ISSN 1068-1620, Russian Journal of Bioorganic Chemistry, 2018, Vol. 44, No. 4, pp. 367–385. © Pleiades Publishing, Ltd., 2018.
Original Russian Text © S.D. Kalmykova, G.P. Arapidi, A.S. Urban, M.S. Osetrova, V.D. Gordeeva, V.T. Ivanov, V.M. Govorun, 2018, published in Bioorganicheskaya Khimiya, 2018, Vol. 44, No. 4, pp. 361–381.

REVIEW ARTICLE

In Silico Analysis of Peptide Potential Biological Functions

S. D. Kalmykova[a, b], G. P. Arapidi[a, b, c, 1], A. S. Urban[a, b], M. S. Osetrova[a, b], V. D. Gordeeva[a, b, c], V. T. Ivanov[a], and V. M. Govorun[a, b, c]

[a] *Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, 119779 Russia*
[b] *Moscow Institute of Physics and Technology, Dolgoprudnyi, 141701 Russia*
[c] *Federal Research and Clinical Center of Physical–Chemical Medicine, Moscow, 119435 Russia*

Received November 15, 2017; in final form, February 9, 2018

Abstract—Over the past decade, tools of omics technologies have generated a large amount of data in various repositories, which are of interest for meta-analysis today. Now, researchers in the field of proteomics and peptidomics focus not on sequencing, but on functions performed by molecules and metabolic interactions, in which the proteins or peptides participate. As a result of a single LC-MS/MS analysis, several thousand unique peptides can be identified, each of which may be bioactive. A classic technique for determining the peptide function is a direct experiment. Bioinformatics approaches as a preliminary analysis of potential biological functions are an important step and are able to significantly reduce time and cost of experimental verification. This article provides an overview of computational methods for predicting biological functions of peptides. Approaches based on machine learning, which are the most popular today, algorithms using structural, evolutionary, or statistical patterns, as well as methods based on molecular docking, are considered. Databases of bioactive peptides are reported, providing information necessary to construct new algorithms for predicting biological functions. Attention is paid to the characteristics of peptides, on the basis of which it is possible to draw conclusions about their bioactivity. In addition, the report provides a list of online services that may be used by researchers to analyze potential activities of peptides with which they work.

Keywords: proteomics, peptidomics, bioinformatics, bioactive peptides, biological functions of peptides, machine learning, molecular docking

DOI: 10.1134/S106816201804009X

INTRODUCTION

Today, proteomics, including the subdivision of peptidomics, is one of the rapidly developing areas of biochemistry. These sciences study the totality of proteins and protein fragments (peptides), their functions, and interactions in living organisms [1–4]. The main methods of analysis in proteomics and peptidomics are mass spectrometry, various electrophoretic techniques, and liquid chromatography [3, 5, 6]. Special algorithms are widely employed to identify and further analyze peptides revealed by mass spectrometry. The data obtained allow for analysis of peptide abundance in various samples (comparative peptidomics), determine the position of a peptide in a protein sequence, elucidate which proteases were involved in the generation of the peptides, and predict structures and functions thereof [2, 3]. One of the possible aims of such studies is the search for and analysis of bioactive peptides that modulate physiological functions through binding with specific receptors or other targets [7].

There are various sources of bioactive peptides. On one hand, there are endogenous bioactive peptides that are synthesized in the cell as part of large preproproteins, which are then cleaved and modified. They comprise hormones and other signal peptides [8], as well as neuropeptides [9]. Products of degradation of the organism proteins, or degradome [10–12], is of interest from the point of view of search for bioactive peptides, as well as biomarkers, especially when active degradation occurs upon a strong external effect on a living system [13]. Moreover, products of translation of short reading frames or products of long noncoding RNAs, analysis of which has been gaining popularity lately, also belong to endogenous peptides [14–16].

Another source of bioactive peptides are exogenous molecules, for example, alimentary peptides [17]: milk [18] and dairy products, e.g. cheese and yogurt [19, 20] are widely studied. Milk proteins are inert for the human organism, however, some peptides formed upon their cleavage by enzymes of the gastrointestinal tract or hydrolysis by intestinal microbiota, are bioactive molecules [21]. Biologically active peptides are found in other food products of animal origin, such as meat [22] and skin gelatin [23]. Plant food also provides bioactive peptides: wheat [24], soybean [25], rice

[1] Corresponding author: phone: +7 (926) 471-14-20, e-mail: arapidi@gmail.com.

[26], and amarantus [27]. Moreover, bioactive peptides are obtained in vitro through modification of existing substances or de novo synthesis thereof. Peptides of crustaceans [28] and algae [29] are often used for introduction of modifications.

Bioactive peptides can affect nearly all systems in the human organism [30]. The cardiovascular system is influenced by antithrombotic peptides, which inhibit platelet aggregation and fibrinogen binding [31], hypotensive peptides, which inhibit angiotensin-transforming enzyme [32], and antioxidant peptides able to scavenge free radicals (produced as by-products of cell oxidative metabolism) and inhibit lipid peroxidation [25]. The nervous system is regulated by opioid peptides, which are similar to enkephalins by their structure and exhibit affinity toward opioid receptors, thus causing opiate-like effects on brain [33]. The digestive system is affected by peptides, which by means of anionic amino acid residues in their structure form soluble complexes with calcium, magnesium, and other minerals [34]. The digestion system is influenced by antimicrobial peptides exhibiting bacteriostatic (inhibiting bacterial proliferation) and bactericidal (causing bacteria death) properties [18]. Such peptides bind bacteria cell surfaces and thus inhibit membrane function or destroy bacterial biofilms. Moreover, some peptides have antiviral or antifungal activity [30].

Antimicrobial peptides can replace antibiotics, which become yet less efficient due to growing bacterial resistance [35]. Antimicrobial peptides affect the immune system. Besides, immunomodulatory peptides are also immune system effectors that increase phagocytosis activity of macrophages, stimulate proliferation and maturation of immune cells, and increase antibody synthesis [36]. Alimentary peptides responsible for immunity modulation in children/infants are due to this group [37]. Another group is represented by immunosuppressing peptides, which are used to suppress autoimmune diseases and lower the risk of transplant rejection. Also, a large fraction of immunomodulatory peptides is represented by peptides capable of tumor-cell recognition and modulation of their activity [38].

Today, studies aimed at searching for peptide drugs are gaining in popularity [39–41]. Natural peptides possess low immunogenicity and are able to penetrate tissues better than proteins. Another advantage of natural bioactive peptides is their higher affinity to tissues compared to synthetic agents, which ensures their slower elimination from the organism [42].

Development of high performance automated systems for screening of peptides is a natural step in the development of the area. Similar to pharmaceutical studies, the search for peptides with certain types of bioactivity should include accumulation of primary data on the list of peptides, among which the target compounds should be searched for, in silico filtration of the most likely active, synthesis of the candidates in vitro, and testing them in experiments in vivo. Both peptides identified in the course of a scientific study and predicted by computational techniques can be analyzed. In the area of automated manipulation with biopolymers, Boles et al. suggested the concept of biological agent production on demand [43]. Bioinformatics approaches to prediction of peptide bioactivity are a natural and important complement to the concept. These approaches are as much in demand as molecular docking in pharmaceuticals.

The biological function of a peptide is established in the course of experimental work: by adding the peptide to cell culture [44] or searching for proteins the peptide binds with by means of dihybrid analysis [45], Förster resonance energy transfer (FRET) [46], or coimmunoprecipitation [47]. These studies are rather labor consuming and expensive. To improve their performance, protein microchips are being used [48].

As a result of peptidome analysis, several thousand unique peptides are being identified [3], each of which potentially can happen to be a bioactive one, and no one knows what its activity might be. Therefore, preliminary filtration with bioinformatics approaches is necessary for isolation of the most probable bioactive peptides [3].

As of today, many algorithms for protein function prediction have been created and are actively used [49–54]. However, far from all of them can be applied to analysis of peptide function. The reason is that most of modern methods of protein function prediction use their tertiary structure for analysis [55, 56]. Usually, peptides are too short to form a stable spatial structure. Attempts to apply methods of analysis based on knowledge of protein–protein interaction networks [54] also meet difficulties, since today little is known on the place of peptides in these networks. Therefore, special services are required to search for bioactive peptides.

The review focuses on bioinformatics methods to predict functions of bioactive peptides. It provides references to databases of bioactive peptides and works describing techniques to design a peptide analysis algorithm that would be of interest to specialists working in the area. In addition, the plethora of examples of peptide function studies and description of services for target peptide function evaluation without the prolonged step of algorithm creation in the review will be of interest for experimental research scientists.

APPLICATION OF STRUCTURAL, EVOLUTIONARY, OR STATISTICAL PATTERNS TO SEARCH FOR FUNCTIONALLY ACTIVE PEPTIDES

Algorithms to search for bioactive peptides and predict their functions can be divided into two large groups, namely algorithms created using machine

learning techniques and those developed without machine learning. From a historical perspective, methods not utilizing machine learning appeared the first. Their work is based on one of the three approaches: search for similar fragments, search for evolutionarily conserved motives, and search for statistical trends.

The first approach relying on search for similar fragments in amino acid sequences of peptides remains the simplest and most often used. Supposedly, similar fragments perform similar functions since it has been demonstrated that if amino acid sequences of proteins are 70% similar, the probability that the proteins will have the same function is over 90% [57]. Besides, today, some short amino acid patterns of fixed length associated with certain functions, or motives, are known [58]. For example, short linear motifs, SLiMs, are responsible for protein–peptide interactions in eukaryotic cells [59], while D box and KEN box are motifs of cyclosome that bind ubiquitin in the process of protein degradation [60]. In general, the minimum length required for an amino acid sequence to be involved in a protein–protein interaction and immune recognition is 5 amino acids [61]. Therefore, peptides 6 to 30 a.a. long with similar sequences are highly likely to perform similar functions. The following algorithms are most often used to search for similar sequences: NCBI BLAST [62] for local alignment, T-Coffee [63] or Clustal [64] for multiple alignment, and HMMER [65] for search of weak homology or sequence patterns. For example, these methods were used to search for new peptides in crustaceans [66]. BLAST was used to align peptides of partially annotated transcriptomes of *Tigriopus californicus* and *Lepeophtheirus salmonis* against peptides of known related species, *Calanus finmarchicus* and *Daphnia pulex*. This allowed for 82 new peptides to be discovered. Another example is the search for bioactive peptides of amarantus by homology with peptides of the BIOPEP database [67]. Work on search for bioactive peptides among peptide fragments obtained by in silico cleavage with the *Cyprinus carpio* fish stomach peptides in the BIOPEP database can also be attributed to this group [68].

Some researchers increase the number of potential bioactive peptides by sequentially applying several algorithms searching for similar motifs. In search for conotoxins, which are toxins of carnivorous marine cone-shaped snails, Lavergne and coauthors used patterns of peptide sequences from known families of conotoxins and the HMMER algorithm to search for homology [69]. When validated on proteins from the UniProtKB (UniProt Knowledgebase, http://uniprot.org/), specificity of the combined technique was 99.9%. Among additional criteria were used to discover bioactive peptides are annotations in protein databases [70] and information on the absence of the constant ordered secondary structure of a peptide, which can be derived for instance from the IUPRED server [71]. UniProtKB annotation was used in the search for an antimicrobial peptide and determination of its type [72]. In the process of a search for homologous sequences, all antimicrobial peptides contained in the database of the Biochemistry Department, University of Trieste, were divided into clusters. The proposed algorithm assigns a protein to a cluster using the HMMER algorithm. Then, proteins for which functional regions having been annotated (proproteins and mature peptides) overlap with homologous regions of the cluster by more than 90% are selected. Tenfold cross validation of the algorithm on 81 peptides from 20 clusters showed an average sensitivity of 81% and a specificity of 98.8%.

Other authors apply additional restrictions for bioactive peptide determination. For example, Liu and coauthors performed in silico proteolysis of *Drosophila melanogaster* proteins at major insect protease restriction sites in search for new bioactive peptides of *D. melanogaster* [73]. Then, local alignment was used to select sequences containing known bioactive peptides or more than a single similar fragment. Besides, initial *D. melanogaster* protein should have contained a signal peptide. The search yielded 118 potential bioactive peptides, 43 of which were known previously. Only a single one of the previously known bioactive peptides was not detected.

The second feature evidencing bioactivity of a certain peptide can be evolutionary conservation of its sequence. For example, in the search for peptides regulating platelet function, a search for orthologs was performed among vertebrates [74]. From almost three thousand proteins expressed in platelets, 47 with transmembrane domains were selected. To find functional peptides, 10-a.a. long sequences located in cytoplasm close to the membrane and having orthologs among vertebrates, were selected. This yielded 78 peptides, 13 of which were experimentally proven to be bioactive. A work by Michael and coauthors can also be referred to this group; they used evolutionary conservation as an additional criterion to search for bioactive peptides using motif resemblance [70].

The third group of algorithms not utilizing machine learning is based on search for statistical trends. Mainly, these trends are looked for among amino acid sequences and physicochemical properties because the latter define the functions performed by bioactive peptides. For example, hypotensive peptides are known to be obliged to have aromatic or basic amino acid at the N terminus and positive and hydrophobic amino acids at the C terminus, since the effect of most hypotensive peptides is based on the ability to inhibit the angiotensin-transforming enzyme, which requires this very structure to bind the enzyme [75]. Most antimicrobial peptides possess amphiphilic structure since they interact with the anionic cell wall and membrane of bacteria [76, 77].

To find peptides exerting antimicrobial properties, a table of amino acid proneness to antimicrobial effect was constructed; the table contained the antimicrobial activity of the bactenecin-derived peptides obtained by sequential substitution of each of its amino acids by one of the 19 remaining amino acids [78]. In a validation on 40 peptides either possessing or not the antimicrobial activity, the algorithm demonstrated 90% sensitivity and 81% specificity.

MACHINE LEARNING-BASED ALGORITHMS

Machine learning algorithms are widely used to search for bioactive peptides. In most cases, design of a prediction algorithm includes the following stages: generation of training, validation, and test sets, selection of parameters that will be used for prediction, choice of machine learning algorithm, its training and validation. Different kinds of software that search for and predict bioactive peptides differ in terms of which algorithm best fits the target, which parameters it utilizes, and how good (regarding diversity and sample size) was the training set. Let us consider each stage separately.

Generation of the training set using open databases. When a machine learning-based algorithm to predict bioactive peptides is being designed, appropriately formed training sets—of sufficient size, as well as diversity—are of great importance. Inappropriately selected data can result in an overfit algorithm, that is, one that produces misleading results when applied to data different from those used for training [79, 80].

Training sets are formed using open databases exemplified in Table 1 or upon analysis of publications. For example, to study peptides involved in quorum sensing (capacity for coordinated behavior in some bacteria), for which there was only a single ready-to-use database available when the algorithm was being designed [81], data from publication [82] were used. While open data on bioactive peptides are rather abundant (Table 1), selection of peptides that surely lack certain biological activity (negative control) is significantly hampered. Few databases contain information on peptides that do not possess certain activity [83–85] and in each work authors utilize a unique approach to form the negative control set. In a number of works, bioactive peptides with an altered amino acid sequence are used as those lacking activity [82]. Besides, peptides can be randomly selected from a database containing a large number of various proteins [86]. The latter approach is sometimes upgraded: only those protein fragments are chosen whose annotation lacks description of the biological activity of interest. For example, when an algorithm predicting peptide toxicity was being elaborated, random peptides from the UniProtKB database containing no words "toxin" or "allergen" in their description were used as a negative control [87]. An algorithm predicting antimicrobial peptides [88] utilizes only nonsecreted peptides as peptides without biological activity since it has been shown that most of natural antimicrobial peptides are secreted [89].

To exclude data redundancy leading to overfit, peptides with similar sequences are removed [90]. One of the examples of software performing such a filtration is the CD-HIT [91]. Moreover, oversampling or undersampling approaches are often used in the process of set generation to adjust sample size between groups being compared. The approach was used by the authors of an algorithm classifying antimicrobial peptides [92].

Choice of parameters for machine learning. Three characteristics are used as parameters to train the algorithms: amino acid composition, presence of a motif associated with a biological function, and physicochemical properties. Each of the parameters can be utilized in several different ways.

The first type of parameter reflects the amino acid composition of a peptide. These parameters can be included in the algorithm in various ways. The simplest variant is represented by the normalized amino acid frequencies in the peptide. For example, an algorithm predicting antibacterial peptides [135] is based on amino acid composition. When being validated on 466 peptides from the UniProtKB database, the algorithm showed 87.55% accuracy. Loss of information on the amino acid sequence is a disadvantage of the method.

In order to retain the information on amino acid sequence in a peptide, a so-called pseudo amino acid composition is used. In addition to amino acid composition the method utilizes hydrophobicity, charge, and mass of the peptides. The method has been applied to predict antimicrobial peptides and their functional type in an algorithm with sensitivity of 97.3% and specificity of 92.1% [92]. Besides, normalized frequencies of di- and tripeptides can be used. An algorithm searching for quorum-sensing peptides [82] utilizes the dipeptide composition approach. On a set of 20 peptides from the Quarumpeps database, its accuracy was 90%. Another approach relies on the use of binary profile of patterns, which are matrices of size 20 by peptide length (in a.a.), where columns correspond to positions in a peptide, and lines to the 20 amino acids. It contains 1 in cells where column and line match and 0 where they do not.

An algorithm presented in a work by Tyagi and coauthors [136] performs a search for anticancer peptides based on binary profile patterns. Validation of an independent set of 50 bioactive and 50 not bioactive peptides demonstrated an accuracy of 89%. An interesting approach was proposed in a paper devoted to the elaboration of an algorithm to search for anticancer peptides [137]. The authors used a vector of dimension 60, where the first 20 positions showed how much an amino acid frequency differs from a random distribution in the peptide, the second 20 positions

Table 1. Open data on bioactive peptides

Source name	Link to a source	Data stored
Antimicrobial and antiviral peptides		
APD [93]	http://aps.unmc.edu/AP/	Antimicrobial peptides and their activity
ParaPep [94]	http://crdd.osdd.net/raghava/parapep/home.php	Antiparasitic peptides and their structures
CAMPR3 [95]	http://www.camp3.bicnirrh.res.in/	Antimicrobial peptides
DBAASP [96]	https://dbaasp.org/home	Antimicrobial peptides, their activities, and structures
AVPdb [97]	http://crdd.osdd.net/servers/avpdb/	Antiviral peptides
AVPpred [83]	http://crdd.osdd.net/servers/avppred/collection.php	antiviral and NOT antiviral peptides
DAMPD [98]	http://apps.sanbi.ac.za/dampd/	Antimicrobial peptides
PhytAMP [99]	http://phytamp.hammamilab.org/about.php	plant antimicrobial peptides
BACTIBASE 2.0 [100]	http://bactibase.hammamilab.org/main.php	Bacteriocins, proteins and peptides produced by bacteria suppressing other strains of the same or related species
Bagel [101]	http://bagel2.molgenrug.nl/index.php/bacteriocin-database	Bacteriocins
DEFENSINS Knowledgebase [102]	http://defensins.bii.a-star.edu.sg/	Defensins, cationic peptides of immune system active against bacteria, fungi, and viruses
DADP [103]	http://split4.pmfst.hr/dadp/	Protective peptides of 167 species of tailless amphibians
BaAMPs [104]	http://www.baamps.it/	Peptides destroying bacterial biofilms
MilkAMP [105]	http://milkampdb.org/home.php	milk antimicrobial peptides
YADAMP [106]	http://yadamp.unisa.it/	Antimicrobial peptides
Anticancer peptides		
CancerPPD [107]	http://crdd.osdd.net/raghava/cancerppd/	Anticancer peptides
TumorHoPe [108]	http://crdd.osdd.net/raghava/tumorhope/	Tumor-migrating peptides
Immunology peptides		
Immune Epitope Database [109]	http://www.iedb.org/	Various epitopes
HPVdb [110]	http://cvc.dfci.harvard.edu/hpv/	Epitopes of human papilloma virus antigens
AgAbDb [111]	http://196.1.114.46:8080/agabdb2/home.jsp	Protein and peptide antigens
SDAP [112]	http://fermi.utmb.edu/SDAP/	Allergens
Allergome [113]	http://www.allergome.org/	Allergens
Protegen [114]	http://www.violinet.org/protegen/	Protective antigens
HSPVdb [115]	http://srs.bioinformatics.nl/hspv/	Epitopes of human MHC and T cells
Hypotensive peptides		
AHTPDB [116]	http://crdd.osdd.net/raghava/ahtpdb/	Hypotensive peptides
ACEpepDB	http://www.cftri.com/pepdb/	Food-derived hypotensive peptides

Table 1. (Contd.)

Source name	Link to a source	Data stored
	Toxins	
HEMOLYTIK [84]	http://crdd.osdd.net/raghava/hemolytik/	Hemolytic and NOT hemolytic peptides
ArachnoServer 2.0 [117]	http://www.arachnoserver.org/mainMenu.html;jsessionid=6A4C5048824C690782B3-BC72EFEB67D3	Spider toxins
ATDB [118]	http://protchem.hunnu.edu.cn/toxin/	Animal toxins
DBETH [119]	http://www.hpppi.iicb.res.in/btox/	Bacteria toxins
ConoServer [120]	http://www.conoserver.org/	Conopeptides, peptides of marine carnivorous cone-shaped snails
	Cell-penetrating peptides	
CPPsite 2.0 [121]	http://crdd.osdd.net/raghava/cppsite/	Cell-penetrating peptides, their secondary and tertiary structures
	Others	
PepBank [122]	http://pepbank.mgh.harvard.edu/	Bioactive peptides
PeptideDB	http://www.peptides.be/	Bioactive peptides of animal origin
StraPep	http://isyslab.info:8090/CBPS/	Bioactive peptides
Peptide Atlas [123]	http://www.peptideatlas.org/	Peptides detected in the course of mass spectrometry studies
BIOPEP [124]	http://www.uwm.edu.pl/biochemia/index.php/en/biopep	Bioactive peptides
SwePep [125]	http://www.swepep.se/	Peptides detected in the course of mass spectrometry studies
SATPdb	http://crdd.osdd.net/raghava/satpdb/	Therapeutic peptides: anticancer, antivirus, antiparasitic, antibacterial, drug delivery, and toxic peptides
WALTZ-DB [126]	http://waltzdb.switchlab.org/	Amyloid fibril-stimulating peptides
Quorumpeps [81]	http://quorumpeps.ugent.be/	Quorum sensing peptides
ELM [127]	http://elm.eu.org/	Short linear motifs of protein–protein interaction
BDB [128]	http://immunet.cn/bdb/	Peptides capable of binding with various molecules, tissues, organs, and organisms
SPdb [129]	http://proline.bic.nus.edu.sg/spdb/	Signal peptides
MBPDB [130]	http://mbpdb.nws.oregonstate.edu/	Milk bioactive peptides
BactPepDB	http://bactpepdb.rpbs.univ-paris-diderot.fr/cgi-bin/links.pl	Bacterial peptides
Brainpeps [131]	http://brainpeps.ugent.be/	Peptides capable of blood–brain barrier penetration
EROP-Moscow [132]	http://erop.inbi.ras.ru/index.html	Human regulatory peptides
HMRbase [133]	http://crdd.osdd.net/raghava/hmrbase/	Hormones and receptors
MAHMI [134]	http://sing.ei.uvigo.es/mahmi/index.php	Bacterial peptides

reflected how much more frequent was an amino acid in a certain fragment of a protein, and the last 20 positions were associated with the presence of clusters in a sequence. Euclidean metrics was used to calculate the average distance between amino acid sequences and make a prediction using a support vector machine. Sensitivity and specificity calculated on a set of 80 peptides were 95 and 97%, respectively.

The second type of parameters used to predict bioactive peptides with machine learning algorithms are motifs associated with a certain function. The algorithms generate an estimate based on the presence of certain motifs in a peptide. The strategy was used by Gupta and coauthors during development of an algorithm to predict anti-inflammatory peptides using a support vector machine [138]. The method accuracy was 78.1%. A similar method was applied to predict peptide toxicity [87]. In a validation on a set of data compiled from proteins of the UniProtKB database using a keyword search, the algorithm showed 99.39% sensitivity and 97.91% specificity.

Algorithms attempting to find one of known motifs in a peptide have been created; if a motif is found, the peptide is considered bioactive, if not, machine learning methods are further applied. The approach is utilized by an algorithm predicting hemolytic peptides [139]. If the peptide contains no known motifs associated with a certain biological activity, the algorithm makes a prediction using a support vector machine over various types of amino acid composition. The algorithm sensitivity was 96.4% and specificity, 99.1%.

An original approach based on motif search was used to elaborate an algorithm to search for peptides binding the main histocompatibility complex (MHC) types I and II [140]. It used a machine learning algorithm to reveal frequently occurring motifs, including not serial ones. If an amino acid occurred often after another amino acid in a frame of preset length, the event was considered a motif. The search proceeded by progressive addition of amino acids, that is, first, amino acids frequently encountered in peptides were considered; then, those that follow them in a frame of certain length; etc. The AUC metrics of the algorithm was 0.897 for 9 and 10-a.a. long peptides. The AUC for 9-a.a. long peptides of MHC type I was 0.908 and MHC type II, 0.779.

The third type of parameters that are used for machine learning models to predict peptide functions—physicochemical properties—usually utilize one of the existing services: AAindex db [99], PaDEL [142], as it has been done by Kumar and coauthors [143], or peptide package in R [144], as in the work by Meher and coauthors [145]. The resulting set of properties can be obtained with a simple averaging of physicochemical values for each of the amino acids contained in the peptide, as in the work of Andreu and Torrent [146]. An algorithm to predict antimicrobial peptides with an artificial neural network with sensitivity and specificity at the level of 85% was created. A more complicated procedure to construct parameters based on physicochemical properties is the utilization of autocorrelation, that is the relationship between the properties of amino acid sequences within a single peptide taken with a shift [147]. For example, in the course of development of an algorithm predicting allergens, physicochemical properties were transformed into a single parameter by means of the auto- and cross-correlation transformation [148]. Autocorrelation can achieve better results through taking into consideration the order of amino acids in the peptide [149]. Typically, secondary structure, hydrophobicity, isoelectric point, total molecular charge [145], proneness to aggregation in vitro and in vivo, and chain length [150] are used as properties.

In their work, Li and Wang proposed using chemical shifts of $^{1}H_{N}$, $^{13}C_{\alpha}$, and ^{15}N in combination with amino acid composition to predict anticancer activity of peptides [151]. The choice of the parameters is supported by the fact that chemical shifts of the nuclei depend on the secondary structure of peptides [152], while peptide functions are determined by their structure. The algorithm has a sensitivity of 89.86% and specificity of 96.12% with a jackknife validation on a data set containing 138 positive and 206 negative examples. Another way to use physicochemical properties was applied in a method to predict antibacterial peptides with further determination whether they are active against gram-positive or gram-negative bacteria [153]. In the method, 20 amino acids were divided into four groups based on their physicochemical properties and assembled into motifs of k letters each. Then, the k-mers were used to construct four types of the parameter: a motif is present in a peptide; a motif is present in a certain position of a peptide; a motif is present in a certain position with a shift allowed; a motif is present in a certain position relative to another motif. Then, an evolutionary algorithm was used to calculate additional parameters derived from the previous ones by means of logical operators. The AUC of the algorithm was 0.95.

Many algorithms utilize not one but several parameter types simultaneously. For example, an algorithm proposed by Kumar and coauthors to determine hypotensive peptides uses physicochemical properties of peptides, as well as their amino acid and atomic composition [143]. Notably, the algorithm utilizes individual sets of parameters for peptides as a function of the peptide length. Cross-validation demonstrated that the algorithm has a sensitivity of 78.14% and specificity of 78.78%. Another algorithm, PeptideLocator, using bidirectional recurrent neural network allows finding amino acids potentially comprising bioactive peptides in a sequence of the protein analyzed [154]. To predict bioactive peptides in this manner, PeptideLocator calculates secondary structure, availability to solvent, structural motifs, and presence of disordered

structures, as well as predicting the domain structure of a protein using the Porter+ [155], IUPred [156], and SMART [157] services. A fivefold cross-validation demonstrated that the algorithm has a sensitivity of 82.33% and specificity of 83.50%.

Some authors build several algorithms based on different parameters and choose the most efficient. For example, developers of the AntiAngioPred [86] service predicting peptides inhibiting angiogenesis trained several algorithms operating the support vector machine using such parameters as amino acid composition, dipeptide composition, and binary profile patterns. The algorithm processed either the whole sequence of a peptide or a certain number of amino acids at its N and/or C end. In a tenfold validation, an algorithm based on amino acid composition at the peptide N-terminus demonstrated the best results. Its sensitivity and specificity were 79.0 and 82.7%, respectively.

In addition, special methods for parameter selection are used. For example, to select an optimal set of parameters among amino acid composition and physicochemical properties to predict cell penetrating peptides, Wei and coauthors used the maximum relevance–maximum distance methods [158]. To make a prediction, the algorithm sorts all parameters by their importance through maximization of importance evaluated using the Pearson coefficient between a parameter and bioactivity. Excessiveness is evaluated using the Euclidean distance between the parameters. Then, using sequential forward search, an optimal set of parameters is selected. Final algorithm utilizes 290 parameters and demonstrates sensitivity of 90.5% and specificity of 92.6%.

To calculate a parameter, the abovementioned properties of peptide sequences are to be derived. To do this, experiments or services for calculation or prediction of molecular characteristics are used. Examples of such services are listed in Table 2.

Design of algorithms to predict bioactive peptides. When parameters are chosen, they are used to build one of many machine learning algorithms, including support vector machines [196], random forests [158], C4.5 [139], naive Bayesian classifiers [86], the k nearest neighbor method [148], neural networks [146], and gradient or adaptive boosting [137] (Table 3). Classification algorithms can be used to predict the presence of bioactivity or choose bioactivity type. In a work by Lin and Xu, a random forest algorithm was used to predict whether a peptide is an antimicrobial one and if yes, whether it possesses antibacterial, antifungal, or antiviral activity [92].

Algorithms utilizing multiple instance learning deserve their own mention. This approach, for example, was used to develop an algorithm predicting MHC type II binding peptides [197]. In contrast to classic machine learning techniques, this method uses sets of parameters instead of individual characteristics. If only one of the parameters of the set belongs to a class, the whole set refers to the class. The peptide is considered a set, while parameters of the set are all 9- or 11-mer sequences built from amino acids comprising the peptide. The similarity of 9- and 11-mers is evaluated using the BLOSUM62 matrix. Some positions in the peptide k-mer, particularly the amino acids at positions 1, 4, 6, 7, and 9, are considered more important and they are used for classification. The support vector machine method is used for classification. Fivefold cross-validation demonstrated that the AUC of the algorithm was 0.777. One of the specific features of the algorithm is that it takes not final parameters, but the amino acid as an input. Algorithms based on N1-NN neural networks also need no preliminary processing of the peptide sequence [198, 199].

Table 3 provides algorithms predicting diverse biological functions of peptides created based on various methods, either relying on machine learning techniques or not.

MOLECULAR DOCKING-BASED METHODS

A separate class of methods is represented by methods of molecular modeling based on information on the spatial structure of molecules allowing prediction of the most probable orientation of a bioactive peptide and its protein target upon interaction. As a rule, these methods are used when the activity of a peptide has already been revealed or predicted and the question is how it works. On the other hand, when the molecular target is known, it is possible to analyze a large number of peptides and reveal those selectively binding the target. Therefore, molecular docking can be used for the purposes of bioactive peptide search.

In contrast to small molecules, for which the space of conformations is limited, the docking procedure for larger, and thus more flexible, peptides is more complicated since their conformation can change considerably upon interaction with protein. Hence, most algorithms elaborated for molecular docking are not applicable for peptides [212]. The algorithm for peptide bioactivity analysis by docking in general comprises the following steps: determination of the most probable conformation of a peptide, search for the peptide binding site in the protein (can be omitted if interaction with a known site is being investigated), and determination of the complex structure and evaluation of the energy of complex formation.

The simplest services allow prediction of a protein fragment that is involved in protein–peptide interaction. For example, the PepSite online service belongs to this type; it utilizes spatial position weight matrices calculated over the large collection of known complexes of peptides and proteins [213]. The service uses peptide amino acid sequence and protein spatial structure data to predict peptide-binding site on protein surface. Similar information can be obtained using the

Table 2. Services for analysis of amino acid sequences and calculation/prediction of certain properties

Source name	Link to the source	Functions
NCBI BLAST [62]	https://blast.ncbi.nlm.nih.gov/Blast.cgi	Local alignment
T-Coffee [63]	http://tcoffee.crg.cat/	Multiple alignment
Clustal [64]	http://www.clustal.org/	Multiple alignment
HMMER [65]	http://hmmer.org/	Weak homology search
Lalign [159]	http://embnet.vital-it.ch/software/LALIGN_form.html	Search for several matching motifs in two amino acid sequences
Peptide Match [160]	http://research.bioinformatics.udel.edu/peptidematch/index.jsp	Search for accurate/precise peptide inclusion in UniProtKB
CompariMotif [161]	http://bioware.ucd.ie/~compass/bioware-web/Server_pages/comparimotif.php	Search for similar regular expressions
AAindex db [141]	http://www.genome.jp/aaindex/	Physicochemical properties of amino acids
APDbase [162]	http://www.roskamps.com/bioinfo/APDbase/APDbase.php	Physicochemical properties of amino acids
PaDEL [142]	http://www.yapcwsoft.com/dd/padeldescriptor/	Physicochemical properties
PepCalc.com [163]	http://pepcalc.com/	Physicochemical properties
ProtParam [164]	http://web.expasy.org/protparam/	Physicochemical properties
Prot pi	https://www.protpi.ch/Calculator/PeptideTool	Physicochemical properties
acACS Web Server [165]	http://202.207.14.87:8032/fuwu/acacs/index.asp	Chemical shift
TheorChromo	http://theorchromo.ru/	HPLC retention time
IPC [166]	http://isoelectric.ovh.org/	Isoelectric point
SMART [157]	http://smart.embl-heidelberg.de/	Search for domains
SLiMSearch 1.5 [167]	http://bioware.ucd.ie/~compass/bioware-web/Server_pages/slimsearch.php	Search for known short linear motifs of protein–protein interactions
SLiMEnrich	http://shiny.slimsuite.unsw.edu.au/SLiMEnrich/	Prediction of short linear motifs of protein–protein interactions
SLiMFinder [168]	http://www.slimsuite.unsw.edu.au/servers/slimfinder.php	Prediction of short linear motifs of protein–protein interactions
SLiMPrints 3.0 [169]	http://bioware.ucd.ie/~compass/bioware-web/Server_pages/slimprints.php	Search for new overrepresented short linear motifs of protein–protein interactions
iELM [170]	http://elmint.embl.de/index/	Prediction of short linear motifs of protein–protein interactions
SLiMPred 0.9 [171]	http://bioware.ucd.ie/~compass/bioware-web/Server_pages/slimpred_legacy.php	Prediction of short linear motifs of protein–protein interaction
DiLiMot [172]	http://dilimot.russelllab.org/	Search for motifs
CD-HIT server [91]	http://weizhongli-lab.org/cd-hit/	Clusterization
GibbsCluster-2.0 [173]	http://www.cbs.dtu.dk/services/GibbsCluster/	Alignment and clusterization
MEME/MAST 4.9.1 [174]	http://meme-suite.org/	Search for motifs
AGGRESCAN [175]	http://bioinf.uab.es/aggrescan/	Aggregation
TANGO [176]	http://tango.crg.es/	Secondary structure, aggregation in solution
IUPRED [71]	http://iupred.enzim.hu/	Presence of a stable secondary structure
PEP-FOLD [177]	http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD/	Prediction of tertiary structure for peptides 5 to 50 a.a. long

Table 2. (Contd.)

Source name	Link to the source	Functions
PEPstrMOD [178]	http://osddlinux.osdd.net/raghava/pepstrmod/	Prediction of tertiary structure for peptides 7 to 25 a.a. long
PROSPER [179]	https://prosper.erc.monash.edu.au/	Protease cleavage sites
CleavPredict [180]	http://cleavpredict.sanfordburnham.org/	Matrix metalloprotease cleavage sites
CutDB [181]	http://cutdb.burnham.org/	Protease cleavage sites
PoPS [182]	http://pops.csse.monash.edu.au/	Protease cleavage sites
Enzyme Predictor [183]	http://bioware.ucd.ie/~enzpred/Enzpred.php	Prediction of proteases that yielded the hydrolyzate
PepControls	http://bioware.ucd.ie/~cyclops/Fergal/tags/PepControls_release_versions/1.3/control_pep_website/control_peptides.html	Prediction of control peptides
Phobius [184]	http://phobius.sbc.su.se/	Prediction of signaling peptides
SignalP 4.1 [185]	http://www.cbs.dtu.dk/services/SignalP/	Prediction of signal peptides in prokaryotic and eukaryotic proteins
sigcleave [186]	http://emboss.bioinformatics.nl/cgi-bin/emboss/sigcleave	Prediction of signal peptides in prokaryotic and eukaryotic proteins
Octopus [184]	http://octopus.cbr.su.se/	Prediction of signal peptides
PrediSi [187]	http://www.predisi.de/	Prediction of signal peptides in prokaryotic and eukaryotic proteins
TatP 1.0 [188]	http://www.cbs.dtu.dk/services/TatP/	Prediction of signal peptides
ChloroP 1.1 [189]	http://www.cbs.dtu.dk/services/ChloroP/	Prediction of chloroplast transit peptide
AACompSim [164]	http://web.expasy.org/aacompsim/	Comparison of amino acid composition vectors
Virtual Ribosome 2.0 [190]	http://www.cbs.dtu.dk/services/VirtualRibosome/	Translation of DNA sequence into peptide sequence, search for open reading frames, start, and stop codon
TargetP [191]	http://www.cbs.dtu.dk/services/TargetP/	Intracellular localization of eukaryotic proteins
Peptide Server [192]	http://bioserver-1.bioacademy.gr/Bioserver/PepServe/	Clusterization and visualization of peptides
Peptigram [193]	http://bioware.ucd.ie/peptigram/	Data visualization: peptide profile and alignment map
BlockLogo [194]	http://research4.dfci.harvard.edu/cvc/blocklogo/index.php	Analysis of consensus sequence motif
WebLogo 3 [195]	http://weblogo.threeplusone.com	Analysis of consensus sequence motif
Blast2logo 1.1	http://www.cbs.dtu.dk/biotools/Blast2logo-1.1/	Analysis of consensus sequence motif

PEP-SiteFinder algorithm, which predicts possible peptide conformations, then performs docking minimizing the binding energy, and evaluates the contribution of each of the amino acid residues into the interaction [214]. PEP-SiteFinder works much faster because it requires fewer calculations since molecules are viewed as rigid bodies.

Some algorithms require additional information for analysis. For example, the proABC service allows us to determine the antibody–antigen interaction site, in addition to the antibody sequence and spatial structure of the molecules, accepts the organism of the antibody origin and embryonic cell line, as well as the length of the antibody hypervariable loops as the input data [215].

The Peptiderive algorithm is also an interesting one; it allows the search for a linear peptide region of interaction in protein–protein complexes, which can be used to design a drug inhibiting the protein–protein interaction [216]. Input data include the structure of the receptor and interaction partner complex. A sliding frame of a preset size moves along the amino acid

Table 3. Algorithms to search for and predict bioactive peptides

Algorithm name	Link to online source, if available	Verified bioactivity	Algorithm	Parameters*	Efficiency of prediction**
Antimicrobial peptides					
AMPer [72]	http://www.cnbi2.com/cgi-bin/amp.pl	Antimicrobial peptides	Alignments	Motifs	Sn 82%, Sp 99.2%, Acc 99.2%
Torrent et al. [77]		Bioactive fragments of antimicrobial proteins	Disposition matrix	Statistical trends	Sn 90%, Sp 80%, Acc 85%
dPABBs [196]	http://ab-open-lab.csir.res.in/abp/antibiofilm/	Peptides destroying bacterial biofilms	Support vector machine	a.a. composition	Sn 95.24%, Sp 92.5%, Acc 97.73%, MCC 0.91
iAMPpred [145]	http://cab-grid.res.in:8080/amppred/	Antimicrobial peptides	Support vector machine	a.a. composition, physicochemical properties	Sn 90.75%, Sp 92.94%, Acc 91.84%, MCC 0.84
AntiBP2 [135]	http://www.imtech.res.in/raghava/antibp2/	Antibacterial peptides	Support vector machine	a.a. composition	Acc 87.55%
MLAMP [92]	http://www.jci-bio-info.cn/MLAMP	Antimicrobial peptides and their functional types	Random forest	a.a. composition	Sn 97.3%, Sp 92.1%, Acc 94.7%, MCC 0.9
Veltri et al. [153]		Antibacterial peptides divided into peptides active against gram-positive and gram-negative bacteria	Logistic regression, C4.5	Physicochemical properties	AUC 0.95
Torrent et al. [150]		Antimicrobial peptides	Neural networks	Physicochemical properties	Acc 90%
AVPpred [83]	http://crdd.osdd.net/servers/avppred/	Antivirus peptides	Support vector machine	Physicochemical properties	Sn 88.3%, Sp 82.2%, Acc 85.7%, MCC 0.71
iAMP-2L [200]	http://www.jci-bio-info.cn/iAMP-2L	Antimicrobial peptides and their functional types	k nearest neighbors	a.a. composition	Sn 87.13%, Sp 86.03%, Acc 86.32%, MCC 0.73
Anticancer peptides					
iACP [201]	http://lin.uestc.edu.cn/server/iACP	Anticancer peptides	Support vector machine	a.a. composition	Sn 93.33%, Sp 92%, Acc 92.67%, MCC 0.85
AntiCP [136]	http://crdd.osdd.net/raghava/anticp/	Antimicrobial peptides with anticancer properties	Support vector machine	a.a. composition	Acc 89%, MCC 0.78
AntiAngioPred [86]	http://clri.res.in/subramanian/tools/antiangiopred/index.html	Angiogenesis-inhibiting peptides	Support vector machine	a.a. composition	Sn 65%, Sp 85%, Acc 75%, MCC 0.51
Li and Wang [151]		Anticancer peptides	Support vector machine	a.a. composition, physicochemical properties	Sn 89.86%, Sp 96.12%, Acc 93.61%, MCC 0.87
ACPP [137]	http://acpp.bicpu.edu.in/predict.php	Anticancer peptides	Support vector machine	a.a. composition	Sn 95%, Sp 97%, Acc 96%, MCC 0.92
TumorHPD [202]	http://crdd.osdd.net/raghava/tumorhpd/	Tumor-migrating peptides	Support vector machine	a.a. composition	Sn 77.11%, Sp 90.36%, Acc 83.73%, MCC 0.68, AUC 0.94

Table 3. (Contd.)

Algorithm name	Link to online source, if available	Verified bioactivity	Algorithm	Parameters*	Efficiency of prediction**
Epitopes					
Meydan et al. [140]		Peptides binding MHC type I or II	Support vector machine	a.a. sequence	AUC 0.90 for MHC type I and 0.78 for MHC type II
MHC2MIL [197]	http://datamining-iip.fudan.edu.cn/service/MHC2MIL/index.html	Peptides binding MHC type II	Support vector machine	a.a. composition, physicochemical properties	AUC 0.78
NetMHCIIpan-3.0 [203]	http://www.cbs.dtu.dk/services/NetMHCIIpan-3.0/	Peptides binding MHC type II	k nearest neighbors	a.a. composition	AUC 0.87
NetMHCstabpan [204]	http://www.cbs.dtu.dk/services/NetMHCstabpan/	Stability of peptide binding with MHC type I	Neural network	a.a. composition	PCC 0.68
NetMHCpan-3.0 [205]	http://www.cbs.dtu.dk/services/NetMHCpan/	Peptides binding MHC type I	Neural network	a.a. composition	AUC 0.89, PCC 0.74
MHC-NP [206]	http://tools.immuneepitope.org/mhcnp/	Peptides binding MHC type I	Support vector machine	a.a. composition, physicochemical properties	AUC 0.86
Toxins					
ConoSorter [69]	https://sourceforge.net/projects/conosorter/	Conotoxins	Regular expressions, alignments	Motifs	Sn 99.9%
ToxinPred [87]	http://crdd.osdd.net/raghava/toxinpred/	Toxic peptides	Support vector machine	a.a. composition, motifs	Sn 99.39%, Sp 97.91%, Acc 98.41%, MCC 0.96
HemoPI [139]	http://crdd.osdd.net/raghava/hemopi/	Hemolytic peptides and their activity	Support vector machine	a.a. composition, motifs	Sn 96.4%, Sp 99.1%, Acc 96.4%, MCC 0.93
Cell-penetrating peptides					
C2Pred [207]	http://lin.uestc.edu.cn/server/C2Pred	Cell-penetrating peptides	Support vector machine	a.a. composition	Sn 97.3%, Sp 76.5%, Acc 92.4%
CellPPD [208]	http://crdd.osdd.net/raghava/cellppd/#thumb	Cell-penetrating peptides	Support vector machine	a.a. composition, motifs	Sn 98.15%, Sp 96.58%, Acc 97.4%, MCC 0.95, AUC 0.99
CPPred-RF [158]	http://server.malab.cn/CPPred-RF	Efficiency of cell penetration by peptides	Random forest	a.a. composition, physicochemical properties	Sn 90.5%, Sp 92.6%, Acc 91.6%, MCC 0.83
CPPred [199]	http://bioware.ucd.ie/~compass/bioware-web/Server_pages/cpppred.php	Efficiency of cell penetration by peptides	Neural network	a.a. sequences	Sn 68.09%, Sp 96.97%, MCC 0.69

Table 3. (Contd.)

Algorithm name	Link to online source, if available	Verified bioactivity	Algorithm	Parameters*	Efficiency of prediction**
Other bioactive peptides					
AHTpin [143]	http://crdd.osdd.net/raghava/ahtpin/	Hypotensive peptides	Support vector machine	a.a. composition, physicochemical properties	Sn 78.14%, Sp 78.78%, Acc 78.46%, MCC 0.57
FP_predict	https://sourceforge.net/projects/fptool/files/?source=navbar	Retrovirus fusion peptides	Hidden Markov model	a.a. sequence	Acc 91.97%
QSPpred [82]	http://crdd.osdd.net/servers/qsppred	Quorum sensing peptides	Support vector machine	a.a. composition	Acc 90%, MCC 0.8
PeptideLocator [154]	http://bioware.ucd.ie/~compass/biowareweb/Server_pages/biopred.php	Bioactive peptides as part of proteins	Neural networks	Motifs, physicochemical properties	Sn 82.33%, Sp 83.5%, Acc 85.8%, MCC 0.71
Peptide Ranker [198]	http://bioware.ucd.ie/~compass/biowareweb/Server_pages/peptideranker.php	Bioactive peptides	Neural network	a.a. sequence	MCC 0.74, AUC 0.94
AllerTOP v. 2.0 [209]	http://www.ddg-pharmfac.net/AllerTOP/	Allergens	k nearest neighbors	Physicochemical properties	Sn 86.7%, Sp 90.7%, Acc 88.7%, MCC 0.78
IL4pred [210]	http://crdd.osdd.net/raghava/il4pred/	Peptides stimulating IL4 production	Support vector machine	a.a. composition, motifs	Sn 78.76%, Sp 72.1%, Acc 75.76%, MCC 0.51
appnn [211]	https://cran.r-project.org/web/packages/appnn/index.html	Peptides stimulating amyloid fibrilla formation	Neural networks	Physicochemical properties	Sn 87.4%, Sp 78.9%, Acc 84.9%

* a.a., amino acid; ** Sn, sensitivity; Sp, specificity; Acc, accuracy; MCC, Mathew's correlation coefficient; AUC, area under the ROC-curve; PCC, Pearson's correlation coefficient.

sequence of the protein partner chain and evaluates the binding energy with each resulting peptide. The output of the algorithm is the peptide with the maximum binding energy. Then, the algorithm verifies whether the complex could be made even stronger by adding disulfide bonds in the peptide under study.

The most widely used methods perform the whole modeling procedure receiving the amino acid sequence of a peptide and protein spatial structure as an input and predicting the protein–peptide complex. An example of such a service is the CABS-dock [217]. The algorithm generates random spatial structures of a peptide that are randomly arranged on a spherical surface, the center of which coincides with the geometric center of a protein. Then, using the replica exchange method of the Monte Carlo algorithm, peptide binding to protein is modeled and 1000 complexes with minimal energy are selected and clusterized. Clusterization yields 10 consensus conformations that are reconstructed.

The spatial structure of a peptide can be generated from three specific conformations: unfolded, alpha helix-containing, and containing a polyproline helix. For example, this approach is used in the pepATTRACT protocol [218]. When the three possible conformations are created, the binding of each of them with the protein is modeled; the molecules are considered as rigid bodies. After 1000 of the most successful conformations have been chosen, the complex structure is optimized. When the algorithm was tested on 80 structures, approximately 70% of them were predicted correctly. A similar approach was used in the HADDOCK method [219]. If additional information about the peptide is available, for example, chemical shifts, intermolecular bonds, or mutagenesis, protein–peptide interactions are subjected to various structural limitations. After docking, the complex structure is refined considering the aqueous environment. Testing of the algorithm on 62 protein–peptide complexes correctly predicted 70% of structures.

Another way to predict peptide structure is to divide its sequence into segments. Analysis of short fragments proceeds faster and resembles the process of progressive binding between a protein with disordered tertiary structure and a protein having a stable spatial structure [220]. However, it also can be used for peptide studies. The protocol of peptide–protein interaction analysis [221] comprises the following steps: description of a peptide with a set of short fragments that have a similar amino acid sequence (short binding motifs); docking of fragments as rigid bodies; and structure optimization of the complex of fragments assembled into a peptide using the Rosetta Flex-PepDock [212] algorithm which can consider conformational changes of the receptor and peptide. The method defined 52% of structures in the course of testing on 27 complexes.

In addition, the spatial structure of a peptide can be determined by similarity with an existing structure, as in the GalaxyTBM algorithm [222]. At the first stage of the algorithm, a template complex is searched for in the PepBind database [223] based on the homology of protein structure and peptide amino acid sequence. At the next stage, a model minimizing the interaction energy is built. The algorithm was tested on 22 complexes and predicted correctly 17 of them.

Molecular docking plays an important role in the analysis of various epitopes, including the MHC type II and antibodies. NetMHCIIpan-3.0 takes peptide amino acid sequences as the input and uses a neural network on pseudosequences of MHC type II to predict protein–peptide complexes [224]. A pseudosequence contains amino acids important for binding that have been determined by alignment of tertiary structures of the molecules. The method's AUC is 0.847. Another algorithm uses the docking score-based quantitative matrices for prediction. The matrix is filled beforehand based on the analysis of various peptides binding with MHC type II [225]. The sensitivity of the algorithm is 70%.

CONCLUSIONS

The study of bioactive peptides is of interest both for searching for new drugs and for determining the action of known bioactive peptides in the human body. To analyze biological functions of peptides, methods of bioinformatics are being elaborated to make the process more efficient. Algorithms determining potential bioactive peptides can be developed using machine learning techniques, however, the basis is mainly provided by the analysis of several peptide characteristics: amino acid composition, physicochemical properties, and functional motifs. Most algorithms allow searching for peptides performing one specific function, although there are methods capable of determination of peptide bioactivity as such and, further, more precise prediction of the bioactivity type.

Some algorithms developed to predict certain functions in practice are more universal than assumed by design [198], which in future will probably allow algorithms can alone determine a wide range of functional activities. One of the important factors limiting the development of bioactive peptide prediction is the lack of information about known bioactive peptides. Analyzing the trends in the development of protein-function prediction algorithms, one may suppose that, as data accumulate, methods based on the analysis of the spatial structure of peptides and proteins as they interact will evolve greatly. On the other hand, algorithms allowing analysis based only on the amino acid sequence of peptides are simpler to use and probably will gain even broader popularity.

ACKNOWLEDGMENTS

We are grateful to Olga Ivanova (Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences) for valuable comments. The work was supported by the Russian Science Foundation (project no. 14-50-00131).

REFERENCES

1. Aebersold, R. and Mann, M., *Nature*, 2016, vol. 537, pp. 347–355.
2. Schrader, M., Schulz-Knappe, P., and Fricker, L.D., *EuPA Open Proteomics*, 2014, vol. 3, pp. 171–182.
3. Dallas, D.C., Guerrero, A., Parker, E.A., Robinson, R.C., Gan, J., German, J.B., Barile, D., and Lebrilla, C.B., *Proteomics*, 2015, vol. 15, pp. 1026–1038.
4. Ivanov, V.T. and Yatskin, O.N., *Expert Rev. Proteomics*, 2005, vol. 2, pp. 463–473.
5. Mann, M., Kulak, N.A., Nagaraj, N., and Cox, J., *Mol. Cell*, 2013, vol. 49, pp. 583–590.
6. Mann, M., *Clin. Chem.*, 2016, vol. 62, pp. 293–294.
7. Fitzgerald, R.J. and Murray, B.A., *Int. J. Dairy Technol.*, 2006, vol. 59, pp. 118–125.
8. Schulz-Knappe, P., Mägert, H.J., Dewald, B., Meyer, M., Cetin, Y., Kubbies, M., Tomeczkowski, J., Kirchhoff, K., Raida, M., Adermann, K., et al., *J. Exp. Med.*, 1996, vol. 183, pp. 295–299.
9. Hernández, E., Benavente, F., Sanz-Nebot, V., and Barbosa, J., *Electrophoresis*, 2008, vol. 29, pp. 3366–3376.
10. Petricoin, E.F., Belluco, C., Araujo, R.P., and Liotta, L.A., *Nat. Rev. Cancer*, 2006, vol. 6, pp. 961–967.
11. Ziganshin, R., Arapidi, G., Azarkin, I., Zaryadieva, E., Alexeev, D., Govorun, V., and Ivanov, V., *J. Proteomics*, 2011, vol. 74, pp. 595–606.
12. Yatskin, O.N., Karelin, A.A., and Ivanov, V.T., *Russ. J. Bioorg. Chem.*, 2009, vol. 35, pp. 426–436.
13. Fesenko, I.A., Arapidi, G.P., Skripnikov, A.Y., Alexeev, D.G., Kostryukova, E.S., Manolov, A.I., Altukhov, I.A., Khazigaleeva, R.A., Seredina, A.V., Kovalchuk, S.I., et al., *BMC Plant Biol.*, 2015, vol. 15, pp. 2–16.
14. Slavoff, S.A., Mitchell, A.J., Schwaid, A.G., Cabili, M.N., Ma, J., Levin, J.Z., Karger, A.D., Budnik, B.A., Rinn, J.L., and Saghatelian, A., *Nat. Chem. Biol.*, 2013, vol. 9, pp. 59–64.
15. Andrews, S.J. and Rothnagel, J.A., *Nat. Rev. Genet.*, 2014, vol. 15, pp. 193–204.
16. Matsumoto, A., Pasut, A., Matsumoto, M., Yamashita, R., Fung, J., Monteleone, E., Saghatelian, A., Nakayama, K.I., Clohessy, J.G., and Pandolfi, P.P., *Nature*, 2017, vol. 541, pp. 228–232.
17. Aito-Inoue, M., Ohtsuki, K., Nakamura, Y., Park, E.Y., Iwai, K., Morimatsu, F., and Sato, K., *J. Agric. Food Chem.*, 2006, vol. 54, pp. 5261–5266.
18. Mohanty, D.P., Mohapatra, S., Misra, S., and Sahu, P.S., *Saudi J. Biol. Sci.*, 2016, vol. 23, pp. 577–583.
19. Pritchard, S.R., Phillips, M., and Kailasapathy, K., *Food Res. Int.*, 2010, vol. 43, pp. 1545–1548.
20. Sforza, S., Cavatorta, V., Lambertini, F., Galaverna, G., Dossena, A., and Marchelli, R., *J. Dairy Sci.*, 2012, vol. 95, pp. 3514–3526.
21. Phelan, M., Aherne, A., Fitzgerald, R.J., and O'Brien, N.M., *Int. Dairy J.*, 2009, vol. 19, pp. 643–654.
22. Ryder, K., Bekhit, A.E.-D., McConnell, M., and Carne, A., *Food Chem.*, 2016, vol. 208, pp. 42–50.
23. Lassoued, I., Mora, L., Barkia, A., Aristoy, M.-C., Nasri, M., and Toldra, F., *J. Proteomics*, 2015, vol. 128, pp. 8–17.
24. Kumagai, H., in *Bioactive Proteins and Peptides as Functional Foods and Nutraceuticals*, Wiley–Blackwell, 2010, pp. 289–303.
25. Singh, B.P., Vij, S., and Hati, S., *Peptides*, 2014, vol. 54, pp. 171–179.
26. Selamassakul, O., Laohakunjit, N., Kerdchoechuen, O., and Ratanakhanokchai, K., *Food Funct.*, 2016, vol. 7, pp. 2635–2644.
27. Silva-Sánchez, C., de la Rosa, A.P.B., León-Galván, M.F., de Lumen, B.O., de León-Rodríguez, A., and de Mejía, E.G., *J. Agric. Food Chem.*, 2008, vol. 56, pp. 1233–1240.
28. Falanga, A., Lombardi, L., Franci, G., Vitiello, M., Iovene, M.R., Morelli, G., Galdiero, M., and Galdiero, S., *Int. J. Mol. Sci.*, 2016, vol. 17, pp. 3–18.
29. Fan, X., Bai, L., Zhu, L., Yang, L., and Zhang, X., *J. Agric. Food Chem.*, 2014, vol. 62, pp. 9211–9222.
30. Sharma, S., Singh, R., and Rana, S., *Int. J. Bioautomation*, 2011, vol. 15, pp. 223–250.
31. Gibbs, B.F., Zougman, A., Masse, R., and Mulligan, C., *Food Res. Int.*, 2004/3, vol. 37, pp. 123–131.
32. Fitzgerald, C., Aluko, R.E., Hossain, M., Rai, D.K., and Hayes, M., *J. Agric. Food Chem.*, 2014, vol. 62, pp. 8352–8356.
33. Yamamoto, N., Ejiri, M., and Mizuno, S., *Curr. Pharm. Des.*, 2003, vol. 9, pp. 1345–1355.
34. Meisel, H. and Fitzgerald, R.J., *Curr. Pharm. Des.*, 2003, vol. 9, pp. 1289–1295.
35. Deshpande, L.M., Fritsche, T.R., and Jones, R.N., *Diagn. Microbiol. Infect. Dis.*, 2004, vol. 49, pp. 231–236.
36. Santiago-López, L., Hernández-Mendoza, A., Vallejo-Cordoba, B., Mata-Haro, V., and González-Córdova, A.F., *J. Sci. Food Agric.*, 2016, vol. 96, pp. 3631–3641.
37. Korhonen, H. and Pihlanto, A., *Curr. Pharm. Des.*, 2003, vol. 9, pp. 1297–1308.
38. Wang, Z. and Zhang, X., *J. Sci. Food Agric.*, 2017, vol. 97, pp. 918–922.
39. Shindo, Y., McDonough, J.S., Chang, K.C., Ramachandra, M., Sasikumar, P.G., and Hotchkiss, R.S., *J. Surg. Res.*, 2017, vol. 208, pp. 33–39.
40. Riedl, S.J. and Pasquale, E.B., *Curr. Drug Targets*, 2015, vol. 16, pp. 1031–1047.
41. Sumi, C.D., Yang, B.W., Yeo, I.-C., and Hahm, Y.T., *Can. J. Microbiol.*, 2015, vol. 61, pp. 93–103.
42. Koyama, M., Hattori, S., Amano, Y., Watanabe, M., and Nakamura, K., *PLoS One*, 2014, vol. 9, pp. 1–8.

43. Boles, K.S., Kannan, K., Gill, J., Felderman, M., Gouvis, H., Hubby, B., Kamrud, K.I., Venter, J.C., and Gibson, D.G., *Nat. Biotechnol.*, 2017, vol. 35, pp. 672–675.
44. Negi, V.S. and Borthakur, D., *Methods Mol. Biol.*, 2016, vol. 1405, pp. 59–77.
45. Miller, J. and Stagljar, I., in *Protein–Protein Interactions*, New Jersey: Humana Press, 247–262
46. Sekar, R.B. and Periasamy, A., *J. Cell Biol.*, 2003, vol. 160, pp. 629–633.
47. Yaciuk, P., *Methods Mol. Med.*, 2007, vol. 131, pp. 103–111.
48. Piehler, J., *Curr. Opin. Struct. Biol.*, 2005, vol. 15, no. 1 (spec. iss.), pp. 4–14.
49. Kihara, D., *Methods*, 2016, vol. 93, pp. 1–2.
50. Kumar, C. and Choudhary, A., *EURASIP J. Bioinform. Syst. Biol.*, 2012, vol. 2012, pp. 1–14.
51. Sael, L., Chitale, M., and Kihara, D., *J. Struct. Funct. Genomics*, 2012, vol. 13, pp. 111–123.
52. Tiwari, A.K. and Srivastava, R., *Int. J. Proteomics*, 2014, vol. 2014, pp. 1–22.
53. Sahraeian, S.M., Luo, K.R., and Brenner, S.E., *Nucleic Acids Res.*, 2015, vol. 43, pp. W141–W147.
54. Cao, R. and Cheng, J., *Methods*, 2016, vol. 93, pp. 84–91.
55. Maheshwari, S. and Brylinski, M., *Methods*, 2016, vol. 93, pp. 64–71.
56. Nakamura, T. and Tomii, K., *Methods*, 2016, vol. 93, pp. 35–40.
57. Joshi, T. and Xu, D., BMC Genomics, 2007, vol. 8, pp. 1–10.
58. Bailey, T.L., *Methods Mol. Biol.*, 2008, vol. 452, pp. 231–251.
59. Davey, N.E., Van Roey, K., Weatheritt, R.J., Toedt, G., Uyar, B., Altenberg, B., Budd, A., Diella, F., Dinkel, H., and Gibson, T.J., *Mol. Biosyst.*, 2012, vol. 8, pp. 268–281.
60. Pfleger, C.M. and Kirschner, M.W., *Genes Dev.*, 2000, vol. 14, pp. 655–665.
61. Lucchese, G., Stufano, A., Trost, B., Kusalik, A., and Kanduc, D., *Amino Acids*, 2007, vol. 33, pp. 703–707.
62. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L., *BMC Bioinf.*, 2009, vol. 10, pp. 1–9.
63. Notredame, C., Higgins, D.G., and Heringa, J., *J. Mol. Biol.*, 2000, vol. 302, pp. 205–217.
64. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., et al., *Mol. Syst. Biol.*, 2011, vol. 7, pp. 1–6.
65. Eddy, S.R., *Bioinformatics*, 1998, vol. 14, pp. 755–763.
66. Christie, A.E., *Gen. Comp. Endocrinol.*, 2014, vol. 201, pp. 87–106.
67. Montoya-Rodríguez, A., Gómez-Favela, M.A., Reyes-Moreno, C., Milán-Carrillo, J., and González de Mejía, E., *Compr. Rev. Food Sci. Food Saf.*, 2015, vol. 14, pp. 139–158.
68. Darewicz, M., Borawska, J., and Pliszka, M., *Czech J. Food Sci.*, 2016, vol. 34, pp. 111–117.
69. Lavergne, V., Dutertre, S., Jin, A.-H., Lewis, R.J., Taft, R.J., and Alewood, P.F., *BMC Genomics*, 2013, vol. 14, pp. 1–12.
70. Michael, S., Travé, G., Ramu, C., Chica, C., and Gibson, T.J., *Bioinformatics*, 2008, vol. 24, pp. 453–457.
71. Dosztányi, Z., Csizmók, V., Tompa, P., and Simon, I., *J. Mol. Biol.*, 2005, vol. 347, pp. 827–839.
72. Fjell, C.D., Hancock, R.E.W., and Cherkasov, A., *Bioinformatics*, 2007, vol. 23, pp. 1148–1155.
73. Liu, F., Baggerman, G., D'Hertog, W., Verleyen, P., Schoofs, L., and Wets, G., *Mol. Cell. Proteomics*, 2006, vol. 5, pp. 510–522.
74. Edwards, R.J., Moran, N., Devocelle, M., Kiernan, A., Meade, G., Signac, W., Foy, M., Park, S.D.E., Dunne, E., Kenny, D., et al., *Nat. Chem. Biol.*, 2007, vol. 3, pp. 108–112.
75. Li, Y. and Yu, J., *J. Med. Food*, 2015, vol. 18, pp. 147–156.
76. Pimenta, A.M.C. and De Lima, M.E., *J. Pept. Sci.*, 2005, vol. 11, pp. 670–676.
77. Torrent, M., Nogués, V.M., and Boix, E., *BMC Bioinf.*, 2009, vol. 10, pp. 1–9.
78. Wu, M. and Hancock, R.E., *Antimicrob. Agents Chemother.*, 1999, vol. 43, pp. 1274–1276.
79. Defernez, M. and Kemsley, E.K., *Analyst*, 1999, vol. 124, pp. 1675–1681.
80. Way, T.W., Sahiner, B., Hadjiiski, L.M., and Chan, H.-P., *Med. Phys.*, 2010, vol. 37, pp. 907–920.
81. Wynendaele, E., Bronselaer, A., Nielandt, J., D'Hondt, M., Stalmans, S., Bracke, N., Verbeke, F., Van De Wiele, C., De Tré, G., and De Spiegeleer, B., *Nucleic Acids Res.*, 2013, vol. 41, pp. D655–D659.
82. Rajput, A., Gupta, A.K., and Kumar, M., *PLoS One*, 2015, vol. 10, pp. 1–16.
83. Thakur, N., Qureshi, A., and Kumar, M., *Nucleic Acids Res.*, 2012, vol. 40, pp. 199–204.
84. Gautam, A., Chaudhary, K., Singh, S., Joshi, A., Anand, P., Tuknait, A., Mathur, D., Varshney, G.C., and Raghava, G.P.S., *Nucleic Acids Res.*, 2014, vol. 42, pp. D444–D449.
85. Smialowski, P., Pagel, P., Wong, P., Brauner, B., Dunger, I., Fobo, G., Frishman, G., Montrone, C., Rattei, T., Frishman, D., et al., *Nucleic Acids Res.*, 2010, vol. 38, pp. D540–D544.
86. Ramaprasad, A.S.E., Singh, S., Gajendra, R.P.S., and Venkatesan, S., *PLoS One*, 2015, vol. 10, pp. 7–12.
87. Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., and Raghava, G.P.S., *PLoS One*, 2013, vol. 8, pp. 1–10.
88. Wang, P., Hu, L., Liu, G., Jiang, N., Chen, X., Xu, J., Zheng, W., Li, L., Tan, M., Chen, Z., et al., *PLoS One*, 2011, vol. 6, pp. 1–9.
89. Bals, R., *Respir. Res.*, 2000, vol. 1, pp. 141–150.
90. Chen, W., Ding, H., Feng, P., Lin, H., and Chou, K.-C., *Oncotarget*, 2016, vol. 7, pp. 16895–16909.
91. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W., *Bioinformatics*, 2012, vol. 28, pp. 3150–3152.
92. Lin, W. and Xu, D., *Bioinformatics*, 2016, vol. 32, no. 24, pp. 3745–3752.

93. Wang, G., Li, X., and Wang, Z., *Nucleic Acids Res.*, 2016, vol. 44, pp. D1087–D1093.
94. Mehta, D., Anand, P., Kumar, V., Joshi, A., Mathur, D., Singh, S., Tuknait, A., Chaudhary, K., Gautam, S.K., Gautam, A., et al., *Database*, 2014, vol. 2014, pp. 1–7.
95. Waghu, F.H., Barai, R.S., Gurung, P., and Idicula-Thomas, S., *Nucleic Acids Res.*, 2016, vol. 44, pp. D1094–D1097.
96. Gogoladze, G., Grigolava, M., Vishnepolsky, B., Chubinidze, M., Duroux, P., Lefranc, M.-P., and Pirtskhalava, M., *FEMS Microbiol. Lett.*, 2014, vol. 357, pp. 63–68.
97. Qureshi, A., Thakur, N., Tandon, H., and Kumar, M., *Nucleic Acids Res.*, 2014, vol. 42, pp. D1147–D1153.
98. Sundararajan, V., Gabere, M.N., Pretorius, A., Adam, S., Christoffels, A., Lehvaslaiho, M., Archer, J.A.C., and Bajic, V.B., *Nucleic Acids Res.*, 2012, vol. 40, pp. D1108–D1112.
99. Hammami, R., Ben Hamida, J., Vergoten, G., and Fliss, I., *Nucleic Acids Res.*, 2009, vol. 37, pp. D963–D968.
100. Hammami, R., Zouhir, A., Le Lay, C., Ben Hamida, J., and Fliss, I., *BMC Microbiol.*, 2010, vol. 10, pp. 1–5.
101. de Jong, A., van Hijum, S.A.F.T., Bijlsma, J.J.E., Kok, J., and Kuipers, O.P., *Nucleic Acids Res.*, 2006, vol. 34, pp. W273–W279.
102. Seebah, S., Suresh, A., Zhuo, S., Choong, Y.H., Chua, H., Chuon, D., Beuerman, R., and Verma, C., *Nucleic Acids Res.*, 2007, vol. 35, pp. D265–D268.
103. Novković, M., Simunić, J., Bojović, V., Tossi, A., and Juretić, D., *Bioinformatics*, 2012, vol. 28, pp. 1406–1407.
104. Di Luca, M., Maccari, G., Maisetta, G., and Batoni, G., *Biofouling*, 2015, vol. 31, pp. 193–199.
105. Théolier, J., Fliss, I., Jean, J., and Hammami, R., *Dairy Sci. Technol.*, 2014, vol. 94, pp. 181–193.
106. Piotto, S.P., Sessa, L., Concilio, S., and Iannelli, P., *Int. J. Antimicrob. Agents*, 2012, vol. 39, pp. 346–351.
107. Tyagi, A., Tuknait, A., Anand, P., Gupta, S., Sharma, M., Mathur, D., Joshi, A., Singh, S., Gautam, A., and Raghava, G.P.S., *Nucleic Acids Res.*, 2015, vol. 43, pp. D837–D843.
108. Kapoor, P., Singh, H., Gautam, A., Chaudhary, K., Kumar, R., and Raghava, G.P.S., *PLoS One,* 2012, vol. 7, pp. 1–6.
109. Vita, R., Zarebski, L., Greenbaum, J.A., Emami, H., Hoof, I., Salimi, N., Damle, R., Sette, A., and Peters, B., *Nucleic Acids Res.*, 2010, vol. 38, pp. D854–D862.
110. Zhang, G.L., Riemer, A.B., Keskin, D.B., Chitkushev, L., Reinherz, E.L., and Brusic, V., *Database*, 2014, vol. 2014, pp. 1–12.
111. Kulkarni-Kale, U., Raskar-Renuse, S., Natekar-Kalantre, G., and Saxena, S.A., *Methods Mol. Biol.*, 2014, vol. 1184, pp. 149–164.
112. Ivanciuc, O., Schein, C.H., and Braun, W., *Nucleic Acids Res.*, 2003, vol. 31, pp. 359–362.
113. Mari, A., Rasi, C., Palazzo, P., and Scala, E., *Curr. Allergy Asthma Rep.*, 2009, vol. 9, pp. 376–383.
114. Yang, B., Sayers, S., Xiang, Z., and He, Y., *Nucleic Acids Res.*, 2011, vol. 39, pp. D1073–D1078.
115. Nijveen, H., Kester, M.G.D., Hassan, C., Viars, A., de Ru, A.H., de Jager, M., Falkenburg, J.H.F., Leunissen, J.A.M., and van Veelen, P.A., *Immunogenetics*, 2011, vol. 63, pp. 143–153.
116. Kumar, R., Chaudhary, K., Sharma, M., Nagpal, G., Chauhan, J.S., Singh, S., Gautam, A., and Raghava, G.P.S., *Nucleic Acids Res.*, 2015, vol. 43, pp. D956–D962.
117. Herzig, V., Wood, D.L.A., Newell, F., Chaumeil, P.-A., Kaas, Q., Binford, G.J., Nicholson, G.M., Gorse, D., and King, G.F., *Nucleic Acids Res.*, 2011, vol. 39, pp. D653–D657.
118. He, Q.-Y., He, Q.-Z., Deng, X.-C., Yao, L., Meng, E., Liu, Z.-H., and Liang, S.-P., *Nucleic Acids Res.*, 2008, vol. 36, pp. D293–D297.
119. Chakraborty, A., Ghosh, S., Chowdhary, G., Maulik, U., and Chakrabarti, S., *Nucleic Acids Res.*, 2012, vol. 40, pp. D615–D620.
120. Kaas, Q., Yu, R., Jin, A.-H., Dutertre, S., and Craik, D.J., *Nucleic Acids Res.*, 2012, vol. 40, pp. D325–D330.
121. Agrawal, P., Bhalla, S., Usmani, S.S., Singh, S., Chaudhary, K., Raghava, G.P.S., and Gautam, A., *Nucleic Acids Res.*, 2016, vol. 44, pp. D1098–D1103.
122. Shtatland, T., Guettler, D., Kossodo, M., Pivovarov, M., and Weissleder, R., *BMC Bioinf.*, 2007, vol. 8, pp. 1–10.
123. Deutsch, E.W., *Methods Mol. Biol.*, 2010, vol. 604, pp. 285–296.
124. Minkiewicz, P., Dziuba, J., Iwaniak, A., Dziuba, M., and Darewicz, M., J. *AOAC Int.*, 2008, vol. 91, pp. 965–980.
125. Fälth, M., Sköld, K., Norrman, M., Svensson, M., Fenyö, D., and Andren, P.E., *Mol. Cell. Proteomics*, 2006, vol. 5, pp. 998–1005.
126. Beerten, J., Van Durme, J., Gallardo, R., Capriotti, E., Serpell, L., Rousseau, F., and Schymkowitz, J., *Bioinformatics*, 2015, vol. 31, pp. 1698–1700.
127. Dinkel, H., Van Roey, K., Michael, S., Kumar, M., Uyar, B., Altenberg, B., Milchevskaya, V., Schneider, M., Kühn, H., Behrendt, A., et al., *Nucleic Acids Res.*, 2016, vol. 44, pp. D294–D300.
128. He, B., Chai, G., Duan, Y., Yan, Z., Qiu, L., Zhang, H., Liu, Z., He, Q., Han, K., Ru, B., et al., *Nucleic Acids Res.*, 2016, vol. 44, pp. D1127–D1132.
129. Choo, K.H., Tan, T.W., and Ranganathan, S., *BMC Bioinf.*, 2005, vol. 6, pp. 1–8.
130. Nielsen, S.D., Beverly, R.L., Qu, Y., and Dallas, D.C., *Food Chem.*, 2017, vol. 232, pp. 673–682.
131. Van Dorpe, S., Bronselaer, A., Nielandt, J., Stalmans, S., Wynendaele, E., Audenaert, K., Van De Wiele, C., Burvenich, C., Peremans, K., Hsuchou, H., et al., *Brain Struct. Funct.*, 2012, vol. 217, pp. 687–718.
132. Zamyatnin, A.A., Borchikov, A.S., Vladimirov, M.G., and Voronina, O.L., *Nucleic Acids Res.*, 2006, vol. 34, pp. D261–D266.
133. Rashid, M., Singla, D., Sharma, A., Kumar, M., and Raghava, G.P.S., *BMC Genomics*, 2009, vol. 10, pp. 1–10.
134. Blanco-Míguez, A., Gutiérrez-Jácome, A., Fdez-Riverola, F., Lourenco, A., and Sánchez, B., *Database*, 2017, vol. 2017, pp. 1–10.
135. Lata, S., Mishra, N.K., and Raghava, G.P.S., *BMC Bioinf.*, 2010, vol. 11, suppl. 1, pp. 1–7.

136. Tyagi, A., Kapoor, P., Kumar, R., Chaudhary, K., Gautam, A., and Raghava, G.P.S., *Sci. Rep.*, 2013, vol. 3, pp. 1–8.

137. Vijayakumar, S. and Ptv, L., *Int. J. Pept. Res. Ther.*, 2015, vol. 21, pp. 99–106.

138. Gupta, S., Sharma, A.K., Shastri, V., Madhu, M.K., and Sharma, V.K., *J. Transl. Med.*, 2017, vol. 15, pp. 1–11.

139. Chaudhary, K., Kumar, R., Singh, S., Tuknait, A., Gautam, A., Mathur, D., Anand, P., Varshney, G.C., and Raghava, G.P.S., *Sci. Rep.*, 2016, vol. 6, pp. 1–13.

140. Meydan, C., Otu, H.H., and Sezerman, O.U., *BMC Bioinf.*, 2013, vol. 14, suppl. 2, pp. 1-12.

141. Kawashima, S., Pokarowski, P., Pokarowska, M., Kolinski, A., Katayama, T., and Kanehisa, M., *Nucleic Acids Res.*, 2008, vol. 36, database issue, pp. D202–D205.

142. Yap, C.W., *J. Comput. Chem.*, 2011, vol. 32, pp. 1466–1474.

143. Kumar, R., Chaudhary, K., Singh, ChauhanJ., Nagpal, G., Kumar, R., Sharma, M., and Raghava, G.P.S., *Sci. Rep.*, 2015, vol. 5, pp. 1–10.

144. Osorio, D., Rondón-Villarreal, P., and Torres, R., R J., 2015, vol. 7, pp. 4–14.

145. Meher, P.K., Sahu, T.K., Saini, V., and Rao, A.R., *Sci. Rep.*, 2017, vol. 7, pp. 1–12.

146. Andreu, D. and Torrent, M., *Methods Mol. Biol.*, 2015, vol. 1260, pp. 101–118.

147. Venkatarajan, M.S. and Braun, W., *J. Mol. Model.*, 2001, vol. 7, pp. 445–453.

148. Dimitrov, I., Bangov, I., Flower, D.R., and Doytchinova, I., *J. Mol. Model.*, 2014, vol. 20, pp. 2–6.

149. Yousefinejad, S., Hemmateenejad, B., and Mehdipour, A.R., *J. Iran. Chem. Soc.*, 2012, vol. 9, pp. 569–577.

150. Torrent, M., Andreu, D., Nogués, V.M., and Boix, E., *PLoS One*, 2011, vol. 6, pp. 1–8.

151. Li, F.-M. and Wang, X.-Q., *Sci. Rep.*, 2016, vol. 6, pp. 1–6.

152. Zhao, Y., Alipanahi, B., Li, S.C., and Li, M., *J. Bioinform. Comput. Biol.*, 2010, vol. 8, pp. 867–884.

153. Veltri, D., Kamath, U., and Shehu, A., *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 2017, vol. 14, pp. 300–313.

154. Mooney, C., Haslam, N.J., Holton, T.A., Pollastri, G., and Shields, D.C., *Bioinformatics*, 2013, vol. 29, pp. 1120–1126.

155. Mooney, C., Vullo, A., and Pollastri, G., *J. Comput. Biol.*, 2006, vol. 13, pp. 1489–1502.

156. Dosztányi, Z., Csizmok, V., Tompa, P., and Simon, I., *Bioinformatics*, 2005, vol. 21, pp. 3433–3434.

157. Letunic, I., Doerks, T., and Bork, P., *Nucleic Acids Res.*, 2009, vol. 37, pp. D229–D232.

158. Wei, L., Xing, P., Su, R., Shi, G., Ma, Z.S., and Zou, Q., *J. Proteome Res.*, 2017, vol. 16, pp. 2044–2053.

159. Huang, X., *Adv. Appl. Math.*, 1991, vol. 12, pp. 337–357.

160. Chen, C., Li, Z., Huang, H., Suzek, B.E., Wu, C.H., and UniProt Consortium, *Bioinformatics*, 2013, vol. 29, pp. 2808–2809.

161. Edwards, R.J., Davey, N.E., and Shields, D.C., *Bioinformatics*, 2008, vol. 24, pp. 1307–1309.

162. Mathura, V.S. and Kolippakkam, D., *Bioinformation*, 2005, vol. 1, pp. 2–4.

163. Lear, S. and Cobb, S.L., *J. Comput. Aided Mol. Des.*, 2016, vol. 30, pp. 271–277.

164. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.E., Wilkins, M.R., Appel, R.D., and Bairoch, A., in *The Proteomics Protocols Handbook*, Walker, J.M., Ed., Humana Press, 2005, pp. 571–607.

165. Fan, G.-L., Liu, Y.-L., Zuo, Y.-C., Mei, H.-X., Rang, Y., Hou, B.-Y., and Zhao, Y., *Sci. World J.*, 2014, vol. 2014, pp. 1–9.

166. Kozlowski, L.P., *Biol. Direct*, 2016, vol. 11, pp. 1–16.

167. Davey, N., Haslam, N., Shields, D., and Edwards, R., *Pattern Recognit. Bioinform.*, 2010, pp. 50–61.

168. Davey, N.E., Haslam, N.J., Shields, D.C., and Edwards, R.J., *Nucleic Acids Res.*, 2010, vol. 38, pp. W534–W539.

169. Davey, N.E., Cowan, J.L., Shields, D.C., Gibson, T.J., Coldwell, M.J., and Edwards, R.J., *Nucleic Acids Res.*, 2012, vol. 40, pp. 10628–10641.

170. Weatheritt, R.J., Jehl, P., Dinkel, H., and Gibson, T.J., *Nucleic Acids Res.*, 2012, vol. 40, pp. W364–W369.

171. Mooney, C., Pollastri, G., Shields, D.C., and Haslam, N.J., *J. Mol. Biol.*, 2012, vol. 415, pp. 193–204.

172. Neduva, V., Linding, R., Su-Angrand, I., Stark, A., de Masi, F., Gibson, T.J., Lewis, J., Serrano, L., and Russell, R.B., *PLoS Biol.*, 2005, vol. 3, pp. 2090–2099.

173. Andreatta, M., Lund, O., and Nielsen, M., *Bioinformatics*, 2013, vol. 29, pp. 8–14.

174. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S., *Nucleic Acids Res.*, 2009, vol. 37, pp. W202–W208.

175. Conchillo-Solé, O., de Groot, N.S., Avilés, F.X., Vendrell, J., Daura, X., and Ventura, S., *BMC Bioinf.*, 2007, vol. 8, pp. 1–17.

176. Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J., and Serrano, L., *Nat. Biotechnol.*, 2004, vol. 22, pp. 1302–1306.

177. Shen, Y., Maupetit, J., Derreumaux, P., and Tufféry, P., *J. Chem. Theory Comput.*, 2014, vol. 10, pp. 4745–4758.

178. Singh, S., Singh, H., Tuknait, A., Chaudhary, K., Singh, B., Kumaran, S., and Raghava, G.P.S., *Biol. Direct*, 2015, vol. 10, pp. 1–19.

179. Song, J., Tan, H., Perry, A.J., Akutsu, T., Webb, G.I., Whisstock, J.C., and Pike, R.N., *PLoS One*, 2012, vol. 7, pp. 1–23.

180. Kumar, S., Ratnikov, B.I., Kazanov, M.D., Smith, J.W., and Cieplak, P., *PLoS One*, 2015, vol. 10, pp. 1–19.

181. Igarashi, Y., Eroshkin, A., Gramatikova, S., Gramatikoff, K., Zhang, Y., Smith, J.W., Osterman, A.L., and Godzik, A., *Nucleic Acids Res.*, 2007, vol. 35, database issue, pp. D546–D549.

182. Boyd, S.E., Pike, R.N., Rudy, G.B., Whisstock, J.C., and Garcia de la Banda, M., *J. Bioinform. Comput. Biol.*, 2005, vol. 3, pp. 551–585.

183. Vijayakumar, V., Guerrero, A.N., Davey, N., Lebrilla, C.B., Shields, D.C., and Khaldi, N., *J. Proteome Res.*, 2012, vol. 11, pp. 6056–6065.

184. Käll, L., Krogh, A., and Sonnhammer, E.L.L., *J. Mol. Biol.*, 2004, vol. 338, pp. 1027–1036.

185. Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H., *Nat. Methods*, 2011, vol. 8, pp. 785–786.

186. Zhang, Z. and Henzel, W.J., *Protein Sci.*, 2004, vol. 13, pp. 2819–2824.

187. Hiller, K., Grote, A., Scheer, M., Munch, R., and Jahn, D., *Nucleic Acids Res.*, 2004, vol. 32, pp. W375–W379.

188. Bendtsen, J.D., Nielsen, H., Widdick, D., Palmer, T., and Brunak, S., *BMC Bioinf.*, 2005, vol. 6, pp. 1–9.

189. Emanuelsson, O., Nielsen, H., and von Heijne, G., *Protein Sci.*, 1999, vol. 8, pp. 978–984.

190. Wernersson, R., *Nucleic Acids Res.*, 2006, vol. 34, pp. W385–W388.

191. Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G., *J. Mol. Biol.*, 2000, vol. 300, pp. 1005–1016.

192. Alexandridou, A., Dovrolis, N., Tsangaris, G.T., Nikita, K., and Spyrou, G., *Nucleic Acids Res.*, 2011, vol. 39, pp. W381–W384.

193. Manguy, J., Jehl, P., Dillon, E.T., Davey, N.E., Shields, D.C., and Holton, T.A., *J. Proteome Res.*, 2017, vol. 16, pp. 712–719.

194. Olsen, L.R., Kudahl, U.J., Simon, C., Sun, J., Schonbach, C., Reinherz, E.L., Zhang, G.L., and Brusic, V., *J. Immunol. Methods*, 2013, vol. 400-401, pp. 37–44.

195. Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E., *Genome Res.*, 2004, vol. 14, pp. 1188–1190.

196. Sharma, A., Gupta, P., Kumar, R., Bhardwaj, A., D'Costa, V.M., Boeckel, T.P.V., Levy, S.B., Marshall, B., Costerton, J.W., Hall-Stoodley, L., et al., *Sci. Rep.*, 2016, vol. 6, pp. 1–13.

197. Xu, Y., Luo, C., Qian, M., Huang, X., and Zhu, S., *BMC Genomics*, 2014, vol. 15 S.

198. Mooney, C., Haslam, N.J., Pollastri, G., and Shields, D.C., *PLoS One*, 2012, vol. 7, pp. 1–12.

199. Holton, T.A., Pollastri, G., Shields, D.C., and Mooney, C., *Bioinformatics*, 2013, vol. 29, pp. 3094–3096.

200. Xiao, X., Wang, P., Lin, W.-Z., Jia, J.-H., and Chou, K.-C., *Anal. Biochem.*, 2013, vol. 436, pp. 168–177.

201. Chen, W., Ding, H., Feng, P., Lin, H., and Chou, K.-C., *Oncotarget*, 2016, vol. 7, pp. 16895–16909.

202. Sharma, A., Kapoor, P., Gautam, A., Chaudhary, K., Kumar, R., Chauhan, J.S., Tyagi, A., and Raghava, G.P.S., *Sci. Rep.*, 2013, vol. 3, pp. 1–7.

203. Karosiene, E., Rasmussen, M., Blicher, T., Lund, O., Buus, S., and Nielsen, M., *Immunogenetics*, 2013, vol. 65, pp. 711–724.

204. Rasmussen, M., Fenoy, E., Harndahl, M., Kristensen, A.B., Nielsen, I.K., Nielsen, M., and Buus, S., *J. Immunol.*, 2016, vol. 197, pp. 1517–1524.

205. Nielsen, M. and Andreatta, M., *Genome Med.*, 2016, vol. 8, pp. 1–9.

206. Giguère, S., Drouin, A., Lacoste, A., Marchand, M., Corbeil, J., and Laviolette, F., *J. Immunol. Methods*, 2013, vol. 400-401, pp. 30–36.

207. Tang, H., Su, Z.-D., Wei, H.-H., Chen, W., and Lin, H., *Biochem. Biophys. Res. Commun.*, 2016, vol. 477, pp. 150–154.

208. Gautam, A., Chaudhary, K., Kumar, R., Sharma, A., Kapoor, P., Tyagi, A., Open Source Drug Discovery Consortium, and Raghava, G.P.S., *J. Transl. Med.*, 2013, vol. 11, pp. 1–12.

209. Dimitrov, I., Bangov, I., Flower, D.R., and Doytchinova, I., *J. Mol. Model.*, 2014, vol. 20, pp. 1–6.

210. Dhanda, S.K., Gupta, S., Vir, P., and Raghava, G.P.S., *Clin. Dev. Immunol.*, 2013, vol. 2013, pp. 1–9.

211. Família, C., Dennison, S.R., Quintas, A., and Phoenix, D.A., *PLoS One*, 2015, vol. 10, pp. 1–16.

212. London, N., Raveh, B., Cohen, E., Fathi, G., and Schueler-Furman, O., *Nucleic Acids Res.*, 2011, vol. 39, pp. 249–253.

213. Trabuco, L.G., Lise, S., Petsalaki, E., and Russell, R.B., *Nucleic Acids Res.*, 2012, vol. 40, pp. W423–W427.

214. Saladin, A., Rey, J., Thévenet, P., Zacharias, M., Moroy, G., and Tufféry, P., *Nucleic Acids Res.*, 2014, vol. 42, pp. W221–W226.

215. Olimpieri, P.P., Chailyan, A., Tramontano, A., and Marcatili, P., *Bioinformatics*, 2013, vol. 29, pp. 2285–2291.

216. Sedan, Y., Marcu, O., Lyskov, S., and Schueler-Furman, O., *Nucleic Acids Res.*, 2016, vol. 44, pp. W536–W541.

217. Kurcinski, M., Jamroz, M., Blaszczyk, M., Kolinski, A., and Kmiecik, S., *Nucleic Acids Res.*, 2015, vol. 43, pp. W419–W424.

218. de Vries, S.J., Rey, J., Schindler, C.E., Zacharias, M., and Tuffery, P., *Nucleic Acids Res.*, 2017, vol. 45, pp. 361–364.

219. Trellet, M., Melquiond, A.S., and Alexandre, M.J., *Comput. Peptidol.*, 2015, vol. 1268, pp. 221–239.

220. Peterson, L.X., Roy, A., Christoffer, C., Terashi, G., and Kihara, D., *PLoS Comput. Biol.*, 2017, vol. 13, pp. 1–28.

221. Alam, N., Goldstein, O., Xia, B., Porter, K.A., Kozakov, D., and Schueler-Furman, O., *PLoS Comput Biol.*, 2017, vol. 13, pp. 1–20.

222. Lee, H., Heo, L., Lee, M.S., and Seok, C., *Nucleic Acids Res.*, 2015, vol. 43, pp. 431–435.

223. Das, A.A., Sharma, O.P., Kumar, M.S., Krishna, R., and Mathur, P.P., *Genomics, Proteomics Bioinf.*, 2013, vol. 11, pp. 241–246.

224. Karosiene, E., Rasmussen, M., Blicher, T., Lund, O., Buus, S., and Nielsen, M., *Immunogenetics*, 2013, vol. 65, pp. 711–724.

225. Atanasova, M., Dimitrov, I., Flower, D.R., and Doytchinova, I., *Mol. Inform.*, 2011, vol. 30, pp. 368–375.

Translated by N. Onishchenko