



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 1, 2021 – 10:09 AM JST

PDB ID : 7VEY  
Title : Crystal structure of Cyclosorus parasiticus chalcone synthase 1 (CpCHS1)  
Authors : Li, J.X.; Cheng, A.X.  
Deposited on : 2021-09-10  
Resolution : 1.90 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 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	:	2.23.2
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)	:	2.23.2

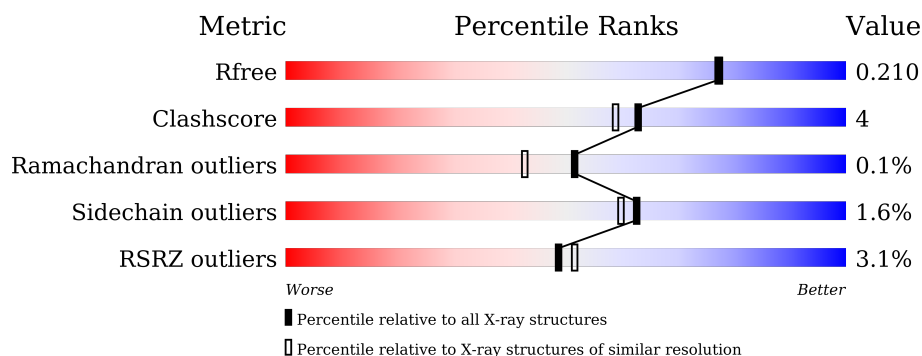
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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	404	 2% 90% 7% ..
1	B	404	 3% 90% 7% ..
1	C	404	 2% 89% 9% ..
1	D	404	 5% 85% 11% ..

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 13325 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 chalcone synthases.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	396	Total	C	N	O	S	0	0	0
			3048	1939	517	570	22			
1	B	398	Total	C	N	O	S	0	0	0
			3061	1946	520	573	22			
1	C	399	Total	C	N	O	S	0	0	0
			3068	1950	521	575	22			
1	D	396	Total	C	N	O	S	0	0	0
			3049	1940	517	570	22			

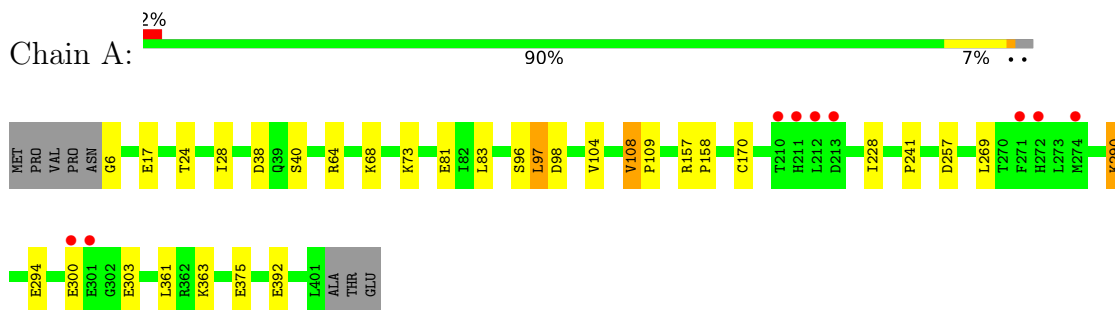
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	286	Total	O	0	0
			286	286		
2	B	258	Total	O	0	0
			258	258		
2	C	305	Total	O	0	0
			305	305		
2	D	250	Total	O	0	0
			250	250		

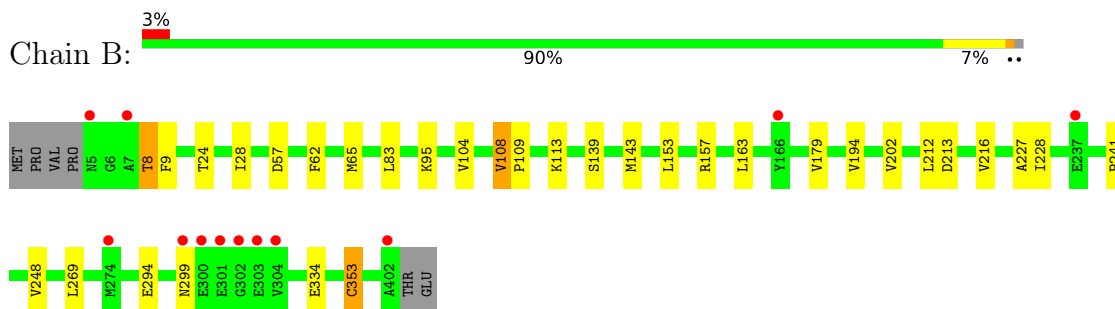
### 3 Residue-property plots [i](#)

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.

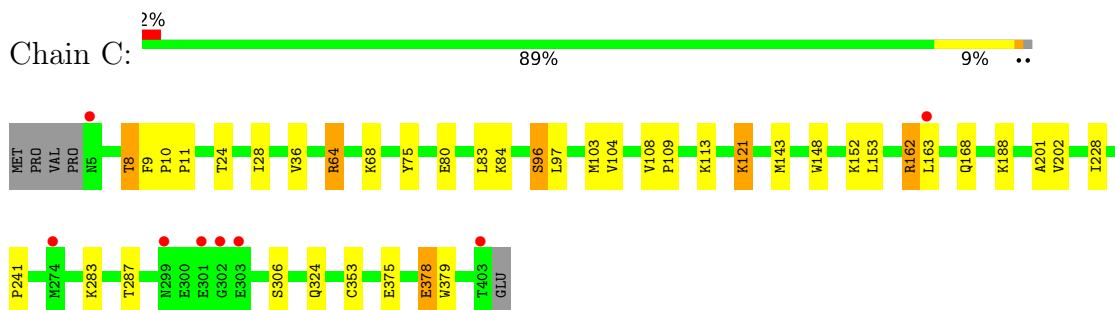
- Molecule 1: chalcone synthases



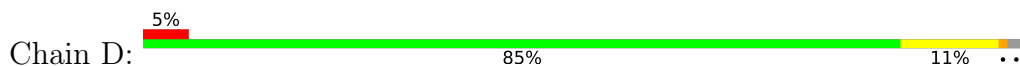
- Molecule 1: chalcone synthases

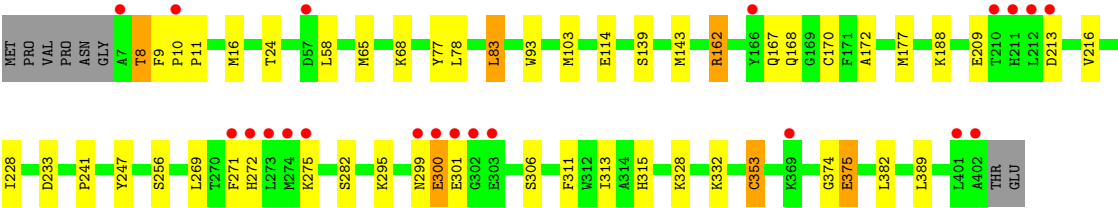


- Molecule 1: chalcone synthases



- Molecule 1: chalcone synthases





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	71.16Å 95.46Å 246.74Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	41.12 – 1.90 46.86 – 1.90	Depositor EDS
% Data completeness (in resolution range)	77.1 (41.12-1.90) 93.9 (46.86-1.90)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.36 (at 1.90Å)	Xtriage
Refinement program	PHENIX 1.10_2155	Depositor
R, $R_{free}$	0.181 , 0.208 0.182 , 0.210	Depositor DCC
$R_{free}$ test set	2000 reflections (1.52%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	18.6	Xtriage
Anisotropy	0.049	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 36.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	13325	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	8.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.33% 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

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	0.44	1/3115 (0.0%)	0.59	0/4221
1	B	0.46	1/3128 (0.0%)	0.56	0/4239
1	C	0.48	3/3135 (0.1%)	0.60	3/4249 (0.1%)
1	D	0.48	2/3116 (0.1%)	0.61	2/4223 (0.0%)
All	All	0.47	7/12494 (0.1%)	0.59	5/16932 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	D	0	2

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	378	GLU	CD-OE2	-5.52	1.19	1.25
1	D	353	CYS	CB-SG	-5.33	1.73	1.81
1	D	315	HIS	C-N	-5.22	1.24	1.34
1	B	353	CYS	CB-SG	-5.18	1.73	1.81
1	C	353	CYS	CB-SG	-5.13	1.73	1.81

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	162	ARG	NE-CZ-NH2	-7.67	116.46	120.30
1	D	275	LYS	N-CA-C	6.40	128.28	111.00
1	C	96	SER	CB-CA-C	5.91	121.32	110.10
1	C	162	ARG	NE-CZ-NH1	5.81	123.20	120.30
1	D	162	ARG	NE-CZ-NH2	-5.44	117.58	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	299	ASN	Peptide
1	D	300	GLU	Peptide

## 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	3048	0	3050	22	0
1	B	3061	0	3061	22	0
1	C	3068	0	3068	29	0
1	D	3049	0	3052	39	0
2	A	286	0	0	7	1
2	B	258	0	0	3	0
2	C	305	0	0	9	1
2	D	250	0	0	9	0
All	All	13325	0	12231	104	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 4.

The worst 5 of 104 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:D:272:HIS:HB3	2:D:726:HOH:O	1.47	1.11
1:B:157:ARG:NH2	2:B:502:HOH:O	2.01	0.92
1:D:177:MET:HA	1:D:177:MET:HE2	1.58	0.84
1:B:334:GLU:OE1	2:B:501:HOH:O	2.01	0.78
1:A:363:LYS:NZ	2:A:504:HOH:O	2.21	0.73

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)
2:A:720:HOH:O	2:C:778:HOH:O[1_655]	2.04	0.16



## 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	394/404 (98%)	383 (97%)	10 (2%)	1 (0%)	41	31
1	B	396/404 (98%)	386 (98%)	10 (2%)	0	100	100
1	C	397/404 (98%)	384 (97%)	13 (3%)	0	100	100
1	D	394/404 (98%)	383 (97%)	10 (2%)	1 (0%)	41	31
All	All	1581/1616 (98%)	1536 (97%)	43 (3%)	2 (0%)	51	42

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	97	LEU
1	D	375	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	331/338 (98%)	327 (99%)	4 (1%)	71	70
1	B	332/338 (98%)	326 (98%)	6 (2%)	59	55
1	C	333/338 (98%)	328 (98%)	5 (2%)	65	62
1	D	331/338 (98%)	325 (98%)	6 (2%)	59	55
All	All	1327/1352 (98%)	1306 (98%)	21 (2%)	62	60

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	306	SER
1	D	139	SER
1	D	353	CYS
1	D	301	GLU
1	D	83	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	167	GLN
1	C	324	GLN

### 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, 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	396/404 (98%)	-0.05	9 (2%) 60 63	3, 7, 17, 34	0
1	B	398/404 (98%)	0.02	12 (3%) 50 53	4, 8, 18, 36	0
1	C	399/404 (98%)	-0.03	8 (2%) 65 68	3, 7, 16, 33	0
1	D	396/404 (98%)	0.13	21 (5%) 26 29	4, 8, 19, 34	0
All	All	1589/1616 (98%)	0.02	50 (3%) 49 51	3, 7, 18, 36	0

The worst 5 of 50 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	300	GLU	5.4
1	D	301	GLU	5.2
1	A	212	LEU	5.2
1	D	402	ALA	4.5
1	A	274	MET	4.2

### 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.