JERZY DZIUBA*), ANNA IWANIAK, PIOTR MINKIEWICZ

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

Computer-aided characteristics of proteins as potential precursors of bioactive peptides

Summary — Proteins, apart from their basic functions, may play the role of precursors of the peptides that regulate the body functions and/or show antihypertensive, antithrombotic, antiamnestic, immunomodulating and other activities. Computer analysis of 109 protein sequences, done with help of a database containing 109 protein sequences and 1100 bioactive peptides sequences, has been presented. Following discriminants have been used: the frequency of occurence of the bioactive fragments in a protein chain *(A)*, potential activity of protein fragments *(B)* and the profile of potential biological activity of a protein. 3781 bioactive fragments present in the protein sequences under consideration have been found. Most commonly occurring bioactive fragments were angiotensin converting enzyme inhibitors (ACE) (EC 3.4.15.1.) and dipeptidyl peptidase IV inhibitors. Bovine β-casein and rice prolamin are the best precursors of antihypertensive peptides. Bioactive fragments as well as their surroundings are hydrophilic and mostly localized at random coil structures of proteins.
Key words: proteins, bioactive peptides, computer-aided analysis of the sequences.

All proteins apart from their basic functions, can be precursors of bioactive peptides [1]. Term "peptidome" describes all peptides present in organism [2]. At present peptidomics covers among others investigations of biological activities of peptides identified in organism or tissue [2]. Peptidomes can be considered in case of peptides from foods as well. Biologically active peptides are used in viral infections, immunological system disorders and cancer treatment [3]. Bioactive peptides can be also applied as components of so-called functional food.

Significance of such food is increasing [4—7]. Food proteins are precursors of many peptides with diverse activities such as antihypertensive, antithrombotic, immunomodulating, antioxidant, opioid and others [4—8].

To date trials of protein classification as potential precursors of bioactive peptides have been undertaken taking into consideration the following criteria: maximal yield of peptide, which may be obtained from 1 g of protein [4], and the ratio of the frequency of occurrence of bioactive fragment to the probability that peptide with the given length, encoded by the random DNA fragment, may contain a sequence corresponding to a given activity [9].

The aim of this work was to propose a criterion of evaluation of proteins as potential precursors of bioactive peptides and to determine protein structural properties associated with the occurrence of bioactive fragments in protein chains.

METHODS

We used the following protein sequences: [Q9TOP5], [PRO1_ORYS1], [KAF2_SORBI], [KAF8_SORBI], [KAF1_SORBI], [PRO5_ORYSA], [PRO7_ORYSA], [PRO6_ORYSA], [PRO4_ORYSA], [11SB_CUCMA], [7SBG_SOYBN], [11S3_HELAN], [LEGA_GOSHI], [LEGB_GOSHI], [AVEN_AVESA], [MONB_DIOCU], [MONA_DIOCU], [THN1_WHEAT], [THG1_WHEAT], [THN2_WHEAT], [THNA_HORVU], [GDA4_WHEAT], [GDA3_WHEAT], [GDA9_WHEAT], [GDA0_WHEAT], [GDA5_WHEAT], [GDA6_WHEAT], [GDA1_WHEAT], [GDA7_WHEAT], [GDA2_WHEAT], [GDA8_WHEAT], [GDB0_WHEAT], [GDB2_WHEAT], [GDBX_WHEAT], [GDBB_WHEAT], [GDB3_WHEAT], [HOG1_HORVU], [ASP_THECC], [GLT_WHEAT], [Q9SIA7], [Q39770], [Q9TOP5], [Q43671], [THNB_WHEAT],

*) To whom all correspondence should be addressed, e-mail: dziuba@uwm.edu.pl

[THG_HORVU], [LCA_CAMDR], [LCA_SHEEP], [LCA_HORSE], [LCA_HUMAN], [LCA_CAVPO], [LCA_RAT], [LCA_RABIT], [LCA_MACEU], [LCA_BOVIN], [LACB_BOVIN], [LACB_CAPHI], [LYCN_BOVIN], [LYC_COLLI], [LYC2_CANFA], [LYC_CHIC], [LYC_HYACE], [LYC_HUMAN], [CHEY_COLI], [MLE1_CHICK], [TRIC_CHICK], [TRCS_CHICK], [TRT1_CHICK], [TPCS_PIG], [CASK_CAPHI], [CASK_HUMAN], [CASK_BOVIN], genetic variant B, [CASB_BOVIN]: genetic variants: A^1, A^2, A^3, B, E, F, [CAS1_BOVIN, genetic variants: A, B, C, D, [CAS2_BOVIN], genetic variant A, [TRT3_CHICK], [TRT2_CHICK], [TPMA_CHICK], [TPMB-CHICK], [Q9UCY5], [ELSC_BOVIN], [ELS_BOVIN], [CA13_BOVIN], [CA11_BOVIN], [CA1_CHICK], [CON1_CHICK], [BBP_PIEBR], [OBP_BOVINA], [RET2_RAT], [Q28129], legumin-like chains from: barley (*Avena sativa*), rape (*Brassica napus*), pumpkin (*Cucurbita maxima*), soybean (*Glycine max*), rice (*Oryza sativa*) [9], fykocyanin, plastocyanin, flavodoxin [10]. Most of the proteins summarized above is decribed using entry names used in the SWISS-PROT/trEMBL database (http://www.expasy.ch) [11].

Protein sequence analysis directed to the presence of peptides with diverse biological activities required introduction of newly created, quantitative criteria for their evaluation as well as creation of new computer application BIOPEP described below.

BIOPEP is a computer application designed at the Chair of Food Biochemistry. It requires Windows environment and MS Access'97 installation. The BIOPEP enables creation of potential protein fragments bioactivity profiles defined as a location and kind of activity of bioactive fragments in protein chain as well as searching for the sites susceptible to the action of endopeptidases, similarly to the previously described PROTEIN program [5, 12].

The main form of BIOPEP application consists of two twin databases: protein sequences and biologically active peptide sequences. Both databases are continuously enriched and contain till now 109 sequences of proteins and 1100 sequences of peptides.

The amino acid sequences are inserted using one-letter code. It is possible to insert peptides containing 20 most frequently occurring amino acids and peptides containing posttranslational modifications (*e.g.* phosphorylation and *C*-terminal amidation).

Both databases besides standard information such as: name, sequence, references contain also the results of calculations performed automatically by application. These are: number of amino acid residues, monoisotopic and chemical molecular mass [13].

Main computer application form also performs "Operations on the records". This option contains programmed algorithms determining the values of proteins as the precursors of bioactive peptides. A choice of this application enables the automatic calculation. It is described below. The results can be transferred to the other Windows applications such as MS Word or MS Excel or can be obtained as a typical MS Access report.

The following discriminants have been designed for evaluation of proteins as the precursors of bioactive peptides:

a) the occurrence frequency of the fragments with a given activity in polypeptide chain (A):

$$A = \frac{a}{N} \tag{1}$$

where: a — *the number of fragments with a given activity;* N — *the number of amino acid residues.*

b) potential biological activity of the protein fragments (B), mM^{-1}:

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

where: a_i — *number of repetitions of i-th bioactive fragment;* EC_{50i} — *the concentration of half-maximum activity of i-th protein fragment, mM;* k — *number of different fragments with a given activity.*

Calculation of the potential activity of protein fragments (B) has been carried out on the basis of data concerning antihypertensive peptides because of known EC_{50} values for all peptides belonging to this group.

We have calculated the hydropathy index [14] and predicted secondary structures [15] of bioactive fragments and fragments with the nearest surroundings consisting of 5 amino acid residues preceding and 5 of them following the bioactive fragment using the PREDICT 7 program [16].

All calculations were carried out for mature proteins (without signal peptides).

RESULTS AND DISCUSSION

The frequency of bioactive fragments occurrence (A) for six selected proteins is shown in Fig. 1. All presented proteins were potential sources of antihypertensive peptides and dipeptidylpeptidase IV inhibitors. Relatively high values of discriminant A were observed especially for two proteins: genetic variant A^2 of β-casein and β-lactoglobulin. The presence of peptides with these activities among the proteolysis products derived from these proteins was found experimentally [4, 6, 17]. The high value of A discriminant calculated for chicken connectin may suggest that this protein besides its basic function (*i.e.* playing role of connective tissue component) may also be a precursor of endogenous inhibitors of angiotensin-converting enzyme (ACE) [EC 3.4.15.1]. This suggestion is in agreement with hypothesis published by Karelin *et al.* [1] assuming that all proteins of organism may act as a reserve of peptides regulating its functions.

The average A values calculated for the most frequently occurring kinds of bioactive fragments are shown in Table 1. Relatively high frequency of occurrence of antihypertensive fragments as compared with

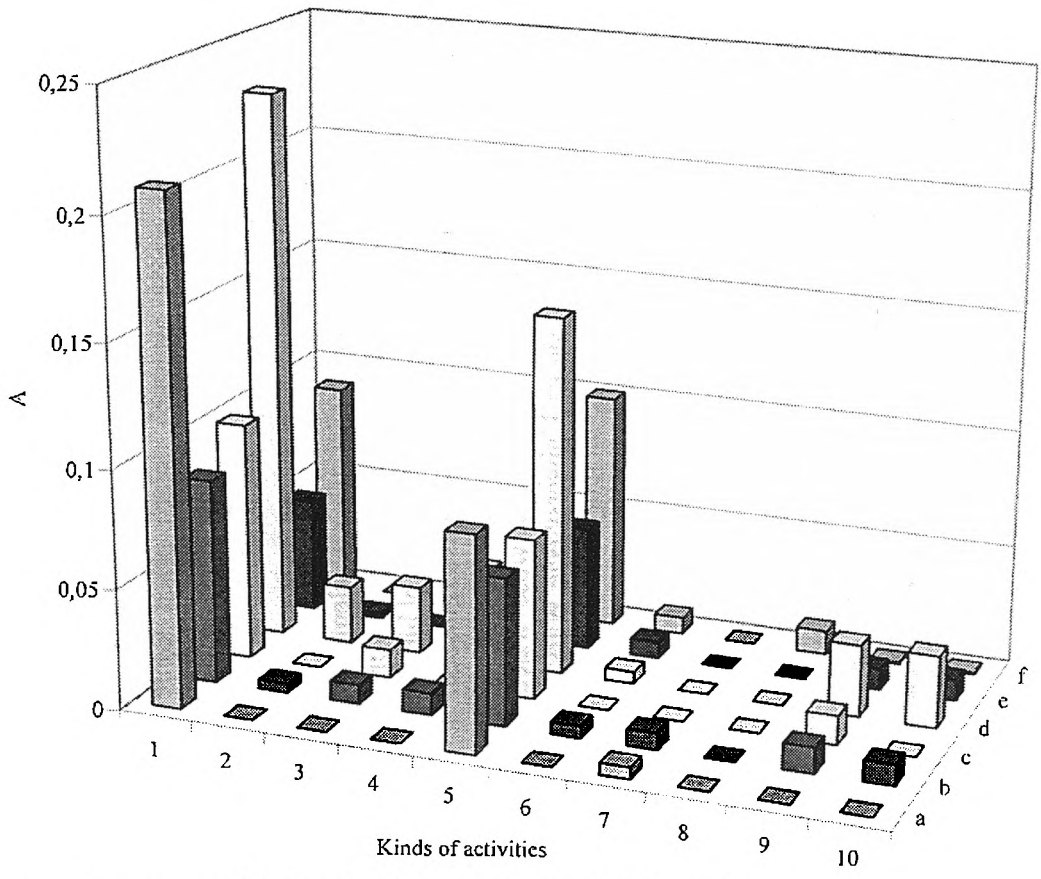

Fig. 1. The frequency of occurrence of the bioactive fragments (A) in selected proteins. Peptide activity: 1 — antihypertensive, 2 — immunomodulating, 3 — opioid, 4 — antiamnestic, 5 — inhibitors of dipeptidylpeptidase IV, 6 — antioxidantive, 7 — inhibitors of dipeptidylcarboxypeptidase, 8 — coeliac toxic peptides, 9 — regulating the ion flow, 10 —neuropeptides; proteins: a) chicken (Gallus gallus) connectin 1 [Q90804], b) chicken (Gallus gallus) myosin light chain [MLE1_CHICK], c) soybean (Glycine max) 7S globulin [7SBG_SOYBN], d) bovine (Bos taurus) β-lactoglobulin [LACB_BOVIN], genetic variant A, e) bovine (Bos taurus) β-casein [CASB_BOVIN], genetic variant A^2, f) wheat (Triticum aestivum) α/β-gliadin, clone PW 1215 [GDA6_WHEAT]

T a b l e 1. Biological activities of the most frequently occurring fragments of investigated proteins

Activity	Number of bioactive fragments	A_{av}±sd
Antihypertensive	1556	0.0715±0.0515
Inhibiting dipeptidylpeptidase IV action	1557	0.046±0.040
Immunomodulating	130	0.0069±0.0080
Opioid	107	0.0062±0.0097
Neuropeptides	106	0.0058±0.016

T a b l e 2. Proteins with the highest frequency of antihypertensive fragments' occurrence

Protein	*A*	*B*, mM^{-1}
Prolamin, clone PPROL 7, rice (*Oryza sativa*) [PRO7_ORYSA]	0.1704	0.02077
Prolamin, clone PPROL 14, rice (*Oryza sativa*) [PRO6_ORYSA]	0.1567	0.02040
β-Casein, genetic variant A^1, cow (*Bos taurus*) [CASB_BOVIN]	0.196	0.0072
β-Casein, genetic variant A^2, cow (*Bos taurus*) [CASB_BOVIN]	0.23	0.0080
β-Casein, genetic variant A^3, cow (*Bos taurus*) [CASB_BOVIN]	0.196	0.0080
β-Casein, genetic variant B, cow (*Bos taurus*) [CASB_BOVIN]	0.191	0.0080
β-Casein, genetic variant C, cow (*Bos taurus*)	0.196	0.0080
β-Casein, genetic variant E, cow (*Bos taurus*)	0.191	0.0080
β-Casein, genetic variant F, cow (*Bos taurus*)	0.220	0.0106

fragments showing other activities may be due to the fact that they are most common peptides in our database. Bioactive fragments acting as dipeptidylpeptidase IV inhibitors create the second group of dominant activity. Their frequent occurrence may be due to a short chain length. They are mainly di- and tripeptides [18].

The *A* and *B* discriminant values for proteins being the richest sources of antihypertensive peptides are shown in Table 2. This table shows that there is no simple relationship between the frequency of bioactive fragments' occurrence in the given protein and activity of

these fragments. Fragments present in rice prolamins are stronger inhibitors of angiotensin-converting enzyme than those of the bovine β-casein.

Table 3. Average values of structural parameters concerning bioactive fragments present in the investigated proteins [1),2)]

Parameter	Bioactive fragments	Bioactive fragments with surroundings
Hydropathy index	-0.275±10.100	-0.393±8.102
α-Helix content, %	13.01±4.90	14.17±3.83
β-Turn content, %	19.11±10.93	25.46±7.14
β-Sheet content, %	23.69±10.50	20.81±7.61
Random coil content, %	44.23±16.76	39.56±7.16

[1)] The results were obtained for 3781 bioactive fragments found in the sequences of the investigated proteins.
[2)] All the differences between the values shown in Table 3 were statistically significant at least at the 0.01 level (*t* Student's test).

The average values of hydropathy index and contents of particular secondary structure calculated for bioactive fragments and their surroundings are shown in Table 3. The calculated results suggest that hydrophilic surroundings of bioactive fragments facilitate their exposure on the protein surface. The protein surfaces are more hydrophilic than their interiors [19]. Such location may facilitate accessibility of proteolytic enzymes to the bioactive fragments. The random coil was predicted to be the dominant secondary structure of bioactive fragments and their surroundings. Antihypertensive fragments, being the most frequently occurring, are characterised by high content of the amino acids destabilising the α-helix and facilitating the random coil formation [6].

CONCLUSIONS

— The *A* (frequency of bioactive fragments' occurrence in protein sequence) and *B* (potential activity of protein fragments) discriminants may be a tool for evaluation of proteins as potential precursors of bioactive peptides. The *A* discriminant is more universal because the EC_{50} values are known only for some peptides.

— The antihypertensive fragments (inhibitors of angiotensin-converting enzyme) and inhibitors of dipeptidylpeptidase IV are the most frequently occurring fragments in the investigated proteins.

— The results concerning hydropathy index of bioactive fragments and their surroundings suggest that hydrophilic neighbourhood of bioactive fragments facilitates the location of these fragments on the protein surface.

ACKNOWLEDGMENT

This work was supported by the State Committee for Scientific Research (KBN) within the project 021/P06/99/08.

REFERENCES

1. Karelin A. A., Blischenko E. Y., Ivanov V. T.: *FEBS Lett.* 1998, **428**, 7.
2. Schrader M., Schulz-Knappe P.: *Trends Biotechnol.* 2001, **19**, S55.
3. Latham P. W.: *Nature Biotechnol.* 1999, **17**, 755.
4. Meisel H., Schlimme E.: *Kieler Milchw. Forsch.* 1996, **48**, 343.
5. Dziuba J., Minkiewicz P., Nałęcz D., Iwaniak A.: *Nahrung/Food* 1999, **43**, 190.
6. Meisel H.: *Biopolymers* 1997, **43**, 119.
7. Dziuba J., Minkiewicz P., Nałęcz D.: *Pol. J. Food Nutr. Sci.* 1999, **8/49** (1), 3.
8. Dziuba J., Minkiewicz P., Plitnik K.: *Pol. J. Food Nutr. Sci.* 1996, **5/46** (4), 85.
9. Krause J.-P., Wüstneck R., Schwenke K. D.: *Nahrung/Food* 1999, **43**, 9.
10. Godzik A.: *Protein Sci.* 1996, **5**, 1325.
11. Junker V. L., Contrino S., Fleischmann W., Hermjakob H., Lang F., Magrane M., Martin M. J., Mitaritonna N., O'Donovan C., Apweiler R.: *J. Biotechnol.* 2000, **78**, 221.
12. Dziuba J., Minkiewicz P., Puszka K., Dąbrowski S.: *Pol. J. Food Nutr. Sci.* 1995, **4/45** (3), 31.
13. Yates J. R. III, McCormack A. L., Link A. J., Schieltz D., Eng J., Hays L.: *Analyst* 1996, **121**, 65R.
14. Kyte J., Doolittle R. F.: *J. Mol. Biol.* 1982, **157**, 105.
15. Garnier J., Osguthorpe D. J., Robson B.: *J. Mol. Biol.* 1978, **120**, 97.
16. Cármenes R. S., Freije J. P., Molina M. M., Martín J. M.: *Biochem. Biophys. Res. Commun.* 1989, **189**, 687.
17. Yamamoto N.: *Biopolymers* 1997, **43**, 129.
18. Yaron A., Naider F.: *Crit. Rev. Biochem. Mol. Biol.* 1993, **28**, 31.
19. Jones S., Thornton J. M.: *Proc. Natl Acad. Sci. USA* 1996, **93**, 13.