



# Full wwPDB X-ray Structure Validation Report ⓘ

May 14, 2020 – 04:51 pm BST

PDB ID : 3PEP  
Title : REVISED 2.3 ANGSTROMS STRUCTURE OF PORCINE PEPSIN. EVIDENCE FOR A FLEXIBLE SUBDOMAIN  
Authors : Abad-Zapatero, C.; Erickson, J.W.  
Deposited on : 1989-10-24  
Resolution : 2.30 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](mailto: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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtrriage (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

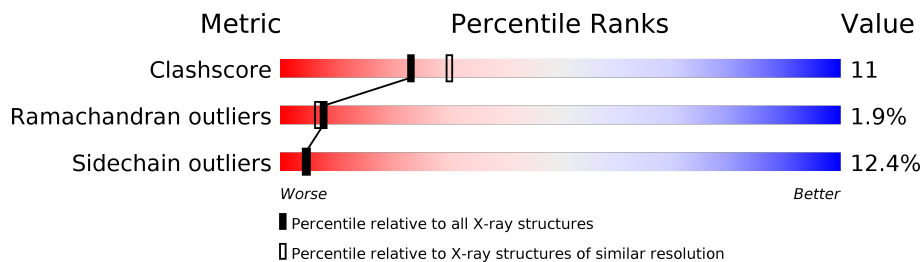
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.30 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	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)

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  $\geq 3$ , 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  $\leq 5\%$ .

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	326	 64% 31% 5%

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 2641 atoms, of which 2 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.

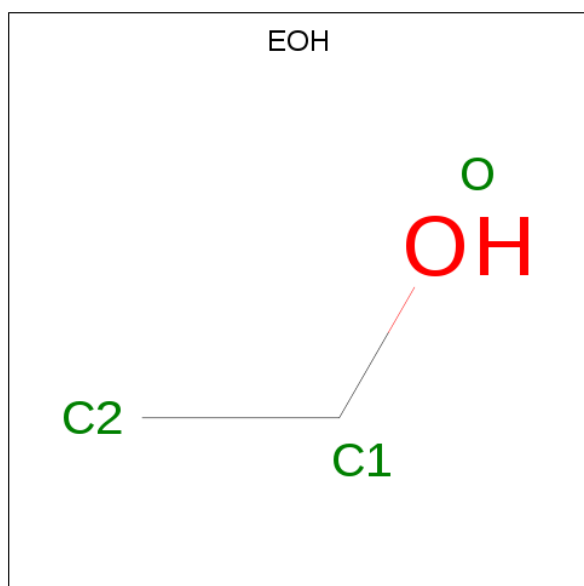
- Molecule 1 is a protein called PEPSIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	326	2429	1530	365	524	10	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	ILE	DELETION	UNP P00791
A	263	ASP	ASN	CONFLICT	UNP P00791

- Molecule 2 is ETHANOL (three-letter code: EOH) (formula: C<sub>2</sub>H<sub>6</sub>O).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	H		
2	A	1	3	2	1	0	0
2	A	1	3	2	1	0	0

- Molecule 3 is water.

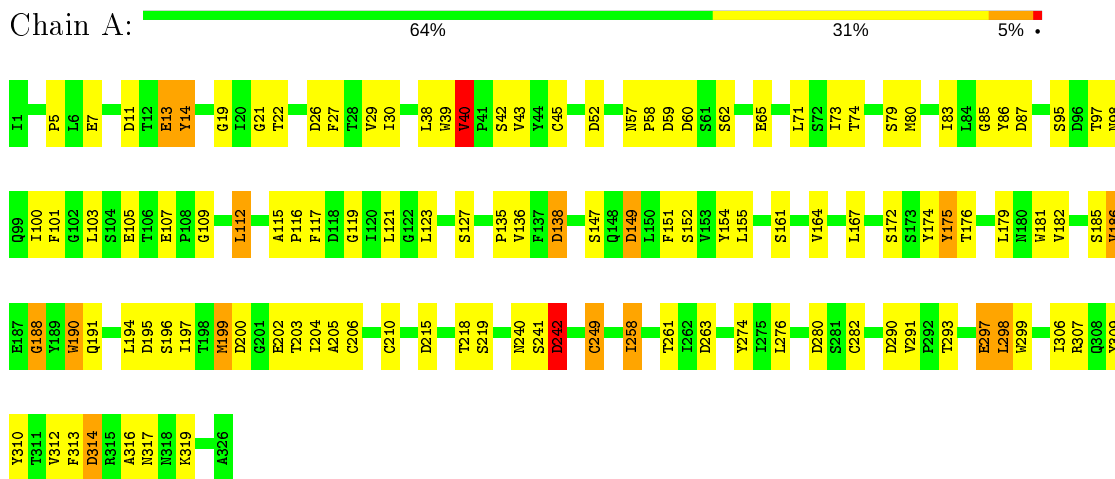
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	206	Total 206	O 206	0	0

### 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: PEPSIN



## 4 Data and refinement statistics

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

Property	Value	Source
Space group	P 1 1 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	55.27Å 73.82Å 36.44Å 90.00° 90.00° 103.38°	Depositor
Resolution (Å)	5.00 – 2.30	Depositor
% Data completeness (in resolution range)	(Not available) (5.00-2.30)	Depositor
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	PROLSQ	Depositor
R, $R_{free}$	0.171 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2641	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	26.0	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: EOH

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	1.01	2/2484 (0.1%)	1.64	32/3400 (0.9%)

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	A	1	5

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	40	VAL	N-CA	-10.16	1.26	1.46
1	A	19	GLY	N-CA	5.13	1.53	1.46

All (32) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	40	VAL	CA-C-O	-10.63	97.77	120.10
1	A	39	TRP	C-N-CA	10.44	147.79	121.70
1	A	40	VAL	CA-C-N	8.80	141.73	117.10
1	A	40	VAL	N-CA-CB	7.30	127.55	111.50
1	A	307	ARG	NE-CZ-NH2	7.07	123.83	120.30
1	A	39	TRP	CB-CG-CD1	-6.81	118.14	127.00
1	A	39	TRP	CD1-CG-CD2	6.81	111.75	106.30
1	A	149	ASP	CB-CG-OD1	6.52	124.17	118.30
1	A	242	ASP	CB-CG-OD2	6.44	124.09	118.30
1	A	86	TYR	CB-CG-CD2	-6.39	117.16	121.00
1	A	22	THR	CA-CB-CG2	-6.25	103.66	112.40

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	249	CYS	CA-CB-SG	-6.05	103.12	114.00
1	A	195	ASP	CB-CG-OD1	6.02	123.72	118.30
1	A	190	TRP	CD1-CG-CD2	5.91	111.03	106.30
1	A	314	ASP	CB-CG-OD1	5.84	123.56	118.30
1	A	39	TRP	CE2-CD2-CG	-5.80	102.66	107.30
1	A	309	TYR	CB-CG-CD1	-5.68	117.59	121.00
1	A	299	TRP	CD1-CG-CD2	5.60	110.78	106.30
1	A	203	THR	CA-CB-CG2	5.57	120.19	112.40
1	A	39	TRP	CG-CD1-NE1	-5.52	104.58	110.10
1	A	215	ASP	CB-CG-OD2	-5.46	113.39	118.30
1	A	14	TYR	CB-CG-CD2	-5.42	117.75	121.00
1	A	280	ASP	CB-CG-OD1	5.42	123.18	118.30
1	A	43	VAL	N-CA-CB	-5.40	99.62	111.50
1	A	40	VAL	CB-CA-C	5.38	121.63	111.40
1	A	39	TRP	CG-CD2-CE3	5.36	138.72	133.90
1	A	190	TRP	CE2-CD2-CG	-5.36	103.01	107.30
1	A	188	GLY	N-CA-C	-5.27	99.93	113.10
1	A	13	GLU	CA-CB-CG	5.14	124.71	113.40
1	A	242	ASP	CA-CB-CG	5.07	124.56	113.40
1	A	182	VAL	CG1-CB-CG2	-5.03	102.85	110.90
1	A	261	THR	CA-CB-OG1	-5.00	98.50	109.00

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	40	VAL	CA

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	154	TYR	Sidechain
1	A	175	TYR	Sidechain
1	A	310	TYR	Sidechain
1	A	40	VAL	Mainchain,Peptide

## 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	2429	0	2252	53	0
2	A	4	2	0	0	0
3	A	206	0	0	1	0
All	All	2639	2	2252	53	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 11.

All (53) 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:27:PHE:CZ	1:A:40:VAL:HG21	2.09	0.86
1:A:27:PHE:CE1	1:A:40:VAL:HG21	2.19	0.77
1:A:149:ASP:HB3	1:A:316:ALA:HB2	1.77	0.67
1:A:27:PHE:CE2	1:A:40:VAL:CG2	2.78	0.66
1:A:199:MET:HB2	1:A:204:ILE:HD13	1.84	0.60
1:A:27:PHE:CE2	1:A:40:VAL:HG21	2.37	0.59
1:A:152:SER:HB3	1:A:312:VAL:HG22	1.85	0.58
1:A:27:PHE:CZ	1:A:40:VAL:CG2	2.86	0.57
1:A:291:VAL:HG23	1:A:298:LEU:HB2	1.88	0.56
1:A:115:ALA:HB1	1:A:117:PHE:CE1	2.40	0.55
1:A:57:ASN:OD1	1:A:59:ASP:HB3	2.07	0.55
1:A:155:LEU:HD22	1:A:306:ILE:CG2	2.39	0.52
1:A:71:LEU:HD11	1:A:73:ILE:HD11	1.91	0.52
1:A:197:ILE:HB	1:A:205:ALA:HB3	1.92	0.52
1:A:181:TRP:HZ3	1:A:314:ASP:HB2	1.75	0.52
1:A:186:VAL:HG12	1:A:191:GLN:OE1	2.10	0.52
1:A:38:LEU:HD13	1:A:121:LEU:HG	1.93	0.51
1:A:27:PHE:CD2	1:A:40:VAL:CG2	2.93	0.51
1:A:107:GLU:HB3	1:A:112:LEU:HD23	1.93	0.50
1:A:194:LEU:HD23	1:A:210:CYS:SG	2.52	0.49
1:A:258:ILE:HG22	1:A:274:TYR:CE2	2.46	0.49
1:A:276:LEU:O	1:A:282:CYS:HA	2.12	0.48
1:A:194:LEU:HB3	1:A:210:CYS:SG	2.53	0.48
1:A:85:GLY:O	1:A:100:ILE:HA	2.14	0.47
1:A:80:MET:C	1:A:80:MET:SD	2.93	0.46
1:A:197:ILE:HG23	1:A:258:ILE:HD11	1.96	0.46
1:A:123:LEU:HB3	1:A:151:PHE:CZ	2.52	0.45
1:A:196:SER:HA	1:A:206:CYS:HB2	1.99	0.45
1:A:191:GLN:NE2	1:A:298:LEU:HD11	2.32	0.44
1:A:155:LEU:HD22	1:A:306:ILE:HG21	1.98	0.44
1:A:83:ILE:HD11	1:A:103:LEU:HD22	1.99	0.44

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:115:ALA:HA	1:A:116:PRO:HD2	1.89	0.44
1:A:181:TRP:HB3	1:A:319:LYS:HD3	2.01	0.43
1:A:135:PRO:O	1:A:138:ASP:HB2	2.19	0.43
1:A:218:THR:HG23	3:A:758:HOH:O	2.18	0.43
1:A:7:GLU:O	1:A:14:TYR:HA	2.18	0.43
1:A:197:ILE:CG2	1:A:258:ILE:HD11	2.48	0.43
1:A:240:ASN:OD1	1:A:242:ASP:HB3	2.19	0.43
1:A:21:GLY:HA2	1:A:87:ASP:OD1	2.19	0.42
1:A:27:PHE:CD2	1:A:40:VAL:HG23	2.55	0.42
1:A:190:TRP:CZ2	1:A:313:PHE:HB3	2.54	0.42
1:A:42:SER:HB3	1:A:103:LEU:HB3	2.02	0.42
1:A:40:VAL:HG12	1:A:101:PHE:CE2	2.56	0.41
1:A:179:LEU:HD13	1:A:181:TRP:CZ2	2.56	0.41
1:A:45:CYS:HA	1:A:105:GLU:O	2.21	0.41
1:A:74:THR:HG23	1:A:79:SER:OG	2.21	0.41
1:A:30:ILE:HG21	1:A:30:ILE:HD13	1.84	0.41
1:A:27:PHE:CD1	1:A:40:VAL:HG21	2.54	0.41
1:A:197:ILE:HD12	1:A:206:CYS:SG	2.61	0.40
1:A:172:SER:HA	1:A:175:TYR:CE1	2.55	0.40
1:A:40:VAL:HG12	1:A:101:PHE:HE2	1.86	0.40
1:A:29:VAL:HA	1:A:119:GLY:O	2.22	0.40
1:A:5:PRO:HA	1:A:164:VAL:HG12	2.02	0.40

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	324/326 (99%)	298 (92%)	20 (6%)	6 (2%)	<b>8</b> <b>7</b>

All (6) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	293	THR
1	A	297	GLU
1	A	109	GLY
1	A	136	VAL
1	A	62	SER
1	A	188	GLY

### 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	275/275 (100%)	241 (88%)	34 (12%)	<b>4</b> <b>5</b>

All (34) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	11	ASP
1	A	13	GLU
1	A	26	ASP
1	A	40	VAL
1	A	52	ASP
1	A	58	PRO
1	A	60	ASP
1	A	65	GLU
1	A	95	SER
1	A	97	THR
1	A	98	ASN
1	A	112	LEU
1	A	127	SER
1	A	138	ASP
1	A	147	SER
1	A	161	SER
1	A	167	LEU
1	A	174	TYR
1	A	176	THR
1	A	185	SER
1	A	186	VAL

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	199	MET
1	A	200	ASP
1	A	202	GLU
1	A	219	SER
1	A	241	SER
1	A	242	ASP
1	A	249	CYS
1	A	258	ILE
1	A	263	ASP
1	A	290	ASP
1	A	297	GLU
1	A	298	LEU
1	A	317	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	230	ASN
1	A	317	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 carbohydrates in this entry.

### 5.6 Ligand geometry [i](#)

2 ligands 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
2	EOH	A	902	-	1,1,2	0.55	0	-		
2	EOH	A	901	-	1,1,2	0.54	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

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.