



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 05:34 am GMT

PDB ID : 4K7H  
Title : Major capsid protein P1 of the Pseudomonas phage phi6  
Authors : Boura, E.; Nemecek, D.; Plevka, P.; Steven, C.A.; Hurley, J.H.  
Deposited on : 2013-04-17  
Resolution : 3.60 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28949

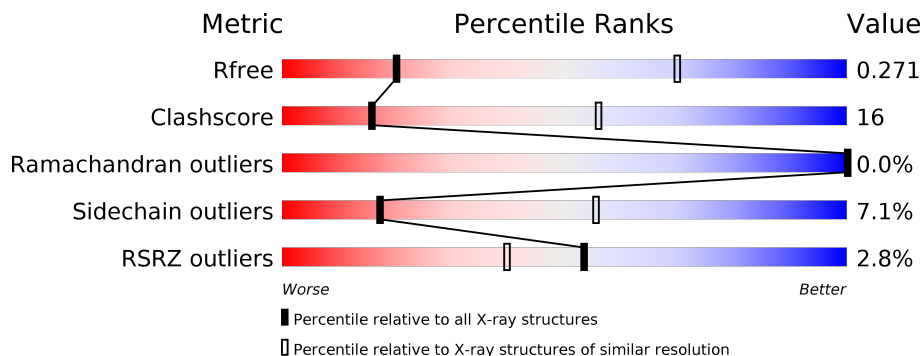
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.60 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))
$R_{free}$	100719	1026 (3.74-3.46)
Clashscore	112137	1036 (3.70-3.50)
Ramachandran outliers	110173	1030 (3.72-3.48)
Sidechain outliers	110143	1030 (3.72-3.48)
RSRZ outliers	101464	1051 (3.74-3.46)

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	775	<div> <div>2%</div> <div> <div></div> <div>59%</div> <div>35%</div> <div>• •</div> </div> </div>
1	B	775	<div> <div>3%</div> <div> <div></div> <div>63%</div> <div>31%</div> <div>• •</div> </div> </div>
1	C	775	<div> <div>3%</div> <div> <div></div> <div>63%</div> <div>32%</div> <div>• •</div> </div> </div>
1	D	775	<div> <div>3%</div> <div> <div></div> <div>62%</div> <div>32%</div> <div>• •</div> </div> </div>
1	E	775	<div> <div>2%</div> <div> <div></div> <div>65%</div> <div>29%</div> <div>• •</div> </div> </div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 29340 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 inner protein P1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	760	Total	C	N	O	S	0	0	0
			5868	3709	1032	1105	22			
1	B	760	Total	C	N	O	S	0	0	0
			5868	3709	1032	1105	22			
1	C	760	Total	C	N	O	S	0	0	0
			5868	3709	1032	1105	22			
1	D	760	Total	C	N	O	S	0	0	0
			5868	3709	1032	1105	22			
1	E	760	Total	C	N	O	S	0	0	0
			5868	3709	1032	1105	22			

There are 35 discrepancies between the modelled and reference sequences:

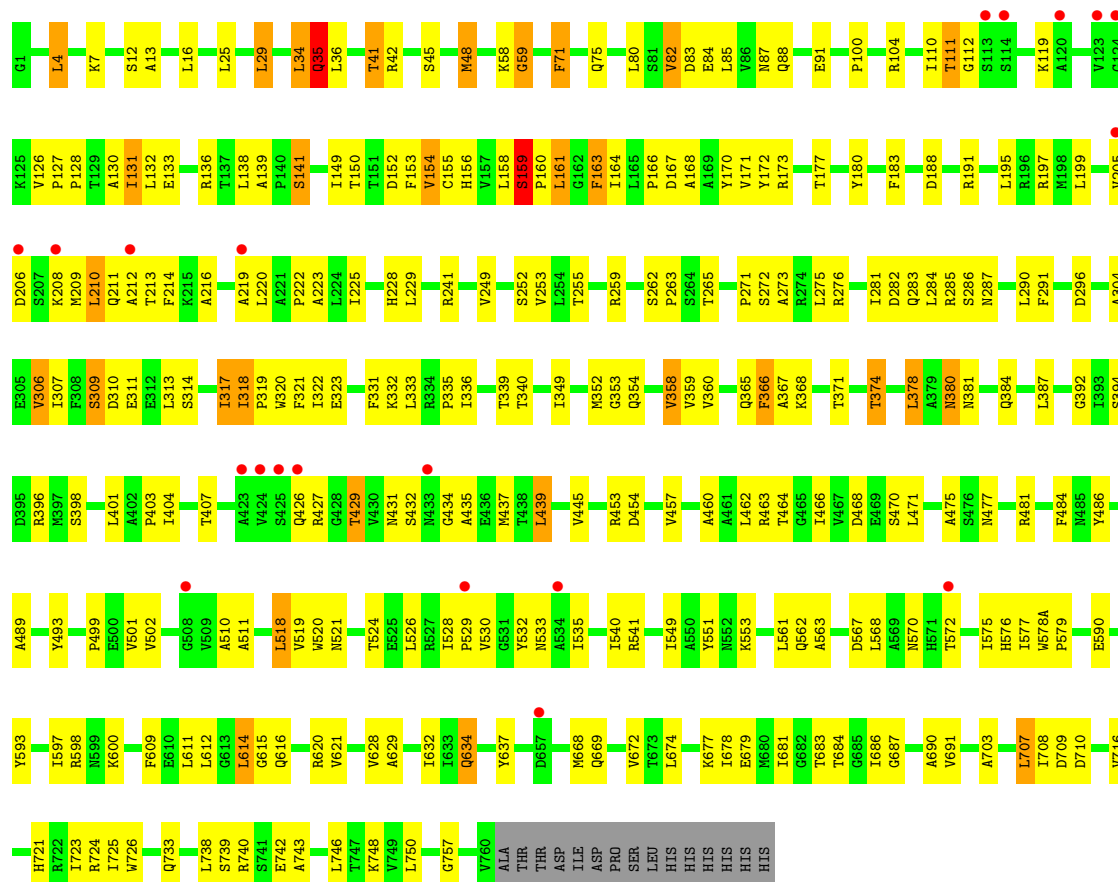
Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	EXPRESSION TAG	UNP P11126
A	770	HIS	-	EXPRESSION TAG	UNP P11126
A	771	HIS	-	EXPRESSION TAG	UNP P11126
A	772	HIS	-	EXPRESSION TAG	UNP P11126
A	773	HIS	-	EXPRESSION TAG	UNP P11126
A	774	HIS	-	EXPRESSION TAG	UNP P11126
A	775	HIS	-	EXPRESSION TAG	UNP P11126
B	1	GLY	-	EXPRESSION TAG	UNP P11126
B	770	HIS	-	EXPRESSION TAG	UNP P11126
B	771	HIS	-	EXPRESSION TAG	UNP P11126
B	772	HIS	-	EXPRESSION TAG	UNP P11126
B	773	HIS	-	EXPRESSION TAG	UNP P11126
B	774	HIS	-	EXPRESSION TAG	UNP P11126
B	775	HIS	-	EXPRESSION TAG	UNP P11126
C	1	GLY	-	EXPRESSION TAG	UNP P11126
C	770	HIS	-	EXPRESSION TAG	UNP P11126
C	771	HIS	-	EXPRESSION TAG	UNP P11126
C	772	HIS	-	EXPRESSION TAG	UNP P11126
C	773	HIS	-	EXPRESSION TAG	UNP P11126

*Continued on next page...*

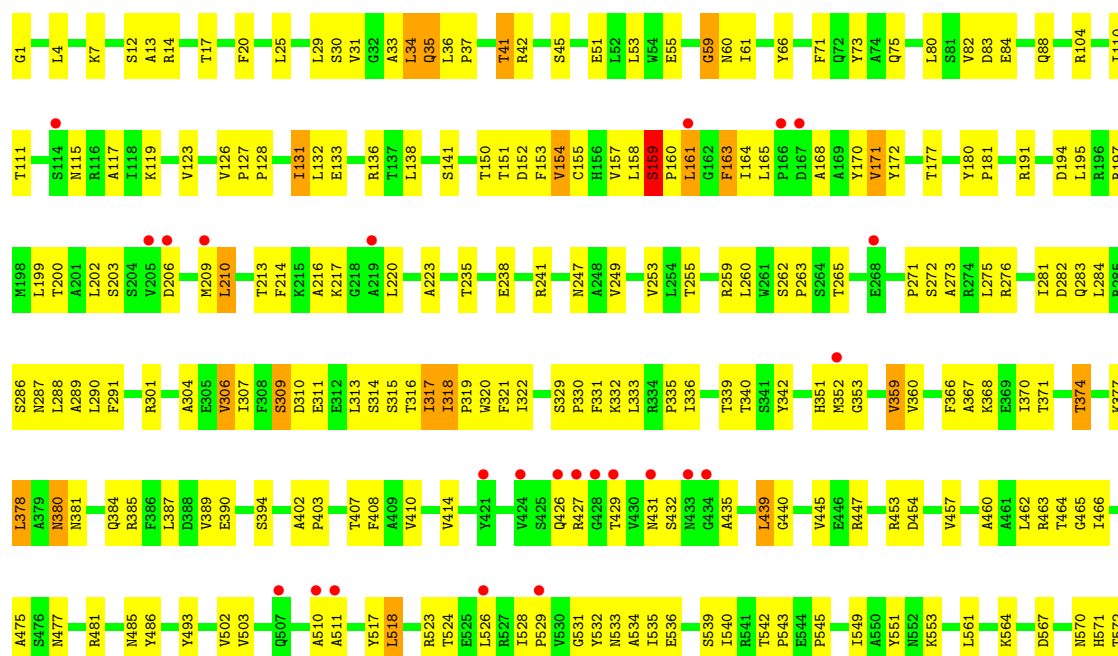
*Continued from previous page...*

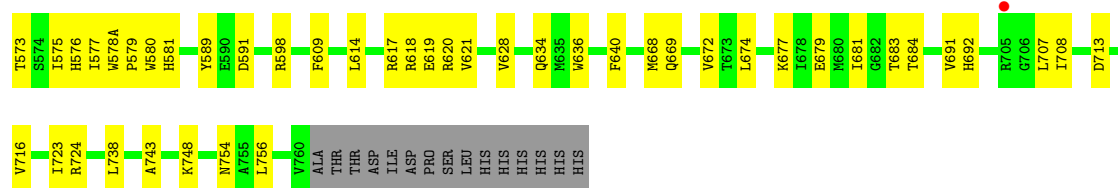
Chain	Residue	Modelled	Actual	Comment	Reference
C	774	HIS	-	EXPRESSION TAG	UNP P11126
C	775	HIS	-	EXPRESSION TAG	UNP P11126
D	1	GLY	-	EXPRESSION TAG	UNP P11126
D	770	HIS	-	EXPRESSION TAG	UNP P11126
D	771	HIS	-	EXPRESSION TAG	UNP P11126
D	772	HIS	-	EXPRESSION TAG	UNP P11126
D	773	HIS	-	EXPRESSION TAG	UNP P11126
D	774	HIS	-	EXPRESSION TAG	UNP P11126
D	775	HIS	-	EXPRESSION TAG	UNP P11126
E	1	GLY	-	EXPRESSION TAG	UNP P11126
E	770	HIS	-	EXPRESSION TAG	UNP P11126
E	771	HIS	-	EXPRESSION TAG	UNP P11126
E	772	HIS	-	EXPRESSION TAG	UNP P11126
E	773	HIS	-	EXPRESSION TAG	UNP P11126
E	774	HIS	-	EXPRESSION TAG	UNP P11126
E	775	HIS	-	EXPRESSION TAG	UNP P11126



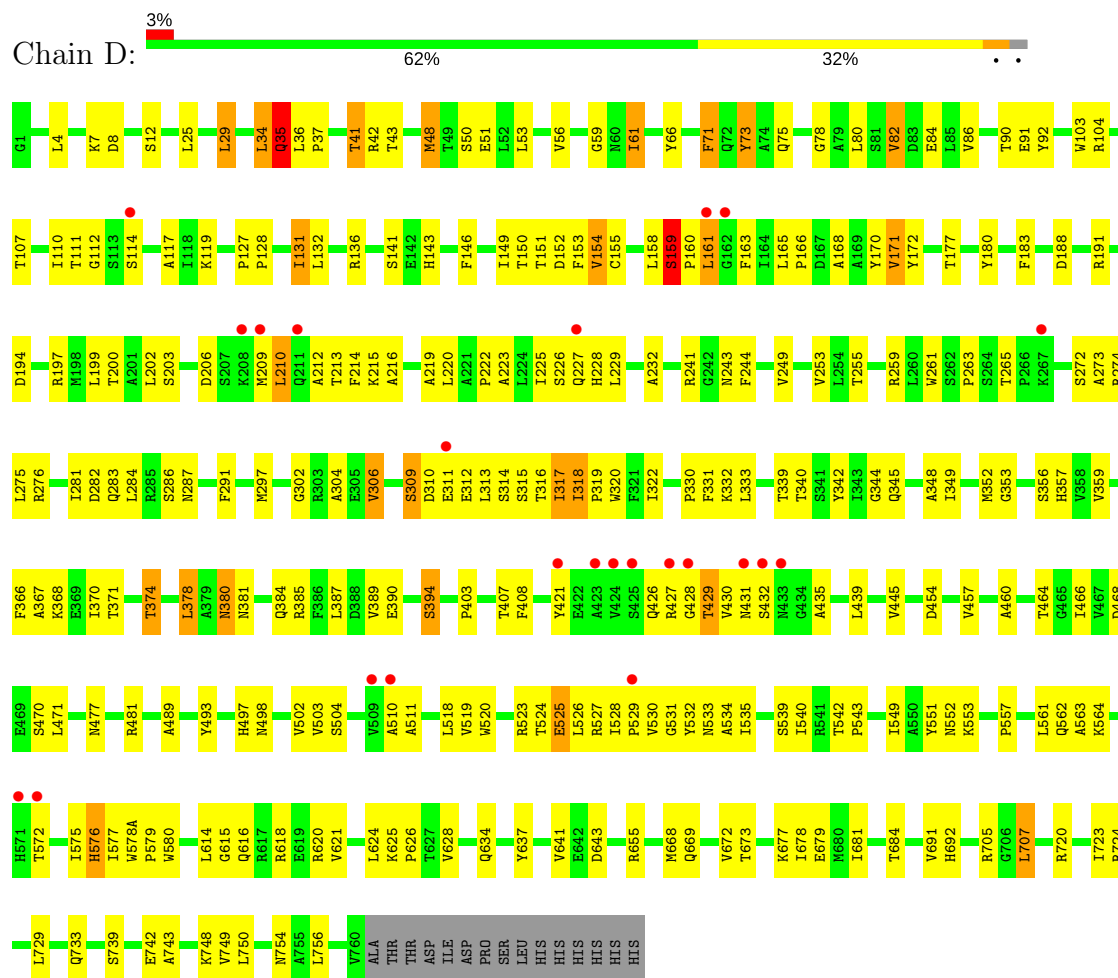


● Molecule 1: Major inner protein P1

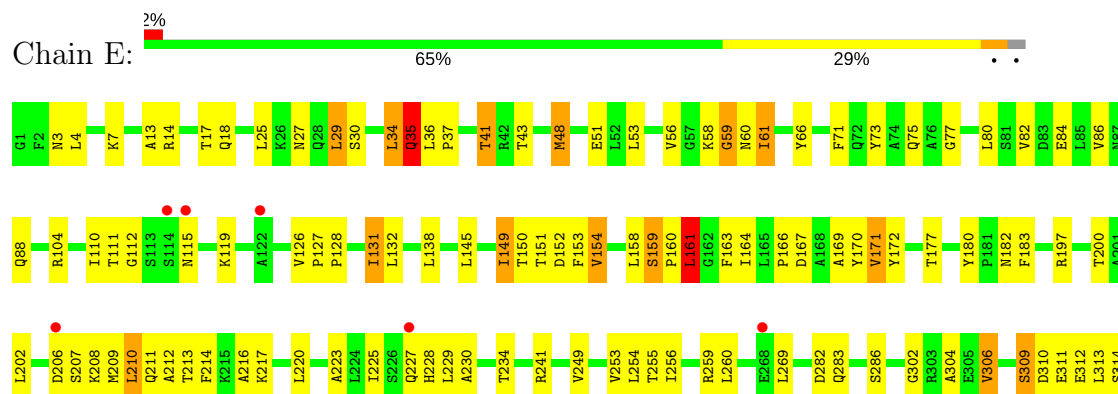


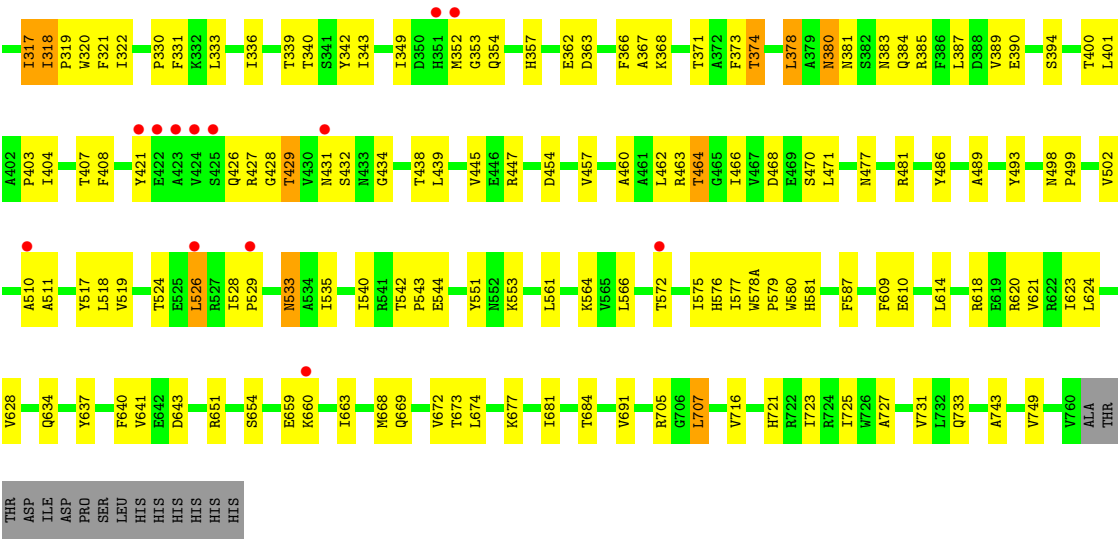


• Molecule 1: Major inner protein P1



• Molecule 1: Major inner protein P1





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	182.59Å 278.85Å 246.47Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.92 – 3.60 40.92 – 3.60	Depositor EDS
% Data completeness (in resolution range)	91.0 (40.92-3.60) 91.0 (40.92-3.60)	Depositor EDS
$R_{merge}$	0.19	Depositor
$R_{sym}$	0.19	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.08 (at 3.57Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, $R_{free}$	0.217 , 0.274 0.212 , 0.271	Depositor DCC
$R_{free}$ test set	3346 reflections (5.03%)	DCC
Wilson B-factor (Å <sup>2</sup> )	118.3	Xtriage
Anisotropy	0.115	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 150.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	29340	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	178.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.79% 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	0.41	0/5988	0.67	5/8146 (0.1%)
1	B	0.42	0/5988	0.67	5/8146 (0.1%)
1	C	0.40	0/5988	0.66	3/8146 (0.0%)
1	D	0.42	0/5988	0.68	4/8146 (0.0%)
1	E	0.43	0/5988	0.69	6/8146 (0.1%)
All	All	0.42	0/29940	0.68	23/40730 (0.1%)

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	4
1	B	0	5
1	C	0	3
1	D	0	5
1	E	0	4
All	All	0	21

There are no bond length outliers.

The worst 5 of 23 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	161	LEU	CA-CB-CG	6.90	131.16	115.30
1	B	161	LEU	CA-CB-CG	6.81	130.97	115.30
1	E	59	GLY	N-CA-C	-6.64	96.49	113.10
1	B	59	GLY	N-CA-C	-6.42	97.06	113.10
1	D	161	LEU	CA-CB-CG	6.32	129.83	115.30

There are no chirality outliers.

5 of 21 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	159	SER	Peptide
1	A	318	ILE	Peptide
1	A	34	LEU	Peptide
1	A	35	GLN	Peptide
1	B	34	LEU	Peptide

## 5.2 Too-close contacts

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	5868	0	5807	213	0
1	B	5868	0	5807	203	0
1	C	5868	0	5807	188	1
1	D	5868	0	5807	195	2
1	E	5868	0	5807	169	1
All	All	29340	0	29035	949	2

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 16.

The worst 5 of 949 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:51:GLU:H	1:C:564:LYS:HE3	1.32	0.92
1:D:367:ALA:HB1	1:D:368:LYS:HA	1.51	0.91
1:A:51:GLU:H	1:A:564:LYS:HE3	1.36	0.90
1:E:318:ILE:HG22	1:E:320:TRP:HB3	1.54	0.89
1:E:367:ALA:HB1	1:E:368:LYS:HA	1.55	0.86

All (2) 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:C:523:ARG:O	1:D:523:ARG:NH2[8_455]	2.12	0.08
1:D:655:ARG:O	1:E:27:ASN:ND2[4_555]	2.18	0.02

## 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 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	758/775 (98%)	717 (95%)	41 (5%)	0	100	100
1	B	758/775 (98%)	720 (95%)	38 (5%)	0	100	100
1	C	758/775 (98%)	718 (95%)	40 (5%)	0	100	100
1	D	758/775 (98%)	717 (95%)	41 (5%)	0	100	100
1	E	758/775 (98%)	722 (95%)	35 (5%)	1 (0%)	55	88
All	All	3790/3875 (98%)	3594 (95%)	195 (5%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	61	ILE

### 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	618/643 (96%)	575 (93%)	43 (7%)	18	56
1	B	618/643 (96%)	576 (93%)	42 (7%)	18	57
1	C	618/643 (96%)	579 (94%)	39 (6%)	21	60
1	D	618/643 (96%)	572 (93%)	46 (7%)	16	53
1	E	618/643 (96%)	569 (92%)	49 (8%)	14	51
All	All	3090/3215 (96%)	2871 (93%)	219 (7%)	17	55

5 of 219 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	309	SER
1	C	716	VAL
1	E	429	THR
1	C	317	ILE
1	C	429	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	533	ASN
1	B	228	HIS
1	D	228	HIS
1	E	156	HIS

### 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 [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	760/775 (98%)	-0.24	19 (2%)	58	43	130, 172, 242, 296	0
1	B	760/775 (98%)	-0.20	20 (2%)	56	42	125, 172, 241, 296	0
1	C	760/775 (98%)	-0.21	25 (3%)	47	35	123, 170, 243, 295	0
1	D	760/775 (98%)	-0.20	23 (3%)	51	37	122, 169, 241, 293	0
1	E	760/775 (98%)	-0.20	19 (2%)	58	43	122, 168, 241, 296	0
All	All	3800/3875 (98%)	-0.21	106 (2%)	53	40	122, 171, 241, 296	0

The worst 5 of 106 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	425	SER	6.1
1	D	529	PRO	5.4
1	D	161	LEU	5.2
1	C	529	PRO	4.9
1	B	426	GLN	4.9

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.