Trends in Analytical Chemistry 162 (2023) 117051

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

journal homepage: www.elsevier.com/locate/trac

Bioinformatics approaches to discovering food-derived bioactive peptides: Reviews and perspectives

Zhenjiao Du [a], Jeffrey Comer [b], Yonghui Li [a, *]

[a] Department of Grain Science and Industry, Kansas State University, Manhattan, KS, 66506, USA
[b] Department of Anatomy and Physiology, Kansas State University, Manhattan, KS, 66506, USA

ARTICLE INFO

Article history:
Received 22 December 2022
Received in revised form
1 April 2023
Accepted 3 April 2023
Available online 6 April 2023

Keywords:
Food proteins
Nutraceutical
Database
QSAR
Molecular docking
Virtual screening

ABSTRACT

Food-derived bioactive peptides (FBPs) are gaining interest due to their great potential in agricultural byproduct valorization and high-activity peptide screening. The introduction of bioinformatics into FBP studies further enhances the prospects of this field. This review provides a comprehensive overview and critical insight into the latest advances in bioinformatics-driven FBPs studies. The roles of databases, proteolysis simulation, bioactivity potency evaluation, quantitative structure-activity relationships (QSAR) models, molecular docking, molecular dynamics simulation, and free energy calculation in FBP studies are covered. Furthermore, critical issues related to QSAR model development, molecular docking, and integrated bioinformatics strategies are highlighted. By leveraging these bioinformatics approaches, researchers can fully utilize existing knowledge about identified peptides for checking novelty, evaluating bioactivity potency as well as rational peptide and protein hydrolysate design. QSAR models and molecular docking enable efficient screening of thousands of peptide candidates and generate new insights into bioactivity mechanisms. Directions for future research and challenges in current studies are also discussed. The employment of bioinformatics will significantly accelerate the process from the identification of high-potential FBPs to product development, assist in wet chemistry experiment design for targeted protein hydrolysates preparation, and ultimately enhance the long-term development of nutraceutical, pharmaceutical, and cosmeceutical industries.

1. Introduction

Biologically active peptides (or bioactive peptides) are protein fragments that contain 2–20 amino acid residues joined by peptide bonds and exhibit positive biological effects. Food proteins are a sustainable source for bioactive peptide preparation and are gaining interests from academic researchers and industries [1,2]. The number of studies on food-derived bioactive peptides (FBPs) in 2022 are expected to quintuple the number ten years ago (Fig. 1). The growing interest in FBPs is driven by the increasing demand for natural nutraceuticals, perceptions of the safety of synthetic products, sustainability, and, most importantly, the diverse bioactivities exhibited by FBPs and their potential to relieve the health burdens [3,4]. Common bioactivities of FBPs include antioxidant, antihypertension, anti-diabetes, and anti-inflammation activities (Fig. 1). Most FBPs exhibit their bioactivities at the protein level by inhibiting enzymes (e.g., angiotensin-converting enzyme inhibition for antihypertension) or through protein-ligand interactions (e.g., the Keap1–Nrf2 interaction for cytoprotective response regulation), although the bioactivities of a few FBPs might not involve proteins (e.g., capturing free radicals) [5–10].

FBPs need to be liberated from protein precursors (animal, plant, edible insect protein) to exert their functions (Fig. 1). Enzymatic hydrolysis is the most common method to generate FBPs. This approach requires only mildly controlled production conditions, results in good cleavage specificity and efficiency, and is free of organic solvents and toxic chemicals. Through wet chemistry experiments and *in vitro* and *in vivo* characterization, thousands of FBPs have been identified and characterized in the last few decades, and some of them (e.g., Val-Pro-Pro) have been commercialized [3,11]. However, these conventional approaches suffer from low efficiency and high cost, and they also rely heavily on advanced instruments and trained personnel. Moreover, most mechanisms of protein–peptide interactions can only be elucidated experimentally by X-ray crystallography or nuclear magnetic resonance spectroscopy (NMR), both of which are hindered by sample

* Corresponding author.
E-mail address: yonghui@ksu.edu (Y. Li).

https://doi.org/10.1016/j.trac.2023.117051
0165-9936/

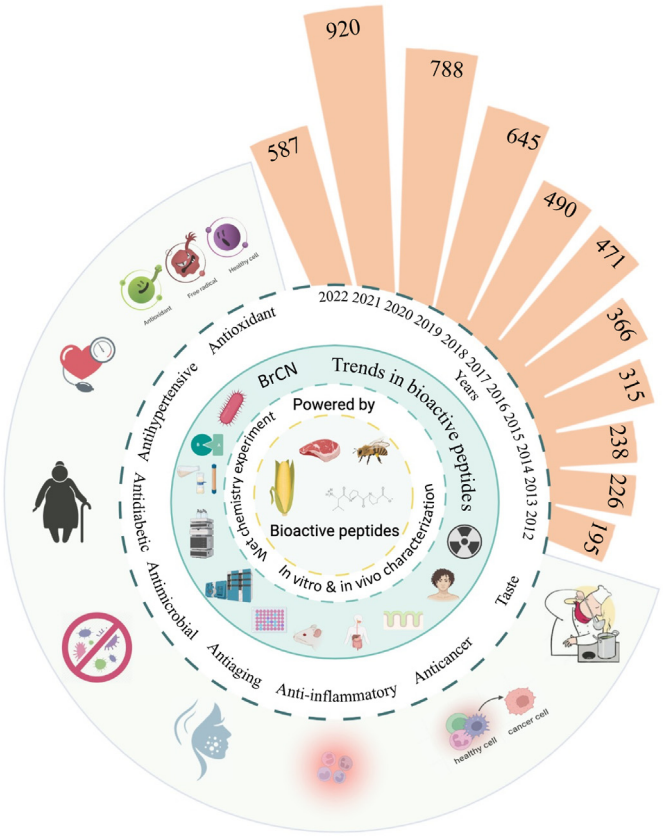

Fig. 1. Status quo of food protein-derived bioactive peptide research. Note: Data on the number of publications per year were obtained from Web of Science in August 2022 using "food & bioactive peptide" as keywords; BrCN: cyanogen bromide.

preparation, and it is impossible to employ such protocols to exhaust all the interaction mechanisms between FBPs and protein receptors [12].

To circumvent these limitations and accelerate FBP screening and mechanism explanation, some researchers turned to bioinformatics approaches [1,4,7,10,13–15]. As an *in silico* method, bioinformatics can fully exploit known FBPs, protein sequences, cleavage sites of enzymes or chemicals, and computational chemistry knowledge to eliminate guesswork and guide experiment design (e.g., enzyme selection or purification condition setting) [7,13,16,17]. In addition, bioinformatics can be used to screen FBP bioactivities; explore interaction mechanisms; evaluate bioavailability, allergenicity, ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties, and taste-evoking properties [2,17–23]. The field of bioinformatics has made considerable progress, including newly developed algorithms and models, web servers, software, and docking strategies, yet its potential in FBPs discovery has not yet been fully explored.

To our knowledge, there is currently no comprehensive or in-depth review for readers who have expertise in conventional FBPs studies but lack knowledge in bioinformatics to accelerate their studies or who would like to explore more possibilities with advanced bioinformatics tools. To address this gap, this review presents the whole spectrum of bioinformatics-aided procedures from FBP preparation to bioactivity, bioavailability, taste, allergenicity, and ADMET evaluation. Specifically, the application of quantitative structure-activity relationships (QSAR), molecular docking, and molecular dynamics simulation in FBP virtual screening is highlighted, and commonly used and advanced databases, web servers, and software are summarized. Furthermore, suggestions for future research are given to accelerate FBPs studies from bench to market, including the needs in database construction, structure construction of unavailable targeted proteins, trends in QSAR model development and molecular docking & molecular dynamics simulation, utilization of hardware for computation-intensive tasks, ensemble strategies for accuracy improvement, dataset clean & augmentation, and multifunctional peptide screening. In short, this review gathers the latest advances in bioinformatics and FBPs studies to advance the long-term, synergistic developments of both fields and guide the sustainable future production of value-added products from agricultural products.

2. Application of databases and proteolytic simulation in FBPs screening and potency evaluation

In the past decades, millions of protein sequences and hundreds of FBPs were identified and characterized by *in vitro* and *in vivo* studies, most being composed of the 20 proteinogenic amino acids. These findings contributed to the creation of bioactive peptide databases for efficient retrieval of information. As of now, dozens of bioactive peptide databases have been built, and each documents one or more bioactivities (e.g., BIOPEP) (Table 1). Users can use these databases to retrieve information (sources of origin, reference articles, bioactivities, etc.) about the peptide of interest using a one-letter sequence or three-letter sequence from the reported peptides [24,25]. In addition, these databases classify peptides into different categories (e.g., based on their source of origin, bioactivity, etc.), which simplifies the data mining procedure for other bioinformatics studies (e.g., QSAR analysis). It should be noted that no existing bioactive peptide database can be expected to contain all the latest data, so manual double-checking is always recommended [26–28]. In addition, government-led databases for proteins (i.e., Uniprot (The Universal Protein Resource), NCBI (The National Center for Biotechnology Information), and RCSB PDB (Research Collaboratory for Structural Bioinformatics Protein Data Bank) provide one-stop portals for protein sequence retrieval with all available sequence information (Table 1). *In silico* proteolysis successfully connects the protein sequence databases and bioactive peptide databases to advance bioactive peptide discovery and bioactivity potency evaluation (Fig. 2).

2.1. Database-driven approaches

The pure database-driven approach has three steps: 1. parent protein retrieval; 2. *in silico* proteolysis; 3. identification of bioactive peptides and additional evaluation (Fig. 2). This technical route was adopted in most studies and was combined with QSAR or molecular docking methods for unknown FBPs [19,24,29,30,30–37]. As shown in Table 2, most database-driven FBPs studies are related to the angiotensin-I-converting enzyme (ACE) inhibitory activity and dipeptidyl peptidase (DPP) IV inhibitory activity since these bioactivities are highly correlated to the most common and urgent chronic diseases (hypertension and diabetes). In addition, database-driven FBP studies tend to limit their data source to BIOPEP database where the number of peptides with ACE and DPP-IV inhibitory activity ranks first and fourth, respectively [25]. This limitation in data can be improved by searching in other peptide databases (Table 1). This strategy was adopted in the study of Martini et al., where BIOPEP and MBPDB were combined to search the identified peptides from LC-MS for further selection [38].

A few studies adopted a partial database-driven approach which employed only peptide bioactivity databases in their FBP studies (Table 2) [48–51]. In the studies of Sayd et al. and Devita et al.,

Table 1
Summary of protein information databases, proteolysis simulation web servers, bioactive peptide databases, mass spectroscopy data processing software, physiochemical property prediction web servers, peptide structure prediction web servers, and software/webserver for protein and peptide structure building, modification, and visualization.

Protein information database		
Name	*Website*	*Description**
UniProt	https://www.uniprot.org/	The most commonly used protein database. Contains Swiss-Prot with 568,002 manually reviewed protein sequences and TrEMBL with 226,771,948 unreviewed protein sequences
NCBI Protein	https://www.ncbi.nlm.nih.gov/protein/	Collection of protein sequences from 7 public databases (including sequences translated from annotated coding regions from genes)
RCSB	https://www.rcsb.org/	Most commonly used protein structure database for molecular docking and molecular dynamics simulation. Contains 194,259 protein three-dimensional structures from X-ray crystallography, NMR, and electron microsopy experiments.
EMDB	https://www.ebi.ac.uk/emdb/	Contains 21,807 entries of electron cryo-microscopy maps and tomograms of macromolecular complexes and subcellular structures
Proteolysis simulation web server		
Name	*Website*	*Description*
PeptideCutter	https://web.expasy.org/peptide_cutter/	The most commonly used tool. Has 34 different enzymes and chemicals available for proteolytic simulation
FeptideDB	http://www4g.biotec.or.th/FeptideDB/enzyme_digestion.php	Has the same enzymes as PeptideCutter
BIOPEP-UWM	https://biochemia.uwm.edu.pl/biopep-uwm/	Has 34 different enzymes from microbes, vegetables, fruits, and humans
AHPP	http://hazralab.iitr.ac.in/ahpp/index.php	Has 10 enzymes from the gastrointestinal tract and 7 enzymes from vegetables and fruits
SpirPep	http://spirpepapp.sbi.kmutt.ac.th/UserGuide.html	Has 10 enzymes with 1−3 miss cleavage options
Bioactive peptide database		
Name	*Website*	*Description***
DFBP	http://www.cqudfbp.net/	Contains a total of 6276 peptide entries in 31 types from different food sources (last updated in 2022)
BIOPEP-UWM	https://biochemia.uwm.edu.pl/	Most commonly used database in FBP studies. Contains 4485 bioactive peptides with 58 bioactivity categories from literature and 533 sensory peptides and amino acids (last updated in 2019)
FeptideDB	http://www4g.biotec.or.th/FeptideDB/index.php	Combination of 12 public bioactive peptide databases and peptides manually extracted from literature (last updated in 2019)
SpirPep	http://spirpepapp.sbi.kmutt.ac.th/BioactivePeptideDB.html	Combination of 13 public peptide databases (28,334 bioactive peptides)
$CAMP_{R3}$	www.camp3.bicnirrh.res.in	Contains 10,247 antimicrobial peptide sequences captured by analysis of 1386 antimicrobial peptide sequences from experiments (last updated in 2019)
BaAMPs	http://www.baamps.it/	Antimicrobial peptides (AMPs) specifically tested against microbial biofilms (last updated in 2015)
YADAMP	http://yadamp.unisa.it/about.aspx	Contains 252 antimicrobial peptides (last updated in 2018)
BioPepDB	http://bis.zju.edu.cn/biopepdbr/index.php	Contains 4807 bioactive peptides (51.07% are antimicrobial peptides, 34.9% are antihypertensive peptides, and 13.21% are anticancer peptides.) (last updated in 2018)
AHTPDB	http://crdd.osdd.net/raghava/ahtpdb/	Contains about 1700 antihypertensive peptides (last updated in 2015)
BERT4Bitter	http://pmlab.pythonanywhere.com/dataset	Contained 256 bitter peptides (last updated in 2021)
DBAASP	https://dbaasp.org/home	Contains 19,902 antimicrobial/cytotoxic peptides with detailed 3D structure information (last updated in 2021)
AVPdb	http://crdd.osdd.net/servers/avpdb/	Contains 2683 antiviral peptides (last updated in 2014)
TumorHoPe	http://crdd.osdd.net/raghava/tumorhope/	Contains 744 tumor homing peptides (last updated in 2012)
CancerPPD	http://crdd.osdd.net/raghava/cancerppd/index.php	Contains 3491 anticancer peptides (last updated in 2015)
CPP site2.0	http://crdd.osdd.net/raghava/cppsite/	Contains 1855 cell penetrating peptides (last updated in 2015)
BRAINPEP	https://brainpeps.ugent.be/	Blood-brain barrier peptide database (last updated in 2012)
NeuroPep	http://isyslab.info/NeuroPep/	Contains 5949 neuropeptides (last updated in 2015)
Hemolytik	http://crdd.osdd.net/raghava/hemolytik/	Contains about 3000 hemolytic and 2000 non-hemolytic peptides (last updated in 2013)
MBPDB	http://mbpdb.nws.oregonstate.edu/	Contains 994 milk-derived bioactive peptides (last updated in 2021)
FermFooDb	https://webs.iiitd.edu.in/raghava/fermfoodb/	Contains 2325 bioactive peptides from fermented food (last updated in 2021)
THPdb	http://crdd.osdd.net/raghava/thpdb/index.html	Contains 1238 FDA-approved therapeutic peptides (last updated in 2017)
StraPep	http://isyslab.info/StraPep/index.php	Contains 3791 bioactive peptide 3D-structures (last updated in 2018)
Mass spectroscopy data processing software		
Name	*Website*	*Description*
Mascot	https://www.matrixscience.com/search_form_select.html	Searching software for peptide sequence identification using MS data and precursor proteins
PEAKS X	https://www.bioinfor.com/peaks-studio/	Searching software for peptide sequence identification using MS data and precursor proteins
Physiochemical property prediction web server		
Name	*Website*	*Description*
PepDraw	http://www2.tulane.edu/∼biochem/WW/PepDraw/	The most commonly used tool. Predicts net charges, isoelectric points, hydrophobicity, molar extinction coefficient, and molecular weight
AhtPom	http://crdd.osdd.net/raghava/ahtpin/allinone.php	Predicts net charges, isoelectric points, hydrophobicity, hydrophilicity, steric hindrance, solvation, net hydrogen, charge, and molecular weight
Compute pI/Mw	https://web.expasy.org/compute_pi/	Predicts isoelectric point and molecular weight
ProtParam	https://web.expasy.org/protparam/	Predicts molar extinction coefficient, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity
PepCalc	https://pepcalc.com/	Predicts molar extinction coefficient, water solubility, net charges at neural pH, isoelectric points, and molecular weight
Peptide solubility calculator	https://pepcalc.com/peptide-solubility-calculator.php	Predicts solubility
Peptide structure prediction web server		
Name	*Website*	*Description*
DEBT	https://comp.chem.nottingham.ac.uk/debt/	Secondary structure prediction
PPIIPred		Secondary structure prediction

(*continued on next page*)

Table 1 (*continued*)

	http://bioware.ucd.ie/∼compass/biowareweb/Server_pages/ppIIpredSEQUENCE.php	
PEPstrMOD	http://osddlinux.osdd.net/raghava/pepstrmod/nat_ss.php	Tertiary structure prediction
APPTEST	https://research.timmons.eu/apptes	Tertiary structure prediction
PEP-FOLD3	https://bioserv.rpbs.univ-paris-diderot.fr/servixes/PEP-FOLD3	*De novo* structure prediction
PEP-FOLD	https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD/	*De novo* structure prediction (less optimized than PEP-FOLD3 but includes options such as disulfide bridges unavailable in the newer version)
Software/web server for protein and peptide structure building, modification, and visualization		
Name	*Availability*	*Website*
MolView	Free	https://molview.org/
VMD	Free for academic use	http://www.ks.uiuc.edu/Research/vmd/
DiscoveryStudio	Free for academic use	https://discover.3ds.com/discovery-studio-visualizer-download
Chimera	Free for academic use	https://www.cgl.ucsf.edu/chimera/
Avogadro2	Free and open source	https://two.avogadro.cc/
PyMol	Free for academic use	https://pymol.org/2/

Note: * The number of sequences was obtained in August 2022; ** The "last updated" time refers to the time of the latest publication of the databases or notice of updates on the database websites.

liquid chromatography-mass spectroscopy (LC-MS) was used to identify the FBPs released from the protein precursor, and the BIOPEP database was employed to search the LC-MS identified FBPs [49]. Proteolytic simulation was used to guide enzyme selection or enzyme combination for protein substrates in order to maximize the value of desired functional food hydrolysis [51,53,54]. There are disadvantages to these studies. Usually, web servers can work only on one type of bioactivity for one protein subject to cleavage by one enzyme at a time, which makes it difficult to conduct large-scale analyses [51]. Employing some automation of data mining tools (e.g., Selenium WebDriver) could significantly reduce repeated work. Besides, the order of enzyme addition for hydrolysis is not taken into consideration by most of the available proteolysis simulation tools (e.g., PeptideCutter, BIOPEP-UWM, etc.). Our lab recently developed and published an improved version (R-PeptideCutter) overcoming this defect. In addition, this improved tool allows researchers to run large-scale simulation and generate results in an easily readable format with Python scripting [55].

Additionally, QSAR web servers are widely employed in database-driven approaches to supplement the evaluation of bioactivity potential, physicochemical properties, allergenicity, toxicity, and ADMET [19,32,37,56,39]. However, most web servers might not update their built-in models with the latest reported data.

Overall, it is highly recommended that researchers employ two or more peptide databases and prediction tools built and developed by different research groups using different strategies to validate *in silico* results and provide more robust predictions. The most advanced and user-friendly web servers that can be used in database-driven studies are summarized in Table 3.

2.2. Bioactivity potency evaluation

Bioactivity potency evaluation of parent proteins is a common step after database searching (Table 2). A significant number of FBP studies aim to valorize byproducts from the food industry, and the most practical FBP production strategy using byproducts is to produce FBPs as a hydrolysis mixture instead of pure FBPs, because the latter incurs high purification cost [2,31,57]. Some indicators—e.g., A (the occurrence frequency of peptides in a protein), A_E (the occurrence frequency of released peptides in a protein under a specific enzyme), DH_t (theoretical degree of hydrolysis)—were proposed to evaluate the potential bioactivity of protein hydrolysis (Table 2). The distinction between indicators A and A_E as well as B (the potential bioactivity of a protein) and B_E (the potential bioactivity of released peptides in a protein under a specific enzyme) should also be heeded. For example, A is based on the occurrence frequency of the FBPs in the protein sequence regardless of the availability of cleavage needs from available enzymes. For some potential FBP sequences, there might be no enzymes capable of releasing them in practice [34,35,39,41,46,47]. Modified versions of indicators (e.g., A_E and B_E) were proposed for such situations [19,32,33]. B_E is valuable because it can be directly used for decision-making on the feasibility of specific protein substrates for bioactive hydrolysate production under optimal enzymes, and it takes into consideration cost, efficiency, and the expected positive effects on health. An interesting study of large-scale protein bioactivity potency evaluation based on B_E was conducted for prediction of 40 pigeon proteins hydrolyzed by pepsin, papain, and thermolysin, and the theoretically generated peptides were predicted by modeling eight parameters using the random forest method; the eight parameters are frequencies of the six amino acid residues (A, P, V, G, L, F), hydrophobicity values, and A_E [58]. Using only these eight parameters might have simplified the model development procedures, but also undermined the prediction power of the model [58].

With reasonable indicators and proteolytic simulation, the last prerequisite for comprehensive bioactivity potency evaluation is an all-inclusive database that can be used to search for the bioactivity of the theoretical peptide. However, purifying or synthesizing all the theoretically possible peptides (i.e., 400 dipeptides, 8000 tripeptides, and 160,000 tetrapeptides, etc.) is not possible in reality, let alone the evaluation of multiple possible distinct bioactivities. Therefore, there will always be FBPs with unknown bioactivity, which will undermine potency evaluation accuracy [14,58,59]. The alternative approach is to combine reported data and predictive data from QSAR models (see Section 3) for potency evaluation. Such attempts are expected to be used in *in silico* studies in the future.

2.3. Challenges in proteolysis simulation

In silico prediction was mainly employed as a qualitative tool for hypothesis proposal and experimental validation, without quantitative validation of predictions for peptides released by degradation of proteins [29,30,34,37,38,51,50,53,41,43]. With the help of QSAR models, database-driven approaches have the potential to become a quantitative tool for predicting the bioactivity of peptides released from proteins; however, there has not yet been a systematic quantitative validation of this approach. A big challenge is that the gap between peptides theoretically predicted using proteolysis simulation and those identified by wet chemistry needs to be bridged in order to advance the quantitative evaluation of protein bioactivities.

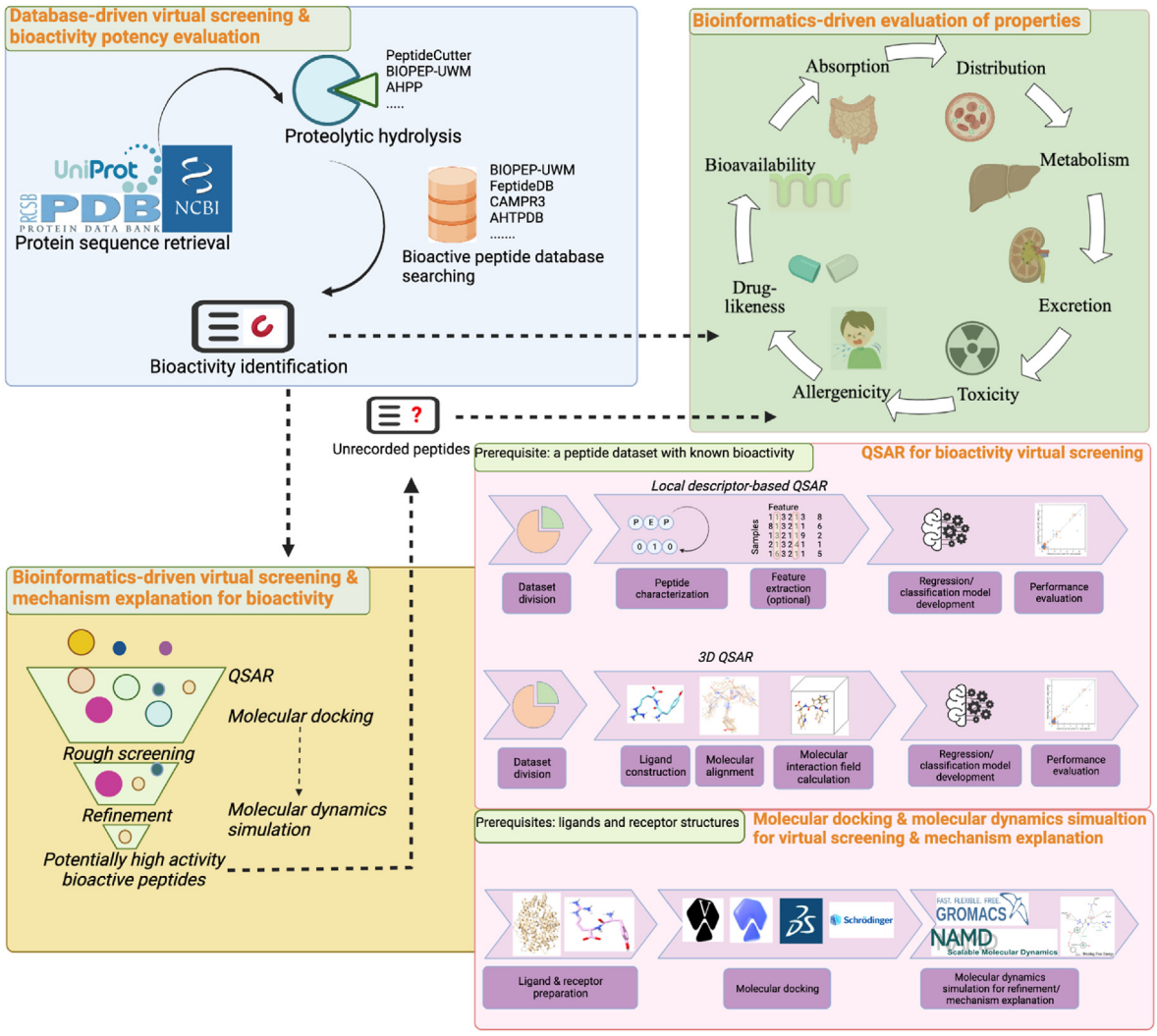

Fig. 2. Overall workflow of bioinformatics application in food protein-derived bioactive peptide studies.

Some experiments have shown discordance between wet chemistry identification and *in silico* proteolysis [16,60,61]. A study comparing *in silico* and *in vitro* analysis conducted by Chatterjee et al. showed deviations not only between *in vitro* identified peptides and *in silico* predicted peptides but also between two different *in silico* strategies [60]. However, the details of this study limit generalization. The experimental sample used for the *in vitro* analysis, whey protein concentrate, was a mixture of various proteins, while the *in silico* proteolytic simulation considered only the two proteins α-lactalbumin and β-lactoglobulin, accounting for some of the disagreement [60]. Considering that purified forms of these two proteins are commercially available, further experiments could more directly compare the *in silico* and *in vitro* methods, allowing for a better understanding of the origin of the disagreement and checking the limitations of current proteolytic simulation methods for predicting experimental results.

There are two major challenges for predicting the peptides released from a given protein source. First, protein naturally occurring in foods or agricultural by-products is invariably of a complex composition. This challenge could be addressed by obtaining better quantitative data on the protein composition of the protein source. Quantitative characterization of protein composition could be obtained from size exclusion chromatography (SEC) or sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) [62,63]. In proteolytic simulation, this data could be used to assign a weight to each of the predominant protein sequences present in a source material.

The second challenge is from limitations of the *in vitro* experiments. Other components in the protein substrates (e.g., lipids and carbohydrates), environmental conditions, enzyme quality, and each protein's secondary, tertiary and quaternary structures contributed to inconsistency in different *in vitro* experiments [53,64]. Some attempts have been made to optimize *in vitro* experiments such as using microwave pre-treatment of protein substrates, heating, high hydrostatic pressure, pulsed electric fields, and ultrasound to unfold protein structure and improve protease accessibility during hydrolysis, and response surface methodology to select the hydrolysis conditions [57,64]. After hydrolysis, the smaller molecular weight fractions generally exhibited higher bioactivity, and the physicochemical properties of these peptides were the main factor for fractionation methods design. Some physicochemical property prediction web servers have been developed (Table 1), which can be used to guide extraction and determination protocols in wet chemistry experiments.

Such efforts have contributed to narrowing the gaps between *in silico* and *in vitro* experiments. Future studies are expected to

Table 2
Database-driven FBP virtual screening and bioactivity potency evaluation studies in the last 5 years.

Pure database-driven approach				
Protein source	*Bioactivity*	*Bioactivity potency evaluation**	*Additional comments*	*Reference*
Chickpea	ACE inhibitory activity	A, B, A_E, B_E	Conducted ADMET evaluation and molecular docking for interaction mechanism explanation	[19]
Sheep milk protein	ACE inhibitory activity DPP-IV inhibitory activity	A, A_E	Used PeptideRanker for bioactivity evaluation; conducted physicochemical property and toxicity evaluation; used molecular docking for virtual screening and mechanism explanation.	[39]
Mammal milk proteins	ACE inhibitory activity	Molar concentration of released peptides	Also searched for DPP-III inhibitory activity, antioxidant activity, hypo-cholesterolemic activity, immunomodulatory activity, and antithrombotic activity in FBPs with ACE inhibitory activity	[40]
Goat casein	ACE inhibitory activity	A, B	Used PeptideRanker and AHTpin for bioactivity evaluation. Conducted *in vitro* and *in vivo* validation	[41]
Pumpkin seed	ACE inhibitory activity	A, A_E, W	Used molecular docking for virtual screening and molecular dynamics simulation for refinement. Conducted ADMET evaluation and *in vitro* validation	[42]
Porcine liver	Antioxidant activity	Not mentioned	Used PeptideRanker for bioactivity evaluation. Conducted *in vitro* and *in vivo* validation	[43]
Flaxseed	ACE inhibitory activity Renin inhibitory activity	Not mentioned	Conducted physicochemical property, ADMET, and drug-likeness evaluation; used molecular docking for mechanism exploration and affinity comparison with common drugs (aliskiren and captopril)	[32]
Tomato seed	ACE inhibitory activity DPP-IV inhibitory activity Antioxidant activity	A, A_E, sum of A_E	Conducted allergenicity and toxicity evaluation	[33]
Animal and fish collagens	ACE inhibitory activity DPP-IV inhibitory activity	A_E, DH_t, W	Conducted ADMET evaluation; used PeptideRanker for bioactivity evaluation; used SwissTargetPrediction to predict potential interaction between selected FBPs and other enzymes and proteins	[24]
Bovine milk protein	ACE inhibitory activity DPP-IV inhibitory activity Antioxidant activity	Molar concentration of released peptides	Used PeptideRanker for bioactivity evaluation and comparison of released FBPs with digestion-resistant peptides	[44]
Milk and meat derived protein	ACE inhibitory activity Antioxidant activity Opioid activity	Not mentioned	Conducted physicochemical property and toxicity evaluation for FBPs and peptides; used PeptideRanker for bioactivity evaluation; conducted *in vitro* and *in vivo* validation	[29]
Tilapia	ACE inhibitory activity	DH_t	Conducted physicochemical property and toxicity evaluation and *in vitro* and *in vivo* validation	[45]
Flaxseed	ACE inhibitory activity DPP-IV inhibitory activity Antioxidant activity	A, A_E, DH_t, W	Conducted physicochemical property and toxicity evaluation	[31]
Oyster	ACE inhibitory activity DPP-IV inhibitory activity	Not mentioned	Validated DH_t of enzymes by *in vitro* hydrolysis; used proteolytic simulation to guide enzyme selection	[30]
Yak milk	ACE inhibitory activity	A	Conducted toxicity evaluation	[46]
Salmo salar	ACE inhibitory activity	Not mentioned	Conducted physicochemical property and toxicity evaluation; used molecular docking screening and *in vitro* study for validation	[37]
Giant grouper roe	ACE inhibitory activity DPP-IV inhibitory activity	A	Conducted *in vitro* study on trypsin in-gel digestion	[34]
Bovine casein	Anticancer activity Antithrombotic activity Anti-inflammatory activity Immunomodulating activity	Not mentioned	Used PeptideRanker for bioactivity, toxicity, and allergenicity evaluation	[36]
Caulerpa RuBisCO	ACE inhibitory activity DPP-IV inhibitory activity Antioxidant activity Neuroprotective activity Antithrombotic activity	A, A_E	Proteins only had 49% sequence homology among 28 species	[47]
Rice bran	ACE inhibitory activity	A, B, A_E, W	Used PeptideRanker for bioactivity, physicochemical property, allergenicity, and toxicity evaluation	[35]

Table 2 (*continued*)

Pure database-driven approach			
	DPP-IV inhibitory activity		
Partial database-driven approach			
Protein source	*Chemistry experiments and other analysis*	*Database-driven analysis*	*Reference*
Tuna skin collagen	FBP identification by LC-MS; physicochemical property evaluation	All the available bioactivities of identified FBPs were searched in BIOPEP	[48]
Cheese	FBP identification by LC-MS; FBPs with anti-diabetic activity were synthesized and confirmed by *in vitro* experiments	Identified FBPs were searched in BIOPEP and MBPDB (Table 1)	[38]
Food matrix	FBP identification by LC-MS; FBP conformation prediction	All the available bioactivities of identified FBPs were searched in BIOPEP	[49]
Bean protein	FBP identification by LC-MS; physicochemical property evaluation	Bioactivities of identified FBPs were searched in BIOPEP	[50]
Porcine products	*In vitro* hydrolysis and bioactivity evaluation	*In silico* hydrolysis simulation and potential bioactivity evaluation were conducted to guide optimal enzyme selection	[51]
Wheat	Peptideranker was used for FBP screening, and opioid activity of FBPs was confirmed by *in vitro* experiment	Used *in silico* hydrolysis simulation	[52]

Notes: *: $A = \frac{a}{N}$ or $A_E = \frac{d}{N}$ where A is the occurrence frequency of peptides in a protein, A_E is the occurrence frequency of released peptides from a protein under a specific enzyme, a is the number of peptides with bioactivity encrypted in the selected protein chain, d is the number of released peptides with bioactivity, and N is the number of amino acid residues in the protein. $DH_t = \frac{d}{N}$ where DH_t is the theoretical degree of hydrolysis, d is the number of hydrolyzed peptide bonds, and D is the total number of peptide bonds in a protein chain. $W = \frac{A_E}{A}$ where W is the relative frequency of the peptide released by given activity by selected enzymes or chemicals, A is the occurrence frequency of peptides, and A_E is the occurrence frequency of released peptides. $B = \frac{\sum_{i=1}^{k} \frac{a_i}{EC_{50i}}}{N}$ or $B_E = \frac{\sum_{i=1}^{k} \frac{a_{Ei}}{EC_{50i}}}{N}$ where B is the potential bioactivity of a protein, B_E is the potential bioactivity of released peptides from a protein under a specific enzyme, a_i is the number of repetition of the i-th bioactive fragment in the protein sequence, a_{Ei} is the number of repetition of the i-th bioactive peptide released from the protein sequence, EC_{50i} is the concentration of the i-th bioactive peptide corresponding to its half-maximal activity (μM), k is the number of different fragments with a given activity, and N is the number of amino acid residues.
Abbreviations: ACE: angiotensin-I-converting enzyme; ADMET: absorption, distribution, metabolism, excretion, and toxicity; DPP-IV: dipeptidyl peptidase IV; RuBisCO: ribulose-1,5-bisphosphate carboxylase; UbMP: activate ubiquitin-mediated proteolysis.

integrate this progress, leading to more robust and reliable proteolytic simulation for complex protein sources. On the other hand, it should be noticed that bioinformatics software such as PEAKS X or Mascot, which is used for mass spectrometry (MS) data deconvolution and interpretation, is limited by predefined settings (e.g., parent protein databases and enzyme specificity) and also has difficulty in identifying small peptides, which are the majority of the peptides in the databases [51,43,61,65,66].

3. QSAR approaches and applications in FBP virtual screening

QSAR is a computational modeling method to interlink the structural characteristics and biological activity of a substance. It is the most popular bioinformatics approach in FBP virtual screening. QSAR models can be categorized as classification models or regression models based on their modeling methods (Tables 3 and 4). A regression model can predict specific bioactivity values (e.g., IC_{50}), while a classification model can only predict relative bioactivity among a group of samples (e.g., bioactivity potential) or binary classification (e.g., toxicity) [10,13,56,67]. The prerequisites for QSAR model development is a set of bioactive peptides with known bioactivity, so QSAR approaches are also classified as a ligand-based virtual screening method [68]. There are three basic steps in QSAR modeling: collection of peptide bioactivity data; peptide representation by different descriptors; and model development (Fig. 2).

3.1. Datasets in QSAR-driven virtual screening

As a knowledge-based method, QSAR modeling relies highly on datasets [67]. Therefore, dataset collection is a common issue that hinders QSAR model performance, especially for researchers without a biochemistry background [20,67,75]. At this time, there is no well-recognized and curated dataset for QSAR model development (Table 1 summarizes the databases used for FBP data retrieval). Some researchers directly retrieve the datasets used in their previous papers without any updates with the latest FBPs [20,67,75]. For example, in the study of Zhou et al., the peptides in the 8 datasets were published in 1986–2008 [67]. Manually collecting FBP data from literature is laborious but effective for dataset enlargement and model performance improvement [5,26]. Some efforts have been made to develop FBP databases. BIOPEP has an option for authors to upload their latest bioactivity data with published references; it then manually checks the data [25]. However, there is still a long way to go.

In addition, the size of current FBP datasets is quite small. All the datasets used in the studies in Table 4 had fewer than 250 entries, although the dataset size slowly increased [5,26,27,69]. In order to gain more peptides for QSAR model development, some researchers have synthesized many pure peptides using chemical approaches, which enlarged the FBP databases and contributed to QSAR model performance improvement [26,27,69,70].

3.2. Peptide representation in QSAR-driven virtual screening

Peptide representation is an essential step in QSAR model development. Basically, there is a need to numerically represent peptide characteristics by different molecular descriptors. Peptide representation can be classified into different types based on the type of descriptors used (e.g., chemical descriptors), descriptor properties (e.g., 1D-, 2D-, and 3D-QSAR), or structure separation (i.e., local descriptors and global descriptors) (Fig. 3). Below, we further discuss peptide representation, considering categorization by structure separation because it can easily differentiate among FBP studies (Table 4), as well as the latest development in natural language processing (NLP)-based peptide representation [14,79].

3.2.1. Local descriptor-based peptide representation

In peptide representation, local descriptors are also known as

Table 3
Summary of QSAR prediction webservers.

Common bioactivity prediction					
Bioactivity	*Web server*	*Website*	*Model development*	*Model performance*	*Release time*
18 different properties (20 datasets)*	UniDL4BioPep	https://nepc2pvmzy.us-east-1.awsapprunner.com/	A protein language model based CNN model	Better performances than the respective state-of-the-art models for 15 out of 20 different bioactivity dataset prediction tasks)	2023
Antioxidant	AnOxPePred	http://services.bioinformatics.dtu.dk/service.php?AnOxPePred-1.0	CNN model	ACC is around 80% MCC is around 0.7	2020
Antioxidant	IDAod	http://antioxidant.weka.cc	Neural network for feature extraction; t-SNE for feature reduction; binary SVM classifier model	ACC = 97.05% MCC = 0.7409	2018
ACE inhibition	pLM4ACE	https://sqzujiduce.us-east-1.awsapprunner.com/	A protein language model based CNN model; confident learning theory for data cleaning	BACC = 88.3% MCC = 0.77	2023
ACE inhibition	MAT-AHT	http://hazralab.iitr.ac.in/ahpp/index.php	No extra feature selection method; regression decision tree model	r = 0.9513	2021
ACE inhibition	PAAP	http://codes.bio/paap/	No extra feature selection method; random forest binary classifier	ACC = 84.73%	2018
ACE inhibition	AHTpin	http://crdd.osdd.net/raghava/ahtpin/di_mat.php	No extra feature selection method; SVM regression model for di/tripeptides; SVM classification	r = 0.701 and 0.543 for dipeptides and tripeptides, respectively; ACC = 76.67%, 72.04%, 77.39, 82.61%, and 84.21% for tetrapeptide, pentapeptide, hexapeptides, medium peptides (7–13 residues), and large peptides (above 13 residues), respectively	2015
DPPIV inhibition	StackDPPIV	http://pmlabstack.pythonanywhere.com/StackDPPIV	GA-SAR for feature selection; random forest binary classifier	ACC = 89.1% MCC = 0.784 AUC = 96.1%	2022
DPPIV inhibition	iDPPIV-SCM	http://camt.pythonanywhere.com/iDPPIV-SCM	No extra feature selection method; Scoring card method for binary classifier	ACC = 69.7% MCC = 0.594 AUC = 84.7	2020
Antimicrobial	ClassAMP	http://www.bicnirrh.res.in/classamp/	SVM or RF classifier	MCC = 0.92, 0.83, and 0.96 for antibacterial, antifungal, and antiviral peptides, respectively	2020
Antimicrobial	MAT-AMP	http://hazralab.iitr.ac.in/ampgp.php	Not available	Not available	2021
Antimicrobial	iAMGpred	http://hazralab.iitr.ac.in/ampgp.php	SVM classifier	AUC = 94% MCC = 0.88	2017
Antimicrobial	ADAM	https://bioinformatics.cs.ntou.edu.tw/ADAM/tool.html	SVM or profile hidden Markov models	Not available	2015
Anticancer	xDeep-AcPEP	https://app.cbbio.online/acpep/home	CNN model	r = 0.8073, 0.8322, 0.7289, 0.8179, 0.8370, and 0.8285 for breast, cervix, skin, prostate, lung, and colon cancer, respectively	2021
Anticancer	MLACP 2.0	https://balalab-skku.org/mlacp2	CNN model	ACC = 76.5% MCC = 0.513 AUC = 0.773	2022
Anticancer	AntiCP 2.0	https://webs.iiitd.edu.in/raghava/anticp2/index.html	Extra trees classifier	ACC = 92.1% MCC = 0.84	2021
Anti-inflammatory	InflamNat	http://www.inflamnat.com/	Multi-tokenization transformer model	AUC = 84.2%	2022
Anti-inflammatory	PreAIP	http://kurata14.bio.kyutech.ac.jp/PreAIP/	RF model	ACC = 77% MCC = 0.512 AUC = 84%	2019
Anti-angiogenic	AntiAngioPred	http://webs.iiitd.edu.in/raghava/antiangiopred/	SVM classifier	ACC = 80.9% MCC = 0.62	2015
Hemolytic	HAPPENN	https://research.timmons.eu/happenn	Neural network	ACC = 85.7% MCC = 0.71	2020
Hemolytic	HemoPred	http://codes.bio/hemopred/	RF model	ACC = 95% MCC = 0.91	2017
Hemolytic	HemoPI	http://crdd.osdd.net/raghava/hemopi/	SVM model	ACC = 96.4% or 75.7% MCC = 0.93 or 0.51 for two different datasets	2016
Anti-tubercular	AtbPpred	http://thegleelab.org/AtbPpred	Neural network model	AAC = 87.3%	2019
Immunomodulatory	NetMHCpan 4.0	https://services.healthtech.dtu.dk/service.php?NetMHCpan-4.0	Neural network model	r = 0.790 AUC = 93.4%	2020

Table 3 (*continued*)

Common bioactivity prediction					
Additional property prediction					
Function	*Web server*	*Website*	*Model development*	*Model performance*	*Release time*
Whether a is bioactive	PeptideRanker	http://distilldeep.ucd.ie/PeptideRanker/	Neural network model	With 0.8 as threshold, FPR = 0.02 and 0.06, MCC = 0.54 and 0.74, and AUC = 0.04 and 0.932 for short peptides (4–20 residues) and long peptides (>20 residues), respectively	2012
Peptide intestinal stability	HLP	http://crdd.osdd.net/raghava/hlp/	SVM regression model	r = 0.70 and 0.98 for two different datasets	2014
Peptide penetrate cell	CPPpred	http://distilldeep.ucd.ie/CPPpred/	Neural network model	MCC = 0.69 FPR = 2.13	2013
Plasma stability	PlifePred	http://webs.iiitd.edu.in//raghava/plifepred/	SVM regression model	r = 0.743	2018
Metabolism prediction	BioTransformer 3.0	https://biotransformer.ca/	A knowledge-based prediction tool and a set of random forest and ensemble models	Jaccard score range from 0.380 to 0.452 for 9 metabolic transformations	2022
ADMET evaluation	ADMETlab 2.0	https://admetmesh.scbdd.com/	40 classification models and 13 regression models were built on MGA framework	For the regression models, R^2 ranges from 0.678 to 0.957, and average R^2 is 0.783; AUC ranges from 0.707 to 0.983, and average AUC is 0.863	2021
ADME evaluation	SwissADME	http://www.swissadme.ch/	19 classification models and 15 regression models were built on existing models	Best model R^2 = 0.75 for water solubility; R^2 = 0.67 for skin permeability coefficient; ACC ranges from 0.78 to 90.8 for pharmacokinetic parameters, and average ACC is 80.5%; r = 0.91 or 0.62 (depending on datasets)	2017
Protein allergenicity	AllerCatPro	https://allercatpro.bii.a-star.edu.sg	Combination of protein clustering program (cd-hit) and BLAST search tool; homology detection and molecular dynamics simulation for 3D structure correction and comparison	ACC = 84% MCC = 0.727	2022
Peptide allergenicity	SORTALLER	http://sortaller.gzhmc.edu.cn/	A substantially optimized SVM binary classification model	AAC = 98.5% MCC = 0.97	2012
Chemical allergenicity	ChAlPred	https://webs.iiitd.edu.in/raghava/chalpred/	Pearson correlation and support vector classifier for feature selection; random forest for classifier	AAC = 83.39% AUC = 0.93	2021
Chemical cytochrome activity	SuperCYPsPred	http://insilico-cyp.charite.de/SuperCYPsPred/	Five RF classifiers for CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4	All the ACC >93% All the AUC >0.93	2020
Peptide toxicity	ProTox-II	http://tox.charite.de/protox_II	31 models were developed by RF, ensembles of SVM and RF, or Bernoulli–Naive Bayes models for different toxicities	Average ACC = 85% Average AUC = 0.83	2018
Peptide toxicity	ToxinPred	https://webs.iiitd.edu.in/raghava/toxinpred/index.html	SVM model	ACC = 96.01% MCC = 0.89	2013
Taste	VirtualTaste	http://virtualtaste.charite.de/VirtualTaste/	RF classifier	ACC = 90% AUC = 98%	2021
Bitterness	BERT4Bitter	http://pmlab.pythonanywhere.com/BERT4Bitter	A BERT-based binary classification model	92.2% accuracy in test dataset; the BERT-model outperformed models built on CNN and LSTM neural network	2021
Bitterness	iBitter-SCM	http://camt.pythonanywhere.com/iBitter-SCM	A SCM-based binary predictor	AAC = 84.38%	2020
Umami	iUmami-SCM	http://camt.pythonanywhere.com/iUmami-SCM	A SCM-based binary predictor	ACC = 86.5% MCC = 0.679	2020

Abbreviation: ACC: accuracy; ADMET: absorption, distribution, metabolism, excretion, and toxicity; AUC: area under the curve of a receiver operating characteristic curve plots; BERT: bidirectional encoder representation from transformer, BLAST: Basic Local Alignment Search Tool; CNN: convolutional neural network; FPR: false positive rate; GA-SAR: genetic algorithm based on self-assessment report; LSTM: long short-term memory; MCC: Matthews correlation coefficient; MGA: multi-task graph attention; r: Pearson correlation coefficient; R2: determination of coefficient; RF: random forest; SCM: scoring card method; SVM: support vector machine; t-SNE: t-distributed stochastic neighbor embedding.

*: The 18 bioactivities include angiotensin-converting enzyme (ACE) inhibitory activity (anti-hypertension), dipeptidyl peptidase IV (DPPIV) inhibitory activity (antidiabetes), bitter, umami, antimicrobial activity, antimalarial activity, quorum-sensing (QS) activity, anticancer activity, anti-methicillin-resistant S. aureus (MRSA) strains activity, tumor T cell antigens (TTCA), blood-brain barrier, antiparasitic activity, neuropeptide, antibacterial activity, antifungal activity, antiviral activity, toxicity and antioxidant activity.

amino acid descriptors (AADs) where peptide residues are separately characterized by their own AADs and assembled depending on the peptide sequence [69]. For example, 5-z scale is a kind of AAD with 5 parameters for amino acids, and then a peptide with n residues will be represented as a vector of $5n$ components where the first 5 elements correspond to the 5 parameters of the first amino acid residue in the 5-z scale (Fig. 2). This approach is the most popular peptide representation approach among FBP studies [26,53,69,72–74,76]. AADs are derived from the basic properties of amino acids, including physicochemical properties, topological properties, 3D structural information and others, using feature extraction methods such as principal component analysis (PCA) [67]. To date, there are up to 80 proposed AADs, and most of them were extracted by PCA for different properties or a mixture of different properties [67]. For instance, the 5-z scale is a physicochemical descriptor extracted from 26 original physicochemical

Table 4
Selected virtual screening strategies using quantitative structure-activity relationship (QSAR) modeling.

Local descriptor-based QSAR model application						
Bioactivity	*Dataset size*	*Amino acid descriptors (AADs)*	*Model*	*Performance*	*Additional comments**	*Reference*
Antioxidant activity (ABTS assay)	133 tripeptides	566 amino acid properties were selected by 6 ML methods as AADs or directly used as AADs	14 ML regression models	$R^2_P = 0.847$ (best model by RFR for feature selection and XGB for regression model)	QAY, PHC, YPQ, VYV, GPE, and YSQ; performance comparison between 98 models	[5]
Antioxidant activity (DPPH assay)	69 peptides	SVRG and SVEEVA were screened by SWR and only for the representation of first five residues at terminus	PLSR	$R^2_P = 0.5536$ by SVRG $R^2_P = 07173$ for SVEEVA	AGWACLVG, IDLAY, YPLDL, IPIGP, and EAFDPLG	[59]
Antioxidant activity (FTC and FRAP assays)	214 tripeptides in FTC dataset and 173 tripeptides in FRAP dataset	16 AADs were integrated by BOSS	PLSR	$R^2_{CV} = 0.7471$ for FTC dataset and $R^2_{CV} = 0.6088$ for FRAP dataset	MPA for outlier detection; performance comparison between the descriptor selector by BOSS and 16 basic AADs	[26]
Antioxidant activity	91 tripeptides	195 physiochemical properties of amino acids were screened by SWR	PLSR, SVMR, RFR, and MLR	$R^2_{CV} = 0.706$ for PLSR $R^2_{CV} = 0.764$ for SVMR $R^2_{CV} = 0.728$ for RFR $R^2_{CV} = 0.798$ for MLR	GHG, LVG, GHT, GHG, GHP, KHP, GVR, ECG, GVW, GKW, GHW, QVW, KVW, NKW, NHW, QHW, KHW, PYW, and YHW	[69]
Antioxidant activity (ORAC and ABTS assays)	48 dipeptides	3-z scale, 5-z scale, DPPS, and ISA-ECI	PLSR	All the R^2_{CV} were below 0.5	The bioactivity of peptides used in this studies were determined in the same laboratory	[70]
ACE inhibitory activity	58 dipeptides	80 different AADs	SVMR and PLSR	R^2_P varied from 0.6 to 0.82 for SVMR and from 0.48 to 0.7 for PLSR	N/A	[67]
ACE inhibitory activity	84 dipeptides, 169 tripeptides, and 15 tetrapeptides	SVHEHS screened by OSC	SVMR	R^2_{CV} in four models were all above 0.995	ACC was employed to unify the feature dimension among different peptide length datasets	[71]
ACE inhibitory activity	141 dipeptides	16 AADs were integrated by BOSS	PLSR	$R^2_{CV} = 0.7151$	MPA for outlier detection; performance comparison between the descriptor selector by BOSS and 16 basic AADs	[72]
ACE inhibitory activity	166 dipeptides and 141 tripeptides	5-z scale	PLSR	$R^2_{CV} = 0.756$ for dipeptides $R^2_{CV} = 0.445$ for tripeptides	YW and LRY	[53]
DPP IV inhibitory activity	30 peptides	5-z scale and v-scale used for the representation of N- and C-terminus resides	PLSR	$R^2_{CV} = 0.775$ for 5-z scale $R^2_{CV} = 0.754$ for v-scale	FP, HP, RP, VP, IPM, LPP, IPPL, IPSK, VPGEIVE, YPFPGP, LPQNIPPLT, IPPLTQT, TPVVVPP, YPVEPF, LPLPLL, QPHQPLPPT, QPLPPT, and LPVPQ	[73]
Antimicrobial activity	196 dodecapeptides	20 different AADs	PLSR	$R^2_{CV} = 0.633$ for FASGAI as AADs	AADs were screened by GA-PLSR before model development	[74]
Bitterness	48 dipeptides	553 physicochemical properties selected by PCA	MLR	$R^2_P = 0.907$	Performance comparison with 7 AAD-based models	[75]
Bitterness	48 dipeptides, 52 tripeptides, and 23 tetrapeptides	16 AADs were integrated by BOSS	PLSR	$R^2_{CV} = 0.941$ for dipeptides $R^2_{CV} = 0.742$ for tripeptides $R^2_{CV} = 0.956$ for tetrapeptides	MPA for outlier detection; performance comparison between the descriptor selector by BOSS and 16 basic AADs	[76]
Global descriptor-based QSAR (3D-QSAR) model application						
Bioactivity	*Dataset size*	*Molecular alignment & MIF calculation*	*Model*	*Performance*	*Additional comments**	*Reference*
ACE inhibitory activity	53 di/tripeptides	Docking-based alignment (Suflex-Dock); CoMFA and CoMSIA for MIF calculation	PLSR	$R^2_{CV} = 0.773$ for CoMFA $R^2_P = 0.664$ for CoMSIA	GEF, VEF, VRF, and VKF were synthesized for validation	[77]
ACE inhibitory activity	40 dipeptides, 32 tripeptides, and 41 tetra/penta/hexapeptides	Template ligand-based alignment; CoMFA and CoMSIA for MIF calculation	PLSR	$R^2_{CV} = 0.862$ for dipeptides (CoMSIA) $R^2_{CV} = 0.848$ for tripeptides (CoMFA) $R^2_{CV} = 0.656$ for longer peptides (CoMFA)	N/A	[78]
Antimicrobial activity	24 nonapeptides			$R^2_{CV} = 0.601$	All peptides are rich in R, P, and F residues	[78]
Bitterness	21 peptides			$R^2_{CV} = 0.530$	All peptides are rich in W residue	[78]

Notes: *: Peptides were synthesized by chemical approach for validation experiments; N/A: not available.
Abbreviation: AADs: amino acid descriptors; ABTS: 2,2′-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid) radical; ACC: auto and cross auto covariances; BOSS: bootstrapping soft shrinkage; CoMFA: comparative molecular field analysis; CoMSIA: comparative molecular similarity indices analysis; DPPH: 2,2-Diphenyl-1-picrylhydrazyl; FTC: ferric thiocyanate; FRAP: ferric ion reducing antioxidant power; GA: genetic algorithm; MPA: model population analysis; MIF: molecular interaction field; ML: machine learning; MLR: multiple linear regression; OSC: orthogonal signal correction; PCA: principal component analysis; PLSR: partial least squares regression; RF: random forest regression; R^2_P: determination of coefficient in test dataset; R^2_{CV}: determination of coefficient of cross-validation; SVMR: support vector machine regression; SVRG: vector of radial distribution function descriptors and geometrical descriptors; SVEEVA: vector of principal component score for electronic eigenvalue descriptors; SWR: stepwise regression; XGB: extreme gradient boost.

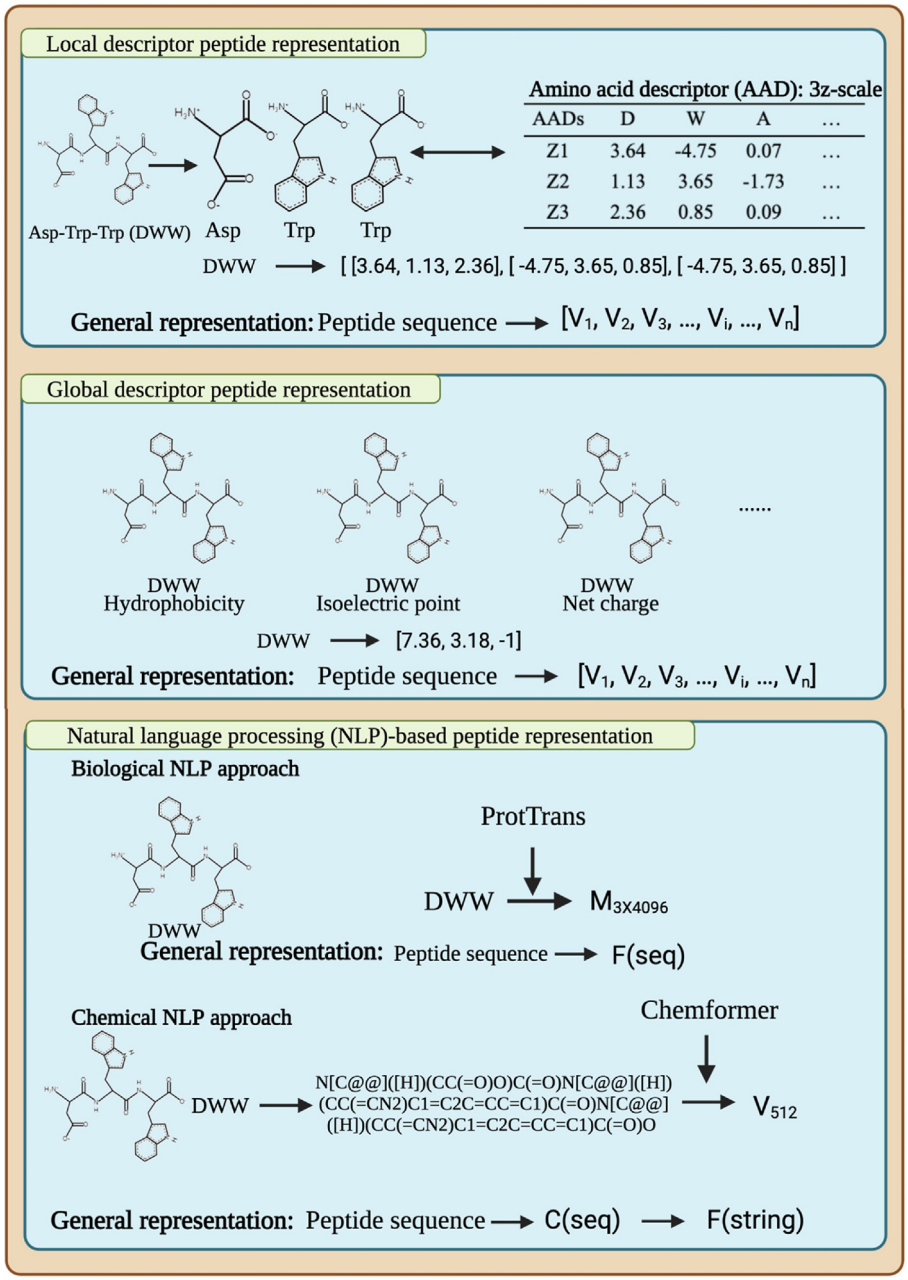

Fig. 3. Overall workflow of peptide representation. Note: Local descriptor peptide representation: V_i is the amino acid descriptor of the ith residue in the peptide sequence, and it is a one-by-k vector; k is the feature dimension of AAD; n is the length of the peptides; the total dimension of one peptide is k*n. Global descriptor peptide representation: V_i is the property of the whole peptide; n is the number of the properties used for peptide encoding. NLP -based peptide representation: the biological NLP approach uses the pre-trained transformer (e.g., ProtTrans) from protein sequences as the dataset for the peptide encoding; the chemical NLP approach uses the pre-trained transformer (e.g., Chemformer) from molecules as the dataset for the peptide encoding. Before the transformation, peptide need to be converted into chemicals strings (e.g., SMILE format).

properties, and the factor analysis scale of generalized amino acid information (FASGAI) is a mixture descriptor extracted from 335 properties including hydrophobicity, alpha and turn propensities, bulk properties, compositional characteristics, local flexibility, and electronic properties [53,70,74]. Combining more AADs can provide more information and better represent peptides. In light of this, the bootstrapping soft shrinkage (BOSS) method was employed to integrate 16 AADs for antioxidant, ACE inhibitory, and bitter peptide representation, and the performance of all QSAR models with integrated descriptors achieved better performance [26,72,76].

Meanwhile, the availability of the original properties for AAD extraction has been gradually improved with the development of amino acid property databases (e.g., AAIndex). Therefore, some researchers turned to extracting AADs from the original dataset [5,75,69]. For example, our lab conducted a comprehensive study [5] where 566 amino acid properties including different physicochemical and biological properties were screened by 6 feature selection methods and 14 machine learning methods for regression model development. The best tripeptide antioxidant activity regression model was obtained by a combination of new AADs extracted from original properties of amino acids by random forest regression and an eXtreme gradient boosting (XGBoost) regression model. This model performed better than the antioxidant activity prediction model that relies only on physiochemical properties and

linear regression methods [5,69]. The 566 amino acid properties and the 16 popular AADs compared in the study of Du, Wang et al., and Deng et al., are included in Supplementary 1 from the previous study [5].

The AAD-based QSAR approach does have an obvious disadvantage. Since descriptors are assigned to each residue, peptides with different lengths will have different feature dimensions after characterization, which results in dataset prerequisites (i.e., peptides with the same length). This limits the utilization of the latest data reported for peptides with different lengths and hinders building a universal QSAR model for a specific bioactivity irrespective of the peptide length [53,70,76,71]. Some researchers proposed modified AAD-based QSAR approaches that only consider residues close to the terminus because some studies showed that these residues had greater influence on bioactivity, and thus peptides with different lengths can be unified to the same feature dimension (Nongonierma & FitzGerald, 2016; Zhu et al., 2022). Peptides with different lengths could then be combined as a larger dataset for model development, although some peptides would suffer from property information loss. The other disadvantage is that the characterization does not consider the peptide as a whole and so does not consider the synergetic effects between different residues.

The information loss mentioned above in AAD-based QSAR modeling was shown in the study of Zhou et al., where support vector machine regression (SVMR) and partial least squares regression (PLSR) were used to build QSAR models based on 80 AADs from 8 different bioactivity datasets [67]. Specifically, two random descriptors generated by a standard normal distribution and a uniform distribution were also used in model development and did not result in a significantly worse performance [67], which meant the prediction power was mainly derived from the modeling method. Even though the authors stated that their AAD-based modeling had almost reached theoretical limits, it should be noted that the datasets used in the study were very small and out-of-date, and accordingly, the study might not reflect the power of AADs very well.

There is an alternative way to overcome the peptide length limitation inspired by one-hot encoding and such attempts have been used to build a cutting-edge bioactivity prediction web server, AnOxPePred (Table 3). Since most reported antioxidant peptides in available antioxidant peptide dataset are composed of the 20 proteinogenic amino acids, a vector with 20 elements (nineteen elements are 0 and one element is 1) was used to represent any amino acids by the position of 1 in the vector. A 20 $\times$ 30 matrix was created for peptide representation (ranging from 2 to 30 residues). A dipeptide was represented by a 20 $\times$ 30 matrix, but only the first and the second row were used. With a large dataset (1404 peptides) and the boost of a convolutional neural network (CNN) for classification model development, this peptide representation approach achieved great performance (accuracy around 80%) [80]. It should be noted that this approach is only feasible when a large dataset is available, and it is not suitable for regression model development.

3.2.2. Global descriptor-based peptide representation

Global descriptors characterize peptides as a whole, which can overcome the peptide length limitation mentioned above and also take the whole peptide structure into consideration for representative vector generation [14]. Theoretically, global descriptors have more advantages than local descriptors.

Molecular descriptors for general chemicals might also be a great alternative for AADs. Descriptor calculation software for chemicals has been well developed (e.g., PaDEL, Dragon, MOE, etc.) [21,28,81,82]. In the study of Li et al., the QSAR model achieved amazing accuracy for IC_{50} (the half maximal inhibitory concentration) of DPP IV inhibitory activity among 39 cysteine-containing dipeptides by using a combination of PaDEL for descriptor calculation, a genetic algorithm for feature selection, and multiple linear regression for model development [21]. A similar attempt was also seen in the study of Wang et al., where 728 peptides with different lengths were collected [28].

Among FBP studies, three-dimensional (3D) QSAR is the most popular approach with global descriptor-based characterization because it reasonably represents 3D characteristics [14,74,82]. In FBP studies, 3D QSAR mostly refers to Comparative Molecular Moment Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA), which have some additional steps compared to AAD-based QSAR model development (Fig. 2). There are also descriptors for 3D QSAR without the need for molecular superposition, but here we specifically focus on CoMFA and CoMSIA analysis. The first step of these analyses is to reconstruct the 3D structure of the peptides and then conduct energy minimization by calculation under force fields. Some software and web servers for structure construction are provided in Table 1. Energy minimization is essential because it makes structure conformation much more reasonable and representative and is generally achieved by search strategies: the steepest gradient descent method (fast) and conjugate gradient method (slow) [9,10,78]. The next step, structural alignment, is the most critical step in developing a reliable 3D-QSAR model [10]. There are, generally speaking, three different alignment strategies: template ligand-based alignment, docking-based (receptor-based) alignment, and scaffold-based alignment (Table 4). The template ligand-based alignment aligns all the peptides in the dataset to the most potent peptides. The docking-based alignment retrieves peptide conformation from molecular docking results and then aligns peptides to the most potent peptides with the consideration of orientation, while scaffold-based alignment does not consider the orientation [10,82,83]. The last step is to calculate the molecular interaction field based on the 3D structure. Based on results from FBP studies (Table 4), it is difficult to judge which method is better: CoMFA analysis or CoMSIA analysis. Some researchers believe CoMSIA was often more robust because it contains additional information from hydrogen bonding groups and hydrophobic regions [14,82].

CoMFA and CoMSIA analyses have all the advantages of global descriptors over local descriptors [9,14]. Furthermore, 3D-QSAR can be easily combined with molecular docking to explain the peptide ligand-receptor interaction mechanism, because the descriptors (e.g., steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor fields in CoMSIA) are the interaction forces in molecular docking and can be visualized as a contour map for the contribution distribution of different fields [10]. However, these approaches also face some challenges. First, peptide molecules are highly flexible, and energy minimization usually only samples a single energy minimum near the initial structure. If there are multiple biologically relevant structures, some may be missed. In addition, compared to AADs, global descriptors cannot reflect how an amino acid residue at a certain site affects bioactivity, so they cannot directly aid rational design at the residue level [5,14,82].

3.2.3. Natural language processing (NLP)-based peptide representation

Natural language processing (NLP) has been applied for encoding peptides by building pre-trained models for feature extraction from peptide sequences. One attempt (BERT4Bitter in Tables 1 and 3) was conducted in 2021 to update the state-of-the-art classification model for bitter peptide prediction, in which three NLP-based peptide encoding methods were employed for peptide representation [84]. A cutting-edge modeling architecture (bidirectional encoder representations from transformers (BERT)) is

emerging recently for protein sequence encoding, which can transform the words in a sentence into a numerical vector with the consideration of the sentence context and has exhibited great performance in various NLP tasks. Correspondingly, each amino acid residue in a protein sequence (sentence) can be considered as a word in the "sentence" [85,86]. Such models are based on the comprehension of the available protein sequences through building self-supervised learning models. This approach relies on more than 200 million protein sequences and theoretically has a better capacity to internalize the information encrypted in protein sequences [85,87]. Several BERT-based protein language models (pLMs) (e.g., ESM- 2, ProteinBERT, ProtTrans) have been released in the past two years and achieved great performance in downstream classification tasks, such as subcellular location, structure prediction, function prediction, etc. [85,87,88]. Very recently, our lab has developed a universal model architecture (UniDL4BioPep) by employing the latest protein language model ESM-2 with a CNN model. It was tested on 20 bioactivity dataset prediction tasks and achieved better performance than the respective state-of-the-art models in 15 out of 20 datasets [89]. Meanwhile, the model development for thee bioactive peptide datasets based UniDL4BioPep does not require feature selection or hyperparameter tuning. This successful attempt supports the feasibility and superiority of pLM-based peptide representation. The great potential of pLMs in peptide representation for model performance improvement is expected to gain more attention in the future, though the model method lacks explainability.

Besides, recent progress in drug discovery might also help advance global descriptor-based characterization. Chemicals are often encoded by the simplified molecular input line entry system (SMILES), which is a string of letters, numbers, and symbols. These strings can in turn be subjected to NLP, which can categorize strings and build models to screen other chemicals [90]. The SMILES format provides sufficient information to construct the molecules in atomic detail, and, therefore, it can be thought to represent the global structural information of the chemical. The latest transformer, Chemformer released in 2022 was pretrained through 100 million SMILES strings and can be used to encode the SMILES string of the peptide sequences. Unlike typical drug-like molecules, which are often smaller and contain mostly rigid groups, peptides with more than a few amino acids are relatively flexible and can have many accessible conformations. Thus, its practical application in peptides may be limited [90]. More information may need to be encoded in descriptors or strings to describe peptide conformational ensembles, especially for long peptides that can form secondary structures (>4 residues) [90–93].

3.3. Model development

Model development is the process to connect representative vectors of peptides and their bioactivity/properties. Recent progress in machine learning has benefited from numerous biochemistry studies that have increased model robustness and accuracy [5,14,94,95]. Generally speaking, model development includes three steps: feature/variable selection, model building, and performance evaluation. Feature/variable selection as well as model building and performance evaluation have no specific prerequisites. A previous review from our lab reviewed in detail popular feature selection methods and model building and performance evaluation methods to compare their advantages and disadvantages [95]. The review here only focuses on the models used in Table 4 (recent QSAR studies on FBPs) and Table 3 (web servers).

All the studies in Table 4 created a regression model, and most of them adopted AADs or global descriptors plus PLSR for model building. Because AADs have gone through a round of feature selection from original properties, it is not necessary to conduct one more feature selection before model building [53,67,73]. In some cases, researchers integrated different AADs to have as much information as possible for model building. In those cases, feature selection would be necessary since there would be a lot of redundant features among different descriptors, which will undermine the model [26,72,76]. For researchers who characterized peptides from the individual properties of amino acids, feature selection would be indispensable to reduce feature dimension and remove redundant features [5,69]. For global descriptors, the criteria for whether to adopt feature selection were the same for AAD descriptors.

As for model building methods, traditional machine learning methods are the mainstream because they work well with limited dataset sizes compared to emerging deep learning approaches. Linear regression methods (e.g., PLSR) can calculate feature importance by variable importance in projection (VIP) values and illustrate which peptide residues or whole peptide properties are more influential [9,28,95,96]. Non-linear regression methods generally perform better than linear regression, but some of them require complex model processes, and sometimes it is difficult to figure out the contribution of each feature to the predicted bioactivity because of their poor explainability (e.g., artificial neural network) [95]. The regression models in Table 4 can give the specific bioactivity values of a peptide, which equips the model with more potential for precise screening. In terms of model performance parameters in antioxidant activity prediction, an improved regression model among available QSAR regression models was achieved by our lab, where the R_p^2 was 0.847 using a non-linear random forest regression (RFR) model [5].

Table 3 collects the latest QSAR web servers developed by bioinformatics researchers. Their feature selection and modeling methods are diverse, and some of them adopt advanced machine learning methods (e.g., CNN) [80,97,98]. Unlike the regression models in Table 4, these QSAR web servers (except AHTpin for dipeptides and tripeptides) are classification models, whose responses (Y) in the model development are labels (e.g., active or not active) instead of the specific activity (e.g., IC_{50} = 1 μM). Such models can only predict which group an unknown peptide belongs to (e.g., high activity (IC_{50} < 10 μM), medium activity (10 μM < IC_{50} < 100 μM), low activity (100 μM < IC_{50} < 1000 μM), or non-activity (IC_{50} > 1000 μM) and, therefore, are much more suitable for rough screening. In addition, these classification models are all based on larger datasets by unifying the feature dimensions of different peptide lengths or by creating new global descriptors [80,99]. For example, PeptideRanker is the most popular binary classification web server among FBP studies. Developed in 2012, it used 1330 short peptides (4–20 amino acids) and 4731 long peptides (>20 amino acids) for its neural network classification model. Its method of peptide representation was not clearly explained. One thing that should be mentioned is that PeptideRanker's dataset for model development was composed of bioactive peptides retrieved from peptide databases (i.e., BIOPEP, PeptideDB, APD2, and CAMP), including peptide hormones, antimicrobial peptides, toxin/venom peptides, antifreeze proteins, antibacterial peptides, antibiotic peptides, and anticancer peptides [99]. It is obvious that the number of peptides with each type of bioactivity was not equally distributed in PeptideRanker's original dataset, which caused bias when it was used to predict the potential for bioactivity. Several web servers have been recently developed to predict specific types of bioactivity and, therefore, are likely to be more accurate when a specific type of bioactivity is of interest (Table 3). Among recent FBP studies, only two studies (L. Wen et al. and Sansi et al.) employed the latest immunomodulatory activity

prediction sever NetMHCpan 4.0 (released in 2020) and antimicrobial bioactivity prediction server $CAMPR_3$ (released in 2014) for virtual screening [100–102]. This is expected to change in the future as the collaboration between biochemistry and bioinformatics studies increases.

3.4. Current status of QSAR application in FBP virtual screening

When employed in practical wet chemistry studies to screen potentially highly bioactive peptides, QSAR models incur a lower cost than traditional methods. The latest QSAR applications in FBP studies are summarized in Table 4, and advanced QSAR web servers are reviewed in Table 3.

There is usually a time lag between the release of advanced QSAR models in the bioinformatics field and their wide application to biochemistry studies [41,43,99]. As shown in Table 1, many database-driven FBP studies adopted the PeptideRanker tool to evaluate the possibility of bioactivity for previously unreported peptides generated by *in silico* proteolytic simulation. Peptides predicted to have a high probability of bioactivity were then synthesized and tested by *in vitro* or *in vivo* experiments. Compared to blind experimental searches, these approaches can accelerate FBP exploration, but the nonspecificity for general bioactivity prediction (e.g., PeptideRanker in Section 3.3) also to some extent hinders their efficiency [24,29,35,36,41,43,100,44,52].

Some researchers with food science backgrounds have participated in QSAR model development and applied the developed models in FBP screening studies. For example, a multi-bioactive tripeptide, IRW, was found to have antihypertensive, anti-inflammatory, and antioxidant properties in both *in vitro* and *in vivo* studies [103–105]. The initial finding of this valuable tripeptide was based on the AAD-based ACE inhibitory activity prediction QSAR model for tripeptides developed in 2006 [106]. Four years later, the model was further applied to ACE inhibitory prediction of 76 theoretical peptides released from egg proteins, and IRW was successfully screened because of its lowest IC_{50} value [16,106]. Another example involves an AAD-based model for predicting DPP IV inhibitory activity developed in 2016 by Nongonierma & FitzGeral. It is used for FBPs liberated from milk protein under the hydrolysis of enzymes in the intestinal tract. With the additional *in vitro* validation for the synthesized peptides, some high-activity DPP IV inhibitory FBPs (e.g., IPM and LPVPQ) were identified [73]. In 2018, a study based on the model was conducted to predict the theoretically generated peptides from camel milk proteins, and the predicted high-activity peptides were synthesized to validate the results of LC-MS/MS from the experimental protein hydrolysates [61]. Both models successfully obtained some valuable FBPs using their QSAR models, but it should be noted that both models used the determination of the coefficient in cross-validation of the final performance evaluation, which strictly speaking is not objective in model development [9,21,73,76,78,96,106,107]. Most biochemistry researchers tend to use wet chemistry experiments as an additional validation approach, which is a better way to demonstrate a model's value in predicting high activity peptides. However, synthesizing peptides with a large range of bioactivity (as predicted by the QSAR model) instead of only those peptides predicted to have high activity can be a better way to demonstrate the overall performance of the model [5,69]. There is no best solution given the cost of synthesis, so the best way is to choose peptide synthesis based on the practical needs for high activity FBPs or FBPs with a specific bioactivity.

For linear and tree-based regression methods (e.g., PLSR, multiple linear regression (MLR), random forest regression, and XGBoost), the feature importance used in peptide representation for model development is available and can be used to reveal deeper structure-activity relationships [5,69]. In the study of Fu et al., AADs (5z-scale) were used to build PLSR models for ACE inhibitory activity, and the results revealed that the hydrophobicity and side chain bulk of residues at the C-terminus contributed more to the bioactivity for dipeptides, while for tripeptides, the hydrophobicity and electronic properties of residues at C-terminus were most important [53]. Another study selected features for peptide representation from 195 physicochemical properties, which resulted in a better explainability [69]. In the 3D QSAR study of Vukic et al., the favorable effects of steric interactions and electronegativity at the C-terminus in ACE inhibitory activity were highlighted by the CoMFA method [9]. Sometimes, it is difficult to achieve such great explainability when the features used for peptide representation are complex intrinsically. In our lab, 566 physicochemical properties and biochemical properties were included for feature selection, and the features retained in the final list for model development (such as “optimized propensity to form a reverse turn”) were difficult to understand intuitively and do not provide an obvious mechanism for the bioactivity [5]. The same dilemma also exists in AAD-based characterization and global descriptor-based QSAR modeling. For example, the 5 features in the T-scale were extracted from 67 structural and topological variables by PCA, but no information was provided to explain each feature. For the QSAR model developed based on the T-scale, it would be very difficult to track back which properties of the peptides contribute to the bioactivity [108].

3.5. Application of QSAR models in evaluating other properties of FBPs

There is a long way from FBPs to commercial products (e.g., functional foods, nutraceuticals, or pharmaceuticals). It is not wise to synthesize all the FBPs predicted to have high activity for *in vitro* or *in vivo* experiments after a single round of *in silico* rough and refined screening. This is because some of the FBPs might not meet the basic requirements for commercial products [42]. The most common strategy is to conduct a second round of filtration for all the theoretical FBPs based on commercial feasibility (e.g., easy to synthesize, allergenicity, toxicity) before wet chemistry validation [21,42]. Then, the final candidates would be limited to peptides that have a high potential for both the desired bioactivity and economical commercial production. Important properties for commercial product development based on FBPs include bioaccessibility, bioavailability, metabolism, toxicity, allergenicity, excretion, bitterness, etc. [2,7,24,109,110].

Evaluation of some of these properties has been achieved by QSAR models using large datasets and has been established as user-friendly web servers [111,112]. In Table 3, the latest web servers for the final round of screening, with their model development methods and model performance in the benchmark dataset, are summarized. These servers examine taste, cell penetration, plasma stability, metabolism prediction, ADMET, allergenicity, toxicity, and drug-likeness [110]. It should be noted that some of these models were initially developed for general chemicals but can be used for peptides since peptides are chemicals in the broad sense. Among these web servers, ADMET evaluation, which is an integrated one-stop evaluation for the absorption, distribution, metabolism, excretion, and toxicity properties of peptides, is the most popular in FBP screening [21,24,33,35,56,69,42]. Among these web servers, SwissADME was released in 2017 and evaluates absorption, distribution, metabolism, and excretion properties, which were the top priorities in FBP studies [111]. A newly developed tool, ADMETlab 2.0, was released in 2021, which integrated prediction of more properties. However, given its relative newness, so far few FBP studies have adopted it [112]. The same situation is also observed

among other latest prediction servers (e.g., those on plasma stability and bitterness).

Given the current state of FBP studies, it is still highly recommended to include external evaluation in FBP screening to validate accuracy. This additional work will also allow researchers who have expertise in other bioactivity studies to use the data for further experiments.

4. Molecular docking approaches and their application in FBP virtual screening

Molecular docking is a structure-based method for virtual screening where the structure information of both peptide ligands and targets are needed. This method overcomes the limitation of bioactivity value availability in QSAR approaches [68,113]. It aims to find the best matching binding mode between a ligand and a targeted receptor using conformational sampling and a binding affinity scoring function [15,114]. The three-dimensional structures of proteins and other biomacromolecules have been determined by X-ray crystallography, NMR, and increasingly by cryo-electron microscopy. Nearly 200,000 such structures are freely available and include pharmaceutical targets related to various common health issues (e.g., hypertension, diabetes, aging, etc.) [115]. These structures, which include many small proteins and peptides, have also provided training data to make prediction of peptide ligand conformations more readily available and accurate. Both of these factors have stimulated the application of molecular docking in FBP discoveries [113,114].

Molecular docking can be used to visualize the interaction mechanism between ligands and receptors at the atomic level, and this ability makes it the most popular bioinformatics method for elucidating the mechanism of action of FBPs [15,23,116,117]. In addition, the scores associated with the best binding mode of ligands can be used for virtual screening by ranking these scores for different ligands. However, these scores are at best a rough estimation of the standard binding free energy, which is directly related to the association (K_a) and dissociation constants (K_d), but is not necessarily correlated with specific bioactivity values (e.g., IC_{50}) [15,118,119]. Molecular docking has demonstrated its great potential in drug discovery (since 1980s) and has also recently employed for FBP virtual screening [114,120–123]. Given the availability of a 3D structure for the receptor and reasonable coordinates for the ligand, molecular docking includes two critical parts. The first is sampling of different locations, orientations, and conformations of the ligand relative to the receptor. In the context of docking, each configuration (location/orientation/conformation) is referred to as a “pose”. The second part is assigning a “score” to each pose that represents its thermodynamic favorability. These scores are the quantity that the optimization algorithm of the docking program seeks to optimize. Well-designed scoring functions can also be used to compare the best-scoring poses of distinct ligands, enabling virtual screening. Table 5 summarizes popular molecular docking software and web servers with detailed information on sampling algorithms, scoring functions, availability, flexibility, maintenance, and more.

4.1. Conformational sampling in molecular docking

Conformational sampling (also known as conformation search) attempts different conformations of the ligand around the whole surface/cavities of the receptor (global docking) or at a designated binding site (local docking) [91,113,124]. The number of accessible conformers for peptides of more than a few amino acids is astronomical; hence, it is not feasible to computationally generate all the possible conformations for scoring. Furthermore, for each conformation of the ligand, many orientations and positions relative to the receptor must be considered. In addition, different conformations of the receptor may be considered. A major difference among docking tools and protocols is which portions of the ligand and receptor are regarded as flexible, meaning they are subjected to the conformational sampling algorithm, and which are treated as rigid, meaning that they retain the conformation originally provided by the user throughout the docking process. Some docking tools, such as those intended for protein–protein docking, treat both the ligand and receptor as rigid and “conformational” sampling entails only overall rotations and displacements of the provided ligand structure [91]. Such an approach is usually unsuitable for docking of peptides, which typically have many internal degrees of freedom and a diverse ensemble of thermodynamically accessible structures. The most commonly used approach in FBP studies is semi-flexible docking where the ligand is flexible, while the receptor is rigid [15]. Many docking programs also support treating chosen side chains of the receptor (typically those near the binding site) as flexible groups. However, side chain flexibility is not sufficient to permit sampling of distinct receptor conformations that involve changes in protein backbone positions, such as inward-facing and outward-facing states of transporters. Hence, in all cases it is essential to choose an initial receptor conformation that is relevant for the bioactivity of interest. For some applications, it may be necessary to perform docking of a ligand over multiple distinct conformational states of the receptor. Indeed, given the inherent flexibility of proteins, improved performance has been found with ensemble docking, where the ligand is docked to multiple snapshots of the receptor protein obtained from molecular dynamics simulation [125].

The commonly used semi-flexible docking approach is a compromise between computational effort and exhaustive sampling of accessible conformations [92,114]. It should be noted that more flexibility does not guarantee a significant improvement in docking performance, especially since greater flexibility usually comes at the cost of less exhaustive sampling. A performance assessment by Huang demonstrated that semi-flexible docking generated bound poses nearer to experimentally derived structures than rigid docking, although ranking of ligands for virtual screening showed no clear difference in performance between rigid and flexible approaches [126]. However, the study of Huang considered drug-like molecules, which are typically more rigid than many-residue peptides. For peptides of more than a few amino acids, it is likely that consideration of multiple conformations of the peptide is necessary unless it is known to have well-defined folded structure.

Various algorithms are used for sampling ligand and receptor conformations with the available computational power [68]. The algorithms can be generally divided into three categories based on their searching strategies: systematic search, stochastic search, and deterministic search (dynamics simulation search), and some docking software combines different strategies into a multi-phase approach (e.g., GLIDE, CDOCKER, and DOCK 6) [124,127–129]. Briefly, the systematic method is more time consuming than the stochastic method (e.g., the Monte-Carlo methods or genetic algorithm) because of it includes an exhaustive search of possible rotamer states for subsets of the molecule. The computational demand of deterministic search is the largest and proportional to the dynamics simulation runtime. This approach is highly sensitive to the initial conformation and spatial position because they can change very slowly in dynamics simulation. Therefore, a deterministic search can easily be trapped in local minima unless long simulation times are used to cross the barriers or tempering strategies to accelerate this crossing are applied, which consume more time [113,130]. Therefore, deterministic search is usually integrated

Table 5
Summary of molecular docking software/web servers, dynamics simulation software, free energy calculation tools, and force fields.

Molecular docking software						
Name	*Sampling algorithm*	*Scoring function*	*Additional comments*	*Successl rates**	*Availability*	*Websites*
AutoDock 4	Stochastic algorithm (Lamarckian genetic algorithm)	Force field-based and empirical scoring function (AD4)	The basic version uses flexible ligands and a rigid receptor, but it also has specific modified solutions for hydrated docking, zinc metalloprotein docking, flexible docking (flexible residues for receptors), and multiple ligand docking. AutoDock4 can be up to	53%	Free for academic use	https://autodock.scripps.edu/download-autodock4/
AutoDock Vina	Stochastic search (Monte-Carlo/BFGS searching)	Knowledge-based and empirical energy scoring function (Vina)	100 × slower than AutoDock Vina. AutoDock Vina has a batch mode for a large number of ligands and also has modified versions (e.g., QuickVina2, Vinardo, and InstaDock). Both AutoDock4 and AutoDock Vina are continually maintained and updated.	80%	Free for academic use	https://vina.scripps.edu/
AutoDockFR	Stochastic algorithm (genetic algorithm and Solis-Wets local search)	Force field-based and empirical scoring function (AD4)	It uses receptors with flexible side chains and flexible ligands.	74%	Free for academic use	https://ccsb.scripps.edu/adfr/
AutoDock CrankPep	Stochastic search (Monte-Carlo searching)	Force field-based and empirical scoring function (AD4)	It uses rigid receptors and flexible ligands. It is specially designed for protein-ligand docking.	85.7%	Free for academic use	https://ccsb.scripps.edu/adcp/documentation/
DOCK 6	Systematic conformational search (anchor-and-grow search algorithm/BFGS)	Empirical energy scoring function	It uses rigid receptors and flexible ligands. There are also six more scoring functions and deterministic search options.	73.3%	Free for academic use	https://dock.compbio.ucsf.edu/DOCK_6/index.htm
CDOCKER	Stochastic search (simulated annealing) and deterministic search (MDS + energy minimization)	Force field-based scoring function (soft-core potential)	It uses rigid receptors and flexible ligands. The last update was made in 2016 and called Flexible CDOCKER (62.7% successful rate), which can set flexibility for receptors and was accelerated by SGLD. Flexible CDOCKER can remove the simulated annealing procedure.	74%	Need to purchase; built in Discovery Studio	https://discover.3ds.com/discovery-studio-visualizer-download
LibDock	Systematic conformational search (geometric hashing algorithm/BFGS)	Empirical energy scoring function (Ligscore)	It uses rigid receptors and flexible ligands. No update was reported after 2007.	46%	Need to purchase; built in Discovery Studio	https://discover.3ds.com/discovery-studio-visualizer-download
LigandFit	Stochastic search (Monte-Carlo searching)	Empirical energy scoring function (Ligscore)	It uses rigid receptors and flexible ligands. No update was reported after 2003.	–	Need to purchase; built in Discovery Studio	https://discover.3ds.com/discovery-studio-visualizer-download
GOLD	Stochastic search (genetic algorithm)	Empirical energy scoring function (Chemscore) or knowledge-based scoring function (GOLD)	It uses partial flexibility for receptors and flexible ligands.	81%	Need to purchase	https://www.ccdc.cam.ac.uk/solutions/csd-discovery/
PLANT	Stochastic search (ant colony algorithm)	Empirical energy scoring function ($LANTS_{CHEMPLP}$ and $PLANTS_{PLP}$)	It uses partial flexibility for receptors and flexible ligands.	72%	Free for academic use	http://www.tcd.uni-konstanz.de/research/plants.php
Glide	Systematic conformational search (hierarchical searching) and deterministic search algorithm	Empirical energy scoring function (GlideScore) or empirical energy scoring function (Emodel)	It uses partial flexibility for receptors and flexible ligands. It uses GlideScore to rank ligands and Emodel to find the best conformation. It is continually maintained and updated.	82%	Need to purchase	https://www.schrodinger.com/products/glide
Surflex-Dock	Systematic conformational search (incremental construction search)	Force field-based and empirical scoring function	It uses a rigid receptor and flexible ligands. It is continually maintained and updated.	78.6%	Need to purchase	https://www.biopharmics.com/downloads/
MOE	Stochastic search (triangle matcher algorithm)	Force field-based scoring function (S value)	It uses a rigid receptor and flexible ligands.	61.2%	Need to purchase	https://www.chemcomp.com/Products.htm

Molecular docking web servers

	Website	*Type*	*Additional comments*	*Availability*
CB-Dock	http://clab.labshare.cn/cb-dock/php/index.php	Global docking	Cavity-detection guided global docking based on Autodock Vina	Free
PepSite2	http://pepsite2.russelllab.org/	Global docking	Stochastic search (spatial position specific scoring matrix) and knowledge-based scoring function (hot spot score)	Free
SwissDock/ EADock DSS	http://www.swissdock.ch/	Local/global docking	Stochastic search (EADock dihedral space sampling) and Force field-based and empirical energy scoring function (EADock2)	Free
HPEPDOCK	http://huanglab.phys.hust.edu.cn/hpepdock/	Global docking	Systematic conformational search (hierarchical searching) and knowledge-based and empirical energy scoring function	Free
HADDOCK	https://wenmr.science.uu.nl/haddock2.4/	Global docking	Systematic conformational search (an experimental knowledge-driven searching method) and energy scoring function (HADDOCK score)	Free
CABS-dock	http://biocomp.chem.uw.edu.pl/CABSdock	Global docking	Fully flexible receptors and ligands; deterministic search (CABS coarse-grained protein model) and empirical scoring function (energy scoring and structural clustering)	Free
PIPER-FlexPepDock	http://piperfpd.furmanlab.cs.huji.ac.il/	Global docking	Systematic conformational search (fragment-based searching) and Force field-based and empirical scoring function (FFT docking algorithm)	Free
FlexPepDock	http://flexpepdock.furmanlab.cs.huji.ac.il/	Local docking	Fully flexible refinement; stochastic search (Monte-Carlo sampling with energy minimization) and Force field-based scoring function (Rosetta score12 and Rosetta centroid score4)	Free
InterEvDock3	http://bioserv.rpbs.univ-paris-diderot.fr/services/InterEvDock3/	Template-based docking	Mainly used for protein-protein docking and can also be used for protein and long peptide docking. Systematic conformational search (FRODOCK algorithm) and Force field energy scoring function (InterEvScore)	Free

Molecular dynamics simulation software

Name	*Website*	*Availability*
NAMD	https://www.ks.uiuc.edu/Research/namd/	Free and open source for academic users
GROMACS	https://www.gromacs.org/	Free and open source
OpenMM	https://openmm.org	Free and open source
CHARMM	http://www.charmm.org/	Free for academic users
LAMMPS	https://lammps.org	Free and open source
AMBER	http://ambermd.org/	Need to purchase

Force fields

Name	*Website*	*Availability*
CHARMM	http://mackerell.umaryland.edu/charmm_ff.shtml	Free
AMBER	https://ambermd.org/AmberModels.php	Free
GROMOS	https://www.igc.ethz.ch/gromos.html	Free
OPLS	http://zarbi.chem.yale.edu/oplsaam.html	Free
KBFF20	https://kbff.chem.k-state.edu/	Free

Free energy calculation tools

Name	*Website*	*Availability*
YANK	https://getyank.org	Free
BFEE2	https://github.com/fhh2626/BFEE2	Free
pAPRika	https://paprika.readthedocs.io	Free
BRIDGE	https://github.com/scientificomputing/bridge	Free

Note: * means the root-mean-square deviation of the difference between the predicted molecular conformation and the experimental conformation is lower than 2.0 Å. The success rate can only be used to compare the performance of different docking tools when they have the same benchmark.

Abbreviations: AIRs: Ambiguous Interaction Restraints; BFGS: Broyden–Fletcher–Goldfarb–Shanno algorithm; FFT: Fast Fourier transform; MDS: molecular dynamics simulation; SGLD: self-guided Langevin dynamics.

Table 6
Selected virtual screening strategies using molecular docking & molecular dynamics simulation.

Protein source	Bioactivity	Molecular docking	Molecular dynamics simulation	Additional comments	In vitro or *in vivo* experiments	Reference
Fermented soy	Keap1−Nrf2 interaction inhibitory activity	CABS-Dock, GalaxyPepDock, and HADDOCK	–	28 peptides identified by LC-MS; CABS-Dock and GalaxyPepDock were used for screening, and HADDOCK was used for conformation refinement	Thirteen peptides were selected for *in vitro* and *in vivo* antioxidant experiments	[136]
Egg protein	Keap1−Nrf2 interaction inhibitory activity	CDOCKER	–	Fluorescence polarization assay was used to further refine the FBPs selected by CDOCKER	HepG2 cell model for oxidative damage, cytotoxicity, and cytoprotection evaluation	[7]
Pea protein	DPP IV inhibitory activity	AutoDock Vina	–	30 peptides identified by LC-MS and contained P/A at the second position at N-terminus were further screened for peptide synthesis	Eight peptides were synthesized for *in vitro* DPP IV inhibitory activity assay	[66]
Pumpkin seed	ACE inhibitory activity	MOE	GROMACS for 30 ns of simulation; Force field: Amber99SB-ILDN	Identified 47 di/tripeptides virtually screened by MOE; ADMET evaluation to select for peptide synthesis	Tripeptide IFA was synthesized for *in vitro* ACE inhibitory activity assay	[42]
–	ACE inhibitory activity	GOLD	AMBER12 for 25 ns of simulation; Force field: AMBER FF03	8000 theoretical tripeptides were ranked by GOLD, and MDS was used to elucidate interaction mechanism	Five peptides (WCW, IWW, WWW, WWI and WLW) were synthesized for *in vitro* ACE inhibitory activity assay	[135]
Amaranth protein	Renin inhibitory activity	CABS-dock and FlexPepRDock	–	CABS-dock for ligand-receptor docking; FlexPepDock for conformation refinement and peptide-protein interaction energy evaluation (Rosseta score)	Four peptides were synthesized for *in vitro* renin inhibitory activity assay	[8]
Mytilus edulis proteins	Thrombin inhibitory activity	CDOCKER	–	39 peptides identified by LC-MS and further screened by CODCKER	KNAQNQLGEVTVR was synthesized for *in vitro* thrombin inhibitory activity assay	[120]
Bovine milk casein	Antithrombotic activity	CDOCKER	–	35 peptides identified by UPLC-Q-TOF-MS/MS and further screened by CODCKER	No peptide synthesis for validation	[123]
Walnut meal	Tyrosinase inhibitory activity	AutoDock 4 and CDOCKER	–	606 peptides identified from LC-MS were screened by AutoDock 4; CDOCKER was used for conformation refinement	FPY was synthesized for *in vitro* experiment	[6]
Sesame seeds	Tyrosinase inhibitory activity	AutoDock 4	–	Eight peptides were selected from 361 peptides reported to exhibit antioxidant activity from *in silico* proteolysis simulation	No peptide synthesis for validation	[56]
Oncorhynchus mykiss Nebulin	Umami	CDOCKER	–	332 peptides were obtained from *in silico* proteolysis simulation and further screened by CODCKER	20 peptides were synthesized and tested by electronic tongue	[23]

with stochastic search or systematic search for the final conformation optimization (e.g., CDOCKER, GLOD, and Glide) (Table 5).

4.2. Scoring function in molecular docking

The scoring function is the conformation selector for the results from the sampling engine. It selects by estimating the difference in energy (or free energy) between the ligand−receptor complex and the isolated unbound molecules [114,124,129]. There are three types of scoring functions: force field-based scoring functions, empirical scoring functions, and knowledge-based scoring functions [118]. Force field-based scoring functions are composed of a series of energy terms from a classical force field, including bonded (intramolecular) and nonbonded (intermolecular) components, and in some cases, the solvation terms are also included, but they are more computationally expensive [68]. Empirical scoring functions have relatively simple energy terms for calculation by assigning weights to different empirical energy terms (e.g., hydrogen bonding, ionic bonding, non-polar interactions, desolvation, and entropic effects), and the weights are obtained by training models for the experimental data [114,124,131]. Knowledge-based scoring functions are computationally simple and based on statistical analysis of interaction atom pairs from complexes with experimentally derived 3D structures; the frequency of the ligand-receptor atom pairs is computed for scoring [114,132]. There are also some integrated strategies (also called consensus scoring) that attempt to take advantage of different scoring functions. An example is AutoDock Vina, which combines knowledge-based and empirical scoring functions for scoring and ranking.

4.3. Other considerations in molecular docking

Docking studies can also be characterized as local, where a known active site of the receptor is targeted, or global (also known as blind docking), where the active site is unknown. Some docking servers (such as CBDock) will first perform an analysis to locate possible binding cavities and then apply local docking in turn to each cavity [133]. Most of the targeted proteins for FBP studies have known active pockets, so there is no need to conduct global docking, but some web servers (e.g., Pepsite2, CABS-dock, HADDOCK, etc.) adopt global docking by default [50,65,96,134]. For all docking software, users can choose whether to conduct docking over an important part of the receptor (local) or over the entire receptor (global), although the latter requires more computational effort or reduced exhaustiveness.

4.4. Application of molecular docking in FBP virtual screening

In FBP studies, the most common molecular docking technique is to identify potential bioactive peptides by LC-MS from the highest activity fractionation or proteolysis simulation and then

conduct molecular docking of these peptides on targeted proteins (e.g., ACE, renin, DPPIV, etc.) (Table 6). The peptides with the most favorable docking scores can then be synthesized for *in vitro* or *in vivo* bioactivity determination, and the docked conformations can help elucidate the interaction mechanism, highlighting particular hydrogen bonds, hydrophobic contacts, π−π stacking, or other such interactions between protein receptors and peptide ligands [66,42,120,122,123]. This approach does not lend itself to large-scale virtual screening because of the difficulty of identifying and synthesizing large numbers of peptides [92]. The study conducted by L. Li et al. on Keap1−Nrf2 interaction inhibitory FBPs was a great example for future molecular docking-driven FBP screening. In the study, a total of 400 dipeptides and 6138 tripeptides were screened by CDOCKER, and six dipeptides and ten tripeptides with stronger binding affinity were synthesized for *in vitro* and *in vivo* experiments. Finally, two tripeptides (DKK and DDW) with strong inhibitory activity were successfully obtained [7]. Another such attempt employed GOLD to screen ACE inhibitory activity from 8000 theoretical tripeptides, of which the five with the highest binding affinity were synthesized for experimental validation [135].

Molecular docking involves many approximations and, even for exhaustive sampling, the resulting poses are not guaranteed to be correct nor is there a guarantee that the docking score closely corresponds to experimental inhibition activity, as shown by the inconsistency between binding affinity rank and experimental activity results in the studies of Panyayai et al., and X. Li et al. [7,135,137]. Some researchers proposed a new docking strategy called consensus docking, which has shown a performance improvement compared to single docking protocols in the reliability of pose prediction for interaction mechanism determination and virtual screening [138,139]. In consensus docking, different molecular docking protocols are used for the same docking task, and then the binding poses from different docking protocols are compared to calculate the root-mean-square deviation (RMSD) value. A molecular docking prediction will be accepted when the RMSD value between different protocols is below 2 Å [139]. There is a well-designed one-stop python package (dockbox, available at https://pypi.org/project/dockbox/) for consensus docking or docking by different protocols [140]. Another more straightforward strategy for consensus docking relies on summation of normalized docking scores from different docking protocols, and corresponding tools have been released for academic free usage (e.g., MolAr software and the DockingPie plugin for PyMol) [141,142].In the studies of Tonolo et al. and Nardo et al., a similar idea was adopted where two docking protocols were employed consecutively to refine the docking results, but neither study adopted RMSD threshold values or normalized summation of docking score for refinement [8,122]. By combining large-scale screening for FBPs and consensus docking, we can hope to build molecular docking binding affinity databases for small peptides.

Although consensus docking can improve the reliability of docking results for unknown peptides, we also need to assess the correlation between identified FBPs with bioactivity values and the binding affinity results from molecular docking. The only one such study was conducted in 2007 by Pripp, where the coefficient is only 0.29 between the docking score of 29 ACE inhibitory FBPs with ACE and the experimentally observed $\log(1/IC_{50})$, which means the prediction power of molecular docking in the virtual screening of ACE inhibitory peptides was poor [143]. With the increasing availability of FBPs and improvements in molecular docking strategies in the past decades, especially the introduction of consensus docking, we may hope to see performance improvement in the near future.

Finally, there are two areas of concern in molecular docking. First, unlike conventional drug-like molecules that may include a wide variety of chemical groups but typically have a limited number of accessible conformations, most peptides are composed of only 20 different amino acids but have high conformational flexibility. As such, more information in the representative descriptors or strings may be needed to describe their conformation flexibility, especially for long peptides that form secondary structures (>4 residues) [37,90,92]. Table 1 includes some secondary structure prediction tools for molecular docking of long peptides and protein receptors when a secondary structure is needed (e.g., CABS-dock). Second, unless an allosteric binding site is known and considered as part of the docking region, molecular docking is not suitable for virtual screening of non-competitive inhibitors, which may be another important factor that causes inconsistency between *in vitro* or *in vivo* studies and molecular docking [144].

5. Molecular dynamics simulation and its application in FBP virtual screening

Molecular dynamics simulation is also a structure-based approach for virtual screening and can explore interaction mechanisms at the atomic level, but is typically much more computationally demanding than molecular docking [22,113]. It is typically considered to be more accurate than docking and better able to give physical insight [145], compensating for its greater computational cost. In molecular dynamics simulations, the behavior of molecules (not limited to ligands and receptors) over a period of time under the constraints of physical laws (classical or quantum theories) are captured [15,114]. Quantum chemistry methods, which explicitly treat all or some electrons at the quantum mechanical level, are prized for yielding highly accurate energies and molecular geometries with few or no empirical parameters. However, despite methodological improvements in the past couple decades, the computational expense of quantum mechanics-based methods remains prohibitive for probing the behavior of proteins surrounded with explicit water molecules on time scales relevant for binding processes (often microseconds or more). On the other hand, classical molecular dynamics simulations can be used to simulate complete proteins and ligands in an aqueous environment for time scales currently reaching many microseconds (or even milliseconds with specialized hardware [146]). Classical molecular dynamics is based on numerical solutions of Newton's equations of motion (or similar equations for constant temperature or pressure thermodynamic ensembles) for collections of atoms. The downside of this method is that the accuracy of the results is dependent on the accuracy of the empirical potential energy functions, termed “force fields”, that describe interactions between atoms (or particles representing groups of atoms in united atom or coarse-grained models). Fortunately for the study of peptide−protein binding, force fields describing polypeptides consisting of the 20 proteinogenic amino acids in water have been developed and constantly improved over the past three decades [147−153].

The most practical consideration for molecular dynamics simulation is the selection of the force field. Commonly used force fields and molecular dynamics simulation software are summarized in Table 5. The most common force fields for protein and peptide simulations are the CHARMM and Amber force fields, which are based on models where each atom is represented as a point particle carrying a partial charge and the network of covalent bonds is fixed at the beginning of the simulation [149,153]. Therefore, while conventional molecular dynamics simulations can represent interactions that give rise to physical bonding, including hydrogen bonds, salt bridges, the hydrophobic effect, and other solvation-dependent interactions, they are unable to directly describe chemical reactions and changes in protonation state and may not accurately represent changes in electrical polarization

between different environments. Techniques and alternate force fields to overcome these limitations have been developed; however, they often incur increased computational cost and, for typical peptide–protein simulations, improved accuracy is not guaranteed. There can be trade-offs between systematic errors, which are reduced by more accurate models, and statistical errors due to insufficient sampling, which can be increased by using more accurate models since these models are often slower and the real time available to researchers is fixed.

5.1. Molecular dynamics simulations for protein–ligand binding

Conventional molecular dynamics simulations (with fixed atomic charges, protonation states, and covalent bonds) have become mainstream for studying peptide-protein binding, including for studying FBPs [154,155]. By putting peptide ligands and protein receptors into a simulation box, typically along with explicit water molecules and dissolved ions, one can simulate protein-ligand interaction for anywhere between femtoseconds to milliseconds [146] at a given temperature and pressure. If the kinetics of binding between the ligand and receptor is sufficiently fast, a protein-ligand complex will be observed after a period of dynamics simulation. In the case of very fast kinetics and low binding affinity, multiple binding and unbinding events can be observed, allowing for direct estimation of the equilibrium constants. However, for peptide ligands, the kinetics will be almost always much too slow to estimate equilibrium constants from brute force simulation. Including multiple ligands in the simulation box can help to observe spontaneous binding, although there is no guarantee that any of the binding poses found are the lowest energy [156,157]. Docking is often a more efficient way to search for and find putative bound conformations of protein–ligand complexes. Hence, a typical molecular dynamics approach to virtual screening is to perform docking to obtain multiple poses of the bound ligand in the protein binding site. Each of the poses can be solvated in explicit water and ions and subjected to molecular dynamics simulation. Restraints are usually applied during the energy minimization and equilibration stages of the simulation to avoid prematurely disrupting the complex while water molecules, ions, and some parts of the protein may be far from equilibrium. Such molecular dynamics simulations can help to identify which poses predicted by docking are truly stable [137,158]. For example, if for a given pose, the peptide dissociates from the protein a few nanoseconds after equilibration for several different initial conditions (different initial atomic velocities or water molecule positions), then it can be assumed that the binding affinity for this pose is marginal.

Numerous methodologies have been used to estimate free energies of protein-ligand binding for virtual screening and the field continues to develop rapidly [145]. The SAMPL challenges provide a snapshot of methods currently used for estimating binding free energies from molecular dynamics simulations (or by other means such as machine learning) and some indication of which methods may work better than others [159–162]. While these SAMPL challenges focus on host-guest systems, which are likely easier than protein-ligand systems due to the relative symmetry and rigidity of the hosts compared to proteins, they reveal aspects of the calculation that need to be considered for accurate results such as the details of the computational protocol, conformational sampling, multiple bound conformations, minor protonation states, and atomic polarizability.

At the present time, molecular models with fixed atomic charges, protonation states, and covalent networks and explicit solvent are the most commonly used for simulating peptide–protein interactions; however, for specific cases, more sophisticated methods and force fields might be needed. For instance, peptide therapeutics with a covalent mechanism of action have recently attracted increased interest [163] and can be studied by hybrid methods combining classical molecular dynamics and quantum chemistry methods [164]. Protein force fields that explicitly represent atomic polarizability have also been developed [165] and are continually being improved [166,167]. However, models with atomic polarizability incur a greater computational cost and have less history of development than force fields with fixed atomic charges and, so far, there is no guarantee of improved accuracy over fixed-charge force fields for all systems. Nonetheless, recent submissions to the SAMPL challenges using a polarizable force field appear have been among the best performers [159,160]. Furthermore, for systems where highly polarizable ions play important roles, force fields including atomic polarizability can be crucial for reasonable results [168].

5.2. Binding free energy estimates with MM-PBSA and MM-GBSA methods

Even if association or dissociation is observed on a time scale accessible to molecular dynamics simulations, a single simulation is not sufficient accurately characterize the equilibrium constant for protein–ligand binding or, equivalently, the binding free energy. As end-point methods, the MM-PBSA (molecular mechanics Poisson–Boltzmann surface area) and MM-GBSA (molecular mechanics generalized Born surface area) methods provide a convenient way to estimate binding free energies from molecular dynamics simulations without the need for observing association or dissociation events. Tools for performing these calculations with the GROMACS and Amber molecular dynamics packages have been developed [169,170]. The basic idea of the methods is use explicit–solvent molecular dynamics simulation to obtain an ensemble of different conformations of bound complex, and separately, ensembles of conformations of the free receptor and ligand [171]. The binding free energy is estimated by post-processing these simulation trajectories using the equation $\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{free ligand}} - G_{\text{free receptor}}$. Calculating these Gibbs free energy terms directly from explicit-solvent simulation is challenging due to the noisiness of the solvent contribution to the enthalpy (water molecules and ions sample many different positions and orientations during the simulation) and the difficulty of calculating changes in the entropy of the solvent [172]. The MM-GBSA and MM-PBSA methods overcome these difficulties by discarding the explicit water molecules and dissolved ions from the trajectory and estimating the solvation contributions to the free energy using continuum models. The polar solvent/ion contribution to free energy is calculated using the continuum Poisson–Boltzmann (PB) or generalized Born (GB) techniques and the nonpolar contribution with a term proportional to the solvent-accessible surface area (SA). Hence, Gibbs free energy terms for each of the three systems (complex, free ligand, free receptor) are calculated by $G = \langle E_{\text{MM}} \rangle + \langle G_{\text{solv}} \rangle + \gamma \langle A \rangle - TS_{\text{conf}}$, where $\langle \dots \rangle$ denotes an average over the simulation trajectory, E_{MM} is the potential energy from the force field for the ligand and receptor (neglecting the explicit water molecules and dissolved ions), G_{solv} is the polar portion of the solvation free energy calculated by Poisson–Boltzmann or generalized Born method and averaged over the trajectory, γA is the nonpolar contribution to the solvation free energy, A is the solvent-accessible surface area of the ligand and receptor molecules, γ is a parameter linearly relating the solvent-accessible surface area and free energy, T is the temperature, S_{conf} is the conformational entropy of the ligand and receptor. The conformation entropy is typically estimated by a normal mode analysis or the quasi-harmonic method, which are both

implemented in the GROMACS package [173–175]. It should be noted that these methods of computing conformational entropy are costly and the inclusion of the conformational entropy term not guaranteed to improve agreement with experiments [176].

5.3. Rigorous binding free energy calculations

MM-PBSA and MM-GBSA methods are not rigorous in the sense that the simulations are typically performed with explicit water, yet the solvation free energy is obtained with a continuum model, and the conformational entropy is not directly calculated, but approximated using normal mode or quasi-harmonic methods. While MM-PBSA and MM-GBSA are more efficient than rigorous methods, they can yield poor predictions for protein-peptide binding free energies where rigorous methods obtain good agreement with experiment [177]. Rigorous methods instead seek to calculate the free energy under a given molecular dynamics model and force field without approximations (although this free energy value may differ from the true experimental value). These methods are not end-point methods and require comprehensive sampling along a path from the bound to unbound states, which requires considerable computational resources for even a single calculation. The path taken between the bound and unbound states may be entirely unphysical, such as in alchemical methods where the atoms of the ligand are gradually deleted or inserted, or the ligand and the receptor may be physically separated along geometric coordinates, which may still not be the most likely path taken in reality [178,179]. However, because free energy is a state function, the path taken between the two states should not affect the change in free energy between the states and, therefore, one can choose an unphysical path that makes the calculation most efficient (or merely feasible). Notably, changes in the conformation and orientation of the ligand relative to the receptor between the bound and unbound states can make convergence of the free energy unacceptably slow. Hence, for both alchemical and geometric pathways, it is often necessary to apply conformational and orientational restraints to the ligand (and possibly parts of the receptor as well) [180]. However, since these restraints are unphysical, their effect must be removed. Therefore, conceptually, one takes the following approach. First, the free energy of applying conformational and orientational restraints to the free ligand in solution is calculated. Next, the free energy of binding to the receptor is calculated under these restraints, which allow relatively rapid convergence. Finally, the free energy of releasing the conformational and orientational restraints is calculated, so that the final free energy is calculated as $\Delta G_{\text{bind}} = \Delta G_{\text{apply restraints}} + \Delta G_{\text{bind-restrained}} + \Delta G_{\text{release restraints}}$. While this approach is quite complex, recently several tools have been released to make simplify performing rigorous free binding calculations, including BFEE [181], BFEE2 [182], YANK, BRIDGE [183], and pAPRika [184].

It should be noted that this rigorous approach requires an initial pose, used as the reference for the conformational and orientational restraints, that is assumed to be an accurate representation of the bound state of the complex. The MM-PBSA/MM-GBSA methods also require an initial pose. Obtaining such an accurate pose can be difficult in the absence of an experimental structure, but can be found by docking and brute-force molecular dynamics simulations. Identifying the lowest free energy bound state may require performing the rigorous or MM-PBSA/MM-GBSA method with different poses. It is also possible that more than one distinct pose contributes significantly to the bound free energy. Hence, it is often beneficial to use a hierarchy of methods for screening. For example, we used FlexPepDock to generate initial poses and docking scores for 17 different peptide sequences, screened them by performing conventional molecular dynamics simulations for up to 2 μs and calculating the MM-GBSA free energies, and, finally, performed rigorous free energy calculations using the BFEE tool for two selected peptides [185].

5.4. Application of molecular dynamics simulation in FBP studies

In FBP studies, the application of molecular dynamics simulation is not as popular as that of molecular docking, which may be caused by the tremendous computational demands and the complexity of the operation [146,186]. It is more likely to be used to supplement molecular docking results by optimizing the conformations of peptide–protein complexes for interaction mechanism elucidation, or it may be used to validate the stability of peptide–protein complexes predicted through RMSD values [18,93,42,121]. For example, two dual-functional peptides (RALP and WYT) were studied by molecular docking and molecular dynamics simulation to understand their difference in inhibitory potency. Both peptide–ACE complexes were intact after 200 ns of simulation, and the two peptides moved out of the active sites of peptide–renin complexes, which demonstrated the instability of the peptide–protein complexes predicted from molecular docking. The instability of this complex was further supported by the low bioactivity [137]. Similar applications were also seen in the studies of Liang et al. and Panyayai et al., where molecular dynamics simulation was used to check the stability of the complex in salt solution under room temperature (300 K) and 1 bar pressure for 30 ns and 25 ns, respectively [42,135]. Such additional work can help explain the inconsistency between predicted and experimental results in FBP studies.

Recently, a user-friendly front-end was proposed to run molecular dynamics simulation protocols in Google Colaboratory. It successfully simplified the running procedures and provided the needed computation power using cloud computing [186]. Such efforts from bioinformatics could enhance the availability of molecular simulations and benefit biochemistry researchers who struggle with limited in-house computing facilities, programming environments, simulation operations, and result analysis.

6. Progress of integrated strategies in FBP screening

The integration of QSAR models, molecular docking, and molecular dynamics simulation has grown popular among biochemistry researchers for FBP virtual screening (Table 7). The most popular strategy is to employ the QSAR model for the first round of screening of peptide sequences from *in silico* proteolysis or LC-MS identification; further screening is then conducted by molecular docking, and sometimes molecular dynamics simulation is used to optimize conformation for the interaction mechanism elucidation [10,21,187–190]. The combination of different virtual screening methods is expected to possess higher accuracy in FBP identification. These methods were established on different theories to describe and characterize peptides for screening, so double screening would be more efficient than refinement under the same theory (e.g., double screening by different molecular docking methods) [8,107,134].

Most of the integrated studies focus on ACE inhibitory activity and DPP IV inhibitory activity [18,21,28,28,37,38,50,65,81,93,96,107,107,134,187,189–191,193]. FBPs with these two inhibitory activities have the potential to relieve two chronic diseases: hypertension and diabetes. As such, they deserve intensive attention from researchers. On the other hand, the bioinformatics approach is a kind of knowledge-based study, which means the more information that is available for one type of bioactivity, the better future bioinformatics studies will be for that type of bioactivity, especially for QSAR model development. Therefore, for rarely studied types of bioactivity, to

initiate bioinformatics-aided studies, we need to conduct large-scale molecular docking screening as well as wet chemistry studies to provide foundational knowledge on structure-activity relationships for QSAR model development.

Besides the general evaluation model (PeptideRanker), some QSAR classification web servers for specific types of activity (e.g., PreAIP, AntiAngioPred, and PlifePred) are employed in FBP studies. In addition, some QSAR models built by biochemistry researchers have outperformed in dataset collection and curation, characterization methods, and model analysis, compared to those models created bioinformatics background laboratories perhaps due to biochemistry researchers' better understanding of biological properties [10,28,81,93,96,107,189]. For example, Kalyan et al. used data mining to retrieve 1687 peptides from two databases, which was significantly larger than most self-built models. In addition, non-linear regression methods such as regression decision tree and back-propagation neural network (BPNN) were employed by some biochemistry researchers [81,189]. However, most of these models suffer from one or several issues mentioned in Section 3, such as small datasets and poor peptide descriptors [10,28,81,93,96,107,189].

Molecular docking tools used in integrated studies are diverse, and some studies even adopted two docking tools for better refinement and docking conformation optimization [8,37,50,65,134,187,195]. It is difficult to differentiate the performance of different molecular docking protocols in specific protein receptor cases, but generally, for the same docking task, performance is proportional to time consumption [186,197]. Compared to the use of two docking tools for refinement, consensus docking would be a better alternative by combining different molecular docking protocols with different sampling algorithms or scoring functions [139]. The use of molecular docking web servers is popular among FBP studies, and brief introductions of these web servers are given in Table 5.

A novel strategy (ensemble docking) was introduced by Aguilar-Toalá et al. for FBP screening, where molecular dynamics simulations of 300 ns were conducted to generate protein receptor conformations. Three representative conformations were selected for further molecular docking tasks, and the average score of the ligand with the three representative receptor conformations was used for virtual screening [18]. Ensemble docking is also an alternative approach to increase molecular docking accuracy. It can help locate high-activity peptides by changing the protein receptor conformation from a single source conformation (experimental or modeled) to a number of conformations and thus enhance docking performance [198,199].

7. Conclusion and directions for future research

This review provided an overview of using bioinformatics to accelerate FBP screening and interaction mechanism exploration. Database-driven virtual screening with proteolysis simulation has been widely used to identify FBPs, to compare them with reported FBPs, and to differentiate unknown peptides for further screening. QSAR, molecular docking, and molecular dynamics simulation are now commonly used for virtual screening of unknown peptides and elucidating interaction mechanisms. This review discussed in detail the limitations of database-driven studies and bioactivity potency evaluation system; the role of the dataset, peptide representation, and model development in QSAR studies; as well as sampling algorithms, scoring functions, force fields, and free energy estimation methods. Although a lot of progress has been made using bioinformatics in FBP studies, there are still many challenges and specific technical expertise can be required to obtain accurate results.

Bioinformatics has the potential to guide the production of value-added products from agricultural byproducts, to promote sustainable agriculture, to evaluate the potency of food bioactivity for precise nutrition, and to rationally design peptides derived from food for nutraceutical, cosmeceutical, and pharmaceutical industries. The collaboration between biochemistry and bioinformatics researchers is essential to achieve this potential.

To advance the application of bioinformatics in FBP studies, we recommend the following areas.

1). Government-led non-profit databases for bioactive peptides should be built, similar to protein databases (e.g., NCBI, RCSB PDB, etc.). Such large and high quality databases are critical to QSAR model development and efficient information retrieval of bioactive peptides.
2). Dataset cleaning is a challenging task when collecting FBPs data from different literature sources, where some bioactivity results are affected by manual operations, experimental conditions, etc., and should be removed for high quality dataset construction. An example is our lab's web server for ACE inhibitory peptide prediction, where a confident learning theory based tool, CleanLab, is employed to clean real-world datasets and generate a high quality ACE inhibitory peptide dataset [200].
3). Limited negative sample datasets are available for virtual screening of FBPs, since most reported FBPs were identified with high activity. A solution is to manually create negative samples. For example, DUD-E server (http://dude.docking.org/) can provide challenging decoys for checking molecular docking or QSAR model performance, which have been used in umami dipeptide screening [17].
4). Protein language models are expected to gradually become the mainstream peptide representation for bioactivity classification model development. Though it exhibited better performance in peptide information representation, it suffers from high feature dimension problem and might undermine its application in small datasets. For QSAR model development, besides the employment of advanced feature selection methods and modeling methods, an stacking framework is another alternative for modeling strategy, where different modeling methods are combined together for decision-making.
5). When the 3D structure of a targeted protein is not available from X-ray crystallography or NMR, *de novo* structure prediction by Alpha-Fold2 can be a great alternative to create the 3D structure [201]. In addition, homology modeling and threading modeling methods can be considered. In the last five years, Electron Microscopy Data Bank (EMDB, https://www.ebi.ac.uk/emdb/search/) has rapidly added new entries and cryo-electron microscopy has rapidly become a major source of experimental structures, overcoming many difficulties with traditional X-ray crystallography and NMR methods [12,202,203].
6). Molecular docking with advanced docking strategies, such as consensus docking and ensemble docking, are expected to be applied in FBPs discovery and improve the docking accuracy. In addition, large-scale virtual screening based on open-source programs have not yet been widely used in FBPs studies. With the rapidly developing cloud computing platforms such as Amazon Web Services (AWS), a user-friendly and easy-operated servers is in great demand and will significantly promote the popularization of molecular docking and molecular dynamics simulation in FBPs studies.
7). Graphic processing unit (GPU) parallel acceleration and high-performance computation clusters have been introduced into molecular docking and molecular dynamics simulation. They can decrease computer time by more than an order of

Table 7
Selected virtual screening strategies integrating quantitative structure-activity relationship (QSAR) modeling, molecular docking, and molecular dynamics simulation.

Protein source	Bioactivity	QSAR model	Molecular docking	Molecular dynamics simulation	Additional comments	Reference
–	DPP-IV inhibitory activity	Self-built regression model	AutoDock Vina	GROMACS for 100 ns of simulation Forcefiled: AMBER14SB and General AMBER	PaDEL descriptors for peptide representation, GA for feature selection, and MLR for regression model	[21]
Sorghum protein	DPP-IV inhibitory activity	PeptideRanker	HPEPDOCK and FlexPepDock	–	PeptideRanker for rough virtual screening, and HPEPDOCK and FlexPepDock for refinement	[65]
Bean protein	DPP-IV inhibitory activity	PeptideRanker	Pepsite2	–	PeptideRanker for rough virtual screening and PepSite2 for refinement and selection for *in vitro* assay peptide synthesis	[50]
Draft beer	DPP-IV inhibitory activity and ACE inhibitory activity	PeptideRanker	Sybyl software		PeptideRanker for rough screening and selection for *in vitro* assay peptide synthesis; Sybylfor interaction mechanism; absorption and toxicity evaluation	[191]
Cheese	Antidiabetic activity	PeptideRanker	Pepsite2 and HPEPDOCK	–	PeptideRanker for rough virtual screening and PepSite2 for refinement and selection for *in vitro* assay peptide synthesis; HPEPDOCK for interaction mechanism	[38]
Rubing cheese	ACE, a-glucosidase, and Keap1−Nrf2 interaction inhibitory activity	PeptideRanker	Pepsite2 and AutoDock Vina		PeptideRanker for rough virtual screening and PepSite2 for refinement and selection for *in vitro* assay peptide synthesis; AutoDock Vina for interaction mechanism	[192]
Chia Seed	ACE inhibitory activity, anti-inflammatory activity, and plasma stability	PeptideRanker, AHTpin, PreAIP, AntiAngioPred, and PlifePred	AutoDock Vina	PMEMD for 300 ns of simulation Force fields: amberff14 and GAFF	PMEMD for conformation generation of protein receptors and AutoDock Vina for molecular scoring	[18]
Egg yolk protein	ACE inhibitory activity	PeptideRanker and AHTpin	Autodock CrankPep	–	PeptideRanker for bioactivity possibility prediction and AHTpin for virtual screening of ACE inhibitory FBPs	[190]
β-Casein	ACE inhibitory activity	Self-built classification model	AutoDock Vina	–	Eight sequence-based and structure-based features for decision tree model	[189]
Camel milk	ACE and renin inhibitory activity	PeptideRanker	PepSite2 and Glide	–	PeptideRanker for rough screening and Pepsite2 for refinement; Glide for interaction mechanism	[187]
–	ACE inhibitory activity	Self-built regression model	AutoDock4	–	Dragon descriptors, AAindex, and 5-z scale for peptide representation; KNN, RFR, MLP, and SVMR for regression model to further select for peptide synthesis and *in vitro* assays	[28]
	ACE inhibitory activity	Self-built regression model	AutoDock4	–	3D QSAR models (PLSR) based on ligand template-based molecular alignment and MIF calculation; AutoDock4 for interaction mechanism	[10]
Honey protein	ACE inhibitory activity	AHTpin	PatchDock	–	AHTpin for rough screening, PatchDock for refinement, and FireDock for interaction mechanism	[193]
Camel milk	ACE inhibitory activity	PeptideRanker and AHTpin	PepSite2 and Glide	–	PeptideRanker and AHTpin for rough screening, Pepsite2 for refinement, and Glide for interaction mechanism; toxicity evaluation	[134]
Salmo salar collagen	ACE inhibitory activity	PeptideRanker	SwissDock and CDOCKER	–	PeptideRanker for rough screening, SwissDock for refinement, and CDOCKER for interaction mechanism; physicochemical property and toxicity evaluation	[37]
Silkworm cocoon	ACE inhibitory activity	Self-built regression model	Surflex-Dock	–	3D QSAR models (PLSR) based on ligand template-based molecular alignment; MIF calculation by CoMFA and CoMSIA; toxicity and digestive stability evaluation to further select peptides for synthesis and *in vitro* assays	[96]
Qula casein	ACE inhibitory activity	Self-built regression models	Discovery Studio	–	5z-scale was used to build PLSR models for penta/hexa/hepta/octapeptide; QSAR model for rough screening of peptides identified from LC-MS and Autotock Vina for refinement and further selection for peptide synthesis and *in vitro* assays	[107]
Bovine blood	ACE inhibitory activity	Self-built regression model	CDOCKER	–	10 descriptors were generated by Discovery Studio and modeled by BPNN for further selection for peptide synthesis and *in vitro* assays; 24 pentapeptides in dataset	[81]
Egg	Aminopeptidase N inhibitory peptides	PeptideRanker	CDOCKER	–	PeptideRanker, physicochemical property evaluation, and AMDET evaluation combined for virtual screening; CDOCKER for interaction mechanism	[194]
Rice	Immunomodulatory activity	NetMHCpan 4.0	CABS-dock	–	NetMHCpan 4.0 for virtual screening by binding affinity prediction and CABS-dock for interaction mechanism	[101]
Donkey collagen	Tyrosinase inhibitory activity	PeptideRanker and prediction for water solubility and toxicity	CDOCKER	GROMACS for 60 ns of simulation; Force field: CHARMM36	PeptideRanker, physicochemical properties, and CDOCKER combined for virtual screening	[195]

(*continued on next page*)

Table 7 (*continued*)

Protein source	Bioactivity	QSAR model	Molecular docking	Molecular dynamics simulation	Additional comments	Reference
Camel milk	Cholesterol esterase inhibitory activity	PeptideRanker	PepSite2 and Glide	–	PeptideRanker and Pepsite2 for rough screening and Glide score and MM-GBSA for refinement; Glide for interaction mechanism	[196]

Abbreviation: BPNN: back-propagation neural network; CoMFA: comparative molecular field analysis; CoMSIA: comparative molecular similarity indices analysis; GA: genetic algorithm; KNN: k-nearest neighbor; MIF: molecular interaction field; ML: machine learning; MLP: multilayer perceptron; MM-GBSA: molecular mechanics-generalized born surface area; MLR: multiple linear regression; PLSR: partial least squares regression; RFR: random forest regression; SVMR: support vector machine regression.

magnitude and enable even expensive molecular dynamics simulations to be performed with commodity computer hardware [204–209].

8). Current virtual screening studies that use molecular docking and molecular dynamics simulation focus on the interaction between different ligands and the same receptor. It would be meaningful to conduct molecular docking with identified FBPs that have specific bioactivities and proteins with other desired bioactivities for the screening of multi-bioactivity FBPs. This strategy has been used in QSAR studies (e.g., SwissTargetPrediction) where various 2D and 3D molecular fingerprints were proposed to encode molecules with unknown bioactivity in digital formats for molecular similarity calculations (Manhattan distance-based similarity). The most matchable molecules among the 376,342 molecules in the dataset were used to predict the potential activity corresponding to 3068 macromolecular targets [210].

9). Biochemistry researchers are encouraged to keep up with the latest progress in bioinformatics fields and its corresponding outcomes, and bioinformatic researchers could make their findings more accessible to biochemistry researchers, such as developing and providing web servers.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This is contribution No. 23-155-J from the Kansas Agricultural Experimental Station. This work was supported in part by the Agriculture and Food Research Initiative Competitive Grant no. 2020-68008-31408 and no. 2021-67021-34495 from the USDA National Institute of Food and Agriculture, and a seed grant from the Global Food Systems initiative of Kansas State University.

References

[1] M. Tu, S. Cheng, W. Lu, M. Du, Advancement and prospects of bioinformatics analysis for studying bioactive peptides from food-derived protein: sequence, structure, and functions, TrAC, Trends Anal. Chem. 105 (2018) 7–17. https://doi.org/10.1016/j.trac.2018.04.005.

[2] C.C. Udenigwe, Bioinformatics approaches, prospects and challenges of food bioactive peptide research, Trends Food Sci. Technol. 36 (2014) 137–143. https://doi.org/10.1016/j.tifs.2014.02.004.

[3] R.J. FitzGerald, M. Cermeño, M. Khalesi, T. Kleekayai, M. Amigo-Benavent, Application of in silico approaches for the generation of milk protein-derived bioactive peptides, J. Funct.Foods 64 (2020), 103636. https://doi.org/10.1016/j.jff.2019.103636.

[4] A. Iwaniak, M. Darewicz, D. Mogut, P. Minkiewicz, Elucidation of the role of in silico methodologies in approaches to studying bioactive peptides derived from foods, J. Funct.Foods 61 (2019), 103486. https://doi.org/10.1016/j.jff.2019.103486.

[5] Z. Du, D. Wang, Y. Li, Comprehensive evaluation and comparison of machine learning methods in QSAR modeling of antioxidant tripeptides, ACS Omega (2022). https://doi.org/10.1021/acsomega.2c03062. acsomega.2c03062.

[6] Y. Feng, Z. Wang, J. Chen, H. Li, Y. Wang, D.-F. Ren, J. Lu, Separation, identification, and molecular docking of tyrosinase inhibitory peptides from the hydrolysates of defatted walnut (Juglans regia L.) meal, Food Chem. 353 (2021), 129471. https://doi.org/10.1016/j.foodchem.2021.129471.

[7] L. Li, J. Liu, S. Nie, L. Ding, L. Wang, J. Liu, W. Liu, T. Zhang, Direct inhibition of Keap1–Nrf2 interaction by egg-derived peptides DKK and DDW revealed by molecular docking and fluorescence polarization, RSC Adv. 7 (2017) 34963–34971. https://doi.org/10.1039/C7RA04352J.

[8] A.E. Nardo, M.C. Añón, A.V. Quiroga, Identification of renin inhibitors peptides from amaranth proteins by docking protocols, J. Funct.Foods 64 (2020), 103683. https://doi.org/10.1016/j.jff.2019.103683.

[9] V.R. Vukic, D.V. Vukic, S.D. Milanovic, M.D. Ilicic, K.G. Kanuric, M.S. Johnson, In silico identification of milk antihypertensive di- and tripeptides involved in angiotensin I–converting enzyme inhibitory activity, Nutr. Res. 46 (2017) 22–30. https://doi.org/10.1016/j.nutres.2017.07.009.

[10] F. Wang, B. Zhou, Investigation of angiotensin-I-converting enzyme (ACE) inhibitory tri-peptides: a combination of 3D-QSAR and molecular docking simulations, RSC Adv. 10 (2020) 35811–35819. https://doi.org/10.1039/D0RA05119E.

[11] Y. Gu, K. Majumder, J. Wu, QSAR-aided in silico approach in evaluation of food proteins as precursors of ACE inhibitory peptides, Food Res. Int. 44 (2011) 2465–2474. https://doi.org/10.1016/j.foodres.2011.01.051.

[12] E. Callaway, Revolutionary cryo-EM is taking over structural biology, Nature 578 (2020). https://doi.org/10.1038/d41586-020-00341-9, 201–201.

[13] D. Agyei, A. Tsopmo, C.C. Udenigwe, Bioinformatics and peptidomics approaches to the discovery and analysis of food-derived bioactive peptides, Anal. Bioanal. Chem. 410 (2018) 3463–3472. https://doi.org/10.1007/s00216-018-0974-1.

[14] W. Bo, L. Chen, D. Qin, S. Geng, J. Li, H. Mei, B. Li, G. Liang, Application of quantitative structure-activity relationship to food-derived peptides: methods, situations, challenges and prospects, Trends Food Sci. Technol. 114 (2021) 176–188. https://doi.org/10.1016/j.tifs.2021.05.031.

[15] X. Tao, Y. Huang, C. Wang, F. Chen, L. Yang, L. Ling, Z. Che, X. Chen, Recent developments in molecular docking technology applied in food science: a review, Int. J. Food Sci. Technol. 55 (2020) 33–45. https://doi.org/10.1111/ijfs.14325.

[16] K. Majumder, J. Wu, A new approach for identification of novel antihypertensive peptides from egg proteins by QSAR and bioinformatics, Food Res. Int. 43 (2010) 1371–1378. https://doi.org/10.1016/j.foodres.2010.04.027.

[17] Y. Xiong, X. Gao, D. Pan, T. Zhang, L. Qi, N. Wang, Y. Zhao, Y. Dang, A strategy for screening novel umami dipeptides based on common feature pharmacophore and molecular docking, Biomaterials (2022), 121697. https://doi.org/10.1016/j.biomaterials.2022.121697.

[18] J.E. Aguilar-Toalá, A. Vidal-Limon, A.M. Liceaga, Multifunctional analysis of chia seed (salvia hispanica L.) bioactive peptides using peptidomics and molecular dynamics simulations approaches, Int. J. Mol. Sci. 23 (2022) 7288. https://doi.org/10.3390/ijms23137288.

[19] J.G. Arámburo-Gálvez, A.A. Arvizu-Flores, F.I. Cárdenas-Torres, F. Cabrera-Chávez, G.I. Ramírez-Torres, L.K. Flores-Mendoza, P.E. Gastelum-Acosta, O.G. Figueroa-Salcido, N. Ontiveros, Prediction of ACE-I inhibitory peptides derived from chickpea (cicer arietinum L.): in silico assessments using simulated enzymatic hydrolysis, molecular docking and ADMET evaluation, Foods 11 (2022) 1576. https://doi.org/10.3390/foods11111576.

[20] Z. Dai, L. Wang, Y. Chen, H. Wang, L. Bai, Z. Yuan, A pipeline for improved QSAR analysis of peptides: physiochemical property parameter selection via BMSF, near-neighbor sample selection via semivariogram, and weighted SVR regression and prediction, Amino Acids 46 (2014) 1105–1119. https://doi.org/10.1007/s00726-014-1667-5.

[21] X. Li, F. Pan, Z. Yang, F. Gao, J. Li, F. Zhang, T. Wang, Construction of QSAR model based on cysteine-containing dipeptides and screening of natural tyrosinase inhibitors, J. Food Biochem. (2022), e14338. https://doi.org/10.1111/jfbc.14338.

[22] A. Vidal-Limon, J.E. Aguilar-Toalá, A.M. Liceaga, Integration of molecular docking analysis and molecular dynamics simulations for studying food proteins and bioactive peptides, J. Agric. Food Chem. 70 (2022) 934–943.

https://doi.org/10.1021/acs.jafc.1c06110.
[23] W. Zhao, L. Su, S. Huo, Z. Yu, J. Li, J. Liu, Virtual screening, molecular docking and identification of umami peptides derived from Oncorhynchus mykiss, Food Sci. Hum. Wellness 12 (2023) 89–93. https://doi.org/10.1016/j.fshw.2022.07.026.
[24] A. Iwaniak, P. Minkiewicz, M. Pliszka, D. Mogut, M. Darewicz, Characteristics of biopeptides released in silico from collagens using quantitative parameters, Foods 9 (2020) 965. https://doi.org/10.3390/foods9070965.
[25] Minkiewicz, Iwaniak, Darewicz, BIOPEP-UWM database of bioactive peptides: current opportunities, IJMS 20 (2019) 5978. https://doi.org/10.3390/ijms20235978.
[26] B. Deng, H. Long, T. Tang, X. Ni, J. Chen, G. Yang, F. Zhang, R. Cao, D. Cao, M. Zeng, L. Yi, Quantitative structure-activity relationship study of antioxidant tripeptides based on model population analysis, IJMS 20 (2019) 995. https://doi.org/10.3390/ijms20040995.
[27] S. Uno, D. Kodama, H. Yukawa, H. Shidara, M. Akamatsu, Quantitative analysis of the relationship between structure and antioxidant activity of tripeptides, J. Pept. Sci. 26 (2020). https://doi.org/10.1002/psc.3238.
[28] Y.-T. Wang, D.P. Russo, C. Liu, Q. Zhou, H. Zhu, Y.-H. Zhang, Predictive modeling of angiotensin I-converting enzyme inhibitory peptides using various machine learning approaches, J. Agric. Food Chem. 68 (2020) 12132–12140. https://doi.org/10.1021/acs.jafc.0c04624.
[29] L. Amigo, D. Martínez-Maqueda, B. Hernández-Ledesma, In silico and in vitro analysis of multifunctionality of animal food-derived peptides, Foods 9 (2020) 991. https://doi.org/10.3390/foods9080991.
[30] H.L.R. Gomez, J.P. Peralta, L.A. Tejano, Y.-W. Chang, In silico and in vitro assessment of Portuguese oyster (Crassostrea angulata) proteins as precursor of bioactive peptides, Int. J. Mol. Sci. 20 (2019) 5191. https://doi.org/10.3390/ijms20205191.
[31] D. Ji, C.C. Udenigwe, D. Agyei, Antioxidant peptides encrypted in flaxseed proteome: an in silico assessment, Food Sci. Hum. Wellness 8 (2019) 306–314. https://doi.org/10.1016/j.fshw.2019.08.002.
[32] D. Ji, M. Xu, C.C. Udenigwe, D. Agyei, Physicochemical characterisation, molecular docking, and drug-likeness evaluation of hypotensive peptides encrypted in flaxseed proteome, Curr. Res. Food Sci. 3 (2020) 41–50. https://doi.org/10.1016/j.crfs.2020.03.001.
[33] C. Kartal, B. Kaplan Türköz, S. Otles, Prediction, identification and evaluation of bioactive peptides from tomato seed proteins using in silico approach, Food Measure 14 (2020) 1865–1883. https://doi.org/10.1007/s11694-020-00434-z.
[34] F.C.A. Panjaitan, H.L.R. Gomez, Y.-W. Chang, In silico analysis of bioactive peptides released from giant grouper (Epinephelus lanceolatus) roe proteins identified by proteomics approach, Molecules 23 (2018) 2910. https://doi.org/10.3390/molecules23112910.
[35] K. Pooja, S. Rani, B. Prakash, In silico approaches towards the exploration of rice bran proteins-derived angiotensin-I-converting enzyme inhibitory peptides, Int. J. Food Prop. 20 (2017) 2178–2191. https://doi.org/10.1080/10942912.2017.1368552.
[36] M. Tu, H. Liu, R. Zhang, H. Chen, F. Fan, P. Shi, X. Xu, W. Lu, M. Du, Bioactive hydrolysates from casein: generation, identification, and in silico toxicity and allergenicity prediction of peptides, J. Sci. Food Agric. 98 (2018) 3416–3426. https://doi.org/10.1002/jsfa.8854.
[37] Z. Yu, Y. Chen, W. Zhao, J. Li, J. Liu, F. Chen, Identification and molecular docking study of novel angiotensin-converting enzyme inhibitory peptides from Salmo salar using in silico methods, J. Sci. Food Agric. 98 (2018) 3907–3914. https://doi.org/10.1002/jsfa.8908.
[38] S. Martini, L. Solieri, A. Cattivelli, V. Pizzamiglio, D. Tagliazucchi, An integrated peptidomics and in silico approach to identify novel anti-diabetic peptides in parmigiano-reggiano cheese, Biology 10 (2021) 563. https://doi.org/10.3390/biology10060563.
[39] D. Iram, M.S. Sansi, S. Zanab, S. Vij, Ashutosh, S. Meena, In silico identification of antidiabetic and hypotensive potential bioactive peptides from the sheep milk proteins—a molecular docking study, J. Food Biochem. (2022). https://doi.org/10.1111/jfbc.14137.
[40] K. Parastouei, Estimation of bioactive peptide content of milk from different species using an in silico method, Amino Acids (2022). https://doi.org/10.1007/s00726-022-03152-6.
[41] E.R. Coscueta, P. Batista, J.E.G. Gomes, R. da Silva, M.M. Pintado, Screening of novel bioactive peptides from goat casein: in silico to in vitro validation, Int. J. Mol. Sci. 23 (2022) 2439. https://doi.org/10.3390/ijms23052439.
[42] F. Liang, Y. Shi, J. Shi, T. Zhang, R. Zhang, A novel Angiotensin-I-converting enzyme (ACE) inhibitory peptide IAF (Ile-Ala-Phe) from pumpkin seed proteins: in silico screening, inhibitory activity, and molecular mechanisms, Eur. Food Res. Technol. 247 (2021) 2227–2237. https://doi.org/10.1007/s00217-021-03783-1.
[43] N.A. Pearman, E. Ronander, A.M. Smith, G.A. Morris, The identification and characterisation of novel bioactive peptides derived from porcine liver, Curr. Res. Food Sci. 3 (2020) 314–321. https://doi.org/10.1016/j.crfs.2020.11.002.
[44] M. Barati, F. Javanmardi, M. Jabbari, A. Mokari-Yamchi, F. Farahmand, I. Eş, H. Farhadnejad, S.H. Davoodi, A. Mousavi Khaneghah, An in silico model to predict and estimate digestion-resistant and bioactive peptide content of dairy products: a primarily study of a time-saving and affordable method for practical research purposes, LWT 130 (2020), 109616. https://doi.org/10.1016/j.lwt.2020.109616.
[45] J. Chen, B. Ryu, Y. Zhang, P. Liang, C. Li, C. Zhou, P. Yang, P. Hong, Z.-J. Qian, Comparison of an angiotensin-I-converting enzyme inhibitory peptide from tilapia (Oreochromis niloticus) with captopril: inhibition kinetics, in vivo effect, simulated gastrointestinal digestion and a molecular docking study, J. Sci. Food Agric. 100 (2020) 315–324. https://doi.org/10.1002/jsfa.10041.
[46] K. Lin, L. Zhang, X. Han, L. Xin, Z. Meng, P. Gong, D. Cheng, Yak milk casein as potential precursor of angiotensin I-converting enzyme inhibitory peptides based on in silico proteolysis, Food Chem. 254 (2018) 340–347. https://doi.org/10.1016/j.foodchem.2018.02.051.
[47] Z. Agirbasli, L. Cavas, In silico evaluation of bioactive peptides from the green algae Caulerpa, J. Appl. Phycol. 29 (2017) 1635–1646. https://doi.org/10.1007/s10811-016-1045-7.
[48] L. Devita, H.N. Lioe, M. Nurilmala, M.T. Suhartono, The bioactivity prediction of peptides from tuna skin collagen using integrated method combining in vitro and in silico, Foods 10 (2021) 2739. https://doi.org/10.3390/foods10112739.
[49] T. Sayd, C. Dufour, C. Chambon, C. Buffière, D. Remond, V. Santé-Lhoutellier, Combined in vivo and in silico approaches for predicting the release of bioactive peptides from meat digestion, Food Chem. 249 (2018) 111–118. https://doi.org/10.1016/j.foodchem.2018.01.013.
[50] S. Martini, A. Cattivelli, A. Conte, D. Tagliazucchi, Application of a combined peptidomics and in silico approach for the identification of novel dipeptidyl peptidase-IV-inhibitory peptides in in vitro digested pinto bean protein extract, Curr. Issues Mol. Biol. 44 (2022) 139–151. https://doi.org/10.3390/cimb44010011.
[51] J. Bechaux, V. Ferraro, T. Sayd, C. Chambon, J.F. Le Page, Y. Drillet, P. Gatellier, V. Santé-Lhoutellier, Workflow towards the generation of bioactive hydrolysates from porcine products by combining in silico and in vitro approaches, Food Res. Int. 132 (2020), 109123. https://doi.org/10.1016/j.foodres.2020.109123.
[52] S. Garg, V. Apostolopoulos, K. Nurgali, V.K. Mishra, Evaluation of in silico approach for prediction of presence of opioid peptides in wheat, J. Funct.Foods 41 (2018) 34–40. https://doi.org/10.1016/j.jff.2017.12.022.
[53] Y. Fu, J.F. Young, M.M. Løkke, R. Lametsch, R.E. Aluko, M. Therkildsen, Revalorisation of bovine collagen as a potential precursor of angiotensin I-converting enzyme (ACE) inhibitory peptides based on in silico and in vitro protein digestions, J. Funct.Foods 24 (2016) 196–206. https://doi.org/10.1021/acs.jafc.0c04624.
[54] F. Luo, Y. Fu, L. Ma, H. Dai, H. Wang, H. Chen, H. Zhu, Y. Yu, Y. Hou, Y. Zhang, Exploration of dipeptidyl peptidase-IV (DPP-IV) inhibitory peptides from silkworm pupae (*Bombyx mori*) proteins based on *in silico* and *in vitro* assessments, J. Agric. Food Chem. 70 (2022) 3862–3871. https://doi.org/10.1021/acs.jafc.1c08225.
[55] Z. Du, Y. Li, Computer-aided approaches for screening antioxidative dipeptides and application to sorghum proteins, ACS Food Sci. Technol. (2022). https://doi.org/10.1021/acsfoodscitech.2c00286.
[56] R. Baskaran, S.S. Chauhan, R. Parthasarathi, N.S. Mogili, In silico investigation and assessment of plausible novel tyrosinase inhibitory peptides from sesame seeds, LWT 147 (2021), 111619. https://doi.org/10.1016/j.lwt.2021.111619.
[57] Z. Du, Y. Li, Review and perspective on bioactive peptides: a roadmap for research, development, and future opportunities, Journal of Agriculture and Food Research 9 (2022), 100353. https://doi.org/10.1016/j.jafr.2022.100353.
[58] R.T. Boachie, F.L. Okoro, K. Imai, L. Sun, S.O. Elom, J.O. Nwankwo, C.E.C.C. Ejike, C.C. Udenigwe, Enzymatic release of dipeptidyl peptidase-4 inhibitors (gliptins) from pigeon pea (*Cajanus cajan*) nutrient reservoir proteins: in silico and in vitro assessments, J. Food Biochem. 43 (2019). https://doi.org/10.1111/jfbc.13071.
[59] Z. Zhu, Y. Chen, N. Jia, W. Zhang, H. Hou, C. Xue, Y. Wang, Identification of three novel antioxidative peptides from Auxenochlorella pyrenoidosa protein hydrolysates based on a peptidomics strategy, Food Chem. 375 (2022), 131849. https://doi.org/10.1016/j.foodchem.2021.131849.
[60] A. Chatterjee, S.K. Kanawjia, Y. Khetra, P. Saini, Discordance between in silico & in vitro analyses of ACE inhibitory & antioxidative peptides from mixed milk tryptic whey protein hydrolysate, J. Food Sci. Technol. 52 (2015) 5621–5630. https://doi.org/10.1007/s13197-014-1669-z.
[61] A.B. Nongonierma, S. Paolella, P. Mudgil, S. Maqsood, R.J. FitzGerald, Identification of novel dipeptidyl peptidase IV (DPP-IV) inhibitory peptides in camel milk protein hydrolysates, Food Chem. 244 (2018) 340–348. https://doi.org/10.1016/j.foodchem.2017.10.033.
[62] Y. Shen, S. Hong, G. Singh, K. Koppel, Y. Li, Improving functional properties of pea protein through “green” modifications using enzymes and polysaccharides, Food Chem. 385 (2022), 132687. https://doi.org/10.1016/j.foodchem.2022.132687.
[63] Y. Shen, Y. Li, Acylation modification and/or guar gum conjugation enhanced functional properties of pea protein isolate, Food Hydrocolloids 117 (2021), 106686.
[64] A.B. Nongonierma, R.J. FitzGerald, Enhancing bioactive peptide release and identification using targeted enzymatic hydrolysis of milk proteins, Anal. Bioanal. Chem. 410 (2018) 3407–3423. https://doi.org/10.1007/s00216-017-0793-9.
[65] A.G. Garzón, F.F. Veras, A. Brandelli, S.R. Drago, Purification, identification and in silico studies of antioxidant, antidiabetogenic and antibacterial peptides obtained from sorghum spent grain hydrolysate, LWT 153 (2022), 112414. https://doi.org/10.1016/j.lwt.2021.112414.
[66] M. Zhang, L. Zhu, G. Wu, T. Liu, X. Qi, H. Zhang, Rapid screening of novel

dipeptidyl peptidase-4 inhibitory peptides from pea (*Pisum sativum L.*) protein using peptidomics and molecular docking, J. Agric. Food Chem. 70 (2022) 10221–10228. https://doi.org/10.1021/acs.jafc.2c03949.
[67] P. Zhou, Q. Liu, T. Wu, Q. Miao, S. Shang, H. Wang, Z. Chen, S. Wang, H. Wang, Systematic comparison and comprehensive evaluation of 80 amino acid descriptors in peptide QSAR modeling, J. Chem. Inf. Model. 61 (2021) 1718–1731. https://doi.org/10.1021/acs.jcim.0c01370.
[68] X.-Y. Meng, H.-X. Zhang, M. Mezei, M. Cui, Molecular docking: a powerful approach for structure-based drug discovery, CAD 7 (2011) 146–157. https://doi.org/10.2174/157340911795677602.
[69] N. Chen, J. Chen, B. Yao, Z. Li, QSAR study on antioxidant tripeptides and the antioxidant activity of the designed tripeptides in free radical systems, Molecules 23 (2018) 1407. https://doi.org/10.3390/molecules23061407.
[70] L. Zheng, Y. Zhao, H. Dong, G. Su, M. Zhao, Structure–activity relationship of antioxidant dipeptides: dominant role of Tyr, Trp, Cys and Met residues, J. Funct.Foods 21 (2016) 485–496. https://doi.org/10.1016/j.jff.2015.12.003.
[71] X. Guan, J. Liu, QSAR study of angiotensin I-converting enzyme inhibitory peptides using SVHEHS descriptor and OSC-SVM, Int. J. Pept. Res. Therapeut. 25 (2019) 247–256. https://doi.org/10.1007/s10989-017-9661-x.
[72] B. Deng, X. Ni, Z. Zhai, T. Tang, C. Tan, Y. Yan, J. Deng, Y. Yin, New quantitative structure–activity relationship model for angiotensin-converting enzyme inhibitory dipeptides based on integrated descriptors, J. Agric. Food Chem. 65 (2017) 9774–9781. https://doi.org/10.1021/acs.jafc.7b03367.
[73] A.B. Nongonierma, R.J. FitzGerald, Structure activity relationship modelling of milk protein-derived peptides with dipeptidyl peptidase IV (DPP-IV) inhibitory activity, Peptides 79 (2016) 1–7. https://doi.org/10.1016/j.peptides.2016.03.005.
[74] Y. Qian, Y. Liang, W. Liu, G. Liang, Comprehensive comparison of twenty structural characterization scales applied as QSAM of antimicrobial dodecapeptides derived from Bac2A against P. aeruginosa, J. Mol. Graph. Model. 71 (2017) 88–95. https://doi.org/10.1016/j.jmgm.2016.11.003.
[75] M. Mahmoodi-Reihani, F. Abbasitabar, V. Zare-Shahabadi, In silico rational design and virtual screening of bioactive peptides based on QSAR modeling, ACS Omega 5 (2020) 5951–5958. https://doi.org/10.1021/acsomega.9b04302.
[76] Chung Xu, Quantitative structure–activity relationship study of bitter di-, tri- and tetrapeptides using integrated descriptors, Molecules 24 (2019) 2846. https://doi.org/10.3390/molecules24152846.
[77] C. Qi, G. Lin, R. Zhang, W. Wu, Studies on the bioactivities of ACE-inhibitory peptides with phenylalanine C-terminus using 3D-QSAR, molecular docking and in vitro evaluation, Mol Inform 36 (2017). https://doi.org/10.1002/minf.201600157.
[78] S. Wu, W. Qi, R. Su, T. Li, D. Lu, Z. He, CoMFA and CoMSIA analysis of ACE-inhibitory, antimicrobial and bitter-tasting peptides, Eur. J. Med. Chem. 84 (2014) 100–106. https://doi.org/10.1016/j.ejmech.2014.07.015.
[79] H.M. Patel, M.N. Noolvi, P. Sharma, V. Jaiswal, S. Bansal, S. Lohan, S.S. Kumar, V. Abbot, S. Dhiman, V. Bhardwaj, Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery, Med. Chem. Res. 23 (2014) 4991–5007. https://doi.org/10.1007/s00044-014-1072-3.
[80] T.H. Olsen, B. Yesiltas, F.I. Marin, M. Pertseva, P.J. García-Moreno, S. Gregersen, M.T. Overgaard, C. Jacobsen, O. Lund, E.B. Hansen, P. Marcatili, AnOxPePred: using deep learning for the prediction of antioxidative properties of peptides, Sci. Rep. 10 (2020), 21471. https://doi.org/10.1038/s41598-020-78319-w.
[81] T. Zhang, S. Nie, B. Liu, Y. Yu, Y. Zhang, J. Liu, Activity prediction and molecular mechanism of bovine blood derived angiotensin I-converting enzyme inhibitory peptides, PLoS One 10 (2015), e0119598. https://doi.org/10.1371/journal.pone.0119598.
[82] X. Zhao, M. Chen, B. Huang, H. Ji, M. Yuan, Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) studies on α1a-adrenergic receptor antagonists based on pharmacophore molecular alignment, Int. J. Mol. Sci. 12 (2011) 7022–7037. https://doi.org/10.3390/ijms12107022.
[83] R.R.S. Pissurlenkar, M.S. Shaikh, E.C. Coutinho, 3D-QSAR studies of Dipeptidyl peptidase IV inhibitors using a docking based alignment, J. Mol. Model. 13 (2007) 1047–1071. https://doi.org/10.1007/s00894-007-0227-2.
[84] P. Charoenkwan, C. Nantasenamat, M.M. Hasan, B. Manavalan, W. Shoombuatong, BERT4Bitter: a bidirectional encoder representations from transformers (BERT)-based model for improving the prediction of bitter peptides, Bioinformatics 37 (2021) 2556–2562. https://doi.org/10.1093/bioinformatics/btab133.
[85] N. Brandes, D. Ofer, Y. Peleg, N. Rappoport, M. Linial, ProteinBERT: a universal deep-learning model of protein sequence and function, Bioinformatics 38 (2022) 2102–2110. https://doi.org/10.1093/bioinformatics/btac020.
[86] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2019. http://arxiv.org/abs/1810.04805. (Accessed 15 November 2022). accessed.
[87] A. Elnaggar, M. Heinzinger, C. Dallago, G. Rehawi, Y. Wang, L. Jones, T. Gibbs, T. Feher, C. Angerer, M. Steinegger, D. Bhowmik, B. Rost, ProtTrans, Towards cracking the language of lifes code through self-supervised deep learning and high performance computing, IEEE Trans. Pattern Anal. Mach. Intell. (2021). https://doi.org/10.1109/TPAMI.2021.3095381, 1–1.
[88] Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. dos Santos Costa, Evolutionary-scale prediction of atomic-level protein structure with a language model, Science 379 (2023). https://doi.org/10.1126/science.ade2574.
[89] Z. Du, X. Ding, Y. Xu, Y. Li, UniDL4BioPep: a universal deep learning architecture for binary classification in peptide bioactivity, Briefings Bioinf. (2023) 1–10. https://doi.org/10.1093/bib/bbad135.
[90] R. Irwin, S. Dimitriadis, J. He, E.J. Bjerrum, Chemformer, A pre-trained transformer for computational chemistry, Mach. Learn.: Sci. Technol. 3 (2022), 015022. https://doi.org/10.1088/2632-2153/ac3ffb.
[91] M. Ciemny, M. Kurcinski, K. Kamel, A. Kolinski, N. Alam, O. Schueler-Furman, S. Kmiecik, Protein–peptide docking: opportunities and challenges, Drug Discov. Today 23 (2018) 1530–1537. https://doi.org/10.1016/j.drudis.2018.05.006.
[92] S. Forli, R. Huey, M.E. Pique, M.F. Sanner, D.S. Goodsell, A.J. Olson, Computational protein–ligand docking and virtual drug screening with the AutoDock suite, Nat. Protoc. 11 (2016) 905–919. https://doi.org/10.1038/nprot.2016.051.
[93] Y. Li, S. Zhang, Z. Bao, N. Sun, S. Lin, Exploring the activation mechanism of alcalase activity with pulsed electric field treatment: effects on enzyme activity, spatial conformation, molecular dynamics simulation and molecular docking parameters, Innovat. Food Sci. Emerg. Technol. 76 (2022), 102918. https://doi.org/10.1016/j.ifset.2022.102918.
[94] Z. Du, X. Zeng, X. Li, X. Ding, J. Cao, W. Jiang, Recent advances in imaging techniques for bruise detection in fruits and vegetables, Trends Food Sci. Technol. 99 (2020) 133–141. https://doi.org/10.1016/j.tifs.2020.02.024.
[95] Z. Du, W. Tian, M. Tilley, D. Wang, G. Zhang, Y. Li, Quantitative assessment of wheat quality using near-infrared spectroscopy: a comprehensive review, Compr. Rev. Food Sci. Food Saf. 21 (2022) 2956–3009. https://doi.org/10.1111/1541-4337.12958.
[96] H. Sun, Q. Chang, L. Liu, K. Chai, G. Lin, Q. Huo, Z. Zhao, Z. Zhao, High-throughput and rapid screening of novel ACE inhibitory peptides from sericin source and inhibition mechanism by using in silico and in vitro prescriptions, J. Agric. Food Chem. 65 (2017) 10020–10028. https://doi.org/10.1021/acs.jafc.7b04043.
[97] J. Chen, H.H. Cheong, S.W.I. Siu, xDeep-AcPEP: deep learning method for anticancer peptide activity prediction based on convolutional neural network and multitask learning, J. Chem. Inf. Model. 61 (2021) 3789–3803. https://doi.org/10.1021/acs.jcim.1c00181.
[98] L. Thi Phan, H. Woo Park, T. Pitti, T. Madhavan, Y.-J. Jeon, B. Manavalan, Mlacp 2.0: an updated machine learning tool for anticancer peptide prediction, Comput. Struct. Biotechnol. J. 20 (2022) 4473–4480. https://doi.org/10.1016/j.csbj.2022.07.043.
[99] C. Mooney, N.J. Haslam, G. Pollastri, D.C. Shields, Towards the improved discovery and design of functional peptides: common features of diverse classes permit generalized prediction of bioactivity, PLoS One 7 (2012), e45012. https://doi.org/10.1371/journal.pone.0045012.
[100] M.S. Sansi, D. Iram, S. Zanab, S. Vij, A.K. Puniya, A. Singh, Ashutosh, S. Meena, Antimicrobial bioactive peptides from goat Milk proteins: in silico prediction and analysis, J. Food Biochem. (2022). https://doi.org/10.1111/jfbc.14311.
[101] L. Wen, L. Huang, Y. Li, Y. Feng, Z. Zhang, Z. Xu, M.-L. Chen, Y. Cheng, New peptides with immunomodulatory activity identified from rice proteins through peptidomic and in silico analysis, Food Chem. 364 (2021), 130357. https://doi.org/10.1016/j.foodchem.2021.130357.
[102] F.H. Waghu, L. Gopi, R.S. Barai, P. Ramteke, B. Nizami, S. Idicula-Thomas, CAMP: collection of sequences and structures of antimicrobial peptides, Nucleic Acids Res. 42 (2014) D1154–D1158. https://doi.org/10.1093/nar/gkt1157.
[103] W. Liao, K.S. Bhullar, S. Chakrabarti, S.T. Davidge, J. Wu, Egg white-derived tripeptide IRW (Ile-Arg-Trp) is an activator of angiotensin converting enzyme 2, J. Agric. Food Chem. 66 (2018) 11330–11336. https://doi.org/10.1021/acs.jafc.8b03501.
[104] W. Liao, H. Fan, S.T. Davidge, J. Wu, Egg white–derived antihypertensive peptide IRW (Ile-Arg-Trp) reduces blood pressure in spontaneously hypertensive rats via the ACE2/ang (1-7)/mas receptor Axis, Mol. Nutr. Food Res. 63 (2019), 1900063. https://doi.org/10.1002/mnfr.201900063.
[105] K. Majumder, S. Chakrabarti, S.T. Davidge, J. Wu, Structure and activity study of egg protein ovotransferrin derived peptides (IRW and IQW) on endothelial inflammatory response and oxidative stress, J. Agric. Food Chem. 61 (2013) 2120–2129. https://doi.org/10.1021/jf3046076.
[106] J. Wu, R.E. Aluko, S. Nakai, Structural requirements of angiotensin I-converting enzyme inhibitory peptides: quantitative Structure–Activity relationship study of di- and tripeptides, J. Agric. Food Chem. 54 (2006) 732–738. https://doi.org/10.1021/jf051263l.
[107] K. Lin, L. Zhang, X. Han, D. Cheng, Novel angiotensin I-converting enzyme inhibitory peptides from protease hydrolysates of Qula casein: quantitative structure-activity relationship modeling and molecular docking study, J. Funct.Foods 32 (2017) 266–277. https://doi.org/10.1016/j.jff.2017.03.008.
[108] F. Tian, P. Zhou, Z. Li, T-scale as a novel vector of topological descriptors for amino acids and its application in QSARs of peptides, J. Mol. Struct. 830 (2007) 106–115. https://doi.org/10.1016/j.molstruc.2006.07.004.
[109] M. Karaś, Influence of physiological and chemical factors on the absorption of bioactive peptides, Int. J. Food Sci. Technol. 54 (2019) 1486–1496. https://doi.org/10.1111/ijfs.14054.
[110] C. Xiao, F. Toldrá, M. Zhao, F. Zhou, D. Luo, R. Jia, L. Mora, In vitro and in silico analysis of potential antioxidant peptides obtained from chicken hydrolysate produced using Alcalase, Food Res. Int. 157 (2022), 111253. https://doi.org/10.1016/j.foodres.2022.111253.

[111] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, Sci. Rep. 7 (2017), 42717. https://doi.org/10.1038/srep42717.
[112] G. Xiong, Z. Wu, J. Yi, L. Fu, Z. Yang, C. Hsieh, M. Yin, X. Zeng, C. Wu, A. Lu, X. Chen, T. Hou, D. Cao, ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties, Nucleic Acids Res. 49 (2021) W5–W14. https://doi.org/10.1093/nar/gkab255.
[113] L. Ferreira, R. dos Santos, G. Oliva, A. Andricopulo, Molecular docking and structure-based drug design strategies, Molecules 20 (2015) 13384–13421. https://doi.org/10.3390/molecules200713384.
[114] V. Salmaso, S. Moro, Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: an overview, Front. Pharmacol. 9 (2018). https://doi.org/10.3389/fphar.2018.00923.
[115] H. Berman, K. Henrick, H. Nakamura, Announcing the worldwide protein Data Bank, Nat. Struct. Mol. Biol. 10 (2003). https://doi.org/10.1038/nsb1203-980, 980–980.
[116] A. Majid, M. Lakshmikanth, N.K. Lokanath, C.G. Poornima Priyadarshini, Generation, characterization and molecular binding mechanism of novel dipeptidyl peptidase-4 inhibitory peptides from sorghum bicolor seed protein, Food Chem. 369 (2022), 130888. https://doi.org/10.1016/j.foodchem.2021.130888.
[117] C. Wen, J. Zhang, H. Zhang, Y. Duan, H. Ma, Plant protein-derived antioxidant peptides: isolation, identification, mechanism of action and application in food systems: a review, Trends Food Sci. Technol. 105 (2020) 308–322. https://doi.org/10.1016/j.tifs.2020.09.019.
[118] I.A. Guedes, F.S.S. Pereira, L.E. Dardenne, Empirical scoring functions for structure-based virtual screening: applications, critical aspects, and challenges, Front. Pharmacol. 9 (2018) 1089. https://doi.org/10.3389/fphar.2018.01089.
[119] X. Li, Y. Li, T. Cheng, Z. Liu, R. Wang, Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes, J. Comput. Chem. 31 (2010) 2109–2125. https://doi.org/10.1002/jcc.21498.
[120] L. Feng, M. Tu, M. Qiao, F. Fan, H. Chen, W. Song, M. Du, Thrombin inhibitory peptides derived from Mytilus edulis proteins: identification, molecular docking and in silico prediction of toxicity, Eur. Food Res. Technol. 244 (2018) 207–217. https://doi.org/10.1007/s00217-017-2946-7.
[121] T. Panyayai, C. Ngamphiw, S. Tongsima, W. Mhuantong, W. Limsripraphan, K. Choowongkomon, O. Sawatdichaikul, FeptideDB: a web application for new bioactive peptides from food protein, Heliyon 5 (2019), e02076. https://doi.org/10.1016/j.heliyon.2019.e02076.
[122] F. Tonolo, A. Folda, L. Cesaro, V. Scalcon, O. Marin, S. Ferro, A. Bindoli, M.P. Rigobello, Milk-derived bioactive peptides exhibit antioxidant activity through the Keap1-Nrf2 signaling pathway, J. Funct.Foods 64 (2020), 103696. https://doi.org/10.1016/j.jff.2019.103696.
[123] M. Tu, L. Feng, Z. Wang, M. Qiao, F. Shahidi, W. Lu, M. Du, Sequence analysis and molecular docking of antithrombotic peptides from casein hydrolysate by trypsin digestion, J. Funct.Foods 32 (2017) 313–323. https://doi.org/10.1016/j.jff.2017.03.015.
[124] E.H.B. Maia, L.C. Assis, T.A. de Oliveira, A.M. da Silva, A.G. Taranto, Structure-based virtual screening: from classical to artificial intelligence, Front. Chem. 8 (2020) 343. https://doi.org/10.3389/fchem.2020.00343.
[125] R.E. Amaro, J. Baudry, J. Chodera, Ö. Demir, J.A. McCammon, Y. Miao, J.C. Smith, Ensemble docking in drug discovery, Biophys. J. 114 (2018) 2271–2278. https://doi.org/10.1016/j.bpj.2018.02.038.
[126] S.-Y. Huang, Comprehensive assessment of flexible-ligand docking algorithms: current effectiveness and challenges, Briefings Bioinf. 19 (2018) 982–994. https://doi.org/10.1093/bib/bbx030.
[127] W.J. Allen, T.E. Balius, S. Mukherjee, S.R. Brozell, D.T. Moustakas, P.T. Lang, D.A. Case, I.D. Kuntz, R.C. Rizzo, Dock 6: impact of new features and current docking performance, J. Comput. Chem. 36 (2015) 1132–1156. https://doi.org/10.1002/jcc.23905.
[128] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, J. Med. Chem. 47 (2004) 1739–1749. https://doi.org/10.1021/jm0306430.
[129] J.K. Gagnon, S.M. Law, C.L. Brooks, Flexible CDOCKER: development and application of a pseudo-explicit structure-based docking method within CHARMM: adding receptor flexibility improves protein-ligand docking within CDOCKER, J. Comput. Chem. 37 (2016) 753–762. https://doi.org/10.1002/jcc.24259.
[130] I. Ugur, M. Schroft, A. Marion, M. Glaser, I. Antes, Predicting the bioactive conformations of macrocycles: a molecular dynamics-based docking procedure with DynaDock, J. Mol. Model. 25 (2019) 197. https://doi.org/10.1007/s00894-019-4077-5.
[131] M.L. Verdonk, J.C. Cole, M.J. Hartshorn, C.W. Murray, R.D. Taylor, Improved protein–ligand docking using GOLD, Proteins: Struct., Funct., Bioinf. 52 (2003) 609–623. https://doi.org/10.1002/prot.10465.
[132] O. Trott, A.J. Olson, AutoDock Vina, Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. (2009). https://doi.org/10.1002/jcc.21334. NA-NA.
[133] Y. Liu, M. Grimm, W. Dai, M. Hou, Z.-X. Xiao, Y. Cao, Cb-Dock, A web server for cavity detection-guided protein–ligand blind docking, Acta Pharmacol. Sin. 41 (2020) 138–144. https://doi.org/10.1038/s41401-019-0228-6.
[134] P. Mudgil, B. Baby, Y.-Y. Ngoh, H. Kamal, R. Vijayan, C.-Y. Gan, S. Maqsood, Molecular binding mechanism and identification of novel anti-hypertensive and anti-inflammatory bioactive peptides from camel milk protein hydrolysates, LWT 112 (2019), 108193. https://doi.org/10.1016/j.lwt.2019.05.091.
[135] T. Panyayai, P. Sangsawad, E. Pacharawongsakda, O. Sawatdichaikul, S. Tongsima, K. Choowongkomon, The potential peptides against angiotensin-I converting enzyme through a virtual tripeptide-constructing library, Comput. Biol. Chem. 77 (2018) 207–213. https://doi.org/10.1016/j.compbiolchem.2018.10.001.
[136] F. Tonolo, L. Moretto, A. Grinzato, F. Fiorese, A. Folda, V. Scalcon, S. Ferro, G. Arrigoni, M. Bellamio, E. Feller, A. Bindoli, O. Marin, M.P. Rigobello, Fermented soy-derived bioactive peptides selected by a molecular docking approach show antioxidant properties involving the keap1/nrf2 pathway, Antioxidants 9 (2020) 1306. https://doi.org/10.3390/antiox9121306.
[137] S. Gunalan, K. Somarathinam, J. Bhattacharya, S. Srinivasan, S.M. Jaimohan, R. Manoharan, S. Ramachandran, S. Kanagaraj, G. Kothandan, Understanding the dual mechanism of bioactive peptides targeting the enzymes involved in Renin Angiotensin System (RAS): an *in-silico* approach, J. Biomol. Struct. Dyn. 38 (2020) 5044–5061. https://doi.org/10.1080/07391102.2019.1695668.
[138] D.R. Houston, M.D. Walkinshaw, Consensus docking: improving the reliability of docking in a virtual screening context, J. Chem. Inf. Model. 53 (2013) 384–390. https://doi.org/10.1021/ci300399w.
[139] X. Ren, Y.-S. Shi, Y. Zhang, B. Liu, L.-H. Zhang, Y.-B. Peng, R. Zeng, Novel consensus docking strategy to improve ligand pose prediction, J. Chem. Inf. Model. 58 (2018) 1662–1668. https://doi.org/10.1021/acs.jcim.8b00329.
[140] J. Preto, F. Gentile, Assessing and improving the performance of consensus docking strategies using the DockBox package, J. Comput. Aided Mol. Des. 33 (2019) 817–829. https://doi.org/10.1007/s10822-019-00227-7.
[141] E.H.B. Maia, L.R. Medaglia, A.M. da Silva, A.G. Taranto, Molecular architect: a user-friendly workflow for virtual screening, ACS Omega 5 (2020) 6628–6640. https://doi.org/10.1021/acsomega.9b04403.
[142] S. Rosignoli, A. Paiardini, DockingPie: a consensus docking plugin for PyMOL, Bioinformatics 38 (2022) 4233–4234. https://doi.org/10.1093/bioinformatics/btac452.
[143] A.H. Pripp, Docking and virtual screening of ACE inhibitory dipeptides, Eur. Food Res. Technol. 225 (2007) 589–592. https://doi.org/10.1007/s00217-006-0450-6.
[144] N.T.P. Nong, J.-L. Hsu, Bioactive peptides: an understanding from current screening methodology, Processes 10 (2022) 1114. https://doi.org/10.3390/pr10061114.
[145] D.L. Mobley, K.A. Dill, Binding of small-molecule ligands to proteins: "what you see" is not always "what you get, Structure 17 (2009) 489–498. https://doi.org/10.1016/j.str.2009.02.010.
[146] D.E. Shaw, R.O. Dror, J.K. Salmon, J.P. Grossman, K.M. Mackenzie, J.A. Bank, C. Young, M.M. Deneroff, B. Batson, K.J. Bowers, E. Chow, M.P. Eastwood, D.J. Ierardi, J.L. Klepeis, J.S. Kuskin, R.H. Larson, K. Lindorff-Larsen, P. Maragakis, M.A. Moraes, S. Piana, Y. Shan, B. Towles, Millisecond-scale molecular dynamics simulations on Anton, in: Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis, 2009, pp. 1–11. https://doi.org/10.1145/1654059.1654126.
[147] R.B. Best, X. Zhu, J. Shim, P.E.M. Lopes, J. Mittal, M. Feig, A.D. MacKerell Jr., Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain $\chi 1$ and $\chi 2$ dihedral angles, J. Chem. Theor. Comput. 8 (2012) 3257–3273. https://doi.org/10.1021/ct300400x.
[148] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, J. Am. Chem. Soc. 117 (1995) 5179–5197. https://doi.org/10.1021/ja00124a002.
[149] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B.L. de Groot, H. Grubmüller, A.D. MacKerell, CHARMM36m: an improved force field for folded and intrinsically disordered proteins, Nat. Methods 14 (2017) 71–73. https://doi.org/10.1038/nmeth.4067.
[150] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, J. Am. Chem. Soc. 118 (1996) 11225–11236. https://doi.org/10.1021/ja9621760.
[151] A.D. MacKerell Jr., D. Bashford, M. Bellott, R.L. Dunbrack Jr., J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B 102 (1998) 3586–3616. https://doi.org/10.1021/jp973084f.
[152] E.A. Ploetz, S. Karunaweera, P.E. Smith, Kirkwood–buff-derived force field for peptides and proteins: applications of KBFF20, J. Chem. Theor. Comput. 17 (2021) 2991–3009. https://doi.org/10.1021/acs.jctc.1c00076.
[153] C. Tian, K. Kasavajhala, K.A.A. Belfon, L. Raguette, H. Huang, A.N. Migues, J. Bickel, Y. Wang, J. Pincay, Q. Wu, C. Simmerling, ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution, J. Chem. Theor. Comput. 16 (2020) 528–552. https://doi.org/10.1021/acs.jctc.9b00591.

[154] T. Feng, M. Li, J. Zhou, H. Zhuang, F. Chen, R. Ye, O. Campanella, Z. Fang, Application of molecular dynamics simulation in food carbohydrate research—a review, Innovat. Food Sci. Emerg. Technol. 31 (2015) 1–13. https://doi.org/10.1016/j.ifset.2015.06.015.
[155] S.A. Hollingsworth, R.O. Dror, Molecular dynamics simulation for all, Neuron 99 (2018) 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011.
[156] J. Comer, M. Bassette, R. Burghart, M. Loyd, S. Ishiguro, E.R. Azhagiya Singam, A. Vergara-Jaque, A. Nakashima, K. Suzuki, B.V. Geisbrecht, M. Tamura, Beta-1,3 oligoglucans specifically bind to immune receptor CD28 and may enhance T cell activation, IJMS 22 (2021) 3124. https://doi.org/10.3390/ijms22063124.
[157] S. Ishiguro, D. Upreti, M. Bassette, E.R.A. Singam, R. Thakkar, M. Loyd, M. Inui, J. Comer, M. Tamura, Local immune checkpoint blockade therapy by an adenovirus encoding a novel PD-L1 inhibitory peptide inhibits the growth of colon carcinoma in immunocompetent mice, Translational Oncology 16 (2022), 101337. https://doi.org/10.1016/j.tranon.2021.101337.
[158] G. Kalyan, V. Junghare, S. Bhattacharya, S. Hazra, Understanding structure-based dynamic interactions of antihypertensive peptides extracted from food sources, J. Biomol. Struct. Dyn. 39 (2021) 635–649. https://doi.org/10.1080/07391102.2020.1715836.
[159] M. Amezcua, L. El Khoury, D.L. Mobley, SAMPL7 Host–Guest Challenge Overview: assessing the reliability of polarizable and non-polarizable methods for binding free energy calculations, J. Comput. Aided Mol. Des. 35 (2021) 1–35. https://doi.org/10.1007/s10822-020-00363-5.
[160] M. Amezcua, J. Setiadi, Y. Ge, D.L. Mobley, An overview of the SAMPL8 host–guest binding challenge, J. Comput. Aided Mol. Des. 36 (2022) 707–734. https://doi.org/10.1007/s10822-022-00462-5.
[161] A. Rizzi, S. Murkli, J.N. McNeill, W. Yao, M. Sullivan, M.K. Gilson, M.W. Chiu, L. Isaacs, B.C. Gibb, D.L. Mobley, J.D. Chodera, Overview of the SAMPL6 host–guest binding affinity prediction challenge, J. Comput. Aided Mol. Des. 32 (2018) 937–963. https://doi.org/10.1007/s10822-018-0170-6.
[162] J. Yin, N.M. Henriksen, D.R. Slochower, M.R. Shirts, M.W. Chiu, D.L. Mobley, M.K. Gilson, Overview of the SAMPL5 host–guest challenge: are we doing better? J. Comput. Aided Mol. Des. 31 (2017) 1–19. https://doi.org/10.1007/s10822-016-9974-4.
[163] V.Y. Berdan, P.C. Klauser, L. Wang, Covalent peptides and proteins for therapeutics, Bioorg. Med. Chem. 29 (2021), 115896. https://doi.org/10.1016/j.bmc.2020.115896.
[164] B.R. Jagger, S.E. Kochanek, S. Haldar, R.E. Amaro, A.J. Mulholland, Multiscale simulation approaches to modeling drug–protein binding, Curr. Opin. Struct. Biol. 61 (2020) 213–221. https://doi.org/10.1016/j.sbi.2020.01.014.
[165] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J.W. Ponder, P. Ren, Polarizable atomic multipole-based AMOEBA force field for proteins, J. Chem. Theor. Comput. 9 (2013) 4046–4063. https://doi.org/10.1021/ct4003702.
[166] F.-Y. Lin, J. Huang, P. Pandey, C. Rupakheti, J. Li, B. Roux, A.D. MacKerell Jr., Further optimization and validation of the classical drude polarizable protein force field, J. Chem. Theor. Comput. 16 (2020) 3221–3239. https://doi.org/10.1021/acs.jctc.0c00057.
[167] J.A. Rackers, Q. Wang, C. Liu, J.-P. Piquemal, P. Ren, J.W. Ponder, An optimized charge penetration model for use with the AMOEBA force field, Phys. Chem. Chem. Phys. 19 (2016) 276–291. https://doi.org/10.1039/C6CP06017J.
[168] A. Vergara-Jaque, P. Fong, J. Comer, Iodide binding in sodium-coupled cotransporters, J. Chem. Inf. Model. 57 (2017) 3043–3055. https://doi.org/10.1021/acs.jcim.7b00521.
[169] B.R.I. Miller, T.D. McGee Jr., J.M. Swails, N. Homeyer, H. Gohlke, A.E. Roitberg, MMPBSA.py: an efficient program for end-state free energy calculations, J. Chem. Theor. Comput. 8 (2012) 3314–3321. https://doi.org/10.1021/ct300418h.
[170] M.S. Valdés-Tresanco, M.E. Valdés-Tresanco, P.A. Valiente, E. Moreno, gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS, J. Chem. Theor. Comput. 17 (2021) 6281–6291. https://doi.org/10.1021/acs.jctc.1c00645.
[171] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, T.E. Cheatham, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models, Acc. Chem. Res. 33 (2000) 889–897. https://doi.org/10.1021/ar000033j.
[172] E.R. Azhagiya Singam, Y. Zhang, G. Magnin, I. Miranda-Carvajal, L. Coates, R. Thakkar, H. Poblete, J. Comer, Thermodynamics of adsorption on graphenic surfaces from aqueous solution, J. Chem. Theor. Comput. 15 (2019) 1302–1316. https://doi.org/10.1021/acs.jctc.8b00830.
[173] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1–2 (2015) 19–25. https://doi.org/10.1016/j.softx.2015.06.001.
[174] S. Genheden, O. Kuhn, P. Mikulskis, D. Hoffmann, U. Ryde, The normal-mode entropy in the MM/GBSA method: effect of system truncation, buffer region, and dielectric constant, J. Chem. Inf. Model. 52 (2012) 2079–2088. https://doi.org/10.1021/ci3001919.
[175] M. Karplus, J.N. Kushick, Method for estimating the configurational entropy of macromolecules, Macromolecules 14 (1981) 325–332. https://doi.org/10.1021/ma50003a019.
[176] H. Sun, L. Duan, F. Chen, H. Liu, Z. Wang, P. Pan, F. Zhu, J.Z.H. Zhang, T. Hou, Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches, Phys. Chem. Chem. Phys. 20 (2018) 14450–14460. https://doi.org/10.1039/C7CP07623A.
[177] H. Fu, W. Cai, J. Hénin, B. Roux, C. Chipot, New coarse variables for the accurate determination of standard binding free energies, J. Chem. Theor. Comput. 13 (2017) 5173–5178. https://doi.org/10.1021/acs.jctc.7b00791.
[178] J.D. Chodera, D.L. Mobley, M.R. Shirts, R.W. Dixon, K. Branson, V.S. Pande, Alchemical free energy methods for drug discovery: progress and challenges, Curr. Opin. Struct. Biol. 21 (2011) 150–160. https://doi.org/10.1016/j.sbi.2011.01.011.
[179] J.C. Gumbart, B. Roux, C. Chipot, Standard binding free energies from computer simulations: what is the best strategy? J. Chem. Theor. Comput. 9 (2013) 794–802. https://doi.org/10.1021/ct3008099.
[180] H.-J. Woo, B. Roux, Calculation of absolute protein–ligand binding free energy from computer simulations, Proc. Natl. Acad. Sci. USA 102 (2005) 6825–6830. https://doi.org/10.1073/pnas.0409005102.
[181] H. Fu, J.C. Gumbart, H. Chen, X. Shao, W. Cai, C. Chipot, BFEE: a user-friendly graphical interface facilitating absolute binding free-energy calculations, J. Chem. Inf. Model. 58 (2018) 556–560. https://doi.org/10.1021/acs.jcim.7b00695.
[182] H. Fu, H. Chen, W. Cai, X. Shao, C. Chipot, BFEE2: automated, streamlined, and accurate absolute binding free-energy calculations, J. Chem. Inf. Model. 61 (2021) 2116–2123. https://doi.org/10.1021/acs.jcim.1c00269.
[183] T. Senapathi, M. Suruzhon, C.B. Barnett, J. Essex, K.J. Naidoo, BRIDGE: an open platform for reproducible high-throughput free energy simulations, J. Chem. Inf. Model. 60 (2020) 5290–5295. https://doi.org/10.1021/acs.jcim.0c00206.
[184] C. Velez-Vega, M.K. Gilson, Overcoming dissipation in the calculation of standard binding free energies by ligand extraction, J. Comput. Chem. 34 (2013) 2360–2371. https://doi.org/10.1002/jcc.23398.
[185] R. Thakkar, D. Upreti, S. Ishiguro, M. Tamura, J. Comer, Computational design of a cyclic peptide that inhibits the CTLA4 immune checkpoint, RSC Medicinal Chemistry (2023). https://doi.org/10.1039/D2MD00409G.
[186] P.R. Arantes, M.D. Polêto, C. Pedebos, R. Ligabue-Braun, Making it rain: cloud-based molecular simulations for everyone, J. Chem. Inf. Model. 61 (2021) 4852–4856. https://doi.org/10.1021/acs.jcim.1c00998.
[187] W.N. Baba, B. Baby, P. Mudgil, C.-Y. Gan, R. Vijayan, S. Maqsood, Pepsin generated camel whey protein hydrolysates with potential antihypertensive properties: identification and molecular docking of antihypertensive peptides, LWT 143 (2021), 111135. https://doi.org/10.1016/j.lwt.2021.111135.
[188] W.N. Baba, P. Mudgil, B. Baby, R. Vijayan, C.-Y. Gan, S. Maqsood, New insights into the cholesterol esterase- and lipase-inhibiting potential of bioactive peptides from camel whey hydrolysates: identification, characterization, and molecular interaction, J. Dairy Sci. 104 (2021) 7393–7405. https://doi.org/10.3168/jds.2020-19868.
[189] G. Kalyan, V. Junghare, M.F. Khan, S. Pal, S. Bhattacharya, S. Guha, K. Majumder, S. Chakrabarty, S. Hazra, Anti-hypertensive peptide predictor: a machine learning-empowered web server for prediction of food-derived peptides with potential angiotensin-converting enzyme-I inhibitory activity, J. Agric. Food Chem. 69 (2021) 14995–15004. https://doi.org/10.1021/acs.jafc.1c04555.
[190] I. Marcet, J. Delgado, N. Díaz, M. Rendueles, M. Díaz, Peptides recovery from egg yolk lipovitellins by ultrafiltration and their in silico bioactivity analysis, Food Chem. 379 (2022), 132145. https://doi.org/10.1016/j.foodchem.2022.132145.
[191] T. Wenhui, H. Shumin, Z. Yongliang, S. Liping, Y. Hua, Identification of in vitro angiotensin-converting enzyme and dipeptidyl peptidase IV inhibitory peptides from draft beer by virtual screening and molecular docking, J. Sci. Food Agric. 102 (2022) 1085–1094. https://doi.org/10.1002/jsfa.11445.
[192] G. Wei, Q. Zhao, D. Wang, Y. Fan, Y. Shi, A. Huang, Novel ACE inhibitory, antioxidant and α-glucosidase inhibitory peptides identified from fermented rubing cheese through peptidomic and molecular docking, LWT 159 (2022), 113196. https://doi.org/10.1016/j.lwt.2022.113196.
[193] R.A. Tahir, A. Bashir, M.N. Yousaf, A. Ahmed, Y. Dali, S. Khan, S.A. Sehgal, In Silico identification of angiotensin-converting enzyme inhibitory peptides from MRJP1, PLoS One 15 (2020), e0228265. https://doi.org/10.1371/journal.pone.0228265.
[194] W. Zhao, D. Zhang, Z. Yu, L. Ding, J. Liu, Aminopeptidase N inhibitory peptides derived from hen eggs: virtual screening, inhibitory activity, and action mechanisms, Food Biosci. 37 (2020), 100703. https://doi.org/10.1016/j.fbio.2020.100703.
[195] W. Xue, X. Liu, W. Zhao, Z. Yu, Identification and molecular mechanism of novel tyrosinase inhibitory peptides from collagen, J. Food Sci. 87 (2022) 2744–2756. https://doi.org/10.1111/1750-3841.16160.
[196] P. Mudgil, B. Baby, Y.-Y. Ngoh, R. Vijayan, C.-Y. Gan, S. Maqsood, Identification and molecular docking study of novel cholesterol esterase inhibitory peptides from camel milk proteins, J. Dairy Sci. 102 (2019) 10748–10759. https://doi.org/10.3168/jds.2019-16520.
[197] S. Attique, M. Hassan, M. Usman, R. Atif, S. Mahboob, K. Al-Ghanim, M. Bilal, M. Nawaz, A molecular docking approach to evaluate the pharmacological properties of natural and synthetic treatment candidates for use against hypertension, IJERPH 16 (2019) 923. https://doi.org/10.3390/ijerph16060923.
[198] W. Evangelista Falcon, S.R. Ellingson, J.C. Smith, J. Baudry, Ensemble docking in drug discovery: how many protein configurations from molecular dynamics simulations are needed to reproduce known ligand binding? J. Phys. Chem. B 123 (2019) 5189–5195. https://doi.org/10.1021/acs.jpcb.8b11491.

[199] O. Korb, T.S.G. Olsson, S.J. Bowden, R.J. Hall, M.L. Verdonk, J.W. Liebeschuetz, J.C. Cole, Potential and limitations of ensemble docking, J. Chem. Inf. Model. 52 (2012) 1262−1274. https://doi.org/10.1021/ci2005934.
[200] Z. Du, X. Ding, W. Hsu, A. Munir, Y. Xu, Y. Li. pLM4Ace: A Protein Language Model-Based Machine Learning Predictor for Screening Peptides with High Antihypertensive Activity. (submitted for publication).
[201] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583−589. https://doi.org/10.1038/s41586-021-03819-2.
[202] E. Alnabati, G. Terashi, D. Kihara, Protein structural modeling for electron microscopy maps using VESPER and MAINMAST, Curr Protoc 2 (2022) e494. https://doi.org/10.1002/cpz1.494.
[203] E. Callaway, The revolution will not be crystallized: a new method sweeps through structural biology, Nature 525 (2015) 172−174. https://doi.org/10.1038/525172a.
[204] H. Chen, J.D.C. Maia, B.K. Radak, D.J. Hardy, W. Cai, C. Chipot, E. Tajkhorshid, Boosting free-energy perturbation calculations with GPU-accelerated NAMD, J. Chem. Inf. Model. 60 (2020) 5301−5307. https://doi.org/10.1021/acs.jcim.0c00745.
[205] P. Eastman, J. Swails, J.D. Chodera, R.T. McGibbon, Y. Zhao, K.A. Beauchamp, L.-P. Wang, A.C. Simmonett, M.P. Harrigan, C.D. Stern, R.P. Wiewiora, B.R. Brooks, V.S. Pande, OpenMM 7: rapid development of high performance algorithms for molecular dynamics, PLoS Comput. Biol. 13 (2017), e1005659. https://doi.org/10.1371/journal.pcbi.1005659.
[206] N. Kondratyuk, V. Nikolskiy, D. Pavlov, V. Stegailov, GPU-accelerated molecular dynamics: state-of-art software performance and porting from Nvidia CUDA to AMD HIP, Int. J. High Perform. Comput. Appl. 35 (2021) 312−324. https://doi.org/10.1177/10943420211008288.
[207] C. Kutzner, S. Páll, M. Fechner, A. Esztermann, B.L. de Groot, H. Grubmüller, More bang for your buck: improved use of GPU nodes for GROMACS 2018, J. Comput. Chem. 40 (2019) 2418−2431. https://doi.org/10.1002/jcc.26011.
[208] S. Páll, A. Zhmurov, P. Bauer, M. Abraham, M. Lundborg, A. Gray, B. Hess, E. Lindahl, Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS, J. Chem. Phys. 153 (2020), 134110. https://doi.org/10.1063/5.0018516.
[209] J.C. Phillips, D.J. Hardy, J.D.C. Maia, J.E. Stone, J.V. Ribeiro, R.C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M.C.R. Melo, B.K. Radak, R.D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L.V. Kalé, K. Schulten, C. Chipot, E. Tajkhorshid, Scalable molecular dynamics on CPU and GPU architectures with NAMD, J. Chem. Phys. 153 (2020), 044130. https://doi.org/10.1063/5.0014475.
[210] A. Daina, O. Michielin, V. Zoete, SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules, Nucleic Acids Res. 47 (2019) W357−W364. https://doi.org/10.1093/nar/gkz382.