



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 03:52 PM GMT

PDB ID : 4DQ9  
Title : Crystal structure of the minor pseudopilin EPSH from the type II secretion system of *Vibrio cholerae*  
Authors : Raghunathan, K.; Vago, F.S.; Grindem, D.; Ball, T.; Wedemeyer, W.J.; Arvidson, D.N.  
Deposited on : 2012-02-15  
Resolution : 1.59 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

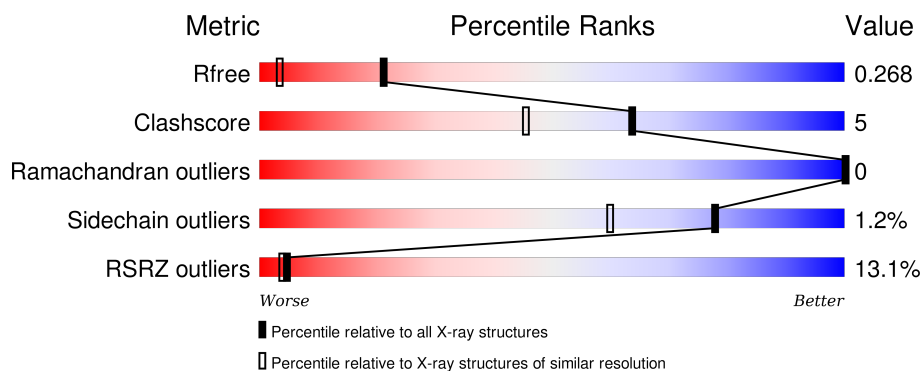
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.59 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	2475 (1.60-1.60)
Clashscore	102246	2732 (1.60-1.60)
Ramachandran outliers	100387	2654 (1.60-1.60)
Sidechain outliers	100360	2653 (1.60-1.60)
RSRZ outliers	91569	2479 (1.60-1.60)

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	170	<div> <div>11%</div> <div> <div></div> <div>75%</div> <div>11%</div> <div>•</div> <div>12%</div> </div> </div>
1	B	170	<div> <div>12%</div> <div> <div></div> <div>78%</div> <div>5%</div> <div>•</div> <div>16%</div> </div> </div>

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 2618 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 General secretion pathway protein H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	149	Total	C	N	O	S	0	6	0
			1225	773	210	240	2			
1	B	142	Total	C	N	O	S	0	2	0
			1136	717	192	225	2			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	189	LEU	-	EXPRESSION TAG	UNP P45774
A	190	GLU	-	EXPRESSION TAG	UNP P45774
A	191	HIS	-	EXPRESSION TAG	UNP P45774
A	192	HIS	-	EXPRESSION TAG	UNP P45774
A	193	HIS	-	EXPRESSION TAG	UNP P45774
A	194	HIS	-	EXPRESSION TAG	UNP P45774
A	195	HIS	-	EXPRESSION TAG	UNP P45774
A	196	HIS	-	EXPRESSION TAG	UNP P45774
B	189	LEU	-	EXPRESSION TAG	UNP P45774
B	190	GLU	-	EXPRESSION TAG	UNP P45774
B	191	HIS	-	EXPRESSION TAG	UNP P45774
B	192	HIS	-	EXPRESSION TAG	UNP P45774
B	193	HIS	-	EXPRESSION TAG	UNP P45774
B	194	HIS	-	EXPRESSION TAG	UNP P45774
B	195	HIS	-	EXPRESSION TAG	UNP P45774
B	196	HIS	-	EXPRESSION TAG	UNP P45774

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	2	Total	Na	0	0
			2	2		

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total 1	Cl 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	149	Total 149	O 149	0	0
4	B	105	Total 105	O 105	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	53.39Å 71.11Å 84.65Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	27.94 – 1.59 27.94 – 1.59	Depositor EDS
% Data completeness (in resolution range)	96.3 (27.94-1.59) 96.3 (27.94-1.59)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.10 (at 1.59Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.201 , 0.249 0.225 , 0.268	Depositor DCC
$R_{free}$ test set	2125 reflections (5.05%)	DCC
Wilson B-factor (Å <sup>2</sup> )	23.9	Xtriage
Anisotropy	0.091	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 33.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	4 of 42102 reflections (0.010%)	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	2618	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.23% 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

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

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	1.24	2/1251 (0.2%)	1.17	6/1691 (0.4%)
1	B	1.10	0/1159	0.98	1/1566 (0.1%)
All	All	1.17	2/2410 (0.1%)	1.09	7/3257 (0.2%)

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

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	155	SER	CB-OG	-7.58	1.32	1.42
1	A	49	GLU	CD-OE2	-5.54	1.19	1.25

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	76	ASP	CB-CG-OD2	-6.04	112.87	118.30
1	A	60	ARG	NE-CZ-NH2	-6.02	117.29	120.30
1	A	56	ASP	CB-CG-OD1	5.76	123.48	118.30
1	A	164	ASP	CB-CG-OD1	-5.63	113.23	118.30
1	A	179	LEU	CB-CG-CD2	-5.47	101.69	111.00
1	A	152	PHE	CB-CG-CD1	5.36	124.56	120.80
1	B	56	ASP	CB-CG-OD1	5.32	123.09	118.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	171	ALA	Mainchain,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	1225	0	1186	19	0
1	B	1136	0	1093	6	0
2	A	2	0	0	0	0
3	B	1	0	0	0	0
4	A	149	0	0	1	0
4	B	105	0	0	0	0
All	All	2618	0	2279	24	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 5.

All (24) 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:174[B]:ASN:OD1	1:A:176[B]:THR:HG23	1.91	0.69
1:A:172[B]:GLN:CG	1:A:176[B]:THR:OG1	2.45	0.64
1:A:173[B]:GLU:C	1:A:175[B]:GLY:N	2.51	0.60
1:A:174[B]:ASN:CG	1:A:176[B]:THR:HG23	2.24	0.58
1:A:174[B]:ASN:OD1	1:A:176[B]:THR:CG2	2.54	0.54
1:A:173[B]:GLU:C	1:A:175[B]:GLY:H	2.11	0.53
1:A:174[B]:ASN:ND2	1:A:176[B]:THR:HG23	2.25	0.51
1:A:94:LYS:HB2	1:A:94:LYS:NZ	2.24	0.51
1:A:172[B]:GLN:HG2	1:A:176[B]:THR:OG1	2.12	0.49
1:A:175[A]:GLY:O	1:A:177:LEU:HD22	2.12	0.48
1:A:56:ASP:OD2	1:B:56:ASP:OD2	2.33	0.47
1:B:179:LEU:O	1:B:180:LEU:HD23	2.15	0.46
1:B:40:TYR:CZ	1:B:44:LEU:HD11	2.51	0.45
1:B:99:LEU:HD12	1:B:99:LEU:C	2.37	0.45
1:A:178[B]:ARG:HH11	1:A:178[B]:ARG:HD3	1.63	0.44

*Continued on next page...*



*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:176[A]:THR:C	1:A:177:LEU:HD13	2.37	0.44
1:A:45:LEU:HD23	1:A:45:LEU:C	2.39	0.43
1:B:167:TRP:CE2	1:B:182:PRO:HD3	2.55	0.42
1:A:173[B]:GLU:O	1:A:174[B]:ASN:C	2.57	0.42
1:A:29:LYS:HG3	4:A:386:HOH:O	2.20	0.42
1:A:44:LEU:HD23	1:A:177:LEU:HD23	2.02	0.41
1:A:172[B]:GLN:HG3	1:A:174[B]:ASN:ND2	2.35	0.41
1:A:37:GLN:O	1:A:41:GLN:HG3	2.21	0.41
1:B:101:PHE:HA	1:B:153:THR:O	2.20	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	151/170 (89%)	150 (99%)	1 (1%)	0	100	100
1	B	138/170 (81%)	137 (99%)	1 (1%)	0	100	100
All	All	289/340 (85%)	287 (99%)	2 (1%)	0	100	100

There are no Ramachandran outliers to report.

### 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	131/149 (88%)	127 (97%)	4 (3%)	47	19
1	B	121/149 (81%)	121 (100%)	0	100	100
All	All	252/298 (85%)	248 (98%)	4 (2%)	78	47

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	94	LYS
1	A	172[A]	GLN
1	A	172[B]	GLN
1	A	177	LEU

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

Mol	Chain	Res	Type
1	B	172	GLN

### 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 ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

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

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	149/170 (87%)	0.66	18 (12%) 6 5	15, 22, 38, 63	0
1	B	142/170 (83%)	0.75	20 (14%) 4 3	17, 28, 52, 86	0
All	All	291/340 (85%)	0.70	38 (13%) 5 4	15, 25, 52, 86	0

All (38) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	126	PHE	6.8
1	A	125	MET	5.5
1	B	185	SER	4.8
1	A	135	GLN	4.6
1	A	124	GLU	4.4
1	B	173	GLU	4.3
1	B	107	ALA	4.3
1	A	175[A]	GLY	3.9
1	A	172[A]	GLN	3.7
1	A	174[A]	ASN	3.6
1	B	122	ASP	3.4
1	A	43	LEU	3.2
1	A	173[A]	GLU	3.2
1	A	176[A]	THR	3.1
1	B	184	GLU	3.0
1	B	174	ASN	3.0
1	B	30	ASP	2.9
1	B	95	GLU	2.9
1	A	44	LEU	2.8
1	A	114	LEU	2.7
1	A	122	ASP	2.7
1	B	172	GLN	2.7
1	B	186	ASP	2.6
1	B	176	THR	2.6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	27	SER	2.5
1	B	59	VAL	2.5
1	A	179	LEU	2.5
1	A	183	GLY	2.4
1	B	105	GLY	2.4
1	A	121	PHE	2.3
1	B	46	LEU	2.3
1	A	123	GLU	2.3
1	B	44	LEU	2.2
1	B	141	LEU	2.1
1	B	127	ALA	2.0
1	A	39	PHE	2.0
1	B	165	GLU	2.0
1	B	64[A]	ASP	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 ⓘ

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	NA	A	201	1/1	0.93	0.08	-1.71	25,25,25,25	0
3	CL	B	201	1/1	0.97	0.05	-2.88	28,28,28,28	0
2	NA	A	202	1/1	0.96	0.06	-	28,28,28,28	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.