



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 13, 2023 – 11:17 PM JST

PDB ID : 5YJS  
Title : Structure of vicilin from Capsicum annuum  
Authors : Shikhi, M.; Nair, D.T.; Salunke, D.M.  
Deposited on : 2017-10-11  
Resolution : 2.16 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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
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

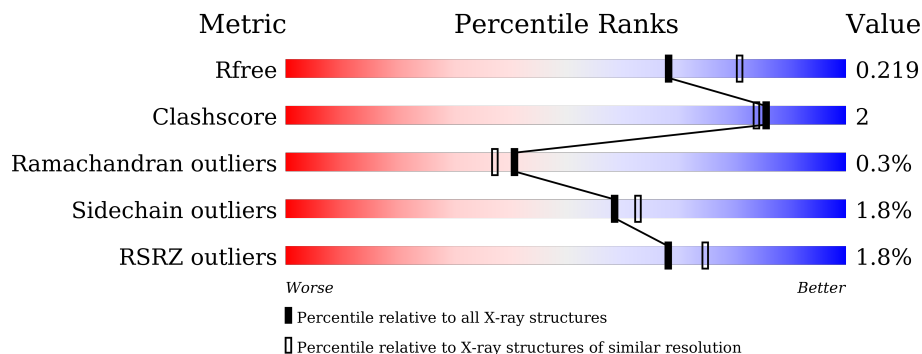
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.16 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	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)
RSRZ outliers	127900	1456 (2.16-2.16)

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	390	 86% 8% 5%
1	B	390	 3% 89% 6% 5%

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 6169 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 vicilin-like antimicrobial peptides 2-2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	370	2904	1855	504	536	9	0	0	0
1	B	370	2904	1855	504	536	9	0	0	0

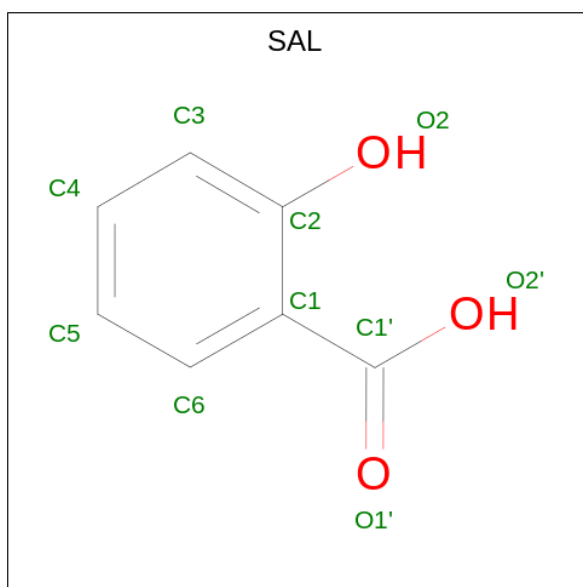
- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Cl	0	0
			1	1		

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

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

- Molecule 4 is 2-HYDROXYBENZOIC ACID (three-letter code: SAL) (formula: C<sub>7</sub>H<sub>6</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 10 7 3	0	0
4	B	1	Total C O 10 7 3	0	0

- Molecule 5 is COPPER (I) ION (three-letter code: CU1) (formula: Cu).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total Cu 1 1	0	0
5	B	1	Total Cu 1 1	0	0

- Molecule 6 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: C<sub>8</sub>H<sub>18</sub>O<sub>5</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	C	O	0	0
			13	8	5		

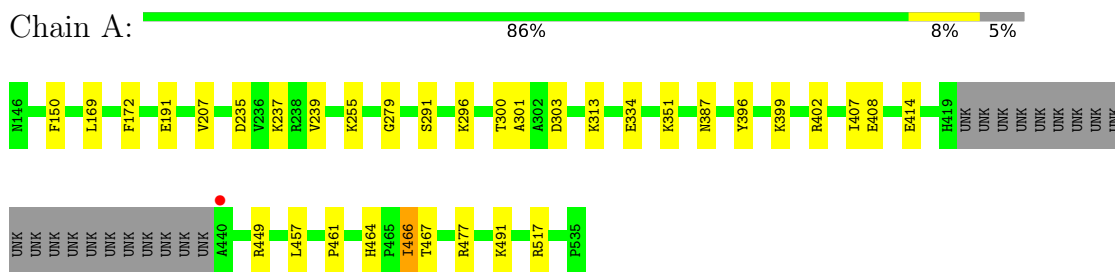
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	240	Total	O	0	0
			240	240		
7	B	84	Total	O	0	0
			84	84		

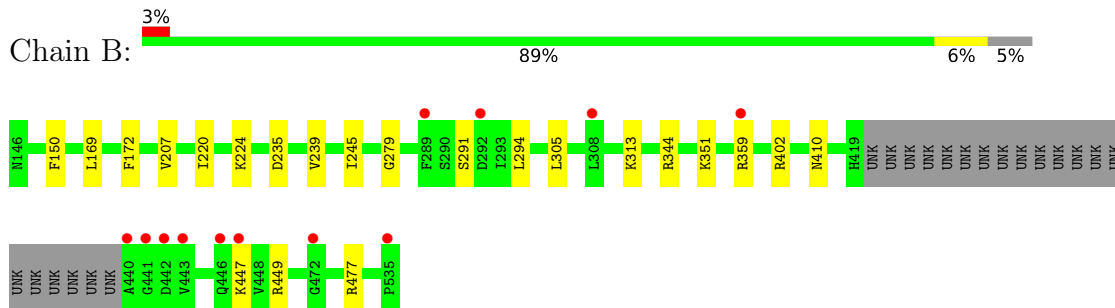
### 3 Residue-property plots [i](#)

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: vicilin-like antimicrobial peptides 2-2



- Molecule 1: vicilin-like antimicrobial peptides 2-2



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 21 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	148.03Å 148.03Å 148.03Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.16 74.02 – 2.16	Depositor EDS
% Data completeness (in resolution range)	100.0 (50.00-2.16) 100.0 (74.02-2.16)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.05 (at 2.16Å)	Xtrriage
Refinement program	REFMAC 5.8.0155	Depositor
R, $R_{free}$	0.175 , 0.217 0.183 , 0.219	Depositor DCC
$R_{free}$ test set	2985 reflections (5.16%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	31.1	Xtrriage
Anisotropy	0.000	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 40.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.024 for l,-k,h	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	6169	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	37.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.27% 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: CU1, PG4, SAL, CL, NA

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	1.00	2/2968 (0.1%)	0.95	5/4002 (0.1%)
1	B	0.81	1/2968 (0.0%)	0.89	4/4002 (0.1%)
All	All	0.91	3/5936 (0.1%)	0.92	9/8004 (0.1%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	237	LYS	CD-CE	-7.18	1.33	1.51
1	B	235	ASP	CB-CG	5.49	1.63	1.51
1	A	517	ARG	CZ-NH1	5.44	1.40	1.33

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	517	ARG	NE-CZ-NH2	-7.32	116.64	120.30
1	A	477	ARG	NE-CZ-NH1	6.67	123.64	120.30
1	A	517	ARG	NE-CZ-NH1	6.52	123.56	120.30
1	A	303	ASP	CB-CG-OD1	6.25	123.92	118.30
1	A	235	ASP	CB-CG-OD1	6.17	123.85	118.30

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	2904	0	2862	17	0
1	B	2904	0	2862	12	0
2	A	1	0	0	0	0
3	A	1	0	0	0	0
4	A	10	0	5	1	0
4	B	10	0	4	0	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
6	A	13	0	18	0	0
7	A	240	0	0	0	0
7	B	84	0	0	1	0
All	All	6169	0	5751	26	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 26 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:207:VAL:HG23	1:A:239:VAL:HB	1.67	0.75
1:A:387:ASN:HD21	1:B:351:LYS:HE3	1.55	0.71
1:A:207:VAL:CG2	1:A:239:VAL:HB	2.20	0.71
1:B:207:VAL:HG23	1:B:239:VAL:HB	1.77	0.66
1:A:407:ILE:HG22	1:A:408:GLU:HG3	1.78	0.64

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	366/390 (94%)	360 (98%)	5 (1%)	1 (0%)	41	37
1	B	366/390 (94%)	359 (98%)	6 (2%)	1 (0%)	41	37

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	732/780 (94%)	719 (98%)	11 (2%)	2 (0%)	41	37

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	279	GLY
1	A	279	GLY

### 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	300/300 (100%)	293 (98%)	7 (2%)	50	53
1	B	300/300 (100%)	296 (99%)	4 (1%)	69	74
All	All	600/600 (100%)	589 (98%)	11 (2%)	59	63

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	224	LYS
1	B	291	SER
1	B	410	ASN
1	B	313	LYS
1	A	351	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	387	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 7 ligands modelled in this entry, 4 are monoatomic - leaving 3 for Mogul analysis.

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
6	PG4	A	605	-	12,12,12	0.61	0	11,11,11	0.46	0
4	SAL	B	601	-	10,10,10	1.51	1 (10%)	13,13,13	1.84	3 (23%)
4	SAL	A	603	-	10,10,10	2.16	3 (30%)	13,13,13	1.04	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '2' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
6	PG4	A	605	-	-	5/10/10/10	-
4	SAL	B	601	-	-	0/4/4/4	0/1/1/1
4	SAL	A	603	-	-	0/4/4/4	0/1/1/1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	603	SAL	C1-C2	5.71	1.49	1.40
4	B	601	SAL	C1-C2	4.20	1.47	1.40
4	A	603	SAL	O2-C2	2.27	1.41	1.36
4	A	603	SAL	O1'-C1'	2.06	1.28	1.22

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	601	SAL	O1'-C1'-C1	-4.24	111.64	121.94
4	B	601	SAL	O2'-C1'-C1	3.77	126.16	115.31
4	B	601	SAL	C2-C1-C1'	2.71	122.92	120.03

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	A	605	PG4	O2-C3-C4-O3
6	A	605	PG4	O4-C7-C8-O5
6	A	605	PG4	O1-C1-C2-O2
6	A	605	PG4	O3-C5-C6-O4
6	A	605	PG4	C1-C2-O2-C3

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	603	SAL	1	0

## 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	370/390 (94%)	-0.23	1 (0%) 94 95	16, 26, 47, 96	0
1	B	370/390 (94%)	-0.02	12 (3%) 47 56	27, 43, 69, 97	0
All	All	740/780 (94%)	-0.13	13 (1%) 68 75	16, 34, 66, 97	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	440	ALA	3.2
1	B	359	ARG	3.0
1	B	446	GLN	2.9
1	B	292	ASP	2.8
1	B	442	ASP	2.7

### 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( $\text{\AA}^2$ )	Q<0.9
2	CL	A	601	1/1	0.84	0.19	70,70,70,70	0
6	PG4	A	605	13/13	0.89	0.17	54,65,75,80	0
5	CU1	B	602	1/1	0.96	0.07	46,46,46,46	0
4	SAL	B	601	10/10	0.96	0.17	34,47,53,66	0
4	SAL	A	603	10/10	0.97	0.14	22,27,32,42	0
3	NA	A	602	1/1	0.97	0.15	43,43,43,43	0
5	CU1	A	604	1/1	1.00	0.08	31,31,31,31	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.