Full wwPDB X-ray Structure Validation Report (i)

May 16, 2020 – 06:44 am BST

PDB ID	:	5PTI
Title	:	STRUCTURE OF BOVINE PANCREATIC TRYPSIN INHIBITOR. RESULTS OF JOINT NEUTRON AND X-RAY REFINEMENT OF CRYSTAL FORM II
Authors	:	Wlodawer, A.; Huber, R.
Deposited on	:	1984-10-05
Resolution	:	1.00 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 (i) symbol.

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

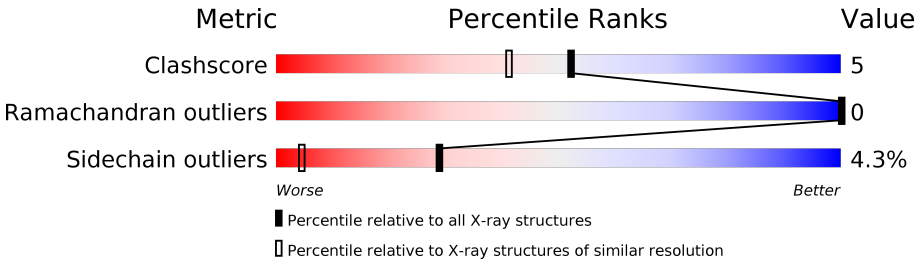
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA,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

The reported resolution of this entry is 1.00 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	1117 (1.06-0.94)
Ramachandran outliers	138981	1043 (1.06-0.94)
Sidechain outliers	138945	1045 (1.06-0.94)

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%

Mol	Chain	Length	Quality of chain
1	A	58	71% 24%

2 Entry composition

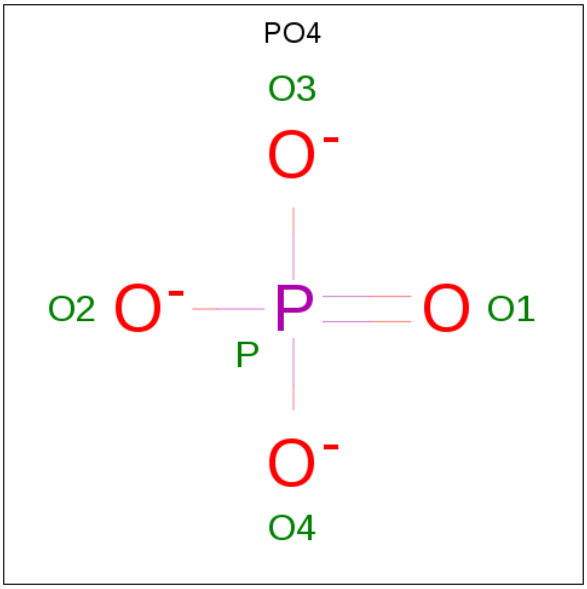
There are 4 unique types of molecules in this entry. The entry contains 1104 atoms, of which 344 are hydrogens and 229 are deuteriums.

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 INHIBITOR.

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
			Total	C	D	H	N	O	S			
1	A	58	909	289	103	344	84	81	8	0	2	0

- Molecule 2 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	P		
2	A	1	5	4	1	0	0

- Molecule 3 is UNKNOWN ATOM OR ION (three-letter code: UNX) (formula: X).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	X		
3	A	1	1	1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	63	Total	D	O	0	0
			189	126	63		

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.

- Molecule 1: TRYPSIN INHIBITOR

Chain A:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	74.10Å 23.40Å 28.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	(Not available) – 1.00 7.12 – 1.00	Depositor EDS
% Data completeness (in resolution range)	(Not available) ((Not available)-1.00) 63.2 (7.12-1.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$	-	Xtriage
Refinement program	unknown	Depositor
R, R_{free}	(Not available) , (Not available) 0.189 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	11.4	Xtriage
Anisotropy	0.185	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.49 , 73.5	EDS
L-test for twinning[1]	$< \|L\| > = 0.42$, $< L^2 > = 0.24$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	1104	wwPDB-VP
Average B, all atoms (Å^2)	19.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 18.72% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: UNX, PO4, DOD

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	1.16	0/482	2.67	22/645 (3.4%)

There are no bond length outliers.

All (22) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	49	GLU	OE1-CD-OE2	-23.72	94.84	123.30
1	A	17	ARG	CD-NE-CZ	21.17	153.24	123.60
1	A	58	ALA	CB-CA-C	12.85	129.38	110.10
1	A	50	ASP	CB-CG-OD2	-12.57	106.99	118.30
1	A	50	ASP	CB-CG-OD1	-11.69	107.78	118.30
1	A	3	ASP	CB-CG-OD2	11.53	128.68	118.30
1	A	50	ASP	OD1-CG-OD2	11.53	145.20	123.30
1	A	49	GLU	CG-CD-OE1	10.44	139.17	118.30
1	A	3	ASP	CB-CA-C	8.43	127.26	110.40
1	A	50	ASP	CB-CA-C	-7.77	94.86	110.40
1	A	56	GLY	C-N-CA	6.92	136.82	122.30
1	A	7[A]	GLU	OE1-CD-OE2	6.85	131.52	123.30
1	A	7[B]	GLU	OE1-CD-OE2	6.85	131.52	123.30
1	A	17	ARG	NE-CZ-NH1	6.67	123.63	120.30
1	A	10	TYR	CB-CG-CD1	-6.30	117.22	121.00
1	A	45	PHE	CB-CG-CD2	-6.24	116.43	120.80
1	A	3	ASP	OD1-CG-OD2	-6.16	111.59	123.30
1	A	49	GLU	CB-CG-CD	6.12	130.72	114.20
1	A	23	TYR	CZ-CE2-CD2	-6.10	114.31	119.80
1	A	25	ALA	CB-CA-C	5.78	118.76	110.10
1	A	57	GLY	CA-C-O	5.66	130.78	120.60
1	A	32	THR	CA-C-O	5.24	131.10	120.10

There are no chirality outliers.

There are no planarity outliers.

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	565	344	445	5	0
2	A	5	0	0	0	0
3	A	1	0	0	0	0
4	A	189	0	0	1	0
All	All	760	344	445	5	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 5.

All (5) 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:GLU:OE1	4:A:304:DOD:O	2.03	0.77
1:A:13:PRO:O	1:A:15:LYS:HE2	2.03	0.54
1:A:4:PHE:O	1:A:7[A]:GLU:CG	2.66	0.44
1:A:4:PHE:O	1:A:7[A]:GLU:HG3	2.15	0.42
1:A:7[B]:GLU:HB3	1:A:8:PRO:HD2	1.92	0.41

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	58/58 (100%)	58 (100%)	0	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	A	48/46 (104%)	46 (96%)	2 (4%)	30 5

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	49	GLU
1	A	50	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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

Of 2 ligands modelled in this entry, 1 is unknown - leaving 1 for Mogul analysis.

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	PO4	A	70	-	4,4,4	0.93	0	6,6,6	0.78	0

There are no bond length outliers.

There are no bond angle outliers.

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

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.