Full wwPDB X-ray Structure Validation Report (i)

May 13, 2020 – 07:28 am BST

PDB ID	:	2IPP
Title	:	Crystal Structure of the tetragonal form of human liver cathepsin B
Authors	:	Huber, C.P.; Campbell, R.L.; Hasnain, S.; Hirama, T.; To, R.
Deposited on	:	2006-10-12
Resolution	:	2.15 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at *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 (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

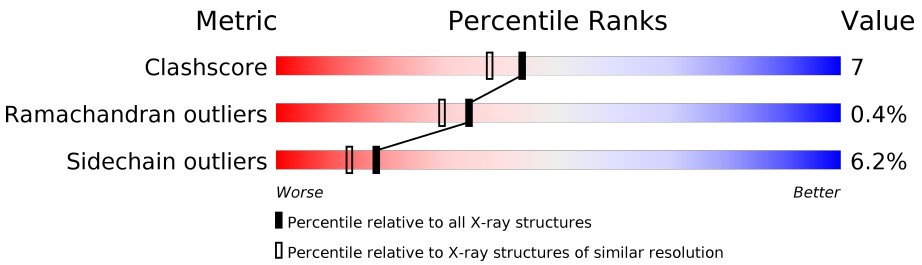
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.15 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	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)

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 >=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 <=5%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	47	74% 23% •
2	B	205	73% 23% • •

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	PYS	A	255	-	X	-	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2097 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.

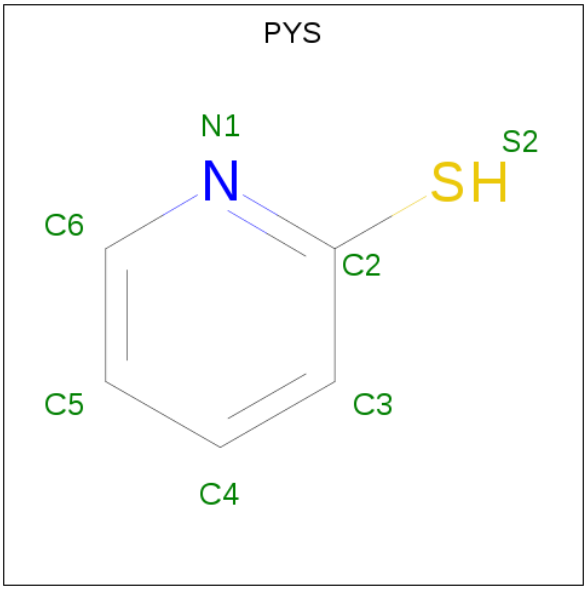
- Molecule 1 is a protein called Cathepsin B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	47	364	225	65	70	4	0	0	0

- Molecule 2 is a protein called Cathepsin B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	205	1574	988	265	307	14	0	0	0

- Molecule 3 is 2-PYRIDINETHIOL (three-letter code: PYS) (formula: C_5H_5NS).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	S		
3	A	1	7	5	1	1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	24	24	24	0	0
			Total	O		
4	B	128	128	128	0	0

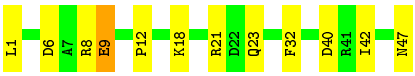
3 Residue-property plots (i)

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. 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. 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.

Note EDS was not executed.

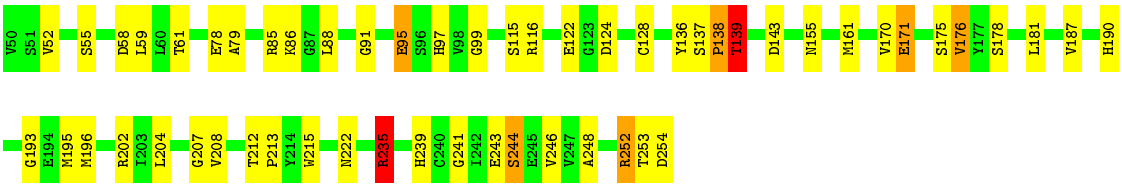
- Molecule 1: Cathepsin B

Chain A:



- Molecule 2: Cathepsin B

Chain B:



4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 43	Depositor
Cell constants a, b, c, α, β, γ	85.58Å 85.58Å 34.39Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	8.00 – 2.15	Depositor
% Data completeness (in resolution range)	(Not available) (8.00-2.15)	Depositor
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	PROLSQ	Depositor
R, R_{free}	0.157 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2097	wwPDB-VP
Average B, all atoms (Å^2)	9.0	wwPDB-VP

5 Model quality (i)

5.1 Standard geometry (i)

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

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	0.85	0/373	1.51	5/505 (1.0%)
2	B	0.85	0/1622	1.71	19/2203 (0.9%)
All	All	0.85	0/1995	1.68	24/2708 (0.9%)

There are no bond length outliers.

All (24) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	116	ARG	NE-CZ-NH1	18.66	129.63	120.30
2	B	235	ARG	CD-NE-CZ	17.85	148.59	123.60
2	B	116	ARG	NE-CZ-NH2	-15.16	112.72	120.30
2	B	116	ARG	CD-NE-CZ	14.67	144.14	123.60
2	B	202	ARG	NE-CZ-NH2	13.20	126.90	120.30
2	B	85	ARG	NE-CZ-NH1	8.00	124.30	120.30
2	B	124	ASP	CB-CG-OD2	-7.78	111.30	118.30
1	A	40	ASP	CB-CG-OD1	7.77	125.29	118.30
2	B	252	ARG	NE-CZ-NH2	-7.28	116.66	120.30
2	B	139	THR	CA-CB-CG2	7.06	122.28	112.40
2	B	122	GLU	CA-CB-CG	7.04	128.88	113.40
2	B	58	ASP	CB-CG-OD1	6.97	124.58	118.30
2	B	122	GLU	N-CA-CB	6.72	122.69	110.60
1	A	21	ARG	NE-CZ-NH2	-6.50	117.05	120.30
2	B	88	LEU	CA-CB-CG	6.13	129.40	115.30
2	B	95	GLU	CA-CB-CG	5.97	126.54	113.40
1	A	8	ARG	NE-CZ-NH1	5.80	123.20	120.30
2	B	171	GLU	CG-CD-OE2	-5.74	106.82	118.30
2	B	143	ASP	CB-CG-OD1	5.66	123.40	118.30
2	B	254	ASP	CB-CG-OD1	5.64	123.37	118.30
2	B	202	ARG	NE-CZ-NH1	-5.62	117.49	120.30
1	A	6	ASP	CB-CG-OD1	5.47	123.22	118.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	97	HIS	CA-CB-CG	-5.32	104.56	113.60
1	A	9	GLU	OE1-CD-OE2	5.26	129.61	123.30

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	364	0	339	6	0
2	B	1574	0	1444	21	1
3	A	7	0	4	0	0
4	A	24	0	0	2	0
4	B	128	0	0	2	1
All	All	2097	0	1787	25	2

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 7.

All (25) 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)
2:B:59:LEU:HD21	2:B:79:ALA:HB1	1.71	0.72
2:B:190:HIS:H	2:B:239:HIS:HE1	1.41	0.69
1:A:42:ILE:HG21	2:B:52:VAL:HG12	1.82	0.61
2:B:190:HIS:H	2:B:239:HIS:CE1	2.21	0.56
2:B:176:VAL:HG22	2:B:196:MET:HG3	1.88	0.56
2:B:178:SER:HB3	2:B:193:GLY:HA3	1.88	0.56
2:B:253:THR:N	4:B:368:HOH:O	2.41	0.53
2:B:139:THR:HG23	4:B:545:HOH:O	2.11	0.50
1:A:47:ASN:ND2	4:A:573:HOH:O	2.44	0.50
2:B:170:VAL:HG12	2:B:248:ALA:HB2	1.95	0.48
1:A:9:GLU:O	1:A:12:PRO:HD3	2.16	0.45
2:B:207:GLY:O	2:B:235:ARG:NH2	2.50	0.45
2:B:241:GLY:O	2:B:244:SER:HB2	2.18	0.44

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:181:LEU:HD11	2:B:196:MET:CE	2.48	0.43
2:B:208:VAL:HA	2:B:213:PRO:HA	2.00	0.43
2:B:99:GLY:HA2	2:B:136:TYR:CE1	2.54	0.43
2:B:55:SER:HB2	2:B:91:GLY:HA3	2.00	0.43
1:A:23:GLN:HA	1:A:32:PHE:CZ	2.55	0.42
2:B:171:GLU:O	2:B:246:VAL:HA	2.19	0.42
2:B:187:VAL:HG11	2:B:212:THR:HG22	2.01	0.41
1:A:1:LEU:HD13	2:B:161:MET:HB3	2.02	0.41
1:A:18:LYS:HE3	4:A:571:HOH:O	2.21	0.40
2:B:155:ASN:HA	2:B:243:GLU:O	2.21	0.40
2:B:61:THR:O	2:B:128:CYS:HB2	2.21	0.40
2:B:215:TRP:CD1	2:B:235:ARG:HB2	2.55	0.40

All (2) 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:B:78:GLU:OE1	2:B:95:GLU:OE2[1_556]	2.09	0.11
4:B:423:HOH:O	4:B:470:HOH:O[1_556]	2.18	0.02

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	45/47 (96%)	45 (100%)	0	0	100	100
2	B	203/205 (99%)	185 (91%)	17 (8%)	1 (0%)	29	22
All	All	248/252 (98%)	230 (93%)	17 (7%)	1 (0%)	34	29

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	138	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	39/39 (100%)	39 (100%)	0	100 100
2	B	169/169 (100%)	156 (92%)	13 (8%)	13 8
All	All	208/208 (100%)	195 (94%)	13 (6%)	18 13

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	86	LYS
2	B	115	SER
2	B	137	SER
2	B	138	PRO
2	B	139	THR
2	B	175	SER
2	B	176	VAL
2	B	195	MET
2	B	204	LEU
2	B	222	ASN
2	B	235	ARG
2	B	244	SER
2	B	252	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	47	ASN
2	B	167	ASN
2	B	190	HIS
2	B	237	GLN
2	B	239	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)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	PYS	A	255	1	6,7,7	2.54	2 (33%)	6,8,8	2.68	5 (83%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	PYS	A	255	1	-	-	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	255	PYS	C2-N1	5.02	1.39	1.33
3	A	255	PYS	C5-C4	2.42	1.44	1.38

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	255	PYS	C6-N1-C2	4.29	124.60	117.10
3	A	255	PYS	C3-C2-N1	-2.53	118.38	122.93
3	A	255	PYS	C5-C4-C3	2.45	123.92	120.19
3	A	255	PYS	C5-C6-N1	-2.17	119.88	123.43
3	A	255	PYS	C4-C5-C6	-2.03	114.69	118.62

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 (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)

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.