Full wwPDB X-ray Structure Validation Report (i)

Nov 2, 2024 – 01:32 PM EDT

PDB ID	:	4PEP
Title	:	THE MOLECULAR AND CRYSTAL STRUCTURES OF MONOCLINIC PORCINE PEPSIN REFINED AT 1.8 ANGSTROMS RESOLUTION
Authors	:	Andreeva, N.; Fedorov, A.A.; Sielecki, A.; James, M.
Deposited on	:	1989-12-18
Resolution	:	1.80 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	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 1.80 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.
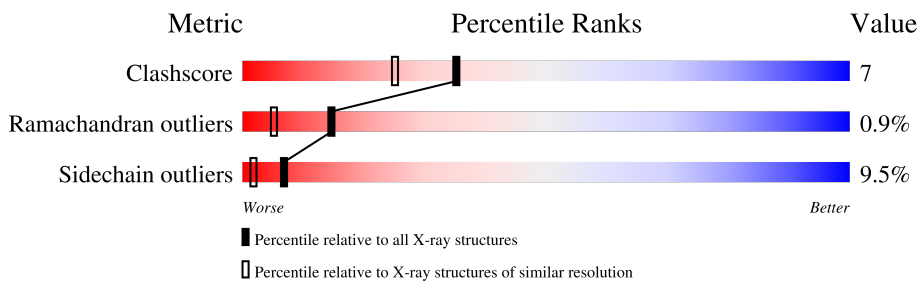
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	180529	8162 (1.80-1.80)
Ramachandran outliers	177936	8077 (1.80-1.80)
Sidechain outliers	177891	8076 (1.80-1.80)

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	326	63% 29% 5% •

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 2620 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 PEPSIN.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	326	Total 2433	C 1530	N 365	O 527	P 1	S 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 water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	187	Total 187	O 187	0	0

3 Residue-property plots ⓘ

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: PEPSIN
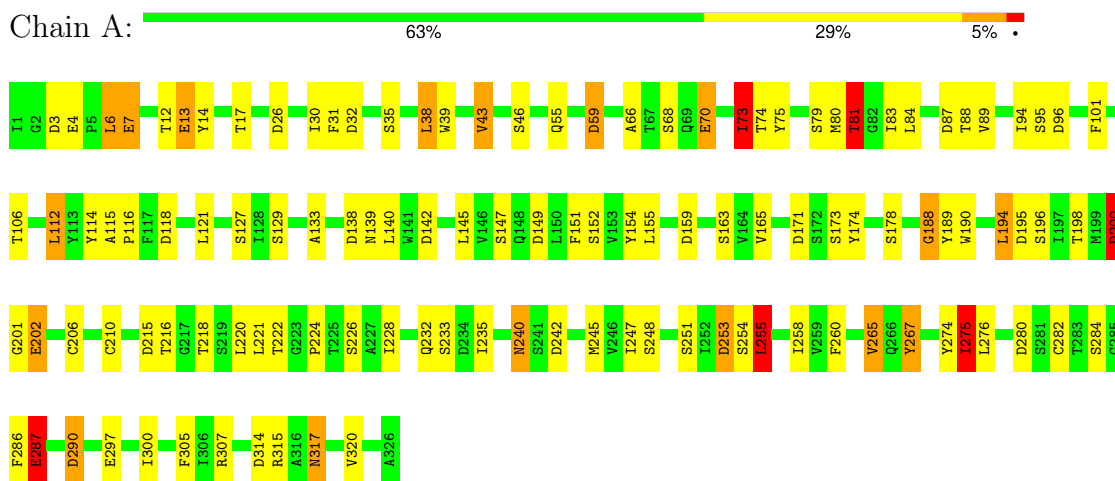
4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	54.83Å 36.44Å 73.68Å 90.00° 103.80° 90.00°	Depositor
Resolution (Å)	8.00 – 1.80	Depositor
% Data completeness (in resolution range)	(Not available) (8.00-1.80)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	PROLSQ	Depositor
R, R_{free}	0.189 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2620	wwPDB-VP
Average B, all atoms (Å^2)	20.0	wwPDB-VP

5   Model quality ⓘ

5.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: SEP

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	1.02	0/2477	2.07	90/3389 (2.7%)

There are no bond length outliers.

All (90) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	215	ASP	CB-CG-OD2	16.73	133.36	118.30
1	A	118	ASP	CB-CG-OD1	11.95	129.06	118.30
1	A	73	ILE	CA-CB-CG2	9.78	130.46	110.90
1	A	87	ASP	CB-CG-OD2	9.39	126.75	118.30
1	A	159	ASP	CB-CG-OD1	9.23	126.61	118.30
1	A	245	MET	CA-CB-CG	9.23	128.99	113.30
1	A	315	ARG	NE-CZ-NH2	-9.07	115.76	120.30
1	A	26	ASP	O-C-N	8.59	136.44	122.70
1	A	142	ASP	CB-CG-OD1	-8.51	110.64	118.30
1	A	253	ASP	CB-CG-OD2	8.42	125.88	118.30
1	A	3	ASP	CB-CG-OD1	-8.20	110.92	118.30
1	A	81	THR	CA-CB-CG2	8.03	123.65	112.40
1	A	215	ASP	CB-CG-OD1	-7.96	111.13	118.30
1	A	84	LEU	CB-CG-CD1	7.74	124.15	111.00
1	A	317	ASN	CB-CA-C	-7.61	95.18	110.40
1	A	138	ASP	O-C-N	7.45	134.62	122.70
1	A	79	SER	CB-CA-C	-7.41	96.02	110.10
1	A	43	VAL	CA-CB-CG1	7.38	121.98	110.90
1	A	138	ASP	CB-CG-OD1	-7.21	111.81	118.30
1	A	240	ASN	CB-CA-C	7.18	124.77	110.40
1	A	307	ARG	NE-CZ-NH2	7.14	123.87	120.30
1	A	188	GLY	N-CA-C	-7.08	95.39	113.10
1	A	66	ALA	CB-CA-C	-7.08	99.49	110.10
1	A	142	ASP	CB-CG-OD2	6.94	124.54	118.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	96	ASP	CB-CG-OD1	6.93	124.53	118.30
1	A	320	VAL	CG1-CB-CG2	-6.86	99.93	110.90
1	A	87	ASP	CB-CG-OD1	-6.83	112.15	118.30
1	A	198	THR	N-CA-CB	-6.75	97.47	110.30
1	A	35	SER	O-C-N	6.73	133.47	122.70
1	A	32	ASP	N-CA-CB	-6.70	98.53	110.60
1	A	284	SER	O-C-N	6.62	134.46	123.20
1	A	112	LEU	CA-CB-CG	6.62	130.53	115.30
1	A	190	TRP	CB-CA-C	6.54	123.48	110.40
1	A	3	ASP	CB-CA-C	6.49	123.38	110.40
1	A	287	GLU	CA-CB-CG	6.30	127.26	113.40
1	A	200	ASP	CA-CB-CG	6.14	126.91	113.40
1	A	55	GLN	CG-CD-OE1	6.11	133.81	121.60
1	A	275	ILE	CB-CA-C	-6.04	99.52	111.60
1	A	17	THR	O-C-N	5.96	132.24	122.70
1	A	133	ALA	CB-CA-C	5.96	119.04	110.10
1	A	13	GLU	CA-CB-CG	5.91	126.40	113.40
1	A	38	LEU	CA-C-N	5.90	130.18	117.20
1	A	87	ASP	N-CA-CB	5.89	121.20	110.60
1	A	198	THR	CA-CB-CG2	5.89	120.64	112.40
1	A	297	GLU	CG-CD-OE2	-5.88	106.54	118.30
1	A	200	ASP	CB-CG-OD1	5.86	123.57	118.30
1	A	75	TYR	CB-CG-CD1	-5.85	117.49	121.00
1	A	151	PHE	CB-CG-CD2	-5.84	116.71	120.80
1	A	13	GLU	OE1-CD-OE2	5.82	130.28	123.30
1	A	121	LEU	CA-C-O	5.80	132.28	120.10
1	A	226	SER	O-C-N	5.73	131.87	122.70
1	A	195	ASP	CB-CG-OD1	5.72	123.44	118.30
1	A	267	TYR	CB-CG-CD2	-5.72	117.57	121.00
1	A	7	GLU	N-CA-CB	5.69	120.84	110.60
1	A	95	SER	N-CA-CB	5.68	119.02	110.50
1	A	305	PHE	CB-CG-CD1	5.67	124.77	120.80
1	A	154	TYR	CA-CB-CG	-5.65	102.67	113.40
1	A	255	LEU	CA-CB-CG	5.64	128.28	115.30
1	A	80	MET	O-C-N	5.62	131.69	122.70
1	A	287	GLU	CA-C-O	5.58	131.82	120.10
1	A	59	ASP	CA-CB-CG	5.57	125.66	113.40
1	A	189	TYR	CB-CG-CD2	-5.55	117.67	121.00
1	A	163	SER	O-C-N	5.45	131.43	122.70
1	A	75	TYR	CA-CB-CG	5.44	123.75	113.40
1	A	275	ILE	N-CA-CB	5.41	123.23	110.80
1	A	129	SER	CB-CA-C	5.37	120.30	110.10

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	149	ASP	CB-CG-OD1	-5.37	113.47	118.30
1	A	121	LEU	O-C-N	-5.36	114.08	123.20
1	A	88	THR	O-C-N	5.36	131.28	122.70
1	A	290	ASP	CB-CG-OD1	5.36	123.12	118.30
1	A	152	SER	CB-CA-C	5.35	120.27	110.10
1	A	43	VAL	N-CA-CB	-5.29	99.87	111.50
1	A	232	GLN	CA-CB-CG	5.27	124.99	113.40
1	A	190	TRP	N-CA-C	-5.27	96.78	111.00
1	A	6	LEU	CB-CG-CD2	-5.24	102.09	111.00
1	A	317	ASN	N-CA-CB	-5.24	101.17	110.60
1	A	30	ILE	C-N-CA	5.21	134.72	121.70
1	A	284	SER	CB-CA-C	-5.21	100.21	110.10
1	A	70	GLU	CB-CG-CD	5.20	128.25	114.20
1	A	159	ASP	CB-CG-OD2	-5.19	113.63	118.30
1	A	114	TYR	CA-CB-CG	-5.14	103.64	113.40
1	A	39	TRP	O-C-N	5.11	130.88	122.70
1	A	39	TRP	CA-C-O	-5.08	109.42	120.10
1	A	151	PHE	N-CA-CB	5.08	119.75	110.60
1	A	112	LEU	CB-CA-C	5.08	119.84	110.20
1	A	89	VAL	O-C-N	5.05	130.78	122.70
1	A	139	ASN	CA-CB-CG	5.04	124.50	113.40
1	A	145	LEU	CA-CB-CG	5.04	126.89	115.30
1	A	12	THR	CA-CB-CG2	5.02	119.42	112.40
1	A	216	THR	C-N-CA	5.02	132.84	122.30

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	2433	0	2251	34	0
2	A	187	0	0	2	0
All	All	2620	0	2251	34	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 7.

All (34) 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:251:SER:HB3	1:A:255:LEU:HD22	1.76	0.68
1:A:171:ASP:HB3	1:A:174:TYR:HD1	1.59	0.67
1:A:235:ILE:O	1:A:247:ILE:HD11	1.95	0.66
1:A:275:ILE:HG21	1:A:282:CYS:HB3	1.79	0.64
1:A:265:VAL:HG13	1:A:267:TYR:CE1	2.33	0.63
1:A:286:PHE:O	2:A:362:HOH:O	2.15	0.63
1:A:6:LEU:HD11	1:A:165:VAL:HG23	1.84	0.59
1:A:194:LEU:HD21	1:A:260:PHE:HD2	1.73	0.54
1:A:81:THR:HB	1:A:106:THR:OG1	2.07	0.54
1:A:171:ASP:HB3	1:A:174:TYR:CD1	2.42	0.54
1:A:275:ILE:CG2	1:A:282:CYS:HB3	2.37	0.53
1:A:7:GLU:O	1:A:14:TYR:HA	2.08	0.53
1:A:222:THR:OG1	1:A:300:ILE:HG12	2.08	0.53
1:A:258:ILE:HD13	1:A:286:PHE:CZ	2.45	0.51
1:A:210:CYS:HB2	2:A:438:HOH:O	2.11	0.51
1:A:224:PRO:HA	1:A:290:ASP:OD1	2.12	0.50
1:A:200:ASP:O	1:A:202:GLU:N	2.48	0.47
1:A:196:SER:HA	1:A:206:CYS:O	2.16	0.46
1:A:275:ILE:HG22	1:A:276:LEU:N	2.30	0.45
1:A:258:ILE:HD12	1:A:274:TYR:CE2	2.52	0.45
1:A:4:GLU:HG2	1:A:31:PHE:CZ	2.52	0.44
1:A:14:TYR:CD2	1:A:155:LEU:HD22	2.52	0.44
1:A:115:ALA:HA	1:A:116:PRO:HD3	1.92	0.43
1:A:274:TYR:CD1	1:A:275:ILE:HD12	2.53	0.43
1:A:276:LEU:O	1:A:282:CYS:HA	2.19	0.43
1:A:221:LEU:HD12	1:A:221:LEU:HA	1.89	0.43
1:A:73:ILE:HG13	1:A:74:THR:N	2.33	0.43
1:A:38:LEU:HG	1:A:101:PHE:HB2	2.01	0.42
1:A:228:ILE:HG13	1:A:287:GLU:O	2.19	0.42
1:A:258:ILE:HD12	1:A:274:TYR:CD2	2.55	0.42
1:A:314:ASP:CG	1:A:317:ASN:HB2	2.40	0.41
1:A:94:ILE:HG21	1:A:140:LEU:HD22	2.04	0.40
1:A:218:THR:HG21	1:A:300:ILE:HG21	2.04	0.40
1:A:218:THR:HG22	1:A:220:LEU:O	2.21	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	323/326 (99%)	307 (95%)	13 (4%)	3 (1%)	14	5

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	200	ASP
1	A	201	GLY
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	274/274 (100%)	248 (90%)	26 (10%)	7	2

All (26) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	13	GLU
1	A	43	VAL
1	A	46	SER
1	A	59	ASP
1	A	70	GLU
1	A	73	ILE
1	A	81	THR
1	A	83	ILE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	112	LEU
1	A	127	SER
1	A	147	SER
1	A	173	SER
1	A	178	SER
1	A	194	LEU
1	A	202	GLU
1	A	233	SER
1	A	240	ASN
1	A	242	ASP
1	A	248	SER
1	A	253	ASP
1	A	254	SER
1	A	255	LEU
1	A	265	VAL
1	A	275	ILE
1	A	280	ASP
1	A	287	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	90	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)

1 non-standard protein/DNA/RNA residue 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	$\#|Z| > 2$
1	SEP	A	68	1	8,9,10	0.99	0	7,12,14	3.15	2 (28%)

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	SEP	A	68	1	-	3/6/8/10	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	68	SEP	OG-P-O1P	7.41	126.47	106.44
1	A	68	SEP	O3P-P-OG	-3.50	97.54	106.67

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	68	SEP	CB-OG-P-O3P
1	A	68	SEP	N-CA-CB-OG
1	A	68	SEP	CB-OG-P-O1P

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates ⓘ

There are no oligosaccharides 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.