



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 2, 2016 – 08:55 PM EDT

PDB ID : 5E1R
Title : Crystal structure of pecan (*carya illinoensis*) vicilin, a new food allergen
Authors : Zhang, Y.Z.
Deposited on : 2015-09-30
Resolution : 2.65 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

<http://wwpdb.org/validation/2016/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
Mogul : 1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20027939
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20027939

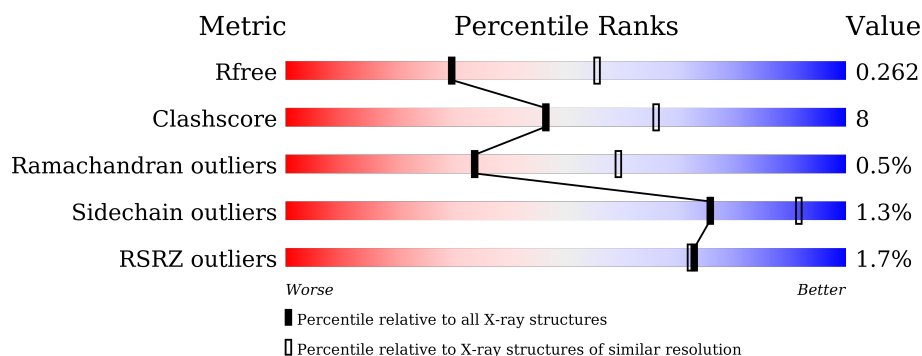
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.65 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}	91344	3152 (2.70-2.62)
Clashscore	102246	3524 (2.70-2.62)
Ramachandran outliers	100387	3469 (2.70-2.62)
Sidechain outliers	100360	3469 (2.70-2.62)
RSRZ outliers	91569	3161 (2.70-2.62)

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. 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	426	
1	B	426	
1	C	426	
1	D	426	
1	E	426	
1	F	426	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	MPD	A	802	-	-	-	X
3	MPD	B	802	-	-	-	X
3	MPD	C	802	-	-	-	X
3	MPD	D	802	-	-	-	X
3	MPD	E	802	-	-	-	X
3	MPD	F	802	-	-	-	X

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 17082 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 7S vicilin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	357	Total	C	N	O	S	0	1	0
			2841	1789	504	539	9			
1	B	355	Total	C	N	O	S	0	0	0
			2799	1766	490	534	9			
1	C	358	Total	C	N	O	S	0	0	0
			2826	1785	499	533	9			
1	D	357	Total	C	N	O	S	0	0	0
			2838	1785	508	536	9			
1	E	358	Total	C	N	O	S	0	0	0
			2866	1799	512	546	9			
1	F	358	Total	C	N	O	S	0	0	0
			2818	1776	496	537	9			

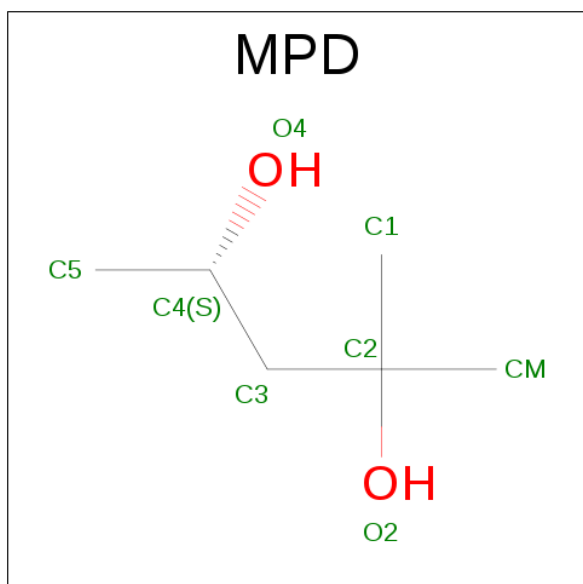
There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	367	MET	-	initiating methionine	UNP B3STU4
A	368	SER	-	expression tag	UNP B3STU4
B	367	MET	-	initiating methionine	UNP B3STU4
B	368	SER	-	expression tag	UNP B3STU4
C	367	MET	-	initiating methionine	UNP B3STU4
C	368	SER	-	expression tag	UNP B3STU4
D	367	MET	-	initiating methionine	UNP B3STU4
D	368	SER	-	expression tag	UNP B3STU4
E	367	MET	-	initiating methionine	UNP B3STU4
E	368	SER	-	expression tag	UNP B3STU4
F	367	MET	-	initiating methionine	UNP B3STU4
F	368	SER	-	expression tag	UNP B3STU4

- Molecule 2 is COPPER (II) ION (three-letter code: CU) (formula: Cu).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	D	1	Total Cu 1 1	0	0
2	E	1	Total Cu 1 1	0	0
2	B	1	Total Cu 1 1	0	0
2	C	1	Total Cu 1 1	0	0
2	A	1	Total Cu 1 1	0	0
2	F	1	Total Cu 1 1	0	0

- Molecule 3 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula: $C_6H_{14}O_2$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 8 6 2	0	0
3	B	1	Total C O 8 6 2	0	0
3	C	1	Total C O 8 6 2	0	0
3	D	1	Total C O 8 6 2	0	0
3	E	1	Total C O 8 6 2	0	0

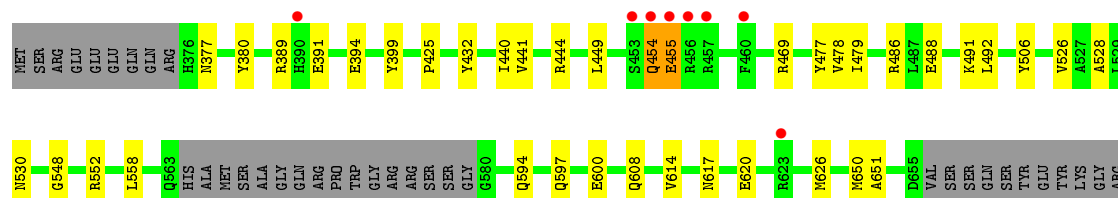
Continued on next page...

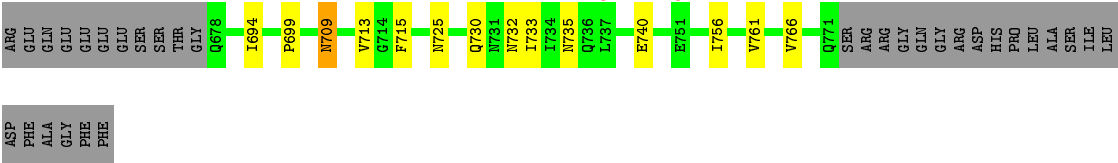
Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	F	1	Total	C	O	0	0
			8	6	2		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	7	Total	O	0	0
			7	7		
4	B	8	Total	O	0	0
			8	8		
4	C	7	Total	O	0	0
			7	7		
4	D	7	Total	O	0	0
			7	7		
4	E	8	Total	O	0	0
			8	8		
4	F	3	Total	O	0	0
			3	3		





4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	42.69Å 151.61Å 342.78Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.60 – 2.65 48.47 – 2.65	Depositor EDS
% Data completeness (in resolution range)	89.0 (46.60-2.65) 89.1 (48.47-2.65)	Depositor EDS
R_{merge}	0.20	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.28 (at 2.65Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.218 , 0.262 0.219 , 0.262	Depositor DCC
R_{free} test set	1996 reflections (3.38%)	DCC
Wilson B-factor (Å ²)	43.8	Xtriage
Anisotropy	1.006	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 26.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.35$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	17082	wwPDB-VP
Average B, all atoms (Å ²)	18.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.81% 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](#)

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

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	0.27	0/2896	0.44	0/3913
1	B	0.24	0/2851	0.44	0/3856
1	C	0.25	0/2881	0.45	0/3895
1	D	0.27	0/2893	0.43	0/3911
1	E	0.26	0/2919	0.46	0/3942
1	F	0.24	0/2872	0.44	0/3887
All	All	0.26	0/17312	0.44	0/23404

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	2841	0	2741	53	0
1	B	2799	0	2684	57	0
1	C	2826	0	2707	53	0
1	D	2838	0	2731	46	0
1	E	2866	0	2769	45	0
1	F	2818	0	2688	35	0
2	A	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	F	1	0	0	0	0
3	A	8	0	14	2	0
3	B	8	0	14	1	0
3	C	8	0	14	1	0
3	D	8	0	14	1	0
3	E	8	0	14	2	0
3	F	8	0	14	0	0
4	A	7	0	0	0	0
4	B	8	0	0	0	0
4	C	7	0	0	0	0
4	D	7	0	0	0	0
4	E	8	0	0	0	0
4	F	3	0	0	0	0
All	All	17082	0	16404	254	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 8.

The worst 5 of 254 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:F:709:ASN:HD22	1:F:709:ASN:H	1.20	0.90
1:E:720:LYS:HG2	1:E:721:ASN:HD22	1.40	0.87
1:B:532:PRO:HG2	1:B:535:ARG:HD3	1.60	0.83
1:B:735:ASN:HD22	1:B:735:ASN:N	1.76	0.83
1:B:389:ARG:NH2	1:B:513:THR:OG1	2.15	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	352/426 (83%)	336 (96%)	15 (4%)	1 (0%)	46	72
1	B	349/426 (82%)	333 (95%)	14 (4%)	2 (1%)	30	54
1	C	352/426 (83%)	334 (95%)	17 (5%)	1 (0%)	46	72
1	D	351/426 (82%)	332 (95%)	16 (5%)	3 (1%)	21	44
1	E	352/426 (83%)	335 (95%)	16 (4%)	1 (0%)	46	72
1	F	352/426 (83%)	332 (94%)	18 (5%)	2 (1%)	30	54
All	All	2108/2556 (82%)	2002 (95%)	96 (5%)	10 (0%)	34	59

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	764	TYR
1	C	764	TYR
1	D	721	ASN
1	D	764	TYR
1	B	764	TYR

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	298/369 (81%)	294 (99%)	4 (1%)	76	91
1	B	292/369 (79%)	289 (99%)	3 (1%)	82	94
1	C	293/369 (79%)	290 (99%)	3 (1%)	82	94
1	D	299/369 (81%)	293 (98%)	6 (2%)	63	86
1	E	304/369 (82%)	299 (98%)	5 (2%)	70	89
1	F	294/369 (80%)	291 (99%)	3 (1%)	82	94
All	All	1780/2214 (80%)	1756 (99%)	24 (1%)	76	91

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	391	GLU
1	D	623	ARG
1	F	614	VAL
1	D	542	GLN
1	D	605	GLU

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

Mol	Chain	Res	Type
1	E	630	HIS
1	F	709	ASN
1	E	680	GLN
1	B	735	ASN
1	E	721	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](#)

Of 12 ligands modelled in this entry, 6 are monoatomic - leaving 6 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
3	MPD	A	802	-	6,7,7	0.32	0	6,10,10	0.33	0
3	MPD	B	802	-	6,7,7	0.33	0	6,10,10	0.55	0
3	MPD	C	802	-	6,7,7	0.33	0	6,10,10	0.30	0
3	MPD	D	802	-	6,7,7	0.34	0	6,10,10	0.34	0
3	MPD	E	802	-	6,7,7	0.33	0	6,10,10	0.66	0
3	MPD	F	802	-	6,7,7	0.35	0	6,10,10	0.35	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. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	MPD	A	802	-	-	0/5/5/5	0/0/0/0
3	MPD	B	802	-	-	0/5/5/5	0/0/0/0
3	MPD	C	802	-	-	0/5/5/5	0/0/0/0
3	MPD	D	802	-	-	0/5/5/5	0/0/0/0
3	MPD	E	802	-	-	0/5/5/5	0/0/0/0
3	MPD	F	802	-	-	0/5/5/5	0/0/0/0

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.

5 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	802	MPD	2	0
3	B	802	MPD	1	0
3	C	802	MPD	1	0
3	D	802	MPD	1	0
3	E	802	MPD	2	0

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 [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	357/426 (83%)	-0.07	2 (0%) 90 91	6, 14, 29, 38	0
1	B	355/426 (83%)	0.10	9 (2%) 61 59	6, 16, 32, 42	0
1	C	358/426 (84%)	0.13	6 (1%) 73 72	6, 15, 33, 50	0
1	D	357/426 (83%)	0.05	7 (1%) 68 67	8, 18, 35, 45	0
1	E	358/426 (84%)	0.03	3 (0%) 87 87	6, 16, 31, 41	0
1	F	358/426 (84%)	0.07	10 (2%) 56 55	9, 19, 36, 48	0
All	All	2143/2556 (83%)	0.05	37 (1%) 73 72	6, 16, 33, 50	0

The worst 5 of 37 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	390	HIS	8.2
1	D	390	HIS	5.8
1	F	455	GLU	4.4
1	D	536	LEU	4.0
1	E	391	GLU	3.9

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](#)

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å ²)	Q<0.9
3	MPD	B	802	8/8	0.88	0.33	9.30	6,13,20,27	0
3	MPD	D	802	8/8	0.86	0.39	8.04	9,12,15,19	0
3	MPD	E	802	8/8	0.88	0.36	6.90	10,12,21,24	0
3	MPD	F	802	8/8	0.86	0.38	6.10	12,19,25,26	0
3	MPD	C	802	8/8	0.89	0.24	2.86	4,15,21,21	0
3	MPD	A	802	8/8	0.88	0.25	2.76	8,15,19,21	0
2	CU	A	801	1/1	0.98	0.12	-1.19	12,12,12,12	0
2	CU	C	801	1/1	0.99	0.09	-2.25	11,11,11,11	0
2	CU	B	801	1/1	0.98	0.08	-3.55	16,16,16,16	0
2	CU	E	801	1/1	0.99	0.07	-3.74	18,18,18,18	0
2	CU	F	801	1/1	0.99	0.07	-	22,22,22,22	0
2	CU	D	801	1/1	0.99	0.07	-	11,11,11,11	0

6.5 Other polymers [i](#)

There are no such residues in this entry.