



# Full wwPDB X-ray Structure Validation Report ⓘ

Jun 14, 2020 – 10:47 pm BST

PDB ID : 3F9U  
Title : Crystal structure of C-terminal domain of putative exported cytochrome C biogenesis-related protein from *Bacteroides fragilis*  
Authors : Chang, C.; Tesar, C.; Cobb, G.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)  
Deposited on : 2008-11-14  
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

<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

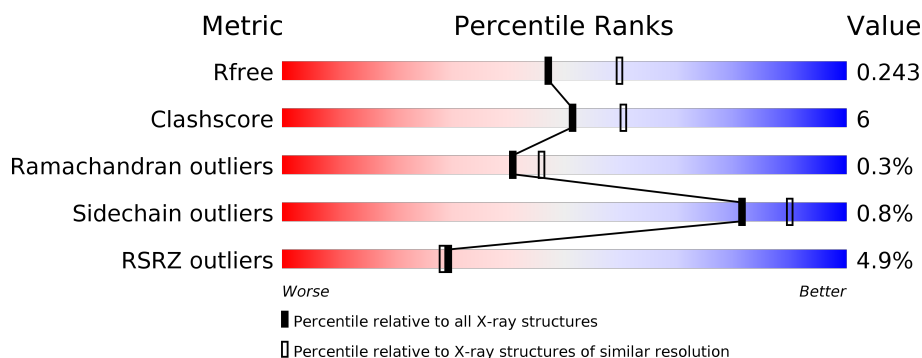
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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))
$R_{free}$	130704	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (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 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	172	<div> <div>3%</div> <div> <div></div> <div>75%</div> <div>8%</div> <div>16%</div> </div> </div>
1	B	172	<div> <div>5%</div> <div> <div></div> <div>77%</div> <div>8%</div> <div>14%</div> </div> </div>

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 2657 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.

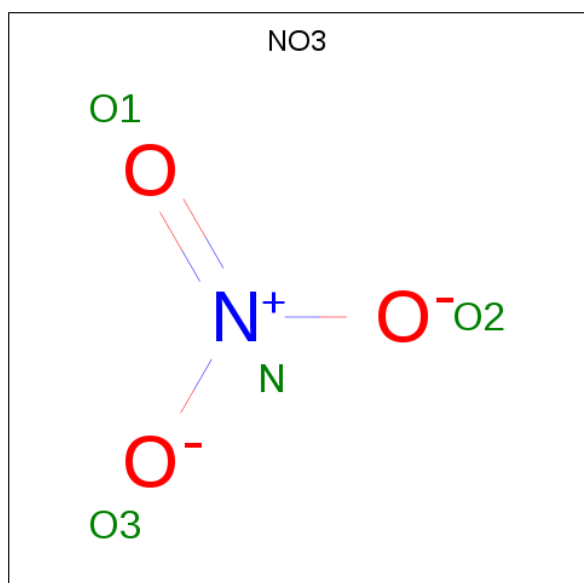
- Molecule 1 is a protein called Putative exported cytochrome C biogenesis-related protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	145	Total	C	N	O	S	Se	0	3	0
			1200	765	196	233	2	4			
1	B	148	Total	C	N	O	S	Se	0	7	0
			1240	789	200	245	2	4			

There are 6 discrepancies between the modelled and reference sequences:

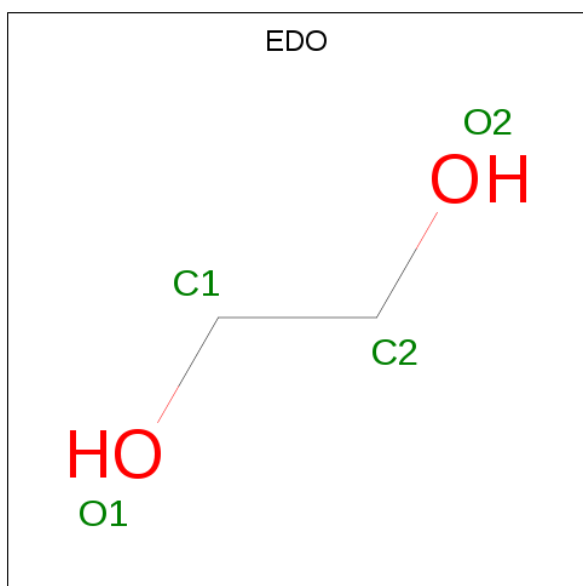
Chain	Residue	Modelled	Actual	Comment	Reference
A	498	SER	-	expression tag	UNP Q5LAZ4
A	499	ASN	-	expression tag	UNP Q5LAZ4
A	500	ALA	-	expression tag	UNP Q5LAZ4
B	498	SER	-	expression tag	UNP Q5LAZ4
B	499	ASN	-	expression tag	UNP Q5LAZ4
B	500	ALA	-	expression tag	UNP Q5LAZ4

- Molecule 2 is NITRATE ION (three-letter code: NO3) (formula: NO<sub>3</sub>).



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

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



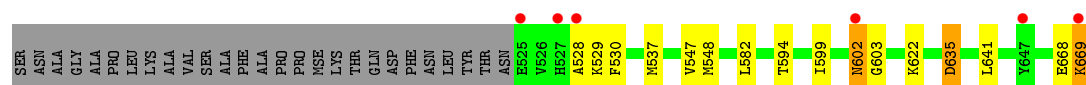
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0

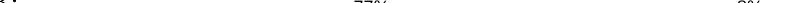
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	92	Total O 92 92	0	0
4	B	97	Total O 97 97	0	0

**i**

- Molecule 1: Putative exported cytochrome C biogenesis-related protein



- Chain B:  5% 77% 8% 14%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	64.16 Å 64.16 Å 181.56 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.20 40.58 – 2.20	Depositor EDS
% Data completeness (in resolution range)	99.7 (50.00-2.20) 99.7 (40.58-2.20)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.79 (at 2.20 Å)	Xtriage
Refinement program	REFMAC 5.5.0054	Depositor
R, $R_{free}$	0.181 , 0.236 0.186 , 0.243	Depositor DCC
$R_{free}$ test set	1022 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	34.2	Xtriage
Anisotropy	0.093	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 39.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	2657	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.12% 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: EDO, NO3

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	0.66	0/1229	0.69	1/1659 (0.1%)
1	B	0.68	0/1281	0.67	0/1730
All	All	0.67	0/2510	0.68	1/3389 (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	1

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	635	ASP	CB-CG-OD1	5.03	122.83	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	668	GLU	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	1200	0	1156	10	1
1	B	1240	0	1198	16	0
2	A	16	0	0	0	0
2	B	4	0	0	0	0
3	A	8	0	12	1	0
4	A	92	0	0	1	0
4	B	97	0	0	3	1
All	All	2657	0	2366	25	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 6.

All (25) 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:B:548:MSE:HE3	1:B:633:LEU:HD21	1.38	1.02
1:B:548:MSE:HE3	1:B:633:LEU:CD2	2.13	0.78
1:A:594:THR:HB	1:B:531[B]:ASP:OD2	1.85	0.76
1:B:528:ALA:HA	1:B:582:LEU:HD23	1.71	0.73
1:A:602:ASN:N	1:A:603:GLY:HA2	2.05	0.71
1:B:531[B]:ASP:OD1	4:B:50:HOH:O	2.13	0.66
1:B:637[B]:GLU:OE1	4:B:159:HOH:O	2.13	0.65
1:B:529:LYS:HB3	1:B:530:PHE:CD2	2.32	0.64
1:B:548:MSE:HE1	1:B:657:PHE:HE2	1.67	0.60
1:B:548:MSE:CE	1:B:633:LEU:HD21	2.24	0.59
1:A:529:LYS:HB3	1:A:530:PHE:CD2	2.38	0.58
1:B:548:MSE:CE	1:B:657:PHE:HE2	2.20	0.54
1:B:663[B]:GLU:HG3	4:B:165:HOH:O	2.07	0.53
1:B:548:MSE:HE2	1:B:631:TYR:HB3	1.91	0.52
1:A:548:MSE:HE2	1:A:582:LEU:HD13	1.94	0.50
1:B:548:MSE:CE	1:B:657:PHE:CE2	2.98	0.46
1:B:599:ILE:HD13	1:B:617:TYR:CE1	2.51	0.45
1:B:602:ASN:HA	1:B:603:GLY:HA2	1.43	0.44
1:A:528:ALA:HA	1:A:582:LEU:HD23	1.98	0.44
1:B:548:MSE:HE1	1:B:657:PHE:CE2	2.51	0.43
1:A:622:LYS:NZ	4:A:93:HOH:O	2.51	0.42
1:A:599:ILE:HG22	3:A:670:EDO:H22	2.01	0.41
1:A:635:ASP:HB3	1:A:641:LEU:HD21	2.03	0.41
1:A:537:MSE:SE	1:A:547:VAL:HG21	2.71	0.41
1:A:529:LYS:HB3	1:A:530:PHE:HD2	1.86	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:669:LYS:NZ	4:B:102:HOH:O[3_554]	2.07	0.13

## 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	146/172 (85%)	142 (97%)	4 (3%)	0	100	100
1	B	153/172 (89%)	147 (96%)	5 (3%)	1 (1%)	22	22
All	All	299/344 (87%)	289 (97%)	9 (3%)	1 (0%)	41	46

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	602	ASN

### 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	132/147 (90%)	130 (98%)	2 (2%)	65	78
1	B	139/147 (95%)	139 (100%)	0	100	100
All	All	271/294 (92%)	269 (99%)	2 (1%)	81	91

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	602	ASN
1	A	669	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	602	ASN
1	B	636	ASN

### 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](#)

7 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	$\# Z  > 2$
3	EDO	A	703	-	3,3,3	0.65	0	2,2,2	0.09	0
2	NO3	A	701	-	1,3,3	3.81	1 (100%)	0,3,3	0.00	-
3	EDO	A	670	-	3,3,3	0.44	0	2,2,2	0.53	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	NO3	A	704	-	1,3,3	3.57	1 (100%)	0,3,3	0.00	-
2	NO3	A	671	-	1,3,3	3.51	1 (100%)	0,3,3	0.00	-
2	NO3	A	702	-	1,3,3	3.77	1 (100%)	0,3,3	0.00	-
2	NO3	B	702	-	1,3,3	3.67	1 (100%)	0,3,3	0.00	-

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	EDO	A	703	-	-	1/1/1/1	-
3	EDO	A	670	-	-	1/1/1/1	-

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	701	NO3	O1-N	3.81	1.41	1.24
2	A	702	NO3	O1-N	3.77	1.41	1.24
2	B	702	NO3	O1-N	3.67	1.41	1.24
2	A	704	NO3	O1-N	3.57	1.40	1.24
2	A	671	NO3	O1-N	3.51	1.40	1.24

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	703	EDO	O1-C1-C2-O2
3	A	670	EDO	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	670	EDO	1	0

## 5.7 Other polymers

There are no such residues in this entry.

## 5.8 Polymer linkage issues

There are no chain breaks in this entry.

## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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	141/172 (81%)	-0.04	6 (4%) 35 33	16, 23, 38, 54	0
1	B	144/172 (83%)	0.09	8 (5%) 24 23	15, 24, 39, 71	0
All	All	285/344 (82%)	0.03	14 (4%) 29 28	15, 24, 39, 71	0

All (14) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	602	ASN	5.7
1	B	523	THR	5.0
1	A	669	LYS	4.4
1	B	522	TYR	4.4
1	B	605	GLU	3.6
1	B	524	ASN	3.5
1	B	528	ALA	3.4
1	B	604	THR	3.4
1	A	528	ALA	3.1
1	A	527	HIS	3.1
1	B	527	HIS	2.9
1	A	602	ASN	2.9
1	A	525	GLU	2.3
1	A	647	TYR	2.0

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates ⓘ

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	NO3	A	701	4/4	0.60	0.33	64,64,64,64	0
2	NO3	B	702	4/4	0.79	0.20	74,75,75,75	0
3	EDO	A	703	4/4	0.83	0.20	62,62,63,63	0
2	NO3	A	702	4/4	0.90	0.14	57,59,60,61	0
2	NO3	A	704	4/4	0.90	0.28	53,54,55,55	0
3	EDO	A	670	4/4	0.91	0.34	53,53,53,55	0
2	NO3	A	671	4/4	0.92	0.13	59,60,60,60	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.