



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 10:14 pm BST

PDB ID : 4QMK  
Title : Crystal structure of type III effector protein ExoU (exoU)  
Authors : Halavaty, A.S.; Tyson, G.H.; Zhang, A.; Hauser, A.R.; Anderson, W.F.; Center  
for Structural Genomics of Infectious Diseases (CSGID)  
Deposited on : 2014-06-16  
Resolution : 2.50 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 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

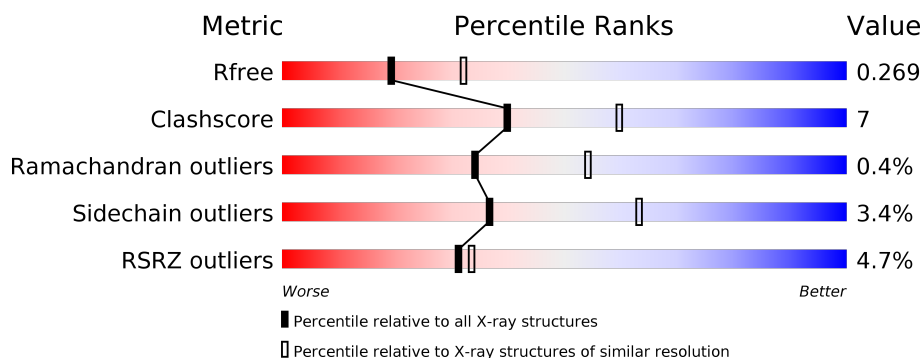
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	677	<div> <div>2%</div> <div> <div></div> <div>64%</div> <div>13%</div> <div>•</div> <div>22%</div> </div> </div>
1	B	677	<div> <div>6%</div> <div> <div></div> <div>64%</div> <div>10%</div> <div>•</div> <div>25%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 8162 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 Type III secretion system effector protein ExoU.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	528	Total	C	N	O	S	0	3	0
			4056	2542	741	762	11			
1	B	509	Total	C	N	O	S	0	2	0
			3895	2438	708	738	11			

There are 76 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-11	MET	-	EXPRESSION TAG	UNP I2BZ03
A	-10	GLY	-	EXPRESSION TAG	UNP I2BZ03
A	-9	LYS	-	EXPRESSION TAG	UNP I2BZ03
A	-8	ALA	-	EXPRESSION TAG	UNP I2BZ03
A	-7	SER	-	EXPRESSION TAG	UNP I2BZ03
A	-6	VAL	-	EXPRESSION TAG	UNP I2BZ03
A	-5	ASP	-	EXPRESSION TAG	UNP I2BZ03
A	-4	GLN	-	EXPRESSION TAG	UNP I2BZ03
A	-3	ILE	-	EXPRESSION TAG	UNP I2BZ03
A	-2	SER	-	EXPRESSION TAG	UNP I2BZ03
A	-1	LYS	-	EXPRESSION TAG	UNP I2BZ03
A	0	LEU	-	EXPRESSION TAG	UNP I2BZ03
A	640	GLY	-	EXPRESSION TAG	UNP I2BZ03
A	641	GLY	-	EXPRESSION TAG	UNP I2BZ03
A	642	ARG	-	EXPRESSION TAG	UNP I2BZ03
A	643	LEU	-	EXPRESSION TAG	UNP I2BZ03
A	644	VAL	-	EXPRESSION TAG	UNP I2BZ03
A	645	PRO	-	EXPRESSION TAG	UNP I2BZ03
A	646	ARG	-	EXPRESSION TAG	UNP I2BZ03
A	647	GLY	-	EXPRESSION TAG	UNP I2BZ03
A	648	SER	-	EXPRESSION TAG	UNP I2BZ03
A	649	PRO	-	EXPRESSION TAG	UNP I2BZ03
A	650	GLY	-	EXPRESSION TAG	UNP I2BZ03
A	651	ALA	-	EXPRESSION TAG	UNP I2BZ03
A	652	ALA	-	EXPRESSION TAG	UNP I2BZ03

*Continued on next page...*

*Continued from previous page...*

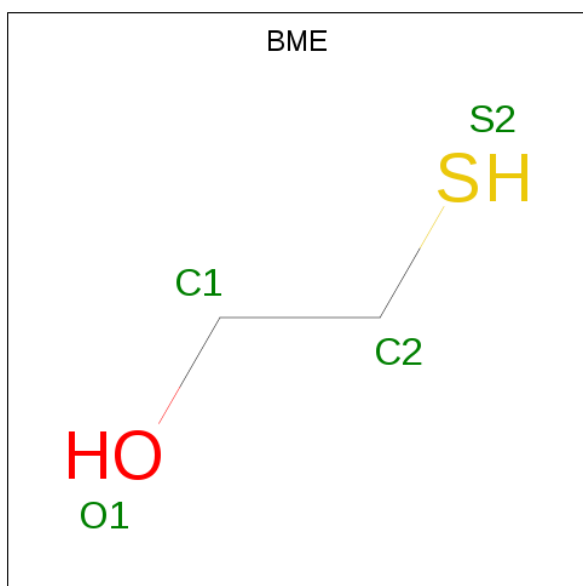
Chain	Residue	Modelled	Actual	Comment	Reference
A	653	GLY	-	EXPRESSION TAG	UNP I2BZ03
A	654	HIS	-	EXPRESSION TAG	UNP I2BZ03
A	655	ASN	-	EXPRESSION TAG	UNP I2BZ03
A	656	HIS	-	EXPRESSION TAG	UNP I2BZ03
A	657	ASN	-	EXPRESSION TAG	UNP I2BZ03
A	658	HIS	-	EXPRESSION TAG	UNP I2BZ03
A	659	ASN	-	EXPRESSION TAG	UNP I2BZ03
A	660	HIS	-	EXPRESSION TAG	UNP I2BZ03
A	661	ASN	-	EXPRESSION TAG	UNP I2BZ03
A	662	HIS	-	EXPRESSION TAG	UNP I2BZ03
A	663	ASN	-	EXPRESSION TAG	UNP I2BZ03
A	664	HIS	-	EXPRESSION TAG	UNP I2BZ03
A	665	ASN	-	EXPRESSION TAG	UNP I2BZ03
B	-11	MET	-	EXPRESSION TAG	UNP I2BZ03
B	-10	GLY	-	EXPRESSION TAG	UNP I2BZ03
B	-9	LYS	-	EXPRESSION TAG	UNP I2BZ03
B	-8	ALA	-	EXPRESSION TAG	UNP I2BZ03
B	-7	SER	-	EXPRESSION TAG	UNP I2BZ03
B	-6	VAL	-	EXPRESSION TAG	UNP I2BZ03
B	-5	ASP	-	EXPRESSION TAG	UNP I2BZ03
B	-4	GLN	-	EXPRESSION TAG	UNP I2BZ03
B	-3	ILE	-	EXPRESSION TAG	UNP I2BZ03
B	-2	SER	-	EXPRESSION TAG	UNP I2BZ03
B	-1	LYS	-	EXPRESSION TAG	UNP I2BZ03
B	0	LEU	-	EXPRESSION TAG	UNP I2BZ03
B	640	GLY	-	EXPRESSION TAG	UNP I2BZ03
B	641	GLY	-	EXPRESSION TAG	UNP I2BZ03
B	642	ARG	-	EXPRESSION TAG	UNP I2BZ03
B	643	LEU	-	EXPRESSION TAG	UNP I2BZ03
B	644	VAL	-	EXPRESSION TAG	UNP I2BZ03
B	645	PRO	-	EXPRESSION TAG	UNP I2BZ03
B	646	ARG	-	EXPRESSION TAG	UNP I2BZ03
B	647	GLY	-	EXPRESSION TAG	UNP I2BZ03
B	648	SER	-	EXPRESSION TAG	UNP I2BZ03
B	649	PRO	-	EXPRESSION TAG	UNP I2BZ03
B	650	GLY	-	EXPRESSION TAG	UNP I2BZ03
B	651	ALA	-	EXPRESSION TAG	UNP I2BZ03
B	652	ALA	-	EXPRESSION TAG	UNP I2BZ03
B	653	GLY	-	EXPRESSION TAG	UNP I2BZ03
B	654	HIS	-	EXPRESSION TAG	UNP I2BZ03
B	655	ASN	-	EXPRESSION TAG	UNP I2BZ03
B	656	HIS	-	EXPRESSION TAG	UNP I2BZ03

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	657	ASN	-	EXPRESSION TAG	UNP I2BZ03
B	658	HIS	-	EXPRESSION TAG	UNP I2BZ03
B	659	ASN	-	EXPRESSION TAG	UNP I2BZ03
B	660	HIS	-	EXPRESSION TAG	UNP I2BZ03
B	661	ASN	-	EXPRESSION TAG	UNP I2BZ03
B	662	HIS	-	EXPRESSION TAG	UNP I2BZ03
B	663	ASN	-	EXPRESSION TAG	UNP I2BZ03
B	664	HIS	-	EXPRESSION TAG	UNP I2BZ03
B	665	ASN	-	EXPRESSION TAG	UNP I2BZ03

- Molecule 2 is BETA-MERCAPTOETHANOL (three-letter code: BME) (formula:  $C_2H_6OS$ ).

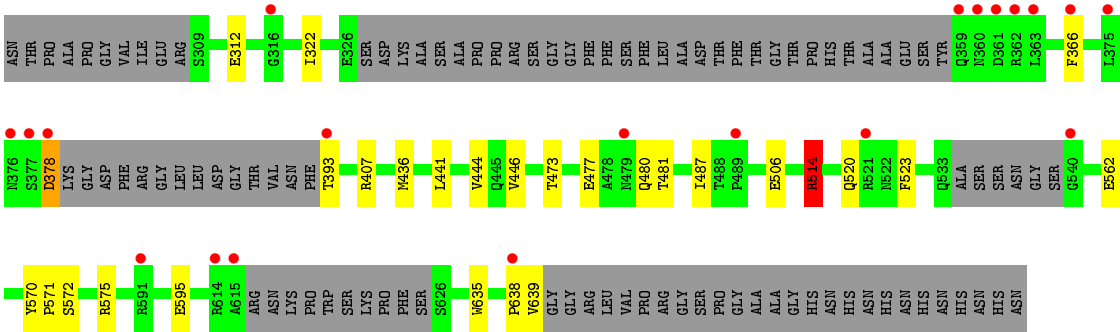


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	O	S	0	0
			4	2	1	1		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	136	Total	O	0	1
			137	137		
3	B	70	Total	O	0	0
			70	70		





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	67.22Å 115.34Å 88.44Å 90.00° 102.77° 90.00°	Depositor
Resolution (Å)	29.86 – 2.50 29.86 – 2.50	Depositor EDS
% Data completeness (in resolution range)	94.8 (29.86-2.50) 94.9 (29.86-2.50)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.45 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, $R_{free}$	0.206 , 0.270 0.208 , 0.269	Depositor DCC
$R_{free}$ test set	2190 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	45.9	Xtriage
Anisotropy	0.606	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 49.5	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	0.94	EDS
Total number of atoms	8162	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.64% 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: BME

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	0.77	0/4120	0.85	4/5576 (0.1%)
1	B	0.62	0/3950	0.80	2/5346 (0.0%)
All	All	0.70	0/8070	0.82	6/10922 (0.1%)

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	514	ARG	NE-CZ-NH1	-6.19	117.20	120.30
1	A	106	ASP	CB-CG-OD1	5.91	123.62	118.30
1	B	514	ARG	NE-CZ-NH2	5.77	123.19	120.30
1	A	245	LEU	CA-CB-CG	5.57	128.12	115.30
1	A	504	ARG	NE-CZ-NH1	5.32	122.96	120.30

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	4056	0	4132	62	0
1	B	3895	0	3963	59	0
2	A	4	0	5	0	0
3	A	137	0	0	5	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	70	0	0	7	0
All	All	8162	0	8100	117	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 7.

The worst 5 of 117 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:251:ASN:HB3	3:B:722:HOH:O	1.53	1.05
1:A:160:THR:HG21	1:B:160:THR:HG21	1.56	0.87
1:B:67:ILE:O	1:B:70:PRO:HD2	1.78	0.83
1:A:64:ALA:HB1	1:A:385:LEU:HD22	1.60	0.82
1:B:570[A]:TYR:HB3	1:B:571:PRO:HD3	1.69	0.75

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	517/677 (76%)	497 (96%)	16 (3%)	4 (1%)	19	35
1	B	493/677 (73%)	477 (97%)	16 (3%)	0	100	100
All	All	1010/1354 (75%)	974 (96%)	32 (3%)	4 (0%)	34	54

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	194	ALA
1	A	385	LEU
1	A	622	LYS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	196	ARG

### 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	433/548 (79%)	419 (97%)	14 (3%)	39	65
1	B	416/548 (76%)	401 (96%)	15 (4%)	35	61
All	All	849/1096 (78%)	820 (97%)	29 (3%)	37	63

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	595	GLU
1	B	139	GLU
1	B	562	GLU
1	B	78	GLU
1	B	177	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	413	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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](#)

1 ligand is 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	BME	A	700	1	3,3,3	0.63	0	1,2,2	0.05	0

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
2	BME	A	700	1	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	700	BME	O1-C1-C2-S2

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 ⓘ

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	528/677 (77%)	-0.16	11 (2%) 63 66	20, 43, 85, 133	0
1	B	509/677 (75%)	0.23	38 (7%) 14 14	29, 63, 106, 151	0
All	All	1037/1354 (76%)	0.03	49 (4%) 31 33	20, 52, 101, 151	0

The worst 5 of 49 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	540	GLY	6.6
1	B	540	GLY	6.4
1	B	615	ALA	5.2
1	B	393	THR	5.2
1	B	359	GLN	4.6

### 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( $\text{\AA}^2$ )	Q<0.9
2	BME	A	700	4/4	0.89	0.19	77,80,81,84	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.