Full wwPDB X-ray Structure Validation Report (i)

May 23, 2020 – 12:55 am BST

PDB ID	:	3EMY
Title	:	Crystal structure of Trichoderma reesei aspartic proteinase complexed with pepstatin A
Authors	:	Nascimento, A.S.; Krauchenco, S.; Golubev, A.M.; Gustchina, A.; Wlodawer, A.; Polikarpov, I.
Deposited on	:	2008-09-25
Resolution	:	1.85 Å(reported)

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

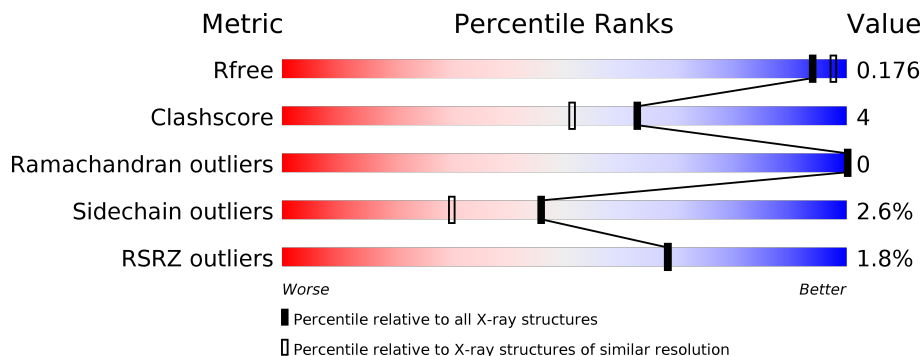|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 1.85 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	Percentile Ranks	Value
Rfree		0.176
Clashscore		4
Ramachandran outliers		0
Sidechain outliers		2.6%
RSRZ outliers		1.8%

Worse ——— *Better*

∎ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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% 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	329	2% / 90% / 9% ·
2	B	6	33% / 50% / 17%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3104 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 Trichoderma reesei Aspartic protease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	329	Total	C	N	O	S	0	5	0
			2434	1516	395	521	2			

- Molecule 2 is a protein called Pepstatin.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	B	6	Total	C	N	O	0	1	0
			52	38	5	9			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	607	Total	O	0	0
			607	607		
3	B	11	Total	O	0	0
			11	11		

3 Residue-property plots ⓘ

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: Trichoderma reesei Aspartic protease
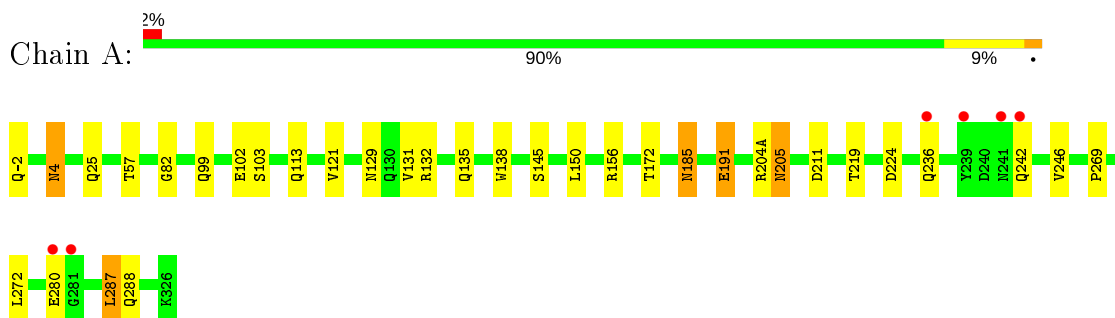
• Molecule 2: Pepstatin

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	74.28Å 74.28Å 160.03Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.47 – 1.85 19.47 – 1.85	Depositor EDS
% Data completeness (in resolution range)	96.5 (19.47-1.85) 96.5 (19.47-1.85)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.14 (at 1.85Å)	Xtriage
Refinement program	PHENIX (phenix.refine)	Depositor
R, R_{free}	0.141 , 0.183 0.133 , 0.176	Depositor DCC
R_{free} test set	1884 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	21.3	Xtriage
Anisotropy	0.015	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 66.2	EDS
L-test for twinning[2]	$<|L|> = 0.49, <L^2> = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	3104	wwPDB-VP
Average B, all atoms (Å^2)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.21% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1] Intensities estimated from amplitudes.
[2] Theoretical values of $<|L|>$, $<L^2>$ 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)

Bond lengths and bond angles in the following residue types are not validated in this section: PCA, STA, IVA

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	1.13	3/2506 (0.1%)	1.00	5/3418 (0.1%)
2	B	1.50	0/24	0.97	0/31
All	All	1.14	3/2530 (0.1%)	1.00	5/3449 (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
2	B	0	4

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	121	VAL	CB-CG1	5.71	1.64	1.52
1	A	191	GLU	CD-OE2	5.50	1.31	1.25
1	A	102	GLU	CB-CG	5.46	1.62	1.52

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	287	LEU	CA-CB-CG	8.27	134.32	115.30
1	A	132	ARG	NE-CZ-NH2	-7.10	116.75	120.30
1	A	156	ARG	NE-CZ-NH1	-5.49	117.56	120.30
1	A	204(A)	ARG	NE-CZ-NH1	-5.40	117.60	120.30
1	A	224	ASP	CB-CG-OD1	5.05	122.84	118.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	327	IVA	Mainchain
2	B	330	STA	Mainchain,Peptide
2	B	331	ALA	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	A	2434	0	2282	20	0
2	B	52	0	69	2	0
3	A	607	0	0	4	2
3	B	11	0	0	0	0
All	All	3104	0	2351	21	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 4.

All (21) 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:211:ASP:HB3	3:A:673:HOH:O	1.41	1.16
1:A:25:GLN:HE22	1:A:57:THR:H	1.07	0.98
1:A:246:VAL:H	1:A:288:GLN:HE22	1.24	0.85
1:A:219:THR:OG1	2:B:328[A]:VAL:HG12	1.76	0.83
1:A:246:VAL:H	1:A:288:GLN:NE2	1.93	0.67
1:A:129:ASN:ND2	1:A:135:GLN:H	1.94	0.65
1:A:150:LEU:C	1:A:150:LEU:HD12	2.20	0.62
2:B:328[A]:VAL:HG22	2:B:330:STA:HG	1.84	0.60
1:A:269:PRO:HD2	1:A:272:LEU:HD12	1.84	0.60
1:A:113:GLN:HG3	3:A:813:HOH:O	2.07	0.55
1:A:191:GLU:OE1	3:A:673:HOH:O	2.17	0.54
1:A:99:GLN:NE2	1:A:138:TRP:HA	2.24	0.53
1:A:4:ASN:C	1:A:4:ASN:HD22	2.11	0.53
1:A:205:ASN:HD22	1:A:205:ASN:N	2.05	0.52
1:A:129:ASN:ND2	1:A:131:VAL:H	2.09	0.50
1:A:185:ASN:HD22	1:A:185:ASN:C	2.14	0.50
1:A:25:GLN:HE22	1:A:57:THR:N	1.91	0.49

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:-2:PCA:HG2	1:A:145:SER:O	2.17	0.45
1:A:236:GLN:HG2	3:A:686:HOH:O	2.16	0.45
1:A:185:ASN:C	1:A:185:ASN:ND2	2.72	0.43
1:A:82:GLY:HA3	1:A:103:SER:O	2.19	0.42

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)
3:A:888:HOH:O	3:A:904:HOH:O[6_565]	2.17	0.03
3:A:666:HOH:O	3:A:869:HOH:O[6_465]	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	332/329 (101%)	328 (99%)	4 (1%)	0	100	100
2	B	4/6 (67%)	4 (100%)	0	0	100	100
All	All	336/335 (100%)	332 (99%)	4 (1%)	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	273/268 (102%)	266 (97%)	7 (3%)	46	30
2	B	3/2 (150%)	3 (100%)	0	100	100
All	All	276/270 (102%)	269 (98%)	7 (2%)	46	31

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	4	ASN
1	A	172	THR
1	A	185	ASN
1	A	205	ASN
1	A	242	GLN
1	A	280	GLU
1	A	287	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (12) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	4	ASN
1	A	25	GLN
1	A	99	GLN
1	A	113	GLN
1	A	129	ASN
1	A	130	GLN
1	A	141	ASN
1	A	185	ASN
1	A	205	ASN
1	A	226	ASN
1	A	236	GLN
1	A	288	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

3 non-standard protein/DNA/RNA residues are 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	#$|Z| > 2$
1	PCA	A	-2	1	7,8,9	1.77	2 (28%)	9,10,12	2.07	6 (66%)
2	STA	B	330	2	10,10,11	1.28	1 (10%)	9,12,14	2.06	2 (22%)
2	STA	B	332	2	8,11,11	1.33	1 (12%)	7,14,14	0.63	0

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
1	PCA	A	-2	1	-	0/0/11/13	0/1/1/1
2	STA	B	330	2	-	3/11/11/12	-
2	STA	B	332	2	-	1/10/12/12	-

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	-2	PCA	CD-N	3.08	1.42	1.34
2	B	330	STA	O-C	3.06	1.37	1.19
1	A	-2	PCA	CA-N	2.92	1.49	1.46
2	B	332	STA	CB-CA	2.64	1.57	1.53

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	330	STA	OH-CH-CM	-4.35	99.75	109.08
2	B	330	STA	O-C-CM	-3.53	115.13	125.43
1	A	-2	PCA	CG-CD-N	2.71	115.41	108.39
1	A	-2	PCA	CA-N-CD	-2.44	105.22	113.58
1	A	-2	PCA	OE-CD-CG	-2.40	122.58	126.76
1	A	-2	PCA	CB-CA-C	-2.33	109.50	112.70
1	A	-2	PCA	O-C-CA	-2.30	118.76	124.78
1	A	-2	PCA	CB-CA-N	2.27	109.83	103.30

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	332	STA	N-CA-CB-CG
2	B	330	STA	O-C-CM-CH
2	B	330	STA	CA-CH-CM-C
2	B	330	STA	OH-CH-CM-C

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	-2	PCA	1	0
2	B	330	STA	1	0

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, 95th 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^2)	Q<0.9
1	A	328/329 (99%)	-0.47	6 (1%)	68	68	12, 18, 38, 71	0
2	B	3/6 (50%)	-0.14	0	100	100	17, 17, 19, 19	0
All	All	331/335 (98%)	-0.47	6 (1%)	68	68	12, 18, 38, 71	0

All (6) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	280	GLU	3.5
1	A	236	GLN	2.8
1	A	241	ASN	2.5
1	A	242	GLN	2.4
1	A	239	TYR	2.1
1	A	281	GLY	2.1

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

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	STA	B	332	12/12	0.90	0.13	19,27,35,38	0
2	STA	B	330	11/12	0.97	0.07	13,14,16,17	0
1	PCA	A	-2	8/9	0.98	0.06	16,18,21,24	0

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.