



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 01:40 PM GMT

PDB ID : 3U95  
Title : Crystal structure of a putative alpha-glucosidase from *Thermotoga neapolitana*  
Authors : Ha, N.C.; Jun, S.Y.; Yun, B.Y.; Yoon, B.Y.; Piao, S.  
Deposited on : 2011-10-17  
Resolution : 2.00 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

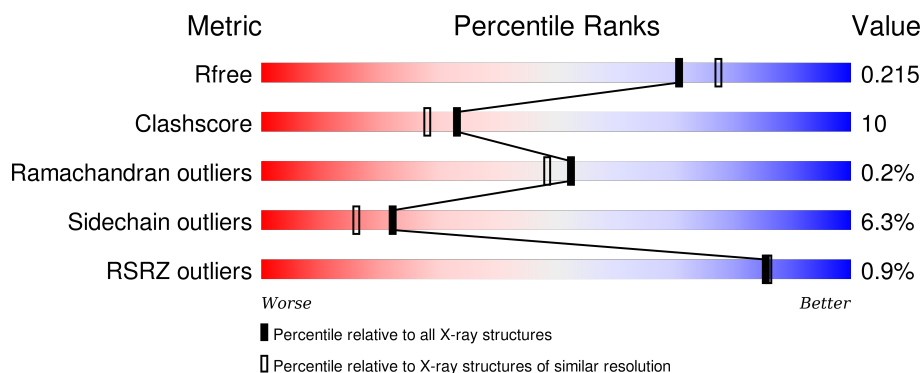
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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	6249 (2.00-2.00)
Clashscore	102246	7340 (2.00-2.00)
Ramachandran outliers	100387	7248 (2.00-2.00)
Sidechain outliers	100360	7247 (2.00-2.00)
RSRZ outliers	91569	6262 (2.00-2.00)

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	477	<div> <div>75%</div> <div>20%</div> <div>..</div> </div>
1	B	477	<div> <div>%</div> <div>78%</div> <div>18%</div> <div>..</div> </div>
1	C	477	<div> <div>%</div> <div>73%</div> <div>22%</div> <div>..</div> </div>
1	D	477	<div> <div>%</div> <div>73%</div> <div>22%</div> <div>..</div> </div>
1	E	477	<div> <div>76%</div> <div>19%</div> <div>..</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	477	<div><div></div><div>2%</div><div>76%</div><div>19%</div><div></div></div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 25816 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 Glycoside hydrolase, family 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	468	Total	C	N	O	S	0	0	0
			3894	2494	679	709	12			
1	B	469	Total	C	N	O	S	0	0	0
			3902	2498	681	711	12			
1	C	464	Total	C	N	O	S	0	0	0
			3860	2473	673	703	11			
1	D	468	Total	C	N	O	S	0	0	0
			3894	2494	679	709	12			
1	E	466	Total	C	N	O	S	0	0	0
			3877	2483	675	707	12			
1	F	469	Total	C	N	O	S	0	0	0
			3902	2498	681	711	12			

There are 54 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	416	THR	MET	CONFLICT	UNP B9KAM3
A	470	LEU	-	EXPRESSION TAG	UNP B9KAM3
A	471	GLU	-	EXPRESSION TAG	UNP B9KAM3
A	472	HIS	-	EXPRESSION TAG	UNP B9KAM3
A	473	HIS	-	EXPRESSION TAG	UNP B9KAM3
A	474	HIS	-	EXPRESSION TAG	UNP B9KAM3
A	475	HIS	-	EXPRESSION TAG	UNP B9KAM3
A	476	HIS	-	EXPRESSION TAG	UNP B9KAM3
A	477	HIS	-	EXPRESSION TAG	UNP B9KAM3
B	416	THR	MET	CONFLICT	UNP B9KAM3
B	470	LEU	-	EXPRESSION TAG	UNP B9KAM3
B	471	GLU	-	EXPRESSION TAG	UNP B9KAM3
B	472	HIS	-	EXPRESSION TAG	UNP B9KAM3
B	473	HIS	-	EXPRESSION TAG	UNP B9KAM3
B	474	HIS	-	EXPRESSION TAG	UNP B9KAM3
B	475	HIS	-	EXPRESSION TAG	UNP B9KAM3
B	476	HIS	-	EXPRESSION TAG	UNP B9KAM3

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	477	HIS	-	EXPRESSION TAG	UNP B9KAM3
C	416	THR	MET	CONFLICT	UNP B9KAM3
C	470	LEU	-	EXPRESSION TAG	UNP B9KAM3
C	471	GLU	-	EXPRESSION TAG	UNP B9KAM3
C	472	HIS	-	EXPRESSION TAG	UNP B9KAM3
C	473	HIS	-	EXPRESSION TAG	UNP B9KAM3
C	474	HIS	-	EXPRESSION TAG	UNP B9KAM3
C	475	HIS	-	EXPRESSION TAG	UNP B9KAM3
C	476	HIS	-	EXPRESSION TAG	UNP B9KAM3
C	477	HIS	-	EXPRESSION TAG	UNP B9KAM3
D	416	THR	MET	CONFLICT	UNP B9KAM3
D	470	LEU	-	EXPRESSION TAG	UNP B9KAM3
D	471	GLU	-	EXPRESSION TAG	UNP B9KAM3
D	472	HIS	-	EXPRESSION TAG	UNP B9KAM3
D	473	HIS	-	EXPRESSION TAG	UNP B9KAM3
D	474	HIS	-	EXPRESSION TAG	UNP B9KAM3
D	475	HIS	-	EXPRESSION TAG	UNP B9KAM3
D	476	HIS	-	EXPRESSION TAG	UNP B9KAM3
D	477	HIS	-	EXPRESSION TAG	UNP B9KAM3
E	416	THR	MET	CONFLICT	UNP B9KAM3
E	470	LEU	-	EXPRESSION TAG	UNP B9KAM3
E	471	GLU	-	EXPRESSION TAG	UNP B9KAM3
E	472	HIS	-	EXPRESSION TAG	UNP B9KAM3
E	473	HIS	-	EXPRESSION TAG	UNP B9KAM3
E	474	HIS	-	EXPRESSION TAG	UNP B9KAM3
E	475	HIS	-	EXPRESSION TAG	UNP B9KAM3
E	476	HIS	-	EXPRESSION TAG	UNP B9KAM3
E	477	HIS	-	EXPRESSION TAG	UNP B9KAM3
F	416	THR	MET	CONFLICT	UNP B9KAM3
F	470	LEU	-	EXPRESSION TAG	UNP B9KAM3
F	471	GLU	-	EXPRESSION TAG	UNP B9KAM3
F	472	HIS	-	EXPRESSION TAG	UNP B9KAM3
F	473	HIS	-	EXPRESSION TAG	UNP B9KAM3
F	474	HIS	-	EXPRESSION TAG	UNP B9KAM3
F	475	HIS	-	EXPRESSION TAG	UNP B9KAM3
F	476	HIS	-	EXPRESSION TAG	UNP B9KAM3
F	477	HIS	-	EXPRESSION TAG	UNP B9KAM3

- Molecule 2 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
-----	-------	----------	-------	---------	---------

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	D	1	Total Mn 1 1	0	0
2	E	1	Total Mn 1 1	0	0
2	B	1	Total Mn 1 1	0	0
2	C	1	Total Mn 1 1	0	0
2	A	1	Total Mn 1 1	0	0
2	F	1	Total Mn 1 1	0	0

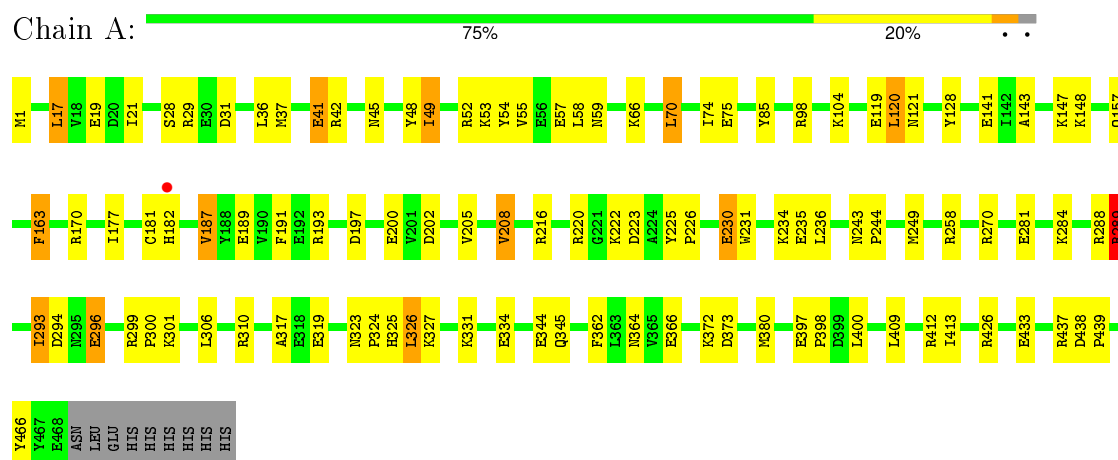
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	444	Total O 444 444	0	0
3	B	402	Total O 402 402	0	0
3	C	393	Total O 393 393	0	0
3	D	450	Total O 450 450	0	0
3	E	431	Total O 431 431	0	0
3	F	361	Total O 361 361	0	0

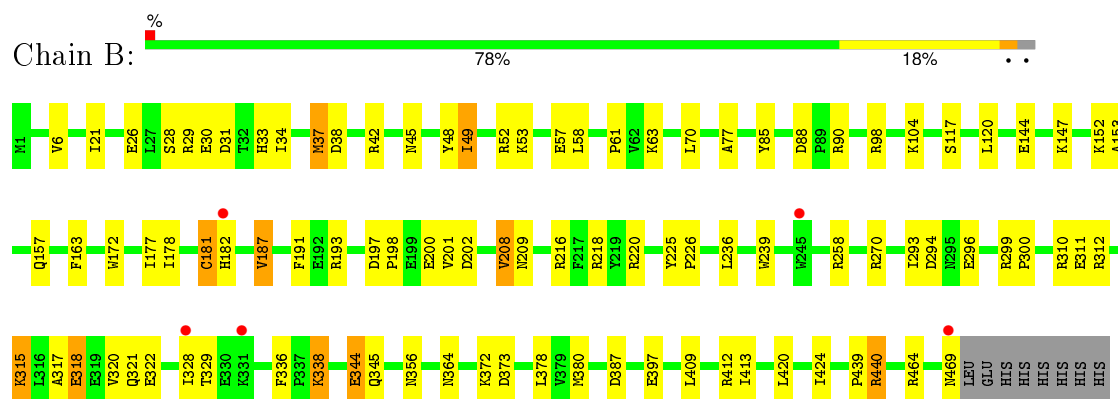
### 3 Residue-property plots

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.

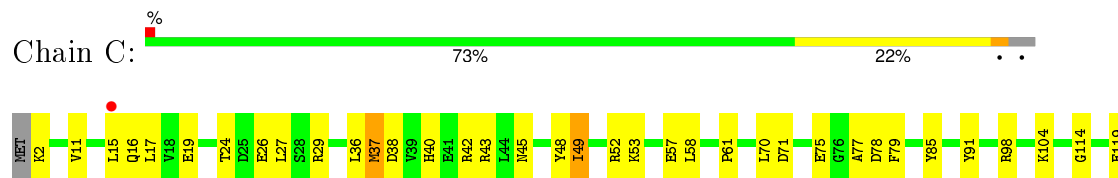
#### • Molecule 1: Glycoside hydrolase, family 4

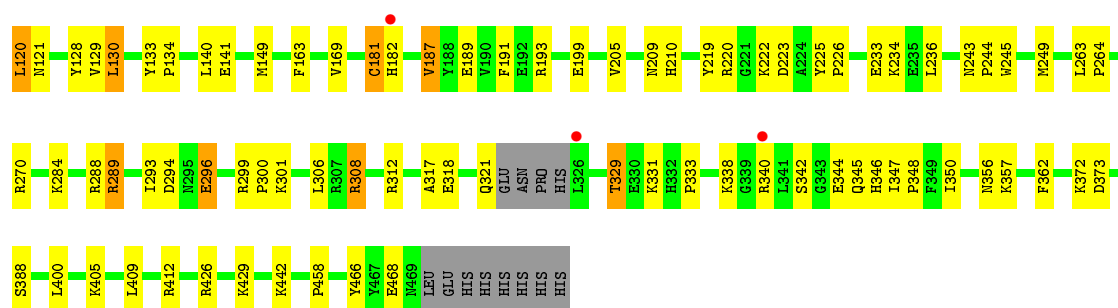


#### • Molecule 1: Glycoside hydrolase, family 4

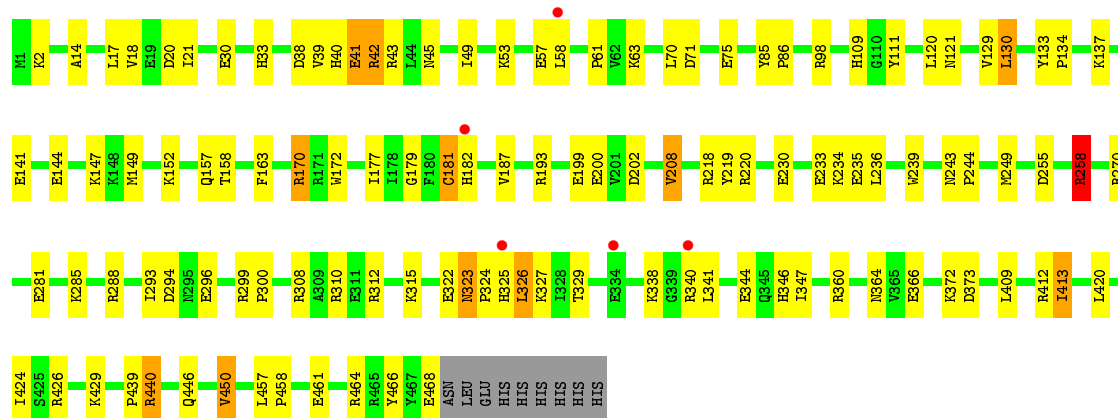
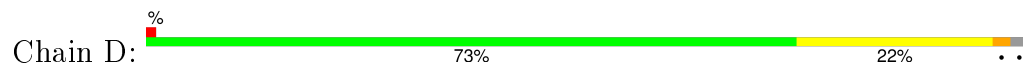


#### • Molecule 1: Glycoside hydrolase, family 4

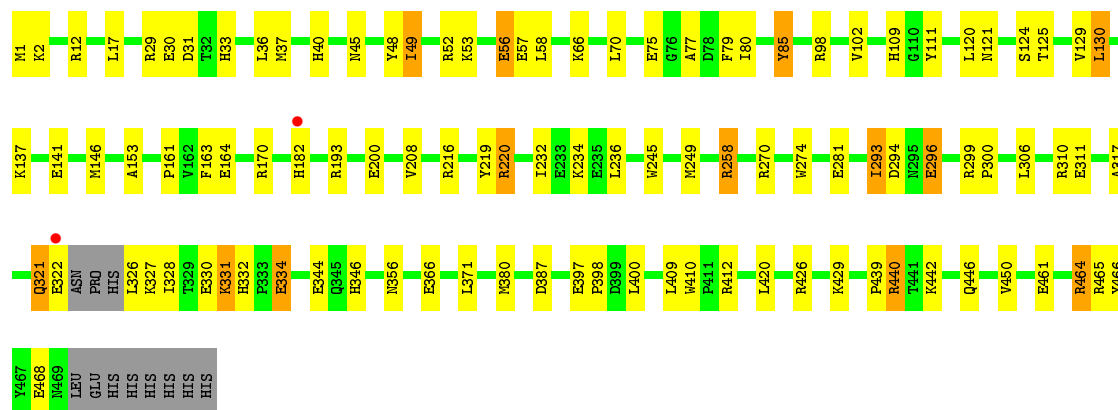
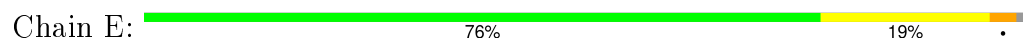




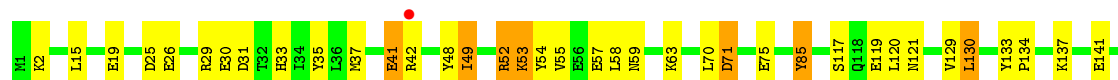
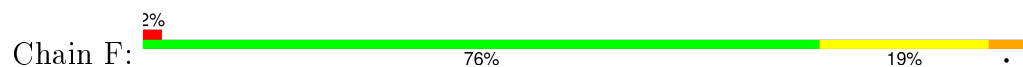
- Molecule 1: Glycoside hydrolase, family 4



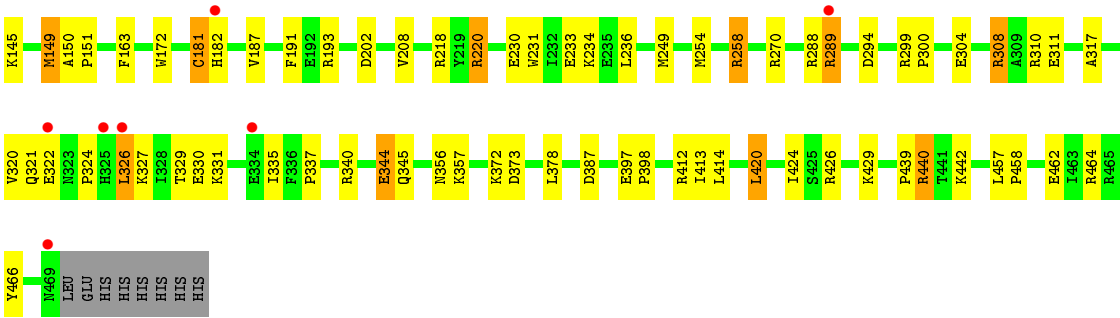
- Molecule 1: Glycoside hydrolase, family 4



- Molecule 1: Glycoside hydrolase, family 4







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	154.53Å 154.53Å 139.06Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	40.90 – 2.00 40.90 – 2.00	Depositor EDS
% Data completeness (in resolution range)	98.8 (40.90-2.00) 98.9 (40.90-2.00)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	0.10	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.03 (at 2.00Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.1_743)	Depositor
R, $R_{free}$	0.185 , 0.222 0.180 , 0.215	Depositor DCC
$R_{free}$ test set	1990 reflections (0.80%)	DCC
Wilson B-factor (Å <sup>2</sup> )	23.4	Xtriage
Anisotropy	0.079	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 32.1	EDS
Estimated twinning fraction	0.009 for -h,-k,l 0.026 for h,-h-k,-l 0.075 for -k,-h,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	0 of 248929 reflections	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	25816	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	31.0	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 23.07 % 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 5.0116e-03. 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: MN

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	0/3994	0.53	0/5407
1	B	0.38	0/4002	0.53	0/5418
1	C	0.38	0/3957	0.52	0/5355
1	D	0.39	0/3994	0.56	2/5407 (0.0%)
1	E	0.39	0/3974	0.55	1/5377 (0.0%)
1	F	0.36	0/4002	0.52	0/5418
All	All	0.38	0/23923	0.54	3/32382 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	440	ARG	NE-CZ-NH2	-5.59	117.50	120.30
1	D	258	ARG	NE-CZ-NH1	5.57	123.08	120.30
1	E	440	ARG	NE-CZ-NH2	-5.17	117.72	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	3894	0	3847	85	0
1	B	3902	0	3853	70	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	3860	0	3814	97	0
1	D	3894	0	3848	73	0
1	E	3877	0	3832	77	0
1	F	3902	0	3853	81	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	F	1	0	0	0	0
3	A	444	0	0	16	0
3	B	402	0	0	12	0
3	C	393	0	0	19	0
3	D	450	0	0	14	0
3	E	431	0	0	14	0
3	F	361	0	0	17	0
All	All	25816	0	23047	476	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 10.

The worst 5 of 476 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:429:LYS:HG3	3:D:2410:HOH:O	1.48	1.09
1:A:289:ARG:HG3	1:A:289:ARG:HH21	1.15	1.09
1:A:182:HIS:CE1	1:A:344:GLU:HG3	1.94	1.03
1:F:289:ARG:HG2	3:F:2213:HOH:O	1.58	1.03
1:A:182:HIS:HE1	1:A:344:GLU:HG3	1.25	1.02

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	466/477 (98%)	453 (97%)	11 (2%)	2 (0%)	39	33
1	B	467/477 (98%)	449 (96%)	18 (4%)	0	100	100
1	C	460/477 (96%)	441 (96%)	18 (4%)	1 (0%)	52	48
1	D	466/477 (98%)	452 (97%)	13 (3%)	1 (0%)	52	48
1	E	462/477 (97%)	445 (96%)	17 (4%)	0	100	100
1	F	467/477 (98%)	449 (96%)	17 (4%)	1 (0%)	52	48
All	All	2788/2862 (97%)	2689 (96%)	94 (3%)	5 (0%)	52	48

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	289	ARG
1	D	323	ASN
1	F	289	ARG
1	C	289	ARG
1	A	324	PRO

### 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	418/427 (98%)	390 (93%)	28 (7%)	20	14
1	B	419/427 (98%)	397 (95%)	22 (5%)	28	22
1	C	414/427 (97%)	392 (95%)	22 (5%)	28	22
1	D	418/427 (98%)	390 (93%)	28 (7%)	20	14
1	E	416/427 (97%)	387 (93%)	29 (7%)	19	12
1	F	419/427 (98%)	391 (93%)	28 (7%)	20	14
All	All	2504/2562 (98%)	2347 (94%)	157 (6%)	22	16

5 of 157 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	340	ARG
1	D	233	GLU
1	F	220	ARG
1	D	42	ARG
1	D	130	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	182	HIS
1	D	33	HIS

### 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 ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 6 ligands modelled in this entry, 6 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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	468/477 (98%)	-0.18	1 (0%) 95 95	15, 26, 50, 72	0
1	B	469/477 (98%)	-0.10	5 (1%) 82 83	14, 28, 56, 69	0
1	C	464/477 (97%)	-0.02	4 (0%) 85 86	15, 29, 60, 74	0
1	D	468/477 (98%)	-0.14	5 (1%) 82 83	14, 27, 53, 71	0
1	E	466/477 (97%)	-0.16	2 (0%) 93 93	14, 27, 52, 74	0
1	F	469/477 (98%)	-0.08	8 (1%) 73 73	15, 29, 57, 80	0
All	All	2804/2862 (97%)	-0.11	25 (0%) 85 86	14, 28, 57, 80	0

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	182	HIS	5.5
1	B	328	ILE	5.2
1	C	326	LEU	5.0
1	F	182	HIS	4.9
1	D	325	HIS	4.0

### 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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	MN	A	601	1/1	0.98	0.04	-0.80	55,55,55,55	0
2	MN	B	602	1/1	0.99	0.03	-1.00	48,48,48,48	0
2	MN	D	600	1/1	0.98	0.04	-1.18	58,58,58,58	0
2	MN	C	500	1/1	0.98	0.03	-1.23	51,51,51,51	0
2	MN	E	600	1/1	0.99	0.03	-1.28	51,51,51,51	0
2	MN	F	600	1/1	0.99	0.04	-	52,52,52,52	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.