Full wwPDB X-ray Structure Validation Report (i)

Jul 26, 2023 – 03:07 AM EDT

PDB ID	:	1AYU
Title	:	CRYSTAL STRUCTURE OF CYSTEINE PROTEASE HUMAN CATHEPSIN K IN COMPLEX WITH A COVALENT SYMMETRIC BISCARBOHYDRAZIDE INHIBITOR
Authors	:	Zhao, B.; Smith, W.W.; Janson, C.A.; Abdel-Meguid, S.S.
Deposited on	:	1997-11-10
Resolution	:	2.20 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

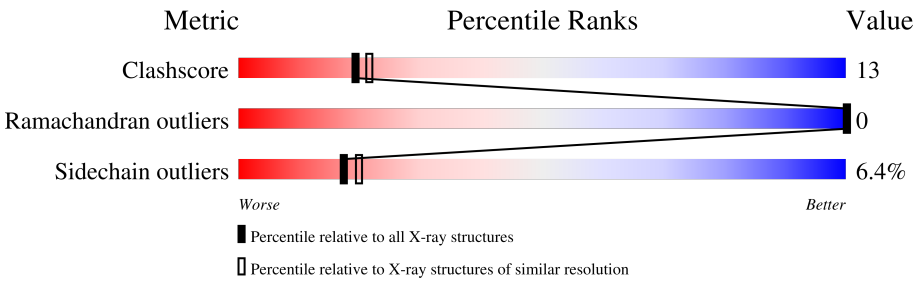
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.34

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 2.20 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	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (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 of 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	215	75% 23% •

2 Entry composition (i)

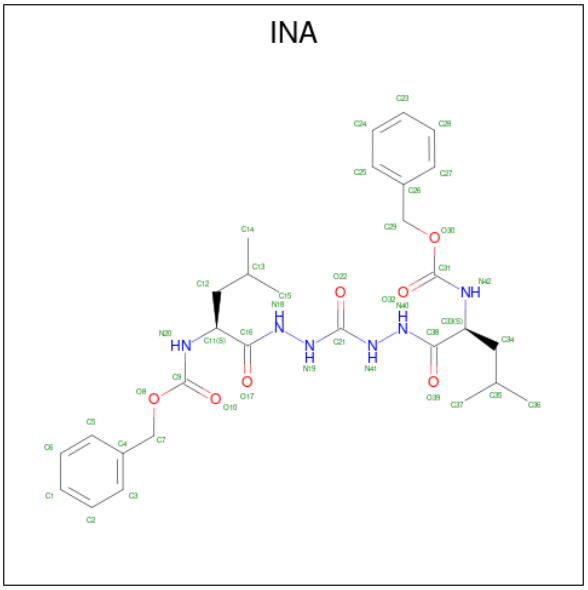
There are 3 unique types of molecules in this entry. The entry contains 1757 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 K.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	215	1649	1030	291	316	12	0	0	0

- Molecule 2 is 1,5-BIS(N-BENZYLOXYCARBONYL-L-LEUCINYL)CARBOHYDRAZIDE (three-letter code: INA) (formula: $C_{29}H_{40}N_6O_7$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	42	29	6	7	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	66	66	66	0	0

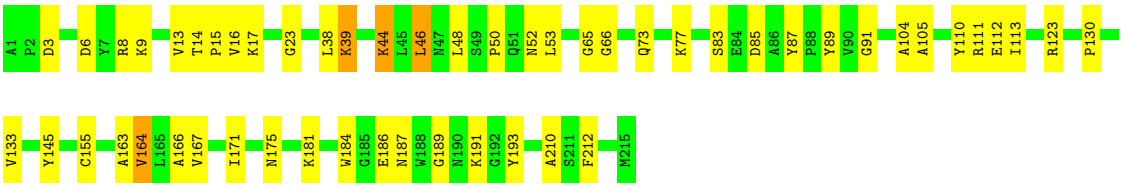
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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 K

Chain A:



4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	38.33Å 51.30Å 103.86Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	6.00 – 2.20	Depositor
% Data completeness (in resolution range)	78.5 (6.00-2.20)	Depositor
R_{merge}	0.11	Depositor
R_{sym}	0.11	Depositor
Refinement program	X-PLOR 3.1	Depositor
R, R_{free}	0.208 , 0.318	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	1757	wwPDB-VP
Average B, all atoms (Å^2)	15.0	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: INA

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	0.42	0/1684	0.62	1/2270 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	46	LEU	CA-CB-CG	5.93	128.94	115.30

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	1649	0	1595	41	0
2	A	42	0	40	8	0
3	A	66	0	0	4	0
All	All	1757	0	1635	43	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.

All (43) 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:66:GLY:HA2	2:A:300:INA:H25	1.42	0.98
1:A:23:GLY:HA3	3:A:309:HOH:O	1.84	0.77
1:A:16:VAL:HB	1:A:186:GLU:HG2	1.69	0.75
1:A:111:ARG:HD2	1:A:212:PHE:CE2	2.27	0.69
1:A:133:VAL:HG23	1:A:164:VAL:HG12	1.76	0.67
1:A:111:ARG:NH1	1:A:113:ILE:HA	2.11	0.66
1:A:65:GLY:C	2:A:300:INA:H292	2.18	0.64
2:A:300:INA:H122	2:A:300:INA:O10	1.98	0.63
1:A:66:GLY:N	2:A:300:INA:H292	2.16	0.60
1:A:111:ARG:HH12	1:A:113:ILE:HG13	1.69	0.57
1:A:66:GLY:O	2:A:300:INA:H35	2.04	0.57
1:A:73:GLN:HG3	1:A:110:TYR:HE2	1.70	0.57
1:A:13:VAL:HG21	1:A:167:VAL:HG21	1.87	0.57
1:A:191:LYS:HB2	1:A:191:LYS:NZ	2.18	0.57
1:A:111:ARG:HH11	1:A:113:ILE:HD12	1.71	0.56
1:A:15:PRO:O	1:A:17:LYS:HE3	2.07	0.54
1:A:145:TYR:O	1:A:189:GLY:HA2	2.10	0.50
1:A:191:LYS:HB2	1:A:191:LYS:HZ3	1.76	0.49
1:A:14:THR:HB	1:A:15:PRO:HD2	1.94	0.49
1:A:83:SER:HB3	1:A:104:ALA:HB2	1.95	0.48
1:A:164:VAL:HG22	1:A:181:LYS:O	2.13	0.48
1:A:133:VAL:HG12	1:A:210:ALA:HB2	1.96	0.48
1:A:13:VAL:CG2	1:A:167:VAL:HG21	2.44	0.47
1:A:73:GLN:CG	1:A:110:TYR:HE2	2.27	0.47
1:A:111:ARG:HB3	1:A:212:PHE:CZ	2.51	0.46
1:A:39:LYS:HA	1:A:44:LYS:O	2.16	0.46
1:A:163:ALA:HB2	2:A:300:INA:H363	1.97	0.46
1:A:73:GLN:O	1:A:77:LYS:HG2	2.16	0.45
1:A:175:ASN:HB3	3:A:313:HOH:O	2.16	0.45
1:A:187:ASN:HB2	3:A:344:HOH:O	2.17	0.44
1:A:8:ARG:HD3	1:A:193:TYR:CZ	2.52	0.44
1:A:111:ARG:HD3	1:A:112:GLU:O	2.16	0.44
2:A:300:INA:O10	2:A:300:INA:C12	2.65	0.44
1:A:130:PRO:HA	1:A:166:ALA:O	2.17	0.44
1:A:184:TRP:CD1	2:A:300:INA:H71	2.52	0.43
1:A:6:ASP:HB3	1:A:9:LYS:HD2	1.99	0.43
1:A:17:LYS:NZ	3:A:322:HOH:O	2.50	0.42
1:A:171:ILE:HA	1:A:175:ASN:O	2.20	0.42
1:A:38:LEU:HD13	1:A:105:ALA:HB2	2.03	0.41
1:A:111:ARG:NH1	1:A:113:ILE:HG13	2.33	0.41
1:A:89:TYR:CZ	1:A:91:GLY:HA2	2.56	0.41
1:A:48:LEU:O	1:A:50:PRO:HD3	2.21	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:111:ARG:HD3	1:A:112:GLU:N	2.36	0.41

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	213/215 (99%)	204 (96%)	9 (4%)	0	100 100

There are no Ramachandran outliers to report.

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	173/173 (100%)	162 (94%)	11 (6%)	17 20

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	3	ASP
1	A	39	LYS
1	A	44	LYS
1	A	46	LEU
1	A	52	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	53	LEU
1	A	85	ASP
1	A	87	TYR
1	A	123	ARG
1	A	155	CYS
1	A	164	VAL

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

Mol	Chain	Res	Type
1	A	60	ASN
1	A	70	ASN
1	A	73	GLN
1	A	143	GLN
1	A	187	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 monosaccharides 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	$\#\|Z\| > 2$
2	INA	A	300	1	43,43,43	2.75	7 (16%)	54,55,55	2.47	13 (24%)

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
2	INA	A	300	1	-	17/44/44/44	0/2/2/2

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	300	INA	N41-N40	8.78	1.49	1.39
2	A	300	INA	N19-N18	8.64	1.49	1.39
2	A	300	INA	C21-N19	7.70	1.53	1.37
2	A	300	INA	C21-N41	6.62	1.50	1.37
2	A	300	INA	O30-C31	5.32	1.45	1.35
2	A	300	INA	O8-C9	4.38	1.43	1.35
2	A	300	INA	C16-N18	2.27	1.38	1.34

All (13) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	300	INA	O22-C21-N19	-10.88	106.15	122.38
2	A	300	INA	O22-C21-N41	-9.19	108.67	122.38
2	A	300	INA	C7-O8-C9	4.23	125.37	115.93
2	A	300	INA	O30-C29-C26	3.77	118.45	109.39
2	A	300	INA	C21-N19-N18	-3.53	110.87	117.70
2	A	300	INA	C29-O30-C31	3.44	123.61	115.93
2	A	300	INA	C11-C16-N18	3.43	119.54	114.36
2	A	300	INA	O8-C9-N20	2.81	116.20	110.50
2	A	300	INA	C16-N18-N19	2.54	123.96	119.95
2	A	300	INA	O30-C31-N42	2.52	115.62	110.50
2	A	300	INA	C21-N41-N40	-2.34	113.18	117.70
2	A	300	INA	C33-C38-N40	2.21	117.70	114.36
2	A	300	INA	C13-C12-C11	2.18	121.43	115.43

There are no chirality outliers.

All (17) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	300	INA	O10-C9-O8-C7
2	A	300	INA	C12-C11-C16-N18
2	A	300	INA	C12-C11-N20-C9
2	A	300	INA	N41-C21-N19-N18
2	A	300	INA	N19-C21-N41-N40
2	A	300	INA	O32-C31-O30-C29
2	A	300	INA	N42-C31-O30-C29
2	A	300	INA	N42-C33-C34-C35
2	A	300	INA	N20-C9-O8-C7
2	A	300	INA	C38-C33-C34-C35
2	A	300	INA	C12-C11-C16-O17
2	A	300	INA	C11-C12-C13-C15
2	A	300	INA	N20-C11-C16-N18
2	A	300	INA	C33-C34-C35-C36
2	A	300	INA	C38-N40-N41-C21
2	A	300	INA	C33-C34-C35-C37
2	A	300	INA	N20-C11-C16-O17

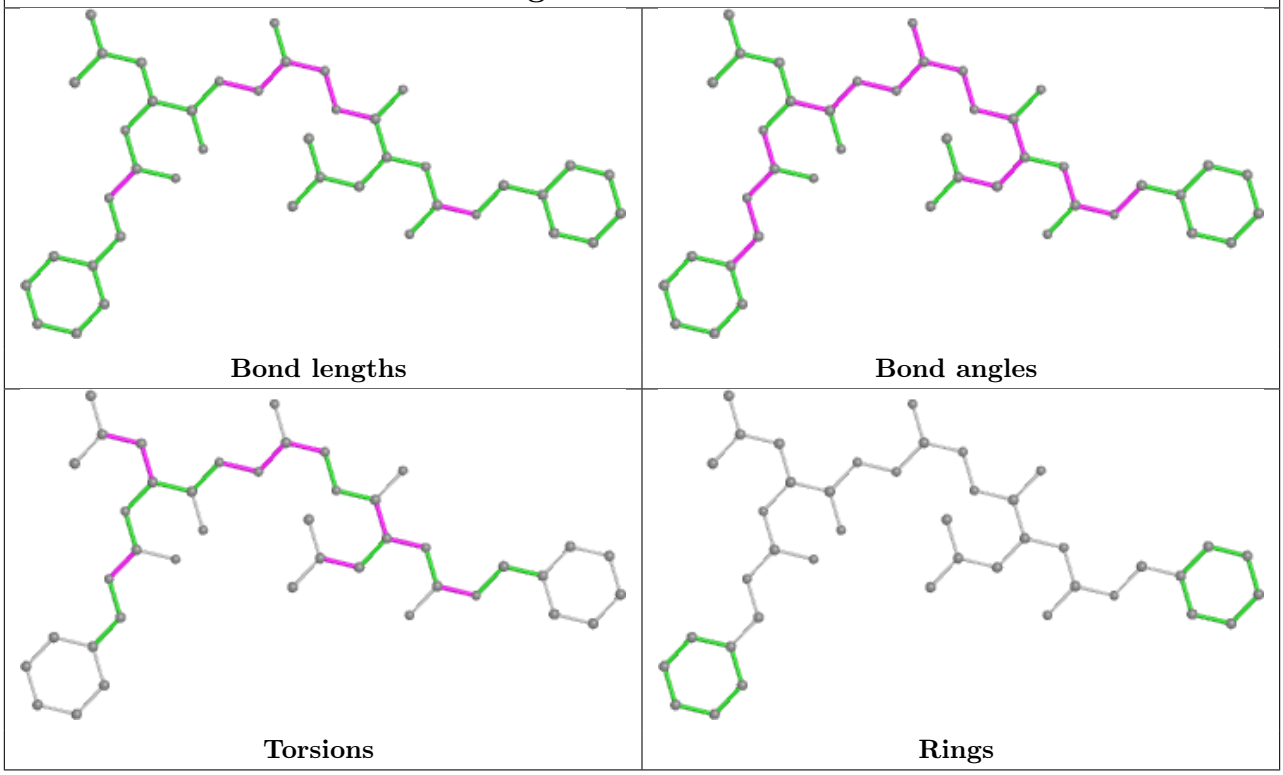
There are no ring outliers.

1 monomer is involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	300	INA	8	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand INA A 300



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.