Supplementary material

Table S1. Description of the data for the entries' dataset. Data calculated using in-house scripts are described with the label "Propedia 26". The scripts can be found on the supplementary material's GitHub repository.

#	Column	Description	Type	Source
0	id	PDB ID and selected chains	String (8)	Propedia 26
1	AAP	Anti-Angiogenic (AAP): This value indicates the probability that the peptide sequence belongs to this class. Propedia 26 uses a minimum cutoff value of 0.9 to indicate a high likelihood of belonging to the Anti-Angiogenic class. About Anti-Angiogenic peptide class – Function: They inhibit angiogenesis, that is, the formation of new blood vessels. Importance: Blocking angiogenesis is a strategy used to prevent tumor growth, since cancer depends on blood supply to obtain nutrients. Example of use: Development of antitumor and antiviral therapies. See the documentation for details on how this machine learning model was developed.	Float	Orange
2	ABP	Antibacterial (ABP): This value indicates the probability that the peptide sequence belongs to this class. Propedia 26 uses a minimum cutoff value of 0.9 to indicate a high likelihood of belonging to the Antibacterial class. About Antibacterial peptide class – Function: They are antimicrobial peptides that destroy or inhibit the growth of bacteria. Common mechanism: They interact with bacterial membranes, leading to cell lysis (rupture). Importance: They are promising alternatives to traditional antibiotics, especially in the face of bacterial resistance. See the documentation for details on how this machine learning model was developed.	Float	Orange
3	ACP	Anticancer (ACP): This value indicates the probability that the peptide sequence belongs to this class. Propedia 26 uses a minimum cutoff value of 0.9 to indicate a high likelihood of belonging to the Anticancer class. About Anticancer peptide class – Function: They induce selective death of tumor cells without significantly affecting normal cells. Mechanism: They can act by altering the permeability of cancer cell membranes, activating apoptosis, or modulating signaling pathways. Application: Development of next-generation antineoplastic therapies. See the documentation for details on how this machine learning model was developed.	Float	Orange
4	AIP	Anti-Inflammatory (AIP): This value indicates the probability that the peptide sequence belongs to this class. Propedia 26 uses a minimum cutoff value of 0.9	Float	Orange

		to indicate a high likelihood of belonging to the Anti-Inflammatory class. About Anti-Inflammatory peptide class – Function: They reduce or regulate exaggerated inflammatory responses. Mechanism: They can inhibit pro-inflammatory cytokines (such as TNF-α, IL-6) or modulate macrophage activity. Application: Treatment of chronic inflammatory and autoimmune diseases. See the documentation for details on how this machine learning model was developed.		
5	ASA_Complex	ASA: Accessible Surface Area (ASA) is the measure of the entire surface area of the molecule that is exposed and can come into contact with the solvent, usually water (value given in Å²)	Float	NACCESS
6	ASA_Peptide	ΔASA (peptide): ΔASA_peptide represents the surface area that is no longer exposed to the solvent upon complex formation and is calculated by the equation: ΔASA = ASA_unbound - ASA_bound. (Value given in Å²)	Float	NACCESS
7	ASA_Protein	ΔASA (protein): ΔASA_protein represents the surface area that is no longer exposed to the solvent upon complex formation and is calculated by the equation: ΔASA = ASA_unbound - ASA_bound. (Value given in Å²)	Float	NACCESS
8	BPP%	Buried Peptide Percentage (%). It can be obtained by: 100 * BPepA / ASA_Peptide.	Int	NACCESS
9	BPepA	Buried peptide area (value given in Å²). Note that BSA = (BPepA + BProA)/2. Then, BPepA = 2*BSA - BProA.	Int	NACCESS
10	BProA	Buried protein area (value given in Å²). Note that BSA = (BPepA + BProA)/2. Then, BProA = 2*BSA - BPepA.	Int	NACCESS
11	BSA	Buried Surface Area represents the area effectively shared at the binding interface and was calculated according to the expression. It can be calculated using the formula: BSA = (ASA_protein + ASA_peptide − ASA_complex) / 2 (value given in Å²)	Int	NACCESS
12	CLASSIFICATION	PDB Classification	String	PDB
13	DEPOSITION_DATE	PDB Deposition Date	String	PDB
14	Interface Residues	Number of Intermolecular Contacts: Total number of atomic contacts between the protein and the peptide within a specified cutoff distance (typically ≤ 5.5 Å). A higher number of contacts usually indicates a more extensive interaction interface.	String	Propedia 26
15	No. of apolar-apolar contacts	Number of Apolar–Apolar Contacts: Number of hydrophobic–hydrophobic interactions (e.g., Leu–Val, Phe–Ile) that promote interface stabilization through the exclusion of water molecules (hydrophobic effect).	Int	Prodigy
16	No. of apolar-polar contacts	Number of Apolar–Polar Contacts: Count of interactions between hydrophobic and polar residues at	Int	Prodigy

		the interface, which can contribute to partial desolvation and interface packing.		
17	No. of charged-apolar contacts	Number of Charged–Apolar Contacts: Number of contacts between charged residues and hydrophobic residues (e.g., Arg–Leu, Lys–Val). These interactions typically contribute less to stability but may influence interface geometry.	Int	Prodigy
18	No. of charged-charged contacts	Number of Charged–Charged Contacts: Number of interactions between oppositely charged residues (e.g., Lys–Asp, Arg–Glu) across the binding interface, contributing significantly to electrostatic stabilization.	Int	Prodigy
19	No. of charged-polar contacts	Number of Charged–Polar Contacts: Count of contacts between charged residues and polar uncharged residues (e.g., Lys–Ser, Asp–Thr), which often form hydrogen bonds or dipole interactions.	Int	Prodigy
20	No. of intermolecular contacts	Number of Intermolecular Contacts: Total number of atomic contacts between the protein and the peptide within a specified cutoff distance (typically ≤ 5.5 Å). A higher number of contacts usually indicates a more extensive interaction interface.	Int	Prodigy
21	No. of polar-polar contacts	Number of Polar–Polar Contacts: Number of interactions between polar uncharged residues (e.g., Ser–Thr, Asn–Gln), frequently involving hydrogen bonding or dipole alignment across the interface.	Int	Prodigy
22	PDB_ID	PDB ID	String (4)	PDB
23	PEPTIDE_CHAIN	Chain: Unique identifier assigned to each molecular chain within the same crystallographic structure or PDB entry.	String (1)	PDB
24	PEPTIDE_DESC	Description: Annotated name or description of the polymer chain, as defined in the PDB file.	String	PDB
25	PEPTIDE_SEQ	Sequence: The primary amino acid structure of the protein or peptide, defining its linear arrangement of residues.	String	PDB
26	PEPTIDE_SIZE	Length (residues): Total number of amino acid residues observed in the polymer chain.	Int	
27	PROTEIN_CHAIN	Chain: Unique identifier assigned to each molecular chain within the same crystallographic structure or PDB entry.	String (1)	PDB
28	PROTEIN_DESC	Description: Annotated name or description of the polymer chain, as defined in the PDB file.	String	PDB
29	PROTEIN_SEQ	Sequence: The primary amino acid structure of the protein or peptide, defining its linear arrangement of residues.	String	PDB
30	PROTEIN_SIZE	Length (residues): Total number of amino acid residues observed in the polymer chain.	String	PDB
31	Percentage of apolar NIS residues	Percentage of Apolar NIS Residues (%): Proportion of residues in the Non-Interacting Surface (NIS) that are classified as apolar, expressed as a percentage. This value helps assess the hydrophobic character of the exposed surface outside the binding interface.	Float	Prodigy

32	Percentage of charged NIS residues	Percentage of Charged NIS Residues (%): Proportion of residues in the Non-Interacting Surface that are charged (either positively or negatively), expressed as a percentage. It reflects the electrostatic nature of the surface not involved in binding.	Float	Prodigy
33	Predicted binding affinity (kcal.mol-1)	Predicted Binding Affinity (kcal·mol^{-1}): Estimated Gibbs free energy of binding (ΔG), in kilocalories per mole. More negative values indicate stronger predicted binding between the protein and peptide.	Float	Prodigy
34	Predicted dissociation constant (M) at 25.0	Predicted Dissociation Constant (M) at 25.0 °C: Predicted equilibrium dissociation constant (K_d), expressed in molar units (M), at 25 °C. It represents the expected concentration of the complex at which half of the binding sites are occupied. Lower values correspond to higher binding affinity.	String	Prodigy
35	QSP	Quorum Sensing (QSP): This value indicates the probability that the peptide sequence belongs to this class. Propedia 26 uses a minimum cutoff value of 0.9 to indicate a high likelihood of belonging to the Quorum Sensing class. About Quorum Sensing peptide class – Function: They participate in bacterial communication (quorum sensing), regulating collective behaviors such as biofilm formation and virulence. Importance: Understanding and manipulating these peptides can lead to strategies to control bacterial infections without necessarily killing the bacteria (reducing selective pressure for resistance). See the documentation for details on how this machine learning model was developed.	Float	Orange
36	RESOLUTION	Structure Resolution	String	PDB
37	SBP	Surface Binding (SBP): This value indicates the probability that the peptide sequence belongs to this class. Propedia 26 uses a minimum cutoff value of 0.9 to indicate a high likelihood of belonging to the Surface Binding class. About Surface Binding peptide class – Function: They bind to biological surfaces or materials, such as metals, polymers, or minerals. Biotechnological use: They can be used to immobilize enzymes, design biomaterials, biosensors, or nanodevices. Example: Peptides that bind strongly to gold, silica, or metal oxides for use in nanotechnology. See the documentation for details on how this machine learning model was developed.	Float	Orange
38	STRUCTURE_METHOD	Experimental method used for determining structure	String	PDB
39	TITLE	PDB entry title	String	PDB
40	binding-cluster	Structures with similar binding site. For more details, see https://doi.org/10.1186/s12859-020-03881-z	String	Propedia 26
41	interface-cluster	Structures with similar interface. For more details, see https://doi.org/10.1186/s12859-020-03881-z	String	Propedia 26
42	is_leader		String	Propedia 26

43	leader_id		String	Propedia 26
44	organism	PDB organism	String	PDB
45	peptide_AliphaticIndex	Aliphatic Index: A measure of the relative volume occupied by aliphatic side chains (Ala, Val, Ile, and Leu). It is often correlated with the thermostability of the protein.	Float	ProtParam
46	peptide_ExtCoeff_Disulfide	Extinction Coefficient (with disulfide): Molar extinction coefficient (in M^{-1} cm^{-1}) calculated assuming all cysteine residues form disulfide bonds (Cys–Cys). This value indicates the protein's absorbance at 280 nm under these conditions.	Int	ProtParam
47	peptide_ExtCoeff_NoDisulfide	Extinction Coefficient (no disulfide): Molar extinction coefficient (in M^{-1} cm^{-1}) calculated assuming no disulfide bond formation, i.e., all cysteine residues remain in the reduced form.	Int	ProtParam
48	peptide_Formula	Atomic Formula: The complete elemental formula representing the protein's overall atomic composition.	String	ProtParam
49	peptide_GRAVY	GRAVY (Grand Average of Hydropathy): The average hydropathy score of all amino acids in the sequence, based on the Kyte-Doolittle scale. Positive values indicate a more hydrophobic protein, while negative values suggest a more hydrophilic character.	Float	ProtParam
50	peptide_HydrophobicPercent	Hydrophobic (%): The proportion of residues in the sequence that are classified as hydrophobic (e.g., Ala, Val, Leu, Ile, Phe, Trp, Met), expressed as a percentage of the total number of residues.	Float	ProtParam
51	peptide_InstabilityIndex	Instability Index: A computed value that estimates the in vitro stability of a protein. Proteins with an instability index greater than 40 are predicted to be unstable, while lower values indicate greater stability.	Float	ProtParam
52	peptide_MW	Molecular Weight (Da): Total molecular mass of the chain, expressed in Daltons (Da), calculated as the sum of the atomic masses of all atoms in the protein.	Float	ProtParam
53	peptide_NegativeResidues	Negative Residues: Total number of negatively charged amino acids in the sequence (Asp and Glu).	Int	ProtParam
54	peptide_PositiveResidues	Positive Residues: Total number of positively charged amino acids in the sequence (Lys, Arg, and His).	Int	ProtParam
55	peptide_TotalAtoms	Total Atoms: The total number of atoms constituting the polypeptide chain.	Int	ProtParam
56	peptide_pI	Isoelectric Point (pI): The pH value at which the protein carries no net electrical charge, resulting in minimal electrophoretic mobility.	Float	ProtParam
57	protein_AliphaticIndex	Aliphatic Index: A measure of the relative volume occupied by aliphatic side chains (Ala, Val, Ile, and Leu). It is often correlated with the thermostability of the protein.	Float	ProtParam
58	protein_ExtCoeff_Disulfide	Extinction Coefficient (with disulfide): Molar extinction coefficient (in M^{-1} cm^{-1}) calculated assuming all cysteine residues form disulfide bonds (Cys–Cys). This value indicates the protein's absorbance at 280 nm under these conditions.	Int	ProtParam

59	protein_ExtCoeff_NoDisulfide	Extinction Coefficient (no disulfide): Molar extinction coefficient (in M^{-1} cm^{-1}) calculated assuming no disulfide bond formation, i.e., all cysteine residues remain in the reduced form.	Int	ProtParam
60	protein_Formula	Atomic Formula: The complete elemental formula representing the protein's overall atomic composition.	String	ProtParam
61	protein_GRAVY	GRAVY (Grand Average of Hydropathy): The average hydropathy score of all amino acids in the sequence, based on the Kyte-Doolittle scale. Positive values indicate a more hydrophobic protein, while negative values suggest a more hydrophilic character.	Float	ProtParam
62	protein_HydrophobicPercent	Hydrophobic (%): The proportion of residues in the sequence that are classified as hydrophobic (e.g., Ala, Val, Leu, Ile, Phe, Trp, Met), expressed as a percentage of the total number of residues.	Float	ProtParam
63	protein_InstabilityIndex	Instability Index: A computed value that estimates the in vitro stability of a protein. Proteins with an instability index greater than 40 are predicted to be unstable, while lower values indicate greater stability.	Float	ProtParam
64	protein_MW	Molecular Weight (Da): Total molecular mass of the chain, expressed in Daltons (Da), calculated as the sum of the atomic masses of all atoms in the protein.	Float	ProtParam
65	protein_NegativeResidues	Negative Residues: Total number of negatively charged amino acids in the sequence (Asp and Glu).	Int	ProtParam
66	protein_PositiveResidues	Positive Residues: Total number of positively charged amino acids in the sequence (Lys, Arg, and His).	Int	ProtParam
67	protein_TotalAtoms	Total Atoms: The total number of atoms constituting the polypeptide chain.	Int	ProtParam
68	protein_pI	Isoelectric Point (pI): The pH value at which the protein carries no net electrical charge, resulting in minimal electrophoretic mobility.	Float	ProtParam
69	seq100_clusters	This category indicates the ID of the complex that contains a peptide with 100% sequence identity.	String	Propedia 26
70	sequence-cluster	Structures with sequences with high identity. For more details, see https://doi.org/10.1186/s12859-020-03881-z	String	Propedia v1