



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 11:28 PM GMT

PDB ID : 1X9Y  
Title : The prostaphopain B structure  
Authors : Filipek, R.; Szczepanowski, R.; Sabat, A.; Potempa, J.; Bochtler, M.  
Deposited on : 2004-08-24  
Resolution : 2.50 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  
<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 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
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) : trunk26865

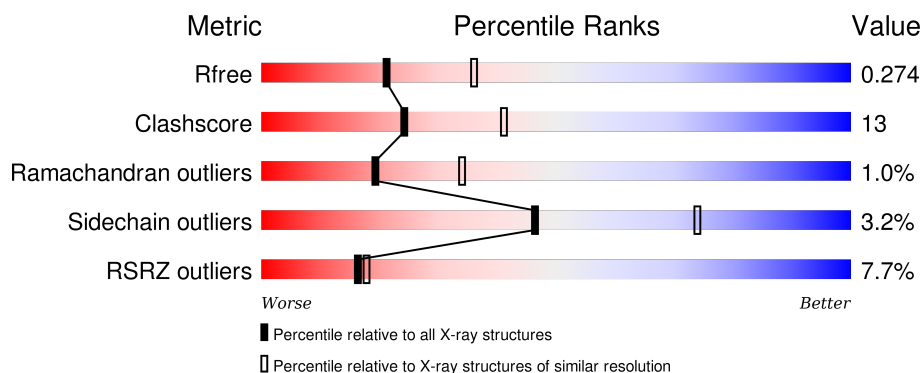
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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	3553 (2.50-2.50)
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)
RSRZ outliers	91569	3562 (2.50-2.50)

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. 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	367	<div> <div>6%</div> <div> <div></div> <div>71%</div> <div>22%</div> <div>• 6%</div> </div> </div>
1	B	367	<div> <div>6%</div> <div> <div></div> <div>71%</div> <div>21%</div> <div>• 6%</div> </div> </div>
1	C	367	<div> <div>9%</div> <div> <div></div> <div>72%</div> <div>20%</div> <div>•• 6%</div> </div> </div>
1	D	367	<div> <div>8%</div> <div> <div></div> <div>72%</div> <div>21%</div> <div>• 6%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 11355 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 cysteine proteinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	346	Total	C	N	O	S	0	0	0
			2778	1756	463	551	8			
1	B	346	Total	C	N	O	S	0	0	0
			2778	1756	463	551	8			
1	C	346	Total	C	N	O	S	0	0	0
			2778	1756	463	551	8			
1	D	346	Total	C	N	O	S	0	0	0
			2778	1756	463	551	8			

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	35	GLY	-	CLONING ARTIFACT	UNP Q70UQ8
A	36	SER	-	CLONING ARTIFACT	UNP Q70UQ8
A	394	LEU	-	CLONING ARTIFACT	UNP Q70UQ8
A	395	GLU	-	CLONING ARTIFACT	UNP Q70UQ8
A	396	VAL	-	CLONING ARTIFACT	UNP Q70UQ8
A	397	PRO	-	CLONING ARTIFACT	UNP Q70UQ8
A	398	ILE	-	CLONING ARTIFACT	UNP Q70UQ8
A	399	HIS	-	CLONING ARTIFACT	UNP Q70UQ8
A	400	ARG	-	CLONING ARTIFACT	UNP Q70UQ8
A	401	ASP	-	CLONING ARTIFACT	UNP Q70UQ8
B	35	GLY	-	CLONING ARTIFACT	UNP Q70UQ8
B	36	SER	-	CLONING ARTIFACT	UNP Q70UQ8
B	394	LEU	-	CLONING ARTIFACT	UNP Q70UQ8
B	395	GLU	-	CLONING ARTIFACT	UNP Q70UQ8
B	396	VAL	-	CLONING ARTIFACT	UNP Q70UQ8
B	397	PRO	-	CLONING ARTIFACT	UNP Q70UQ8
B	398	ILE	-	CLONING ARTIFACT	UNP Q70UQ8
B	399	HIS	-	CLONING ARTIFACT	UNP Q70UQ8
B	400	ARG	-	CLONING ARTIFACT	UNP Q70UQ8
B	401	ASP	-	CLONING ARTIFACT	UNP Q70UQ8
C	35	GLY	-	CLONING ARTIFACT	UNP Q70UQ8

*Continued on next page...*

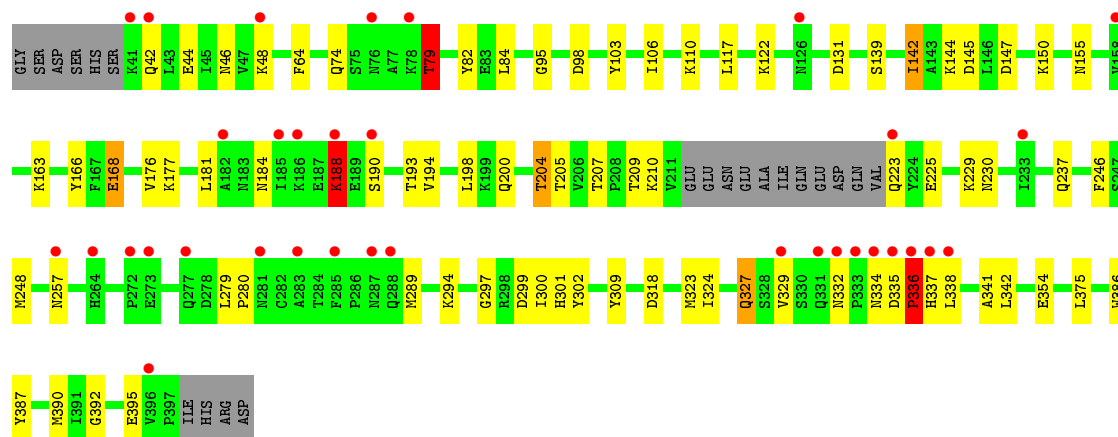
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	36	SER	-	CLONING ARTIFACT	UNP Q70UQ8
C	394	LEU	-	CLONING ARTIFACT	UNP Q70UQ8
C	395	GLU	-	CLONING ARTIFACT	UNP Q70UQ8
C	396	VAL	-	CLONING ARTIFACT	UNP Q70UQ8
C	397	PRO	-	CLONING ARTIFACT	UNP Q70UQ8
C	398	ILE	-	CLONING ARTIFACT	UNP Q70UQ8
C	399	HIS	-	CLONING ARTIFACT	UNP Q70UQ8
C	400	ARG	-	CLONING ARTIFACT	UNP Q70UQ8
C	401	ASP	-	CLONING ARTIFACT	UNP Q70UQ8
D	35	GLY	-	CLONING ARTIFACT	UNP Q70UQ8
D	36	SER	-	CLONING ARTIFACT	UNP Q70UQ8
D	394	LEU	-	CLONING ARTIFACT	UNP Q70UQ8
D	395	GLU	-	CLONING ARTIFACT	UNP Q70UQ8
D	396	VAL	-	CLONING ARTIFACT	UNP Q70UQ8
D	397	PRO	-	CLONING ARTIFACT	UNP Q70UQ8
D	398	ILE	-	CLONING ARTIFACT	UNP Q70UQ8
D	399	HIS	-	CLONING ARTIFACT	UNP Q70UQ8
D	400	ARG	-	CLONING ARTIFACT	UNP Q70UQ8
D	401	ASP	-	CLONING ARTIFACT	UNP Q70UQ8

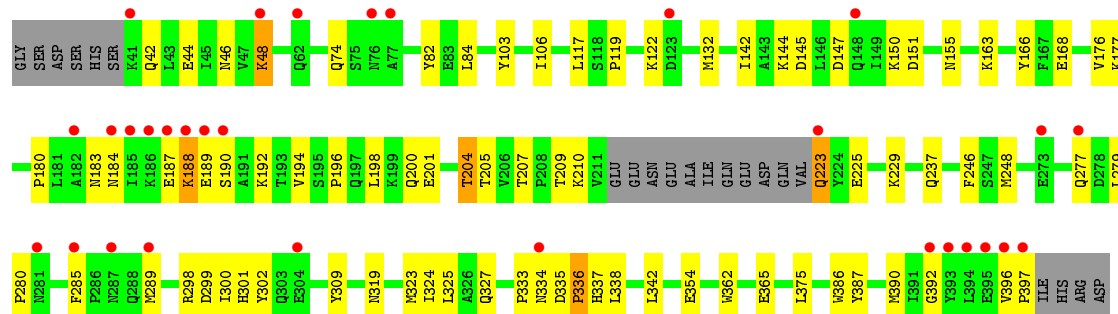
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	68	Total O 68 68	0	0
2	B	58	Total O 58 58	0	0
2	C	49	Total O 49 49	0	0
2	D	68	Total O 68 68	0	0





• Molecule 1: cysteine proteinase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	84.62Å 104.92Å 173.12Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.86 – 2.50 19.86 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.1 (19.86-2.50) 99.2 (19.86-2.50)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.22 (at 2.50Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, $R_{free}$	0.244 , 0.276 0.242 , 0.274	Depositor DCC
$R_{free}$ test set	2730 reflections (5.10%)	DCC
Wilson B-factor (Å <sup>2</sup> )	36.5	Xtriage
Anisotropy	0.721	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 39.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 53496 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	11355	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 15.07% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.39	0/2833	0.60	0/3831
1	B	0.39	0/2833	0.61	0/3831
1	C	0.38	0/2833	0.60	0/3831
1	D	0.39	0/2833	0.61	0/3831
All	All	0.39	0/11332	0.61	0/15324

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	2778	0	2724	79	0
1	B	2778	0	2724	83	0
1	C	2778	0	2724	75	0
1	D	2778	0	2724	72	0
2	A	68	0	0	10	0
2	B	58	0	0	6	0
2	C	49	0	0	6	0
2	D	68	0	0	7	0
All	All	11355	0	10896	287	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 13.



The worst 5 of 287 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:195:SER:HB3	1:B:197:GLN:HG3	1.31	1.09
1:C:122:LYS:HE2	1:C:329:VAL:HG13	1.46	0.96
1:B:308:SER:HB3	1:C:257:ASN:CG	1.86	0.95
1:A:204:THR:HG22	1:A:205:THR:H	1.33	0.94
1:B:332:ASN:HD22	1:B:335:ASP:HB3	1.34	0.92

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	342/367 (93%)	323 (94%)	16 (5%)	3 (1%)	21	37
1	B	342/367 (93%)	321 (94%)	18 (5%)	3 (1%)	21	37
1	C	342/367 (93%)	321 (94%)	15 (4%)	6 (2%)	11	18
1	D	342/367 (93%)	322 (94%)	18 (5%)	2 (1%)	30	50
All	All	1368/1468 (93%)	1287 (94%)	67 (5%)	14 (1%)	19	34

5 of 14 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	188	LYS
1	A	336	PRO
1	B	190	SER
1	C	190	SER
1	C	336	PRO

### 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	309/328 (94%)	301 (97%)	8 (3%)	54	81
1	B	309/328 (94%)	299 (97%)	10 (3%)	46	74
1	C	309/328 (94%)	297 (96%)	12 (4%)	39	66
1	D	309/328 (94%)	300 (97%)	9 (3%)	50	77
All	All	1236/1312 (94%)	1197 (97%)	39 (3%)	46	74

5 of 39 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	397	PRO
1	C	168	GLU
1	D	336	PRO
1	C	48	LYS
1	C	79	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 37 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	311	GLN
1	C	126	ASN
1	D	311	GLN
1	B	327	GLN
1	B	332	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](#)

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	346/367 (94%)	0.36	22 (6%) 23 25	19, 33, 64, 108	0
1	B	346/367 (94%)	0.32	21 (6%) 25 27	19, 34, 63, 100	0
1	C	346/367 (94%)	0.40	34 (9%) 10 10	20, 35, 65, 100	0
1	D	346/367 (94%)	0.38	30 (8%) 13 13	19, 35, 64, 94	0
All	All	1384/1468 (94%)	0.37	107 (7%) 16 18	19, 34, 65, 108	0

The worst 5 of 107 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	335	ASP	13.4
1	B	333	PRO	12.6
1	B	334	ASN	12.1
1	A	332	ASN	12.0
1	A	333	PRO	11.1

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.