JERZY DZIUBA*), MARTA NIKLEWICZ, ANNA IWANIAK,
MAŁGORZATA DAREWICZ, PIOTR MINKIEWICZ

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

Structural properties of proteolytic-accessible bioactive fragments of selected animal proteins

Summary — The computer simulation of the distribution of biologically active fragments and bonds which are predicted to be susceptible to the action of endopeptidases of known specificity, hydropathy index and prediction of secondary structures were performed. The average values of hydropathy index calculated for bioactive fragments of selected animal proteins predicted *in silico* to be released by proteolytic enzymes as well as surroundings of such fragments show that they are hydrophilic. The most frequently occurring structure in the animal proteins bioactive fragments with surroundings is β-strand. There is no preferable secondary structure in the bioactive fragments encrypted in the animal protein sequences. Our findings suggest that the distribution of bioactive fragments may favour their release by proteinases.
Key words: proteins, bioactive peptides, hydropathy index, proteolysis, secondary structure, computer-aided analysis of the sequences.

All proteins may be considered to be precursors of the bioactive peptides [1, 2], which are inactive in the protein sequence and may reveal various types of biological activity after being released by proteolytic enzymes. All peptides present in organism are described by term "peptidome" [3]. At present peptidomics covers among others investigations of biological activities of peptides identified in organism or tissue. Peptidomes can be considered in case of peptides from foods as well. Oligopeptides derived from these proteins may be absorbed from the digest tract and interact with organism receptors [4]. These peptides can, for example, regulate blood pressure, stimulate or suppress the action of the immune system, affect the activity of the nervous system, inhibit the growth of bacteria or fungi and reveal many other functions [2, 4, 5]. Relatively short oligopeptides, obtained from protein hydrolyzates, play an important role in modelling of sensory properties of food [6]. Biologically active peptides can be recommended as components of functional food, *i.e.* food with desirable and designed biological activity as well as in production of nutraceuticals [7—9].

In the recent years the massive growth of computer techniques applied in life science had been observed. Such techniques are often used in prediction of protein function from its primary and/or secondary structure. Thanks to these techniques we were able to elaborate the highly specialized database of protein and bioactive peptide sequences — BIOPEP [10]. BIOPEP allows to evaluate the protein value which is based on the following criteria: the profile of potential biological activity of protein (*i.e.* type and location of bioactive fragment in protein sequence), the frequency of the occurrence of bioactive fragment in protein chain (*A*) and potential biological activity of protein (*B*) [11]. Our database also contains the option allowing the prediction of the potential release of bioactive peptides from their protein precursors. It is possible due to the additionally designed database of proteolytic enzymes available in BIOPEP.

Biopeptides as components of food with desired features become an interesting issue for scientific research. As yet no research has been conducted concerning the molecular parameters of biologically active proteins fragments released *in silico* by proteolytic enzymes. The aim of the research was to analyse structural properties of bioactive fragments encrypted in animal protein chains which are predicted to be accessible for endopeptidases of known specificity. Additionally, the bioinformatic-aided determination of hydropathy index was carried out. Such research can be meaningful from the point of designing the proteolytic processes in aspect of bioactive peptides obtaining on the industrial scale.

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

METHODS

BIOPEP is a computer application designed in the Chair of Food Biochemistry of our university. It requires Windows environment and MS Access'97 installation. At present the database BIOPEP contains information on 1516 peptides and is continuously updated with new peptides with known sequences and various activities. It comprises both, peptides with 20 most commonly occurring amino acids and peptides whose sources were post--translationally modified (*e.g.* phosphorylated or with amidated C-terminal amino acid residue) proteins. Main attention was paid to di- and tripeptides which are most easily adsorbed from the alimentary tract to blood and which can be found in many protein sequences with relatively high probability [6]. The amino acid sequences are inserted using one-letter code.

Fifty three animal protein sequences from the database of the Chair of Food Biochemistry were studied. Their list is presented in Table 1. Most of the sequences are also comprised in the SWISS-PROT/trEMBL database [13] which is indicated by their entry names. Our database contains protein sequences of the same species but varying in genetic variant (*e.g.* casein) or protein sequences from different animal species.

The database is also supplied with recognition and specificity sequences information on proteolytic enzymes. Recognition sequence is the protein fragment which is recognised by proteolytic enzyme. Some recognition sequences contain more than one amino acid residues [2]. Peptide bonds, predictably susceptible to the action of endopeptidases with known specificity, were located in the amino acid sequences of used animal proteins. We used the data concerning the specificity of the following enzymes: Chymotrypsin EC 3.4.21.1; Trypsin EC 3.4.21.4; Proteinase K EC 3.4.21.14; Prolyl endopeptidase EC 3.4.21.26; Plasminogen activator EC 3.4.21.31; Elastase EC 3.4.21.36; Papain EC 3.4.22.2.; Ficin EC 3.4.22.3; Bromelain EC 3.4.22.4; Clostripain EC 3.4.22.8; Pepsin EC 3.4.23.1; Cathepsin D EC 3.4.23.5; Collagenase *Clostridium histolyticum* EC 3.4.24.3; *Bacillus subtilis* neutral proteinase EC 3.4.24.4. The enzyme specificity was taken from the MEROPS Internet database [14] and the reference [15—17].

The secondary structure of proteins was predicted by GOR algorithm [18]. The hydropathy index was calculated according to Kyte and Doolittle [19]. The PREDICT 7 program [20] was used for structural calculations. The contents of the secondary structures in bioactive fragments within protein chains and bioactive fragments with surrounding containing 5 residues preceding and 5 residues following a given fragment were calculated. Hydropathy index was expressed as an average hydropathy index per residue in a bioactive fragment or a fragment with surrounding. Only the bioactive fragments preceded and followed by the bonds predictably susceptible to the action of endopeptidases were consi-

Table 1. List of proteins used for computer simulated proteolysis by means of BIOPEP computer application

No.	Protein	Identification No. in SWISS-PROT database
1	α-Lactalbumin, camel (*Camelus dromedariu*)	LCA_CAMDR
2	α-Lactalbumin, sheep (*Ovis aries*)	LCA_SHEEP
3	α-Lactalbumin, goat (*Capra hircus*)	LCA_CAPHI
4	α-Lactalbumin, mare (*Equus caballus*)	LCA_HORSE
5	α-Lactalbumin, man (*Homo sapiens*)	LCA_HUMAN
6	α-Lactalbumin, guinea pig (*Cavia porcellus*)	LCA_CAVPO
7	α-Lactalbumin, rat (*Rattus norvegicus*)	LCA_RAT
8	α-Lactalbumin, rabbit (*Oryctolagus cuniculus*)	LCA_RABIT
9	α-Lactalbumin, kangaroo (*Macropus eugenii*)	LCA_MACEU
10	α-Lactalbumin, bovine (*Bos taurus*)	LCA_BOVIN
11	β-Lactoglobulin, variant A, cow (*Bos taurus*)	LACB_BOVIN
12	β-Lactoglobulin, goat (*Capra hircus*)	LACB_CAPHI
13	β-Lactoglobulin, sheep (*Ovis aries*)	LACB_SHEEP
14	Lysosyme, cow (*Bos taurus*)	LYCN_BOVIN
15	Lysosyme, pigeon (*Columba livia*)	LYC_COLLI
16	Lysosyme, dog (*Canis familiares*)	LYC2_CANFA
17	Lysosyme, hen (*Gallus gallus*)	LYC_CHICK
18	Lysosyme, moth (*Hydophora cecropia*)	LYC_HYACE
19	Lysosyme, man (*Homo sapiens*)	LYC_HUMAN
20	Myosin, light chain (I), hen (*Gallus gallus*)	KMLS_CHICK
21	Troponin 1, hen (*Gallus gallus*)	TRIC_CHICK
22	Troponin C, skeleton muscles, hen (*Gallus gallus*)	TPCS_CHICK
23	Troponin T, hen (*Gallus gallus*)	TRT2_CHICK
24	Troponin C, swine (*Sus scrofa*)	TPCS_PIG
25	κ-Casein, goat (*Capria hircus*)	CASK_CAPHI
26	κ-Casein, man (*Homo sapiens*)	CASK_HUMAN
27	κ-Casein, variant A, bovine (*Bos taurus*)	CASK_BOVIN
28	κ-Casein, variant B, bovine (*Bos taurus*)	CASK_BOVIN
29	α_{S1}-Casein, variant A, bovine (*Bos taurus*)	CAS1_BOVIN
30	α_{S1}-Casein, variant B, bovine (*Bos taurus*)	CAS1_BOVIN
31	α_{S1}-Casein, variant C, bovine (*Bos taurus*)	CAS1_BOVIN
32	α_{S1}-Casein, variant D, bovine (*Bos taurus*)	CAS1_BOVIN
33	α_{S2}-Casein, variant A, bovine (*Bos taurus*)	CAS2_BOVIN
34	β-Casein, variant A^1, bovine (*Bos taurus*)	CASB_BOVIN
35	β-Casein, variant A^2, bovine (*Bos taurus*)	CASB_BOVIN
36	β-Casein, variant A^3, bovine (*Bos taurus*)	CASB_BOVIN
37	β-Casein, variant B, bovine (*Bos taurus*)	CASB_BOVIN
38	β-Casein, variant C, bovine (*Bos taurus*)	CASB_BOVIN
39	β-Casein, variant E, bovine (*Bos taurus*)	CASB_BOVIN
40	β-Casein, variant F, bovine (*Bos taurus*)	CASB_BOVIN
41	Troponin T, hen (*Gallus gallus*)	TRT3_CHICK
42	Troponin T, hen (*Gallus gallus*)	TRT2_CHICK
43	Tropomyosin-α, hen (*Gallus gallus*)	TPMA_CHICK
44	Tropomyosin-β, hen (*Gallus gallus*)	TPMB-CHICK
45	Lactoferrin, man (*Homo sapiens*)	Q9UCY5
46	Elastin C, bovine (*Bos taurus*)	ELSC_BOVIN
47	Elastin, bovine (*Bos taurus*)	ELS_BOVIN
48	α-Collagen (III), bovine (*Bos taurus*)	CA13_BOVIN
49	α-Collagen(I), bovine (*Bos taurus*)	CA11_BOVIN
50	α-Collagen (I), hen (*Gallus gallus*)	CA1_CHICK
51	Connectin 1, hen (*Gallus gallus*)	CON1_CHICK
52	Azurin A	Ref. [12]
53	Lipocalin, bovine (*Bos taurus*)	OBP_BOVIN

Table 2. Structural parameters and biological activity concerning bioactive peptides without and with surrounding released *in silico* from bovine α_{s1}-casein and β-casein; all differences between the values were statistically significant at least at the 0.01 level (*t*-test)

Protein	Bioactive peptide	Biological activity	Bioactive peptide					Bioactive peptide with surrounding				
			contents of particular secondary structure, %				hydropathy index	contents of particular secondary structure, %				hydropathy index
			α-helix	β-turn	β-strand	coil		α-helix	β-turn	β-strand	coil	
α_{S1}-Casein, A genetic variant, bovine	RL [87—88]	antihypertensive	100	—	—	—	-0.575	58.3	25.0	8.3	8.4	-0.652
	GY [80—81]	antihypertensive	50.0	—	—	50.0	0.000	50.0	25.0	—	25.0	-0.387
	AY [130—131]	antihypertensive	—	100	—	—	0.541	—	25.0	—	75.0	-0.241
	VPL [154—156]	inhibitor of dipeptidyl-peptidase	—	33.3	66.6	—	0.511	—	25.0	75.0	—	-0.311
	TTMPLW [181—186]	opioid ACE inhibitor	—	16.7	—	83.3	0.069	—	27.3	—	72.7	-0.950
	QP [13—14]	antihypertensive	—	—	—	100	-0.891	—	16.7	33.3	50.0	-0.561
	LW [185—186]	antihypertensive	—	—	—	100	0.616	—	28.6	—	71.4	-0.123
	FY [132—133]	antihypertensive	—	—	—	100	-0.350	—	41.7	—	58.3	-0.198
	FGK [19—21]	antihypertensive	25.0	—	—	75.0	-0.761	46.1	—	30.7	23.2	-0.766
	FP [15—16]	inhibitor of dipeptidyl-peptidase	—	—	100	—	-0.008	8.4	—	33.3	58.3	-0.531
	YL [81—82]	antihypertensive opioid	100	—	—	—	-0.183	58.3	25.0	—	16.7	-0.295
β-Casein, B genetic variant, bovine	HL [134—135]	antioxidative	—	—	—	100	0.225	—	—	—	100	0.297
	VY [59—60]	antihypertensive	—	—	100	—	1.116	—	25.0	33.4	41.6	0.052
	PF [61—62]	antihypertensive	—	—	—	100	-0.133	—	8.3	33.4	58.3	0.066
	EMPFPK [108—113]	antihypertensive	16.7	66.6	—	16.7	-0.877	37.5	25.0	—	37.5	-1.098
	AVPYPQR [177—183]	ACE inhibitor	—	42.9	—	57.1	-1.388	—	29.4	47.0	23.6	-0.622
	FAQTQSLVYP [52—61]	antihypertensive	—	50.0	40.0	10.0	0.026	10.0	25.0	20.0	45.0	-0.258
	GP [64—65]	inhibitor of dipeptidyl-peptidase antiamnestic antithrombotic regulating stomach mucous membrane	—	—	—	100	-0.442	—	—	16.6	83.4	-0.220
	GP [199—200]	inhibitor of dipeptidyl-peptidase antiamnestic antithrombotic regulating stomach mucous membrane	—	—	—	100	0.566	—	16.6	8.4	75.0	0.172
	FP [62—63]	inhibitor of dipeptidyl-peptidase	—	—	—	100	-0.133	—	—	33.3	66.7	0.062
	FP [111—112]	inhibitor of dipeptidyl-peptidase	—	100	—	—	-0.908	25.1	33.3	—	41.6	-1.066
	FP [205—206]	inhibitor of dipeptidyl-peptidase	—	—	100	—	1.750	—	—	50.0	50.0	1.264
	VP [173—174]	inhibitor of dipeptidyl-peptidase	—	—	100	—	-0.600	—	33.4	41.6	25.0	-0.077
	YP [180—181] [60—61]	antihypertensive	—	50.0	—	50.0	-2.025	—	41.6	25.0	33.4	-1.048
			—	—	50.0	50.0	0.700	—	16.6	33.4	50.0	0.037
	LP [151—152]	inhibitor of dipeptidyl-peptidase	—	—	—	100	-0.229	—	16.6	33.4	50.0	-0.475
	PG [9—10]	antiamnestic antithrombotic regulating stomach mucous membrane	—	—	—	100	0.591	25.0	—	—	75.0	0.142
	YPFPG [60—64]	opioid	—	—	20.0	80.0	0.243	—	13.3	26.7	60.0	-0.085
	MA [102—103]	inhibitor of dipeptidyl-peptidase	100	—	—	—	-0.558	100	—	—	—	-1.113
	YQQPVL [193—198]	antihypertensive	—	33.4	16.6	50.0	0.066	—	12.5	37.5	50.0	0.422
	LH [133—134]	antioxidative	—	—	—	100	-0.541	—	—	—	100	0.205
	LPP [135—137]	antihypertensive	—	—	—	100	1.222	—	15.4	—	84.6	0.335
	QP [117—118]	antihypertensive	50.0	—	—	50.0	0.292	50.0	—	16.6	33.4	0.614
	QP [117—118]	antihypertensive	—	—	—	100	-0.225	—	33.3	—	66.7	-1.087
	AP [103—104]	inhibitor of dipeptidyl-peptidase	100	—	—	—	-1.008	91.6	—	—	8.4	-1.44

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

The *t*-test was used to check the statistical significance of differences.

RESULTS AND DISCUSSION

We found 1212 bioactive fragments in the sequences of studied animal proteins. The enzymes we considered covered a broad spectrum of specificities. Table 2 shows an example of potential possibilities of bioactive peptides releasing as a result of specific endopeptidases action on bovine α_{s1}- and β-casein. The obtained fragments were mainly peptides with antihypertensive and inhibitor of dipeptidylpeptidase activities. In several cases the fragment with the same sequence had several activities *e.g.* peptide with the sequence PG shows the following functions: antiamnestic, antithrombotic, regulating the activity of the stomach mucous membrane. Di- and tripeptides were mainly released from all animal proteins studied. Short peptides are easily released by proteolytic enzymes and adsorbed from alimentary tract to blood [21, 22].

The examples of structural properties of bioactive peptides alone and with surrounding are presented in Table 2 as well. The mean values of the predicted secondary structures and the average values of hydropathy indices calculated for bioactive fragments and bioactive fragments with surroundings in studied animal proteins are presented in Table 3. All the differences presented in this table are statistically significant, at least at the level $p < 0.01$. There is no preferable secondary structure in the bioactive fragments encrypted in the animal protein sequences. β-Sheet is the most frequently occurring structure in the surroundings of bioactive fragments of animal proteins.

T a b l e 3. Average values of structural parameters concerning bioactive fragments present in studied animal proteins obtained for 1212 of bioactive peptides sequences

	Bioactive fragments	Bioactive fragments with surroundings
Hydropathy index	-0.45 ± 0.06	-0.014 ± 0.743
α-Helix content, %	25.0 ± 1.01	25.0 ± 0.89
β-Turn content, %	25.0 ± 0.77	12.50 ± 0.50
β-Sheet content, %	25.0 ± 1.00	37.50 ± 0.76
Random coil content, %	25.0 ± 1.30	25.0 ± 0.90

The low value of hydropathy index of bioactive fragments suggests that they tend to be located on the surface of a protein molecule rather than in its hydrophobic core. The proteins surfaces are more hydrophilic than their interiors [23]. Dziuba *et al.* [24] found that bioactive fragments found in β structures had been located in internal part of the protein molecule. Taking into account our earlier results we found that the bioactive fragments of animal proteins are relatively more hydrophobic than those of plant proteins [12]. This effect may result from the fact that some of the animal proteins we analysed (caseins, collagen or elastin) do not form compact globular structure. Hydrophobic fragments of such proteins are more susceptible to a proteolytic attack than the hydrophobic fragments of proteins with compact globular structure.

Molecular properties of bioactive fragments of protein molecule can be important aspect in production of nutraceuticals as well as functional food. Quite often the identification of bioactive sequences focuses on the elaboration of combinatorial libraries containing the information on the structure-activity relationship of bioactive peptides [5].

CONCLUSIONS

The results concerning hydropathy index of bioactive fragments and bioactive fragments with surroundings from animal proteins suggest that hydrophilic nature of bioactive fragments facilitates the location of these fragments on the protein surface. The most frequent structure of the bioactive fragments with surrounding present in animal proteins is β-strand. There is no preferable secondary structure of the bioactive fragments of the animal protein. The structural properties of bioactive fragments of animal protein chains and their surroundings suggest that the distribution of such fragments favours their susceptibility to the proteolytic enzymes. These information may be valuable in aspect of bioactive peptides obtaining on the industrial scale as well as functional food designing.

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. Dziuba J., Minkiewicz P., Nałęcz D., Iwaniak A.: *Nahrung/Food* 1999, **43**, 190.
3. Schrader M., Schulz-Knappe P.: *Trends Biotechnol.* 2001, **19**, S55.
4. Schlimme E., Meisel H.: *Nahrung/Food* 1995, **39**, 1.
5. Meisel H.: *Biopolymers* 1997, **43**, 119.
6. Dziuba J., Minkiewicz P., Nałęcz D.: *Pol. J. Food Nutr. Sci.* 1999, **8/49**, 3.
7. Korhonen H., Pilhlanto-Leppälä A., Rantamaki P., Tupasela T.: *Trends Food Sci. Technol.* 1998, **9**, 307.
8. Dziuba J., Darewicz M.: *Natur. Sci.* 2000, **4**, 257.
9. Darewicz M., Dziuba J., Panfil T.: *Żywność, Nauka, Technologia, Jakość* 2003, **4/37**, 36.
10. Internet: www.uwm.edu.pl/biochemia

11. Dziuba J., Iwaniak A., Minkiewicz P.: *Polimery* 2003, **48/1**, 50.
12. Dziuba J., Iwaniak A., Niklewicz M., Darewicz M., Minkiewicz P.: *Acta Aliment.* 2004, **33**, 227.
13. Internet: www.expasy.org
14. Internet: www.merops/index.html
15. Arai S., Fujimaki M.: "Food Enzymology" vol. 2, Elsevier, London — New York 1991, p. 83.
16. Park S. Y., Gibbs B. F., Lee B. H.: *Korean. J. Dairy Sci.* 1996, **18**, 237.
17. Yamamoto N., Akino A., Takano T.: *Biosci. Biotech. Biochem.* 1994, **58**, 776.
18. Garnier J., Osguthorpe D., Robson B.: *J. Mol. Biol.* 1978, **120**, 97.
19. Kyte J., Doolittle R. K.: *J. Mol. Biol.* 1982, **157**, 105.
20. Cármenes R. S., Freije J. P., Molina M. M., Martín J. M.: *Biochem. Biophys. Res. Commun.* 1989, **189**, 687.
21. Siemensma A. D., Weijer W. F., Bak H. J.: *Trends Food Sci. Technol.* 1993, **4**, 16.
22. Dziuba J., Minkiewicz P., Puszka K., Dąbrowski S.: *Pol. J. Food Nutr. Sci.* 1995, **4/45**, 30.
23. Jones S., Thornton J. M.: *Proc. Natl. Acad. Sci. USA* 1996, **93**, 13.
24. Dziuba J., Iwaniak A., Niklewicz M., Minkiewicz P.: *Curr. Top. Pept. & Prot. Res.* 2003, **5**, 101.

Received 18 V 2004.