



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 18, 2024 – 09:26 AM EST

PDB ID : 4EO2  
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 : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.36  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.36

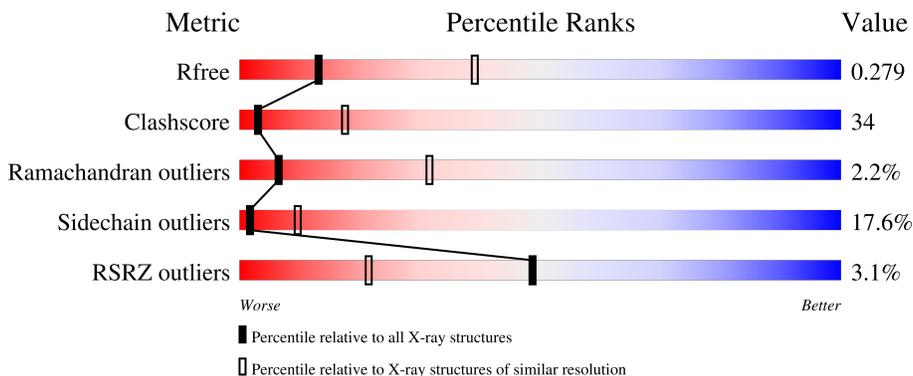
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.01 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.



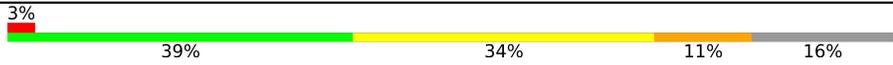
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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	
1	B	583	
1	C	583	
1	D	583	
1	E	583	

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	583	 <p>3% 39% 34% 11% 16%</p>

## 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
			Total	C	N	O	S			
1	A	488	3948	2504	665	761	18	0	0	0
1	B	488	3948	2504	665	761	18	0	0	0
1	C	488	3948	2504	665	761	18	0	0	0
1	D	488	3948	2504	665	761	18	0	0	0
1	E	488	3948	2504	665	761	18	0	0	0
1	F	488	3948	2504	665	761	18	0	0	0

There are 54 discrepancies between the modelled and reference sequences:

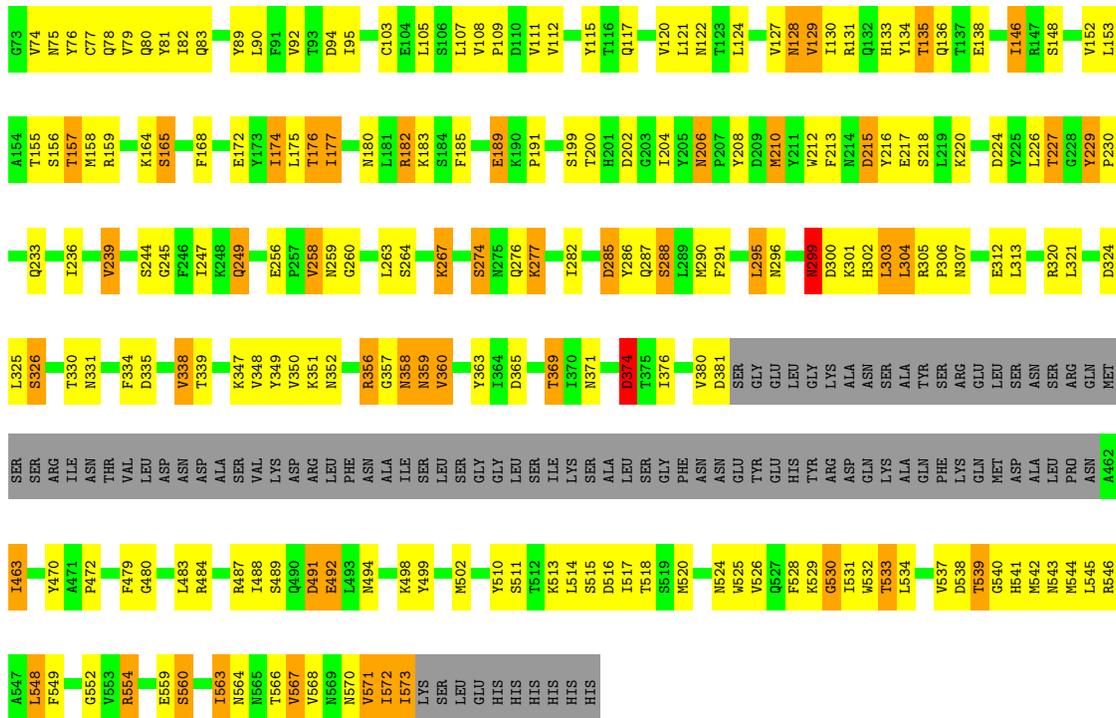
Chain	Residue	Modelled	Actual	Comment	Reference
A	575	SER	-	expression tag	UNP Q7Y3F0
A	576	LEU	-	expression tag	UNP Q7Y3F0
A	577	GLU	-	expression tag	UNP Q7Y3F0
A	578	HIS	-	expression tag	UNP Q7Y3F0
A	579	HIS	-	expression tag	UNP Q7Y3F0
A	580	HIS	-	expression tag	UNP Q7Y3F0
A	581	HIS	-	expression tag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

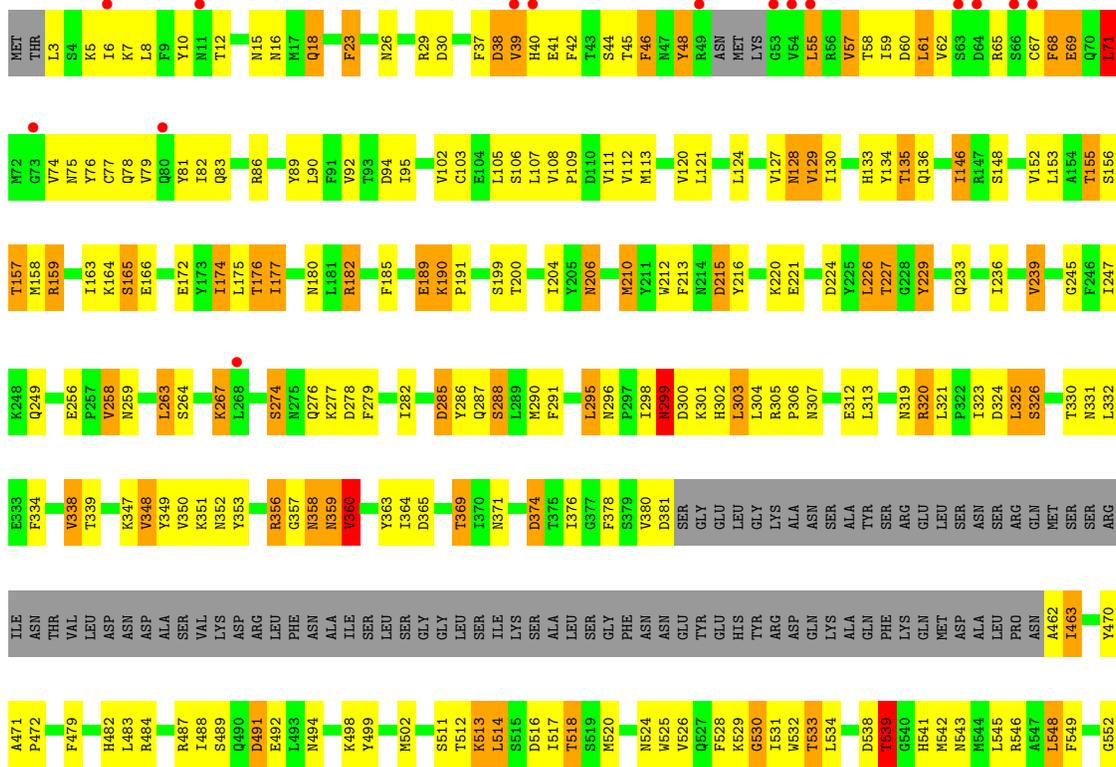
*Continued from previous page...*

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

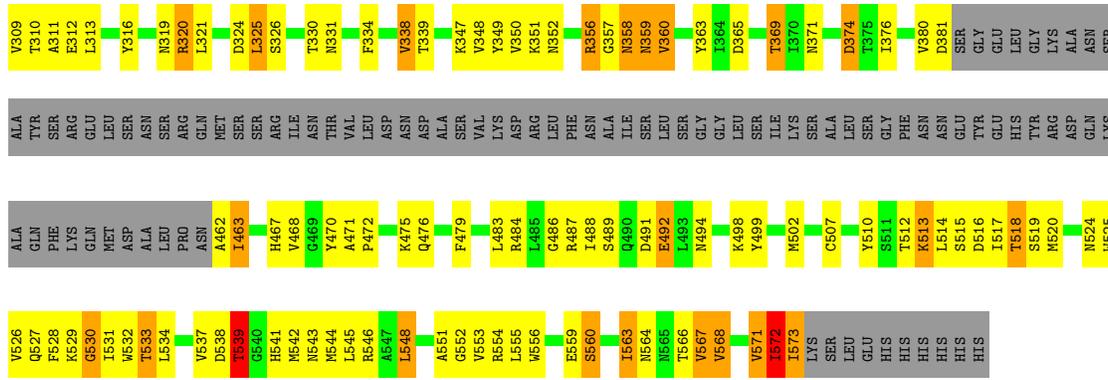




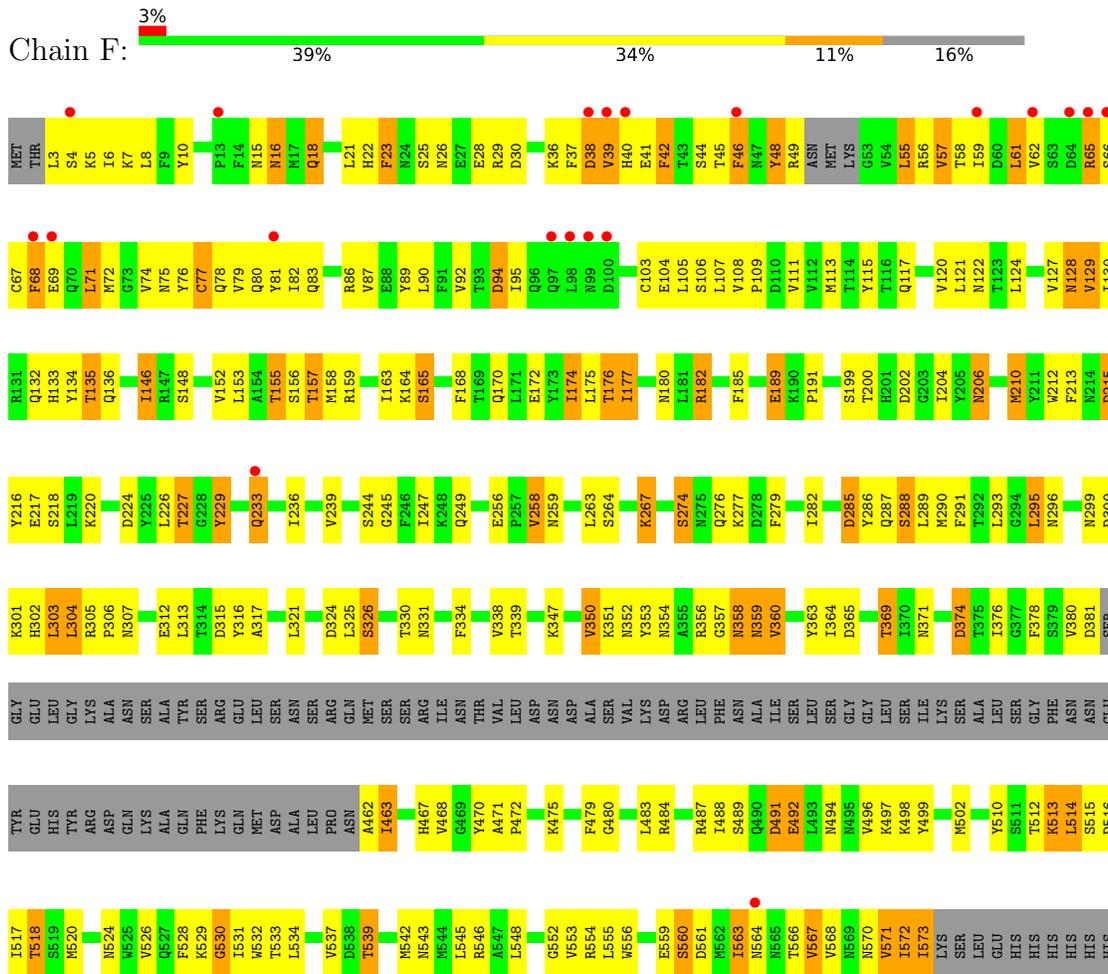
• Molecule 1: Major tail protein







● Molecule 1: Major tail protein



## 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) 98.1 (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)	Xtrriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.238 , 0.282 0.235 , 0.279	Depositor DCC
$R_{free}$ test set	4390 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	75.3	Xtrriage
Anisotropy	0.837	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 98.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.000 for k,h,-l	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	23688	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	108.0	wwPDB-VP

Xtrriage'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 [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	0.90	0/4030	0.97	4/5464 (0.1%)
1	B	0.89	3/4030 (0.1%)	0.94	3/5464 (0.1%)
1	C	0.80	1/4030 (0.0%)	0.89	2/5464 (0.0%)
1	D	0.77	0/4030	0.88	1/5464 (0.0%)
1	E	0.81	2/4030 (0.0%)	0.90	1/5464 (0.0%)
1	F	0.87	1/4030 (0.0%)	0.94	1/5464 (0.0%)
All	All	0.84	7/24180 (0.0%)	0.92	12/32784 (0.0%)

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 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	299	ASN	CG-ND2	-9.03	1.10	1.32
1	B	299	ASN	CG-OD1	-7.23	1.08	1.24
1	C	299	ASN	CG-ND2	-6.63	1.16	1.32
1	F	492	GLU	CG-CD	5.74	1.60	1.51
1	E	551	ALA	CA-CB	-5.69	1.40	1.52

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	F	304	LEU	CB-CG-CD1	-8.30	96.89	111.00
1	B	374	ASP	CB-CG-OD2	-6.55	112.40	118.30
1	C	304	LEU	CB-CG-CD1	-5.99	100.81	111.00
1	B	304	LEU	CB-CG-CD1	-5.98	100.83	111.00
1	D	365	ASP	CB-CG-OD2	5.83	123.55	118.30

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	264	3
1	B	3948	0	3832	264	3
1	C	3948	0	3832	282	1
1	D	3948	0	3832	284	0
1	E	3948	0	3832	271	0
1	F	3948	0	3832	279	1
All	All	23688	0	22992	1585	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 34.

The worst 5 of 1585 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:C:299:ASN:N	1:C:299:ASN:HD22	1.63	0.89

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%)	4	24
1	B	482/583 (83%)	390 (81%)	82 (17%)	10 (2%)	7	33
1	C	482/583 (83%)	387 (80%)	86 (18%)	9 (2%)	8	36
1	D	482/583 (83%)	388 (80%)	81 (17%)	13 (3%)	5	26
1	E	482/583 (83%)	390 (81%)	85 (18%)	7 (2%)	10	42
1	F	482/583 (83%)	387 (80%)	85 (18%)	10 (2%)	7	33
All	All	2892/3498 (83%)	2328 (80%)	501 (17%)	63 (2%)	6	31

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 [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	446/529 (84%)	368 (82%)	78 (18%)	2	10
1	B	446/529 (84%)	369 (83%)	77 (17%)	2	10
1	C	446/529 (84%)	369 (83%)	77 (17%)	2	10
1	D	446/529 (84%)	367 (82%)	79 (18%)	2	9
1	E	446/529 (84%)	365 (82%)	81 (18%)	1	9
1	F	446/529 (84%)	367 (82%)	79 (18%)	2	9
All	All	2676/3174 (84%)	2205 (82%)	471 (18%)	2	10

5 of 471 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	573	ILE
1	F	303	LEU
1	D	330	THR
1	F	274	SER
1	F	55	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	299	ASN
1	F	233	GLN
1	D	40	HIS
1	F	564	ASN
1	E	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 monosaccharides 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%)	-0.12	11 (2%) 60 31	48, 91, 189, 275	0
1	B	488/583 (83%)	-0.31	4 (0%) 86 65	50, 92, 180, 276	0
1	C	488/583 (83%)	-0.23	15 (3%) 49 21	58, 102, 192, 278	0
1	D	488/583 (83%)	-0.12	25 (5%) 28 10	65, 104, 205, 279	0
1	E	488/583 (83%)	-0.19	15 (3%) 49 21	58, 103, 198, 275	0
1	F	488/583 (83%)	-0.24	20 (4%) 37 14	52, 94, 194, 281	0
All	All	2928/3498 (83%)	-0.20	90 (3%) 49 21	48, 99, 194, 281	0

The worst 5 of 90 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	39	VAL	9.7
1	F	39	VAL	8.7
1	B	39	VAL	7.5
1	D	40	HIS	7.2
1	A	68	PHE	7.2

### 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 monosaccharides 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.