



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 09:17 PM UTC

PDB ID : 4EO2 / pdb\_00004eo2  
Title : Structure of the bacteriophage C1 tail knob protein, gp12  
Authors : Aksyuk, A.A.; Rossmann, M.G.  
Deposited on : 2012-04-13  
Resolution : 3.01 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

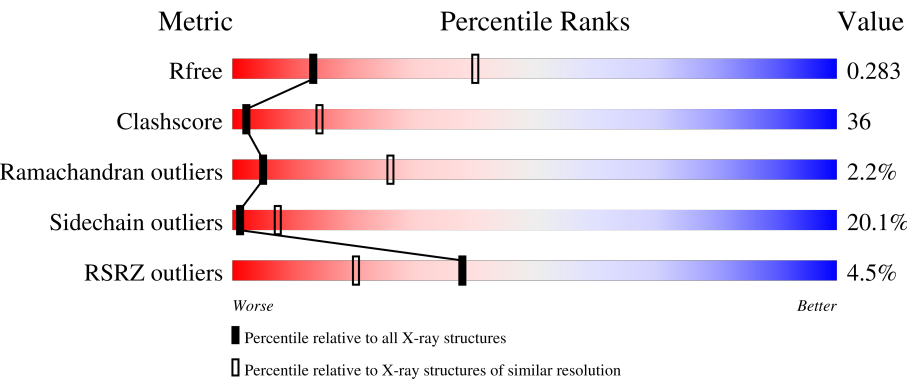
MolProbity	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:  
*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.01 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))
$R_{free}$	180053	2672 (3.00-3.00)
Clashscore	190562	2977 (3.00-3.00)
Ramachandran outliers	187476	2877 (3.00-3.00)
Sidechain outliers	187428	2880 (3.00-3.00)
RSRZ outliers	180081	2671 (3.00-3.00)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. 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	583	<div><div>5%</div><div><div></div><div>37%</div><div>32%</div><div>13%</div><div>•</div><div>16%</div></div></div>
1	B	583	<div><div>6%</div><div><div></div><div>38%</div><div>31%</div><div>13%</div><div>•</div><div>16%</div></div></div>
1	C	583	<div><div>2%</div><div><div></div><div>37%</div><div>32%</div><div>13%</div><div>•</div><div>16%</div></div></div>
1	D	583	<div><div>3%</div><div><div></div><div>35%</div><div>35%</div><div>13%</div><div>•</div><div>16%</div></div></div>
1	E	583	<div><div>3%</div><div><div></div><div>35%</div><div>34%</div><div>13%</div><div>•</div><div>16%</div></div></div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	583	<div><div></div><div>4%</div><div>34%</div><div>37%</div><div>12%</div><div>•</div><div>16%</div></div>

## 2 Entry composition [i](#)

There is only 1 type of molecule in this entry. The entry contains 23688 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 Major tail protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	488	Total	C	N	O	S	0	0	0
			3948	2504	665	761	18			
1	B	488	Total	C	N	O	S	0	0	0
			3948	2504	665	761	18			
1	C	488	Total	C	N	O	S	0	0	0
			3948	2504	665	761	18			
1	D	488	Total	C	N	O	S	0	0	0
			3948	2504	665	761	18			
1	E	488	Total	C	N	O	S	0	0	0
			3948	2504	665	761	18			
1	F	488	Total	C	N	O	S	0	0	0
			3948	2504	665	761	18			

There are 54 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	575	SER	-	expression tag	UNP Q7Y3F0
A	576	LEU	-	expression tag	UNP Q7Y3F0
A	577	GLU	-	expression tag	UNP Q7Y3F0
A	578	HIS	-	expression tag	UNP Q7Y3F0
A	579	HIS	-	expression tag	UNP Q7Y3F0
A	580	HIS	-	expression tag	UNP Q7Y3F0
A	581	HIS	-	expression tag	UNP Q7Y3F0
A	582	HIS	-	expression tag	UNP Q7Y3F0
A	583	HIS	-	expression tag	UNP Q7Y3F0
B	575	SER	-	expression tag	UNP Q7Y3F0
B	576	LEU	-	expression tag	UNP Q7Y3F0
B	577	GLU	-	expression tag	UNP Q7Y3F0
B	578	HIS	-	expression tag	UNP Q7Y3F0
B	579	HIS	-	expression tag	UNP Q7Y3F0
B	580	HIS	-	expression tag	UNP Q7Y3F0
B	581	HIS	-	expression tag	UNP Q7Y3F0
B	582	HIS	-	expression tag	UNP Q7Y3F0

*Continued on next page...*

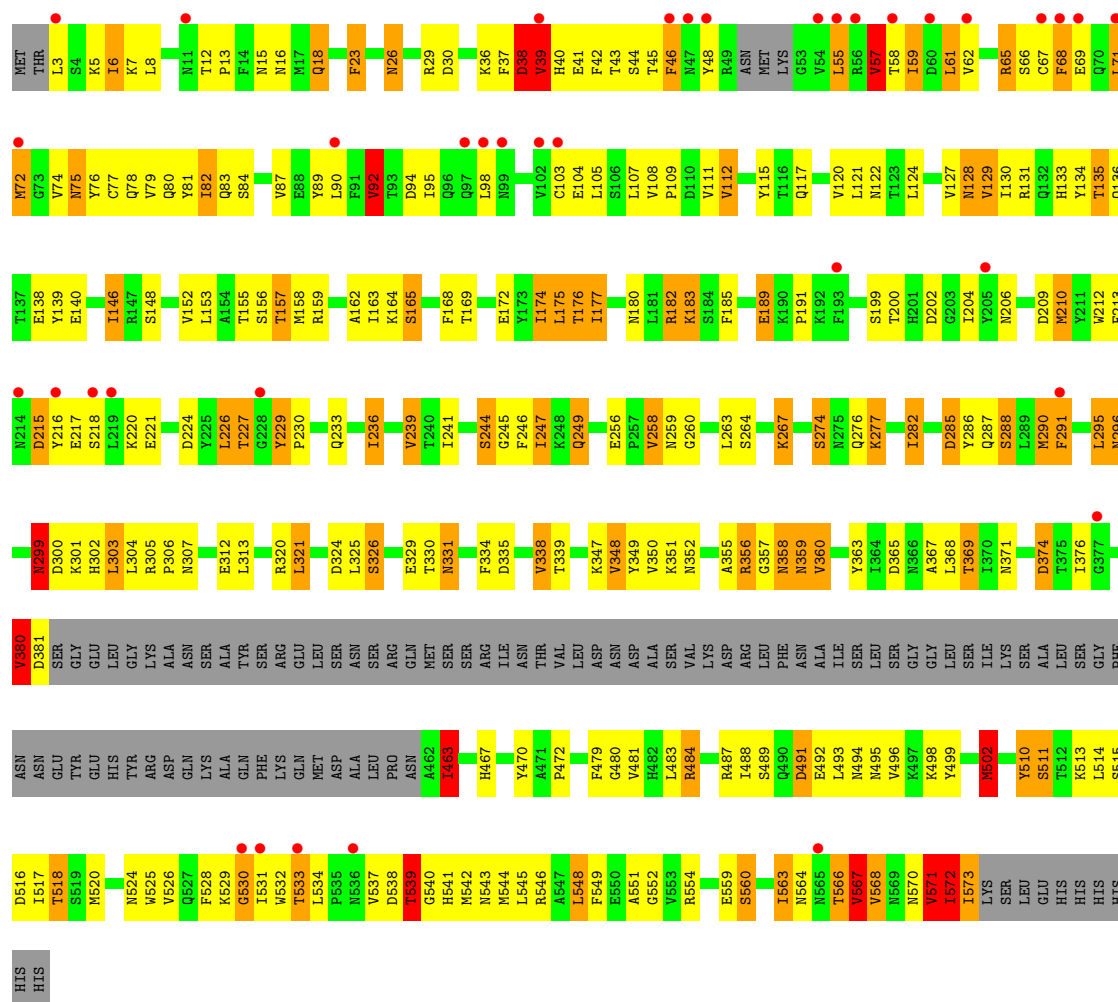
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	583	HIS	-	expression tag	UNP Q7Y3F0
C	575	SER	-	expression tag	UNP Q7Y3F0
C	576	LEU	-	expression tag	UNP Q7Y3F0
C	577	GLU	-	expression tag	UNP Q7Y3F0
C	578	HIS	-	expression tag	UNP Q7Y3F0
C	579	HIS	-	expression tag	UNP Q7Y3F0
C	580	HIS	-	expression tag	UNP Q7Y3F0
C	581	HIS	-	expression tag	UNP Q7Y3F0
C	582	HIS	-	expression tag	UNP Q7Y3F0
C	583	HIS	-	expression tag	UNP Q7Y3F0
D	575	SER	-	expression tag	UNP Q7Y3F0
D	576	LEU	-	expression tag	UNP Q7Y3F0
D	577	GLU	-	expression tag	UNP Q7Y3F0
D	578	HIS	-	expression tag	UNP Q7Y3F0
D	579	HIS	-	expression tag	UNP Q7Y3F0
D	580	HIS	-	expression tag	UNP Q7Y3F0
D	581	HIS	-	expression tag	UNP Q7Y3F0
D	582	HIS	-	expression tag	UNP Q7Y3F0
D	583	HIS	-	expression tag	UNP Q7Y3F0
E	575	SER	-	expression tag	UNP Q7Y3F0
E	576	LEU	-	expression tag	UNP Q7Y3F0
E	577	GLU	-	expression tag	UNP Q7Y3F0
E	578	HIS	-	expression tag	UNP Q7Y3F0
E	579	HIS	-	expression tag	UNP Q7Y3F0
E	580	HIS	-	expression tag	UNP Q7Y3F0
E	581	HIS	-	expression tag	UNP Q7Y3F0
E	582	HIS	-	expression tag	UNP Q7Y3F0
E	583	HIS	-	expression tag	UNP Q7Y3F0
F	575	SER	-	expression tag	UNP Q7Y3F0
F	576	LEU	-	expression tag	UNP Q7Y3F0
F	577	GLU	-	expression tag	UNP Q7Y3F0
F	578	HIS	-	expression tag	UNP Q7Y3F0
F	579	HIS	-	expression tag	UNP Q7Y3F0
F	580	HIS	-	expression tag	UNP Q7Y3F0
F	581	HIS	-	expression tag	UNP Q7Y3F0
F	582	HIS	-	expression tag	UNP Q7Y3F0
F	583	HIS	-	expression tag	UNP Q7Y3F0

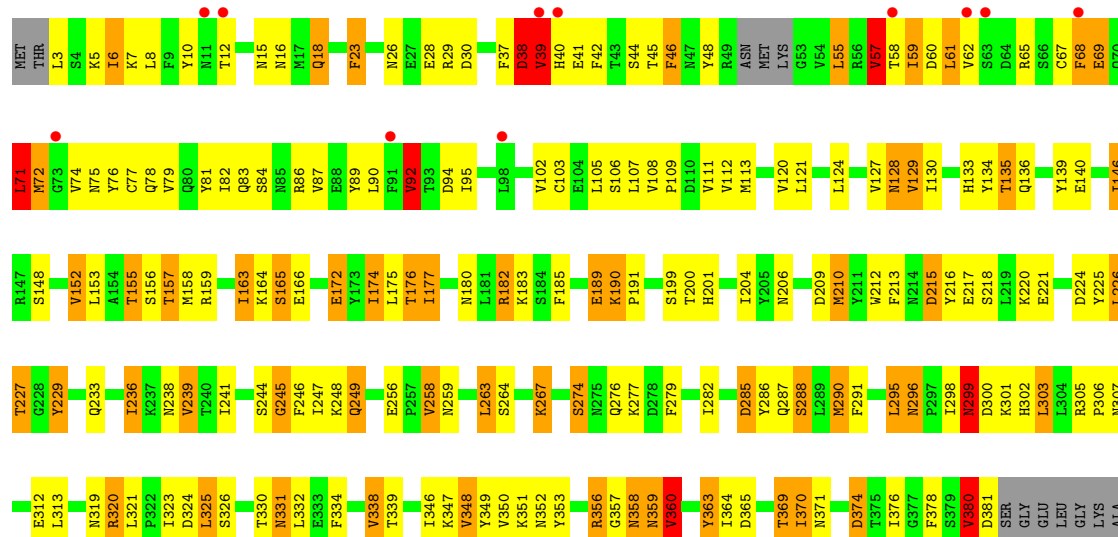


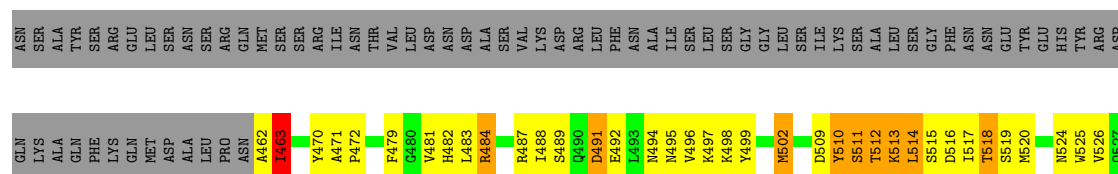
- Molecule 1: Major tail protein



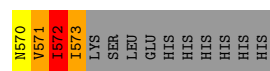
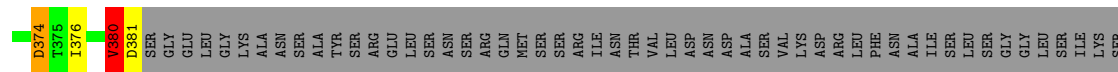
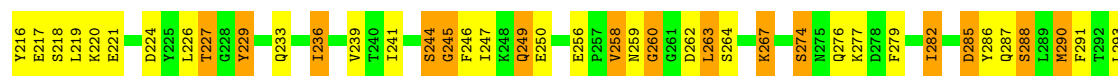
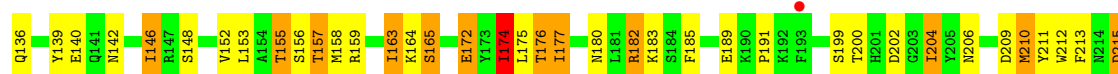


• Molecule 1: Major tail protein





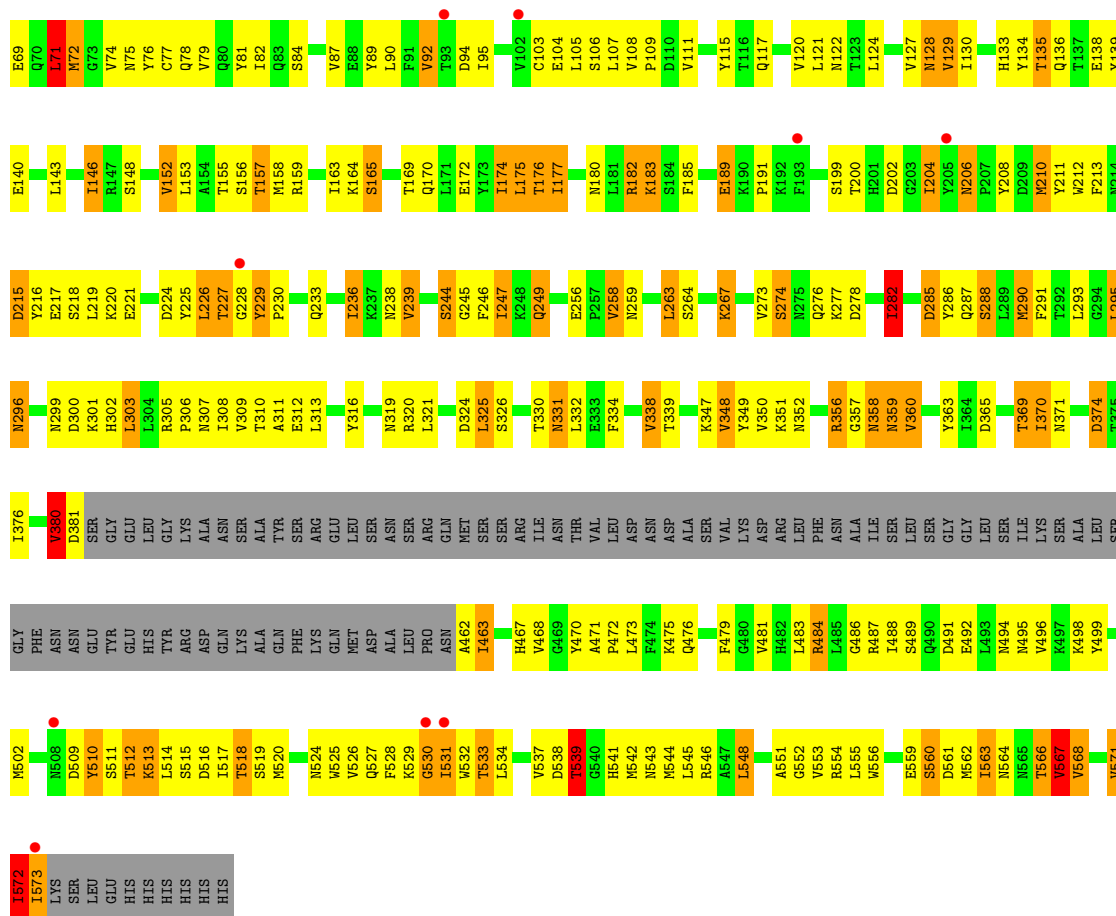
• Molecule 1: Major tail protein



• Molecule 1: Major tail protein









## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	205.61Å 209.64Å 102.97Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.01 50.00 – 3.01	Depositor EDS
% Data completeness (in resolution range)	98.0 (50.00-3.01) 98.0 (50.00-3.01)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.21 (at 3.01Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.238 , 0.282 0.238 , 0.283	Depositor DCC
$R_{free}$ test set	4404 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	75.3	Xtriage
Anisotropy	0.837	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 105.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.000 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	23688	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	108.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.10% 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, 0.333 respectively for untwinned datasets, and 0.375, 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	1.16	13/4030 (0.3%)	1.30	36/5464 (0.7%)
1	B	1.15	15/4030 (0.4%)	1.28	26/5464 (0.5%)
1	C	1.04	6/4030 (0.1%)	1.20	29/5464 (0.5%)
1	D	0.98	2/4030 (0.0%)	1.20	26/5464 (0.5%)
1	E	1.07	9/4030 (0.2%)	1.22	23/5464 (0.4%)
1	F	1.11	9/4030 (0.2%)	1.25	24/5464 (0.4%)
All	All	1.09	54/24180 (0.2%)	1.24	164/32784 (0.5%)

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	A	0	2
1	B	0	3
1	C	0	2
1	D	0	3
1	E	0	3
1	F	0	2
All	All	0	15

The worst 5 of 54 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	299	ASN	CG-ND2	-10.94	1.10	1.33
1	E	551	ALA	CA-CB	-8.42	1.40	1.53
1	E	152	VAL	CA-CB	-8.22	1.46	1.55
1	B	299	ASN	CG-OD1	-8.16	1.08	1.23
1	C	299	ASN	CG-ND2	-8.09	1.16	1.33

The worst 5 of 164 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	F	511	SER	N-CA-C	9.52	122.92	111.02
1	A	511	SER	N-CA-C	8.74	121.94	111.02
1	B	511	SER	N-CA-C	8.61	120.98	110.91
1	A	246	PHE	N-CA-C	-8.39	103.06	113.38
1	E	511	SER	N-CA-C	8.07	121.11	111.02

There are no chirality outliers.

5 of 15 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	356	ARG	Peptide
1	A	71	LEU	Peptide
1	B	183	LYS	Peptide
1	B	356	ARG	Peptide
1	B	71	LEU	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	3948	0	3832	280	3
1	B	3948	0	3832	281	3
1	C	3948	0	3832	303	1
1	D	3948	0	3832	303	0
1	E	3948	0	3832	285	0
1	F	3948	0	3832	294	1
All	All	23688	0	22992	1682	4

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 36.

The worst 5 of 1682 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:A:573:ILE:O	1:A:573:ILE:HG22	1.66	0.93
1:F:573:ILE:O	1:F:573:ILE:HG22	1.70	0.91

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:158:MET:O	1:D:159:ARG:HD3	1.70	0.91
1:E:573:ILE:HG22	1:E:573:ILE:O	1.71	0.91
1:A:290:MET:HE3	1:A:295:LEU:HB3	1.52	0.90

All (4) 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)
1:A:255:HIS:O	1:B:291:PHE:CZ[2_665]	1.96	0.24
1:C:295:LEU:O	1:F:65:ARG:NH2[1_556]	2.07	0.13
1:A:255:HIS:O	1:B:291:PHE:CE2[2_665]	2.09	0.11
1:A:255:HIS:C	1:B:291:PHE:CZ[2_665]	2.19	0.01

## 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	482/583 (83%)	386 (80%)	82 (17%)	14 (3%)	3	20
1	B	482/583 (83%)	390 (81%)	82 (17%)	10 (2%)	5	27
1	C	482/583 (83%)	387 (80%)	86 (18%)	9 (2%)	6	30
1	D	482/583 (83%)	388 (80%)	81 (17%)	13 (3%)	4	22
1	E	482/583 (83%)	390 (81%)	85 (18%)	7 (2%)	8	35
1	F	482/583 (83%)	387 (80%)	85 (18%)	10 (2%)	5	27
All	All	2892/3498 (83%)	2328 (80%)	501 (17%)	63 (2%)	5	26

5 of 63 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	258	VAL
1	A	530	GLY

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	539	THR
1	B	258	VAL
1	B	530	GLY

### 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	446/529 (84%)	358 (80%)	88 (20%)	1	8
1	B	446/529 (84%)	354 (79%)	92 (21%)	1	7
1	C	446/529 (84%)	357 (80%)	89 (20%)	1	7
1	D	446/529 (84%)	361 (81%)	85 (19%)	1	8
1	E	446/529 (84%)	351 (79%)	95 (21%)	1	6
1	F	446/529 (84%)	358 (80%)	88 (20%)	1	8
All	All	2676/3174 (84%)	2139 (80%)	537 (20%)	1	7

5 of 537 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	74	VAL
1	F	176	THR
1	F	72	MET
1	F	491	ASP
1	C	72	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 40 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	302	HIS
1	F	96	GLN
1	E	366	ASN
1	E	564	ASN
1	F	233	GLN

### 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 oligosaccharides 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 [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	488/583 (83%)	0.41	28 (5%)	29	15	48, 91, 189, 275	0
1	B	488/583 (83%)	0.40	37 (7%)	20	10	50, 92, 180, 276	0
1	C	488/583 (83%)	0.23	12 (2%)	58	35	58, 102, 192, 278	0
1	D	488/583 (83%)	0.26	17 (3%)	47	27	65, 104, 205, 279	0
1	E	488/583 (83%)	0.17	15 (3%)	51	30	58, 103, 198, 275	0
1	F	488/583 (83%)	0.21	23 (4%)	36	19	52, 94, 194, 281	0
All	All	2928/3498 (83%)	0.28	132 (4%)	38	20	48, 99, 194, 281	0

The worst 5 of 132 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	98	LEU	6.3
1	D	58	THR	5.9
1	F	13	PRO	4.8
1	A	40	HIS	4.6
1	A	69	GLU	4.3

### 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 oligosaccharides 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.