



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 28, 2020 – 10:03 pm BST

PDB ID : 2JIJ  
Title : Crystal structure of the apo form of Chlamydomonas reinhardtii prolyl- 4  
hydroxylase type I  
Authors : Koski, M.K.; Hieta, R.; Bollner, C.; Kivirikko, K.I.; Myllyharju, J.; Wierenga,  
R.K.  
Deposited on : 2007-06-28  
Resolution : 2.90 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	:	2.11
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)	:	2.11

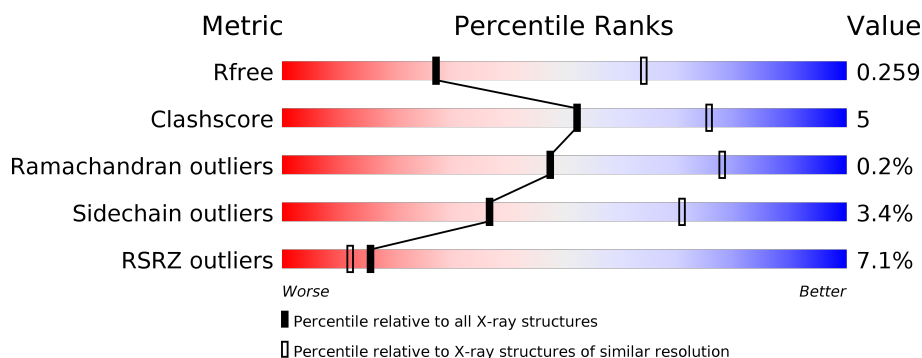
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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 on 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	233	<div> <div>4%</div> <div> <div></div> <div>71%</div> <div>15%</div> <div>•</div> <div>12%</div> </div> </div>
1	B	233	<div> <div>7%</div> <div> <div></div> <div>71%</div> <div>12%</div> <div>•</div> <div>15%</div> </div> </div>
1	C	233	<div> <div>7%</div> <div> <div></div> <div>73%</div> <div>7%</div> <div>20%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4673 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 PROLYL-4 HYDROXYLASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	206	Total	C	N	O	S	0	0	0
			1623	1034	278	302	9			
1	B	199	Total	C	N	O	S	0	0	0
			1561	997	263	292	9			
1	C	187	Total	C	N	O	S	0	0	0
			1456	928	250	269	9			

- 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		
2	C	1	Total	Cl	0	0
			1	1		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	14	Total	O	0	0
			14	14		
3	B	12	Total	O	0	0
			12	12		
3	C	5	Total	O	0	0
			5	5		

i

• Molecule 1: PROLYL-4 HYDROXYLASE



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	98.32Å 117.46Å 72.04Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	33.83 – 2.90 33.82 – 2.90	Depositor EDS
% Data completeness (in resolution range)	98.8 (33.83-2.90) 98.8 (33.82-2.90)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.83 (at 2.90Å)	Xtriage
Refinement program	REFMAC 5.3.0028	Depositor
R, $R_{free}$	0.212 , 0.261 0.212 , 0.259	Depositor DCC
$R_{free}$ test set	960 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	48.2	Xtriage
Anisotropy	0.088	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 57.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	4673	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.94% 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 ⓘ

### 5.1 Standard geometry ⓘ

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

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	0.81	1/1666 (0.1%)	0.91	7/2253 (0.3%)
1	B	0.87	3/1602 (0.2%)	0.79	0/2168
1	C	0.68	0/1493	0.72	0/2020
All	All	0.79	4/4761 (0.1%)	0.82	7/6441 (0.1%)

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	A	0	3
1	B	0	1
All	All	0	4

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	63	CYS	CB-SG	-5.91	1.72	1.81
1	B	173	GLU	CG-CD	5.46	1.60	1.51
1	A	68	GLU	CG-CD	5.36	1.59	1.51
1	B	33	GLU	CG-CD	5.05	1.59	1.51

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	173	GLU	N-CA-C	-9.34	85.78	111.00
1	A	233	LEU	N-CA-C	-7.86	89.77	111.00
1	A	39	ARG	CG-CD-NE	7.75	128.09	111.80
1	A	39	ARG	NE-CZ-NH2	-6.52	117.04	120.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	39	ARG	NE-CZ-NH1	6.00	123.30	120.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	172	VAL	Peptide
1	A	232	THR	Peptide
1	A	35	LYS	Peptide
1	B	189	GLY	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	1623	0	1581	25	0
1	B	1561	0	1515	17	0
1	C	1456	0	1425	10	1
2	A	1	0	0	0	0
2	C	1	0	0	0	0
3	A	14	0	0	0	0
3	B	12	0	0	0	0
3	C	5	0	0	0	0
All	All	4673	0	4521	50	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 5.

The worst 5 of 50 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:B:76:LYS:HA	1:B:96:THR:HG22	1.58	0.85
1:A:76:LYS:HA	1:A:96:THR:HG22	1.57	0.84
1:A:157:HIS:ND1	1:C:157:HIS:ND1	2.36	0.73
1:A:141:GLU:OE1	1:A:141:GLU:HA	1.90	0.72
1:B:37:GLU:OE1	1:B:37:GLU:HA	1.89	0.72

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)
1:C:39:ARG:NH1	1:C:184:GLU:OE2[2_555]	2.14	0.06

## 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	200/233 (86%)	186 (93%)	14 (7%)	0	100	100
1	B	191/233 (82%)	182 (95%)	8 (4%)	1 (0%)	29	61
1	C	179/233 (77%)	172 (96%)	7 (4%)	0	100	100
All	All	570/699 (82%)	540 (95%)	29 (5%)	1 (0%)	47	78

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	148	HIS

### 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	174/199 (87%)	169 (97%)	5 (3%)	42	76
1	B	168/199 (84%)	159 (95%)	9 (5%)	22	54
1	C	157/199 (79%)	154 (98%)	3 (2%)	57	84

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	499/597 (84%)	482 (97%)	17 (3%)	37 71

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	76	LYS
1	B	77	SER
1	B	190	ASP
1	B	75	VAL
1	C	105	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	227	HIS
1	B	227	HIS
1	C	160	GLN
1	C	221	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 2 ligands modelled in this entry, 2 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 ⓘ

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, 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	206/233 (88%)	0.21	10 (4%) 29 26	33, 38, 48, 61	0
1	B	199/233 (85%)	0.25	16 (8%) 12 9	33, 38, 52, 76	0
1	C	187/233 (80%)	0.41	16 (8%) 10 8	36, 38, 41, 63	0
All	All	592/699 (84%)	0.29	42 (7%) 16 12	33, 38, 48, 76	0

The worst 5 of 42 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	188	THR	4.2
1	C	29	VAL	4.1
1	A	35	LYS	4.0
1	B	144	TYR	4.0
1	B	77	SER	3.8

### 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. 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	C	2	1/1	0.84	0.14	60,60,60,60	0
2	CL	A	1	1/1	0.91	0.21	59,59,59,59	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.