



# Full wwPDB/EMDatabank EM Map/Model Validation Report ⓘ

Feb 5, 2020 – 11:50 AM EST

PDB ID : 6V1I  
EMDB ID: : EMD-21012  
Title : Cryo-EM reconstruction of the thermophilic bacteriophage P74-26 small terminase- symmetric  
Authors : Hayes, J.A.; Hilbert, B.J.; Gaubitz, C.; Stone, N.P.; Kelch, B.A.  
Deposited on : 2019-11-20  
Resolution : 3.80 Å(reported)  
Based on PDB ID : 4XVN

This is a Full wwPDB/EMDatabank EM Map/Model Validation Report  
for a publicly released PDB/EMDB 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/EMValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

MolProbity : 4.02b-467  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.4

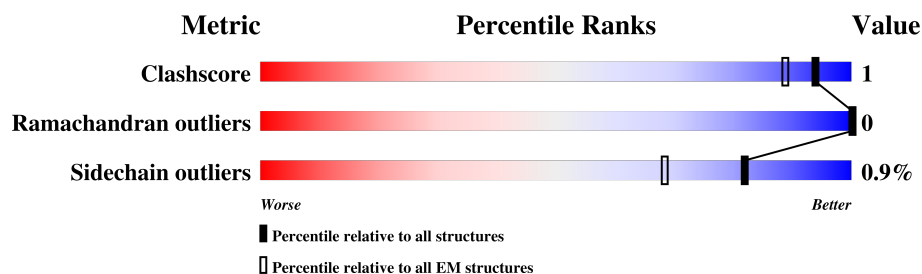
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 3.80 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)	EM structures (#Entries)
Clashscore	136327	1886
Ramachandran outliers	132723	1663
Sidechain outliers	132532	1531

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	174	75% 21%
1	B	174	75% 21%
1	C	174	75% 21%
1	D	174	76% 21%
1	E	174	75% 21%
1	F	174	75% 21%
1	G	174	75% 21%
1	H	174	75% 21%
1	I	174	75% 21%

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 9711 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 Small terminase protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	137	Total	C	N	O	S	0	0
			1079	681	189	203	6		
1	B	137	Total	C	N	O	S	0	0
			1079	681	189	203	6		
1	C	137	Total	C	N	O	S	0	0
			1079	681	189	203	6		
1	D	137	Total	C	N	O	S	0	0
			1079	681	189	203	6		
1	E	137	Total	C	N	O	S	0	0
			1079	681	189	203	6		
1	F	137	Total	C	N	O	S	0	0
			1079	681	189	203	6		
1	G	137	Total	C	N	O	S	0	0
			1079	681	189	203	6		
1	H	137	Total	C	N	O	S	0	0
			1079	681	189	203	6		
1	I	137	Total	C	N	O	S	0	0
			1079	681	189	203	6		

There are 27 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP A7XXR0
A	-1	PRO	-	expression tag	UNP A7XXR0
A	0	HIS	-	expression tag	UNP A7XXR0
B	-2	GLY	-	expression tag	UNP A7XXR0
B	-1	PRO	-	expression tag	UNP A7XXR0
B	0	HIS	-	expression tag	UNP A7XXR0
C	-2	GLY	-	expression tag	UNP A7XXR0
C	-1	PRO	-	expression tag	UNP A7XXR0
C	0	HIS	-	expression tag	UNP A7XXR0
D	-2	GLY	-	expression tag	UNP A7XXR0
D	-1	PRO	-	expression tag	UNP A7XXR0
D	0	HIS	-	expression tag	UNP A7XXR0

*Continued on next page...*

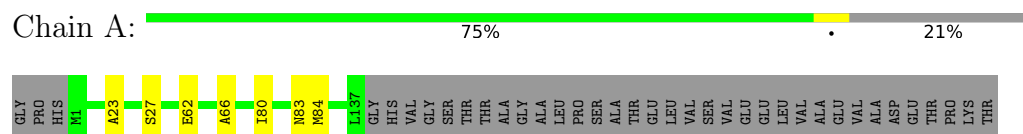
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
E	-2	GLY	-	expression tag	UNP A7XXR0
E	-1	PRO	-	expression tag	UNP A7XXR0
E	0	HIS	-	expression tag	UNP A7XXR0
F	-2	GLY	-	expression tag	UNP A7XXR0
F	-1	PRO	-	expression tag	UNP A7XXR0
F	0	HIS	-	expression tag	UNP A7XXR0
G	-2	GLY	-	expression tag	UNP A7XXR0
G	-1	PRO	-	expression tag	UNP A7XXR0
G	0	HIS	-	expression tag	UNP A7XXR0
H	-2	GLY	-	expression tag	UNP A7XXR0
H	-1	PRO	-	expression tag	UNP A7XXR0
H	0	HIS	-	expression tag	UNP A7XXR0
I	-2	GLY	-	expression tag	UNP A7XXR0
I	-1	PRO	-	expression tag	UNP A7XXR0
I	0	HIS	-	expression tag	UNP A7XXR0

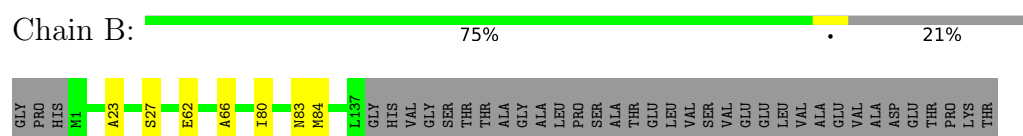
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. 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. 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.

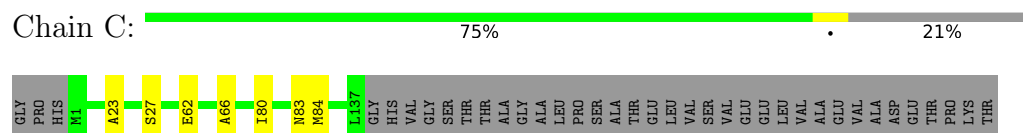
- Molecule 1: Small terminase protein



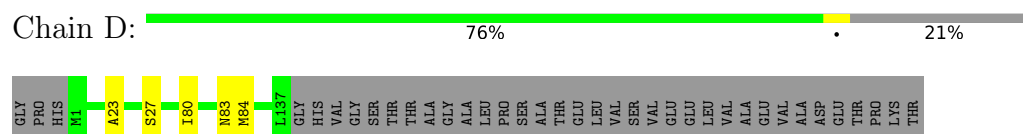
- Molecule 1: Small terminase protein



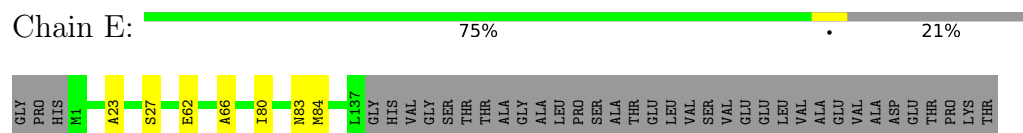
- Molecule 1: Small terminase protein



- Molecule 1: Small terminase protein

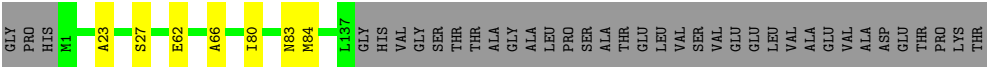


- Molecule 1: Small terminase protein

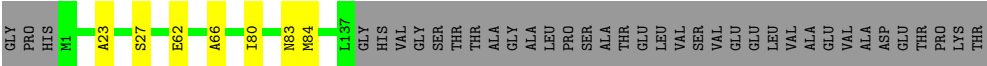


- Molecule 1: Small terminase protein

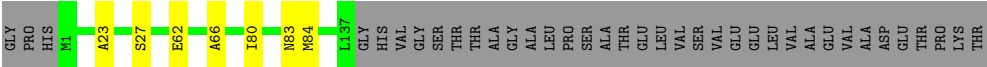




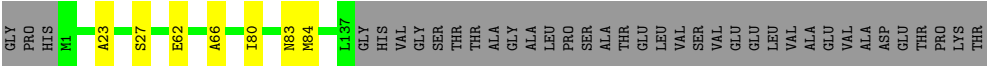
● Molecule 1: Small terminase protein



● Molecule 1: Small terminase protein



● Molecule 1: Small terminase protein



## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C9	Depositor
Number of particles used	84860	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	TFS KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	50	Depositor
Minimum defocus (nm)	1400	Depositor
Maximum defocus (nm)	2600	Depositor
Magnification	13000	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

## 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  >2	RMSZ	# Z  >2
1	A	0.35	0/1091	0.49	0/1472
1	B	0.35	0/1091	0.49	0/1472
1	C	0.34	0/1091	0.49	0/1472
1	D	0.35	0/1091	0.49	0/1472
1	E	0.35	0/1091	0.49	0/1472
1	F	0.35	0/1091	0.49	0/1472
1	G	0.35	0/1091	0.49	0/1472
1	H	0.34	0/1091	0.49	0/1472
1	I	0.35	0/1091	0.49	0/1472
All	All	0.35	0/9819	0.49	0/13248

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	1079	0	1123	3	0
1	B	1079	0	1123	3	0
1	C	1079	0	1123	3	0
1	D	1079	0	1123	2	0
1	E	1079	0	1123	3	0
1	F	1079	0	1123	3	0
1	G	1079	0	1123	3	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	H	1079	0	1123	3	0
1	I	1079	0	1123	3	0
All	All	9711	0	10107	26	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 1.

All (26) 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:62:GLU:O	1:E:66:ALA:N	2.43	0.48
1:H:80:ILE:O	1:H:84:MET:N	2.50	0.45
1:B:80:ILE:O	1:B:84:MET:N	2.50	0.45
1:F:80:ILE:O	1:F:84:MET:N	2.50	0.45
1:I:62:GLU:O	1:I:66:ALA:N	2.43	0.45
1:A:80:ILE:O	1:A:84:MET:N	2.50	0.45
1:D:80:ILE:O	1:D:84:MET:N	2.50	0.45
1:H:62:GLU:O	1:H:66:ALA:N	2.43	0.45
1:G:80:ILE:O	1:G:84:MET:N	2.50	0.44
1:I:80:ILE:O	1:I:84:MET:N	2.50	0.44
1:C:80:ILE:O	1:C:84:MET:N	2.50	0.44
1:E:80:ILE:O	1:E:84:MET:N	2.50	0.44
1:G:62:GLU:O	1:G:66:ALA:N	2.43	0.43
1:A:62:GLU:O	1:A:66:ALA:N	2.43	0.43
1:A:23:ALA:O	1:A:27:SER:N	2.52	0.43
1:C:23:ALA:O	1:C:27:SER:N	2.52	0.43
1:E:23:ALA:O	1:E:27:SER:N	2.52	0.42
1:F:23:ALA:O	1:F:27:SER:N	2.52	0.42
1:B:23:ALA:O	1:B:27:SER:N	2.52	0.42
1:G:23:ALA:O	1:G:27:SER:N	2.52	0.42
1:I:23:ALA:O	1:I:27:SER:N	2.52	0.42
1:F:62:GLU:O	1:F:66:ALA:N	2.43	0.42
1:H:23:ALA:O	1:H:27:SER:N	2.52	0.42
1:D:23:ALA:O	1:D:27:SER:N	2.52	0.41
1:B:62:GLU:O	1:B:66:ALA:N	2.43	0.41
1:C:62:GLU:O	1:C:66:ALA:N	2.43	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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 EM 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	135/174 (78%)	133 (98%)	2 (2%)	0	100	100
1	B	135/174 (78%)	133 (98%)	2 (2%)	0	100	100
1	C	135/174 (78%)	133 (98%)	2 (2%)	0	100	100
1	D	135/174 (78%)	133 (98%)	2 (2%)	0	100	100
1	E	135/174 (78%)	133 (98%)	2 (2%)	0	100	100
1	F	135/174 (78%)	133 (98%)	2 (2%)	0	100	100
1	G	135/174 (78%)	133 (98%)	2 (2%)	0	100	100
1	H	135/174 (78%)	133 (98%)	2 (2%)	0	100	100
1	I	135/174 (78%)	133 (98%)	2 (2%)	0	100	100
All	All	1215/1566 (78%)	1197 (98%)	18 (2%)	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 PDB entries followed by that with respect to all EM 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	117/145 (81%)	116 (99%)	1 (1%)	81	91
1	B	117/145 (81%)	116 (99%)	1 (1%)	81	91
1	C	117/145 (81%)	116 (99%)	1 (1%)	81	91
1	D	117/145 (81%)	116 (99%)	1 (1%)	81	91
1	E	117/145 (81%)	116 (99%)	1 (1%)	81	91

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	F	117/145 (81%)	116 (99%)	1 (1%)	81	91
1	G	117/145 (81%)	116 (99%)	1 (1%)	81	91
1	H	117/145 (81%)	116 (99%)	1 (1%)	81	91
1	I	117/145 (81%)	116 (99%)	1 (1%)	81	91
All	All	1053/1305 (81%)	1044 (99%)	9 (1%)	82	91

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	83	ASN
1	B	83	ASN
1	C	83	ASN
1	D	83	ASN
1	E	83	ASN
1	F	83	ASN
1	G	83	ASN
1	H	83	ASN
1	I	83	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA [i](#)

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

There are no such residues in this entry.

## 5.8 Polymer linkage issues

There are no chain breaks in this entry.