WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Dec 16, 2024 – 02:10 PM JST

PDB ID	:	8Z7K
Title	:	Crystal structure of Hemolysin co-regulated protein 1 (Hcp1) VariantB from Burkholderia pseudomallei
Authors	:	Lebedev, A.A.; Charoenwattanasatien, R.; Tandhavanant, S.; Chantratita, N.
Deposited on	:	2024-04-20
Resolution	:	1.58 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

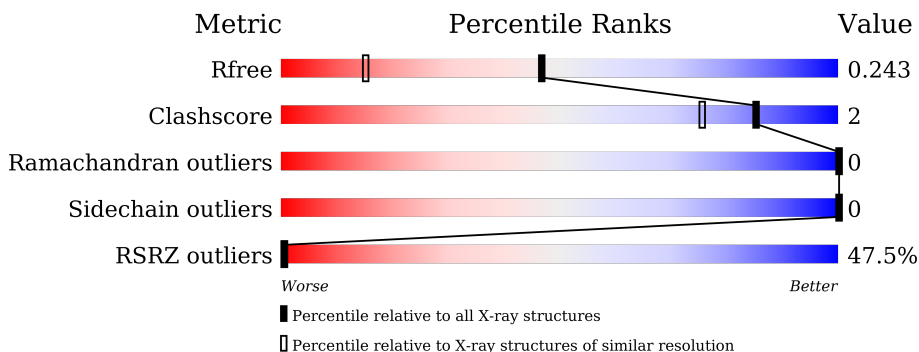|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

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 1.58 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	Percentile Ranks	Value
Rfree		0.243
Clashscore		2
Ramachandran outliers		0
Sidechain outliers		0
RSRZ outliers		47.5%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	164625	7165 (1.60-1.56)
Clashscore	180529	1026 (1.58-1.58)
Ramachandran outliers	177936	1005 (1.58-1.58)
Sidechain outliers	177891	1004 (1.58-1.58)
RSRZ outliers	164620	7163 (1.60-1.56)

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 of 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 <=5%

Mol	Chain	Length	Quality of chain
1	A	178	36% / 76% · 22%
1	B	178	39% / 78% · 19%

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 4475 atoms, of which 2118 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 Hcp family type VI secretion system effector.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	138	Total 2133	C 701	H 1045	N 176	O 205	S 6	0	4	0
1	B	144	Total 2192	C 720	H 1073	N 182	O 211	S 6	0	2	0

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-7	MET	-	initiating methionine	UNP G4XEK9
A	-6	ALA	-	expression tag	UNP G4XEK9
A	-5	ALA	-	expression tag	UNP G4XEK9
A	-4	HIS	-	expression tag	UNP G4XEK9
A	-3	HIS	-	expression tag	UNP G4XEK9
A	-2	HIS	-	expression tag	UNP G4XEK9
A	-1	HIS	-	expression tag	UNP G4XEK9
A	0	HIS	-	expression tag	UNP G4XEK9
A	1	HIS	-	expression tag	UNP G4XEK9
A	14	THR	GLN	engineered mutation	UNP G4XEK9
A	90	TYR	HIS	engineered mutation	UNP G4XEK9
A	91	SER	ARG	engineered mutation	UNP G4XEK9
A	94	ALA	THR	engineered mutation	UNP G4XEK9
A	96	LYS	THR	engineered mutation	UNP G4XEK9
A	97	GLU	-	insertion	UNP G4XEK9
A	168	LYS	THR	engineered mutation	UNP G4XEK9
B	-7	MET	-	initiating methionine	UNP G4XEK9
B	-6	ALA	-	expression tag	UNP G4XEK9
B	-5	ALA	-	expression tag	UNP G4XEK9
B	-4	HIS	-	expression tag	UNP G4XEK9
B	-3	HIS	-	expression tag	UNP G4XEK9
B	-2	HIS	-	expression tag	UNP G4XEK9
B	-1	HIS	-	expression tag	UNP G4XEK9
B	0	HIS	-	expression tag	UNP G4XEK9
B	1	HIS	-	expression tag	UNP G4XEK9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	14	THR	GLN	engineered mutation	UNP G4XEK9
B	90	TYR	HIS	engineered mutation	UNP G4XEK9
B	91	SER	ARG	engineered mutation	UNP G4XEK9
B	94	ALA	THR	engineered mutation	UNP G4XEK9
B	96	LYS	THR	engineered mutation	UNP G4XEK9
B	97	GLU	-	insertion	UNP G4XEK9
B	168	LYS	THR	engineered mutation	UNP G4XEK9

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).
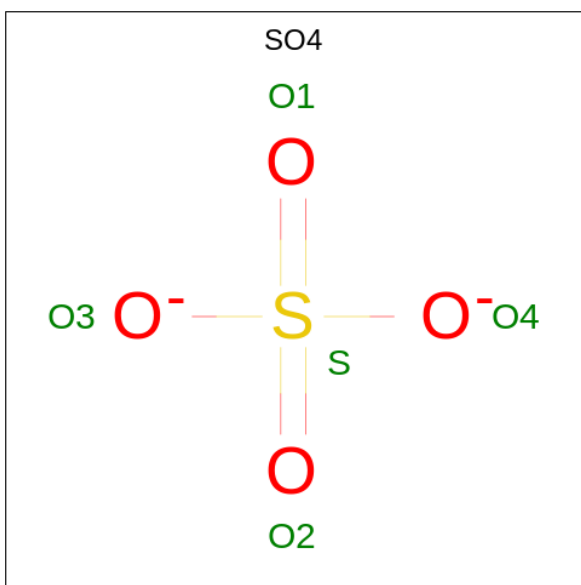
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total 5	O 4	S 1	0	0
2	A	1	Total 5	O 4	S 1	0	0
2	B	1	Total 5	O 4	S 1	0	0
2	B	1	Total 5	O 4	S 1	0	0

- Molecule 3 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	66	Total 66	O 66	0	0
3	B	64	Total 64	O 64	0	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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. 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. 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: Hcp family type VI secretion system effector

Chain A:

• Molecule 1: Hcp family type VI secretion system effector

Chain B:
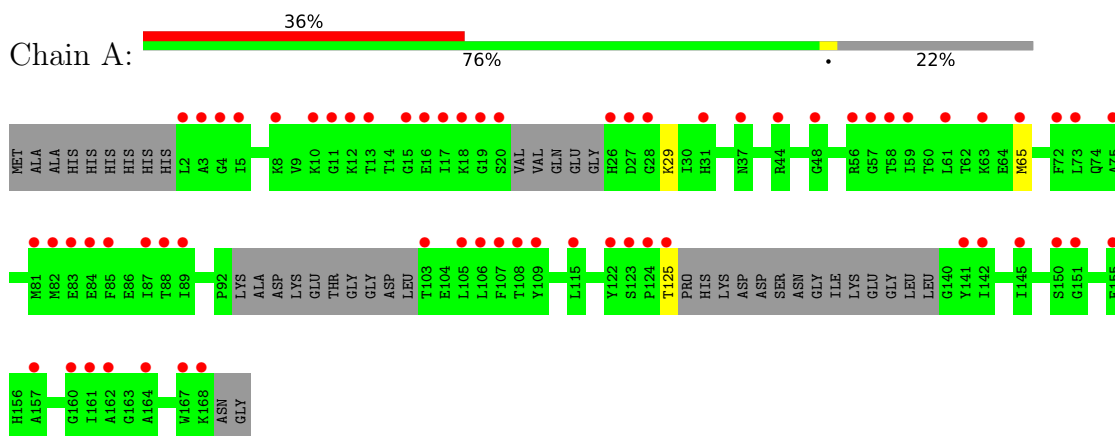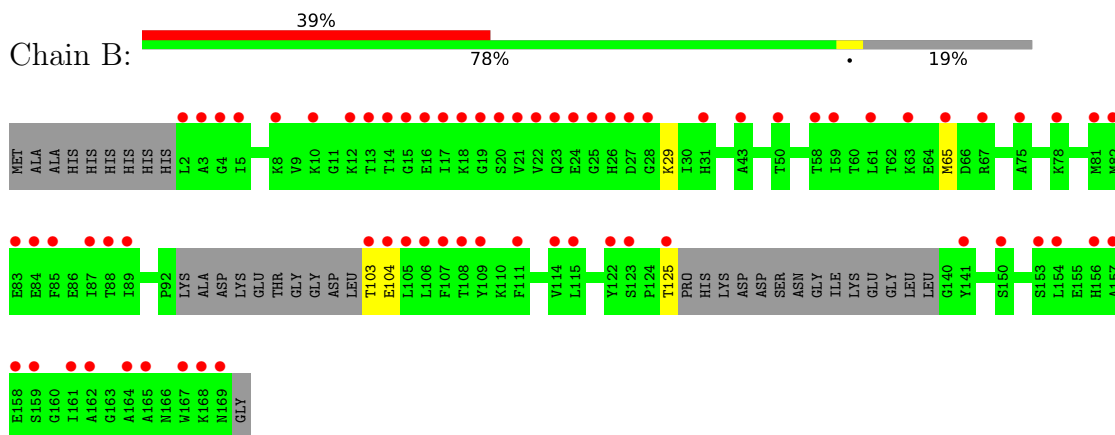
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 6	Depositor
Cell constants a, b, c, α, β, γ	82.72Å 82.72Å 64.07Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	47.76 – 1.58 47.76 – 1.58	Depositor EDS
% Data completeness (in resolution range)	95.9 (47.76-1.58) 95.9 (47.76-1.58)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.17 (at 1.58Å)	Xtriage
Refinement program	REFMAC 5.8.0425	Depositor
R, R_{free}	0.225 , 0.243 0.224 , 0.243	Depositor DCC
R_{free} test set	1558 reflections (4.76%)	wwPDB-VP
Wilson B-factor (Å^2)	27.1	Xtriage
Anisotropy	0.584	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.22 , 60.8	EDS
L-test for twinning[2]	$< \lvert L \rvert > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.052 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.86	EDS
Total number of atoms	4475	wwPDB-VP
Average B, all atoms (Å^2)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.13% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< \lvert L \rvert >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 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: SO4

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	0.50	0/1128	0.85	0/1514
1	B	0.51	0/1149	0.89	0/1545
All	All	0.50	0/2277	0.87	0/3059

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	1088	1045	1054	3	0
1	B	1119	1073	1091	6	0
2	A	10	0	0	0	0
2	B	10	0	0	0	0
3	A	66	0	0	3	0
3	B	64	0	0	6	0
All	All	2357	2118	2145	9	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 2.

The worst 5 of 9 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:125:THR:O	3:B:301:HOH:O	1.60	1.17
1:B:103:THR:N	3:B:304:HOH:O	2.06	0.88
1:B:29:LYS:O	3:B:303:HOH:O	2.00	0.80
1:B:65:MET:HE1	3:B:302:HOH:O	1.85	0.76
1:A:125:THR:O	3:A:303:HOH:O	2.12	0.66

There are no symmetry-related clashes.

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	134/178 (75%)	132 (98%)	2 (2%)	0	100	100
1	B	140/178 (79%)	136 (97%)	4 (3%)	0	100	100
All	All	274/356 (77%)	268 (98%)	6 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	113/145 (78%)	113 (100%)	0	100	100
1	B	115/145 (79%)	115 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	228/290 (79%)	228 (100%)	0	100 100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 oligosaccharides in this entry.

5.6   Ligand geometry ⓘ

4 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	$\#|Z| > 2$
2	SO4	B	202	-	4,4,4	0.42	0	6,6,6	0.25	0
2	SO4	B	201	-	4,4,4	0.32	0	6,6,6	0.05	0
2	SO4	A	202	-	4,4,4	0.41	0	6,6,6	0.26	0
2	SO4	A	201	-	4,4,4	0.28	0	6,6,6	0.08	0

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 ⓘ

Warning: The R factor obtained from EDS is 0.3019, which does not match the depositor's R factor of 0.2247. Please interpret the results in this section carefully.

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(Å²)	Q<0.9
1	A	138/178 (77%)	2.41	64 (46%)	0	0	8, 16, 29, 44	138 (100%)
1	B	144/178 (80%)	2.46	70 (48%)	0	0	10, 16, 31, 54	144 (100%)
All	All	282/356 (79%)	2.43	134 (47%)	0	0	8, 16, 31, 54	282 (100%)

The worst 5 of 134 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	19	GLY	11.0
1	B	4	GLY	10.0
1	B	21	VAL	9.2
1	A	27	ASP	9.1
1	B	27	ASP	9.0

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 monosaccharides 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. The B-factors column lists the minimum, median, 95th 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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	SO4	A	201	5/5	0.88	0.11	51,55,63,65	5
2	SO4	A	202	5/5	0.92	0.17	68,73,78,80	5
2	SO4	B	201	5/5	0.93	0.09	50,50,60,63	5
2	SO4	B	202	5/5	0.94	0.12	41,42,50,50	5

6.5 Other polymers ⓘ

There are no such residues in this entry.