



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 7, 2017 – 03:07 AM EST

PDB ID : 4KSS  
Title : Crystal Structure of Vibrio cholerae ATPase GspsE Hexamer  
Authors : Hol, W.G.; Turley, S.; Lu, C.Y.; Park, Y.J.  
Deposited on : unknown  
Resolution : 7.58 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

<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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	rb-20030345

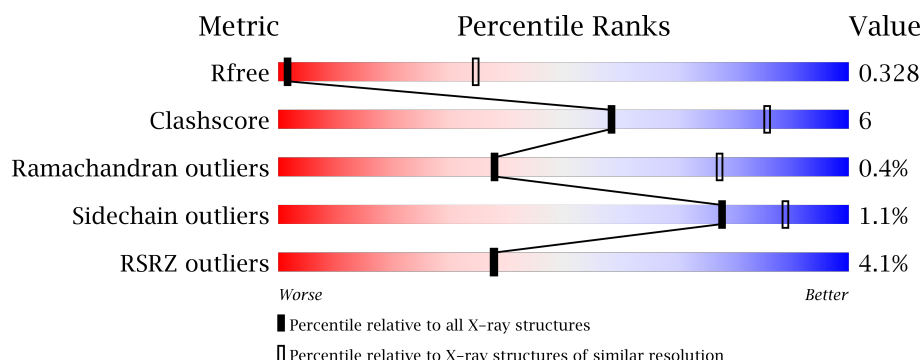
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 7.58 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))
$R_{free}$	100719	1100 (10.00-3.70)
Clashscore	112137	1035 (10.00-3.80)
Ramachandran outliers	110173	1003 (10.00-3.76)
Sidechain outliers	110143	1098 (10.00-3.70)
RSRZ outliers	101464	1003 (10.00-3.72)

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	581	<div> <div>4%</div> <div>83% 7% • 10%</div> </div>
1	B	581	<div> <div>3%</div> <div>83% 6% • 10%</div> </div>
1	C	581	<div> <div>4%</div> <div>83% 6% • 10%</div> </div>
1	D	581	<div> <div>5%</div> <div>83% 6% • 10%</div> </div>
1	E	581	<div> <div>4%</div> <div>84% 5% • 10%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	581	 A horizontal bar chart showing the quality of chain F. The bar is divided into four segments: a small red segment at the beginning labeled '3%', a large green segment labeled '83%', a small yellow segment labeled '6%', and a grey segment at the end labeled '10%'. The segments are separated by small dots.

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 24126 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 II secretion system protein E, hemolysin-coregulated protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	521	Total	C	N	O	S	0	0	0
			4021	2508	724	767	22			
1	B	521	Total	C	N	O	S	0	0	0
			4021	2508	724	767	22			
1	C	521	Total	C	N	O	S	0	0	0
			4021	2508	724	767	22			
1	D	521	Total	C	N	O	S	0	0	0
			4021	2508	724	767	22			
1	E	521	Total	C	N	O	S	0	0	0
			4021	2508	724	767	22			
1	F	521	Total	C	N	O	S	0	0	0
			4021	2508	724	767	22			

There are 96 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	99	MET	-	EXPRESSION TAG	UNP P37093
A	504	GLY	-	LINKER	UNP P37093
A	505	SER	-	LINKER	UNP P37093
A	506	GLY	-	LINKER	UNP P37093
A	507	SER	-	LINKER	UNP P37093
A	508	GLY	-	LINKER	UNP P37093
A	509	SER	-	LINKER	UNP P37093
A	510	GLY	-	LINKER	UNP P37093
A	672	LEU	-	EXPRESSION TAG	UNP Q02UZ4
A	673	GLU	-	EXPRESSION TAG	UNP Q02UZ4
A	674	HIS	-	EXPRESSION TAG	UNP Q02UZ4
A	675	HIS	-	EXPRESSION TAG	UNP Q02UZ4
A	676	HIS	-	EXPRESSION TAG	UNP Q02UZ4
A	677	HIS	-	EXPRESSION TAG	UNP Q02UZ4
A	678	HIS	-	EXPRESSION TAG	UNP Q02UZ4
A	679	HIS	-	EXPRESSION TAG	UNP Q02UZ4

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	99	MET	-	EXPRESSION TAG	UNP P37093
B	504	GLY	-	LINKER	UNP P37093
B	505	SER	-	LINKER	UNP P37093
B	506	GLY	-	LINKER	UNP P37093
B	507	SER	-	LINKER	UNP P37093
B	508	GLY	-	LINKER	UNP P37093
B	509	SER	-	LINKER	UNP P37093
B	510	GLY	-	LINKER	UNP P37093
B	672	LEU	-	EXPRESSION TAG	UNP Q02UZ4
B	673	GLU	-	EXPRESSION TAG	UNP Q02UZ4
B	674	HIS	-	EXPRESSION TAG	UNP Q02UZ4
B	675	HIS	-	EXPRESSION TAG	UNP Q02UZ4
B	676	HIS	-	EXPRESSION TAG	UNP Q02UZ4
B	677	HIS	-	EXPRESSION TAG	UNP Q02UZ4
B	678	HIS	-	EXPRESSION TAG	UNP Q02UZ4
B	679	HIS	-	EXPRESSION TAG	UNP Q02UZ4
C	99	MET	-	EXPRESSION TAG	UNP P37093
C	504	GLY	-	LINKER	UNP P37093
C	505	SER	-	LINKER	UNP P37093
C	506	GLY	-	LINKER	UNP P37093
C	507	SER	-	LINKER	UNP P37093
C	508	GLY	-	LINKER	UNP P37093
C	509	SER	-	LINKER	UNP P37093
C	510	GLY	-	LINKER	UNP P37093
C	672	LEU	-	EXPRESSION TAG	UNP Q02UZ4
C	673	GLU	-	EXPRESSION TAG	UNP Q02UZ4
C	674	HIS	-	EXPRESSION TAG	UNP Q02UZ4
C	675	HIS	-	EXPRESSION TAG	UNP Q02UZ4
C	676	HIS	-	EXPRESSION TAG	UNP Q02UZ4
C	677	HIS	-	EXPRESSION TAG	UNP Q02UZ4
C	678	HIS	-	EXPRESSION TAG	UNP Q02UZ4
C	679	HIS	-	EXPRESSION TAG	UNP Q02UZ4
D	99	MET	-	EXPRESSION TAG	UNP P37093
D	504	GLY	-	LINKER	UNP P37093
D	505	SER	-	LINKER	UNP P37093
D	506	GLY	-	LINKER	UNP P37093
D	507	SER	-	LINKER	UNP P37093
D	508	GLY	-	LINKER	UNP P37093
D	509	SER	-	LINKER	UNP P37093
D	510	GLY	-	LINKER	UNP P37093
D	672	LEU	-	EXPRESSION TAG	UNP Q02UZ4
D	673	GLU	-	EXPRESSION TAG	UNP Q02UZ4

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	674	HIS	-	EXPRESSION TAG	UNP Q02UZ4
D	675	HIS	-	EXPRESSION TAG	UNP Q02UZ4
D	676	HIS	-	EXPRESSION TAG	UNP Q02UZ4
D	677	HIS	-	EXPRESSION TAG	UNP Q02UZ4
D	678	HIS	-	EXPRESSION TAG	UNP Q02UZ4
D	679	HIS	-	EXPRESSION TAG	UNP Q02UZ4
E	99	MET	-	EXPRESSION TAG	UNP P37093
E	504	GLY	-	LINKER	UNP P37093
E	505	SER	-	LINKER	UNP P37093
E	506	GLY	-	LINKER	UNP P37093
E	507	SER	-	LINKER	UNP P37093
E	508	GLY	-	LINKER	UNP P37093
E	509	SER	-	LINKER	UNP P37093
E	510	GLY	-	LINKER	UNP P37093
E	672	LEU	-	EXPRESSION TAG	UNP Q02UZ4
E	673	GLU	-	EXPRESSION TAG	UNP Q02UZ4
E	674	HIS	-	EXPRESSION TAG	UNP Q02UZ4
E	675	HIS	-	EXPRESSION TAG	UNP Q02UZ4
E	676	HIS	-	EXPRESSION TAG	UNP Q02UZ4
E	677	HIS	-	EXPRESSION TAG	UNP Q02UZ4
E	678	HIS	-	EXPRESSION TAG	UNP Q02UZ4
E	679	HIS	-	EXPRESSION TAG	UNP Q02UZ4
F	99	MET	-	EXPRESSION TAG	UNP P37093
F	504	GLY	-	LINKER	UNP P37093
F	505	SER	-	LINKER	UNP P37093
F	506	GLY	-	LINKER	UNP P37093
F	507	SER	-	LINKER	UNP P37093
F	508	GLY	-	LINKER	UNP P37093
F	509	SER	-	LINKER	UNP P37093
F	510	GLY	-	LINKER	UNP P37093
F	672	LEU	-	EXPRESSION TAG	UNP Q02UZ4
F	673	GLU	-	EXPRESSION TAG	UNP Q02UZ4
F	674	HIS	-	EXPRESSION TAG	UNP Q02UZ4
F	675	HIS	-	EXPRESSION TAG	UNP Q02UZ4
F	676	HIS	-	EXPRESSION TAG	UNP Q02UZ4
F	677	HIS	-	EXPRESSION TAG	UNP Q02UZ4
F	678	HIS	-	EXPRESSION TAG	UNP Q02UZ4
F	679	HIS	-	EXPRESSION TAG	UNP Q02UZ4







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 4 <sub>2</sub> 2 <sub>1</sub> 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	205.06 Å   205.06 Å   234.98 Å 90.00°   90.00°   90.00°	Depositor
Resolution (Å)	50.00 – 7.58 49.95 – 7.58	Depositor EDS
% Data completeness (in resolution range)	99.0 (50.00-7.58) 99.6 (49.95-7.58)	Depositor EDS
$R_{merge}$	0.24	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.22 (at 7.37 Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.349   ,   0.360 0.328   ,   0.328	Depositor DCC
$R_{free}$ test set	308 reflections (4.92%)	DCC
Wilson B-factor (Å <sup>2</sup> )	281.2	Xtriage
Anisotropy	0.061	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 108.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.71	EDS
Total number of atoms	24126	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	166.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.06% 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](#)

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	0.26	0/4074	0.47	2/5491 (0.0%)
1	B	0.26	0/4074	0.47	2/5491 (0.0%)
1	C	0.26	0/4074	0.47	2/5491 (0.0%)
1	D	0.26	0/4074	0.47	2/5491 (0.0%)
1	E	0.26	0/4074	0.47	2/5491 (0.0%)
1	F	0.26	0/4074	0.47	2/5491 (0.0%)
All	All	0.26	0/24444	0.47	12/32946 (0.0%)

There are no bond length outliers.

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	430	CYS	CA-CB-SG	7.34	127.22	114.00
1	F	430	CYS	CA-CB-SG	7.33	127.19	114.00
1	C	430	CYS	CA-CB-SG	7.32	127.17	114.00
1	B	430	CYS	CA-CB-SG	7.31	127.16	114.00
1	E	430	CYS	CA-CB-SG	7.31	127.15	114.00

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	4021	0	4099	70	0
1	B	4021	0	4099	59	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	4021	0	4097	76	0
1	D	4021	0	4098	80	0
1	E	4021	0	4099	55	0
1	F	4021	0	4099	70	0
All	All	24126	0	24591	269	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 6.

The worst 5 of 269 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:C:464:GLU:CG	1:D:378:PHE:HE1	1.07	1.66
1:C:464:GLU:HG3	1:D:378:PHE:CE1	0.97	1.50
1:C:463:GLY:CA	1:D:459:HIS:CE1	1.92	1.50
1:C:463:GLY:HA2	1:D:459:HIS:CE1	1.47	1.47
1:E:464:GLU:HG3	1:F:378:PHE:CE1	1.47	1.45

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	509/581 (88%)	490 (96%)	17 (3%)	2 (0%)	38	77
1	B	509/581 (88%)	488 (96%)	19 (4%)	2 (0%)	38	77
1	C	509/581 (88%)	490 (96%)	17 (3%)	2 (0%)	38	77
1	D	509/581 (88%)	490 (96%)	17 (3%)	2 (0%)	38	77
1	E	509/581 (88%)	490 (96%)	17 (3%)	2 (0%)	38	77
1	F	509/581 (88%)	484 (95%)	23 (4%)	2 (0%)	38	77

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	3054/3486 (88%)	2932 (96%)	110 (4%)	12 (0%)	38 77

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	282	ASN
1	B	282	ASN
1	C	282	ASN
1	D	282	ASN
1	E	282	ASN

### 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	444/491 (90%)	439 (99%)	5 (1%)	78 89
1	B	444/491 (90%)	438 (99%)	6 (1%)	71 86
1	C	444/491 (90%)	439 (99%)	5 (1%)	78 89
1	D	444/491 (90%)	440 (99%)	4 (1%)	82 91
1	E	444/491 (90%)	440 (99%)	4 (1%)	82 91
1	F	444/491 (90%)	440 (99%)	4 (1%)	82 91
All	All	2664/2946 (90%)	2636 (99%)	28 (1%)	78 89

5 of 28 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	169	ARG
1	C	456	GLU
1	F	195	ASP
1	C	195	ASP
1	C	313	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 39 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	325	GLN
1	D	465	GLN
1	F	465	GLN
1	D	346	GLN
1	D	390	GLN

### 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](#)

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	521/581 (89%)	0.43	25 (4%) 31 34	119, 174, 258, 384	0
1	B	521/581 (89%)	0.32	17 (3%) 47 46	75, 139, 253, 363	0
1	C	521/581 (89%)	0.48	22 (4%) 37 37	88, 170, 268, 376	0
1	D	521/581 (89%)	0.43	28 (5%) 26 30	69, 160, 262, 407	0
1	E	521/581 (89%)	0.38	21 (4%) 39 38	69, 153, 250, 404	0
1	F	521/581 (89%)	0.36	15 (2%) 52 50	89, 155, 247, 400	0
All	All	3126/3486 (89%)	0.40	128 (4%) 38 38	69, 161, 259, 407	0

The worst 5 of 128 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	605	GLU	4.8
1	D	446	GLU	4.7
1	C	625	GLY	4.6
1	A	557	ALA	4.3
1	F	268	SER	3.9

### 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.