ISSN 1330-9862
(FTB-2424)

scientific note

BIOPEP-PBIL Tool for the Analysis of the Structure of Biologically Active Motifs Derived from Food Proteins

Anna Iwaniak and Jerzy Dziuba*

University of Warmia and Mazury in Olsztyn, Chair of Food Biochemistry, Pl. Cieszynski 1, PL-10-726 Olsztyn-Kortowo, Poland

Received: January 21, 2010
Accepted: May 26, 2010

Summary

This work describes a flexible technique for the analysis of protein sequences as a source of motifs affecting bodily functions. The BIOPEP database, along with the Pôle Bioinformatique Lyonnais (PBIL) server, were applied to define which activities of peptides dominated in their protein precursors and which structure of the protein contained the most of the revealed activities. Such an approach could be helpful in finding some structural requirements for peptide(s) to be regarded as biologically active (bioactive). It was found that apart from the activities of peptides that commonly occur in the majority of proteins (*e.g.* ACE inhibitors), all analyzed proteins can be a source of motifs involved in *e.g.* activation of ubiquitin-mediated proteolysis. This could be important in designing diets for patients who suffer from neural diseases. The structure and bioactivity analyses revealed that if peptides were to be 'bioactive', it is essential that they assume the position of a coil (or combination of coil and α-helix) in the sequence of their protein precursors. However, it is recommended to consider the factors such as the length of peptide chains, the number of peptides in the database as well as the repeatability of the occurrence of characteristic amino acids, both in the peptide and in the protein when studying the bioactivity and structure of biomolecules.

Key words: databases, proteins, bioactive peptides, bioinformatics

Introduction

Food-derived bioactive proteins are considered to be components that reduce the risk of disease or enhance a certain physiological function (*1*). Peptides encrypted in the protein molecules are believed to attribute the physicochemical and sensory functions to proteins, which are also regarded as functional and health-promoting ingredients. There are many activities of peptides described in the literature, *e.g.* antioxidative, antimicrobial, antithrombotic, opioid or immunomodulating as well as the inhibition of ACE (EC 3.4.15.1), involved in blood pressure reduction (*2*). Data concerning the peptide activities are often available in databases, *e.g.* BIOPEP, SwePep, or ANTIMIC (*3*). Such databases are interconnected with other computer resources that allow for the identification of protein function based on its secondary structure prediction (*4*), homology level (*5*) or topological features (*6*). According to Blythe *et al.* (*7*) databases are the *lingua franca* of bioinformatics – a discipline created to understand the biological processes by applying specially designed mathematical and statistical algorithms combined with information technology (*8*).

The properties of molecules, including peptides, are described by their structure as well as physicochemical or pharmacological attributes defined as descriptors. They can be computed based on the structure of the molecule, *e.g.* molecular mass, charge, solubility (molecular descriptors) or derived by means of multivariate statistics (pharmacological descriptors) (*9*). These descriptors (if expressed in numbers) are applicable in the QSAR (quantitative structure–activity relationship) approach, which can be described as the mathematical function:

$$P=f(x_1, x_2,\ldots x_n) \qquad /1/$$

where P means molecular property (activity) of a molecule and x_1, x_2,…x_n are the *n* molecular descriptors (*10*).

*Corresponding author; Phone: ++48 89 523 3722; Fax: ++48 89 523 4945; E-mail: ami@uwm.edu.pl

The above-mentioned technique has successfully been applied in drug design, bioactivity prediction and biological response of chemical compounds (*11*). In food science, QSAR is used as a method to discover the relationships between the structure and activity of ACE inhibitors and bitter taste peptides when, in both cases, hydrophobicity was the main factor affecting the activity of peptides (*12, 13*).

The protein sequence can also be evaluated as a good/bad precursor of peptides with biological activity (*14*). Such an evaluation can be performed by the application of discriminants developed by Dziuba and Iwaniak (*15*). They are available in the BIOPEP database (*16*). It is well-known that the activity of peptide depends on its amino acid composition and structure (*17*). Due to the fact that the sequences of bioactive peptides can be easily retrieved from any database, this study focuses on the activities of peptides that dominate in the proteins and the structure of the protein that is the most 'friendly' for acquiring the most of the revealed activities. Such an approach could be helpful in finding some structural requirements for peptides, which are regarded as biologically active (bioactive).

Materials and Methods

The detailed procedure of data selection and analysis is shown in Fig. 1. Briefly, twenty sequences of food proteins were taken from the BIOPEP database (bar named 'Proteins') at *www.uwm.edu.pl/biochemia* (*16,18*). These proteins represent food components which are commonly consumed by the population, *i.e.* milk, meat, fish, egg, crops, seeds and leguminous plants. The full names of the analyzed proteins are shown in Fig. 2. To find the bioactivities occurring in the above-mentioned proteins, discriminant *A* was calculated, which is defined as the occurrence frequency of the fragments with a given activity in a polypeptide chain described by the equation given below:

$$A=\frac{a}{N} \qquad /2/$$

where *a* is the number of fragments with given activity in the protein chain, and *N* is the number of amino acid residues in the protein molecule (*19*).

The secondary structures of proteins were predicted by the use of GOR1 algorithm available at Pôle Bioinformatique Lyonnais (PBIL) server (*20*). Pôle Bioinformatique Lyonnais provides many tools to analyze protein and peptide sequences. It is helpful to predict *e.g.* the secondary structures of a protein based on its amino acid composition (*21*).

The last stage of the studies (see 'Results' in Fig. 1) was to identify particular sequences of bioactive fragments with their locations in the protein. It was performed by running the function called 'profile of the potential biological activity of proteins', *i.e.* type and location of bioactive fragments in a protein chain (*19*). Based on the match of the peptide(s) to the appropriate secondary structure(s) present in the protein, it was then possible to try to find some structural requirements for motifs to be defined as bioactive.

Although the profiles of potential biological activity were performed for all protein sequences, this paper presents only one of them generated for carp (*Cyprinus carpio*) parvalbumin (Table 1). Both the profile of potential biological activity of a protein and the frequency of occurrence of fragments with a given activity are the equivalent criteria of protein evaluation in the context of being a good or bad source of biopeptides (*19*).

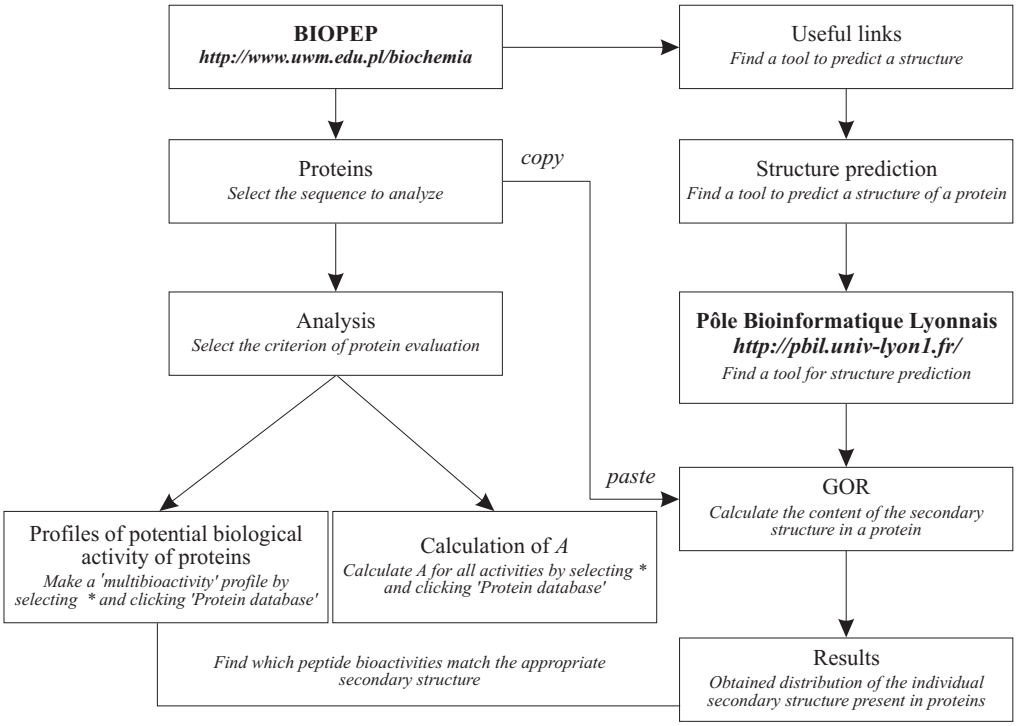

Fig. 1. Data analysis procedure combining BIOPEP database and Pôle Bioinformatique Lyonnais server

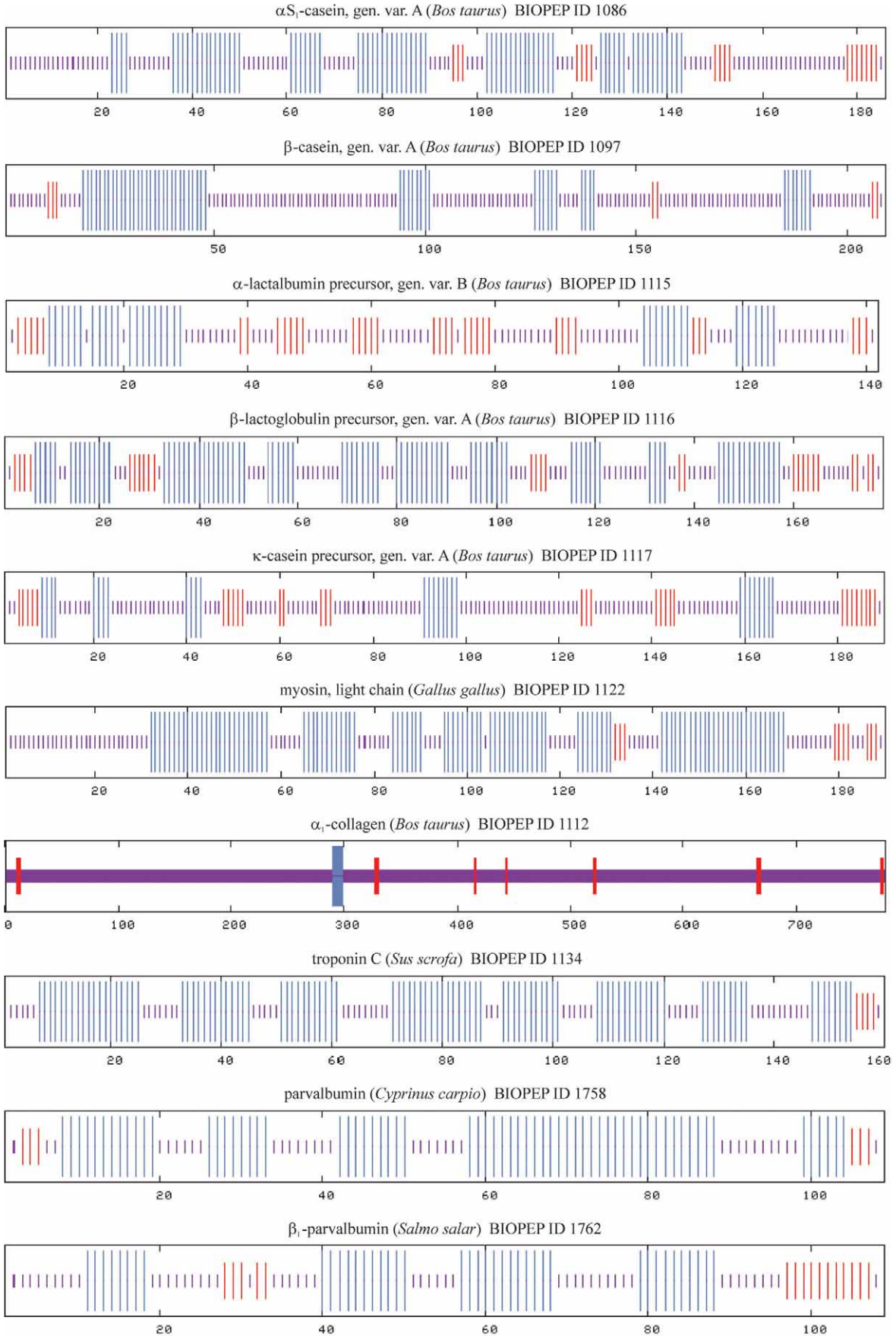

Fig. 2. Distribution of the secondary structures in protein chains using the Pôle Bioinformatique Lyonnais server blue – α-helix, red – extended strand, violet – random coil

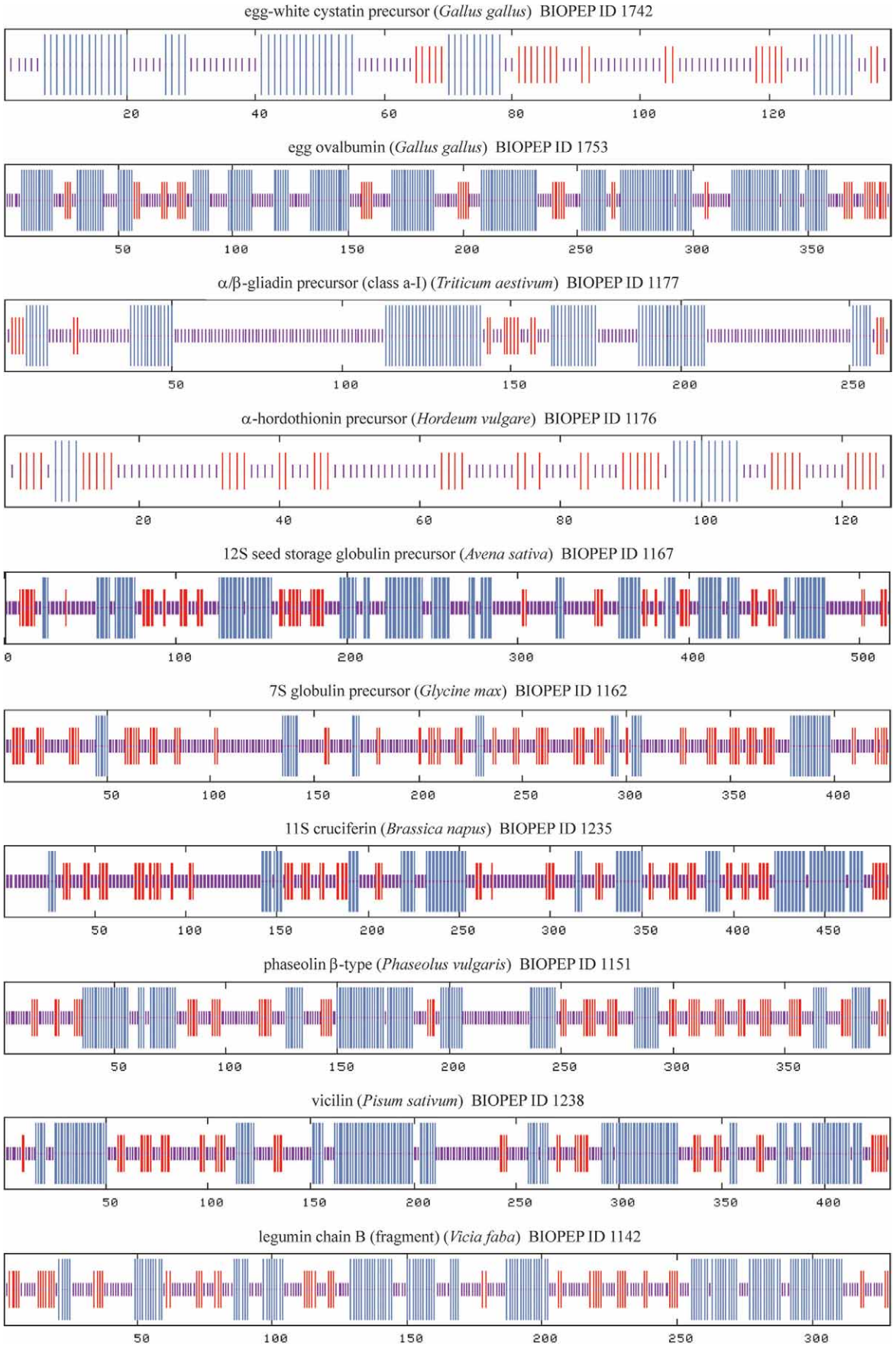

Fig. 2. – continued: blue – α-helix, red – extended strand, violet – random coil

Table 1. An example of the biological activities revealed in the peptide profile of carp (*Cyprinus carpio*) parvalbumin. Analysis performed by the use of BIOPEP database

Activity	Sequence and its location in a protein chain
ACE inhibitor	**LKL** [64-66]; **LF** [66-67]; **VK** [107-108]; **LKA** [87-89]; **AF** [2-3],[47-48],[85-86]; **RA** [76-77]; **AA** [14-15],[21-22],[104-105]; **GF** [57-58]; **VG** [34-35]; **IG** [98-99]; **GI** [5-6]; **GA** [74-75]; **GL** [35-36]; **AG** [4-5],[73-74],[89-90]; **DA** [9-10],[80-81]; **GV** [99-100]; **GK** [96-97]; **QG** [17-18]; **SG** [56-57]; **GD** [90-91],[94-95]; **DG** [93-94], [95-96]; **NF** [70-71]; **SF** [24-25],[29-30]; **KL** [65-66]; **YK** [27-28]; **AR** [75-76]; **KA** [46-47],[84-85],[88-89],[108-109]; **IE** [59-60]; **LQ** [16-17],[68-69]; **LN** [7-8]
Dipeptidyl peptidase-IV inhibitor	**MA** [1-2]; **KA** [46-47],[84-85],[88-89],[108-109]; **FA** [3-4],[31-32],[48-49],[103-104]
Bacterial permease ligand	**KK** [45-46]
Regulation of ion flow	**DY** [26-27]
Activation of ubiquitin-mediated proteolysis	**RA** [76-77]

Results and Discussion

The values of frequency of the occurrence of fragments with the given activity (*A*) for all analyzed sequences are shown in Table 2. Nineteen biological activities can be distinguished in the protein sequences taken for the analysis. Six of them, *i.e.* antibacterial, coeliac-toxic, chemotactic, contracting smooth muscles, neuropeptide

Table 2. The values of frequency of the occurrence (*A*) of fragments with a given activity in the analyzed food proteins

Activity	Protein/***A* value**
ACE inhibitor (antihypertensive)	αS_1-casein/**0.462**; β-casein/**0.545**; α-lactalbumin/**0.317**; β-lactoglobulin/**0.421**; κ-casein/**0.363**; myosin/**0.395**; α_1-collagen/**0.874**; troponin C/**0.375**; carp parvalbumin/**0.404**; salmon β_1-parvalbumin/**0.376**; egg-white cystatin/**0.302**; egg ovalbumin/**0.340**; α/β-gliadin precursor/**0.29**; α-hordothionin precursor/**0.315**; 12S seed storage globulin precursor/**0.315**; 7S globulin precursor/**0.315**; 11S cruciferin/**0.356**; phaseolin, β-type/**0.312**; vicilin/**0.310**; legumin B chain/**0.331** ***total=20***
Activation of ubiquitin-mediated proteolysis	αS_1-casein/**0.005**; α-lactalbumin/**0.007**; β-lactoglobulin/**0.022**; κ-casein/**0.005**; myosin/**0.011**; α_1-collagen/**0.874**; troponin C/**0.019**; carp parvalbumin/**0.009**; egg-white cystatin/**0.022**; egg ovalbumin/**0.008**; α/β-gliadin precursor/**0.012**; α-hordothionin precursor/**0.016**; 12S seed storage globulin precursor/**0.315**; 7S globulin precursor/**0.012**; 11S cruciferin/**0.014**; phaseolin, β-type/**0.010**; vicilin/**0.009**; legumin B chain/**0.018** ***total=18***
Antiamnestic	αS_1-casein/**0.005**; β-casein/**0.043**; α_1-collagen/**0.267**; egg ovalbumin/**0.003**; α/β-gliadin precursor/**0.004**; 12S seed storage globulin precursor/**0.006**; 7S globulin precursor/**0.006**; 11S cruciferin/**0.006**; phaseolin, β-type/**0.002**; vicilin/**0.070**; legumin B chain/**0.003** ***total=11***
Antioxidative	αS_1-casein/**0.005**; β-casein/**0.014**; α-lactalbumin/**0.007**; κ-casein/**0.032**; salmon β_1-parvalbumin/**0.009**; egg ovalbumin/**0.008**; α/β-gliadin precursor/**0.008**; 12S seed storage globulin precursor/**0.006**; 7S globulin precursor/**0.006**; phaseolin, β-type/**0.005**; vicilin/**0.002** ***total=11***
Antithrombotic	β-casein/**0.029**; κ-casein/**0.021**; α_1-collagen/**0.298**; egg ovalbumin/**0.003**; 12S seed storage globulin precursor/**0.006**; 7S globulin precursor/**0.006**; 11S cruciferin/**0.006**; phaseolin, β-type/**0.002**; vicilin/**0.007**; legumin B chain/**0.331** ***total=10***
Bacterial permease ligand	αS_1-casein/**0.005**; β-casein/**0.005**; α-lactalbumin/**0.007**; β-lactoglobulin/**0.012**; κ-casein/**0.005**; myosin/**0.016**; carp parvalbumin/**0.009**; salmon β_1-parvalbumin/**0.009**; 12S seed storage globulin precursor/**0.002**; vicilin/**0.002** ***total=10***
Immunomodulating	αS_1-casein/**0.005**; β-casein/**0.014**; α-lactalbumin/**0.021**; κ-casein/**0.010**; α_1-collagen/**0.001**; phaseolin, β-type/**0.312** ***total=6***
Inhibition of dipeptidyl peptidase-IV	αS_1-casein/**0.069**; β-casein/**0.148**; α-lactalbumin/**0.042**; β-lactoglobulin/**0.090**; κ-casein/**0.089**; myosin/**0.112**; α_1-collagen/**0.268**; troponin C/**0.025**; carp parvalbumin/**0.083**; salmon β_1-parvalbumin/**0.073**; egg-white cystatin/**0.094**; egg ovalbumin/**0.049**; α/β-gliadin precursor/**0.072**; α-hordothionin precursor/**0.024**; 12S seed storage globulin precursor/**0.075**; 7S globulin precursor/**0.075**; 11S cruciferin/**0.086**; phaseolin, β-type/**0.050**; vicilin/**0.062**; legumin B chain/**0.061** ***total=20***
Opioid	αS_1-casein/**0.027**; β-casein/**0.009**; α-lactalbumin/**0.021**; β-lactoglobulin/**0.006**; κ-casein/**0.016**; α_1-collagen/**0.001**; troponin C/**0.006**; egg ovalbumin/**0.010**; α/β-gliadin precursor/**0.012**; α-hordothionin precursor/**0.008**; 11S cruciferin/**0.002**; phaseolin, β-type/**0.007**; vicilin/**0.005**; legumin B chain/**0.009** ***total=14***
Regulation of a stomach mucosal membrane activity	β-casein/**0.029**; α_1-collagen/**0.267**; 12S seed storage globulin precursor/**0.006**; 7S globulin precursor/**0.006**; 11S cruciferin/**0.008**; phaseolin, β-type/**0.002**; vicilin/**0.007**; legumin B chain/**0.003** ***total=8***
Regulation of ion flow	β-lactoglobulin/**0.006**; carp parvalbumin/**0.009**; α-hordothionin precursor/**0.008**; phaseolin, β-type/**0.002**; vicilin/**0.002** ***total=5***
Other activities*: antibacterial (1), coeliac-toxic (1), chemotactic (1), contracting smooth muscles (1), neuropeptide (1), stimulating γ-interferon production (1), embryotoxic (2), regulating phosphoinositole mechanism (3)	

*Other activities revealed in some of the analyzed proteins. The numbers in parentheses indicate the total number of the protein sequences in which the given activity was found

and stimulating γ-interferon production occurred only in α-lactalbumin, α/β-gliadin precursor (coeliac-toxic and stimulating γ-interferon production), α_1-collagen, κ-casein and β-lactoglobulin, respectively. The relatively low values of *A* calculated for these activities reflect their small number of fragments (maximum 3) that occur in the proteins, which it is an obstacle when discussing the structural character of the motif in the context of its bioactivity. The highest value of *A* (0.244) among the above-mentioned activities was obtained for coeliac-toxic activity in the α/β-gliadin precursor. Some grain proteins, including wheat, possess motifs that initiate pathophysiological processes, which are a factor affecting intestinal epithelial damage. Wheat gliadins contain fragments released during the protein digestion and lead to coeliac disease. These fragments have characteristic amino acids such as proline, tyrosine and glutamine, responsible for coeliac disease and they are primarily located in the N-terminal domain of the gliadin chain (*22*).

All twenty analyzed protein sequences showed the presence of ACE and dipeptidyl peptidase-IV (DPP-IV) inhibitors. Motifs with the ACE inhibitory activity are the best represented group among the sequences of food proteins (*23*). BIOPEP analysis revealed that the best precursor of ACE inhibitors can be bovine collagen with the *A* value of 0.874. It is due to the length of the chain and the higher probability of the occurrence of bioactive motif. ACE peptide inhibitors were identified in many food raw materials and products such as fish, milk proteins and cereals (*24*). Milk proteins derived from different species are a rich source of ACE inhibitors (*25,26*). Iwaniak and Dziuba (*2*) compared the profiles of potential ACE inhibitory activity of milk proteins and revealed the presence of many fragments with such an activity (*e.g.* from 45 in case of bovine α-lactalbumin to 121 in genetic variant B of β-casein). The impact of the structure of ACE inhibitors on their activity depends on the C-terminal amino acid residue. The presence of tyrosine (Tyr), phenylalanine (Phe), tryptophan (Trp) and proline (Pro) at the C-terminus of a peptide contributes to its higher activity. Tryptophan exerts the strongest influence on the ACE inhibitory activity of peptides. The conformation of the peptide is also an important factor affecting the efficiency of the interactions of ACE inhibitors (*25*).

Dipeptidyl peptidase-IV (DPP-IV) inhibitors were revealed in all analyzed proteins. DPP-IV (EC 3.4.14.5) is an enzyme belonging to the serine proteases (*27*) and is responsible for the breakdown of glucagon-like peptide (GLP-1) and gastric inhibitory polypeptide (GIP). Inhibition of DPP-IV prolongs the *in vivo* half-life of GLP-1 and GIP, leads to the enhancement of insulin secretion and improves glucose tolerance, which makes DPP-IV inhibitors valuable in treating type 2 diabetes (*27,28*). Food containing proteins with DPP-IV inhibitors could represent a good diet for patients suffering from diabetes. The best sources of peptides with this activity were β-casein, α_1-collagen and myosin (*A* values were: 0.148, 0.268 and 0.112, respectively).

All twenty sequences taken for analysis that showed the presence of fragments with the above-mentioned activities, *i.e.* inhibitors of ACE as well as DPP-IV, can be found in BIOPEP. Generally, peptides in the BIOPEP database described as 'inhibitors' are the most numerous (671 in total). Currently, the number of ACE inhibitors (peptides involved in blood pressure reduction) is 530 and DPP-IV inhibitors is 11. The remaining 130 peptides listed in the database concern other kinds of inhibitors which were not found in the sequences of proteins taken for analysis.

Although the BIOPEP database contains only eleven motifs acting as dipeptidyl peptidase-IV inhibitors (compared to the number of ACE inhibitors), they occurred in all the analyzed proteins because of the structural similarities between them and ACE inhibitors. The similarity relied on the occurrence of proline residues in the sequences of ACE and DPP-IV inhibitors. As described above, proline is one of the crucial amino acid residues involved in ACE inhibitory activity (*25*) and also it seems important in DPP-IV inhibitory activity. Most DPP-IV inhibitors are called proline-like compounds (*27*).

The *in silico* studies showed that all food-derived protein sequences can be a potential source of peptides responsible for the ubiquitin-mediated proteolysis (see Table 2). Proteins are mainly a source of amino acids, which are essential for cell maintenance and growth of human body (*29*). Due to the ageing processes or other factors affecting their proper functioning, these proteins can cause serious damage. A small molecule peptide called ubiquitin recognizes such proteins and controls the process of their degradation (*30*). Ubiquitin is a 76-residue polypeptide which exists in all eukaryotic cells. The characteristic motif of this protein is seven lysine residues at the C-terminus. The C-terminal glycine of ubiquitin is activated by ATP to a high-energy intermediate during a reaction catalyzed by a ubiquitin-activating enzyme (*31*). Research on ubiquitin-mediated proteolysis can lead to a better understanding of the pathogenesis of neural diseases such as Alzheimer's or Parkinson's (*32*) as well as elucidation of the mechanism of regulation and degradation of specific proteins defined as oncoproteins (*31*).

Twelve proteins possessed motifs with anti-amnestic activity. Peptides with this activity are prolyl endopeptidase (PEP) inhibitors (*33*). They are involved in learning and memory abilities and are found in food proteins, *e.g.* corn and rice. A typical motif for inhibitors of prolyl endopeptidases is repetitive proline residue (*34*). The highest value of frequency of the occurrence of anti-amnestic fragments was obtained for bovine collagen (*A*=0.267). This protein is built up from 779 residues, with many prolines that match the sequences of PEP inhibitors (anti-amnestic peptides).

The frequency of the occurrence of fragments with antioxidative activity was obtained for 11 protein sequences. The values of this parameter were relatively low (the highest value of *A*=0.014 was for β-casein). The main components of the antioxidative peptides derived from different proteins are mainly the amino acids known for their antioxidative properties, such as histidine or tyrosine. Similar properties can also be assigned to methionine, lysine and tryptophan. The activity of these peptides affects their primary structure as well as configuration (*34*). The presence of hydrophobic residues fosters the interaction of peptides with linoleic acid and the num-

ber of prolines (the more, the better) contributes to an increase in the antioxidant activity of the peptide (*34*).

Bovine collagen was the best source of peptides with antithrombotic activity (A=0.298) amongst the 10 protein sequences, which also contained fragments with the above-mentioned activity. The effect of peptides derived from porcine skin collagen on fibrin polymerization and platelet aggregation was studied by Nonaka *et al.* (*35,36*). They found that peptides with the G-protein regulatory (GPR) motif, including their analogues such as GPRG, GPRGP, GPRPP and GPRPPP, suppressed the platelet aggregation in humans. The occurrence of proline and glycine is typical for fragments with an anticoagulant effect. These amino acid residues are present in the C-terminal sequence of fibrin α-chain and are responsible for its polymerization. One of the milk proteins – casein – is also known as a good source of anticoagulant fragments. In this study, β- and κ-casein had relatively high values for the frequency of the occurrence of antithrombotic fragments (A=0.029 and A=0.021, respectively). The presence of motifs with antithrombotic activity results from the fact that κ-casein and γ-fibrin evolved from a common ancestor. This is confirmed by the similarities in the mechanisms of chymosin-induced milk and thrombin-induced blood coagulation (*37*).

Twelve proteins (of all twenty studied) showed the presence of peptides which are ligands of bacterial permeases. Oligopeptide permeases are responsible for the uptake of peptides in some species of bacteria and archaea. These permeases are involved in many processes such as: cell wall synthesis, adhesion to host cells and proteins, gene regulation of intercellular signalling, competence development and sporulation. One of the components of oligopeptide permease is periplasmic peptide-binding component (OppA). OppAs can bind oligopeptides of different sizes: from three (*e.g. Salmonella typhimurium* OppA orthologues) to eighteen (*e.g. Lactococcus lactis* OppA paralogues). The binding activity of OppAs of bacterial species may affect cell envelope permeability restrictions and the nutritional strategy required for survival in a specific environment (*38*).

The majority of proteins (fourteen out of twenty) shown in Table 2 were relatively rich in opioid fragments. According to Meisel and FitzGerald (*39*), many opioid peptides are very potent – even insignificant amounts of released peptides exert a physiological effect. Opioid peptides are rich in tyrosine. Tyr is the characteristic residue for endogenous opiates such as enkephalin, dynorphin and endorphin. They all possess a YGGF motif located in the N-terminal part of the chain (*35*). The presence of Tyr and Phe at the third or fourth position fits to the binding sites of opioid receptors. Removal of tyrosine is the cause of the lack of bioactivity (*39*). The presence of proline at the second position favours opioid activity due to the maintenance of the proper orientation of the Pro- and Tyr-side chains (*39*).

Some peptides possess several bioactivities. For example, glycylproline peptides (peptides assigned to the glyproline family) are responsible for antithrombotic effects and the protection of gastric mucosa against various ulcerogenic factors as well as having some anorectic actions (*38*). Eight proteins, *i.e.* β-casein; α_1-collagen; 12S seed storage globulin precursor; 7S globulin precursor; 11S cruciferin; phaseolin, β-type; vicilin and legumin B chain showed the presence of motifs with an activity defined as the regulation of stomach mucosa membrane. The above-mentioned proteins appear to be a source of antithrombotic peptides (see Table 2). According to Samonina *et al.* (*40*), glyprolines can be formed from the precursors like collagen and elastin. Peptides with regulatory activities are usually liberated from specialized precursors, but there is data concerning their formation from various unspecific precursors (*40*). The bioactivity of peptides derived from specific precursors are active in very low concentrations (the highest floor), and those derived from non-specific sources have moderate bioactivity (the lowest floor). Such division of regulatory peptides was described as the 'multi-floor peptide regulatory system' by Karelin *et al.* (*41*).

Some of the analyzed proteins showed that they could be a potential source of peptides with immunomodulating effect and for regulation of ion flow. Peptides with immunomodulating activity occurred in milk proteins as well as in α-collagen and β-phaseolin. The values of frequency of the occurrence of fragments with immunomodulating activity for milk proteins were relatively low (from 0.005 to 0.021). The best source of these fragments was phaseolin (A=0.312). Peptides regulating ion flows were observed only in six proteins and the A values for this activity were low. The explanation is that BIOPEP contains only two peptides with this activity. Based on such limited information concerning the peptides that regulate ion flow, it is hard to discuss the structural requirements of their bioactivity.

Fig. 2 shows the distribution of the individual secondary structures performed for the BIOPEP-generated protein sequences by means of the PBIL server (*12*). The secondary structures of proteins were predicted by the GOR (Garnier-Osguthorpe-Robson) method (*42*). The GOR algorithm combines both information theory and Bayesian statistics for predicting the secondary structure of a molecule (*43*).

As mentioned above, all analyzed proteins revealed the presence of the inhibitors of angiotensin-converting enzyme (ACE) and dipeptidyl peptidase-IV (DPP-IV). Great majority of these peptides matched the random coiled part of the protein chains. Based on the profiles of potential biological activity of proteins (see Materials and Methods), in the case of the porcine troponin C, most of the ACE inhibitors were situated in the α-helical structure of the protein (over 30 fragments), and a coil was revealed in fourteen peptides matching this structure. In the case of the structure of DPP-IV inhibitors – their structure was more diversified. In nine out of twenty proteins, peptides had a random coil structure (β- and κ-caseins, collagen, globulins 12S and 7S from oat and soybean, rapeseed cruciferin, kidney bean phaseolin, garden pea vicilin and fava bean legumin). A combination of helices and coils appeared in peptides present in α-lactalbumin, β-lactoglobulin, chicken myosin, porcine troponin, carp and salmon parvalbumins, egg cystatin and ovalbumin, and α/β-wheat gliadin. DPP-IV inhibitors found in α-casein and barley hordothionin were a combination of coils, extended strands and α-helices.

Almost all protein sequences (see Table 2), excluding β-casein and β-parvalbumin, can be regarded as a potential source of fragments activating ubiquitin-mediated proteolysis. They were located in the parts of the protein chains containing α-helix, random coils and extended strands. The frequency of the occurrence of fragments possessing this bioactivity (*A*) was from 0.005 for α-casein to 0.018 for fava bean legumin. The value of *A* calculated for specific bioactivity depends on the number of peptides gathered in the BIOPEP database (the smaller number of peptides, the lower the value of *A*). BIOPEP database contains only three peptides that are responsible for the activation of ubiquitin-mediated proteolysis (RA, LA, WA) and that may explain the relatively small values of frequency of the occurrence of fragments with the above-mentioned activity. The existence of a small number of these peptides in the proteins, as well as their random occurrence in them (from one to a maximum of three in the whole protein chain) does not provide any clues as to which type of the peptide structure might be responsible for the activation of ubiquitin-mediated proteolysis (due to the 'coincidental' locations of the peptides in different structures of the protein chains). The same line of reasoning could be applied to all proteins found as precursors of peptides defined as bacterial permease ligands (three of them in the BIOPEP; their sequences are: KK, KKK, KKKA), regulators of ion flow (two peptides in BIOPEP: DY and TSLYR), and regulators of phosphoinositole mechanism (five peptides in BIOPEP: GFW, GFL, LGY, GLY and GLF) (*16*).

Different types of structures were found in fragments with opioid and antibacterial activities, or with mechanisms of ion flow and phosphoinositole regulation. In the case of opioid and antibacterial effects of peptides present in some of the analyzed proteins, their number in BIOPEP was high enough (158 and 419, respectively) to discuss the structure and bioactivity relationship. However, some of them were built up from *e.g.* 17 (YGGFLRARKSARKLANQ) to 34 amino acid residues (GICACRRRFCPNSERFSGYCRVNGARYVRCCSRR) or occurred as C-terminal amides, which presented an obstacle in matching the protein sequence for identification according to the profile of the potential biological activity of the protein (*16*,*20*).

The information obtained from BIOPEP using PBIL concerning the bioactivities of peptides related to particular secondary structures of the proteins is summarized in Table 3. Peptides which represent one activity were first summarized and then put in three groups occupying one of the three secondary structures. The mean percentages of peptides matching the individual secondary structure were then calculated for all peptide activities.

Some of the biopeptides completely matched the random coil parts of the precursor protein chain (Table 3). These fragments represented the following activities: antiamnestic, antithrombotic, coeliac-toxic, responsible for the action of stomach mucosa membrane, chemotactic, neuropeptide, stimulating production of γ-interferon, involved in the smooth muscle contraction and embryotoxic activities. Fragments representing the last five activities appeared in one or a maximum of two proteins and the maximum number of peptides present in these proteins ranged from one to three. Coeliac-toxic peptides were identified only in α/β-wheat gliadin precursor (over sixty of them). Although coeliac-toxic peptides are relatively long, they possess some common motifs such as proline and glutamine – amino acids that are repetitive in the whole protein sequence and are mostly located in the coiled structure of wheat gliadin. Antiamnestic, antithrombotic and immunomodulating fragments which fully matched

Table 3. The level of structural matching of peptides calculated for all activities of the analyzed proteins

Activity	Level of structural matching of peptides/%			Total number of peptides
	Helix	Extended-strand	Random coil	
ACE inhibitor (antihypertensive)	19.58 (333*)	2.78 (5)	77.64 (1362)	1700
Activation of ubiquitin-mediated proteolysis	76.19 (48)	19.04 (12)	4.41 (3)	63
Antiamnestic	0 (–)	0 (–)	100 (232)	232
Antibacterial	40.00 (2)	20.00 (1)	40.00 (2)	5
Antioxidative	51.85 (14)	7.41 (2)	40.74 (11)	27
Antithrombotic	0 (–)	0 (–)	100 (258)	258
Bacterial permease ligand	53.85 (7)	0 (–)	46.15 (6)	13
Coeliac-toxic	0 (–)	0 (–)	100 (64)	64
Chemotactic	0 (–)	0 (–)	100 (1)	1
Contraction of smooth muscle	0 (–)	0 (–)	100 (2)	2
Embryotoxic	0 (–)	0 (–)	100 (2)	2
Immuno-modulation	0 (–)	9.10 (1)	90.90 (10)	11
Inhibition of dipeptidyl peptidase–IV	9.01 (45)	0.81 (40)	90.18 (414)	499
Neuropeptide	0 (–)	0 (–)	100 (1)	1
Opioid	30.30 (10)	6.07 (2)	63.63 (21)	33
Regulation of a stomach mucosal membrane activity	0 (–)	0 (–)	100 (229)	229
Regulation of ion flow	33.33 (2)	16.67 (1)	50.00 (3)	6
Regulation of phospho-inositole mechanism	33.33 (1)	33.33 (1)	33.33 (1)	3
Stimulation of γ-interferon production	0 (–)	0 (–)	100 (3)	3

*number of peptides found in the particular structure of the protein chain

the coiled structure were relatively short with repetitive prolines and glycines. This suggests the probability of their frequent occurrence in the appropriate protein structure. This rule can be applied to ACE and DPP-IV inhibitory activity (1700 and 499 of peptides in total, respectively, were random-coil dominated).

When discussing the secondary structures of fragments with specific activity, particular attention should be paid to the length of the peptide chain. Many of bioactive peptides are short sequences (di- or tripeptides) (*25*). Knowing the fact that one of the major secondary structures of proteins, α-helix, requires 3.6 amino acids per one turn of a helical chain (*44*), it should be noted that defining the structure matching the short peptides means locating them in a longer fragment possessing a particular structure or at the border of fragment matching different structures. However, an analysis of the structure of bioactive motifs by means of BIOPEP combined with PBIL can be a useful strategy to look for some structural requirements determining the bioactivity of peptide(s).

To conclude, a structural analysis of peptides representing a variety of bioactivities showed that bioactive motifs mostly occupy random coil or helical parts of the protein chain. Some of the peptides matched a combination of coil and helix structures. These peptides were rich in proline or aromatic residues, which also seems to be crucial in the context of the structure-activity relationship of peptides.

Conclusions

The BIOPEP and Pôle Bioinformatique Lyonnais analysis tools provide valuable data on food protein precursors as a source of bioactive peptides combined with the structural aspects of bioactivities encrypted in proteins. The most recent analysis of food proteins has found that they could be a source of peptides affecting neural diseases (*e.g.* ubiquitin-mediated proteolysis peptides). This could be significant in terms of the relations between the diet and the health of potential patients. The structure--bioactivity relationship of peptides analyzed *via* the method involving the BIOPEP-Pôle Bioinformatique Lyonnais server showed that in order to be bioactive, it is essential for peptides to assume the position of a coil (or a combination of coil and α-helix) in the sequence of their protein precursors. However, it should be noted that the number of peptides in the data resources (databases), the length of the peptide chain and the repeatability of the occurrence of characteristic motifs (*i.e.* amino acids), both in the peptide and in the protein, affect the accuracy of studies of the bioactivity of molecules.

Acknowledgements

This study was supported by project no 528-0712--0809, State Committee for Scientific Research, Poland.

References

1. H. Meisel, Food-derived bioactive proteins and peptides as potential components of nutraceuticals, *Curr. Pharm. Design*, *13* (2007) 771–772.
2. A. Iwaniak, J. Dziuba, Animal and plant proteins as precursors of peptides with ACE inhibitory activity – An *in silico* stategy of protein evaluation, *Food Technol. Biotechnol. 47* (2009) 441–449.
3. P. Minkiewicz, J. Dziuba, M. Darewicz, A. Iwaniak, M. Dziuba, D. Nałęcz, Food peptidomics, *Food Technol. Biotechnol. 46* (2008) 1–10.
4. R. Apweiler, Functional information in SWISS-PROT: The basis for large-scale characterisation of protein sequences, *Brief. Bioinform. 2* (2001) 9–18.
5. A.J. Reid, C. Yeats, C.A. Orengo, Methods of remote homology detection can be combined to increase coverage by 10 % in the midnight zone, *Bioinformatics, 23* (2007) 2353–2360.
6. P. Bork, C.A. Orengo, Sequences and topology: Genes and structures in context, *Curr. Opin. Struct. Biol. 14* (2004) 261–263.
7. M.J. Blythe, I.A. Doytchinova, D.R. Flower, JenPep: A database of quantitative functional peptide data for immunology, *Bioinformatics, 18* (2007) 434–439.
8. A.D. Baxevanis, B.F.F. Ouellette: *Bioinformatics*, PWN, Warsaw, Poland (2004) p. 1 (in Polish).
9. M. Karelson: Empirical Molecular Descriptors. In: *Molecular Descriptors in QSAR/QSPR*, M. Karelson (Ed.), Wiley--Interscience, New York, NY, USA (2000) pp. 13–19.
10. R. Todeschini, V. Consonni: QSAR/QSPR Modeling. In: *Molecular Descriptors for Chemoinformatics. Vol. I: Alphabetical Listing*, R. Todeschini, V. Consonni (Eds.), Wiley-VCH Verlag GmbH, Weinheim, Germany (2006) p. XXIX.
11. O. Mekenyan, Dynamic QSAR techniques: Applications in drug design and toxicology, *Curr. Pharm. Design*, *8* (2002) 1605–1621.
12. A.H. Pripp, R. Sørensen, L. Stepaniak, T. Sørhaug, Relationship between proteolysis and angiotensin-I–converting enzyme inhibition in different cheeses, *LWT-Food Sci. Technol. 39* (2006) 677–683.
13. A.H. Pripp, Y. Ardö, Modelling relationship between angiotensin-(I)-converting enzyme inhibition and the bitter taste of peptides, *Food Chem. 102* (2007) 880–888.
14. A. Iwaniak, J. Dziuba, M. Niklewicz, The BIOPEP database – A tool for the *in silico* method of classification of food proteins as the source of peptides with antihypertensive activity, *Acta Aliment. 34* (2005) 417–425.
15. J. Dziuba, A. Iwaniak: Database of Protein and Bioactive Peptide Sequences. In: *Nutraceutical Proteins and Peptides in Health and Disease*, Y. Mine, F. Shahidi (Eds.), CRC Press, Boca Raton, FL, USA (2006) pp. 543–564.
16. J. Dziuba, A. Iwaniak, M. Niklewicz, Database of protein and bioactive peptide sequences – BIOPEP (2003) (*http://www.uwm.edu.pl/biochemia*).
17. F. Shahidi, Y. Zhong, Bioactive peptides, *J. AOAC Int. 91* (2008) 914–931.
18. P. Minkiewicz, J. Dziuba, A. Iwaniak, M. Dziuba, M. Darewicz, BIOPEP database and other programs processing bioactive peptide sequences, *J. AOAC Int. 91* (2008) 965–980.
19. J. Dziuba, A. Iwaniak, P. Minkiewicz, Computer-aided characteristics of proteins as potential precursors of bioactive peptides, *Polimery, 48* (2003) 50–53.
20. Pôle Bioinformatique Lyonnais (PBIL) server (*http://pbil.univ-lyon1.fr/*).
21. G. Perrière, C. Combet, S. Penel, C. Blanchet, J. Thioulouse, C. Geourjon, J. Grassot, C. Charavay, M. Gouy, L. Duret, G. Deléage, Integrated databanks access and sequence/structure analysis services at the PBIL, *Nucl. Acids. Res. 31* (2003) 3393–3399.
22. M. Dziuba, J. Dziuba, A. Iwaniak, Bioinformatics-aided characteristics of the structural motifs of selected potentially celiac-toxic proteins of cereals and leguminous plants, *Pol. J. Food Nutr. Sci. 57* (2007) 405–414.

23. M. Dziuba, M. Darewicz, Food proteins as precursors of bioactive peptides – Classification into families, *Food Sci. Technol. Int. 13* (2007) 393–404.

24. H. Korhonen, A. Pihlanto, Bioactive peptides: Production and functionality, *Int. Dairy J. 16* (2006) 945–960.

25. H. Meisel, Biochemical properties of peptides encrypted in bovine milk proteins, *Curr. Med. Chem. 12* (2005) 1905–1919.

26. M. Dziuba, B. Dziuba, A. Iwaniak, Milk proteins as precursors of bioactive peptides, *Acta Sci. Pol. Technol. Aliment. 8* (2009) 71–90.

27. I.L. Lu, K.C. Tsai, Y.K. Chiang, W.T. Jiaang, S.H. Wu, N. Mahindroo *et al.*, A three-dimensional pharmacophore model for dipeptidyl peptidase IV inhibitors, *Eur. J. Med. Chem. 43* (2008) 1603–1611.

28. R.N. Bergman, D.T. Finegood, S.E. Kahn, The evolution of beta-cell dysfunction and insulin resistance in type 2 diabetes, *Eur. J. Clin. Invest.* (Suppl. 3), *32* (2002) 35–45.

29. H. Korhonen, A. Pihlanto, Food-derived bioactive peptides – Opportunities for designing future foods, *Curr. Pham. Des. 9* (2003) 1297–1308.

30. R.C. Piper, J.P. Luzio, Ubiquitin-dependent sorting of integral membrane proteins for degradation in lysosomes, *Curr. Opin. Cell Biol. 19* (2007) 459–465.

31. A. Ciechanover, A.L. Schwartz, The ubiquitin-mediated proteolytic pathway: Mechanisms of recognition of the proteolytic substrate and involvement in the degradation of native cellular proteins, *FASEB J. 8* (1994) 182–191.

32. InterPro database (*http://www.ebi.ac.uk/interpro*).

33. A. Iwaniak, P. Minkiewicz, Biologically active peptides derived from proteins – A review, *Pol. J. Food Nutr. Sci. 58* (2008) 289–294.

34. A. Iwaniak, P. Minkiewicz, Proteins as the source of physiologically and functionally active peptides, *Acta Sci. Pol. Technol. Aliment. 6* (2007) 5–15.

35. I. Nonaka, S. Katsuda, T. Ohmori, T. Shigehisa, T. Nakagami, S. Maruyama, *In vitro* and *in vivo* anti-platelet effects of enzymatic hydrolysates of collagen and collagen-related peptides, *Biosci. Biotechnol. Biochem. 61* (1997) 772–775.

36. I. Nonaka, H. Tanaka, S. Maruyama, Production of fibrin polymerization inhibitor from collagen by some proteases, *Ann. NY Acad. Sci. 750* (1995) 412–414.

37. A.M. Fiat, D. Migliore-Samour, P. Jollès, L. Drouet, C.B.D. Sollier, J. Caen, Biologically active peptides from milk proteins with emphasis on two examples concerning antithrombotic and immunomodulating activities, *J. Dairy Sci. 76* (1993) 301–310.

38. A. Moutran, A. Balan, L.C.S Ferreira, A. Giorgetti, A. Tramontano, R.C.C. Ferreira, Structural model and ligand interactions of the *Xanthomonas axonopodis* pv. *citri* oligopeptide-binding protein, *Genet. Mol. Res. 6* (2007) 1169–1177.

39. H. Meisel, R.J. FitzGerald, Opioid peptides encrypted in milk protein sequences, *Brit. J. Nutr.* (Suppl. 1), *84* (2007) 27–31.

40. G. Samonina, I. Ashmarin, L. Lyapina, Glyproline peptide family: Review on bioactivity and possible origins, *Pathophysiology, 8* (2002) 229–234.

41. A.A. Karelin, E.Y. Blishchenko, V.T. Ivanov: Peptide Pools: Species Conservative and Tissue Specific. In: *Peptides 1998, Proceedings of the 25th European Peptide Symposium*, S. Bajusz, F. Hudecz (Eds.), Akadémiai Kiadó, Budapest, Hungary (1999) pp. 726–727.

42. J. Garnier, D.J. Osguthorpe, B. Robson, Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins, *J. Mol. Biol. 120* (1978) 97–120.

43. A. Kloczkowski, K.L. Ting, R.L. Jernigan, J. Garnier, Protein secondary structure prediction based on the GOR algorithm incorporating multiple sequence alignment information, *Polymer, 43* (2002) 441–449.

44. C. Branden, J. Tooze: *Introduction to Protein Structure*, Garland Publishing Inc., New York, NY, USA (1999) p. 15.