



# wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 17, 2023 – 03:38 PM EST

PDB ID : 4W8J  
Title : Structure of the full-length insecticidal protein Cry1Ac reveals intriguing details of toxin packaging into in vivo formed crystals  
Authors : Evdokimov, A.G.; Moshiri, F.; Sturman, E.J.; Rydel, T.J.; Zheng, M.; Seale, J.W.; Franklin, S.  
Deposited on : 2014-08-25  
Resolution : 2.78 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

<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 : 2.36  
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) : 2.36

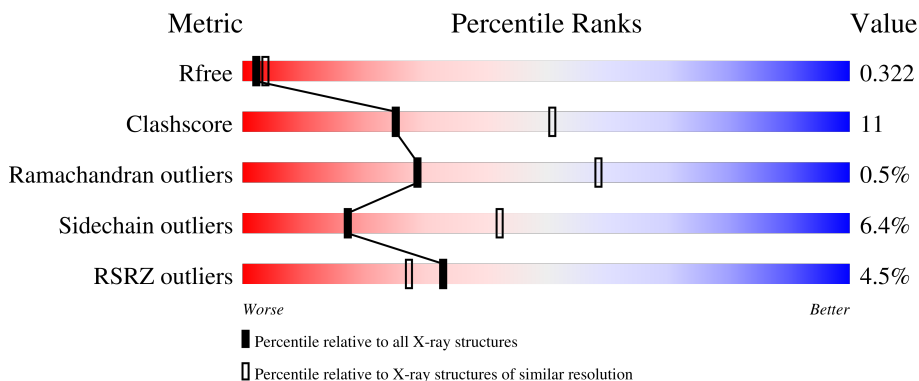
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.78 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	4107 (2.80-2.76)
Clashscore	141614	4575 (2.80-2.76)
Ramachandran outliers	138981	4487 (2.80-2.76)
Sidechain outliers	138945	4489 (2.80-2.76)
RSRZ outliers	127900	4027 (2.80-2.76)

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  $\geq 3$ , 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	1184	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 8206 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 Pesticidal crystal protein cry1Ac.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1017	8152	5170	1406	1567	9	0	0	0

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	148	LEU	PHE	conflict	UNP M4LET9
A	462	VAL	PHE	conflict	UNP M4LET9
A	661	SER	CYS	engineered mutation	UNP M4LET9
A	730	SER	CYS	engineered mutation	UNP M4LET9
A	796	SER	CYS	engineered mutation	UNP M4LET9
A	802	SER	CYS	engineered mutation	UNP M4LET9
A	814	SER	CYS	engineered mutation	UNP M4LET9
A	816	SER	CYS	engineered mutation	UNP M4LET9
A	822	SER	CYS	engineered mutation	UNP M4LET9
A	837	SER	CYS	engineered mutation	UNP M4LET9
A	990	SER	CYS	engineered mutation	UNP M4LET9
A	1025	SER	CYS	engineered mutation	UNP M4LET9
A	1045	SER	CYS	engineered mutation	UNP M4LET9
A	1063	SER	CYS	engineered mutation	UNP M4LET9
A	1076	SER	CYS	engineered mutation	UNP M4LET9
A	1125	SER	CYS	engineered mutation	UNP M4LET9
A	1179	HIS	-	expression tag	UNP M4LET9
A	1180	HIS	-	expression tag	UNP M4LET9
A	1181	HIS	-	expression tag	UNP M4LET9
A	1182	HIS	-	expression tag	UNP M4LET9
A	1183	HIS	-	expression tag	UNP M4LET9
A	1184	HIS	-	expression tag	UNP M4LET9

- Molecule 2 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	3	Total K 3 3	0	0

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Ca 1 1	0	0

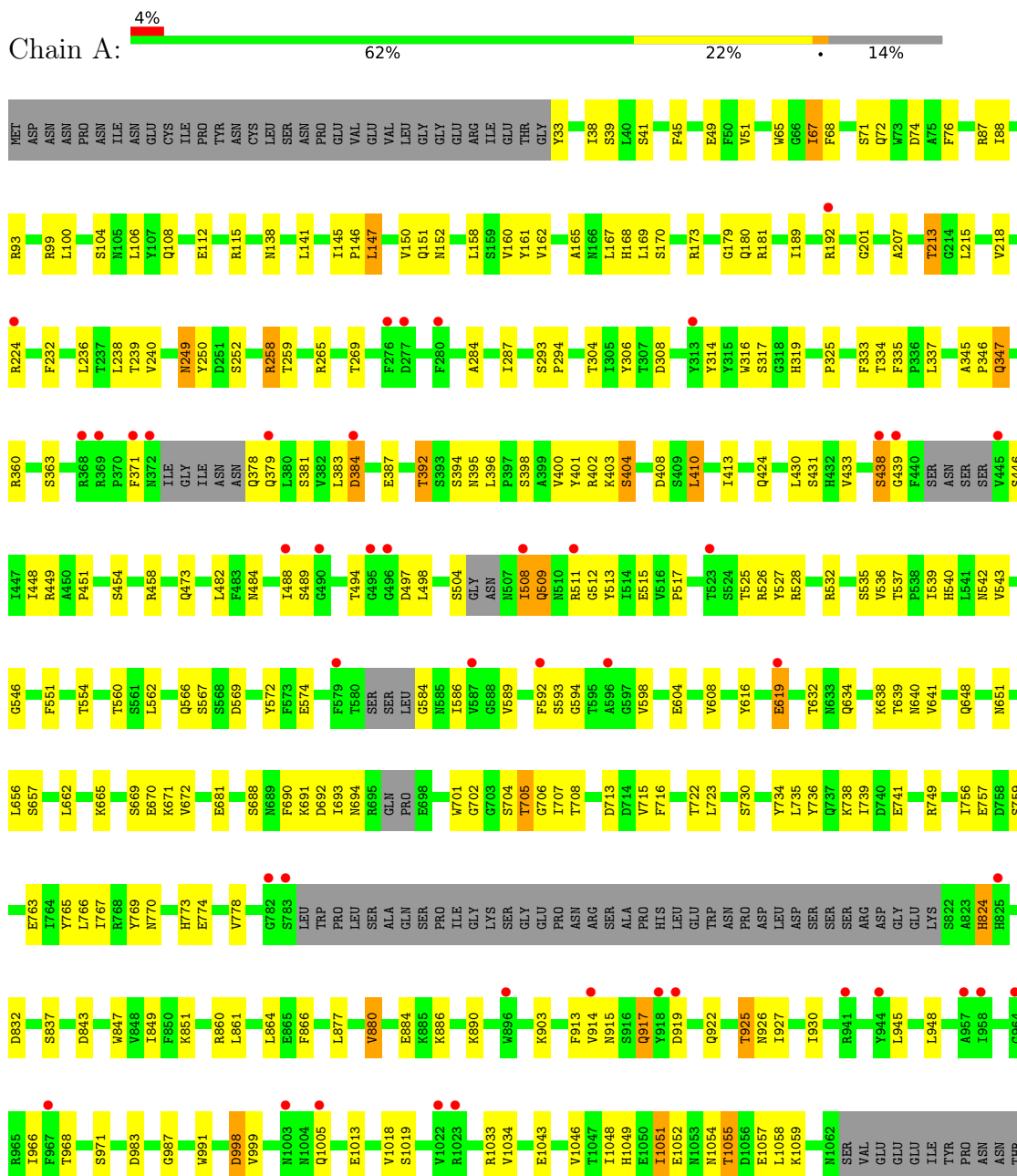
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	50	Total O 50 50	0	0

### 3 Residue-property plots [i](#)

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: Pesticidal crystal protein cry1Ac



VAL  
THR  
SER  
ASN  
ASP  
TYR  
THR  
VAL  
ASN  
GLN  
GLU  
GLU  
TYR  
GLY  
GLY  
ALA  
TYR  
THR  
SER  
SER  
ASN  
ARG  
ASN  
ARG  
GLY  
TYR  
ASN  
GLU  
ALA  
PRO  
VAL  
SER  
PRO  
ALA  
ASP  
TYR  
ALA  
SER  
VAL  
TYR  
GLU  
LYS  
SER  
TYR  
THR  
ASP  
GLY  
ARG  
ARG  
GLU  
ASN  
PRO  
SER  
GLU  
PHE  
ASN  
ARG  
GLY  
TYR  
ARG  
ASP

TYR  
THR  
PRO  
LEU  
P1138  
L1146  
D1153  
K1154  
V1155  
W1156  
Y1157  
E1158  
I1159  
T1162  
E1163  
E1172  
E1178  
HIS  
HIS  
HIS  
HIS  
HIS  
HIS

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	87.31Å 87.31Å 266.40Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	82.97 – 2.78 39.18 – 2.78	Depositor EDS
% Data completeness (in resolution range)	98.8 (82.97-2.78) 98.9 (39.18-2.78)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.33 (at 2.77Å)	Xtrriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.226 , 0.332 0.217 , 0.322	Depositor DCC
$R_{free}$ test set	1340 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	53.4	Xtrriage
Anisotropy	0.002	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 40.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	8206	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	53.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.04% 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](#)

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

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	0.61	0/8335	0.71	0/11317

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 [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	8152	0	7896	171	0
2	A	3	0	0	0	0
3	A	1	0	0	0	0
4	A	50	0	0	8	0
All	All	8206	0	7896	171	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 11.

The worst 5 of 171 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:922:GLN:HB2	1:A:925:THR:CG2	2.03	0.88
1:A:722:THR:HG22	1:A:860:ARG:HG2	1.54	0.87
1:A:100:LEU:HD11	1:A:160:VAL:HG11	1.56	0.86
1:A:715:VAL:HG21	1:A:757:GLU:HG3	1.57	0.84
1:A:213:THR:HG22	4:A:1329:HOH:O	1.80	0.80

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	1001/1184 (84%)	950 (95%)	46 (5%)	5 (0%)	29 58

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	347	GLN
1	A	917	GLN
1	A	1054	ASN
1	A	1052	GLU
1	A	1163	GLU

### 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	888/1036 (86%)	831 (94%)	57 (6%)	17 42

5 of 57 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	619	GLU
1	A	1055	THR
1	A	691	LYS
1	A	1051	ILE
1	A	966	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	770	ASN
1	A	855	GLN
1	A	1002	GLN
1	A	917	GLN
1	A	301	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](#)

Of 4 ligands modelled in this entry, 4 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 [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	1017/1184 (85%)	0.38	46 (4%) 33 27	38, 52, 68, 92	0

The worst 5 of 46 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	371	PHE	6.7
1	A	783	SER	6.3
1	A	1138	PRO	5.6
1	A	379	GLN	4.8
1	A	369	ARG	4.4

### 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](#)

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, 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	B-factors(Å <sup>2</sup> )	Q<0.9
2	K	A	1202	1/1	0.87	0.10	88,88,88,88	0
2	K	A	1203	1/1	0.87	0.11	88,88,88,88	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	K	A	1201	1/1	0.92	0.19	84,84,84,84	0
3	CA	A	1204	1/1	0.96	0.07	90,90,90,90	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.