frontiers | Frontiers in Education

TYPE Review
PUBLISHED 23 November 2023
DOI 10.3389/feduc.2023.1290287

OPEN ACCESS

EDITED BY
Krishna Prasad Rajan,
Royal Commission for Jubail and Yanbu,
Saudi Arabia

REVIEWED BY
Aravinthan Gopanna,
Royal Commission for Jubail and Yanbu,
Saudi Arabia
Abdel-Karim Musa,
Yanbu Industrial College, Saudi Arabia

*CORRESPONDENCE
Tomas Lafarga
✉ tomas.lafarga@ual.es

RECEIVED 20 September 2023
ACCEPTED 13 November 2023
PUBLISHED 23 November 2023

CITATION
Villaró-Cos S and Lafarga T (2023) Online tools to support teaching and training activities in chemical engineering: enzymatic proteolysis. *Front. Educ.* 8:1290287.
doi: 10.3389/feduc.2023.1290287

Online tools to support teaching and training activities in chemical engineering: enzymatic proteolysis

Silvia Villaró-Cos[1,2] and Tomas Lafarga[1,2]*

[1]Department of Chemical Engineering, University of Almería, Almeria, Spain, [2]CIESOL Solar Energy Research Centre, Joint Centre UAL-CIEMAT, Almeria, Spain

The practical teaching or training of enzymatic proteolysis can prove challenging because of the lengthy duration of the process, the complexity of identifying short amino acid sequences, the high cost of the enzymes, and the need to use very specific equipment. There are several freely-available online tools that, despite being employed by scientists to help identify bioactive peptides, are not commonly used for teaching and training activities. This work summarises the most common protein and peptide databases along with other tools that allow one to simulate enzymatic hydrolysis of a given protein and to study the structure, physicochemical properties, bioactivity, toxicity, allergenicity, and even the bitterness of the resulting peptides. Overall, *in silico* tools can be used during the teaching and training of chemical engineers as innovative alternatives to conventional laboratory work and theoretical classes.

KEYWORDS

education, bioactive peptides, enzymes, BIOPEP, databases, proteins

1 Introduction

Proteins are large macromolecules made up of amino acids. They are the building blocks of life and are needed by all living things to function properly. In humans, for example, proteins are the basis of body structures, act as messengers, maintain the proper pH, transport and store nutrients, provide energy, and are responsible for most biochemical reactions. At the industrial level, proteins are used in many areas. For example, proteins are one of the key ingredients in many human foods and most animal feeds, amino acids are common agricultural products, and enzymes (proteins) are common ingredients in the food industry (e.g., baked goods and drinks); they are also widely used during the production of pulp, paper, and leather.

For most commercial protein applications, it is important to retain the protein structure given that their activity is structure dependent. However, other protein applications and properties are enhanced after the protein structure of the protein is disrupted. For example, the cleavage of proteins might result in the release of bioactive fragments or might improve the techno-functional properties of proteins with poor functionality (Vogelsang-O'dwyer et al., 2022). Bioactive peptides are short amino acid sequences that have biological effects when they are released from their parent proteins. Most of the bioactive peptides reported to date are of between 2 and 20 amino acids in length and several have been shown to have different modes of action and to exert more than one form of bioactivity (Ulug et al., 2021). Enzymatic hydrolysis is the most widely used technology to produce bioactive peptides, usually from food sources. This process is carried out under controlled conditions (ideally

under the optimal ones for the enzyme being used) and is generally preferred to microbial fermentation because of its high specificity. To obtain quality products with high bioactivity, it is essential to understand the enzyme cleavage sites and the specificity of the bioactive peptides as well as to apply adequate process control. This knowledge is key for chemical engineers and food engineers working in the field of bioactive peptide generation. The use of enzymes in industrial processes is covered by different subjects comprising most chemical engineering curricula, for example, Bioreactors, Reactor Design, Reaction Kinetics, Biochemical Engineering, and Chemistry. The use of *in silico* strategies, meaning strategies conducted via computer simulations, has become increasingly important in the field and has been used to predict toxic (Gupta et al., 2013), antiviral (Charoenkwan et al., 2020), cell penetrating (Gautam et al., 2013), antihypertensive (Kumar et al., 2015), and antidiabetic (Lafarga et al., 2014) peptides, amongst other bioactivities. Moreover, being able to predict the cleavage of a given protein by means of a computer is also useful for estimating the potential bioavailability of amino acids upon gastrointestinal digestion (Sayd et al., 2018). Because of their potential to save both time and money, *in silico* tools have gained increasing importance in the scientific literature. The practical teaching or training on enzymatic proteolysis can be challenging because of the long duration of the process, the complexity of identifying short amino acid sequences, the high cost of enzymes, and the need to use of very specific equipment including liquid chromatographs and mass spectrometers. For these reasons, at most universities, the teaching of enzymatic proteolysis is generally theoretical and training in this topic is rare. The use of computer simulations in training is also beneficial as they give the students the time necessary to understand the concept (Rodrigues et al., 2010).

Over the last decade, several freely available online tools have been developed that are now being implemented by scientists all around the world. Such tools can also be used to teach and facilitate the understanding of proteins, bioactive peptides, and proteolysis. The results have a strong graphic impact and allow the rapid simulation of reactions and processes that would otherwise be impossible in a training environment. Despite being common in scientific environments, *in silico* tools are not yet widely implemented in teaching and training activities. For this reason, the goal of our review is to summarise the main online tools that can potentially be used in educational and training environments although they are currently only being used by scientists to simulate enzymatic hydrolysis and the release of bioactive peptides. The most comprehensive protein and bioactive peptide databases will be explained together with various tools that enable enzymatic proteolysis to be estimated and understood. Other tools that predict bioactivity, toxicity, or allergenicity will also be discussed suggesting, with tips on their use in science lessons. These tools could be introduced into various Undergraduate and Masters' Degrees, including Chemistry, Chemical Engineering, Food Science, and Biotechnology. A recent work revealed that chemical engineering students were very satisfied with simulator-based learning approaches (Borreguero et al., 2019; Roman et al., 2020). Furthermore, professors teaching a range of subjects (for example, Organic Chemistry, Biochemistry, Bioreactors, Bioprocesses, Food Engineering, and Biotechnology), could benefit from these easy-to-use and freely available tools.

2 Protein databases

2.1 UniProt

The UniProt database, available at https://www.uniprot.org, is a collection of protein sequences covering more than 120 million proteins (Bateman, 2019). This database is the result of a collaboration between the European Bioinformatics Institute, the SIB Swiss Institute of Bioinformatics, and the Protein Information Resource. It counts on the participation of over 100 people who oversee database curation and software development, and who give support. This tool could be used by students to search for protein sequences (and much more) using the central "Find your protein" search bar; it also supports advanced searches that allow keywords and taxonomies, etc. to be included (Figure 1). We searched for bovine serum albumin (P02769, ALBU_BOVIN) and were immediately directed to a vast amount of information including publications related to the protein, its amino acid sequence, 3D structure, subcellular location, potential allergenic properties, features, and function in the body (Figure 2). This information is very useful for students and facilitates access to scientific information on any given protein. The tool also searches for related proteins, which are either obtained from the same organism (e.g., bovine fibrinogen) or a related one (e.g., porcine serum albumin).

UniProt includes other tools, for example, BLAST, which stands for Basic Local Alignment Search Tool, which finds regions of local similarity between protein sequences (Zaru and Orchard, 2023). In the case of BSA (ALBU_BOVIN), the software was able to identify many similar sequences including sheep serum albumin (ALBU_SHEEP, 92.4% similarity), porcine serum albumin (ALBU_PIG, 79.9% similarity), cat serum albumin (ALBU_CAT, 78.5% similarity), and dog serum albumin (ALBU_CANLF, 76.6% similarity). The similarity between the amino acid sequence of bovine and human serum albumin (ALBU_HUMAN) was 76.6%. This tool allows the user, in this case students, to estimate the release of bioactive peptides from a given protein source based on the similarity of the protein to a known source of bioactive peptides. This is useful when estimating the differences that can occur in the final product if the raw material (the protein source) is changed to a similar one. This tool can be complemented with the Align tool, available at https://www.uniprot.org/align/, which allows two or more protein sequences to be aligned and to observe where the main difference occur (Figure 3). This option can identify which peptides will be different if the protein source is changed, offering a possible key to a process in which not all the peptides generated are of interest.

The Peptide Search option (available at UniProt) is also very useful because it allows one to search for a known bioactive peptide from the proteins available in the database. For example, the tripeptides IPP and VPP are known to attenuate the development of hypertension and may also have beneficial effects on vascular function (Turpeinen et al., 2009). Both peptides were identified for the first time from fermented milk proteins although they are present in other protein sequences as well. We used the Peptide Search option to look for protein that include the IPP sequence. The software was able to identify almost 1,000,000 proteins that contain this peptide sequence, suggesting that its production might not be limited to milk proteins. As an example, Peptide Search identified the IPP sequence in the proteins FOGA_ASPRC, FRA17_FRAAN, and LHY_PETHY, which are found in *Aspergillus ruber*, *Fragaria ananassa* (strawberry), and *Petunia hybrida*.

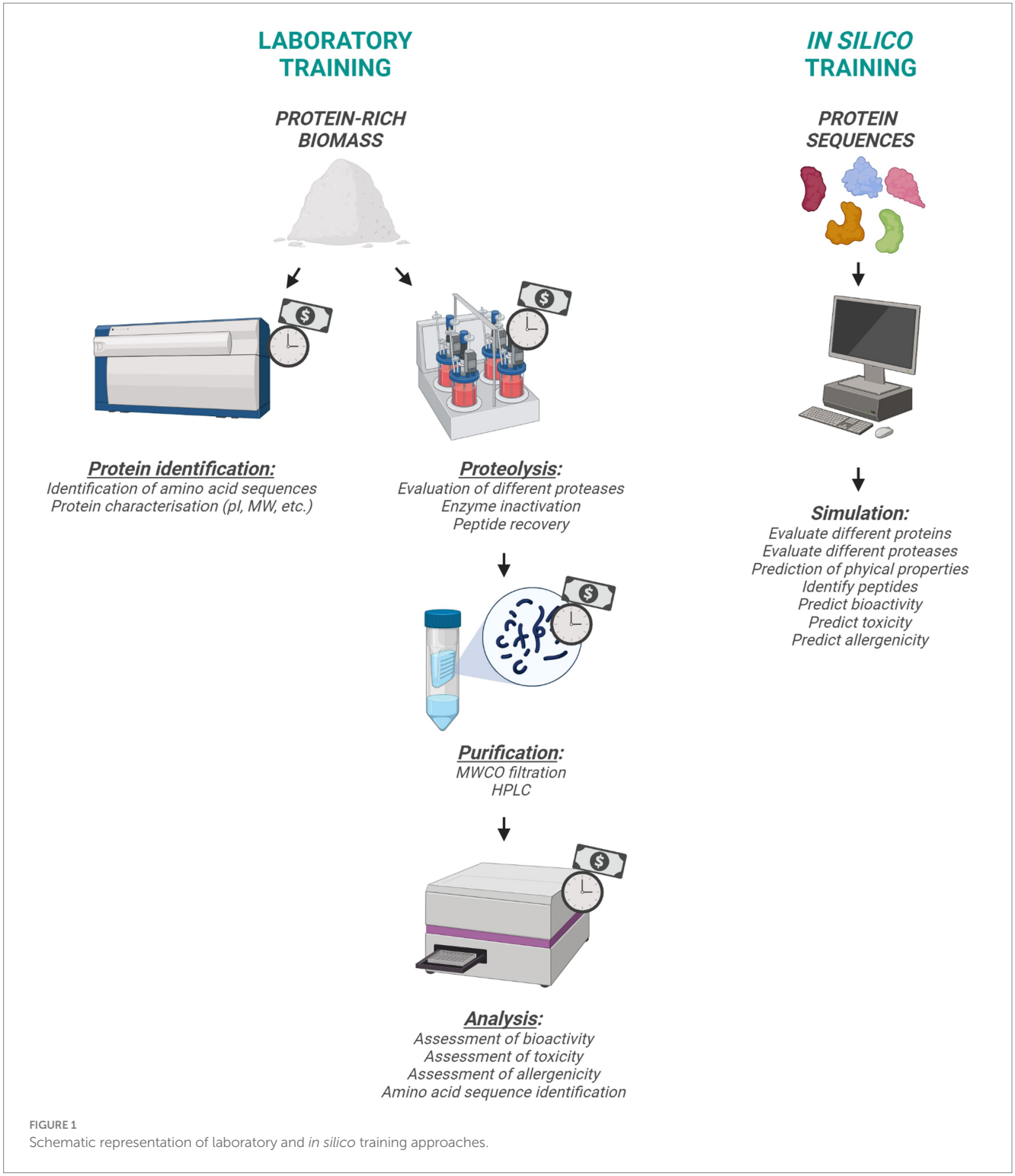


FIGURE 1
Schematic representation of laboratory and *in silico* training approaches.

2.2 RCSB protein data Bank

The Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) is available at http://rcsb.org/ (Rose et al., 2016). The webpage includes public data with no usage limitations and offers several tools for structure query, analysis, and visualisation. In terms of analysis, the webpage offers different tools including: (i) Pairwise Structure Alignment, which allows one to calculate pairwise structure alignments using different methods, (ii) Structure Quality, which shows a slider graphic that compares important global quality indicators for a given structure, and (iii) Symmetry Resources, which include exploratory tools that display global, local, and helical symmetry amongst subunits. Users can search using the top menu bar by introducing the name of the protein or the source organism, or even build complex search combinations with the "Advanced Search" interface. The platform also allows one to search for scientific

A

Function

Binds water, Ca^{2+}, Na^{+}, K^{+}, fatty acids, hormones, bilirubin and drugs. Its main function is the regulation of the colloidal osmotic pressure of blood. Major zinc transporter in plasma, typically binds about 80% of all plasma zinc (By similarity).
Major calcium and magnesium transporter in plasma, binds approximately 45% of circulating calcium and magnesium in plasma (Probable). Potentially has more than two calcium-binding sites and might additionally bind calcium in a non-specific manner (PubMed:22677715).
The shared binding site between zinc and calcium at residue Asp-272 suggests a crosstalk between zinc and calcium transport in the blood (Probable). The rank order of affinity is zinc > calcium > magnesium (Probable). Binds to the bacterial siderophore enterobactin and inhibits enterobactin-mediated iron uptake of E.coli, and may thereby limit the utilization of iron and growth of enteric bacteria such as E.coli (PubMed:6234017).
Does not prevent iron uptake by the bacterial siderophore aerobactin (PubMed:6234017). By Similarity | 2 Publications | 1 Publication

Caution

A peptide arising from positions 165 to 173 was originally termed neurotensin-related peptide (NRP) and was thought to regulate fat digestion, lipid absorption, and blood flow. 1 Publication

B

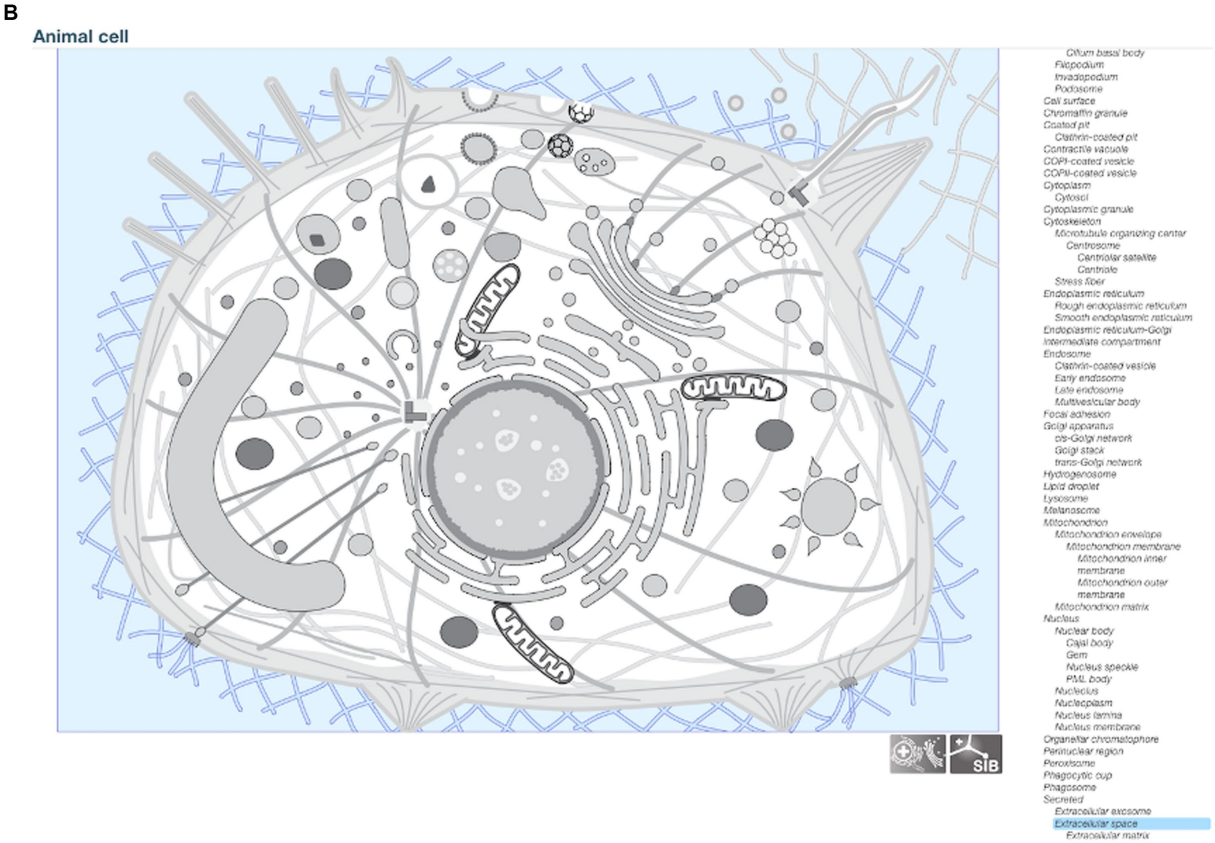


FIGURE 2
(A) Function and **(B)** cellular location of bovine serum albumin. Data accessed from UniProt, available at https://www.uniprot.org on 30/04/2023.

publications related to a given protein and access and download valuable information. As an example, we searched for bovine serum albumin (BSA), which is often used as a protein standard in laboratory experiments. BSA is also a rich source of bioactive peptides (Lafarga et al., 2017) and a relatively small protein, thus facilitating its study. The protein is classified in RCSB PDB as a transport protein included as part of the *Bos taurus* organism. Besides the protein's amino acid sequence, this online tool enables one to access the literature related to the protein and the experimental data available on it, including its structure validation, a 3D structural representation, and ligand interactions, amongst other useful information (Figure 4). The RCSB PDB is a simple tool that can be used by students to access amino acid sequences when carrying out *in silico* simulations, as described in the following sections. Furthermore, RCSB PDB includes educational resources including training courses, guides, and other free resources to promote the exploration of the world of proteins (e.g., colouring books, 3D printing models, paper models, flyers, and posters).

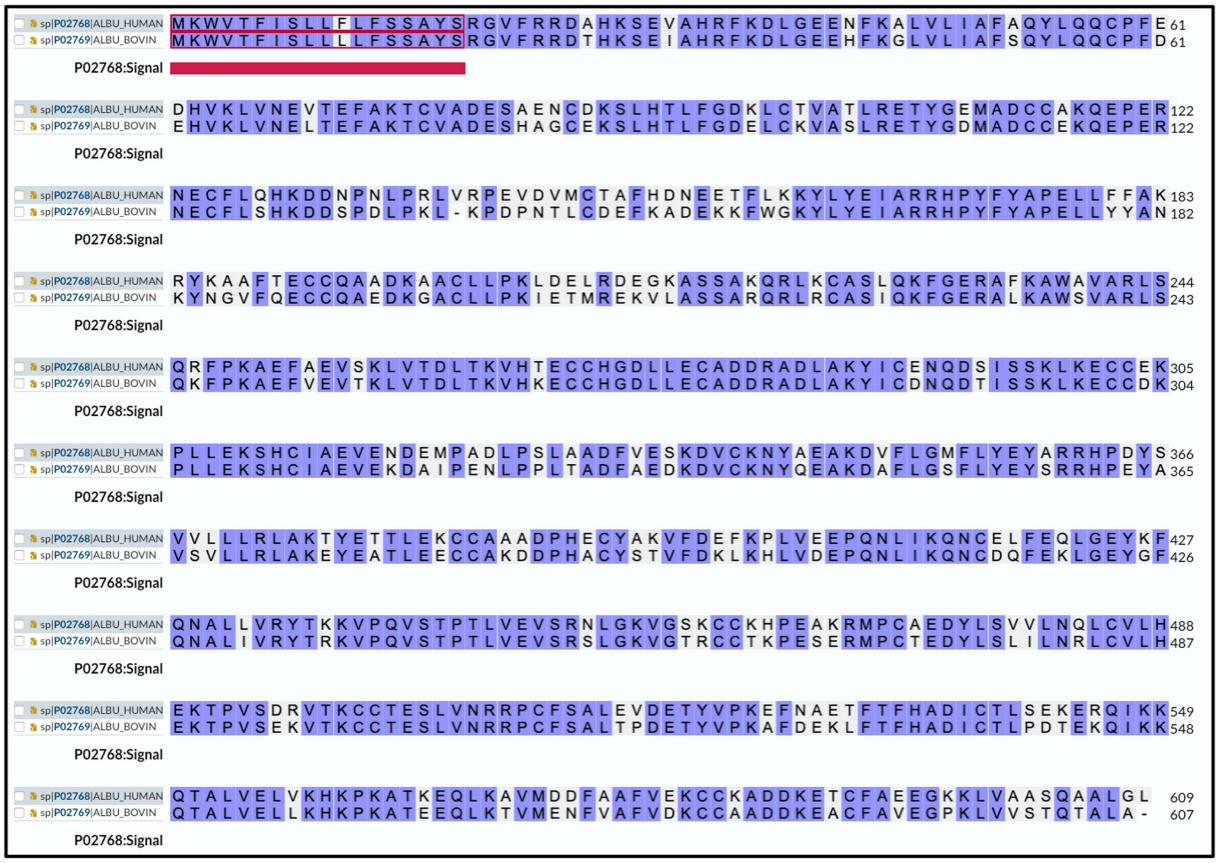


FIGURE 3
Comparison between bovine and human serum albumin using the Align tool at https://www.uniprot.org/align.

2.3 Other protein databases

There are other protein databases that contain thousands of sequences. These include PDBe (Protein Data Bank in Europe) available at https://www.ebi.ac.uk/pdbe/and wwPDB (Worldwide Protein Data Bank), available at https://www.wwpdb.org. The goal of the latter keep data and metadata for biological macromolecules freely accessible to promote basic and applied research and education. Other online protein databases include the NIH protein database of the National Library of Medicine (USA), available at https://www.ncbi.nlm.nih.gov/protein, where protein sequences and different tools can be accessed. All these protein databases, together with the ones mentioned above, allow many different processes to be simulated; these would otherwise be extremely complex to study in a laboratory. For example, working online allows one to simulated processes using raw materials in a location where they are uncommon or not unavailable raw materials are used in a location where they are uncommon or are not available.

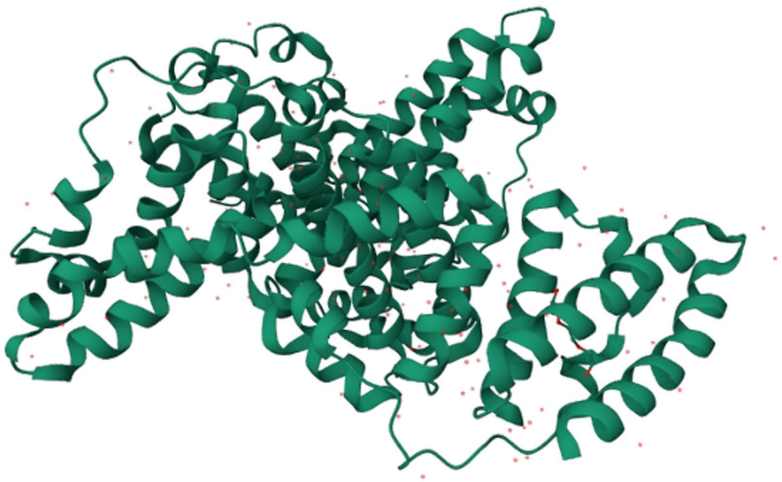
3 *In silico* proteolysis prediction

3.1 Expasy PeptideCutter and PeptideMass

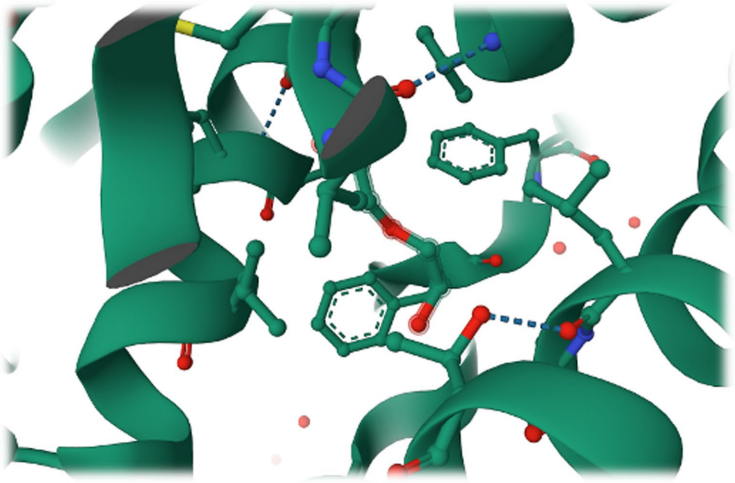
Expasy is the bioinformatics resource portal of the SIB Swiss Institute of Bioinformatics. It is a comprehensive portal that provides access to over 160 databases and online software tools to support different scientific activities including proteomics and medical chemistry (Duvaud et al., 2021). The following section discusses several useful tools that are available via the Expasy portal. PeptideCutter[1] predicts potential cleavage sites cleaved by proteases in a given protein sequence. This can be accessed either via an amino acid sequence or an identified protein such as ALBU_BOVIN. The software allows one to select an enzyme from a list of various commercially available proteases and returns the query sequence with the possible cleavage sites, as shown in Figure 5A. The tool includes a summary of the cleavage sites for the different enzymes selected. These options offer great potential for use in teaching activities, especially since the cleavage sites of a specific enzyme are key to obtaining the desired peptide fractions. Furthermore, these tools allow one to observe the differences between enzymes as a graphic presentation. For example, for the enzyme Thermolysin (EC 3.4.24.27), the Expasy portal reads "Thermolysin preferentially cleaves sites with bulky and aromatic residues (Ile, Leu, Val, Met, Phe) in position P1'. Cleavage is favoured with aromatic sites in position P1 but hindered with acidic residues in position P1. Pro blocks when located in position P2' but not when found in position P1" (see Figure 5B). PeptideCutter can be used together with PeptideMass, also developed by the SIB Swiss Institute of Bioinformatics and available at https://www.expasy.org/

1 https://web.expasy.org/peptide_cutter/

A



B



C

>4F5S_1|Chains A, B|Serum albumin|Bos taurus (9913)
DTHKSEIAHRFKDLGEEHFKGLVLIAFSQYLQQCPFDEHVKLVNELTEFAKTCVADESHAGCEKSLHTLFGDELCKV
ASLRETYGDMADCCEKQEPERNECFLSHKDDSPDLPKLKPDPNTLCDEFKADEKKFWGKYLYEIARRHPYFYAPELL
YYANKYNGVFQECCQAEDKGACLLPKIETMREKVLTSSARQRLRCASIQKFGERALKAWSVARLSQKFPKAEFVEVT
KLVTDLTKVHKECCHGDLLECADDRADLAKYICDNQDTISSKLKECCDKPLLEKSHCIAEVEKDAIPENLPPLTADF
AEDKDVCKNYQEAKDAFLGSFLYEYSRRHPEYAVSVLLRLAKEYEATLEECCAKDDPHACYSTVFDKLKHLVDEPQN
LIKQNCDQFEKLGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGTRCCTKPESERMPCTEDYLSLILNRLCVL
HEKTPVSEKVTKCCTESLVNRRPCFSALTPDETYVPKAFDEKLFTFHADICTLPDTEKQIKKQTALVELLKHKPKAT
EEQLKTVMENFVAFVDKCCAADDKEACFAVEGPKLVVSTQTALA

FIGURE 4
(A) 3D structure, **(B)** ligand interaction, and **(C)** amino acid sequence of bovine serum albumin. Data accessed from RCSB PDB at https://www.rcsb.org on 30/04/2023.

resources/peptidemass. This tool enables one to cleave a protein sequence with a chosen enzyme and computes the masses and amino acid sequences of the generated peptides. The user can select a protein sequence by either introducing the entire amino acid sequence or entering the UniProt protein identifier ID (e.g., ALBU_BOVIN). The user can select different options including the protease being tested, the maximum number of missed cleavages permitted, ot the size of the peptides shown, as one might be interested in either small or big peptides, but not both (Figure 6).

3.2 Other tools to simulate enzymatic hydrolysis

DeepDigest predicts protein cleavage using deep learning (Yang et al., 2021). It was designed to discriminate between correct and incorrect peptide identification in analytical determination but can also be used to help select proteases, not just for shotgun proteomic experiments but also for the release of bioactive peptides. In addition, the RPG Rapid Peptides

A

PeptideCutter

You have selected the protein ALBU_BOVIN (P02769) from UniProtKB/Swiss-Prot :

Albumin precursor (BSA) (Allergen Bos d 6)

The sequence to investigate:

```
       10         20         30         40         50         60
MKWVTFISLL LLFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA FSQYLQQCPF

       70         80         90        100        110        120
DEHVKLVNEL TEFAKTCVAD ESHAGCEKSL HTLFGDELCK VASLRETYGD MADCCEKQEP

      130        140        150        160        170        180
ERNECFLSHK DDSPDLPKLK PDPNTLCDEF KADEKKFWGK YLYEIARRHP YFYAPELLYY

      190        200        210        220        230        240
ANKYNGVFQE CCQAEDKGAC LLPKIETMRE KVLASSARQR LRCASIQKFG ERALKAWSVA

      250        260        270        280        290        300
RLSQKFPKAE FVEVTKLVTD LTKVHKECCH GDLLECADDR ADLAKYICDN QDTISSKLKE

      310        320        330        340        350        360
CCDKPLLEKS HCIAEVEKDA IPENLPPLTA DFAEDKDVCK NYQEAKDAFL GSFLYEYSRR

      370        380        390        400        410        420
HPEYAVSVLL RLAKEYEATL EECCAKDDPH ACYSTVFDKL KHLVDEPQNL IKQNCDQFEK

      430        440        450        460        470        480
LGEYGFQNAL IVRYTRKVPQ VSTPTLVEVS RSLGKVGTRC CTKPESERMP CTEDYLSLIL

      490        500        510        520        530        540
NRLCVLHEKT PVSEKVTKCC TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP

      550        560        570        580        590        600
DTEKQIKKQT ALVELLKHKP KATEEQLKTV MENFVAFVDK CCAADDKEAC FAVEGPKLVV

STQTALA
```

The sequence is 607 amino acids long.

Name of enzyme	No. of cleavages	Positions of cleavage sites
Thermolysin	165	3 5 6 8 9 10 11 12 15 20 21 31 34 42 45 46 47 48 49 50 54 59 63 65 66 73 77 78 83 89 92 93 100 101 103 111 125 126 138 145 151 156 161 165 171 177 180 186 187 193 198 200 204 207 211 212 213 216 220 223 225 228 232 233 235 238 239 241 248 251 256 257 263 273 276 280 283 286 293 297 305 306 312 313 327 329 332 348 349 352 353 364 365 367 368 369 371 372 379 384 390 395 396 399 402 403 409 410 417 420 425 428 429 430 431 440 445 446 452 455 475 477 478 479 482 484 485 491 495 503 504 510 512 513 523 524 528 529 531 533 545 550 551 552 555 561 566 569 570 573 574 575 576 577 582 583 590 591 592 597 598 599 604 605 606

B

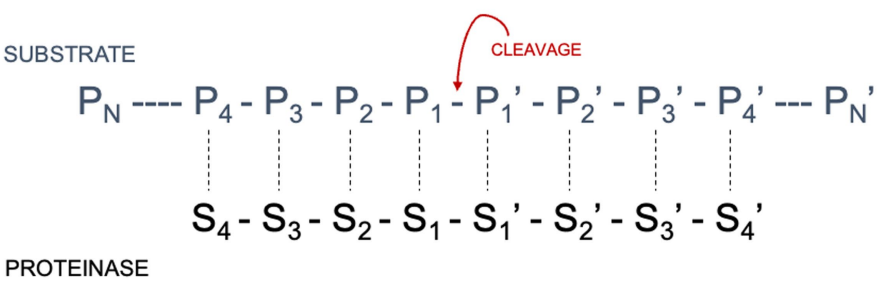


FIGURE 5
(A) Estimation via PeptideCutter of the cleavage sites of bovine serum albumin using thermolysin. **(B)** Schematic representation of enzyme-substrate complex with eight binding sites according to Schechter and Berger (1968).

Generator is software dedicated to predicting the cleavage sites of different enzymes on a given protein. It is a python tool that follows the standards for software development with continuous Gitlab integration.[2] One interesting aspect of this tool is that the hydrolysis can be carried out either in concurrent mode, assuming that several enzymes are available at the same time in the reactor, or in sequential mode, where the protein will be digested by the different enzymes one by one. Both methods are commonly used in the scientific literature (Lu et al., 2010; Lafarga et al., 2016). The option of selecting a concurrent mode is not available in other tools. The RPG Rapid Peptides Generator currently incorporates 43 different enzymes (and chemicals) including bromelain (EC 3.4.22.33) and papain (EC 3.4.22.2), which are not available in the Expasy PeptideCutter. In addition, there are some other tools that allow the release of bioactive peptides from a given protein to be estimated; usually, these tools are included in a protein or peptide database such as those described in the following section.

2 https://gitlab.pasteur.fr/nmaillet/rpg

You have selected ALBU_BOVIN (P02769) from UniProtKB/Swiss-Prot:

Albumin precursor (BSA) (Allergen Bos d 6)
Signal and propep in positions 1-24 have been removed.

- Chain Albumin at positions 25 - 607 [Theoretical pI: 5.60 / Mw (average mass): 66432.96 / Mw (monoisotopic mass): 66389.86]

mass	position	#MC	modifications		peptide sequence
2304.9716	456-475	0			VGTRCCTKPESERMPCTEDY
1826.8225	333-348	0			AEDKDVCKNYQEAKDA
1653.6301	112-125	0			ADCCEKQEPERNEC
1521.7795	314-327	0			AEVEKDAIPENLPP
1512.7230	354-364	0			LYEYSRRHPEY
1351.6852	127-138	0			LSHKDDSPDLPK
1333.6304	534-545	0			ADICTLPDTEKQ
1162.5990	514-523	0			LTPDETYVPK
1146.6742	432-440	0			VRYTRKVPQ
1140.4924	264-273	0			VHKECCHGDL
984.4455	104-111	0	PHOS: 107	1064.4118	LRETYGDM
974.4577	35-42	0			FKDLGEEH
935.4325	298-305	0			LKECCDKP
870.3696	496-503	0			VTKCCTES
851.4145	67-73	0			VNELTEF
848.3931	411-417	0			IKQNCDQ
847.5036	242-248	0			LSQKFPK
829.4050	25-31	0	PHOS: 29	909.3713	DTHKSEI
811.3832	373-379	0			AKEYEAT
811.3654	94-100	0			FGDELCK
808.3141	287-293	0			ICDNQDT
799.4322	166-171	0			ARRHPY
784.4199	139-145	0			LKPDPNT
757.2644	188-193	0			FQECCQ
754.3440	146-151	0			LCDEFK
751.2927	584-590	0	SUCC: 587	851.3087	ADDKEAC
750.4984	556-561	0	METH: 557	764.5141	LKHKPK
744.3933	505-510	0			VNRRPC
739.3661	172-177	0			FYAPEL
724.3988	486-491	0			LHEKTP
716.3396	307-312	0			LEKSHC
701.3100	404-409	0			VDEPQN
700.3453	157-161	0			FWGKY
682.3154	385-390	0			AKDDPH
676.3876	258-263	0			VTDLTK
676.3624	447-452	0			VEVSRS
666.3205	181-186	0			ANKYNG
617.3981	546-550	0			IKKQT
596.2054	380-384	0			LEECC
594.2552	84-89	0	PHOS: 89	674.2215	AGCEKS
590.3144	152-156	0			ADEKK
588.2810	55-59	0			LQQCP
577.2464	562-566	0			ATEEQ
575.3399	252-256	0			VEVTK
567.2265	578-582	0			VDKCC
563.2970	208-211	0			MREK
558.2154	79-83	0	PHOS: 82	638.1817	ADESH

mass	position	#MC	modifications		peptide sequence
547.2147	60-63	0			FDEH
544.2402	51-54	0			FSQY
544.2072	391-395	0			ACYST
538.2507	421-425	0			LGEYG
538.2507	525-528	0			FDEK
537.2919	162-165	0			LYEI
535.2722	600-604	0			VSTQT
530.3157	217-220	0			ARQR
529.2980	593-597	0			VEGPK
519.2409	194-198	0			AEDKG
508.2514	229-232	0			FGER
504.2664	441-445	0	PHOS: 442, 443,445	744.1654	VSTPT
476.2099	277-280	0			ADDR
470.3337	201-204	0			LLPK
462.2558	492-495	0			VSEK
458.2285	178-180	0			LYY
434.2609	294-297	0	PHOS: 296	514.2272	ISSK
423.2238	418-420	0			FEK
422.2068	74-77	0			AKTC
409.2081	397-399	0			FDK
408.1877	426-428	0			FQN
402.2459	480-482	0			LNR
397.2558	400-402	0			LKH
393.1438	571-573	0			MEN
391.2122	221-223	0			LRC
388.2554	226-228	0	SUCC: 228	488.2715	IQK
383.2150	32-34	0			AHR
381.2132	284-286	0			AKY
370.2085	90-92	0			LHT
366.1659	249-251	0			AEF
364.1537	274-276	0			LEC
363.1663	236-238	0			AWS
362.1922	205-207	0			IET
361.2445	567-569	0	PHOS: 569	441.2109	LKT
360.2129	553-555	0			VEL
352.1503	330-332	0			ADF
351.2027	43-45	0			FKG
318.1659	281-283	0			ADL
317.2183	453-455	0			LGK
303.1451	532-533	0			FH
288.2030	370-371	0			LR
276.1554	350-352	0			LGS
267.1339	530-531	0			FT
264.1190	214-216	0			ASS
260.1968	234-235	0			LK
253.1183	511-512	0	PHOS: 512	333.0846	FS
246.1812	64-65	0			VK

FIGURE 6
Peptides released after the *in silico* cleavage of bovine serum albumin using thermolysin. Estimation carried out using PeptideMass, available at https://www.expasy.org/resources/peptidemass.

4 Bioactive peptide databases

Research interest in bioactive peptides has increased exponentially over the last two decades. This has led to a vast number of bioactive sequences being discovered, mainly from food sources and food co-products. These sequences are stored in bioactive peptide databases, which are powerful sources of information that aid in the selection of raw materials, proteases, and hydrolysates to obtain bioactive peptides. In terms of training, these databases enable the students to observe the differences in the amino acid sequences between different bioactivity groups (e.g., antihypertensive, antioxidant, or antidiabetic peptides). There are generic databases, which include information on peptides with different bioactivities and from a wide range of raw materials, and others that are more specific and focus on just one type of peptide (e.g., antimicrobial peptides) or a specific parent protein (e.g., peptides derived from milk proteins). Some of the more widely used peptide databases are listed below.

4.1 MBPDB: milk bioactive peptide database

This database focuses on peptides that are encrypted within dairy proteins. Dairy proteins are one of the main sources of peptides because milk, apart from basic nutrients, provides proteins that are degraded by native proteases into peptides that have biological activity. The MBPDB database, which is available online at https://mbpdb.nws.oregonstate.edu, was created from hundreds of published scientific papers that identified peptides with *in vitro* and *in vivo* biological activity following the hydrolysis of milk proteins (Nielsen et al., 2017). The database includes an MBPDB Search option, where the user can include any amino acid sequence and the software returns a list of publications where that peptide was used together with the main outcomes of each study, particularly the proven bioactivity. As an example, after searching for the tri-peptide IPP, which is a known bioactive sequence, the software returned the list of publications shown in Table 1. The webpage is easy to navigate and provides useful information on bioactive peptides. However, it is limited to peptides derived from milk proteins.

TABLE 1 Information about the peptide IPP available in the MBPDB database.

Protein ID	Peptide	Category	Protein source	Species	Interval	Function	Reference
P02668	IPP	Dairy	Kappa-casein	*Bos taurus*	129–131	Stimulates trabecular bone growth	Narva et al. (2004)
P02668	IPP	Dairy	Kappa-casein	*Bos taurus*	129–131	Stimulates proliferation	Huttunen et al. (2007)
P02668	IPP	Dairy	Kappa-casein	*Bos taurus*	129–131	ACE-inhibitory*	Adams et al. (2020), Donkor et al. (2007), Jäkälä et al. (2010), Jing et al. (2014), Nakamura et al. (1995), Tagliazucchi et al. (2016), Yamada et al. (2015)
P02668	IPP	Dairy	Kappa-casein	*Bos taurus*	129–131	Anti-inflammatory	Adams et al. (2020), Chakrabarti and Wu (2015)
P02668	IPP	Dairy	Kappa-casein	*Bos taurus*	129–131	Nitric oxide liberation	Adams et al. (2020)
P02668	IPP	Dairy	Kappa-casein	*Bos taurus*	129–131	Enhances insulin signalling	Chakrabarti and Wu (2015)
P02668	IPP	Dairy	Kappa-casein	*Bos taurus*	129–131	Antioxidant	Chakrabarti et al. (2017)
P02668	IPP	Dairy	Kappa-casein	*Bos taurus*	129–131	Antihypertensive	Okamoto et al. (2020)

*ACE-inhibitory peptides inhibit the enzyme ACE-I (EC 3.4.15.1), which is related with high blood pressure.

4.2 DFBP: database of food-derived bioactive peptides

DFBP is freely available at http://www.cqudfbp.net. It consists of a database containing over 6,200 peptides isolated from different sources (Qin et al., 2022). In this database, the bioactive peptides are divided into 31 bioactivities, which include inhibitors of the enzymes ACE-I (EC 3.4.15.1), renin (EC 3.4.23.15), DPP-IV (EC 3.4.14.5), α-amylase (EC 3.2.1.1), and β-glucosidase (EC 3.2.1.20) as well as antioxidant, antimicrobial, antithrombotic, and antiviral peptides. The most abundant peptides are ACE inhibitors (N = 1,961), antioxidants (N = 1,032), and antimicrobial peptides (N = 476). DFBP is more than just a list of bioactive peptides as every peptide in the database is characterised by 30 different attributes including their main physicochemical and functional properties, as well as by their stability and toxicity. It also includes a special section containing those peptides that have two or more proven bioactivities, known as multifunctional peptides. When searching for the IPP peptide, the software identifies it as a multifunctional peptide with ACE-inhibitory and antihypertensive properties. Along with the peptide's main physicochemical properties (e.g., its theoretical mass, net charge, isoelectric point, GRAVY, hydrophilic residue ration, and length), the software returns information on its predicted bitterness and a list of different proteins in which IPP has been identified. In this case, more than 1,400 proteins that include IPP in their amino acid sequence are available on the database. The list is extensive, including proteins from animal, microbial, and plant sources. Finally, the tool returns a table with cross-references so that the one can see the results for this same peptide in a different database such as the above-mentioned MBPDB, or in BIOPEP (described below).

In addition to the peptide database, DFBP has a range of prediction and calculation tools that include HotSpot Search,[3] which can be used to search for specific sequences in a list of proteins or group of proteins selected by the user, and the EHP-Tool,[4] which is used to simulate the enzyme hydrolysis of proteins. We hydrolysed bovine serum albumin using papain and the software returned a list of all the potential peptides generated, highlighting those that are already available in their database as bioactive peptides. Other tools include the BPP-Tool,[5] which can be used to predict the bitterness of a given amino acid sequence, the Peptide Calculator,[6] which estimates the different useful properties of the peptides, and AASD-Tool,[7] which is to obtain the peptide descriptors. These can be then used for molecular modelling and bioinformatics prediction.

4.3 BIOPEP-UWM

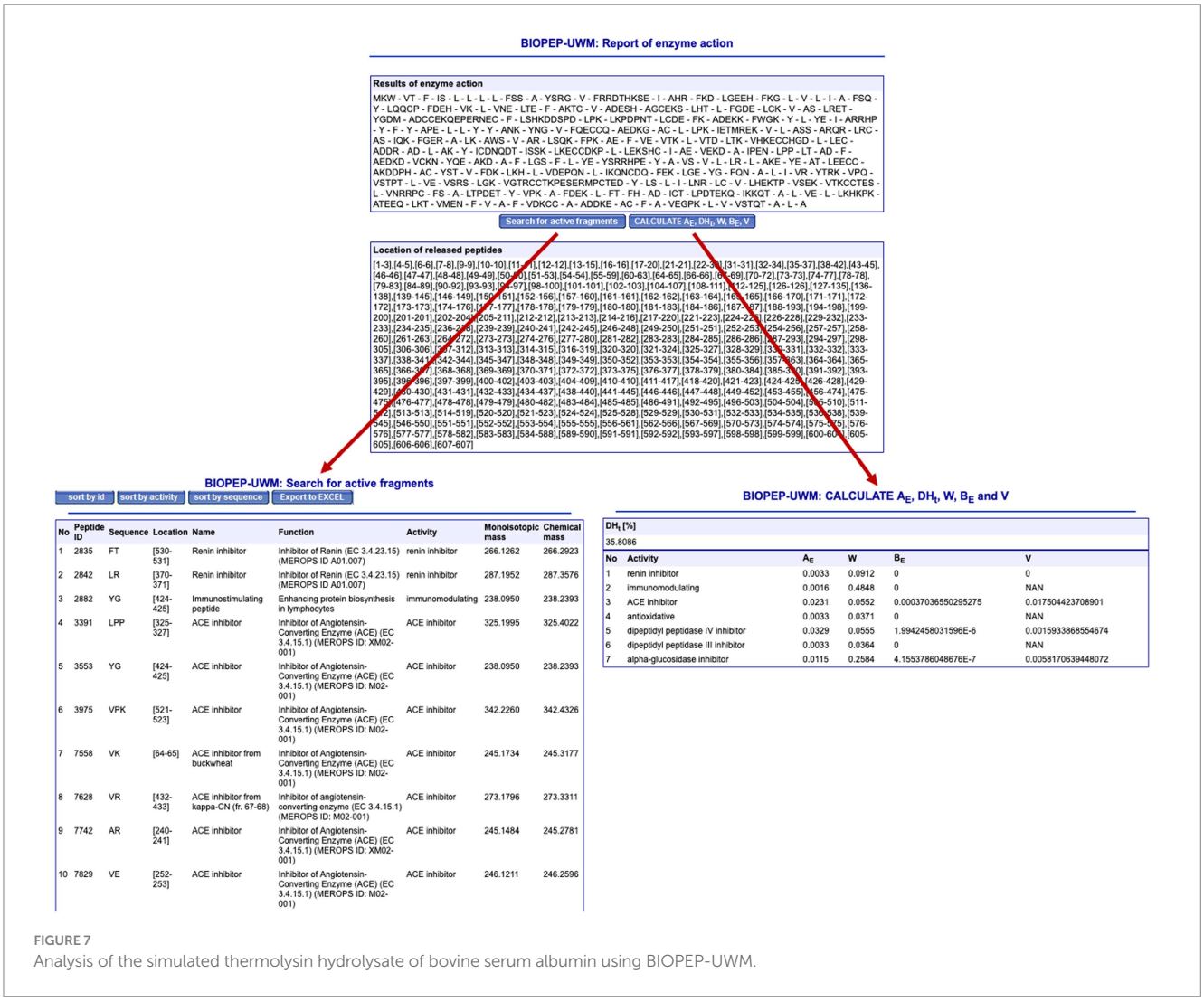
The BIOPEP-UWM database, formerly known as BIOPEP, is probably the most popular peptide database. It focuses especially on

3 http://www.cqudfbp.net/blast/data_input.jsp
4 http://www.cqudfbp.net/enzymes/hydrolysis_tools/dataInput.jsp
5 http://www.cqudfbp.net/bitterPrediction/tools/DataInput.jsp
6 http://www.cqudfbp.net/peptide_calculator/data_input.jsp
7 http://www.cqudfbp.net/uploadanddownload/upload.jsp

bioactive peptides derived from foods and contains over 4,600 bioactive peptide sequences and more than 750 proteins (although the database is being continuously updated). It was developed by the Chair of Food Biochemistry at Warmia and Mazury University in Poland and is freely available at https://biochemia.uwm.edu.pl/en/biopep-uwm-2/ (Minkiewicz et al., 2019).

The database is divided into four sections: proteins, bioactive peptides, allergenic proteins, and sensory peptides. When searching for the tripeptide IPP in the bioactive peptides section of BIOPEP-UWM, the system returned 32 peptide sequences that include not only IPP but also other longer bioactive peptides in which the IPP sequence is encrypted (e.g., AIPP, IAIPP, MAIPPKK). Focusing just on IPP, the database identified four bioactivities for this peptide including α-amylase inhibition (ID 10286), β-glucosidase inhibition (ID 10311), anti-inflammatory properties (ID 9537), and ACE-inhibitory properties (ID 3522), together with the scientific publications that support their bioactivity. In terms of proteins, the tool allows one to search for useful information on the more than 750 proteins registered, including their chemical mass, number of amino acid residues, sequence of amino acids, and references for the scientific publications in which the protein was used. In addition, BIOPEP-UWM contains a list of over 130 allergenic proteins with their epitopes. In terms of sensory peptides and amino acids, BIOPEP-UWM includes a database with information on their taste and bioactivity data, amongst other useful details (Iwaniak et al., 2016).

One of the most interesting aspects of BIOPEP-UWM is the "Analysis" area, where the user can perform various calculations and estimations on a selected protein. These includes calculating profiles of potential biological activity, which are lists of known bioactive peptides contained within a protein sequence. We calculated the profile of potential ACE-inhibitory activity for bovine serum albumin (ID 1729) and obtained a list with all the ACE-inhibitors that are contained within it, together with their position in the protein (Figure 7). These include, amongst other, ALKAWSVAR, RL, RY, LY, VF, LVL, LPP, and AY. Another interesting option provided in the "Analysis" area is the calculation of the quantitative A and B values. The A parameter refers to the occurrence frequency of bioactive fragments in the protein chain, calculated as the number of fragments with a given activity in the protein divided by the number of amino acid residues in the protein chain. In turn, the B parameter refers to the potential biological activity of the protein fragments; this is calculated using the number of repetitions of bioactive fragments in a protein, their half-maximal activity, their half-maximal inhibition, and the total number of amino acid residues.



No	Peptide ID	Sequence	Location	Name	Function	Activity	Monoisotopic mass	Chemical mass
1	2835	FT	[530-531]	Renin inhibitor	Inhibitor of Renin (EC 3.4.23.15) (MEROPS ID A01.007)	renin inhibitor	266.1262	266.2923
2	2842	LR	[370-371]	Renin inhibitor	Inhibitor of Renin (EC 3.4.23.15) (MEROPS ID A01.007)	renin inhibitor	287.1952	287.3576
3	2882	YG	[424-425]	Immunostimulating peptide	Enhancing protein biosynthesis in lymphocytes	immunomodulating	238.0950	238.2393
4	3391	LPP	[325-327]	ACE inhibitor	Inhibitor of Angiotensin-Converting Enzyme (ACE) (EC 3.4.15.1) (MEROPS ID: XM02-001)	ACE inhibitor	325.1995	325.4022
5	3553	YG	[424-425]	ACE inhibitor	Inhibitor of Angiotensin-Converting Enzyme (ACE) (EC 3.4.15.1) (MEROPS ID: M02-001)	ACE inhibitor	238.0950	238.2393
6	3975	VPK	[521-523]	ACE inhibitor	Inhibitor of Angiotensin-Converting Enzyme (ACE) (EC 3.4.15.1) (MEROPS ID: M02-001)	ACE inhibitor	342.2260	342.4326
7	7558	VK	[64-65]	ACE inhibitor from buckwheat	Inhibitor of Angiotensin-Converting Enzyme (ACE) (EC 3.4.15.1) (MEROPS ID: M02-001)	ACE inhibitor	245.1734	245.3177
8	7628	VR	[432-433]	ACE inhibitor from kappa-CN (fr. 67-68)	Inhibitor of angiotensin-converting enzyme (EC 3.4.15.1) (MEROPS ID: M02-001)	ACE inhibitor	273.1796	273.3311
9	7742	AR	[240-241]	ACE inhibitor	Inhibitor of Angiotensin-Converting Enzyme (ACE) (EC 3.4.15.1) (MEROPS ID: XM02-001)	ACE inhibitor	245.1484	245.2781
10	7829	VE	[252-253]	ACE inhibitor	Inhibitor of Angiotensin-Converting Enzyme (ACE) (EC 3.4.15.1) (MEROPS ID: M02-001)	ACE inhibitor	246.1211	246.2596

No	Activity	A_E	W	B_E	V
1	renin inhibitor	0.0033	0.0912	0	0
2	immunomodulating	0.0016	0.4848	0	NAN
3	ACE inhibitor	0.0231	0.0552	0.00037036550295275	0.017504423708901
4	antioxidative	0.0033	0.0371	0	NAN
5	dipeptidyl peptidase IV inhibitor	0.0329	0.0555	1.9942458031596E-6	0.0015933868554674
6	dipeptidyl peptidase III inhibitor	0.0033	0.0364	0	NAN
7	alpha-glucosidase inhibitor	0.0115	0.2584	4.1553786048676E-7	0.0058170639448072

FIGURE 7
Analysis of the simulated thermolysin hydrolysate of bovine serum albumin using BIOPEP-UWM.

BIOPEP-UWM can also be used to simulate the release of bioactive fragments from a given parent protein. This is carried out in the "Enzyme(s) Action" section. The software allows you to use one of the protein sequences already available in the database or to introduce a new protein for analysis. Several proteases are already included and up to three different enzymes can be used simultaneously. Figure 6 shows the simulated hydrolysis of bovine serum albumin using thermolysin. The software calculates the potential cleavage bonds and gives the user two options. Firstly, the user can identify which peptides (from those that have been predicted) have shown bioactivity in the past, independently of their source or bioactivity type. In this case, the software identified 48 bioactive peptides (only the first 10 are shown) which have a variety of bioactivities; these include inhibitors of the enzymes DPP-III, ACE, renin, α-glucosidase, and DPP-IV, immunostimulant peptides, and antioxidant peptides. The second option provided by BIOPEP-UWM is to calculate several parameters that have been developed to facilitate protein comparison. These include W, which refers to the relative frequency of fragments released for a given activity by the selected enzyme (in this case thermolysin) and AE, which refers to the frequency of fragments released for a given bioactivity by the selected enzyme (Iwaniak et al., 2020). The calculation of other parameters, such as BE and V, involves the use of their EC50 or IC50 values, respectively, and therefore might not be available for many of the peptides. All these options could be used by students to compare different proteins, proteases, and protease combinations to identify which raw material, and which enzyme, might induce the release of a greater number of bioactive peptides.

5 Other useful tools

5.1 PepDraw

PepDraw is available at https://www2.tulane.edu/~biochem/WW/PepDraw/. It was developed by the Wimley Lab at Tulane University (LA, United States). This tool can be used to draw the primary structure of a peptide and calculate certain theoretical properties, including its isoelectric point, net charge, hydrophobicity, and mass, amongst other parameters. The isoelectric point of a protein is the key parameter that needs to be calculated to recover proteins by precipitation at the industrial scale. Estimating this parameter online allows students to design processes and/or estimate protein recoveries *in silico*. The images created can be downloaded in png format at high quality. As an example, the 2D structure of the peptide VGTRCCTKPESERMPCTEDY is shown; this, was predicted to be released during the hydrolysis of bovine serum albumin using thermolysin in the previous section (Figure 8A). The tool is available free of charge for academic use.

5.2 PEP-FOLD

PEP-FOLD is available at https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/. It is a *de novo* approach that predicts the structure of a peptide from its amino acid sequence (Lamiable et al., 2016). Using this tool, it is also possible to generate native-like conformations of peptides interacting with a protein when the

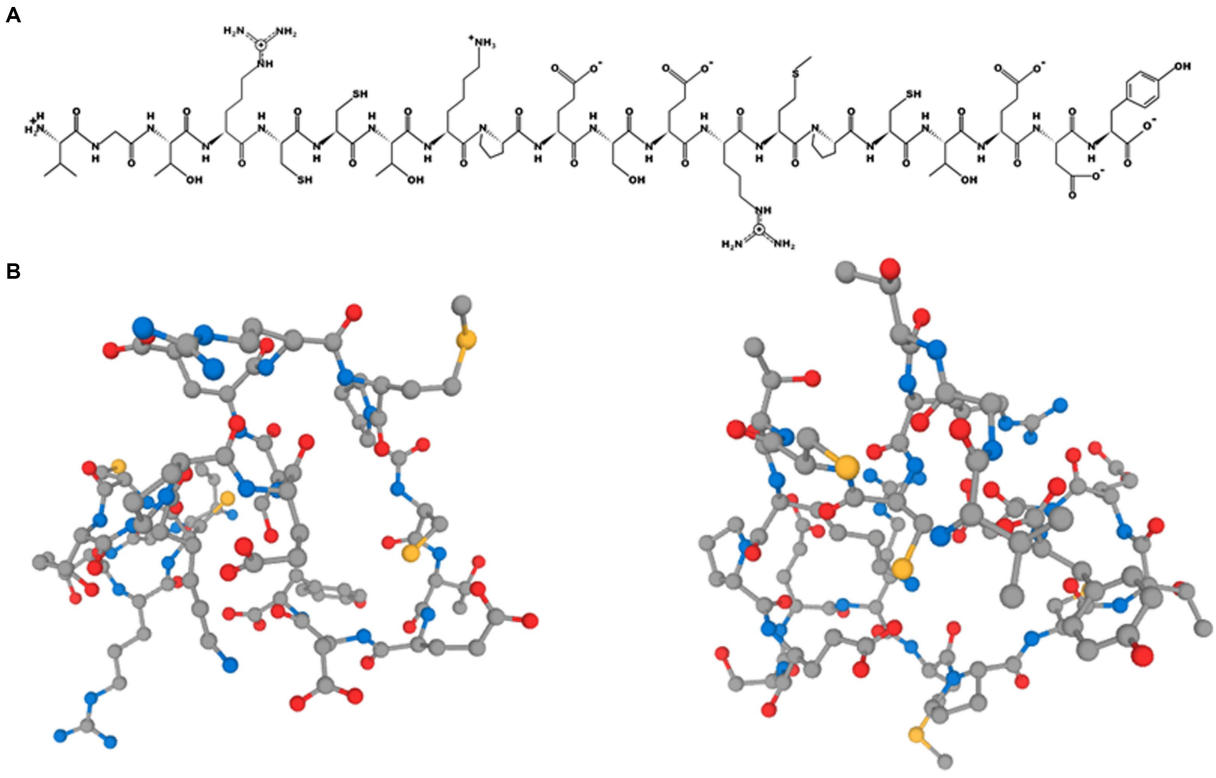


FIGURE 8
(A) 2D and **(B)** 3D representation of the peptide VGTRCCTKPESERMPCTEDY using PepDraw and PEP-FOLD, respectively.

interaction site is known in advance. The process takes a few minutes and returns 3D images as well as useful information on the peptide sequence. As an example, the 3D structure of the peptide VGTRCCTKPESERMPCTEDY is shown; this was predicted to be released during the hydrolysis of bovine serum albumin using thermolysin (Figure 8B). PEP-FOLD is limited to amino acid sequences between 5 and 50 residues long, which is enough for teaching enzymatic proteolysis and various topics related to bioactive peptides, given that most of them are much shorter than 50 residues in length.

5.3 Expasy tools

Several useful tools for the teaching and training of enzymatic proteolysis have been developed by the SIB Swiss Institute of Bioinformatics, all of which are available at the Expasy portal. Expasy Compute pI/MW[8] is a tool that allows one to calculate the isoelectric point and the molecular weight of a protein sequence. This information is useful for developing protein isolation processes based on isoelectric precipitation (Villaró et al., 2023) or for estimating the approximate region where a protein may be found in an electrophoresis gel. ProtParam,[9] computes several physicochemical parameters for a given amino acid sequence including its molecular weight, amino acid composition, extinction coefficient, aliphatic index, or grand average of hydropathicity (GRAVY), amongst others. For bovine serum albumin, the software returned its molecular weight (69190.28), atomic composition (C3068H4821N815O926S39), total number of atoms (9669), estimated half-life (10–30 h), instability index (40.33, unstable), aliphatic index (77.59), GRAVY (−0.434), and amino acid composition (7.9% Ala, 4.3% Arg, 2.3% Asn, 6.6% Asp., 5.6% Cys, 3.3% Gln, 9.7% Glu, 2.8% Gly, 2.8% His, 2.5% Ile, 10.7% Leu, 9.9% Lys, 0.8% Met, 5.0% Phe, 4.6% Pro, 5.3% Ser, 5.6% Thr, 0.5% Trp, 3.5% Tyr, and 6.3% Val). The tool also provided a summary of the total number of negatively charged (Asp + Glu) and positively charged (Arg + Lys) residues: 99 and 86, respectively.

5.4 PeptideRanker

PeptideRanker, available at http://distilldeep.ucd.ie/PeptideRanker/, gives a score for a given peptide (or list of peptides) that predicts the probability of it being bioactive (Mooney et al., 2012). PeptideRanker was trained using different peptide databases mainly antimicrobial peptide databases; these included approximately 19,000 unique peptide sequences. It includes two independent neural network predictors for short and long peptides (larger than 20 amino acids); these predict those peptides that are more likely to be bioactive from amongst a set of peptides. The software is available free of charge for academic use and has been effectively used to aid in the identification of bioactive peptides from different food sources (Lafarga et al., 2015; de Fátima Garcia et al., 2020).

5.5 Prediction of toxicity

Peptides released during food digestion do not pose a risk to human health. However, certain peptides are toxic and their oral administration at high doses (or even at low concentrations via intravenous administration) might be dangerous. For this reason, several online tools have been developed to estimate the potential toxicity of a peptide sequence. ToxIBTL, freely accessible at https://server.wei-group.net/ToxIBTL/, is based on novel deep learning frameworks (Wei et al., 2022). ToxinPred, available free of charge at http://crdd.osdd.net/raghava/toxinpred/, was developed to predict potentially toxic peptides using a dataset of over 1,800 toxic peptides shorter than 35 residues in length (Gupta et al., 2013). This software is based on models using machine learning techniques and a quantitative matrix using of the different peptide properties related to toxicity. A third online software tool for predicting toxicity is CSM-TOXIN, which is available at https://biosig.lab.uq.edu.au/csm_toxin/. The software relies solely on the primary protein structure and encodes the sequence information using deep learning techniques (Morozov et al., 2023). It is important to point out that, although these tools enable one to estimate potential toxicity, the safety of a peptide must be validated *in vitro* and *in vivo* prior to being administered to humans. This tool allows students to estimate which enzymes are more suitable for producing or avoiding the release of toxic peptides from a given protein source. Furthmore, students can identify (when designing the process) which proteins might pose a potential toxicity risk.

5.6 Prediction of allergenicity

Proteins and peptides must be evaluated for their allergenic potential prior to being commercialised because both pose a risk of inducing allergic responses. In this regard, there are various online tools that can be used during the training and education of chemical engineers. These include (i) AllerCatPro, available at https://allercatpro.bii.a-star.edu.sg, which predicts potential allergenicity based on the similarity of the 3D structure of proteins and their amino acid sequences (Maurer-Stroh et al., 2019), (ii) AlgPred, available at http://crdd.osdd.net/raghava/algpred/, which is based on support vector machine methods using the amino acid and dipeptide composition (Saha and Raghava, 2006), and (iii) Allermatch,[10] which predicts allergenicity following the recommendations given in the Codex Alimentarius and the FAO/WHO Expert consultation on the allergenicity of foods (Fiers et al., 2004). In addition, ChAIPred, which is available at https://webs.iiitd.edu.in/raghava/chalpred/, can be used to estimate the potential allergenicity of chemical compounds that are not proteins or peptides (Sharma et al., 2021). These tools can be used for training and educational activities and during the initial screening of allergens. However, potential allergenicity must be further validated as the results given by these online tools are only estimates and predictions.

8 https://web.expasy.org/compute_pi/

9 https://www.expasy.org/resources/protparam

10 https://allermatch.org

The present paper provides a summary of several protein/peptide databases, as well as various online tools that are easy-to-use and freely available. These can be used to predict the enzymatic cleavage of proteins and the potential bioactivity, allergenicity, bioaccessibility, and other useful characteristics of proteolytic processes. Such tools could be used during training activities for chemical engineers, food scientists, biotechnologists, or chemists. Amongst the advantages of using these tools during training is the fact that they allow one to observe what happens during the process and to obtain data that would not be possible with conventional technical training. This has been recently discussed in another work, in which algorithms and design equations were suggested to allow results that chemical engineering students would otherwise find hard to identify and lead to a superficial understanding of the problem (Roman et al., 2020). Their use could be evaluated by collecting post-session student feedback and by evaluating what was learned during the training, for example, via oral presentations or a written exam. Assessing not only the academic performance of students but also their learning experience would be a useful way of evaluating the teaching and learning experience. Moreover, simulating proteolysis online would allow students to continue the simulations and estimations at home with no need for laboratory reagents and equipment.

Author contributions

SV-C: Conceptualization, Visualization, Writing – original draft. TL: Conceptualization, Funding acquisition, Visualization, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work forms part of the CLEAN·AIR (TED2021-131511A-I00) and SOLAR·FOODS (PID2022-136292OB-I00) projects, both funded by the Spanish Ministry of Science and Innovation - MCIN/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR. It was also funded by the BLUE·FUTURE (PCM_00083) project, overseen by the Regional Government of Andalusia and the European Union NextGenerationEU/PRTR. TL would like to thank PPIT-UAL, Junta de Andalucía-FEDER 2021-2027 (Programme 54.A) and the Ramon y Cajal Programme (RYC2021-031061-I) funded by MCIN/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Adams, C., Sawh, F., Green-Johnson, J. M., Jones Taggart, H., and Strap, J. L. (2020). Characterization of casein-derived peptide bioactivity: differential effects on angiotensin-converting enzyme inhibition and cytokine and nitric oxide production. *J. Dairy Sci.* 103, 5805–5815. doi: 10.3168/JDS.2019-17976

Bateman, A. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. doi: 10.1093/NAR/GKY1049

Borreguero, A. M., Valverde, J. L., García-Vargas, J. M., and Sánchez-Silva, L. (2019). Simulator-based learning in the teaching of chemical engineering. *Comput. Appl. Eng. Educ.* 27, 1267–1276. doi: 10.1002/CAE.22150

Chakrabarti, S., Liao, W., Davidge, S. T., and Wu, J. (2017). Milk-derived tripeptides IPP (Ile-pro-pro) and VPP (Val-pro-pro) differentially modulate angiotensin II effects on vascular smooth muscle cells. *J. Funct. Foods* 30, 151–158. doi: 10.1016/J.JFF.2016.12.022

Chakrabarti, S., and Wu, J. (2015). Milk-derived tripeptides IPP (Ile-pro-pro) and VPP (Val-pro-pro) promote adipocyte differentiation and inhibit inflammation in 3T3-F442A cells. *PLoS One* 10:e0117492. doi: 10.1371/JOURNAL.PONE.0117492

Charoenkwan, P., Anuwongcharoen, N., Nantasenamat, C., Hasan, M. M., and Shoombuatong, W. (2020). In silico approaches for the prediction and analysis of antiviral peptides: a review. *Curr. Pharm. Des.* 27, 2180–2188. doi: 10.2174/1381612826666201102105827

de Fátima Garcia, B., de Barros, M., and de Souza Rocha, T. (2020). Bioactive peptides from beans with the potential to decrease the risk of developing noncommunicable chronic diseases. *Crit Rev Food Sci Nutr* 61, 2003–2021. doi: 10.1080/10408398.2020.1768047

Donkor, O. N., Henriksson, A., Singh, T. K., Vasiljevic, T., and Shah, N. P. (2007). ACE-inhibitory activity of probiotic yoghurt. *Int. Dairy J.* 17, 1321–1331. doi: 10.1016/J.IDAIRYJ.2007.02.009

Duvaud, S., Gabella, C., Lisacek, F., Stockinger, H., Ioannidis, V., and Durinx, C. (2021). Expasy, the Swiss bioinformatics resource portal, as designed by its users. *Nucleic Acids Res.* 49, W216–W227. doi: 10.1093/NAR/GKAB225

Fiers, M. W. E. J., Kleter, G. A., Nijland, H., Peijnenburg, A. A. C. M., Nap, J. P., and van Ham, R. C. H. J. (2004). Allermatch™, a webtool for the prediction of potential allergenicity according to current FAO/WHO codex alimentarius guidelines. *BMC Bioinformatics* 5, 1–6. doi: 10.1186/1471-2105-5-133/TABLES/2

Gautam, A., Chaudhary, K., Kumar, R., Sharma, A., Kapoor, P., Tyagi, A., et al. (2013). In silico approaches for designing highly effective cell penetrating peptides. *J. Transl. Med.* 11:74. doi: 10.1186/1479-5876-11-74/FIGURES/6

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., and Raghava, G. P. S. (2013). In silico approach for predicting toxicity of peptides and proteins. *PLoS One* 8:e73957. doi: 10.1371/JOURNAL.PONE.0073957

Huttunen, M. M., Pekkinen, M., Ahlström, M. E. B., and Lamberg-Allardt, C. J. E. (2007). Effects of bioactive peptides isoleucine-proline-proline (IPP), valine-proline-proline (VPP) and leucine-lysine-proline (LKP) on gene expression of osteoblasts differentiated from human mesenchymal stem cells. *Br. J. Nutr.* 98, 780–788. doi: 10.1017/S0007114507744434

Iwaniak, A., Minkiewicz, P., Darewicz, M., Sieniawski, K., and Starowicz, P. (2016). BIOPEP database of sensory peptides and amino acids. *Food Res. Int.* 85, 155–161. doi: 10.1016/J.FOODRES.2016.04.031

Iwaniak, A., Minkiewicz, P., Pliszka, M., Mogut, D., and Darewicz, M. (2020). Characteristics of biopeptides released in silico from collagens using quantitative parameters. *Foods* 9:965. doi: 10.3390/FOODS9070965

Jäkälä, P., Turpeinen, A. M., Rajakari, K., Korpela, R., and Vapaatalo, H. (2010). Biological effects of casein-derived tripeptide powders are not affected by fermentation process. *Int. Dairy J.* 20, 366–370. doi: 10.1016/J.IDAIRYJ.2009.11.017

Jing, P., Qian, B., He, Y., Zhao, X., Zhang, J., Zhao, D., et al. (2014). Screening milk-derived antihypertensive peptides using quantitative structure activity relationship (QSAR) modelling and in vitro/in vivo studies on their bioactivity. *Int. Dairy J.* 35, 95–101. doi: 10.1016/J.IDAIRYJ.2013.10.009

Kumar, R., Chaudhary, K., Chauhan, J. S., Nagpal, G., Kumar, R., Sharma, M., et al. (2015). An in silico platform for predicting, screening and designing of antihypertensive peptides. *Nat. Publ. Group* 5:12512. doi: 10.1038/srep12512

Lafarga, T., Aluko, R. E., Rai, D. K., O'Connor, P., and Hayes, M. (2016). Identification of bioactive peptides from a papain hydrolysate of bovine serum albumin and assessment of an antihypertensive effect in spontaneously hypertensive rats. *Food Res. Int.* 81, 91–99. doi: 10.1016/j.foodres.2016.01.007

Lafarga, T., Álvarez, C., and Hayes, M. (2017). Bioactive peptides derived from bovine and porcine co-products: a review. *J. Food Biochem.* 41:e12418. doi: 10.1111/jfbc.12418

Lafarga, T., O'Connor, P., and Hayes, M. (2014). Identification of novel dipeptidyl peptidase-IV and angiotensin-I-converting enzyme inhibitory peptides from meat proteins using in silico analysis. *Peptides* 59, 53–62. doi: 10.1016/j.peptides.2014.07.005

Lafarga, T., O'Connor, P., and Hayes, M. (2015). In silico methods to identify meat-derived prolyl endopeptidase inhibitors. *Food Chem.* 175, 337–343. doi: 10.1016/j.foodchem.2014.11.150

Lamiable, A., Thevenet, P., Rey, J., Vavrusa, M., Derreumaux, P., and Tuffery, P. (2016). PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.* 44, W449–W454. doi: 10.1093/NAR/GKW329

Lu, J., Ren, D. F., Xue, Y. L., Sawano, Y., Miyakawa, T., and Tanokura, M. (2010). Isolation of an antihypertensive peptide from alcalase digest of spirulina platensis. *J. Agric. Food Chem.* 58, 7166–7171. doi: 10.1021/jf100193f

Maurer-Stroh, S., Krutz, N. L., Kern, P. S., Gunalan, V., Nguyen, M. N., Limviphuvadh, V., et al. (2019). AllerCatPro—prediction of protein allergenicity potential from the protein sequence. *Bioinformatics* 35, 3020–3027. doi: 10.1093/BIOINFORMATICS/BTZ029

Minkiewicz, P., Iwaniak, A., and Darewicz, M. (2019). BIOPEP-UWM database of bioactive peptides: current opportunities. *Int. J. Mol. Sci.* 20:5978. doi: 10.3390/IJMS20235978

Mooney, C., Haslam, N. J., Pollastri, G., and Shields, D. C. (2012). Towards the improved discovery and Design of Functional Peptides: common features of diverse classes permit generalized prediction of bioactivity. *PLoS One* 7:e45012. doi: 10.1371/JOURNAL.PONE.0045012

Morozov, V., Rodrigues, C. H. M., and Ascher, D. B. (2023). CSM-Toxin: a web-server for predicting protein toxicity. *Pharmaceutics* 15:431. doi: 10.3390/PHARMACEUTICS15020431/S1

Nakamura, Y., Yamamoto, N., Sakai, K., Okubo, A., Yamazaki, S., and Takano, T. (1995). Purification and characterization of angiotensin I-converting enzyme inhibitors from sour Milk. *J. Dairy Sci.* 78, 777–783. doi: 10.3168/JDS.S0022-0302(95)76689-9

Narva, M., Collin, N., Jauhiainen, T., Vapaatalo, J., and Korpela, R. (2004). Effects of *Lactobacillus helveticus* fermented milk and its bioactive peptides on bone parameters in spontaneously hypertensive rats. *Milchwissenschaft* 59, 359–363.

Nielsen, S. D., Beverly, R. L., Qu, Y., and Dallas, D. C. (2017). Milk bioactive peptide database: a comprehensive database of milk protein-derived bioactive peptides and novel visualization. *Food Chem.* 232, 673–682. doi: 10.1016/J.FOODCHEM.2017.04.056

Okamoto, K., Kawamura, S., Tagawa, M., Mizuta, T., Zahid, H. M., and Nabika, T. (2020). Production of an antihypertensive peptide from milk by the brown rot fungus Neolentinus lepideus. *Eur. Food Res. Technol.* 246, 1773–1782. doi: 10.1007/S00217-020-03530-Y/TABLES/3

Qin, D., Bo, W., Zheng, X., Hao, Y., Li, B., Zheng, J., et al. (2022). DFBP: a comprehensive database of food-derived bioactive peptides for peptidomics research. *Bioinformatics* 38, 3275–3280. doi: 10.1093/BIOINFORMATICS/BTAC323

Rodrigues, R., Soares, R. P., and Secchi, A. R. (2010). Teaching chemical engineering using EMSO simulator. *Comput. Appl. Eng. Educ.* 18, 607–618. doi: 10.1002/cae.20255

Roman, C., Delgado, M. A., and García-Morales, M. (2020). Using process simulators in chemical engineering education: is it possible to minimize the "black box" effect? *Comput. Appl. Eng. Educ.* 28, 1369–1385. doi: 10.1002/cae.22307

Rose, P., Prlic, A., Altunkaya, A., Bi, C., Bradley, A., Christie, C., et al. (2016). The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* 45, D271–D281. doi: 10.1093/nar/gkw1000

Saha, S., and Raghava, G. P. S. (2006). AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res.* 34, W202–W209. doi: 10.1093/NAR/GKL343

Sayd, T., Dufour, C., Chambon, C., Buffière, C., Remond, D., and Santé-Lhoutellier, V. (2018). Combined in vivo and in silico approaches for predicting the release of bioactive peptides from meat digestion. *Food Chem.* 249, 111–118. doi: 10.1016/J.FOODCHEM.2018.01.013

Schechter, I., and Berger, A. (1968). On the active site of proteases. III. Mapping the active site of papain; specific peptide inhibitors of papain. *Biochem. Biophys. Res. Commun.* 32, 898–902. doi: 10.1016/0006-291X(68)90326-4

Sharma, N., Patiyal, S., Dhall, A., Devi, N. L., and Raghava, G. P. S. (2021). ChAlPred: a web server for prediction of allergenicity of chemical compounds. *Comput. Biol. Med.* 136:104746. doi: 10.1016/J.COMPBIOMED.2021.104746

Tagliazucchi, D., Shamsia, S., and Conte, A. (2016). Release of angiotensin converting enzyme-inhibitory peptides during in vitro gastro-intestinal digestion of camel milk. *Int. Dairy J.* 56, 119–128. doi: 10.1016/J.IDAIRYJ.2016.01.009

Turpeinen, A. M., Kumpu, M., Rönnback, M., Seppo, L., Kautiainen, H., Jauhiainen, T., et al. (2009). Antihypertensive and cholesterol-lowering effects of a spread containing bioactive peptides IPP and VPP and plant sterols. *J. Funct. Foods* 1, 260–265. doi: 10.1016/J.JFF.2009.03.001

Ulug, S. K., Jahandideh, F., and Wu, J. (2021). Novel technologies for the production of bioactive peptides. *Trends Food Sci. Technol.* 108, 27–39. doi: 10.1016/J.TIFS.2020.12.002

Villaró, S., Jiménez-Márquez, S., Musari, E., Bermejo, R., and Lafarga, T. (2023). Production of enzymatic hydrolysates with in vitro antioxidant, antihypertensive, and antidiabetic properties from proteins derived from Arthrospira platensis. *Food Res. Int.* 163:112270. doi: 10.1016/J.FOODRES.2022.112270

Vogelsang-O'dwyer, M., Sahin, A. W., Arendt, E. K., and Zannini, E. (2022). Enzymatic hydrolysis of pulse proteins as a tool to improve techno-functional properties. *Foods* 11:1307. doi: 10.3390/FOODS11091307

Wei, L., Ye, X., Sakurai, T., Mu, Z., and Wei, L. (2022). ToxIBTL: prediction of peptide toxicity based on information bottleneck and transfer learning. *Bioinformatics* 38, 1514–1524. doi: 10.1093/BIOINFORMATICS/BTAC006

Yamada, A., Sakurai, T., Ochi, D., Mitsuyama, E., Yamauchi, K., and Abe, F. (2015). Antihypertensive effect of the bovine casein-derived peptide met-Lys-pro. *Food Chem.* 172, 441–446. doi: 10.1016/J.FOODCHEM.2014.09.098

Yang, J., Gao, Z., Ren, X., Sheng, J., Xu, P., Chang, C., et al. (2021). DeepDigest: prediction of protein proteolytic digestion with deep learning. *Anal. Chem.* 93, 6094–6103. doi: 10.1021/ACS.ANALCHEM.0C04704/SUPPL_FILE/AC0C04704_SI_001.PDF

Zaru, R., and Orchard, S. (2023). UniProt tools: BLAST, align, peptide search, and ID mapping. *Current Protocols* 3:e697. doi: 10.1002/CPZ1.697