



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 29, 2022 – 12:27 AM EST

PDB ID : 8HHE
Title : Crystal structure of Cry5B from *Bacillus thuringiensis* at 4.5 Å resolution
Authors : Li, J.; Chan, M.K.
Deposited on : 2022-11-16
Resolution : 4.50 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

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.31.2
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)	:	2.31.2

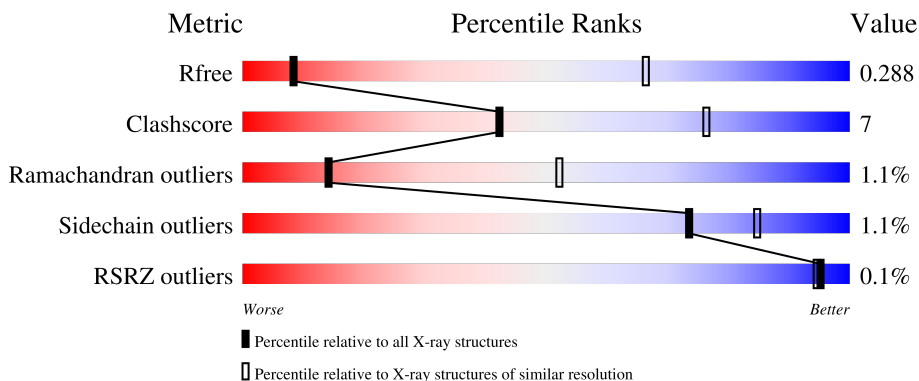
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION



The reported resolution of this entry is 4.50 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	1055 (5.20-3.80)
Clashscore	141614	1123 (5.20-3.80)
Ramachandran outliers	138981	1069 (5.20-3.80)
Sidechain outliers	138945	1050 (5.20-3.80)
RSRZ outliers	127900	1101 (5.30-3.70)

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 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 $\leq 5\%$. 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	808	 65% 14% • 20%
1	B	808	 65% 14% • 20%

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 10254 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 Crystalline entomocidal protoxin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	646	Total	C	N	O	S	0	0	0
			5123	3241	871	994	17			
1	B	647	Total	C	N	O	S	0	0	0
			5131	3247	872	995	17			

There are 72 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-35	MET	-	initiating methionine	UNP B7SB31
A	-34	GLY	-	expression tag	UNP B7SB31
A	-33	SER	-	expression tag	UNP B7SB31
A	-32	SER	-	expression tag	UNP B7SB31
A	-31	HIS	-	expression tag	UNP B7SB31
A	-30	HIS	-	expression tag	UNP B7SB31
A	-29	HIS	-	expression tag	UNP B7SB31
A	-28	HIS	-	expression tag	UNP B7SB31
A	-27	HIS	-	expression tag	UNP B7SB31
A	-26	HIS	-	expression tag	UNP B7SB31
A	-25	SER	-	expression tag	UNP B7SB31
A	-24	SER	-	expression tag	UNP B7SB31
A	-23	GLY	-	expression tag	UNP B7SB31
A	-22	LEU	-	expression tag	UNP B7SB31
A	-21	VAL	-	expression tag	UNP B7SB31
A	-20	PRO	-	expression tag	UNP B7SB31
A	-19	ARG	-	expression tag	UNP B7SB31
A	-18	GLY	-	expression tag	UNP B7SB31
A	-17	SER	-	expression tag	UNP B7SB31
A	-16	HIS	-	expression tag	UNP B7SB31
A	-15	MET	-	expression tag	UNP B7SB31
A	-14	ALA	-	expression tag	UNP B7SB31
A	-13	SER	-	expression tag	UNP B7SB31
A	-12	MET	-	expression tag	UNP B7SB31
A	-11	THR	-	expression tag	UNP B7SB31

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-10	GLY	-	expression tag	UNP B7SB31
A	-9	GLY	-	expression tag	UNP B7SB31
A	-8	GLN	-	expression tag	UNP B7SB31
A	-7	GLN	-	expression tag	UNP B7SB31
A	-6	MET	-	expression tag	UNP B7SB31
A	-5	GLY	-	expression tag	UNP B7SB31
A	-4	ARG	-	expression tag	UNP B7SB31
A	-3	ASP	-	expression tag	UNP B7SB31
A	-2	PRO	-	expression tag	UNP B7SB31
A	-1	ASN	-	expression tag	UNP B7SB31
A	0	SER	-	expression tag	UNP B7SB31
B	-35	MET	-	initiating methionine	UNP B7SB31
B	-34	GLY	-	expression tag	UNP B7SB31
B	-33	SER	-	expression tag	UNP B7SB31
B	-32	SER	-	expression tag	UNP B7SB31
B	-31	HIS	-	expression tag	UNP B7SB31
B	-30	HIS	-	expression tag	UNP B7SB31
B	-29	HIS	-	expression tag	UNP B7SB31
B	-28	HIS	-	expression tag	UNP B7SB31
B	-27	HIS	-	expression tag	UNP B7SB31
B	-26	HIS	-	expression tag	UNP B7SB31
B	-25	SER	-	expression tag	UNP B7SB31
B	-24	SER	-	expression tag	UNP B7SB31
B	-23	GLY	-	expression tag	UNP B7SB31
B	-22	LEU	-	expression tag	UNP B7SB31
B	-21	VAL	-	expression tag	UNP B7SB31
B	-20	PRO	-	expression tag	UNP B7SB31
B	-19	ARG	-	expression tag	UNP B7SB31
B	-18	GLY	-	expression tag	UNP B7SB31
B	-17	SER	-	expression tag	UNP B7SB31
B	-16	HIS	-	expression tag	UNP B7SB31
B	-15	MET	-	expression tag	UNP B7SB31
B	-14	ALA	-	expression tag	UNP B7SB31
B	-13	SER	-	expression tag	UNP B7SB31
B	-12	MET	-	expression tag	UNP B7SB31
B	-11	THR	-	expression tag	UNP B7SB31
B	-10	GLY	-	expression tag	UNP B7SB31
B	-9	GLY	-	expression tag	UNP B7SB31
B	-8	GLN	-	expression tag	UNP B7SB31
B	-7	GLN	-	expression tag	UNP B7SB31
B	-6	MET	-	expression tag	UNP B7SB31
B	-5	GLY	-	expression tag	UNP B7SB31

Continued on next page...

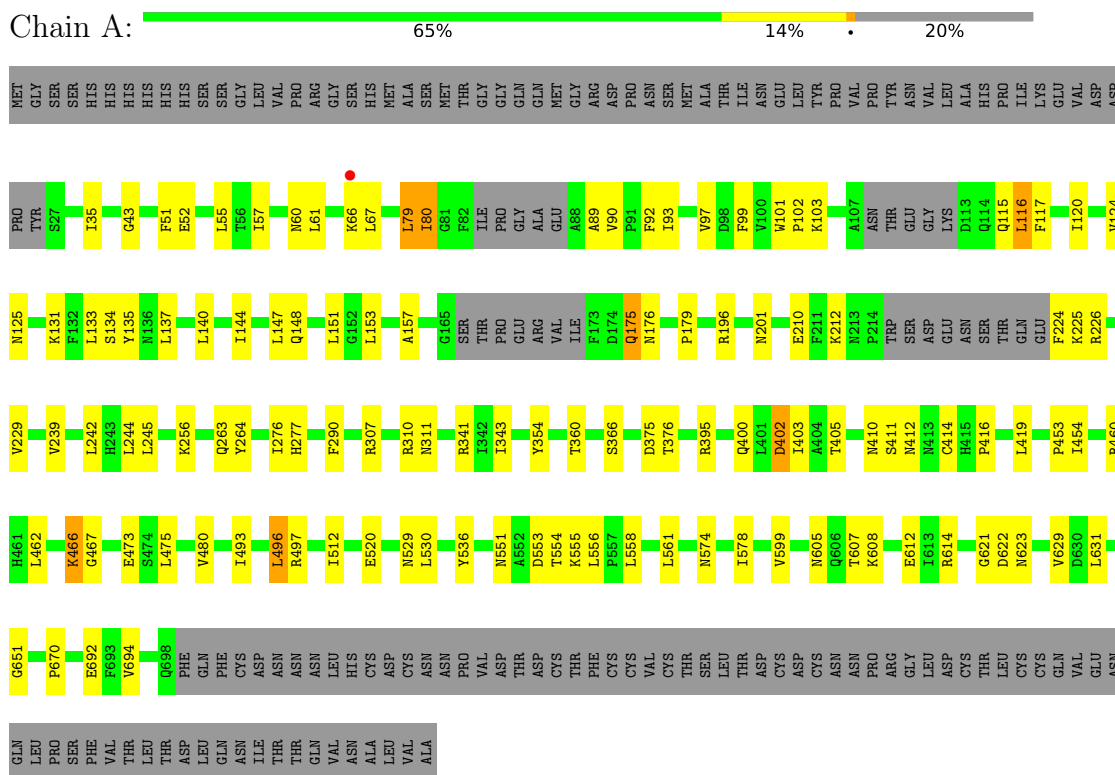
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-4	ARG	-	expression tag	UNP B7SB31
B	-3	ASP	-	expression tag	UNP B7SB31
B	-2	PRO	-	expression tag	UNP B7SB31
B	-1	ASN	-	expression tag	UNP B7SB31
B	0	SER	-	expression tag	UNP B7SB31

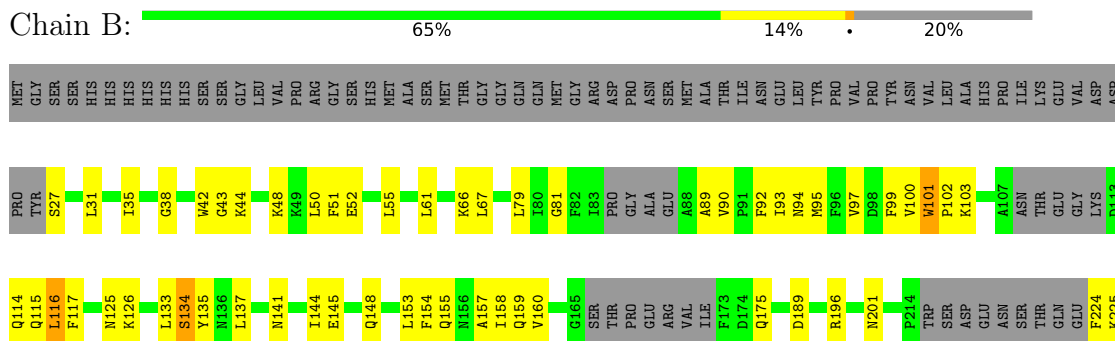
3 Residue-property plots

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 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: Crystalline entomocidal protoxin



• Molecule 1: Crystalline entomocidal protoxin



GLN	PHE	G467	R226
ASN	CYS	█	█
ILE	ASP	E473	V229
THR	ASN	█	█
THR	ASN	V477	V239
GLN	ASN	█	█
VAL	LEU	S478	L242
ASN	HIS	V479	H243
ASN	CYS	V480	L244
ALA	ASP	█	L245
LEU	ASP	I493	█
VAL	CYS	█	█
ALA	ASN	L496	K266
	ASN	R497	█
	PRO	█	K269
	VAL	I512	█
	ASP	█	L276
	THR	N529	█
	ASP	█	F290
	CYS	Y536	█
	THR	█	H329
	PHE	E540	█
	CYS	█	D338
	CYS	D553	█
VAL	VAL	T554	R341
CYS	CYS	K555	I342
ASN	THR	L556	I343
	SER	█	█
LEU	LEU	Q560	P370
THR	THR	L561	█
ASP	ASP	█	D375
	CYS	I578	T376
	ASP	█	P377
	CYS	V599	S378
	ASN	█	█
	ASN	N605	F385
PRO	PRO	█	D402
ARG	ARG	K608	█
GLY	GLY	█	T405
LEU	LEU	E612	█
ASP	ASP	█	█
CYS	CYS	S619	N410
THR	THR	█	S411
LEU	LEU	V629	█
CYS	CYS	D630	C414
CYS	CYS	L631	H415
GLN	GLN	█	P416
VAL	VAL	S641	█
GLU	GLU	█	R428
ASN	ASN	T645	█
GLN	GLN	█	P453
LEU	LEU	I669	I454
PRO	PRO	P670	█
SER	SER	█	V457
PHE	PHE	V676	█
VAL	VAL	█	█
THR	THR	V694	R460
LEU	LEU	█	H461
THR	THR	Q698	L462
ASP	ASP	█	Y463
LEU	LEU	PHE	█
THR	THR	GLN	V465

4 Data and refinement statistics

Property	Value	Source
Space group	P 4 21 2	Depositor
Cell constants a, b, c, α , β , γ	114.40Å 114.40Å 263.35Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.99 – 4.50 19.99 – 4.50	Depositor EDS
% Data completeness (in resolution range)	99.7 (19.99-4.50) 99.7 (19.99-4.50)	Depositor EDS
R_{merge}	0.27	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.11 (at 4.54Å)	Xtriage
Refinement program	PHENIX 1.18_3861, CNS	Depositor
R, R_{free}	0.238 , 0.288 0.237 , 0.288	Depositor DCC
R_{free} test set	590 reflections (5.44%)	wwPDB-VP
Wilson B-factor (Å ²)	104.1	Xtriage
Anisotropy	0.067	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.25 , 36.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.87	EDS
Total number of atoms	10254	wwPDB-VP
Average B, all atoms (Å ²)	107.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.11% 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	0.24	0/5239	0.43	1/7122 (0.0%)
1	B	0.24	0/5247	0.43	1/7133 (0.0%)
All	All	0.24	0/10486	0.43	2/14255 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	175	GLN	C-N-CA	5.31	134.97	121.70
1	A	175	GLN	C-N-CA	5.27	134.87	121.70

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	5123	0	4958	66	0
1	B	5131	0	4969	74	1
All	All	10254	0	9927	139	1

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 139 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:141:ASN:ND2	1:B:145:GLU:OE2	2.04	0.88
1:B:55:LEU:HD11	1:B:99:PHE:HA	1.62	0.82
1:B:89:ALA:HB3	1:B:256:LYS:HD2	1.67	0.75
1:A:89:ALA:HB3	1:A:256:LYS:HD2	1.67	0.74
1:A:117:PHE:HB3	1:A:137:LEU:HD11	1.72	0.72

All (1) 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:B:134:SER:OG	1:B:402:ASP:OD2[4_445]	1.99	0.21

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	636/808 (79%)	569 (90%)	59 (9%)	8 (1%)	12	48
1	B	637/808 (79%)	571 (90%)	60 (9%)	6 (1%)	17	56
All	All	1273/1616 (79%)	1140 (90%)	119 (9%)	14 (1%)	14	52

5 of 14 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	496	LEU
1	A	79	LEU
1	B	496	LEU
1	A	134	SER
1	A	412	ASN

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	576/725 (79%)	569 (99%)	7 (1%)	71	84
1	B	577/725 (80%)	571 (99%)	6 (1%)	76	86
All	All	1153/1450 (80%)	1140 (99%)	13 (1%)	73	85

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	101	TRP
1	B	126	LYS
1	B	641	SER
1	B	338	ASP
1	B	341	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	311	ASN
1	A	461	HIS
1	A	471	ASN
1	A	529	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 monosaccharides 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	646/808 (79%)	-0.39	1 (0%) 95 93	66, 104, 134, 170	0
1	B	647/808 (80%)	-0.36	0 100 100	71, 107, 137, 190	0
All	All	1293/1616 (80%)	-0.38	1 (0%) 95 94	66, 106, 137, 190	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	66	LYS	2.5

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 monosaccharides 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.