecule 1: Aminopeptidase

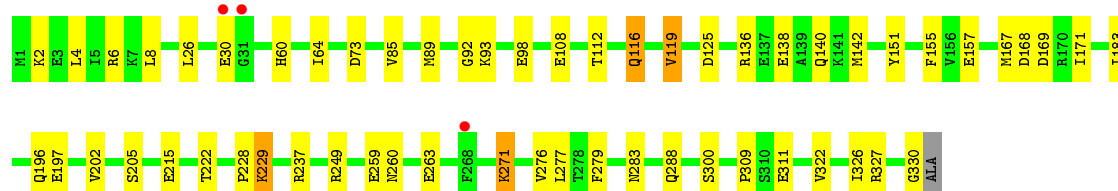
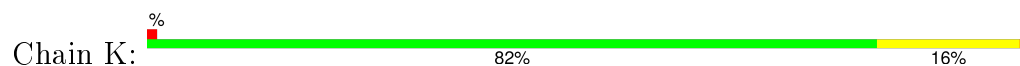




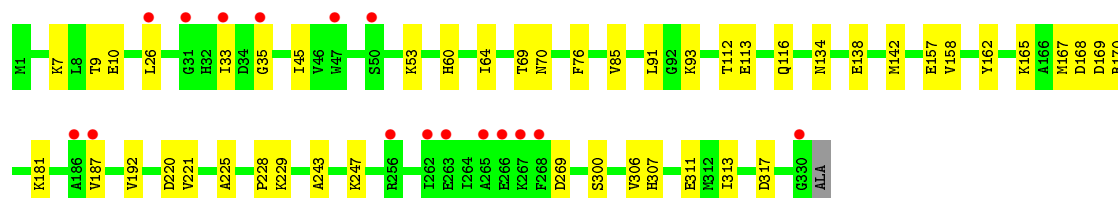
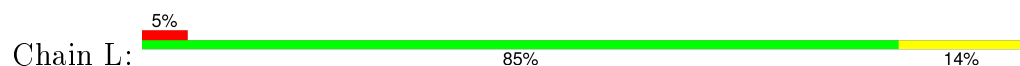
• Molecule 1: Aminopeptidase



• Molecule 1: Aminopeptidase



• Molecule 1: Aminopeptidase





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	114.26Å 114.57Å 114.04Å 114.46° 91.71° 105.69°	Depositor
Resolution (Å)	44.05 – 2.20 47.53 – 2.20	Depositor EDS
% Data completeness (in resolution range)	93.5 (44.05-2.20) 78.6 (47.53-2.20)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.60 (at 2.20Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1539)	Depositor
R, $R_{free}$	0.212 , 0.247 0.212 , 0.246	Depositor DCC
$R_{free}$ test set	11854 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	32.5	Xtriage
Anisotropy	0.164	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 43.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 237152 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	32192	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.89% 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 [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	0.44	0/2551	0.62	0/3442
1	B	0.42	0/2529	0.60	0/3415
1	C	0.44	0/2539	0.61	0/3427
1	D	0.45	0/2534	0.63	0/3420
1	E	0.46	0/2551	0.63	0/3442
1	F	0.47	0/2560	0.64	0/3453
1	G	0.43	0/2530	0.64	1/3416 (0.0%)
1	H	0.42	0/2518	0.61	0/3403
1	I	0.43	0/2532	0.62	0/3418
1	J	0.41	0/2531	0.61	0/3417
1	K	0.42	0/2520	0.61	0/3404
1	L	0.42	0/2530	0.60	0/3417
All	All	0.44	0/30425	0.62	1/41074 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	327	ARG	NE-CZ-NH2	-7.26	116.67	120.30

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	2513	0	2526	38	0
1	B	2491	0	2500	49	0
1	C	2501	0	2512	47	0
1	D	2496	0	2507	46	0
1	E	2513	0	2535	49	0
1	F	2522	0	2542	39	0
1	G	2492	0	2510	46	0
1	H	2480	0	2481	53	0
1	I	2494	0	2511	50	0
1	J	2493	0	2509	62	0
1	K	2482	0	2488	47	0
1	L	2492	0	2503	51	0
2	A	221	0	0	16	1
2	B	168	0	0	11	0
2	C	192	0	0	8	0
2	D	206	0	0	9	0
2	E	204	0	0	11	0
2	F	176	0	0	5	0
2	G	195	0	0	12	0
2	H	165	0	0	11	0
2	I	199	0	0	17	1
2	J	147	0	0	13	0
2	K	176	0	0	12	0
2	L	174	0	0	17	0
All	All	32192	0	30124	509	1

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 8.

The worst 5 of 509 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:E:120:ARG:NH1	2:E:502:HOH:O	1.85	1.09
1:J:316:ASP:HB2	2:J:545:HOH:O	1.53	1.09
1:J:89:MET:O	1:J:93:LYS:HE2	1.54	1.07
1:I:184:LYS:HG2	2:I:598:HOH:O	1.56	1.05
1:L:225:ALA:HA	2:L:570:HOH:O	1.60	0.99

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:A:425:HOH:O	2:I:404:HOH:O[1_655]	2.18	0.02

## 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	329/331 (99%)	320 (97%)	9 (3%)	0	100	100
1	B	328/331 (99%)	320 (98%)	8 (2%)	0	100	100
1	C	329/331 (99%)	322 (98%)	7 (2%)	0	100	100
1	D	329/331 (99%)	320 (97%)	9 (3%)	0	100	100
1	E	328/331 (99%)	320 (98%)	8 (2%)	0	100	100
1	F	329/331 (99%)	320 (97%)	9 (3%)	0	100	100
1	G	328/331 (99%)	319 (97%)	9 (3%)	0	100	100
1	H	328/331 (99%)	320 (98%)	8 (2%)	0	100	100
1	I	328/331 (99%)	319 (97%)	9 (3%)	0	100	100
1	J	328/331 (99%)	321 (98%)	7 (2%)	0	100	100
1	K	328/331 (99%)	320 (98%)	8 (2%)	0	100	100
1	L	328/331 (99%)	319 (97%)	9 (3%)	0	100	100
All	All	3940/3972 (99%)	3840 (98%)	100 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	267/273 (98%)	265 (99%)	2 (1%)	88	94
1	B	263/273 (96%)	263 (100%)	0	100	100
1	C	265/273 (97%)	264 (100%)	1 (0%)	93	97
1	D	263/273 (96%)	261 (99%)	2 (1%)	86	93
1	E	269/273 (98%)	265 (98%)	4 (2%)	72	84
1	F	269/273 (98%)	266 (99%)	3 (1%)	80	89
1	G	264/273 (97%)	264 (100%)	0	100	100
1	H	260/273 (95%)	258 (99%)	2 (1%)	86	93
1	I	264/273 (97%)	263 (100%)	1 (0%)	93	97
1	J	264/273 (97%)	262 (99%)	2 (1%)	86	93
1	K	261/273 (96%)	255 (98%)	6 (2%)	58	71
1	L	264/273 (97%)	263 (100%)	1 (0%)	93	97
All	All	3173/3276 (97%)	3149 (99%)	24 (1%)	86	93

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	187	VAL
1	H	119	VAL
1	K	288	GLN
1	F	327	ARG
1	H	91	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	L	116	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 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 [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	331/331 (100%)	-0.09	3 (0%) 85 85	20, 32, 46, 60	0
1	B	330/331 (99%)	0.26	8 (2%) 62 61	24, 35, 49, 62	0
1	C	331/331 (100%)	-0.14	0 100 100	21, 32, 45, 53	0
1	D	331/331 (100%)	0.15	8 (2%) 62 61	21, 30, 44, 55	0
1	E	330/331 (99%)	-0.03	1 (0%) 94 94	23, 32, 45, 56	0
1	F	331/331 (100%)	-0.06	2 (0%) 90 90	22, 32, 45, 56	0
1	G	330/331 (99%)	-0.06	2 (0%) 90 90	25, 33, 45, 59	0
1	H	330/331 (99%)	-0.02	4 (1%) 81 80	22, 34, 46, 58	0
1	I	330/331 (99%)	0.02	7 (2%) 67 65	24, 32, 45, 54	0
1	J	330/331 (99%)	-0.13	5 (1%) 76 75	25, 34, 46, 62	0
1	K	330/331 (99%)	-0.02	3 (0%) 85 85	25, 36, 48, 57	0
1	L	330/331 (99%)	0.12	16 (4%) 34 34	28, 37, 52, 63	0
All	All	3964/3972 (99%)	-0.00	59 (1%) 76 75	20, 33, 48, 63	0

The worst 5 of 59 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	331	ALA	4.9
1	D	331	ALA	4.6
1	B	268	PHE	3.7
1	D	268	PHE	3.6
1	L	265	ALA	3.3

### 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.