



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 07:40 AM GMT

PDB ID : 3BQW  
Title : Crystal structure of the putative capsid protein of prophage (E.coli CFT073)  
Authors : Zhang, R.; Hatzos, C.; Abdullah, J.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)  
Deposited on : 2007-12-20  
Resolution : 2.20 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  
<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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

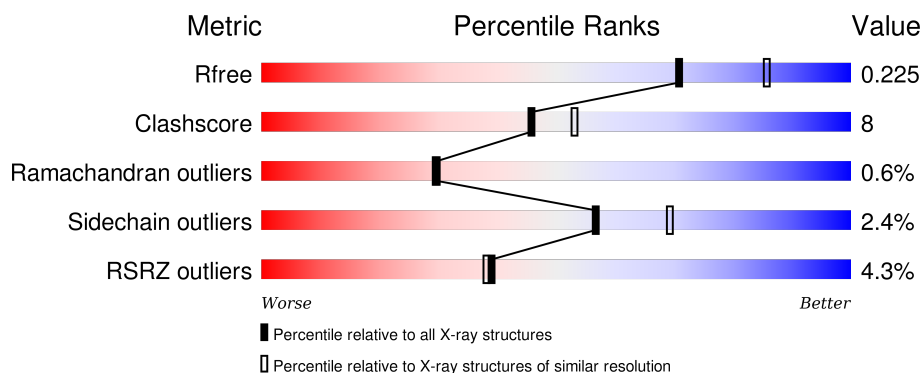
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*


The reported resolution of this entry is 2.20 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))
$R_{free}$	91344	3774 (2.20-2.20)
Clashscore	102246	4477 (2.20-2.20)
Ramachandran outliers	100387	4404 (2.20-2.20)
Sidechain outliers	100360	4405 (2.20-2.20)
RSRZ outliers	91569	3781 (2.20-2.20)

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	351	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3038 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 Putative capsid protein of prophage.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	347	Total	C	N	O	S	0	0	0
			2752	1749	460	527	16			

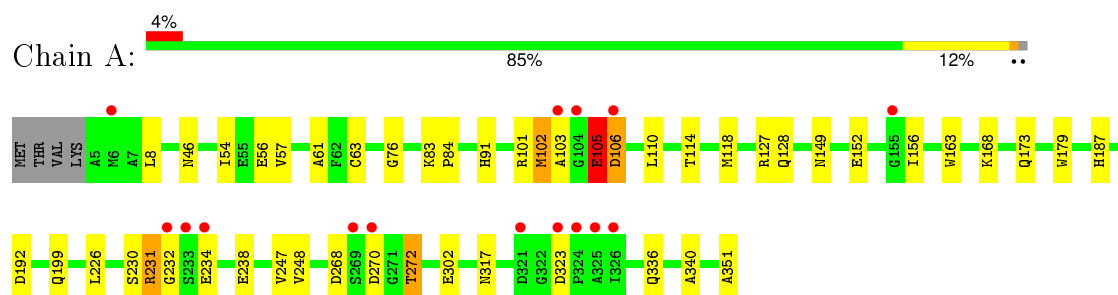
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	286	Total	O	0	0
			286	286		

### 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 errors 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: Putative capsid protein of prophage



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	113.62Å 113.62Å 58.86Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	98.53 – 2.20 37.75 – 2.20	Depositor EDS
% Data completeness (in resolution range)	99.4 (98.53-2.20) 99.4 (37.75-2.20)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.65 (at 2.20Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.175 , 0.225 0.178 , 0.225	Depositor DCC
$R_{free}$ test set	1124 reflections (5.37%)	DCC
Wilson B-factor (Å <sup>2</sup> )	29.3	Xtriage
Anisotropy	0.383	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 46.9	EDS
Estimated twinning fraction	0.046 for h,-h-k,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 22083 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3038	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.50% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.64	0/2811	0.70	1/3806 (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	3

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	A	127	ARG	NE-CZ-NH1	5.44	123.02	120.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	102	MET	Peptide
1	A	105	GLU	Peptide
1	A	323	ASP	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	2752	0	2713	42	1
2	A	286	0	0	5	1
All	All	3038	0	2713	42	1

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 8.

All (42) 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:105:GLU:HG3	1:A:106:ASP:H	1.09	1.07
1:A:102:MET:HB3	1:A:103:ALA:HB3	1.36	1.04
1:A:105:GLU:OE2	2:A:649:HOH:O	1.77	1.02
1:A:46:ASN:ND2	1:A:83:LYS:HD2	1.80	0.95
1:A:91:HIS:HD2	1:A:128:GLN:HE21	1.24	0.86
1:A:105:GLU:HG3	1:A:106:ASP:N	1.74	0.84
1:A:46:ASN:HD21	1:A:83:LYS:HD2	1.38	0.83
1:A:8:LEU:H	1:A:317:ASN:HD22	1.26	0.80
1:A:102:MET:CB	1:A:103:ALA:HB3	2.13	0.79
1:A:105:GLU:CG	1:A:106:ASP:N	2.44	0.79
1:A:46:ASN:HD21	1:A:83:LYS:CD	2.00	0.74
1:A:149:ASN:HD22	1:A:336:GLN:HE22	1.40	0.66
1:A:268:ASP:OD1	1:A:272:THR:HG23	1.96	0.65
1:A:57:VAL:HG22	2:A:552:HOH:O	2.02	0.60
1:A:54:ILE:HG23	1:A:56:GLU:HG3	1.84	0.59
1:A:268:ASP:OD1	1:A:272:THR:CG2	2.51	0.59
1:A:187:HIS:HE1	1:A:192:ASP:OD2	1.86	0.58
1:A:91:HIS:CD2	1:A:128:GLN:HE21	2.13	0.57
1:A:226:LEU:HD11	2:A:631:HOH:O	2.05	0.56
1:A:8:LEU:H	1:A:317:ASN:ND2	2.02	0.56
1:A:173:GLN:HE21	1:A:351:ALA:H	1.52	0.54
1:A:114:THR:O	1:A:118:MET:HG2	2.07	0.54
1:A:168:LYS:NZ	1:A:199:GLN:O	2.42	0.53
1:A:231:ARG:HB2	1:A:238:GLU:OE2	2.09	0.52
1:A:83:LYS:CG	1:A:156:ILE:HG21	2.39	0.52
1:A:149:ASN:ND2	1:A:336:GLN:HE22	2.06	0.51
1:A:83:LYS:HG2	1:A:156:ILE:HG21	1.94	0.49
1:A:173:GLN:NE2	1:A:179:TRP:H	2.13	0.46
1:A:272:THR:HG21	2:A:569:HOH:O	2.15	0.45
1:A:173:GLN:HE22	1:A:179:TRP:H	1.65	0.45
1:A:46:ASN:HD22	1:A:83:LYS:HD2	1.74	0.45
1:A:302:GLU:HG2	2:A:498:HOH:O	2.17	0.44

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:118:MET:HA	1:A:118:MET:HE3	2.00	0.44
1:A:61:ALA:HB2	1:A:163:TRP:CZ3	2.53	0.43
1:A:54:ILE:HG21	1:A:57:VAL:HG13	1.99	0.43
1:A:101:ARG:HB3	1:A:110:LEU:HD11	2.00	0.43
1:A:83:LYS:NZ	1:A:84:PRO:O	2.52	0.42
1:A:173:GLN:NE2	1:A:351:ALA:H	2.16	0.42
1:A:54:ILE:CG2	1:A:57:VAL:HG13	2.50	0.41
1:A:63:CYS:O	1:A:76:GLY:HA2	2.21	0.40
1:A:57:VAL:HA	1:A:340:ALA:HA	2.02	0.40
1:A:247:VAL:HG23	1:A:248:VAL:HG23	2.04	0.40

All (1) 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:103:ALA:CB	2:A:683:HOH:O[6_555]	1.96	0.24

## 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	345/351 (98%)	335 (97%)	8 (2%)	2 (1%)	30	29

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	105	GLU
1	A	232	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	295/299 (99%)	288 (98%)	7 (2%)	57 69

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	106	ASP
1	A	152	GLU
1	A	230	SER
1	A	231	ARG
1	A	234	GLU
1	A	270	ASP
1	A	272	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (10) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	46	ASN
1	A	91	HIS
1	A	130	ASN
1	A	149	ASN
1	A	169	ASN
1	A	173	GLN
1	A	187	HIS
1	A	288	ASN
1	A	317	ASN
1	A	332	HIS

### 5.3.3 RNA ⓘ

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	347/351 (98%)	-0.17	15 (4%) 39 38	19, 30, 56, 68	0

All (15) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	104	GLY	6.2
1	A	323	ASP	4.9
1	A	155	GLY	3.9
1	A	103	ALA	3.5
1	A	233	SER	3.3
1	A	106	ASP	3.1
1	A	324	PRO	3.0
1	A	270	ASP	2.9
1	A	269	SER	2.5
1	A	325	ALA	2.4
1	A	232	GLY	2.4
1	A	321	ASP	2.4
1	A	234	GLU	2.3
1	A	326	ILE	2.1
1	A	6	MET	2.1

### 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.