



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 21, 2024 – 12:03 PM EST

PDB ID : 8TH2  
Title : Structure of the isoflavene-forming dirigent protein PsPTS2  
Authors : Smith, C.A.; Meng, Q.; Lewis, N.G.; Davin, L.B.  
Deposited on : 2023-07-13  
Resolution : 2.60 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

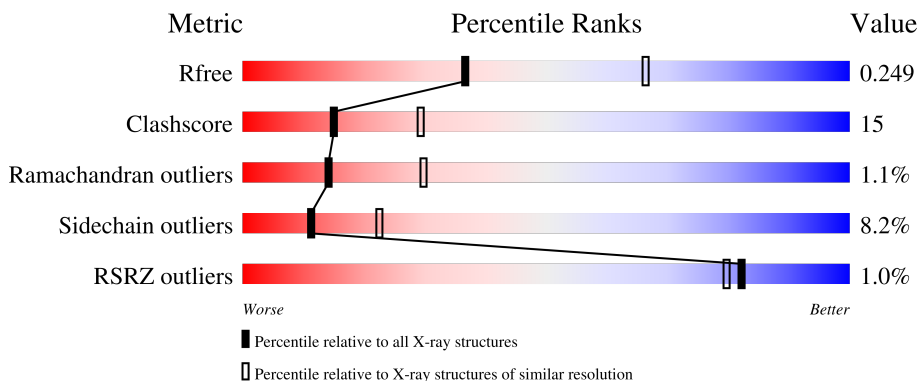
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.60 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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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	218	
1	B	218	
1	C	218	
1	D	218	
1	E	218	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	218	 48% 22% 26%

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 7778 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 Dirigent protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	160	1246	813	199	230	4	0	0	0
1	B	165	1282	833	207	238	4	0	0	0
1	C	173	1350	878	217	251	4	0	0	0
1	D	162	1259	820	201	234	4	0	0	0
1	E	165	1282	833	207	238	4	0	0	0
1	F	162	1272	828	206	234	4	0	1	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	179	GLU	LYS	engineered mutation	UNP A0A9D4XSX2
A	204	GLU	ALA	engineered mutation	UNP A0A9D4XSX2
A	214	ILE	MET	engineered mutation	UNP A0A9D4XSX2
B	179	GLU	LYS	engineered mutation	UNP A0A9D4XSX2
B	204	GLU	ALA	engineered mutation	UNP A0A9D4XSX2
B	214	ILE	MET	engineered mutation	UNP A0A9D4XSX2
C	179	GLU	LYS	engineered mutation	UNP A0A9D4XSX2
C	204	GLU	ALA	engineered mutation	UNP A0A9D4XSX2
C	214	ILE	MET	engineered mutation	UNP A0A9D4XSX2
D	179	GLU	LYS	engineered mutation	UNP A0A9D4XSX2
D	204	GLU	ALA	engineered mutation	UNP A0A9D4XSX2
D	214	ILE	MET	engineered mutation	UNP A0A9D4XSX2
E	179	GLU	LYS	engineered mutation	UNP A0A9D4XSX2
E	204	GLU	ALA	engineered mutation	UNP A0A9D4XSX2
E	214	ILE	MET	engineered mutation	UNP A0A9D4XSX2
F	179	GLU	LYS	engineered mutation	UNP A0A9D4XSX2
F	204	GLU	ALA	engineered mutation	UNP A0A9D4XSX2

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
F	214	ILE	MET	engineered mutation	UNP A0A9D4XSX2

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	12	Total O 12 12	0	0
2	B	11	Total O 11 11	0	0
2	C	21	Total O 21 21	0	0
2	D	14	Total O 14 14	0	0
2	E	13	Total O 13 13	0	0
2	F	16	Total O 16 16	0	0





## 4 Data and refinement statistics

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	98.14Å 162.18Å 232.24Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	90.40 – 2.60 90.40 – 2.50	Depositor EDS
% Data completeness (in resolution range)	69.3 (90.40-2.60) 63.1 (90.40-2.50)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.89 (at 2.51Å)	Xtrriage
Refinement program	PHENIX (1.20.1_4487: ???)	Depositor
R, $R_{free}$	0.196 , 0.252 0.201 , 0.249	Depositor DCC
$R_{free}$ test set	2046 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	38.6	Xtrriage
Anisotropy	0.119	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 50.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	7778	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 3.72% 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	0.51	0/1278	0.71	1/1732 (0.1%)
1	B	0.50	0/1315	0.72	0/1783
1	C	0.48	0/1387	0.68	1/1883 (0.1%)
1	D	0.50	0/1291	0.70	1/1750 (0.1%)
1	E	0.48	0/1315	0.69	1/1783 (0.1%)
1	F	0.51	0/1308	0.73	0/1772
All	All	0.50	0/7894	0.70	4/10703 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	68	LEU	CA-CB-CG	5.84	128.72	115.30
1	C	168	LEU	CA-CB-CG	5.43	127.79	115.30
1	A	35	LEU	CA-CB-CG	-5.17	103.41	115.30
1	E	68	LEU	CA-CB-CG	5.07	126.97	115.30

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	1246	0	1239	61	0
1	B	1282	0	1266	54	0
1	C	1350	0	1323	32	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	D	1259	0	1248	40	0
1	E	1282	0	1266	38	0
1	F	1272	0	1263	62	0
2	A	12	0	0	2	0
2	B	11	0	0	0	0
2	C	21	0	0	1	0
2	D	14	0	0	1	0
2	E	13	0	0	0	0
2	F	16	0	0	0	0
All	All	7778	0	7605	231	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 15.

The worst 5 of 231 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:165:ARG:HB3	1:F:165[B]:ARG:HH21	1.03	1.08
1:E:165:ARG:HG2	1:F:165[B]:ARG:NH1	1.81	0.96
1:F:74:SER:HB2	1:F:114:VAL:HG21	1.43	0.96
1:A:155:ILE:HG23	1:B:155:ILE:HG13	1.49	0.93
1:D:165:ARG:HB3	1:F:165[B]:ARG:NH2	1.84	0.93

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	156/218 (72%)	143 (92%)	12 (8%)	1 (1%)	25	47
1	B	161/218 (74%)	147 (91%)	12 (8%)	2 (1%)	13	27

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	171/218 (78%)	158 (92%)	12 (7%)	1 (1%)	25	47
1	D	158/218 (72%)	149 (94%)	6 (4%)	3 (2%)	8	15
1	E	161/218 (74%)	151 (94%)	10 (6%)	0	100	100
1	F	159/218 (73%)	144 (91%)	11 (7%)	4 (2%)	5	9
All	All	966/1308 (74%)	892 (92%)	63 (6%)	11 (1%)	14	30

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	F	114	VAL
1	A	116	GLU
1	D	56	PRO
1	D	114	VAL
1	F	69	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	137/190 (72%)	126 (92%)	11 (8%)	12	24
1	B	140/190 (74%)	130 (93%)	10 (7%)	14	29
1	C	147/190 (77%)	134 (91%)	13 (9%)	10	19
1	D	138/190 (73%)	124 (90%)	14 (10%)	7	14
1	E	140/190 (74%)	132 (94%)	8 (6%)	20	41
1	F	140/190 (74%)	126 (90%)	14 (10%)	7	14
All	All	842/1140 (74%)	772 (92%)	70 (8%)	11	22

5 of 70 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	54	THR
1	F	65	GLU

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type
1	F	165[A]	ARG
1	C	94	LEU
1	C	49	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	183	HIS
1	F	69	ASN
1	D	103	GLN
1	C	46	HIS
1	D	143	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, 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	160/218 (73%)	-0.15	3 (1%) 66 62	22, 35, 62, 89	0
1	B	165/218 (75%)	-0.16	2 (1%) 79 76	20, 35, 66, 84	0
1	C	173/218 (79%)	-0.26	0 100 100	21, 30, 48, 64	0
1	D	162/218 (74%)	-0.15	3 (1%) 66 62	22, 39, 71, 93	0
1	E	165/218 (75%)	-0.27	1 (0%) 89 88	21, 32, 59, 75	0
1	F	162/218 (74%)	-0.11	1 (0%) 89 88	20, 35, 72, 96	0
All	All	987/1308 (75%)	-0.19	10 (1%) 82 80	20, 34, 65, 96	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	53	VAL	5.0
1	A	56	PRO	3.4
1	A	53	VAL	3.0
1	D	54	THR	2.7
1	B	53	VAL	2.6

### 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.