

BIOPEP Database and Other Programs for Processing Bioactive Peptide Sequences

PIOTR MINKIEWICZ, JERZY DZIUBA,¹ ANNA IWANIAK, MARTA DZIUBA, and MAŁGORZATA DAREWICZ

University of Warmia and Mazury in Olsztyn, Chair of Food Biochemistry, Plac Cieszyński 1, 10-726 Olsztyn-Kortowo, Poland

This review presents the potential for application of computational tools in peptide science based on a sample BIOPEP database and program as well as other programs and databases available via the World Wide Web. The BIOPEP application contains a database of biologically active peptide sequences and a program enabling construction of profiles of the potential biological activity of protein fragments, calculation of quantitative descriptors as measures of the value of proteins as potential precursors of bioactive peptides, and prediction of bonds susceptible to hydrolysis by endopeptidases in a protein chain. Other bioactive and allergenic peptide sequence databases are also presented. Programs enabling the construction of binary and multiple alignments between peptide sequences, the construction of sequence motifs attributed to a given type of bioactivity, searching for potential precursors of bioactive peptides, and the prediction of sites susceptible to proteolytic cleavage in protein chains are available via the Internet as are other approaches concerning secondary structure prediction and calculation of physicochemical features based on amino acid sequence. Programs for prediction of allergenic and toxic properties have also been developed. This review explores the possibilities of cooperation between various programs.

Peptides belong to a major group of compounds that play an important role in the regulation of functions of living organisms (1). Food products may be a source of biologically active peptides. The composition of the peptide pool affects physicochemical, functional, and sensory properties, and data on peptides can be used in the study of food immunology. The presence and abundance of peptides may provide information about product history and authenticity (for a review, *see ref. 2*).

Recently there has been an “explosion” of *in silico* (computer simulation) methods in biology, biochemistry, and chemistry (3–5). The available Internet databases of protein sequences have become a commonly used tool in biological sciences, medicine, biotechnology, and food science (6–9). Some applications of bioinformatic methods for studies into peptides of food origin have been summarized by Pripp *et al.* (10), Minkiewicz *et al.* (2), and Darewicz *et al.* (11). The allergenic properties of proteins and peptides have become a hot topic for bioinformatics (12–14), and achievements in this approach are very important for scientists working in the area of food and nutrition. The possible role of computational methods and their links with experiments in peptide science have also been discussed by Paulson *et al.* (15), Dziuba *et al.* (16), and Minkiewicz *et al.* (2). *In silico* methods should provide preliminary answers to questions such as, Are possible products of the hydrolysis of these proteins bioactive; what are the possible biological, physicochemical, functional, or sensory properties of these peptides; is this peptide potentially allergenic; how are peptides with desired properties liberated from these proteins; how can the release of potentially allergenic, toxic, or other undesired peptides be avoided; and can these peptides provide information about product authenticity? The answers would indicate directions for further experimental research.

The use of an increasing number of Internet databases and programs requires their organization according to clear criteria enabling their efficient application (3). Peptide science needs specific methods. This postulate concerning analytical methods (1, 17) may be extended to the computational approach. Computational tools should be able to give valuable information based on short amino acid sequences. The aim of this review is to show the possibilities of meeting this postulate based on the example of the BIOPEP database and program developed by the authors as well as other programs and databases. The second aim is to explore the possibilities of cooperation between various bioinformatic tools in the area of peptide science.

BIOPEP Database and Program

BIOPEP is a database developed at the Chair of Food Biochemistry of the University of Warmia and Mazury in Olsztyn, Poland (18). BIOPEP can be divided into

Guest edited as a special report on “Accurate Methodology for Amino Acids and Bioactive Peptides in Functional Foods and Dietary Supplements for Assessing Protein Adequacy and Health Effects” by G. Sarwar Gilani and Paul J. Moughan.

¹ Author to whom correspondence should be addressed; e-mail: jerzy.dziuba@uwm.edu.pl

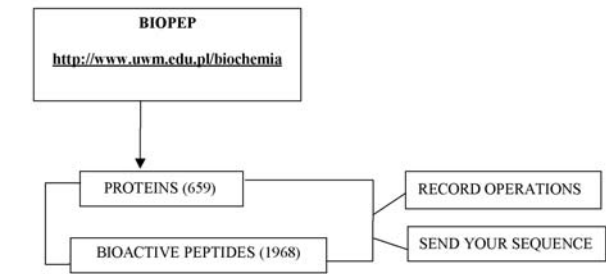


Figure 1. Organization of BIOPEP database.

3 fundamental parts: 2 databases of sequences (Proteins and Bioactive peptides) and an integrated Record operations (Figure 1). Both databases gather the following data: sequence, number of amino acid residues, chemical and monoisotopic mass, and references. In a protein database, one can find a data on the protein function as well as the accession number of the protein in the Swiss-Prot database. In the case of peptides, the activity of peptide, EC₅₀ value (the concentration of *ith* bioactive fragment corresponding to its half-maximal activity, expressed in μM) is also inserted into the database (Figure 2). Both proteins and bioactive peptides possess their own identification number given automatically after the insertion of the sequence into the appropriate database.

At present in BIOPEP, data concerning 659 proteins and 1968 peptides with 37 types of bioactivities have been gathered and are presented in Table 1. Peptides with the same sequence which have different activities have been put into BIOPEP more than once, e.g., sequence Glycyl-Proline (GP) with an antithrombotic (regulating stomach mucosal membrane and dipeptidyl peptidase IV inhibitor functions) can be found in the database 3 times under the ID numbers 3283, 2752, and 3169, respectively. Peptides and proteins are inserted into the BIOPEP by the use of

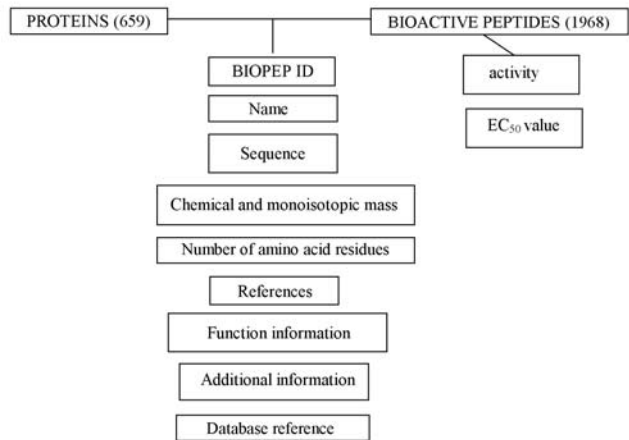


Figure 2. Type of data concerning protein and peptide sequences gathered in BIOPEP.

Table 1. Type of activities of peptides in a BIOPEP database

No.	Types of peptides activities
1	Accelerating cell adhesion
2	Activating ubiquitin-mediated proteolysis
3	Anorectic
4	Antiamnestic
5	Antibacterial
6	Antibiotic
7	Anticancer
8	Antifungal
9	Antihypertensive
10	Anti-inflammatory
11	Antinociceptive
12	Antioxidative
13	Antithrombotic
14	Antiviral
15	Bacterial permease ligand
16	Celiac toxic
17	Chemotactic
18	Contracting smooth muscles
19	Dipeptidyl peptidase IV inhibitor
20	Embryotoxic
21	Fertilization-activating peptide
22	Hemolytic
23	Hypotensive
24	Immunomodulating
25	Immunostimulating
26	Metal binding
27	Natriuretic
28	Neuroactive
29	Neuropeptide
30	Neuropeptide-orphanin receptor GPR14 agonist
31	Opioid
32	Opioid agonist
33	Opioid antagonist
34	Regulating (ion flow, stomach mucosal membrane)
35	Sensory (bitter taste peptides)
36	Stimulating
37	Vasoconstrictor

Table 2. Examples of peptides with antibacterial activity, containing C-terminal amide group, present in the BIOPEP database

BIOPEP ID	Peptide name	Sequence
2579	Tenecin-1 (35-43) (ref. 20)	NGKRVCVCR~
2580	Tenecin-1 (35-43) (ref. 20)	NGKRVCVCRG ^a
2595	Tachylepsin III from homocytes of horseshoe crab (ref. 21)	KWCFRVCYRGICYRRCR~
2596	Tachylepsin III from homocytes of horseshoe crab (ref. 21)	KWCFRVCYRGICYRRCRG ^a
3880	Temporin A (ref. 22)	FLPLIGRVLSGIL~
3881	Temporin A (ref. 22)	FLPLIGRVLSGILG ^a
3882	Temporin B (ref. 22)	LLPIVGNLLKSLL~
3883	Temporin B (ref. 22)	LLPIVGNLLKSLLG ^a
3884	Temporin C (ref. 22)	LLPILGNLLNGLL~
3885	Temporin C (ref. 22)	LLPILGNLLNGLLG ^a

^a According to Bradbury and Smyth (ref. 19), precursors of peptides containing C-terminal amide groups should have glycine residue at C-terminus.

a single letter code. Apart from single letters meaning the amino acid residues, other symbols can be entered into BIOPEP, e.g., peptides including post-translational modifications. For example, the symbol “~” added to the C-end of the peptide describes a C-terminal amide group (currently 116 peptides in BIOPEP). Such peptides appear in the BIOPEP twice: as precursors containing glycine residue at the C-terminus and as mature peptides. The first option allows finding peptides in the sequences of their precursor proteins, the second allows identification on the basis of mass spectrometry. Such annotation is consistent with the mechanism of C-terminal amidation as reviewed by Bradbury and Smyth (19). Some examples of amidated peptides (20–22) are presented in Table 2.

There are 3 possible ways to find the protein/peptide in the BIOPEP database (Figure 3). Proteins can be found by common name, Latin name, and range of molecular mass. Bioactive peptides are searched for by giving a common name of peptide (if such exists), activity, and range of molecular mass. The commands entitled “Search your sequence in the protein database,” or “Search your sequence in the bioactive peptide database,” which appear after opening the bookmark defined as “Record operations,” also enable finding protein sequences or motifs with biological activity. A practical example of using the above-mentioned option is shown in Figure 4. The peptide to find is MAIPPK. By using the option “copy” and “paste” to the “Search your sequence in a bioactive peptide database” menu option, the BIOPEP database shows links to 4 peptides containing the MAIPPK motif.

BIOPEP is able to evaluate the proteins as a potential source of bioactive peptides. It can be performed by means of the following criteria: (1) the profile of potential biological activity of the protein defined as the type and location of the bioactive fragment in a protein chain; (2) the frequency of the occurrence of bioactive fragment in a protein chain (*A*); and (3) the potential biological activity of the protein (*B*) (18).

The frequency of the occurrence of a bioactive fragment in a protein chain (*A*) is defined as:

$$A = \frac{a}{N} \quad (1)$$

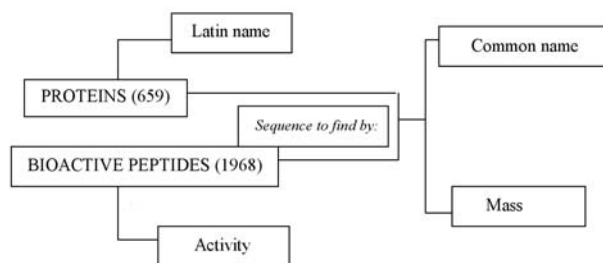
where *a* = number of fragments with a given activity in a protein sequence; *N* = the number of amino acid residues in a protein.

The second discriminant is the potential biological activity of protein (*B*) [mM⁻¹] defined according to equation:

$$B = \frac{\sum_{i=1}^k \frac{a_i}{EC_{50i}}}{N} \quad (2)$$

where *a_i* = the number of repetitions of *i*th bioactive fragment in protein sequence; *EC*_{50*i*} = the concentration of *i*th bioactive peptide corresponding to its half-maximal activity [mM]; *k* = the number of different fragments with a given activity; *N* = the number of amino acid residues in protein (23–25).

Calculation of discriminant *B* requires a knowledge of the *EC*₅₀ value. After this value is determined, it is inserted into

**Figure 3. Criteria of searching proteins and bioactive fragments in a BIOPEP database.**

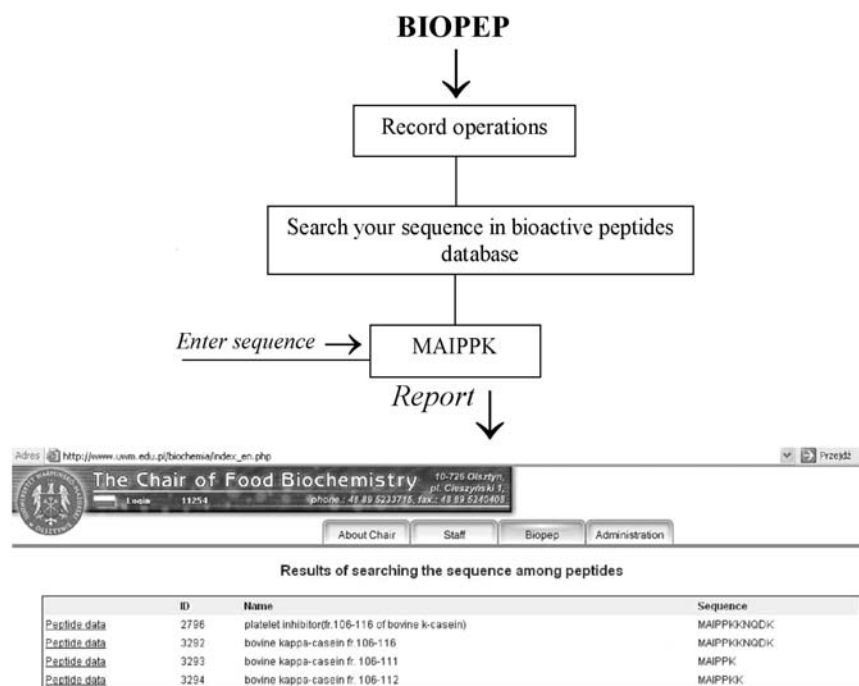


Figure 4. Example describing how to find the bioactive sequence of peptide using the link “The record operations.”

the BIOPEP bioactive peptide database. If the EC_{50} value is not known, then the calculation of B is not performed. All of the above-mentioned discriminants of protein evaluation are described on the website and they are calculated automatically after the selection of a protein for analysis.

Users can create a profile of potential biological activity of protein for one activity as well as all possible activities that can be found in a protein sequence. It can be made by the selection of one activity from a toolbar or an asterisk meaning all activities. The example of the profile of potential activity of the garden pea vicilin (*Pisum sativum*) is shown in Figure 5, and the values of A and B calculated for this protein are presented in Figure 6. Within the vicilin sequence consisting of 124 amino acid residues, BIOPEP identified the presence of peptides with the following activities: antihypertensive, antiemetic, antioxidative, regulating the action of stomach mucosal membrane, activating ubiquitin-mediated proteolysis, dipeptidyl peptidase IV inhibitors, sensory (bitter taste) peptides, and free amino acids involved in featuring the bitter taste in foods.

The profile of vicilin revealed that encrypted peptides are di- or tripeptides, which is important for their absorption from the digestive tract to the blood (more effective than in the case of free amino acids; 26). It should be noted that owing to the high probability of the occurrence of single amino acids, as well as short fragments with bitter taste properties, single amino acids were found in all protein sequences gathered in BIOPEP. Their production from a protein is not a major concern, however, since hydrolysis contributes to the exposition of hydrophobic fragments responsible for the

bitterness. The bitterness depends on factors such as conditions of hydrolytic processes, type of enzymes used, and hydrolysis degree. The properties of the final product such as the size, amino acid composition, and position of hydrophobic residues in a peptide chain play an important role in taste development (27). The values of A confirmed that vicilin may be a good source of peptides with antihypertensive and sensory properties. B discriminant could be calculated only for antihypertensive activity because of the familiarity of EC_{50} value. Some other examples of profiles and quantitative parameter calculations have been published (16, 24, 25, 27).

Bioactive peptides can be produced by enzymatic hydrolysis by digestive tract enzymes; fermentation with proteolytic starter cultures (for example, milk proteins); proteolysis by enzymes originating from microorganisms and plants; and proteolytic action by indigenous proteases present in protein. It is possible to identify proteins that contain fragments with potential functional properties as well as to find specific enzymes able to release these fragments (28). Apart from the qualitative and quantitative criteria of protein evaluation, BIOPEP can show which enzyme can release peptides with biological activity due to the option called “Enzymes action” (Figure 7). Users can select the protein sequence from the database and then from 1 to a maximum of 3 enzymes gathered in BIOPEP (currently out of 25 enzymes) to predict if such a combination of enzymes can produce biologically active peptides. The proteolysis simulation is based on the so-called recognition sequence defined as the fragment of protein recognized by an enzyme as well as the cut sequence—meaning specific amino acid residues bound



Figure 5. Exemplary profile of potential biological activity of garden pea (*Pisum sativum*) vicilin.

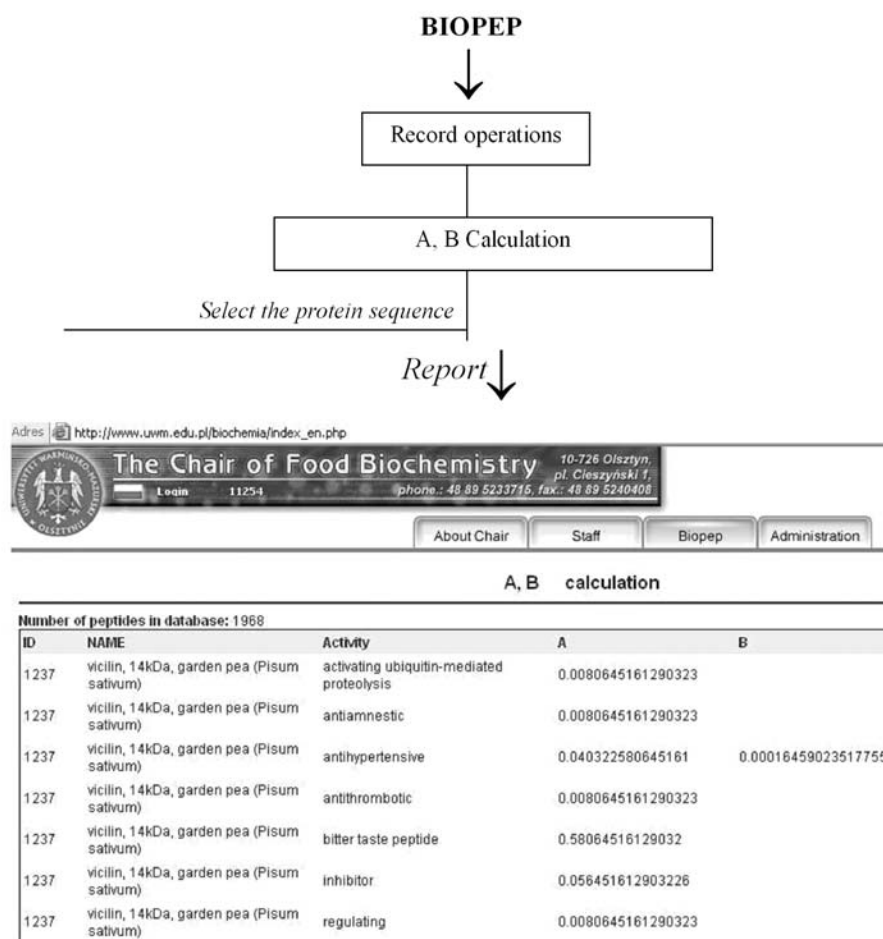


Figure 6. Values of discriminants *A* and *B* calculated for garden pea (*Pisum sativum*) vicilin.

by peptide linkage hydrolyzed (cut) by an enzyme. The example presented in Figure 7 shows that the combination of digestive tract enzymes such as chymotrypsin [EC 3.4.21.1], pepsin [EC 3.4.23.1], and trypsin [3.4.21.4] can potentially produce fragments with antioxidative, antihypertensive, and sensory activities.

Other Databases and Programs that Process Peptide Sequences

The number of amino acid sequence databases and sequence-processing programs is increasing rapidly. One of the most important topics (1) concerns tools enabling peptide identification. Bioinformatic tools supporting the identification of peptides via MS have recently been described in several reviews (29–35).

The databases of peptides revealing various biological activity and programs processing peptide sequences (18, 36–87) are summarized in Table 3. Information about open-source programs able to process sequences containing 20 amino acid residues is given in Table 3. This information was chosen arbitrarily. All the programs can work in a Windows environment without any

additional software installation. They may be easily operated by nonspecialists and do not require registration for noncommercial use.

There are 4 major peptide sequence databases: BIOPEP (Olsztyn, Poland; 18), EROP (Moscow, Russia; 37), SwePep (Uppsala, Sweden; 38), and PepBank (Cambridge, MA; 39). Of these, EROP, SwePep, and PepBank are focused on endogenous peptides, whereas BIOPEP concentrates mainly on peptides of food origin. The first 3 use resources mainly from the area of life sciences; the fourth uses predominantly literature from the area of food science. BIOPEP and EROP contain a relative high number of the di- and tripeptide sequences. The shortest peptides are the most interesting for food scientists because they are relatively easily absorbed from the digestive tract (23). EROP, SwePep, and PepBank all provide links to protein databases with data on proteins being their precursors. On the other hand, the shortest peptides (di- and tripeptides) cannot be attributed to individual precursors because of the relatively high probability of their occurrence in random sequences (88). EROP, PepBank, and BIOPEP provide search engines finding fragments identical to a given sequence within the database. PepBank is also linked to BLAST and SSEARCH programs, enabling screening of major protein databases.

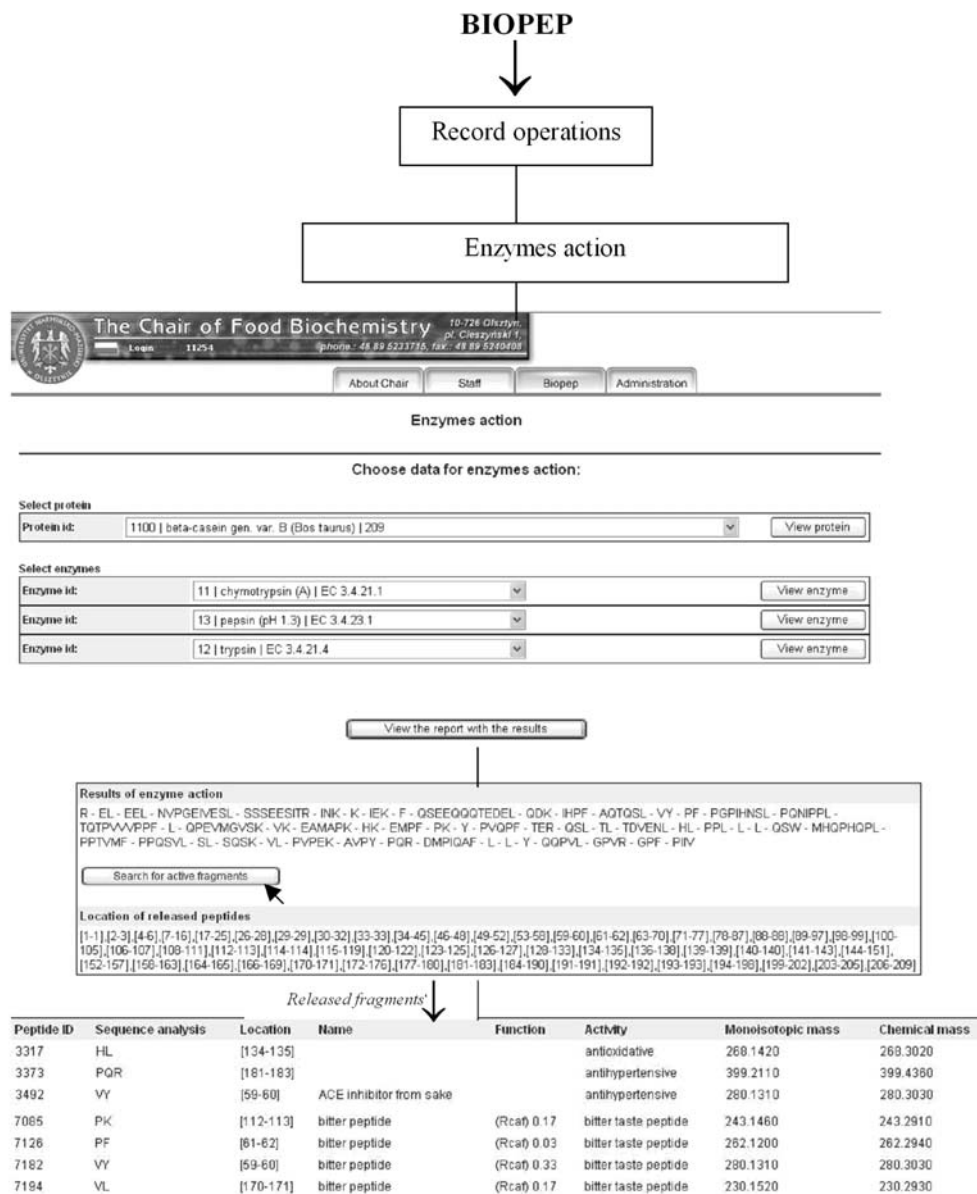


Figure 7. Scheme of simulation of proteolysis in BIOPEP.

SwePep contains the most advanced tools for identification of peptides using MS. BIOPEP and PepBank may allow the construction of a profile of potential biological activity of peptide fragments. PepBank may be screened using BLAST or SSEARCH programs, using protein sequence as a query (39). The other peptide sequence databases are focused on antimicrobial or allergenic peptides (Table 3).

Proteolysis is the most common process for obtaining biologically active peptides. Data concerning the specificity of proteolytic enzymes is thus crucial for the design of technological processes for obtaining protein hydrolysates or individual peptides. The programs BIOPEP (18) and Peptide Cutter (40) offer simple tools for prediction bonds in protein chains susceptible to the action of most common proteolytic enzymes. Recently, more sophisticated strategies involving hydrophobicity of protein fragments as well as secondary and

tertiary structures or protein unfolding have been involved (89–92). Tsai et al. (89) found that unfolded or disordered protein fragments are more susceptible to proteolysis than fragments with well defined secondary structures. Proteolysis process involves demasking of peptide bonds which are buried in a hydrophobic core in native protein structures (89). Thus, it could be expected that more Web-available programs modeling proteolysis belong to the new generation, such as POPS (49) or NeuroPred (50). Enzyme databases such as MEROPS (51), BRENDA (52), and CutDB (53) would contain information about specificity and/or optimal conditions of enzyme action.

Programs designed to search for homology between protein sequences may also serve for protein database screening aimed at finding potential precursors of biologically active peptides. Such programs are complementary to

Table 3. Databases and programs used for peptide sequence processing, available via the Internet

Database or program	Website	Ref.	Comment
Bioinformatics Links Directory	http://bioinformatics.ca/links_directory/	(36)	Website providing access to programs described on the Web Server Issues of the journal <i>Nucleic Acids Research</i> and to some other programs processing protein and peptide sequences
BIOPEP	http://www.uwm.edu.pl/biochemia	(18)	Database of biologically active peptide sequences, program determining profiles of potential biological activity of proteins, calculating quantitative parameters describing the potential biological activity of protein fragments as well as simulating proteolysis
EROP-Moscow	http://erop.inbi.ras.ru	(37)	Database of biologically active peptides.
SwePep	http://www.sweep.org	(38)	Database of biologically active peptides. Contains sophisticated computing tools for peptide mass spectra interpretation
PepBank	http://pepbank.mgh.harvard.edu	(39)	Database of biologically active peptides. Provides program searching for fragments with sequences similar to peptides in the database
UniProt (SwissProt/TrEMBL)	http://www.expasy.org	(40, 41)	Major database of protein sequences. Some biologically active fragments of proteins are also annotated. Website also contains tools for secondary structure prediction, binary and multiple alignments, and calculation of physicochemical features
ANTIMIC	http://research.i2r.a-star.edu.sg/Templar/DB/ANTIMIC	(42)	Database of antimicrobial peptides enabling, e.g., sequence similarity searching via the BLAST program
APD	http://aps.unmc.edu/AP/main.html	(43)	Database of antimicrobial and anticancer peptides
Defensins knowledgebase	http://defensins.bii.aster.edu.sg	(44)	Database of antimicrobial peptides from the defensin family
AMPer	http://www.cnbi2.com/cgi-bin/amp.pl	(45)	Database of antimicrobial peptides and program searching for fragments with sequences similar to antimicrobial peptides in query protein sequence
EPIMHC	http://bio.dfci.harvard.edu/epimhc/	(46)	Database of allergenic peptides. Program comparing sequences of proteins and potentially allergenic peptides. Also accepts peptide sequences as queries
SVMHC	http://www-bs.informatik.uni-tuebingen.de/SVMHC	(47)	Database of allergenic peptides. Program comparing sequences of proteins and potentially allergenic peptides. Also accepts peptide sequences as queries
EpiJen	http://www.jenner.ac.uk/EpiJen	(48)	Database of allergenic peptides. Program comparing sequences of proteins and potentially allergenic peptides. Also accepts peptide sequences as queries
Peptide-Cutter	http://www.expasy.org	(40)	Program for proteolysis simulation
POPS	http://pops.csse.monash.edu.au	(49)	Prediction of proteolytic cleavage sites
NeuroPred	http://neuroproteomics.scs.uiuc.edu/neuropred.html	(50)	Prediction of sites of cleavage leading to the release of neuropeptides
MEROPS	http://merops.sanger.uk	(51)	Database of proteolytic enzymes
BRENDA	http://www.brenda.uni-koeln.de	(52)	Database containing data on enzymes, e.g., proteolytic. Contains e.g., data concerning optimal condition of enzyme action
CutDB	http://cutdb.burnham.org	(53)	Database concerning data on the specificity of less-known endopeptidases
Pôle Bioinformatique Lyonnais	http://pbil.univ-lyon1.fr/	(54)	Server giving access to many programs able to process peptide sequences, e.g., constructing multiple alignments between sequences, secondary structure prediction, calculating amino acid composition, and profiles of physicochemical features
FASTA	http://www.ebi.ac.uk/fasta33/	(55)	Program designed for sequence alignments between proteins. The version presented is also sufficient for alignments between peptide and protein sequences.

Table 3. (continued)

Database or program	Website	Ref.	Comment
BLAST	http://dove.embl-heidelberg.de/Blast2	(56)	Program designed for sequence alignments between proteins. The version presented is also sufficient for alignments between peptide and protein sequences. The program screens protein sequence databases for fragments with sequences similar to the peptide sequence submitted as a query.
SSEARCH	http://www.ddbj.nig.ac.jp/search	(59, 60)	Program designed for sequence alignments between proteins. The version presented is also sufficient for alignments between peptide and protein sequences. The program screens protein sequence databases for fragments with sequences similar to the peptide sequence submitted as a query.
MS-BLAST	http://dove.embl-heidelberg.de/Blast2 MS-BLAST software is selected by clicking icon "New MS-BLAST"	(61)	Program designed for interpretation of mass spectrometric data of proteins with unknown sequences, also useful for alignments between oligopeptide and protein sequences. The program screens protein sequence databases for fragments with sequences similar to the peptide sequence submitted as a query.
PARALIGN	http://www.paralign.org	(62)	Program designed for sequence alignments between proteins. The version presented is also sufficient for alignments between peptide and protein sequences. The program screens protein and patented peptide sequence databases for fragments with sequences similar to the peptide sequence submitted as a query.
CLUSTAL W	http://www.ebi.ac.uk/clustalw/ ; http://www.ddbj.nig.ac.jp/search	(63)	Program designed for multiple sequence alignments between proteins. Also sufficient for aligning peptide sequences.
T-Coffee/M-Coffee	http://www.tcoffee.org	(64, 65)	Server designed for multiple sequence alignments between proteins. Also sufficient for aligning peptide sequences.
Kalign	http://msa.cgb.ki.se	(66, 67)	Program designed for multiple sequence alignments between proteins. Also sufficient for aligning peptide sequences.
Lalign	http://www.expasy.org	(68)	Program designed for construction of binary alignments between protein sequences. Also accepts peptide sequences as queries. Program is able to construct more than one possible local alignment between 2 sequences.
Pratt	http://www.ebi.ac.uk/pratt	(69)	Program finding sequence motifs in peptides or proteins
MEME	http://www.sdsc.edu/MEME	(70, 71)	Program finding sequence motifs in peptides or proteins
RE _E MUS	http://bioinfo.life.nthu.edu.tw/REMUS or http://spider.cs.ntou.edu.tw/BioTools/REMUS	(72)	Program searching for unique motifs in protein and peptides sequences, complementary to programs constructing consensus sequences, useful for highlighting differences between proteins or peptides
SCANMOT	http://caps.ncbs.res.in/scanmot/scanmot.html	(73)	Program searching for characteristic motifs in protein sequences. Also accepts peptide sequences as queries.
EVA	http://cubic.bioc.columbia.edu/eva	(74)	Metaserver enabling access to programs predicting secondary structures of proteins and peptides. The SABLE and PORTER programs are sufficient for oligopeptide secondary structure prediction. List of participating servers may be enriched in the future.
GOR V	http://gor.bb.iastate.edu/	(75)	Server for protein and peptide secondary structure prediction
PreSSAPro	http://bioinformatica.isa.cnr.it/PRESSAPRO/	(76)	Server for protein and peptide secondary structure prediction
FoldUnfold	http://skuld.protres.ru/~mlobanov/ogu/ogu.cgi	(77)	Program predicting the presence of disordered regions in proteins. Also accepts peptide sequences as queries.

Table 3. (continued)

Database or program	Website	Ref.	Comment
POODLE_S	http://mbs.cbrc.jp/poodle/poodle-s.html	(78)	Program predicting presence of disordered regions in proteins. Also accepts peptide sequences as queries.
HSA/HSS	ftp://ftp.agsci.ubc.ca/foodsci/HSA HSS software is selected by clicking on "go to homo ver" option (see the user instruction available on the Website)	(79, 80)	Homology Similarity Analysis (HSA) computes pattern similarity constants and average side chain property index values of segments in sequences. Homology Similarity Search (HSS), a step-wise program initiated from N-terminus of query sequences by shifting the reference segment towards C-terminus. Software based on similarity of segments in terms of similarity constant and average property values compared to template sequences.
PROFEAT	http://jing.cz3.nus.edu.sg/cgi-bin/prof/prof.cgi	(81)	Program for prediction structural and physicochemical features of proteins and peptides
pI/MW	http://www.expasy.org	(40, 82)	Program calculating peptide isoelectric points and molecular masses on the basis of amino acid sequences
ProtParam	http://www.expasy.org	(40)	Program for computing physicochemical features such as molecular mass, isoelectric point, amino acid composition, absorbance coefficient at 280 nm, predicted half-life in vivo, instability index, aliphatic index, and average hydropathy
AAindex	http://www.genome.ad.jp/aaindex ; ftp://ftp.genome.ad.jp/db/genomenet/aaindex	(83)	Database of indexes describing physicochemical properties, sufficient for prediction of peptide properties, e.g., QSAR approach
AlgPred	http://www.imtech.res.in/raghava/algpred	(84)	Program predicting the allergenicity of proteins and peptides
POPI	http://iclab.life.nctu.edu.tw/POPI	(85)	Program predicting the allergenicity of peptides
NTXPred	http://www.imtech.res.in/raghava/ntxpred	(86)	Program predicting the neurotoxic activity of peptides.
BTXPred	http://www.imtech.res.in/raghava/btxpred	(87)	Program classifying peptides and proteins as bacterial toxins

BIOPEP. All the programs from this group meet the principle that the longer the sequence the better, because it provides more information for finding homology. The evaluation of results is based on the statistical criterion: probability of the existence of random sequence similarity (for a review, *see* ref. 93). For example, some versions of BLAST cannot use sequences containing fewer than 20 amino acid residues. The version presented in Table 3, as well as those cooperating with peptide databases, can process such sequences. The program MS BLAST appears to be a sufficient tool for finding precursors of oligopeptides containing fewer than 10 amino acid residues. The application of this program to process very short sequences is discussed by Habermann et al. (94), Wielsch et al. (95), and Darewicz et al. (96).

Reviews describing the state of art of multiple sequence alignment creation with respect to protein sequences have been published by Wallace et al. (97) and Notredame (98). Limitations of multiple alignments have been summarized by Pearson and Sierk (99). Available programs are usually able to process short sequences and construct consensus sequences between various peptides, e.g., to find common amino acid residues in peptides with the same activity. Another option is the construction of motifs attributed to individual activities. Examples of finding motifs associated with neuropeptides

have been published by Baggerman et al. (100) and Liu et al. (101, 102). The above-mentioned programs serve to find the common properties of amino acid sequences. The RE MUS (72), a program designed to find unique motifs, may be useful to highlight the differences between peptide sequences.

Secondary structure is one of the important factors affecting the biological activity of peptides. Methods of secondary structure prediction have been reviewed (103, 104). Many secondary structure-predicting programs are available via servers such as UniProt (41), EVA (74), or Pôle Bioinformatique Lyonnais (54). Some programs are limited in that they do not process sequences containing fewer than 20 amino acids. Progress needs to be made in the reliability of theoretical predictions compared to experiments (105, 106).

Physicochemical properties of individual amino acid residues may be crucial for peptide properties. Insertion, deletion, or substitution of one amino acid residue may have a drastic consequence for peptide properties. Computational methods comparing hydrophobicity profiles of proteins and/or peptides have been successfully used for prediction of susceptibility to proteolytic enzyme action (90) or emulsifying properties (107). Predictions taking into account the physicochemical properties of amino acids and/or

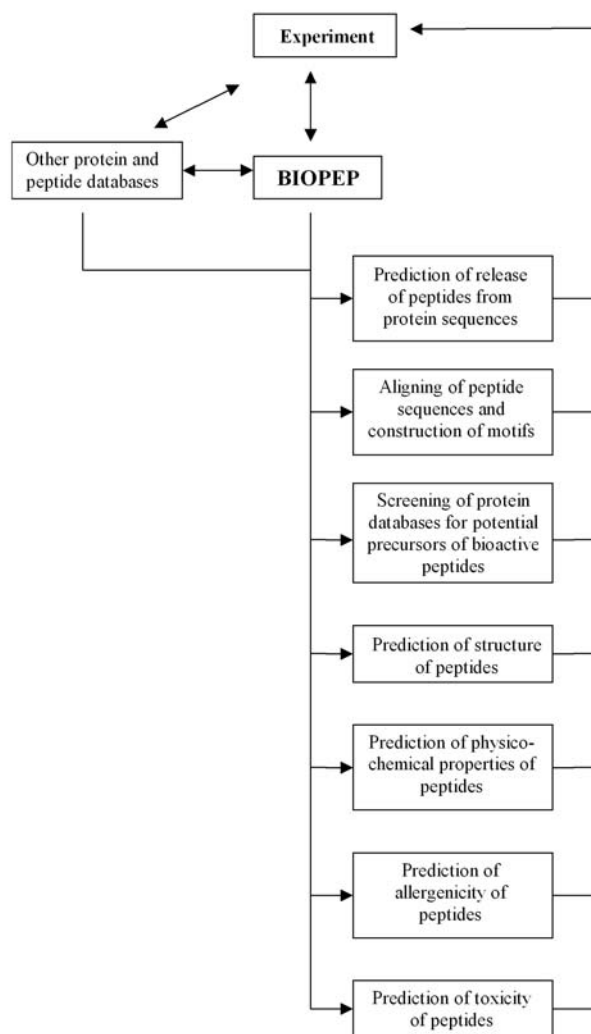


Figure 8. Possible scheme of cooperation of the BIOPEP program with other bioinformatic tools available via the Internet (see Table 3).

secondary structure appear to be more promising than sequence homology searching, in the case of short sequences. Predicting protein function on the basis of structure similarity without sequence homology (108) or similarity of hydrophobicity pattern (109, 110) is also in progress.

A database containing indexes describing the physicochemical properties of amino acids such as hydrophobicity, charge, and secondary structure propensities has been developed (83). The data presented in this database may be sufficient, e.g., for predicting the biological activity of peptides using Quantitative Structure-Activity Relationship (QSAR; for a review, *see ref. 10*). Predicting which protein fragments can be good markers for presence and abundance of proteins (111) may be an excellent example of the application of this database.

Bioinformatic studies predict the allergenic properties of peptides by comparing their sequences with those of known peptides with allergenic properties. Comparison of physicochemical properties of amino acid residues at

appropriate positions has recently been recommended as a promising strategy for allergenicity prediction (112). Prediction algorithms use sequence matching, motif searching, and comparison of physicochemical properties of amino acid residues consisting of peptides (84, 85). Similar approaches have been applied for predicting peptide toxicity (86, 87).

Possible Use of BIOPEP with Other Databases and Programs

The possibilities of using BIOPEP with other sequence databases and sequence-processing programs are presented in Figure 8. It is possible to process sequences from other databases using BIOPEP as well as other programs, as shown in Table 3. There is a possibility of cooperation between the user and BIOPEP administrator to extend the content of our database with the new bioactive peptide sequences (Figure 1). Clicking the option "Send your sequence" allows the user to paste a new sequence with the description of its activity as well as references. After data verification, a new sequence is inserted into the BIOPEP database.

Users can evaluate protein sequences as sources of bioactive peptides as well as design their proteolytic processes when the potentially analyzed sequence cannot be found in BIOPEP. A "new" protein sequence can be analyzed by using the following options: "Calculate *A*, *B*, *Y* for your sequence"; "Calculate the profile of potential biological activity for your sequence"; and "Enzymes action for your sequence." All computations work by copying the protein sequence from the file or available online protein sequence databases and pasting it into the appropriate toolbar using the BIOPEP "Record operations" option (Figure 9).

Amino acid sequence may be transferred from BIOPEP to other programs by using the simple copy-paste option provided by Internet Explorer or equivalent program. "The REMUS" (72), a program designed to find unique motifs highlights the differences between peptide sequences. Differences between the format used by BIOPEP and the FASTA format are highlighted in Figure 10.

An example of cooperation between BIOPEP and other peptide sequence databases is presented in Table 4. The table contains data about peptides containing the query sequence PFPG extracted from 3 peptide sequence databases (113–127). The query sequence is known as a fragment of opioid peptides from bovine β -casein, named β -casomorphins (113, 120, 121). Database screening resulted in 35 hits (18 from BIOPEP, 6 from EROP-Moscow, and 11 from PepBank) concerning 20 peptide sequences. Only 14 hits (defined as information about 1 peptide and one kind of activity, e.g., opioid) occurred in more than one database; 21 hits were found only in one of the databases. Most of the sequences found were fragments of β -casein or related synthetic peptides, although the query sequence was found also in peptides from other sources. The query sequence is a fragment of opioid peptides, toxins as well as inhibitors of prolyl endopeptidase, angiotensin-converting enzyme (ACE), and cathepsin B. Data on the taste of some peptides and their susceptibility to proteolysis are also

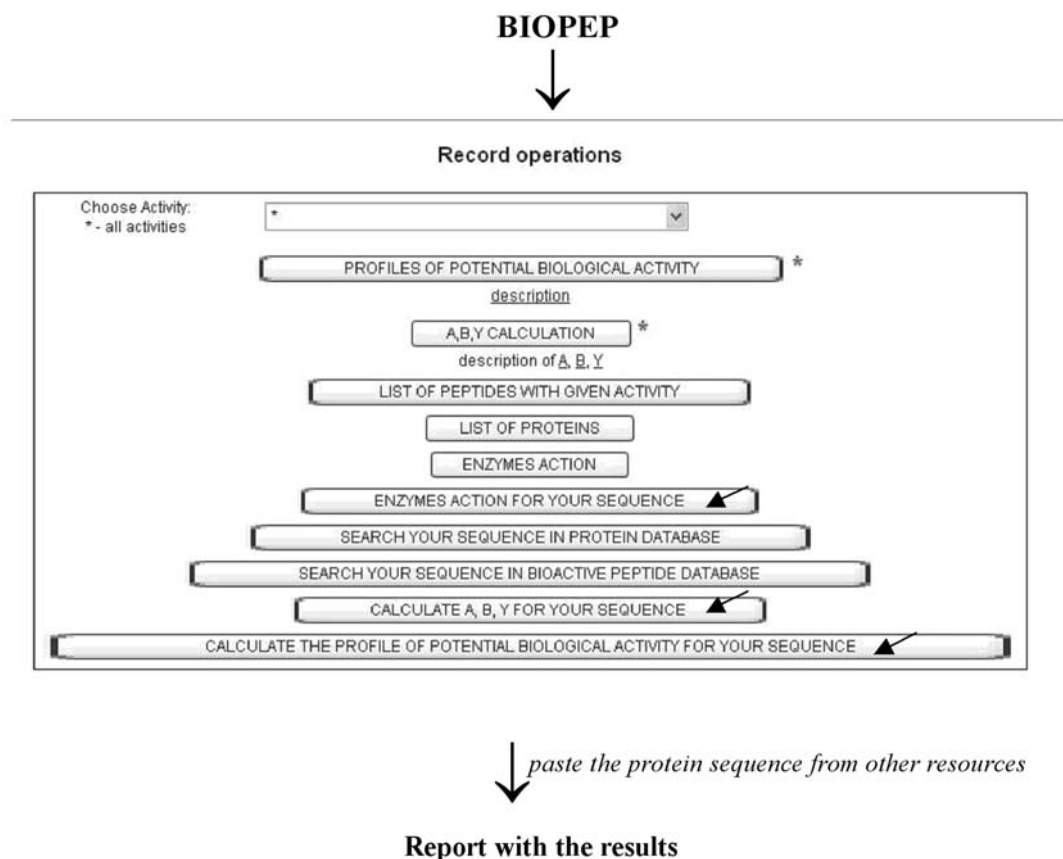


Figure 9. Scheme of processing sequences from other resources by BIOPEP.

available. The example presented shows that the peptide data included in 3 databases only partially overlap. Use of more than one peptide database may thus multiply the possibilities of searching for data concerning peptides.

Final Remarks

The last few years have seen a rapid increase in the number of available bioinformatic tools used for the description and prediction of peptide physicochemical and biological properties. The databases that focus on biologically active peptides, such as BIOPEP, EROP, SwePep, and PepBank, as well as more specialized databases, can become an important source of information on peptides, e.g., for researchers working in the area of food and nutrition. There has also been rapid progress in the area of programs able to process peptide sequences, although most of them have been designed as tools for protein science. The next generation of programs should thus be expected.

QSAR appears to be a promising area of work aimed at predicting the properties of peptides with short amino acid chains (10). This computational method has recently been applied for predicting ACE inhibition (128–130), prolyl endopeptidase inhibition (131), and antimicrobial activity (132), bitterness (133), and modeling of the relationship between ACE inhibition and bitterness of

a)

2807	PEP inhibitor	MPPPLPARVDFSLAGALN
------	---------------	--------------------

b)

```
> 2807 PEP inhibitor
MPPPLPARVDFSLAGALN
```

c)

```
> 2807 PEP inhibitor
MPPPLPARVDFSLAGALN

> 2815 PEP inhibitor
MTPPPLPARVDFSLAGALN

> 2824 PEP inhibitor
MPPPLPTRVDFSLAGALN
```

Figure 10. Example of peptide annotation using different formats: (a) Inhibitor of prolyl endopeptidase (PEP; BIOPEP ID 2807) annotated using BIOPEP format; (b) the same peptide annotated using FASTA format; (c) the same peptide data in FASTA format together with data of 2 other peptides with the same activity (BIOPEP ID 2815 and 2824, respectively) arranged as input for multiple alignment construction using programs such as ClustalW, T-Coffee, or Kalign.

Table 4. Peptides containing fragment PFPG, found in 3 databases: BIOPEP, EROP-Moscow, and PepBank^a

No.	Sequence	BIOPEP ID No./activity	EROP-Moscow ID No./activity	PepBank ID No. ^b /activity
1	YPFPG	3263/opioid (ref. 113)	E00234/opioid (ref. 113)	511111/opioid (ref. 113) 10376245/proteinase substrate (ref. 114) 1967253/proteinase substrate (ref. 115) 2882693/proteinase substrate (ref. 116)
2	YPFPGPI	3262/opioid (ref. 113) 7294/bitter (ref. 117)	E00233/opioid (ref. 113)	511111/opioid (ref. 113) 10480087/ACE inhibitor (ref. 118) 10959001/cathepsin B inhibitor (ref. 119)
3	YPFPGPIPN SL	3251/opioid (ref. 120)	E00610/opioid (ref. 121)	
4	PFPGPI	7289/bitter (ref. 117)	E02217/cathepsin B inhibitor (ref. 119)	10959001/cathepsin B inhibitor (ref. 119)
5	VYPFPGPI	2615/PEP inhibitor (ref. 122)		10959001/cathepsin B inhibitor (ref. 119)
6	IYPFPGPI P	2608/PEP inhibitor (ref. 122)		
7	IYPFPGPI H	2610/PEP inhibitor (ref. 122)		
8	IYPFPGPI	2611/PEP inhibitor (ref. 122)		
9	VYPFPGPI A	2612/PEP inhibitor (ref. 122)		
10	VYPFPGPI P	2613/PEP inhibitor (ref. 122)		
11	VYPFPGPI H	2614/PEP inhibitor (ref. 122)		
12	VYPFPG	3504/ACE inhibitor (ref. 123)		9891260/ACE inhibitor (ref. 123)
13	YPFPGPIPN			10908049/ACE inhibitor (ref. 124)
14	DELQDKIHPFAQSLV YPFPGPI PN	3351/ACE inhibitor (ref. 125)		
15	HPPCCLYGSCR PFPG CYNALCCRKz ^c		E05071/toxic (ref. 126)	
16	GG PFPGY		E04008/toxic (ref. 127)	
17	PFPGPI P	7292/bitter (ref. 117)		
18	VIPFGR	7291/bitter (ref. 117)		
19	VIIIPFGR	7286/bitter (ref. 117)		
20	VIIIPFG	7287/bitter (ref. 117)		

^a ACE = Angiotensin-converting enzyme; PEP = prolyl endopeptidase. Fragment used as a query sequence is indicated in bold letters. Searching was performed October 23, 2007.

^b ID numbers from PepBank, indicated in the table, are attributed to references in PubMed literature database.

^c C-terminal amide groups are annotated in EROP-Moscow using symbol z, which is equivalent to ~ in BIOPEP.

peptides (134, 135). Another approach potentially applicable for work on peptide sequences is machine learning (136, 137). An artificial neural network has been applied for prediction of amylase inhibition by peptides (138) or for prediction of intestinal permeability of peptides (139). It is likely that more programs using such algorithms will be available via the Internet within the next few years.

Compatibility between various programs is a dominant trend in bioinformatics and chemoinformatics (140). This review has illustrated some possibilities of cooperation between programs that work with peptide sequences using the BIOPEP database and another program as an example. The design of systems enabling automated data transfer between various programs can be expected in the future.

Acknowledgments

This work was supported by the Ministry of Science and Higher Education within project No. PBW N N 312 286333 (2007-2009).

References

- Schulz-Knappe, P., Schrader, M., & Zucht, H.-D. (2005) *Combinatorial Chem. High Throughput Screen.* **8**, 697–704
- Minkiewicz, P., Dziuba, J., Darewicz, M., Iwaniak, A., Dziuba, M., & Nałęcz, D. (2008) *Food Technol. Biotechnol.* **46**, 1–10
- Cannata, N., Merelli, E., & Altman, R.B. (2005) *PLoS Comput. Biol.* **1**, 531–533
- Kumar, S., & Dudley, J. (2007) *Bioinformatics* **23**, 1713–1717
- Perez-Iratxeta, C., Andrade-Navarro, M.A., & Wren, J.D. (2007) *Brief. Bioinform.* **8**, 88–95
- Jelić, D., Toth, D., & Verbanac, D. (2003) *Food Technol. Biotechnol.* **41**, 269–286
- Apweiler, R., Bairoch, A., & Wu, C.H. (2004) *Curr. Opin. Chem. Biol.* **8**, 76–80
- Akalin, P.K. (2006) *Mol. Nutr. Food Res.* **50**, 610–619
- Whitfield, E.J., Pruess, M., & Apweiler, R. (2006) *J. Biotechnol.* **124**, 629–639
- Pripp, A.H., Isakson, T., Stepaniak, L., Sørhaug, T., & Ardö, Y. (2005) *Trends Food Sci. Technol.* **16**, 484–494
- Darewicz, M., Dziuba, J., & Minkiewicz, P. (2008) *Food Rev. Int.* **24**(3) (in press)
- Gendel, S.M. (2004) *J. AOAC Int.* **87**, 1417–1422
- Korber, B., LaBute, M., & Yusin, K. (2006) *PLoS Comput. Biol.* **2**, 484–492
- Tong, J.C., Tan, T.W., & Ranganathan, S. (2007) *Brief. Bioinform.* **8**, 96–108
- Paulson, L., Persson, R., Karlsson, G., Silberring, J., Bierczyńska-Krzysik, A., Ekman, R., & Westman-Brinkmalm, A. (2005) *J. Mass Spectrom.* **40**, 202–213
- Dziuba, M., Dziuba, J., & Minkiewicz, P. (2006) *Pol. J. Nat. Sci.* **21**, 999–1020
- Hu, L., Ye, M., Jiang, X., Feng, S., & Zou, H. (2007) *Anal. Chim. Acta* **598**, 193–204
- Dziuba, J., & Iwaniak, A. (2006) in *Nutraceutical Proteins and Peptides in Health and Disease*, Y. Mine & F. Shahidi (Eds), CRC Press, Boca Raton, FL, pp 543–563
- Bradbury, A.F., & Smyth, D.G. (1991) *Trends Biochem. Sci.* **16**, 112–115
- Lee, K.H., Hong, Y.S., Kwon, M., Yoon, J.H., Lee, B.L., & Moon, H.M. (1998) *Biochem. J.* **334**, 99–105
- Saberwal, G., & Nagaraj, R. (1994) *Biochim. Biophys. Acta* **1197**, 109–131
- Simmaco, M., Mignogna, G., Canofeni, S., Miele, R., Mangoni, M.L., & Barra, D. (1996) *Eur. J. Biochem.* **242**, 788–792
- Dziuba, J., Iwaniak, A., & Niklewicz, M. (2003) <http://www.uwm.edu.pl/biochemia>
- Dziuba, J., Iwaniak, A., & Minkiewicz, P. (2003) *Polimery* **48**, 50–53
- Iwaniak, A., Dziuba, J., & Niklewicz, M. (2005) *Acta Aliment.* **34**, 417–425
- Vermeirssen, V., Van Camp, J., & Verstraete, W. (2004) *Br. J. Nutr.* **92**, 357–366
- Wang, W., & Gonzalez de Mejia, E. (2005) *Comp. Rev. Food Sci. Food Saf.* **4**, 63–78
- Korhonen, H., & Pihlanto, A. (2007) *Curr. Pharm. Des.* **13**, 829–843
- Johnson, R.S., Davis, M.T., Taylor, J.A., & Patterson, S.D. (2005) *Methods* **35**, 223–236
- Listgarten, J., & Emily, A. (2005) *Mol. Cell. Proteom.* **4**, 419–434
- Droit, A., Fillon, J., Morissette, J., & Poirier, G.G. (2006) *Curr. Proteom.* **3**, 119–128
- Lisacek, F. (2006) *Proteomics* **6** (Suppl. 1–2), 22–33
- Xu, C., & Ma, B. (2006) *Drug Discov. Today* **11**, 595–600
- Forner, F., Foster, L.J., & Toppo, S. (2007) *Curr. Bioinform.* **2**, 63–93
- Colinge, J., & Bennett, K.L. (2007) *PLoS Comput. Biol.* **3**, 1151–1160
- Fox, J.A., McMillan, S., & Ouellette, B.F.F. (2007) *Nucl. Acids Res.* **35**, W3–W5
- Zamyatnin, A.A., Borchikov, A.S., Vladimirov, M.G., & Voronina, O.L. (2006) *Nucl. Acids Res.* **34**, D261–D266
- Fälth, M., Sköld, K., Norrman, M., Svensson, M., Fenyö, D., & Andren, P.E. (2006) *Mol. Cell. Proteom.* **5**, 998–1005
- Shtatland, T., Guettler, D., Kossodo, M., Pivovarov, M., & Weissleder, R. (2007) *BMC Bioinform.* **8**, 280
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., & Bairoch, A. (2005) in *The Proteomics Protocols Handbook*, J.M. Walker (Ed.), Humana Press Inc., Totowa, NJ, pp 571–607
- The UniProt Consortium (2007) *Nucl. Acids Res.* **35**, D193–D197
- Brahmachary, M., Krishnan, S.P.T., Koh, J.L.Y., Khan, A.M., Seah, S.H., Tan, T.W., Brusica, V., & Bajic, A.V. (2004) *Nucl. Acids Res.* **32**, D586–D589
- Wang, Z., & Wang, G. (2004) *Nucl. Acids Res.* **32**, D590–D592
- Seebah, S., Suresh, A., Zhuo, S., Choong, Y.H., Chua, H., Chuon, D., Beuerman, R., & Verma, C. (2007) *Nucl. Acids Res.* **35**, D265–D268
- Fjell, C.D., Hancock, R.E.W., & Cherkasov, A. (2007) *Bioinformatics* **23**, 1148–1155

- (46) Reche, P.A., Zhang, H., Glutting, J.P., & Reinherz, E.L. (2005) *Bioinformatics* **21**, 2140–2141
- (47) Dönnes, P., & Kohlbacher, O. (2006) *Nucl. Acids Res.* **34**, W194–W197
- (48) Doytchinova, I.A., Guan, P., & Flower, D.R. (2006) *BMC Bioinform.* **7**, 131
- (49) Boyd, S.E., Pike, R.N., Rudy, G.B., Whisstock, J.C., & Garcia de la Banda, M. (2005) *J. Bioinform. Comput. Biol.* **3**, 551–585
- (50) Southey, B.R., Amare, A., Zimmerman, T.A., Rodriguez-Zas, S.L., & Sweedler, J.V. (2006) *Nucl. Acids Res.* **34**, W267–W272
- (51) Rawlings, N.D., Morton, F.R., & Barrett, A.J. (2006) *Nucl. Acids Res.* **34**, D270–D272
- (52) Barthelmes, J., Ebeling, C., Chang, A., Schomburg, I., & Schomburg, D. (2007) *Nucl. Acids Res.* **35**, D511–D514
- (53) Igarashi, Y., Eroshkin, A., Gramatikova, S., Gramatikoff, K., Zhang, Y., Smith, J.W., Osterman, A.L., & Godzik, A. (2007) *Nucl. Acids Res.* **35**, D546–D549
- (54) Combet, C., Blanchet, C., Geourjon, C., & Deleagé, G. (2000) *Trends Biochem. Sci.* **25**, 147–150
- (55) Pearson, W.R. (2000) *Methods Mol. Biol.* **132**, 185–219
- (56) Altschul, S.F., Gisk, W., Miller, W., Meyers, E.W., & Lipman, D.J. (1990) *J. Mol. Biol.* **215**, 403–410
- (57) Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D.J. (1997) *Nucl. Acids Res.* **25**, 3389–3402
- (58) Ye, J., McGinnis, S., & Madden, T.L. (2006) *Nucl. Acids Res.* **34**, W6–W9
- (59) Smith, T.F., & Waterman, M.S. (1981) *J. Mol. Biol.* **147**, 195–197
- (60) Pearson, W.R. (1991) *Genomics* **11**, 635–650
- (61) Shevchenko, A., Sunyaev, S., Loboda, A., Shevchenko, A., Bork, P., Ens, W., & Standing, K.G. (2001) *Anal. Chem.* **73**, 1917–1926
- (62) Sæbo, P.E., Andersen, S.M., Myrseth, J., Laerdahl, J.K., & Rognes, T. (2005) *Nucl. Acids Res.* **33**, W535–W539
- (63) Thompson, J.D., Higgins, D.G., & Gibson, T.J. (1994) *Nucl. Acids Res.* **22**, 4673–4680
- (64) Notredame, C., Higgins, D.G., & Heringa, J. (2000) *J. Mol. Biol.* **302**, 205–217
- (65) Moretti, S., Armougom, F., Wallace, I.M., Higgins, D.G., Jongeneel, C.V., & Notredame, C. (2007) *Nucl. Acids Res.* **35**, W645–W648
- (66) Lassmann, T., & Sonnhammer, E.L.L. (2005) *BMC Bioinform.* **6**, 298
- (67) Lassmann, T., & Sonnhammer, E.L.L. (2006) *Nucl. Acids Res.* **34**, W596–W599
- (68) Huang, X., & Miller, W. (1991) *Adv. Appl. Math.* **12**, 373–381
- (69) Jonassen, I., Collins, J.F., & Higgins, D.G. (1995) *Protein Sci.* **4**, 1587–1595
- (70) Bailey, T.L., & Elkan, C. (1994) in *Proc. Second Conference on Intelligent Systems for Molecular Biology*, AAAI Press, Menlo Park, CA, pp 28–36
- (71) Bailey, T.L., & Gribskov, M. (1998) *J. Comput. Biol.* **5**, 211–221
- (72) Pai, T.-W., Chang, M.D.-T., Tzou, W.-S., Su, B.-H., Wu, P.-C., Chang, H.-T., & Chou, W.-I. (2006) *Nucl. Acids Res.* **34**, W198–W201
- (73) Chakrabarti, S., Anand, A.P., Bhardwaj, N., Pugalanthi, G., & Sowdhamini, R. (2005) *Nucl. Acids Res.* **33**, W274–W277
- (74) Eyrich, V.A., Martí-Renom, M.A., Przybylski, D., Madhusudhan, M.S., Fiser, A., Pazos, F., Valencia, A., Šali, A., & Rost, B. (2001) *Bioinformatics* **17**, 1242–1243
- (75) Sen, T.Z., Jernigan, R.L., Garnier, J., & Kloczkowski, A. (2005) *Bioinformatics* **21**, 2787–2788
- (76) Costantini, S., Colonna, G., & Facchiano, A.M. (2007) *Comput. Biol. Chem.* **31**, 389–392
- (77) Galzitskaya, O.V., Garbuzynskiy, S.D., & Lobanov, M.Y. (2006) *Bioinformatics* **22**, 2948–2949
- (78) Shimizu, K., Hirose, H., & Noguchi, T. (2007) *Bioinformatics* **23**, 2337–2338
- (79) Nakai, S., Chan, J.C.K., Li-Chan, E.C.Y., Dou, J., & Ogawa, M. (2003) *J. Agric. Food Chem.* **51**, 1215–1223
- (80) Nakai, S., & Alizadeh-Pasdar, N. (2006) in *Nutraceutical Proteins and Peptides in Health and Disease*, Y. Mine & F. Shahidi (Eds), CRC Press, Boca Raton, FL, pp 565–582
- (81) Li, Z.R., Lin, H.H., Han, L.Y., Chen, L.J.X., & Chen, Y.Z. (2006) *Nucl. Acids Res.* **34**, W32–W37
- (82) Bjellqvist, B., Hughes, G.J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.-C., Frutiger, S., & Hochstrasser, D.F. (1993) *Electrophoresis* **14**, 1023–1031
- (83) Kawashima, S., & Kanehisa, M. (2000) *Nucl. Acids Res.* **28**, 374
- (84) Saha, S., & Raghava, G.P.S. (2006) *Nucl. Acids Res.* **34**, W202–W209
- (85) Tung, C.-W., & Ho, S.-Y. (2007) *Bioinformatics* **23**, 942–949
- (86) Saha, S., & Raghava, G.P.S. (2007) *In Silico Biol.* **7**, 0025
- (87) Saha, S., & Raghava, G.P.S. (2007) *In Silico Biol.* **7**, 0028
- (88) Dziuba, J., Minkiewicz, P., & Plitnik, K. (1996) *Pol. J. Food Nutr. Sci.* **5**, 85–96
- (89) Tsai, C.J., Polverino de Laureto, P., Fontana, A., & Nussinov, R. (2002) *Protein Sci.* **11**, 1753–1770
- (90) Callebaut, I., Schoentgen, F., Prat, K., Mornon, J.-P., & Jollès, P. (2005) *Biochim. Biophys. Acta* **1749**, 75–80
- (91) Amare, A., Hummon, A.B., Southey, B.R., Zimmerman, T.A., Rodriguez-Zas, S.L., & Sweedler, J.V. (2006) *J. Proteome Res.* **5**, 1162–1167
- (92) Ozer, N., Haliloglu, T., & Schiffer, C.A. (2006) *Proteins* **64**, 444–456
- (93) Mitrophanov, A.Y., & Borodovsky, M. (2006) *Brief. Bioinform.* **7**, 2–24
- (94) Habermann, B., Oegema, J., Sunyaev, S., & Shevchenko, A. (2004) *Mol. Cell. Proteom.* **3**, 238–249
- (95) Wielsch, N., Thomas, H., Surendranath, V., Waridel, P., Frank, A., Pevzner, P., & Shevchenko, A. (2006) *J. Proteome Res.* **5**, 2448–2456
- (96) Darewicz, M., Dziuba, J., & Minkiewicz, P. (2007) *Food Sci. Technol. Int.* **13**, 125–133
- (97) Wallace, I.M., Blackshields, G., & Higgins, D.G. (2005) *Curr. Opin. Struct. Biol.* **15**, 261–266
- (98) Notredame, C. (2007) *PLoS Comput. Biol.* **3**, 1405–1408
- (99) Pearson, W.R., & Sierk, M.L. (2005) *Curr. Opin. Struct. Biol.* **15**, 254–260
- (100) Baggerman, G., Liu, F., Wets, G., & Schoofs, L. (2005) *Ann. NY Acad. Sci.* **1040**, 59–65
- (101) Liu, F., Baggerman, G., D’Hertog, W., Verleyen, P., Schoofs, L., & Wets, G. (2006) *Mol. Cell. Proteom.* **5**, 510–522

- (102) Liu, F., Baggerman, G., Schoofs, L., & Wets, G. (2006) *Peptides* **27**, 3137–3154
- (103) Simossis, V.A., & Heringa, J. (2004) *Curr. Prot. Pept. Sci.* **5**, 249–266
- (104) Dunbrack, R.L. (2006) *Curr. Opin. Struct. Biol.* **16**, 374–384
- (105) Thomas, A., Deshayes, S., Decaffmeyer, M., Van Eyck, M.H., Charlotiaux, B., & Brasseur, R. (2006) *Proteins* **65**, 889–897
- (106) Kazemian, M., Moshiri, B., Nikbakht, H., & Lucas, C. (2007) *Comput. Biol. Chem.* **31**, 44–47
- (107) Nakai, S., Alizadeh-Pasdar, N., Dou, J., Buttinov, R., Roussou, D., & Poulson, A. (2004) *J. Agric. Food Chem.* **52**, 927–935
- (108) Dobson, P.D., Cai, Y.D., Stapley, B.J., & Doig, A.J. (2004) *Curr. Med. Chem.* **11**, 2135–2142
- (109) Bryliński, M., Kochańczyk, M., Konieczny, L., & Roterman, I. (2006) *In Silico Biol.* **6**, 0052
- (110) Bryliński, M., Prymula, K., Jurkowski, W., Kochańczyk, M., Stawowczyk, E., Konieczny, L., & Roterman, I. (2007) *PLoS Comput. Biol.* **3**, 909–923
- (111) Mallick, P., Schirle, M., Chen, S.S., Flory, M.R., Martin, D., Ranish, J., Raught, B., Schmitt, R., Werner, T., Kuster, B., & Aebersold, R. (2007) *Nat. Biotechnol.* **25**, 125–131
- (112) Zhang, C., Bickis, M.G., Wu, F.-X., & Kusalik, A.J. (2006) *J. Bioinform. Comput. Biol.* **4**, 959–980
- (113) Henschen, A., Lottspeich, F., Brantl, V., & Teschemacher, H. (1979) *Hoppe Seyler's Z. Physiol. Chem.* **360**, 1217–1224
- (114) McDonnell, M., Bouchier, P., FitzGerald, R.J., & O'Cuinn, G. (1999) *J. Dairy Res.* **66**, 257–270
- (115) Tiruppathi, C., Miyamoto, Y., Ganapathy, V., Roesel, R.A., Whitford, G.M., & Leibach, F.H. (1990) *J. Biol. Chem.* **265**, 1476–1483
- (116) Miyamoto, Y., Ganapathy, V., Barlas, A., Neubert, K., Barth, A., & Leibach, F.H. (1987) *Am. J. Physiol. Renal Physiol.* **252**, F670–F677
- (117) Shinoda, I., Tada, M., Okai, H., & Fukui, S. (1986) *Agric. Biol. Chem.* **50**, 1247–1254
- (118) Haileselassie, S.S., Lee, B.H., & Gibbs, B.F. (1999) *J. Dairy Sci.* **82**, 1612–1617
- (119) Lee, H.S., & Lee, K.J. (2000) *Peptides* **21**, 807–809
- (120) Meisel, H. (1998) *Int. Dairy J.* **8**, 363–373
- (121) Meisel, H. (1986) *FEBS Lett.* **196**, 223–227
- (122) Asano, M., Nio, N., & Ariyoshi, Y. (1992) *Biosci. Biotech. Biochem.* **56**, 976–977
- (123) Abubakar, A., Saito, T., Kitazawa, H., Kawai, Y., & Itoh, T. (1998) *J. Dairy Sci.* **81**, 3131–3139
- (124) Saito, T., Nakamura, T., Kitazawa, H., Kawai, Y., & Itoh, T. (2000) *J. Dairy Sci.* **83**, 1434–1440
- (125) Yokoyama, K., Chiba, H., & Yoshikawa, M. (1992) *Biosci. Biotech. Biochem.* **56**, 1541–1545
- (126) Van Wagoner, R.M., Jacobsen, R.B., Olivera, B.M., & Ireland, C.M. (2003) *Biochemistry* **42**, 6353–6362
- (127) Morita, H., Shishido, A., Kayashita, T., Shimomura, M., Takeya, K., & Itokawa, H. (1994) *Chem. Lett.* **12**, 2415–2418
- (128) Pripp, A.H. (2005) *Eur. Food Res. Technol.* **221**, 712–716
- (129) Wu, J.P., Aluko, R.E., & Nakai, S. (2006) *J. Agric. Food Chem.* **54**, 732–738
- (130) Wu, J.P., Aluko, R.E., & Nakai, S. (2006) *QSAR Comb. Sci.* **25**, 873–880
- (131) Pripp, A.H. (2006) *J. Agric. Food Chem.* **54**, 224–228
- (132) Fernández, M., & Caballero, J. (2007) *Mol. Simulat.* **33**, 689–702
- (133) Kim, H.O., & Li-Chan, E.C.Y. (2006) *J. Agric. Food Chem.* **54**, 10102–10111
- (134) Wu, J.P., & Aluko, R.E. (2007) *J. Peptide Sci.* **13**, 63–69
- (135) Pripp, A.H., & Ardö, Y. (2007) *Food Chem.* **102**, 880–888
- (136) Cui, J., Han, L., Lin, H., Tang, Z., Ji, Z., Cao, Z., Li, Y., & Chen, Y. (2007) *Curr. Bioinform.* **2**, 95–112
- (137) Tarca, A.L., Carey, V.J., Chen, X.-W., Romero, R., & Drăgici, S. (2007) *PLoS Comput. Biol.* **8**, 953–963
- (138) Yagi, Y., Terada, K., Noma, T., Ikebukuro, K., & Sode, K. (2007) *BMC Bioinform.* **8**, 11
- (139) Jung, E., Kim, J., Kim, M., Jung, D.H., Rhee, H., Shin, J.-M., Choi, Y.-J., & Choi, S.-H. (2007) *BMC Bioinform.* **8**, 245
- (140) Tiwari, A., & Sekhar, A.K.T. (2007) *Comput. Biol. Chem.* **31**, 305–319