



wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 12:42 am BST

PDB ID : 4KSR
Title : Crystal Structure of the Vibrio cholerae ATPase GspE Hexamer
Authors : Hol, W.G.; Turley, S.; Lu, C.Y.; Park, Y.J.; Marionni, S.T.; Lee, K.; Patrick, M.; Sandkvist, M.; Bush, M.; Shah, R.
Deposited on : 2013-05-17
Resolution : 4.20 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

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
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

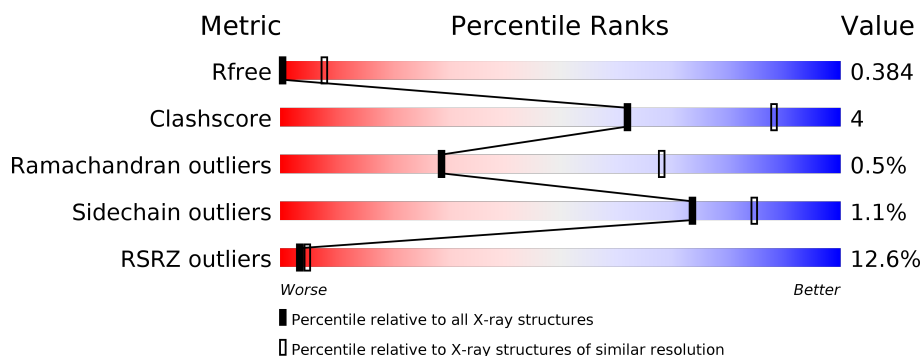
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 4.20 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	1005 (4.62-3.78)
Clashscore	141614	1044 (4.60-3.80)
Ramachandran outliers	138981	1000 (4.60-3.80)
Sidechain outliers	138945	1007 (4.62-3.78)
RSRZ outliers	127900	1063 (4.70-3.70)

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 ≥ 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	583	 13% 82% 7% • 11%
1	B	583	 10% 82% 7% • 11%
1	C	583	 10% 83% 6% • 11%

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 12063 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			

There are 51 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	99	MET	-	EXPRESSION TAG	UNP P37093
A	504	LYS	-	LINKER	UNP P37093
A	505	LEU	-	LINKER	UNP P37093
A	506	ALA	-	LINKER	UNP P37093
A	507	SER	-	LINKER	UNP P37093
A	508	GLY	-	LINKER	UNP P37093
A	509	ALA	-	LINKER	UNP P37093
A	510	GLY	-	LINKER	UNP P37093
A	511	HIS	-	LINKER	UNP P37093
A	674	LEU	-	EXPRESSION TAG	UNP Q02UZ4
A	675	GLU	-	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
A	680	HIS	-	EXPRESSION TAG	UNP Q02UZ4
A	681	HIS	-	EXPRESSION TAG	UNP Q02UZ4
B	99	MET	-	EXPRESSION TAG	UNP P37093
B	504	LYS	-	LINKER	UNP P37093
B	505	LEU	-	LINKER	UNP P37093
B	506	ALA	-	LINKER	UNP P37093
B	507	SER	-	LINKER	UNP P37093

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	508	GLY	-	LINKER	UNP P37093
B	509	ALA	-	LINKER	UNP P37093
B	510	GLY	-	LINKER	UNP P37093
B	511	HIS	-	LINKER	UNP P37093
B	674	LEU	-	EXPRESSION TAG	UNP Q02UZ4
B	675	GLU	-	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
B	680	HIS	-	EXPRESSION TAG	UNP Q02UZ4
B	681	HIS	-	EXPRESSION TAG	UNP Q02UZ4
C	99	MET	-	EXPRESSION TAG	UNP P37093
C	504	LYS	-	LINKER	UNP P37093
C	505	LEU	-	LINKER	UNP P37093
C	506	ALA	-	LINKER	UNP P37093
C	507	SER	-	LINKER	UNP P37093
C	508	GLY	-	LINKER	UNP P37093
C	509	ALA	-	LINKER	UNP P37093
C	510	GLY	-	LINKER	UNP P37093
C	511	HIS	-	LINKER	UNP P37093
C	674	LEU	-	EXPRESSION TAG	UNP Q02UZ4
C	675	GLU	-	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
C	680	HIS	-	EXPRESSION TAG	UNP Q02UZ4
C	681	HIS	-	EXPRESSION TAG	UNP Q02UZ4

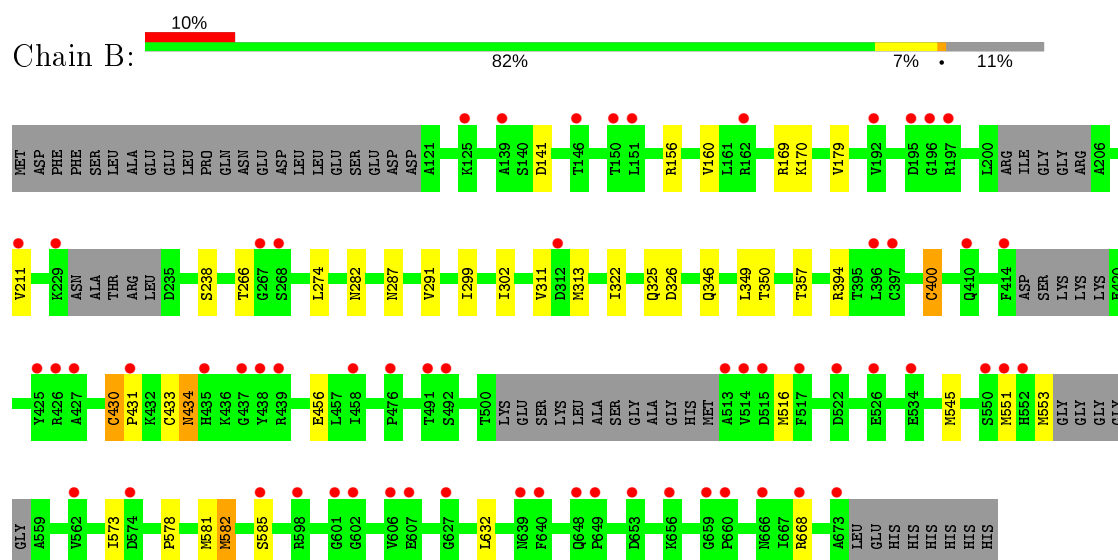
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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.

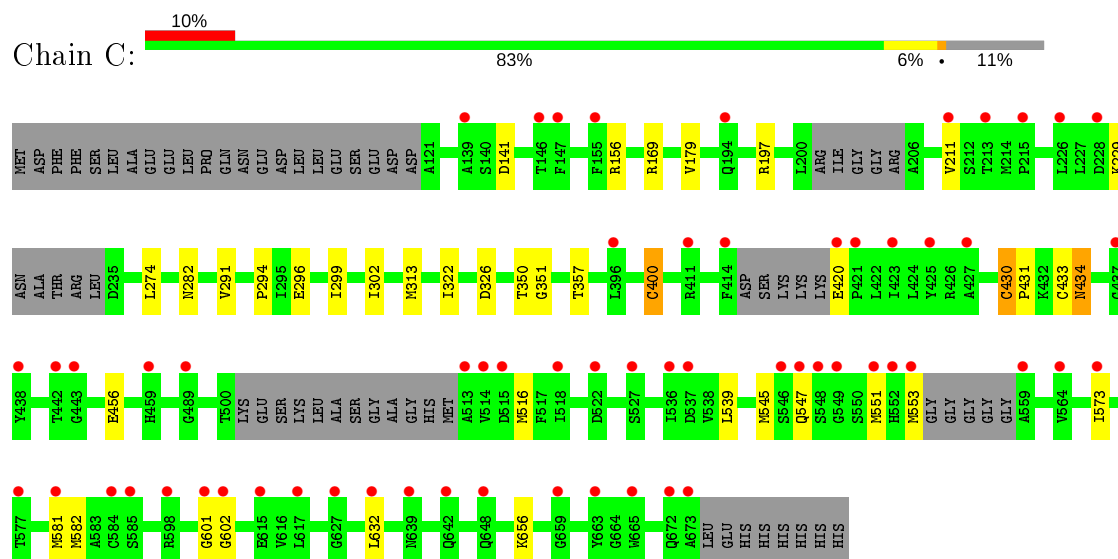
- Molecule 1: Type II secretion system protein E, Hemolysin-coregulated protein



- Molecule 1: Type II secretion system protein E, Hemolysin-coregulated protein



- Molecule 1: Type II secretion system protein E, Hemolysin-coregulated protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 2 ₁ 2 ₁ 2 ₁	Depositor
Cell constants a, b, c, α , β , γ	112.47Å 132.91Å 142.75Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.72 – 4.20 38.69 – 3.78	Depositor EDS
% Data completeness (in resolution range)	98.8 (38.72-4.20) 98.2 (38.69-3.78)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.59 (at 3.76Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.384 , 0.376 0.389 , 0.384	Depositor DCC
R_{free} test set	1107 reflections (5.13%)	wwPDB-VP
Wilson B-factor (Å ²)	170.3	Xtriage
Anisotropy	0.568	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 83.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.42$, $\langle L^2 \rangle = 0.24$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.84	EDS
Total number of atoms	12063	wwPDB-VP
Average B, all atoms (Å ²)	138.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.00% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0.38	6/4074 (0.1%)	0.48	2/5491 (0.0%)
1	B	0.38	6/4074 (0.1%)	0.48	2/5491 (0.0%)
1	C	0.38	6/4074 (0.1%)	0.48	2/5491 (0.0%)
All	All	0.38	18/12222 (0.1%)	0.48	6/16473 (0.0%)

The worst 5 of 18 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	582	MET	CG-SD	6.52	1.98	1.81
1	C	545	MET	CG-SD	6.38	1.97	1.81
1	C	582	MET	CG-SD	6.32	1.97	1.81
1	A	582	MET	CG-SD	6.30	1.97	1.81
1	A	553	MET	CG-SD	6.29	1.97	1.81

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	430	CYS	CA-CB-SG	7.32	127.18	114.00
1	C	430	CYS	CA-CB-SG	7.31	127.16	114.00
1	A	430	CYS	CA-CB-SG	7.29	127.11	114.00
1	A	400	CYS	CA-CB-SG	6.46	125.63	114.00
1	B	400	CYS	CA-CB-SG	6.45	125.61	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	44	4
1	B	4021	0	4099	60	2
1	C	4021	0	4099	36	4
All	All	12063	0	12297	107	5

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 4.

The worst 5 of 107 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:430:CYS:SG	1:B:431:PRO:HD2	1.63	1.37
1:A:430:CYS:SG	1:A:431:PRO:HD2	1.63	1.37
1:C:430:CYS:SG	1:C:431:PRO:HD2	1.63	1.36
1:B:266:THR:HG1	1:C:350:THR:C	1.32	1.32
1:B:266:THR:OG1	1:C:350:THR:C	1.74	1.21

All (5) 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:326:ASP:OD2	1:C:156:ARG:NE[2_555]	1.73	0.47
1:A:125:LYS:NZ	1:B:170:LYS:NZ[3_554]	1.89	0.31
1:B:238:SER:OG	1:C:656:LYS:NZ[3_454]	2.04	0.16
1:A:584:CYS:O	1:C:547:GLN:NE2[2_555]	2.08	0.12
1:A:627:GLY:N	1:C:539:LEU:O[2_555]	2.16	0.04

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	509/583 (87%)	486 (96%)	20 (4%)	3 (1%)	25 65

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	509/583 (87%)	487 (96%)	20 (4%)	2 (0%)	34	72
1	C	509/583 (87%)	490 (96%)	17 (3%)	2 (0%)	34	72
All	All	1527/1749 (87%)	1463 (96%)	57 (4%)	7 (0%)	29	68

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	282	ASN
1	B	282	ASN
1	C	282	ASN
1	A	434	ASN
1	B	434	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	444/493 (90%)	440 (99%)	4 (1%)	78	87
1	B	444/493 (90%)	438 (99%)	6 (1%)	67	80
1	C	444/493 (90%)	440 (99%)	4 (1%)	78	87
All	All	1332/1479 (90%)	1318 (99%)	14 (1%)	73	84

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	313	MET
1	B	456	GLU
1	C	169	ARG
1	B	169	ARG
1	C	141	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	465	GLN
1	C	246	HIS
1	C	565	GLN
1	B	459	HIS
1	C	604	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](#)

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, 95th 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 ²)	Q<0.9
1	A	521/583 (89%)	0.71	75 (14%) 2 3	49, 135, 243, 274	0
1	B	521/583 (89%)	0.66	61 (11%) 4 5	40, 146, 221, 255	0
1	C	521/583 (89%)	0.63	61 (11%) 4 5	44, 137, 219, 296	0
All	All	1563/1749 (89%)	0.67	197 (12%) 3 5	40, 138, 234, 296	0

The worst 5 of 197 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	437	GLY	10.4
1	A	518	ILE	8.7
1	A	617	LEU	6.9
1	B	396	LEU	6.8
1	C	514	VAL	6.6

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.