



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 07:20 pm BST

PDB ID : 5FPW  
Title : proCathepsin B S9 from Trypanosoma congolense  
Authors : Sevajol, M.; Biteau, N.; Baltz, T.; Franzetti, B.; Vellieux, F.M.D.  
Deposited on : 2015-12-03  
Resolution : 2.10 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

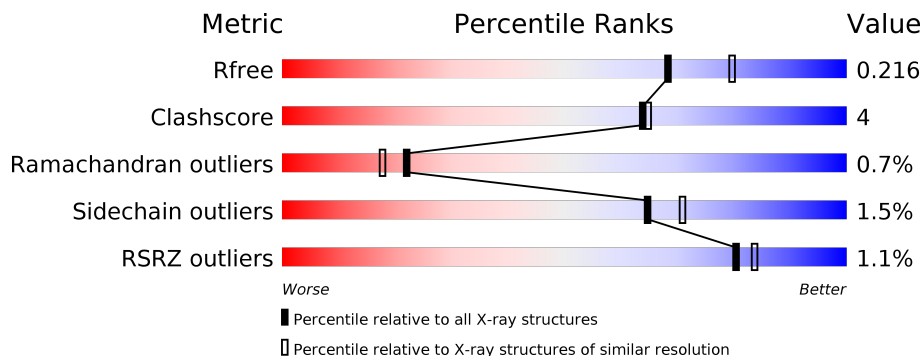
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	321	 82% 9% 9%
1	B	321	 80% 10% 8%
1	C	321	 84% 7% 9%

## 2 Entry composition i

There are 2 unique types of molecules in this entry. The entry contains 7566 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 PRO CATHEPSIN B S9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	293	Total 2302	C 1474	N 389	O 424	S 15	0	0	1
1	B	294	Total 2308	C 1477	N 390	O 426	S 15	5	0	1
1	C	293	Total 2300	C 1473	N 388	O 424	S 15	5	0	1

There are 33 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	ALA	-	expression tag	UNP B2C331
A	-2	GLU	-	expression tag	UNP B2C331
A	-1	PHE	-	expression tag	UNP B2C331
A	0	HIS	-	expression tag	UNP B2C331
A	1	HIS	-	expression tag	UNP B2C331
A	2	HIS	-	expression tag	UNP B2C331
A	3	HIS	-	expression tag	UNP B2C331
A	4	HIS	-	expression tag	UNP B2C331
A	5	HIS	-	expression tag	UNP B2C331
A	192	GLN	ASN	engineered mutation	UNP B2C331
A	208	SER	ASN	engineered mutation	UNP B2C331
B	997	ALA	-	expression tag	UNP B2C331
B	998	GLU	-	expression tag	UNP B2C331
B	999	PHE	-	expression tag	UNP B2C331
B	1000	HIS	-	expression tag	UNP B2C331
B	1001	HIS	-	expression tag	UNP B2C331
B	1002	HIS	-	expression tag	UNP B2C331
B	1003	HIS	-	expression tag	UNP B2C331
B	1004	HIS	-	expression tag	UNP B2C331
B	1005	HIS	-	expression tag	UNP B2C331
B	1192	GLN	ASN	engineered mutation	UNP B2C331
B	1208	SER	ASN	engineered mutation	UNP B2C331
C	1997	ALA	-	expression tag	UNP B2C331

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	1998	GLU	-	expression tag	UNP B2C331
C	1999	PHE	-	expression tag	UNP B2C331
C	2000	HIS	-	expression tag	UNP B2C331
C	2001	HIS	-	expression tag	UNP B2C331
C	2002	HIS	-	expression tag	UNP B2C331
C	2003	HIS	-	expression tag	UNP B2C331
C	2004	HIS	-	expression tag	UNP B2C331
C	2005	HIS	-	expression tag	UNP B2C331
C	2192	GLN	ASN	engineered mutation	UNP B2C331
C	2208	SER	ASN	engineered mutation	UNP B2C331

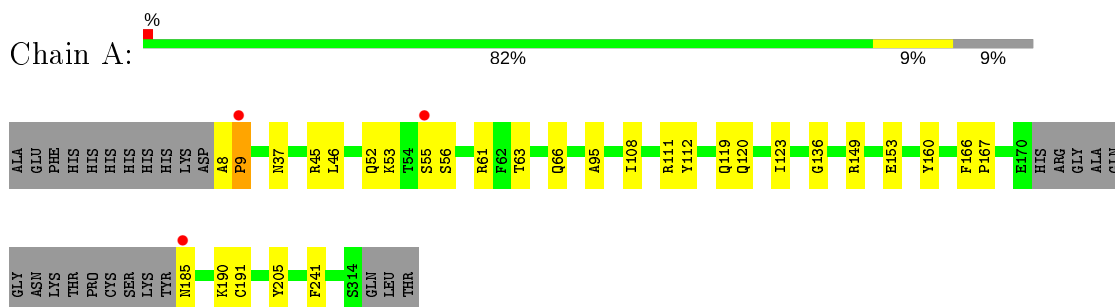
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	239	Total O 239 239	0	0
2	B	220	Total O 220 220	0	0
2	C	197	Total O 197 197	0	0

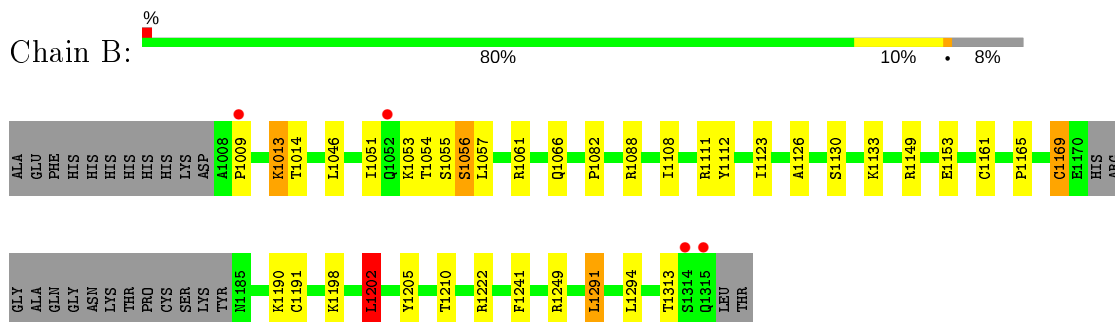
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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.

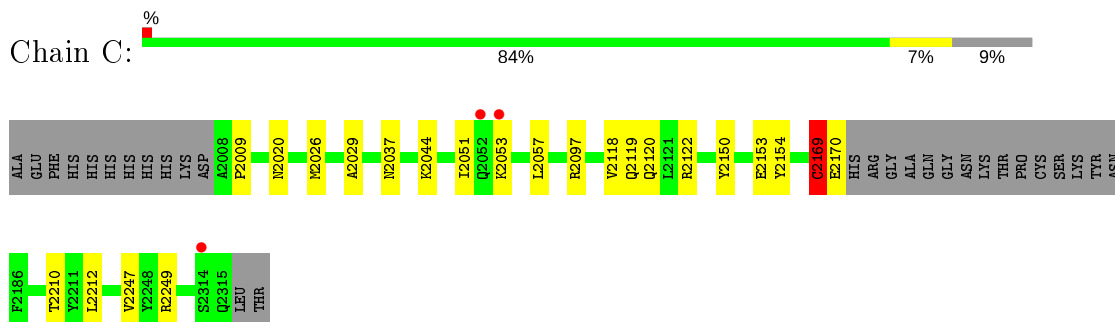
- Molecule 1: PRO CATHEPSIN B S9



- Molecule 1: PRO CATHEPSIN B S9



- Molecule 1: PRO CATHEPSIN B S9



## 4 Data and refinement statistics i

Property	Value	Source
Space group	C 2 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	128.79Å 223.22Å 102.89Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.03 – 2.10 39.03 – 2.10	Depositor EDS
% Data completeness (in resolution range)	98.5 (39.03-2.10) 98.3 (39.03-2.10)	Depositor EDS
$R_{merge}$	0.01	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.16 (at 2.10Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.179 , 0.216 0.179 , 0.216	Depositor DCC
$R_{free}$ test set	4263 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	32.5	Xtriage
Anisotropy	0.374	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 41.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.467 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.477 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	7566	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.05% 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	1/2369 (0.0%)	0.58	1/3221 (0.0%)
1	B	0.57	1/2375 (0.0%)	0.63	3/3229 (0.1%)
1	C	0.53	1/2367 (0.0%)	0.59	0/3218
All	All	0.54	3/7111 (0.0%)	0.60	4/9668 (0.0%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	1009	PRO	N-CD	-14.73	1.27	1.47
1	A	9	PRO	N-CD	-12.79	1.29	1.47
1	C	2009	PRO	N-CD	-11.69	1.31	1.47

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	111	ARG	NE-CZ-NH1	-6.26	117.17	120.30
1	B	1202	LEU	CA-CB-CG	5.57	128.10	115.30
1	B	1291	LEU	CA-CB-CG	5.27	127.41	115.30
1	B	1111	ARG	NE-CZ-NH1	-5.14	117.73	120.30

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	2302	0	2199	20	0
1	B	2308	0	2204	23	0
1	C	2300	0	2198	16	1
2	A	239	0	0	6	2
2	B	220	0	0	6	0
2	C	197	0	0	2	2
All	All	7566	0	6601	57	4

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 57 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:A:95:ALA:HB3	1:A:185:ASN:HB2	1.49	0.93
1:B:1153:GLU:OE1	2:B:4138:HOH:O	1.99	0.81
1:C:2020:ASN:HD21	1:C:2029:ALA:H	1.28	0.79
1:C:2153:GLU:HG3	1:C:2154:TYR:CD2	2.25	0.72
1:A:63:THR:H	1:A:66:GLN:HE21	1.37	0.70

All (4) 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:4056:HOH:O	2:A:4056:HOH:O[2_655]	1.94	0.26
1:C:2053:LYS:NZ	2:C:4178:HOH:O[6_555]	2.05	0.15
2:C:4109:HOH:O	2:C:4109:HOH:O[6_555]	2.13	0.07
2:A:4050:HOH:O	2:A:4058:HOH:O[2_655]	2.17	0.03

## 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	289/321 (90%)	274 (95%)	13 (4%)	2 (1%)	22	18
1	B	290/321 (90%)	278 (96%)	9 (3%)	3 (1%)	15	11
1	C	289/321 (90%)	278 (96%)	10 (4%)	1 (0%)	41	41
All	All	868/963 (90%)	830 (96%)	32 (4%)	6 (1%)	22	18

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	56	SER
1	C	2169	CYS
1	B	1055	SER
1	B	1056	SER
1	A	55	SER

### 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	242/267 (91%)	241 (100%)	1 (0%)	91	94
1	B	243/267 (91%)	237 (98%)	6 (2%)	47	52
1	C	242/267 (91%)	238 (98%)	4 (2%)	60	67
All	All	727/801 (91%)	716 (98%)	11 (2%)	65	71

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	1202	LEU
1	B	1291	LEU
1	C	2118	VAL
1	B	1169	CYS
1	C	2051	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	66	GLN
1	B	1119	GLN
1	C	2020	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	293/321 (91%)	-0.40	3 (1%) 82 85	23, 34, 60, 99	0
1	B	294/321 (91%)	-0.32	4 (1%) 75 78	23, 34, 60, 119	1 (0%)
1	C	293/321 (91%)	-0.41	3 (1%) 82 85	23, 34, 59, 110	1 (0%)
All	All	880/963 (91%)	-0.37	10 (1%) 80 84	23, 34, 60, 119	2 (0%)

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1314	SER	6.2
1	A	185	ASN	4.0
1	B	1052	GLN	3.1
1	A	9	PRO	3.0
1	C	2052	GLN	2.8

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.