



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 12:20 pm GMT

PDB ID : 1CPJ  
Title : CRYSTAL STRUCTURES OF RECOMBINANT RAT CATHEPSIN B  
AND A CATHEPSIN B-INHIBITOR COMPLEX: IMPLICATIONS FOR  
STRUCTURE-BASED INHIBITOR DESIGN  
Authors : Huber, C.P.; Jia, Z.  
Deposited on : 1995-07-17  
Resolution : 2.20 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28949

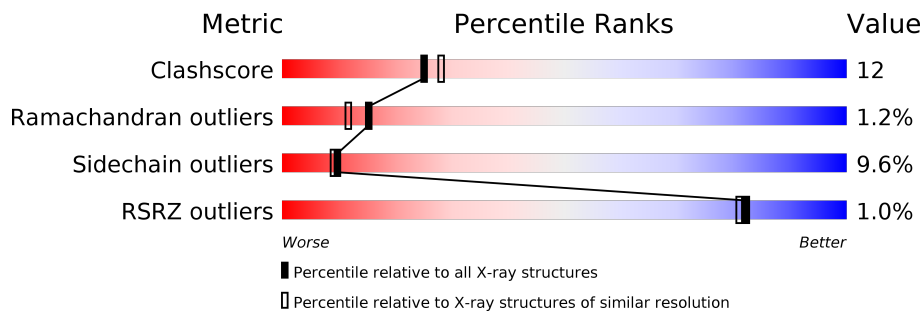
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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))
Clashscore	112137	4730 (2.20-2.20)
Ramachandran outliers	110173	4656 (2.20-2.20)
Sidechain outliers	110143	4657 (2.20-2.20)
RSRZ outliers	101464	4033 (2.20-2.20)

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	260	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red, orange, yellow, green, grey);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>%</span> <span>66%</span> <span>26%</span> <span>5%</span> <span>.</span> </div> </div>
1	B	260	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red, orange, yellow, green, grey);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>%</span> <span>65%</span> <span>29%</span> <span>.</span> <span>.</span> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4058 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 CATHEPSIN B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	253	Total	C	N	O	S	17	0	0
			1934	1210	332	374	18			
1	B	253	Total	C	N	O	S	20	0	0
			1934	1210	332	374	18			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	115	ALA	SER	ENGINEERED	UNP P00787
A	223	ALA	VAL	CLONING ARTIFACT	UNP P00787
B	115	ALA	SER	ENGINEERED	UNP P00787
B	223	ALA	VAL	CLONING ARTIFACT	UNP P00787

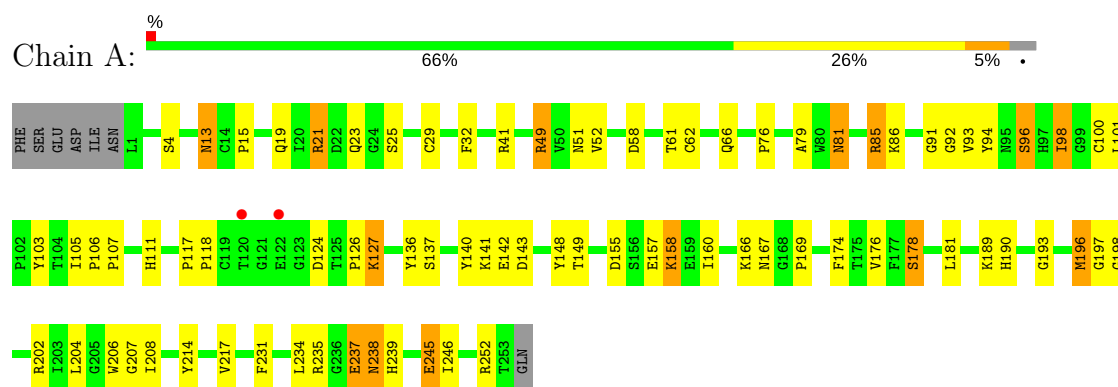
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	96	Total	O	0	0
			96	96		
2	B	94	Total	O	0	0
			94	94		

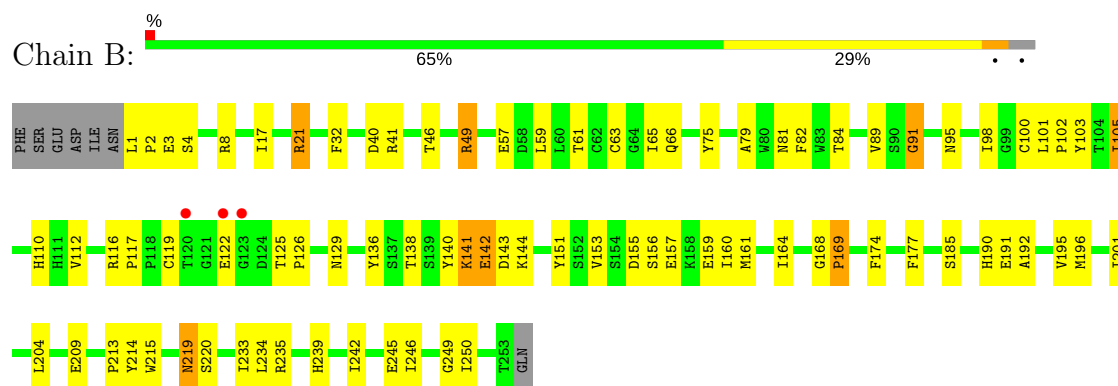
### 3 Residue-property plots [i](#)

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: CATHEPSIN B



#### • Molecule 1: CATHEPSIN B



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	47.09Å 90.25Å 62.10Å 90.00° 97.29° 90.00°	Depositor
Resolution (Å)	8.00 – 2.20 7.99 – 2.20	Depositor EDS
% Data completeness (in resolution range)	93.9 (8.00-2.20) 91.1 (7.99-2.20)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$	-	Xtriage
Refinement program	X-PLOR 3.0	Depositor
R, $R_{free}$	0.175 , (Not available) 0.173 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	11.2	Xtriage
Anisotropy	0.209	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.23 , 82.1	EDS
L-test for twinning <sup>1</sup>	$\langle  L  \rangle = 0.33$ , $\langle L^2 \rangle = 0.17$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	4058	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.96% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</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	0.86	3/1989 (0.2%)	0.81	3/2701 (0.1%)
1	B	0.65	1/1989 (0.1%)	0.67	0/2701
All	All	0.76	4/3978 (0.1%)	0.74	3/5402 (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	9
1	B	0	4
All	All	0	13

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	127	LYS	CB-CG	-26.36	0.81	1.52
1	B	122	GLU	CD-OE2	7.61	1.34	1.25
1	A	127	LYS	CD-CE	5.16	1.64	1.51
1	A	158	LYS	CD-CE	5.16	1.64	1.51

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	127	LYS	CA-CB-CG	15.93	148.45	113.40
1	A	142	GLU	OE1-CD-OE2	-12.89	107.83	123.30
1	A	142	GLU	CG-CD-OE2	7.47	133.24	118.30

There are no chirality outliers.

5 of 13 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	103	TYR	Sidechain
1	A	32	PHE	Sidechain
1	A	85	ARG	Sidechain
1	A	94	TYR	Sidechain
1	A	96	SER	Mainchain

## 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	1934	0	1783	39	0
1	B	1934	0	1783	50	0
2	A	96	0	0	6	0
2	B	94	0	0	3	0
All	All	4058	0	3566	89	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 12.

The worst 5 of 89 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:49:ARG:HD2	1:A:49:ARG:H	1.31	0.95
1:A:49:ARG:HD2	1:A:49:ARG:N	1.90	0.82
1:A:176:VAL:HG12	1:A:196:MET:HG3	1.61	0.81
1:A:4:SER:HB3	1:A:207:GLY:HA2	1.73	0.71
1:A:100:CYS:HB2	1:A:143:ASP:OD2	1.95	0.66

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	251/260 (96%)	211 (84%)	36 (14%)	4 (2%)	11	8
1	B	251/260 (96%)	226 (90%)	23 (9%)	2 (1%)	22	21
All	All	502/520 (96%)	437 (87%)	59 (12%)	6 (1%)	15	12

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	155	ASP
1	A	238	ASN
1	B	155	ASP
1	A	178	SER
1	B	169	PRO

### 5.3.2 Protein sidechains ⓘ

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	204/211 (97%)	184 (90%)	20 (10%)	9	8
1	B	204/211 (97%)	185 (91%)	19 (9%)	10	10
All	All	408/422 (97%)	369 (90%)	39 (10%)	10	9

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

Mol	Chain	Res	Type
1	A	237	GLU
1	B	3	GLU
1	B	204	LEU
1	A	238	ASN
1	A	245	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	238	ASN
1	A	239	HIS
1	B	219	ASN
1	A	190	HIS
1	B	190	HIS

### 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	253/260 (97%)	-0.26	2 (0%) 86 85	2, 19, 44, 75	4 (1%)
1	B	253/260 (97%)	-0.19	3 (1%) 79 77	2, 20, 44, 68	5 (1%)
All	All	506/520 (97%)	-0.22	5 (0%) 82 81	2, 19, 44, 75	9 (1%)

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	120	THR	3.2
1	B	123	GLY	3.1
1	B	122	GLU	2.9
1	A	122	GLU	2.3
1	B	120	THR	2.1

### 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.