Full wwPDB X-ray Structure Validation Report ⓘ

May 16, 2020 – 05:16 am BST

PDB ID	:	1PPE
Title	:	THE REFINED 2.0 ANGSTROMS X-RAY CRYSTAL STRUCTURE OF THE COMPLEX FORMED BETWEEN BOVINE BETA-TRYPSIN AND CMTI-I, A TRYPSIN INHIBITOR FROM SQUASH SEEDS (CUCURBITA MAXIMA): TOPOLOGICAL SIMILARITY OF THE SQUASH SEED INHIBITORS WITH THE CARBOXYPEPTIDASE A INHIBITOR FROM POTATOES
Authors	:	Bode, W.; Huber, R.
Deposited on	:	1991-10-24
Resolution	:	2.00 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 ⓘ symbol.

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

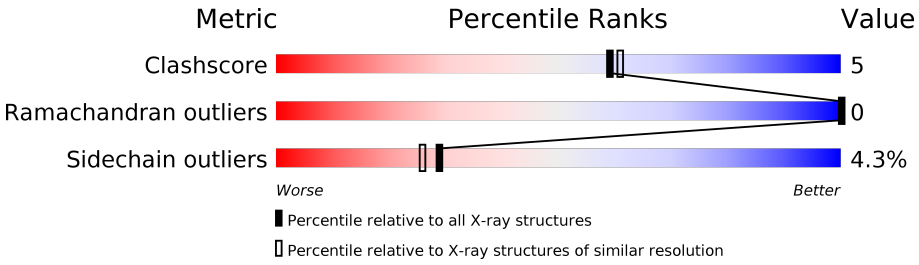
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.00 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	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)

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	E	223	78% 19% .
2	I	29	72% 21% 7%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1991 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 TRYPSIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	E	223	1629	1012	279	324	14	13	0	0

- Molecule 2 is a protein called TRYPSIN INHIBITOR CMTI-I.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	I	29	222	134	39	42	7	4	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	E	126	126	126	0	0
			Total	O		
3	I	14	14	14	0	0

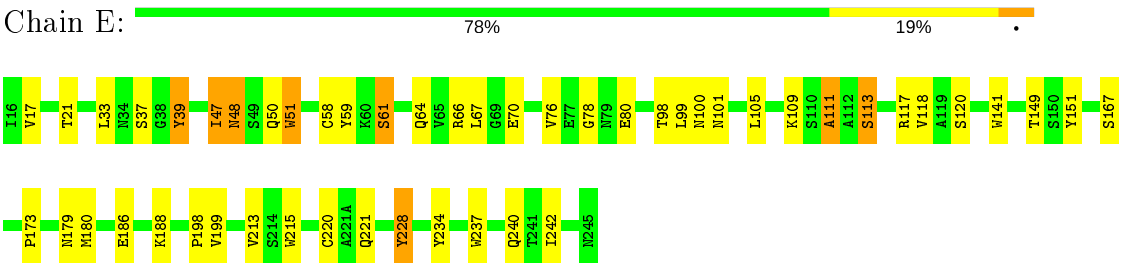
3 Residue-property plots (i)

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.

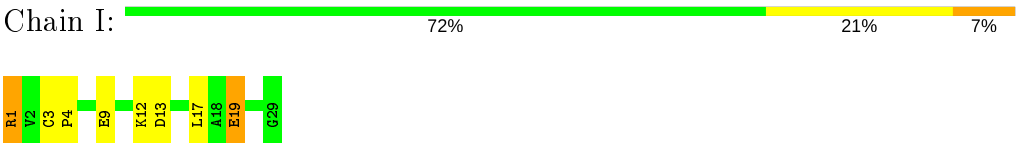
- Molecule 1: TRYPSIN

Chain E:



- Molecule 2: TRYPSIN INHIBITOR CMTI-I

Chain I:



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	59.28Å 55.47Å 74.59Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	6.00 – 2.00	Depositor
% Data completeness (in resolution range)	(Not available) (6.00-2.00)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	EREF	Depositor
R, R_{free}	0.151 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	1991	wwPDB-VP
Average B, all atoms (Å^2)	14.0	wwPDB-VP

5 Model quality

5.1 Standard geometry

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	E	1.21	5/1660 (0.3%)	1.51	8/2250 (0.4%)
2	I	1.06	0/224	2.09	4/297 (1.3%)
All	All	1.19	5/1884 (0.3%)	1.59	12/2547 (0.5%)

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	E	0	18
2	I	0	2
All	All	0	20

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	215	TRP	NE1-CE2	-8.23	1.26	1.37
1	E	237	TRP	NE1-CE2	-7.25	1.28	1.37
1	E	141	TRP	NE1-CE2	-6.41	1.29	1.37
1	E	186	GLU	CD-OE2	5.93	1.32	1.25
1	E	51	TRP	NE1-CE2	-5.44	1.30	1.37

All (12) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	I	1	ARG	NE-CZ-NH2	-18.70	110.95	120.30
2	I	1	ARG	NE-CZ-NH1	14.96	127.78	120.30
1	E	151	TYR	CB-CG-CD1	-7.33	116.60	121.00
1	E	173	PRO	C-N-CA	-6.33	109.01	122.30
1	E	117	ARG	NE-CZ-NH1	6.22	123.41	120.30
2	I	9	GLU	OE1-CD-OE2	-6.21	115.84	123.30
1	E	61	SER	N-CA-CB	-6.08	101.37	110.50

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	80	GLU	OE1-CD-OE2	-5.50	116.71	123.30
1	E	113	SER	N-CA-CB	5.46	118.69	110.50
1	E	39	TYR	N-CA-CB	-5.39	100.90	110.60
2	I	13	ASP	CB-CG-OD1	5.17	122.95	118.30
1	E	113	SER	CB-CA-C	-5.11	100.40	110.10

There are no chirality outliers.

All (20) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	E	109	LYS	Mainchain
1	E	111	ALA	Mainchain
1	E	120	SER	Mainchain
1	E	167	SER	Mainchain
1	E	180	MET	Mainchain
1	E	198	PRO	Mainchain
1	E	21	THR	Mainchain
1	E	228	TYR	Sidechain
1	E	33	LEU	Mainchain
1	E	37	SER	Mainchain
1	E	39	TYR	Sidechain
1	E	47	ILE	Mainchain
1	E	58	CYS	Mainchain
1	E	59	TYR	Sidechain
1	E	70	GLU	Mainchain,Peptide
1	E	76	VAL	Mainchain
1	E	78	GLY	Mainchain
2	I	17	LEU	Mainchain
2	I	19	GLU	Mainchain

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	E	1629	0	1588	16	0
2	I	222	0	214	1	0
3	E	126	0	0	0	1

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	I	14	0	0	0	0
All	All	1991	0	1802	17	1

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 5.

All (17) 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:E:47:ILE:HD11	1:E:242:ILE:HD11	1.84	0.60
1:E:48:ASN:HD22	1:E:50:GLN:H	1.52	0.57
1:E:64:GLN:NE2	1:E:66:ARG:HE	2.03	0.56
1:E:48:ASN:HD22	1:E:48:ASN:C	2.09	0.56
1:E:64:GLN:HE22	1:E:66:ARG:HH21	1.56	0.53
1:E:100:ASN:HD21	1:E:179:ASN:HD22	1.57	0.53
1:E:98:THR:O	1:E:99:LEU:HB2	2.13	0.49
1:E:17:VAL:O	1:E:188:LYS:HA	2.14	0.48
1:E:199:VAL:HG21	1:E:228:TYR:CD1	2.49	0.47
2:I:3:CYS:HA	2:I:4:PRO:HD2	1.89	0.46
1:E:50:GLN:HG3	1:E:111:ALA:HA	1.98	0.46
1:E:48:ASN:HD21	1:E:51:TRP:HD1	1.64	0.45
1:E:67:LEU:HD23	1:E:118:VAL:CG1	2.47	0.44
1:E:101:ASN:ND2	1:E:234:TYR:OH	2.50	0.43
1:E:220:CYS:H	1:E:221:GLN:NE2	2.18	0.42
1:E:48:ASN:ND2	1:E:50:GLN:H	2.16	0.42
1:E:213:VAL:HG22	1:E:228:TYR:HE1	1.86	0.40

All (1) 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)
3:E:968:HOH:O	3:E:970:HOH:O[3_555]	0.15	2.05

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	E	221/223 (99%)	214 (97%)	7 (3%)	0	100	100
2	I	27/29 (93%)	25 (93%)	2 (7%)	0	100	100
All	All	248/252 (98%)	239 (96%)	9 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	E	184/184 (100%)	178 (97%)	6 (3%)	38	37
2	I	26/26 (100%)	23 (88%)	3 (12%)	5	3
All	All	210/210 (100%)	201 (96%)	9 (4%)	29	26

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	48	ASN
1	E	61	SER
1	E	105	LEU
1	E	113	SER
1	E	149	THR
1	E	240	GLN
2	I	1	ARG
2	I	12	LYS
2	I	19	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (10) such sidechains are listed below:

Mol	Chain	Res	Type
1	E	30	GLN
1	E	48	ASN
1	E	64	GLN
1	E	97	ASN
1	E	100	ASN
1	E	101	ASN
1	E	175	GLN
1	E	210	GLN
1	E	221	GLN
1	E	240	GLN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

6.4 Ligands

EDS was not executed - this section is therefore empty.

6.5 Other polymers

EDS was not executed - this section is therefore empty.