



# Full wwPDB X-ray Structure Validation Report ⓘ

May 22, 2020 – 05:24 pm BST

PDB ID : 4A0U  
Title : Structure of the carboxy-terminal domain of bacteriophage T7 fibre gp17 containing residues 371-553, C2221 crystal form.  
Authors : Garcia-Doval, C.; van Raaij, M.J.  
Deposited on : 2011-09-12  
Resolution : 2.00 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
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) : 2.11

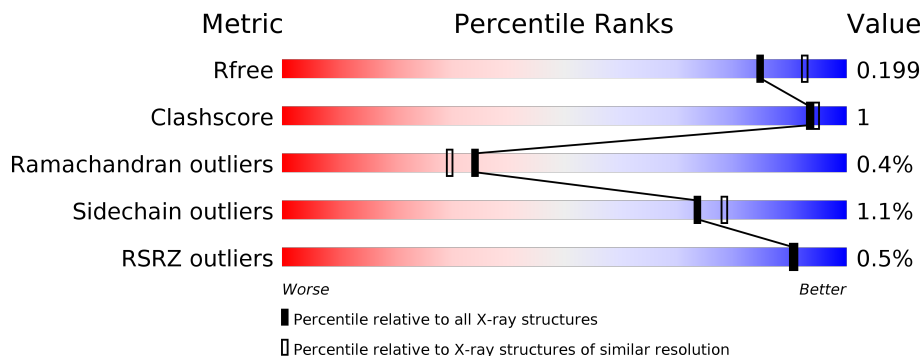
# 1 Overall quality at a glance

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))
$R_{free}$	130704	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (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 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	228	 79% 20%
1	B	228	 78% 18%
1	C	228	 76% 20%

## 2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 5278 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 TAIL FIBER PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	183	1442	909	262	268	3	0	0	0
1	B	186	1476	929	269	274	4	0	2	0
1	C	183	1442	909	262	268	3	0	0	0

There are 138 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	326	MET	-	expression tag	UNP P03748
A	327	HIS	-	expression tag	UNP P03748
A	328	HIS	-	expression tag	UNP P03748
A	329	HIS	-	expression tag	UNP P03748
A	330	HIS	-	expression tag	UNP P03748
A	331	HIS	-	expression tag	UNP P03748
A	332	HIS	-	expression tag	UNP P03748
A	333	SER	-	expression tag	UNP P03748
A	334	SER	-	expression tag	UNP P03748
A	335	GLY	-	expression tag	UNP P03748
A	336	LEU	-	expression tag	UNP P03748
A	337	VAL	-	expression tag	UNP P03748
A	338	PRO	-	expression tag	UNP P03748
A	339	ARG	-	expression tag	UNP P03748
A	340	GLY	-	expression tag	UNP P03748
A	341	SER	-	expression tag	UNP P03748
A	342	GLY	-	expression tag	UNP P03748
A	343	MET	-	expression tag	UNP P03748
A	344	LYS	-	expression tag	UNP P03748
A	345	GLU	-	expression tag	UNP P03748
A	346	THR	-	expression tag	UNP P03748
A	347	ALA	-	expression tag	UNP P03748
A	348	ALA	-	expression tag	UNP P03748

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	349	ALA	-	expression tag	UNP P03748
A	350	LYS	-	expression tag	UNP P03748
A	351	PHE	-	expression tag	UNP P03748
A	352	GLU	-	expression tag	UNP P03748
A	353	ARG	-	expression tag	UNP P03748
A	354	GLN	-	expression tag	UNP P03748
A	355	HIS	-	expression tag	UNP P03748
A	356	MET	-	expression tag	UNP P03748
A	357	ASP	-	expression tag	UNP P03748
A	358	SER	-	expression tag	UNP P03748
A	359	PRO	-	expression tag	UNP P03748
A	360	ASP	-	expression tag	UNP P03748
A	361	LEU	-	expression tag	UNP P03748
A	362	GLY	-	expression tag	UNP P03748
A	363	THR	-	expression tag	UNP P03748
A	364	ASP	-	expression tag	UNP P03748
A	365	ASP	-	expression tag	UNP P03748
A	366	ASP	-	expression tag	UNP P03748
A	367	ASP	-	expression tag	UNP P03748
A	368	LYS	-	expression tag	UNP P03748
A	369	ALA	-	expression tag	UNP P03748
A	370	MET	-	expression tag	UNP P03748
A	371	GLY	-	expression tag	UNP P03748
B	326	MET	-	expression tag	UNP P03748
B	327	HIS	-	expression tag	UNP P03748
B	328	HIS	-	expression tag	UNP P03748
B	329	HIS	-	expression tag	UNP P03748
B	330	HIS	-	expression tag	UNP P03748
B	331	HIS	-	expression tag	UNP P03748
B	332	HIS	-	expression tag	UNP P03748
B	333	SER	-	expression tag	UNP P03748
B	334	SER	-	expression tag	UNP P03748
B	335	GLY	-	expression tag	UNP P03748
B	336	LEU	-	expression tag	UNP P03748
B	337	VAL	-	expression tag	UNP P03748
B	338	PRO	-	expression tag	UNP P03748
B	339	ARG	-	expression tag	UNP P03748
B	340	GLY	-	expression tag	UNP P03748
B	341	SER	-	expression tag	UNP P03748
B	342	GLY	-	expression tag	UNP P03748
B	343	MET	-	expression tag	UNP P03748
B	344	LYS	-	expression tag	UNP P03748

*Continued on next page...*

*Continued from previous page...*

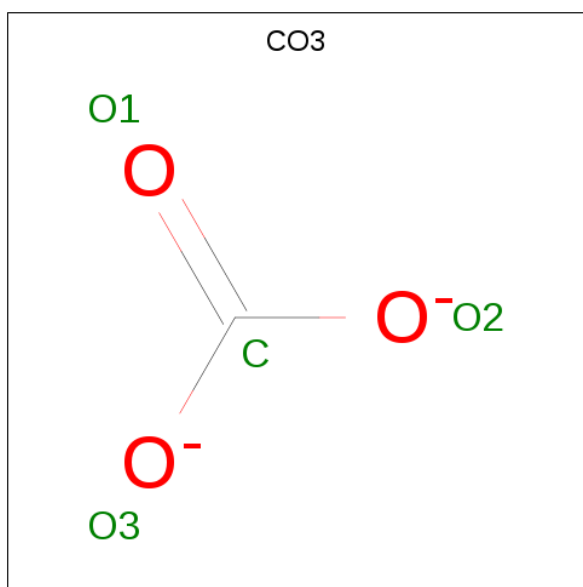
Chain	Residue	Modelled	Actual	Comment	Reference
B	345	GLU	-	expression tag	UNP P03748
B	346	THR	-	expression tag	UNP P03748
B	347	ALA	-	expression tag	UNP P03748
B	348	ALA	-	expression tag	UNP P03748
B	349	ALA	-	expression tag	UNP P03748
B	350	LYS	-	expression tag	UNP P03748
B	351	PHE	-	expression tag	UNP P03748
B	352	GLU	-	expression tag	UNP P03748
B	353	ARG	-	expression tag	UNP P03748
B	354	GLN	-	expression tag	UNP P03748
B	355	HIS	-	expression tag	UNP P03748
B	356	MET	-	expression tag	UNP P03748
B	357	ASP	-	expression tag	UNP P03748
B	358	SER	-	expression tag	UNP P03748
B	359	PRO	-	expression tag	UNP P03748
B	360	ASP	-	expression tag	UNP P03748
B	361	LEU	-	expression tag	UNP P03748
B	362	GLY	-	expression tag	UNP P03748
B	363	THR	-	expression tag	UNP P03748
B	364	ASP	-	expression tag	UNP P03748
B	365	ASP	-	expression tag	UNP P03748
B	366	ASP	-	expression tag	UNP P03748
B	367	ASP	-	expression tag	UNP P03748
B	368	LYS	-	expression tag	UNP P03748
B	369	ALA	-	expression tag	UNP P03748
B	370	MET	-	expression tag	UNP P03748
B	371	GLY	-	expression tag	UNP P03748
C	326	MET	-	expression tag	UNP P03748
C	327	HIS	-	expression tag	UNP P03748
C	328	HIS	-	expression tag	UNP P03748
C	329	HIS	-	expression tag	UNP P03748
C	330	HIS	-	expression tag	UNP P03748
C	331	HIS	-	expression tag	UNP P03748
C	332	HIS	-	expression tag	UNP P03748
C	333	SER	-	expression tag	UNP P03748
C	334	SER	-	expression tag	UNP P03748
C	335	GLY	-	expression tag	UNP P03748
C	336	LEU	-	expression tag	UNP P03748
C	337	VAL	-	expression tag	UNP P03748
C	338	PRO	-	expression tag	UNP P03748
C	339	ARG	-	expression tag	UNP P03748
C	340	GLY	-	expression tag	UNP P03748

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	341	SER	-	expression tag	UNP P03748
C	342	GLY	-	expression tag	UNP P03748
C	343	MET	-	expression tag	UNP P03748
C	344	LYS	-	expression tag	UNP P03748
C	345	GLU	-	expression tag	UNP P03748
C	346	THR	-	expression tag	UNP P03748
C	347	ALA	-	expression tag	UNP P03748
C	348	ALA	-	expression tag	UNP P03748
C	349	ALA	-	expression tag	UNP P03748
C	350	LYS	-	expression tag	UNP P03748
C	351	PHE	-	expression tag	UNP P03748
C	352	GLU	-	expression tag	UNP P03748
C	353	ARG	-	expression tag	UNP P03748
C	354	GLN	-	expression tag	UNP P03748
C	355	HIS	-	expression tag	UNP P03748
C	356	MET	-	expression tag	UNP P03748
C	357	ASP	-	expression tag	UNP P03748
C	358	SER	-	expression tag	UNP P03748
C	359	PRO	-	expression tag	UNP P03748
C	360	ASP	-	expression tag	UNP P03748
C	361	LEU	-	expression tag	UNP P03748
C	362	GLY	-	expression tag	UNP P03748
C	363	THR	-	expression tag	UNP P03748
C	364	ASP	-	expression tag	UNP P03748
C	365	ASP	-	expression tag	UNP P03748
C	366	ASP	-	expression tag	UNP P03748
C	367	ASP	-	expression tag	UNP P03748
C	368	LYS	-	expression tag	UNP P03748
C	369	ALA	-	expression tag	UNP P03748
C	370	MET	-	expression tag	UNP P03748
C	371	GLY	-	expression tag	UNP P03748

- Molecule 2 is CARBONATE ION (three-letter code: CO3) (formula: CO<sub>3</sub>).



Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	C O	0	0
			4	1 3		
2	B	1	Total	C O	0	0
			4	1 3		

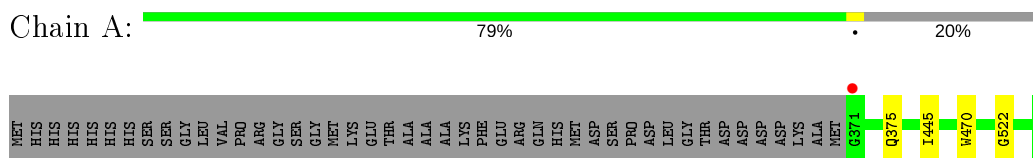
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	362	Total	O	0	0
			362	362		
3	B	301	Total	O	0	0
			301	301		
3	C	247	Total	O	0	0
			247	247		

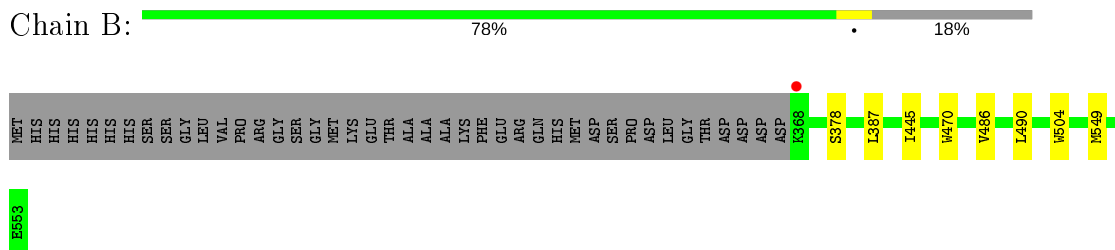
### 3 Residue-property plots [i](#)

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 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.

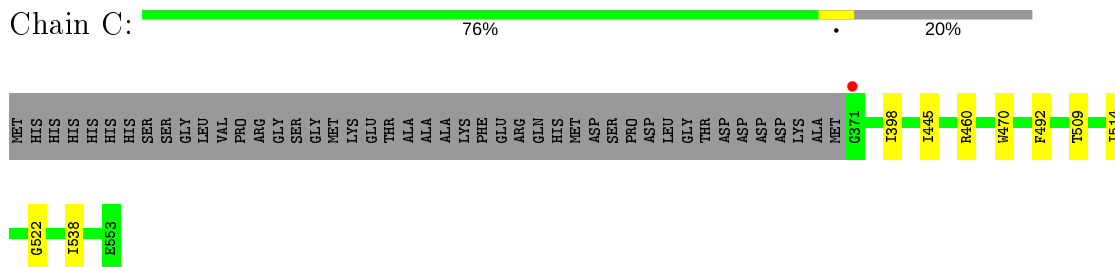
- Molecule 1: TAIL FIBER PROTEIN



- Molecule 1: TAIL FIBER PROTEIN



- Molecule 1: TAIL FIBER PROTEIN





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	68.32Å 145.56Å 172.05Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.95 – 2.00 19.95 – 2.00	Depositor EDS
% Data completeness (in resolution range)	100.0 (19.95-2.00) 100.0 (19.95-2.00)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.64 (at 2.01Å)	Xtrriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.149 , 0.197 0.150 , 0.199	Depositor DCC
$R_{free}$ test set	2274 reflections (3.91%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	17.9	Xtrriage
Anisotropy	0.462	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 57.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	5278	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	19.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.79% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CO3

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	0.75	0/1481	0.67	0/2007
1	B	0.80	0/1515	0.68	0/2051
1	C	0.74	0/1481	0.71	0/2007
All	All	0.76	0/4477	0.69	0/6065

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	1442	0	1362	5	0
1	B	1476	0	1397	7	0
1	C	1442	0	1362	7	0
2	A	4	0	0	0	0
2	B	4	0	0	0	0
3	A	362	0	0	0	0
3	B	301	0	0	0	0
3	C	247	0	0	1	0
All	All	5278	0	4121	12	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 (12) 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:445:ILE:HD11	1:B:445:ILE:CD1	2.16	0.76
1:A:445:ILE:HD13	1:C:445:ILE:HD11	1.79	0.65
1:A:445:ILE:HD11	1:B:445:ILE:HD13	1.79	0.64
1:C:514:ILE:HD12	1:C:538:ILE:HD11	1.85	0.57
1:B:445:ILE:HD11	1:C:445:ILE:HD13	1.87	0.56
1:B:445:ILE:HD11	1:C:445:ILE:CD1	2.40	0.52
1:A:445:ILE:CD1	1:C:445:ILE:HD11	2.40	0.51
1:A:375:GLN:NE2	1:B:387:LEU:HD23	2.29	0.47
1:B:504:TRP:CZ3	1:B:549:MET:HG2	2.52	0.45
1:C:460:ARG:NH2	3:C:2094:HOH:O	2.51	0.43
1:C:492:PHE:HA	1:C:509:THR:O	2.19	0.42
1:B:486:VAL:HG21	1:B:490:LEU:HG	2.02	0.41

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	181/228 (79%)	175 (97%)	5 (3%)	1 (1%)	25	19
1	B	186/228 (82%)	179 (96%)	7 (4%)	0	100	100
1	C	181/228 (79%)	176 (97%)	4 (2%)	1 (1%)	25	19
All	All	548/684 (80%)	530 (97%)	16 (3%)	2 (0%)	34	30

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	522	GLY

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	522	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	150/187 (80%)	149 (99%)	1 (1%)	84	88
1	B	153/187 (82%)	151 (99%)	2 (1%)	69	74
1	C	150/187 (80%)	148 (99%)	2 (1%)	69	74
All	All	453/561 (81%)	448 (99%)	5 (1%)	73	78

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	470	TRP
1	B	378	SER
1	B	470	TRP
1	C	398	ILE
1	C	470	TRP

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](#)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	CO3	A	1554	-	0,3,3	0.00	-	0,3,3	0.00	-
2	CO3	B	1554	-	0,3,3	0.00	-	0,3,3	0.00	-

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	183/228 (80%)	-0.70	1 (0%) 91 90	11, 16, 23, 38	0
1	B	186/228 (81%)	-0.69	1 (0%) 91 90	10, 16, 24, 35	0
1	C	183/228 (80%)	-0.78	1 (0%) 91 90	11, 16, 24, 38	0
All	All	552/684 (80%)	-0.72	3 (0%) 91 90	10, 16, 24, 38	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	371	GLY	3.4
1	C	371	GLY	3.1
1	B	368	LYS	2.3

### 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 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. 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	B-factors(Å <sup>2</sup> )	Q<0.9
2	CO3	B	1554	4/4	0.94	0.19	39,39,40,41	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	CO3	A	1554	4/4	0.95	0.20	42,42,42,43	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.