POLISH JOURNAL OF FOOD AND NUTRITION SCIENCES
www.pan.olsztyn.pl/journal/
e-mail: joan@pan.olsztyn.pl
Pol. J. Food Nutr. Sci.
2007, Vol. 57, No. 4, pp. 405–414

BIOINFORMATICS-AIDED CHARACTERISTICS OF THE STRUCTURAL MOTIFS OF SELECTED POTENTIALLY CELIAC-TOXIC PROTEINS OF CEREALS AND LEGUMINOUS PLANTS

Marta Dziuba, Jerzy Dziuba, Anna Iwaniak

Department of Food Biochemistry, University of Warmia and Mazury in Olsztyn, Olsztyn

Key words: celiac disease, extended structural motif, celiac-toxic peptides, protein and peptide sequences database BIOPEP, *in silico* methods

A structural analysis was conducted of peptides responsible for inducing celiac disease, as well as of the structural fragments they are located in, *i.e.* the so-called "extended structural motifs". These motifs originated from wheat A-gliadin and constituted a standard for the analysis of the other proteins. Experiments were carried out based on *in silico* methods.

Analyses covered a total of 403 sequences of selected cereal and seed proteins. Of all the analysed sequences, 155 were found to contain tetrapeptides (potentially known as toxic for celiac disease), namely: QQQP, QQPY, PSQQ, QPYP, including 29 proteins in which the tetrapeptides were constituents of a structural motif with a longer sequence. These were proteins of wheat, barley and oat. All the extended motifs occurred in a hydrophilic surrounding, attaining the structure of beta-turn and random coil.

Computer-aided design of the proteolysis process of selected proteins was carried out as well in the aspect of celiac-toxic peptide release. Active fragments were released from twenty-eight proteins by thermolysin, K proteinase and prolyl oligopeptidase.

INTRODUCTION

Allergy is a common health problem that affects people worldwide. Foods of animal origin are not as rich a source of allergens (except milk and egg protein) as those of plant origin [Breteneder, 1998; Sampson, 2004; Bruijzeel-Koomen *et al.,* 1995]. Allergy to wheat proteins is an example of a surprising multitude of factors inducing hypersensitivity. Depending on a number of factors, a person sensitive to wheat proteins is likely to suffer from atopic dermatitis, anaphylaxis, asthma or gluten-induced celiac disease [Sicherer, 2002; Sampson & Anderson, 2000; Egan *et al.,* 2001]. Celiac disease is the most extensively studied gastrointestinal disease of autoimmunological origin induced by the presence of wheat *(Triticum aestivum)*, barley *(Hordeum vulgare)* or rye *(Secale cereale)* proteins in a diet [Farell & Kelly, 2002; McLachlan *et al.,* 2002; Cornell, 1996; Kasarda, 2002, Dewar *et al.,* 2004]. In childhood, it may be manifested by *e.g.* growth inhibition, whereas in consecutive years by chronic fatigue, neurological symptoms, absorption disorders or even susceptibility to development of selected types of neoplasms [Nieuwenhuizen *et al.,* 2003]. In the etiology of this disease, meticulous attention is paid to the impact of gluten protein modification by tissue transglutaminase. Simultaneously, over 95% of patients are observed to have human leukocyte antigen DQ2 or –DQ8 conferring susceptibility to celiac disease [van Belzen *et al.,* 2001; Sollid, 2002]. Celiac disease patients have been reported to demonstrate an elevated concentration of tissue transglutaminase, which was postulated to be of key significance in the etiology of this disease [Shan *et al.,* 2002]. Upon the activity of tissue transglutaminase, gliadins gain a resultant negative charge. Some of those negatively-charged peptides of gliadins attach more effectively to HLA DQ2 or –DQ8 at the surface of cells with antibodies than the native peptides do. This, in turn, results in the multiplication of a specific response of T cells [Schuppan & Ciccocioppo, 2002]. The activation of T cells evokes a cascade of reactions, thus leading to the formation of highly specific IgA antibodies to transglutaminase and less-specific ones to gluten.

As mentioned above, the major cause of celiac disease is body gluten intolerance. Gluten proteins differ in the size of subunits that are built of 250 to 850 amino acid residues [Kasarda, 2003]. In addition, these proteins are quite atypical due to a high content of glutamine and proline residues (35% and 15%, respectively). The presence of proline disturbs the linear structure of a polypeptide chain (proline residues form the "knots"). The occurrence of proline in the vicinity of hydrolysed bonds restricts the range of proteolytic enzymes activity. Hence, the presence of these residues often determines the sequence of fragments released as a result of proteolysis [Dewar *et al.,* 2004].

A quantitative evaluation of the toxicity of a given cereal species is extremely difficult. However, according to Kasarda [2003] it is at least possible to state that, compared to wheat, a lack of A-gliadin (one of the best-recognized fractions occurring in this cereal species) in rice *(Oryza sativa)* or barley reduces their toxicity in the grain of the above cereals. Investigations into celiac disease carried out so far have

Author's address for correspondence: prof. dr hab. Jerzy Dziuba, Department of Food Biochemistry, University of Warmia and Mazury in Olsztyn, Pl. Cieszyński 1, 10-726 Olsztyn, Poland; tel. (48 89) 523 37 15; e-mail: dziuba@uwm.edu.pl

demonstrated that the toxicity of cereal proteins is linked with the release of peptides upon digestion of these proteins. Peptides originating from wheat gliadins have been best characterised to date. Cornell [1996] and De Ritis [1988] found motifs with the following sequences: QQQP, QQPY, PSQQ, QPYP, to be responsible for celiac disease. These motifs also occur in a number of the so-called "non-toxic proteins", thus the incidence of celiac disease may be related to the so-called "extended structural motif" [McLachlan *et al.,* 2002]. According to McLachlan *et al.* [2002], extended structural motif is defined as a peptide sequence that contains toxic motifs and have been extended by means of including neighboring amino acid residues selected from the sequence of A-gliadin. Vader *et al.* [2002a] described motifs susceptible to transglutaminase action. The same researchers found celiac toxic fragments in barley, rye and oat protein sequences on the basis of computer-aided homology searching [Vader *et al.,* 2002b]. The computer analysis showed that 128 wheat, oat and rye proteins contain potentially toxic fragments, which can be released by digestive tract enzymes such as: chymotrypsin, trypsin and elastase [Shan *et al.,* 2005].

A study was undertaken, therefore, to estimate the potential toxicity of selected proteins of cereals and leguminous plants based on the designed parameters available in a database [Dziuba & Iwaniak, 2006]. The protein sequences examined were subjected to a structural analysis. The evolutionary (structural) similarity was determined between proteins commonly regarded as toxic and proteins of the analysed cereals and leguminous plants. In addition, *in silico* proteolysis processes were designed to detect which of the endopeptidases can potentially release the celiac-toxic peptides and cannot be recommended in food for patients suffering from this disease.

MATERIALS AND METHODS

Evaluation of the potential toxicity of selected proteins of cereals and leguminous plants based on computational methods. Experiments were carried out on amino acid sequences of wheat *(Triticum aestivum)*, oat *(Avena sativa)*, rice *(Oryza sativa)*, barley *(Hordeum vulgare)*, buckwheat *(Fagopyrum sagitatum)* and pea *(Pisum sativum)* originated from the Internet SWISS-PROT database [www.expasy.org] [Apweiler *et al.,* 2004] and inserted into the BIOPEP database [http://www.uwm.edu.pl/biochemia] [Dziuba & Iwaniak, 2006]. The first stage of investigations involved analysis of 403 protein sequences, including: wheat (144), oat (23), rice (76), barley (58), buckwheat (13), pea (53), sequences of trypsin and alpha-amylase inhibitors (24), 12 sequences of thaumatins as well as sequences of HMW (high molecular weight) and LMW (low molecular weight) glutenins. To answer the question which of them are the strongest precursors of celiac toxic peptides, the following qualitative and quantitative factors of protein evaluation were applied: location of the toxic motif in a protein sequence (the so-called profile of the potential activity of a given protein) and the occurrence frequency of toxic peptides in a protein chain (factor A) [Dziuba & Iwaniak, 2006]. A detailed description of the above-mentioned discriminants is given below.

Detection of fragments responsible for celiac disease. To detect the presence of motifs responsible for celiac disease (Table 1) in the sequences of the analysed proteins, use was made of the option "Profile of the potential biological activity of proteins" available in the BIOPEP database [Dziuba *et al.*, 2004]. This profile is defined as the type and location of a bioactive fragment in a protein sequence.

Evolutionary similarity of potentially-toxic proteins. The degree of the evolutionary similarity of sequences of the proteins examined was determined on the basis of a BLAST (Basic Local Alignment Search Tool) algorithm available on the Internet at: http://www.expasy.org/tools/blast/ [Korf, 2003]. This algorithm requires the insertion of a standard of a sequence, in respect of which the similarity degree is determined. The standard was estimated by means of the criterion defined as the occurrence frequency (A) of toxic fragments in a protein. Discriminant A is described as follows: A=a/N, where: a – the number of fragments with a given activity in a protein chain, and N – the number of amino acid residues.

The protein which possessed the highest value of A parameter calculated for tetrapeptides as well as extended motifs was used as the standard.

Secondary structure of structural motifs. Prediction of a secondary structure of structural motifs used was performed by the computer program PREDICT 7 [Carménes *et al.*, 1989]. Due to the fact that 3.6 amino acid residues are ascribed to one turn of a helix, the secondary structure was determined only for the extended sequential motifs of wheat which contained 14 to 15 amino acid residues (Table 1). The results of secondary structure prediction were to be used in the search for relationships between the structure and potential toxicity of the structural motifs.

TABLE 1. Sequence of celiac-toxic tetrapeptides and their extended structural motifs in A-gliadin of wheat (*Triticum aestivum)* – BIOPEP database.

No.	ID- BIOPEP database	Sequence	Number of amino acid residues
Celiac toxic tetrapeptides			
1.	6986	QQQP	4
2.	6987	PSQQ	4
3.	6988	QQPY	4
4.	6989	QPYP	4
Extended structural motifs			
5.	6990	PQN**PSQQQP**QEQVP	14
6.	6991	QQFLG**QQQP**FPPQQ	14
7.	6992	QQQKQ**QQQP**SSQVS	14
8.	6993	LQPQN**PSQQ**QPQEQ	14
9.	6994	QPQPF**PSQQ**QNPQAQ	15
10.	6995	QPFPPQ**QPYP**QPQP	14
11.	6996	QPFR**PSQQPY**LQLQP	15
12.	6997	QPFRP**QQPYP**QPQP	14
13.	6998	PFPP**QQPYP**QPQPF	14
14.	6999	PFRP**QQPYP**QPQPQ	14
15.	7044	QPQPF**PSQQ**NPQAQ	14

Design of the processes of proteolytic release of toxic peptides. Design of the proteolysis of selected proteins of wheat, oat and barley was performed with the use of the Internet database of proteins and biologically active peptides BIOPEP, developed at the Department of Food Biochemistry. Analyses were carried out following the scheme presented in Figure 1.

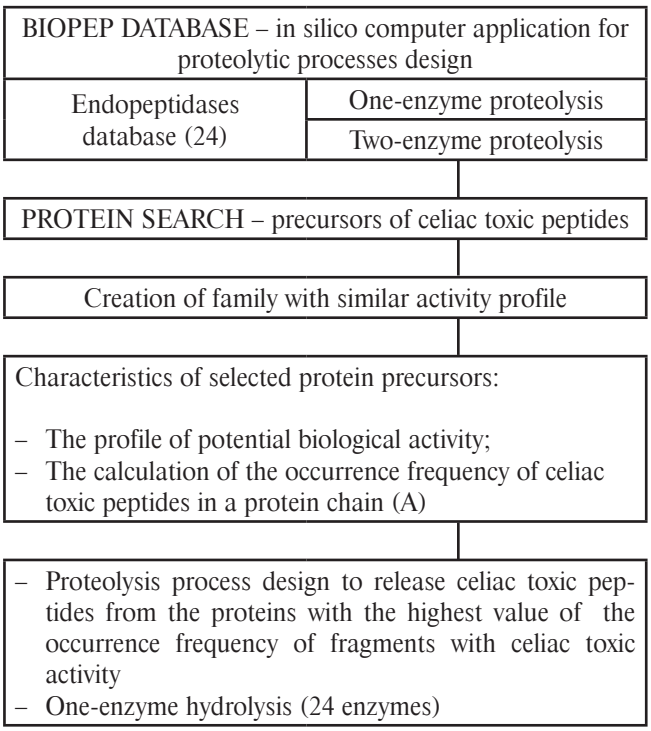

FIGURE 1. Scheme of protein hydrolysis design in the BIOPEP database.

In the BIOPEP database, data was compiled on 24 following proteolytic enzymes: chymotrypsin A, trypsin, pepsin, proteinase K, pancreatic elastase, prolyl oligopeptidase, V-8 protease (glutamyl endopeptidase), thermolysin, plasmin, cathepsin G, clostripain, chymase, papain, ficain, leukocyte elastase, chymotrypsin C, metridin, trombin, bromelain, pancreatic elastase II, glutamyl endopeptidase II, oligopeptidase B, calpain, and glycyl endopeptidase. The following data concerning the enzyme is provided in the BIOPEP database: identity number – ID; classification number – EC; identification sequence – a sequence of amino acid recognized by a given enzyme; cut sequence – amino acid residue typical for a given enzyme that forms a bond hydrolysed by that enzyme, and the specificity (hydrolysis in C/N – terminus). The results are obtained automatically from the BIOPEP database after selection of protein sequence and enzyme (maximum three). As a result of this action, a list of bioactive fragments released from a precursor protein by a given enzyme/enzymes is generated. All enzymes available in BIOPEP were applied to design proteolytic processes to find out which of them should be avoided and not recommended in the functional food production as the most efficient in celiac-toxic peptides liberation.

RESULTS AND DISCUSSION

Detection of fragments responsible for celiac disease Out of the 403 analysed amino acid sequences of select-

TABLE 2. The highest values of the occurrence frequency (A) of tetrapeptides and extended structural motifs in wheat *(Triticum aestivum)* proteins.

BIOPEP ID	Name	Number of amino acid residues	$A_{tetrapep}$	$A_{extended}$
1435	alpha-gliadin fragment (celiac active peptide), wheat *(Triticum aestivum)*	54	0.16	0.055
1189	omega-gliadin fragment, wheat *(Triticum aestivum)*	28	0.11	0
1436	alpha-gliadin fragment (celiac active peptide), wheat *(Triticum aestivum)*	54	0.11	0.037
1303	LMW glutenin, storage protein, wheat *(Triticum aestivum)*	359	0.05	0
1332	LMW glutenin, wheat, *(Triticum aestivum)*	356	0.05	0
1385	LMW glutenin, wheat *(Triticum aestivum)* subunit group 4 type II	340	0.05	0
1334	LMW glutenin, wheat *(Triticum aestivum)*	384	0.049	0
1298	LMW glutenin 3, fragment, wheat *(Triticum aestivum)*	373	0.048	0
1335	LMW glutenin, wheat *(Triticum aestivum)*	376	0.048	0
1314	LMW glutenin, wheat *(Triticum aestivum)*	276	0.047	0
1333	LMW glutenin, wheat *(Triticum aestivum)*	376	0.047	0
1313	LMW glutenin, wheat *(Triticum aestivum)*	261	0.046	0
1329	LMW glutenin, wheat *(Triticum aestivum)*	388	0.046	0
1331	LMW glutenin, wheat *(Triticum aestivum)*	388	0.046	0
1418	LMW glutenin, wheat *(Triticum aestivum)*	280	0.046	0
1358	LMW glutenin, wheat *(Triticum aestivum)*, fragment	220	0.045	0
1384	LMW glutenin, wheat *(Triticum aestivum)* subunit group 4 type II	297	0.044	0
1421	alpha-gliadin, wheat *(Triticum aestivum)*	269	0.044	0.033
Other wheat proteins analysed by BIOPEP				
1372*, 1373, 1308, 1433, 1283, 1338, 1388, 1390, 1391, 1393, 1420, 1423, 1424, 1425, 1429, 1430, 1178, 1306, 1307, 1427, 1427, 1439, 1392, 1179, 1353, 1422, 1428, 1180, 1312, 1337, 1350, 1395, 1351, 1383, 1386, 1183, 1184, 1311, 1425, 1181, 1310, 1382, 1278, 1305, 1394, 1182, 1304, 1145, 1285, 1301, 1302, 1355, 1378, 1387, 1389, 1254, 1300, 1336, 1375, 1399, 1414, 1297, 1339, 1354, 1187, 1349, 1365, 1381, 1410, 1417, 1318, 1409, 1411, 1412, 1414, 1415, 1147, 1185, 1321, 1366, 1368, 1369, 1380, 1413, 1299, 1315, 1330, 1377, 1398, 1186, 1367, 1379, 1374, 1345, 1346, 1376, 1403, 1404, 1419, 1309, 1317, 1320, 1400, 1405, 1281, 1322, 1327, 1328, 1402, 1401, 1361, 1407, 1148, 1348, 1396, 1408, 1406, 1364 **(total = 136 sequences)**				

* The identification numbers of other wheat proteins tested by BIOPEP software.

ed cereals and seeds, 155 were found to contain tetrapeptides known to be responsible for celiac disease (sequences: QQQP, QQPY, PSQQ, QPYP). 29 out of 403 sequences contained tetrapeptides which were the constituents of a structural motif with a longer sequence. Such fragments had not been detected in the remaining 248 sequences and were rejected for further analysis. The presence of tetrapeptides was confirmed in 136 sequences of wheat proteins (Table 2), 9 sequences of oat proteins (Table 3), and 10 sequences of barley proteins (Table 4). The highest number of tetrapeptides responsible for the induction of celiac disease was found in sequences of wheat (*Triticum aestivum*) alpha-gliadins. They occurred at various frequencies, depending on the length of the protein amino acid sequence (from 6 to 11 tetrapeptides) and were usually located in the N-terminal fragment of the chain. The number of extended structural motifs (the so-called "standard") occurring in these sequences ranged from 0 to 7. Their number was lower, compared with that of tetrapeptides, since often several shorter sequences occurred in one longer motif. For instance, in a motif with the sequence QPFPPQQPYPQPQP (wheat alpha-gliadin, identity number in the BIOPEP database - 1420), residues 36-49, the following fragments occurred: QQPY (residues 41-44) and QPYP (residues 42-45). Tetrapeptides were also detected in wheat alpha/beta gliadins (from 2 to 10, depending on the length of the protein chain), *i.e.* wheat alpha/beta gliadin with ID number 1278 contained five extended structural motifs, whereas alpha/beta gliadin with ID number 1183 appeared not to contain any motifs with longer sequences. Wheat omega-gliadin (ID number 1418) was found to contain all celiac-toxic sequences, *i.e.*: QPYP, QQPY, PSQQ, QQQP, that occurred in the number of: 4, 5, 2 and 2 respectively. No motifs with an extended sequence were detected. The protein sequences of wheat gamma-gliadins were observed to be predominated by the motif with the QQQP sequence – in class b-I gamma-gliadins, this sequence appeared six times (*i.e.* the most frequently of all analysed sequences of that protein). In addition, fragments with sequences QPYP, QQPY also appeared as a part of the longer sequence QQPYPQ. None of the gamma-gliadins contained the PSQQ motif. In the case of the HMW glutenins, the repeating "common" sequence was the QQQP fragment (depending on the length of a polypeptide chain, this sequence occurs from 1 to 5 times and is located in the C-terminal fragment of protein), whereas the appearance of the PSQQ sequence (one fragment per entire protein sequence) was typical of HMW glutenin subunits. Fragments with the same sequences predominated also in the LMW glutenins. The profile of the potential activity of these proteins revealed a maximum of 19 QQQP fragments, both in the N- and C-terminal fragment of protein (minimally 2). Moreover, the PSQQ sequence was observed to occur sporadically in the longer fragment PSQQQP. Sequences of barley proteins – hordeins *(Hordeum vulgare)* contained mainly tetrapeptides with the sequences QPYP, QQPY, and QQQP. The first two were present in a longer sequential fragment QQPYPQ. The QPYP and QQQP fragments occurred in oat avenins *(Avena sativa)*. Sequences of barley proteins (hordeins) were reported not to contain extended structural motifs typical of wheat A-gliadin.

TABLE 3. Occurrence frequency (A) of tetrapeptides and extended structural motifs in oat *(Avena sativa)* proteins.

BIOPEP ID	Name	Number of amino acid residues	$A_{tetrapep}$	$A_{extended}$
1450	gamma 3 avenin, fragment, oat *(Avena sativa)*	29	0.038	0
1451	avenin-3 precursor, oat *(Avena sativa)*	220	0.038	0
1452	avenin precursor, oat *(Avena sativa)*	214	0.038	0
1457	avenin, precursor, oat *(Avena sativa)*	181	0.038	0
1458	avenin, precursor, oat *(Avena sativa)*	222	0.038	0
1459	avenin A, fragment, oat *(Avena sativa)*	36	0.038	0
1460	avenin E, fragment, oat *(Avena sativa)*	52	0.038	0
1461	avenin F, fragment, oat *(Avena sativa)*	43	0.038	0
1462	Avenin N9, oat *(Avena sativa)*	182	0.038	0

TABLE 4. Occurrence frequency (A) of tetrapeptides and extended structural motifs in barley *(Hordeum vulgare)* proteins.

BIOPEP ID	Name	Number of amino acid residues	$A_{tetrapep}$	$A_{extended}$
1602	gamma-hordein 3, barley *(Hordeum vulgare)*	289	0.031	0
1607	C-hordein (Clone PC-919), fragment, barley *(Hordeum vulgare)*	68	0.029	0
1629	gamma-hordein 3, fragment, barley *(Hordeum vulgare)*	40	0.025	0
1630	gamma-hordein 1 and 2, fragment, barley *(Hordeum vulgare)*	35	0.025	0
1603	B1-hordein, precursor, barley *(Hordeum vulgare)*	293	0.017	0
1623	B hordein, precursor, barley *(Hordeum vulgare)*	290	0.017	0
1628	putative gamma 2 hordein, fragment, barley *(Hordeum vulgare)*	255	0.016	0
1604	B3-hordein, fragment, barley *(Hordeum vulgare)*	264	0.015	0
1150	gamma-hordein 1 precursor, barley *(Hordeum vulgare)*	305	0.013	0
1627	C-hordein, barley *(Hordeum vulgare)*	310	0.0065	0

The presence of tetrapeptides containing Glu, Pro or Tyr residues in cereal proteins is linked with the occurrence of a toxic effect [McLachlan *et al.,* 2002]. Their presence seems to be insufficient since a number of proteins of different species, commonly acknowledged as non-toxic, contain such motifs as QQQP and PSQQ [Carménes *et al.,* 1989]. Based on their studies, Cornell & Mothes [1993] as well as MacLachlan *et al.* [2002], postulated that in the evaluation of the toxicity of a given protein, consideration should be given not only to the presence of toxic tetrapeptides but also to the sequence of an extended motif containing these tetrapeptides. The results of these studies [McLachlan *et al.,* 2002; Cornell & Mothes, 1993] indicated that the toxic activity of sequences of peptides with longer chains was higher than that of shorter sequences, and that the mucosal membrane was affected to a greater extent by extended motifs. This means that the occurrence of the toxic effect is dependent upon the presence of lateral chains formed by specific amino acids, *e.g.* arginine, N-terminal amino acid in the sequence RPQQPYPQPQPQ, probably responsible for the interaction with the receptor.

Evolutionary similarity of potentially-toxic proteins

Tables 2–4 present the results of calculations of parameter A for the toxic activity in celiac disease patients. At the initial stage of the experiment, parameter A was computed only for the extended structural motif displaying toxic properties. Since the value of this factor equaled 0 for the majority of proteins other than wheat gliadins, it was also calculated for tetrapeptides inducing celiac disease. The highest values for the A discriminant calculations made for the wheat proteins are shown in Table 2 (total number of protein sequences = 136). Value A was also made for the other wheat proteins. They are mentioned in the Table 2 by their BIOPEP identification numbers. The highest value of parameter A ($A_{tetrapeptide}$=0.16, $A_{extended}$=0.0055) was observed for wheat alpha-gliadin – ID number in the BIOPEP database – 1435. However, this sequence was made up by 54 amino acid residues only. Out of the wheat proteins, the highest A value was reported for LMW glutenin (subunit 4, BIOPEP ID number – 1303), yet for the so-called extended motif, the value of parameter A was equal to 0. The protein with the highest occurrence frequency of toxic tetrapeptides and extended motifs, *i.e.* 0.044 and 0.033 respectively, appeared to be wheat alpha-gliadin with a chain length equal to 269 amino acid residues (ID number 1421). That protein was accepted as a standard sequence for the analysis of sequential similarity between the other analysed proteins with the use of the BLAST algorithm, and the results for the all calculations are available in the BIOPEP database. Depending on the values of parameter A, the other proteins were in the following descending order: wheat alpha/beta gliadins, wheat gamma-gliadins, glutenins, oat avenins, and barley hordeins. Except for alpha- and alpha/beta gliadins of wheat, the value of parameter A for the extended motif was equal to 0. *In vitro* and *in vivo* analysis confirmed that, apart from wheat proteins, celiac disease is induced by other proteins, *e.g.* prolamins of barley and rice. They are responsible for damage to the mucous membrane as well as for improper absorption of carbohydrates, proteins, lipids, and vitamins [Rocher *et al.,* 1995]. Oat avenins demonstrated a higher value of parameter $A_{tetra-peptide}$, compared to that of barley and rice proteins. Rocher *et al.* [1995] identified a few prolamins of wheat and oat as proteins with a positive immunological response, and classified them based on the N-terminal sequence as α- and γ-prolamins. According to Kasarda [2003], oat avenins show affinity to the C-terminal fragment of alpha- and gamma-gliadin, yet they do not contain domains rich in the proline of glutamine residues that are associated with toxicity. The toxicity of oat is likely to result from problems of a practical origin. Oat is cultivated interchangeably with, or in the vicinity of, wheat. The same containers are used for storage of both the species, which may lead to oat contamination by wheat [Cornell, 1996].

Sequential similarity between the proteins was determined by means of the BLAST algorithm [http://www.ncbi.nhm.nih.gov/blast]. Table 5 contains a compilation of 26 proteins grouped in a descending order in terms of sequential similarity. Compared to wheat alpha-gliadins (BIOPEP

TABLE 5. Structural homology between proteins – precursors of toxic peptides and wheat A-gliadin (BIOPEP ID - 1421).

No.	Protein	Number of amino acid residues	Similarity
1.	alpha/beta-gliadin precursor (Prolamin)	286	80%
2.	alpha/beta-gliadin clone PTO-A10 (Prolamin)	186	79%
3.	alpha/beta-gliadin A-II precursor (Prolamin)	291	69
4.	alpha/beta-gliadin MM1 precursor (Prolamin)	307	71%
5.	alpha/beta-gliadin A-IV precursor (Prolamin)	297	69%
6.	alpha/beta-gliadin clone PW1215 precursor (Prolamin)	296	73%
7.	alpha/beta-gliadin A-III precursor (Prolamin)	282	71%
8.	alpha/beta-gliadin clone PW8142 precursor (Prolamin)	313	63%
9.	alpha/beta-gliadin A-V precursor (Prolamin)	319	60%
10.	alpha/beta-gliadin A-I precursor (Prolamin)	262	74%
11.	prolamin PPROL 17 precursor	149	47%
12.	gamma-gliadin precursor	251	40%
13.	avenin-3 precursor (Prolamin)	220	40%
14.	gamma-gliadin precursor	327	39%
15.	gamma-gliadin B precursor	291	39%
16.	gamma-gliadin precursor	302	39%
17.	gamma-hordein 3	289	36%
18.	13 kDa prolamin precursor	156	36%
19.	gamma-hordein 1 precursor	305	35%
20.	B1-hordein precursor	293	35%
21.	gamma-gliadin (Gliadin B-III)	244	35%
22.	glutenin, subunit 1D1 precursor	307	34%
23.	B3-hordein	264	34%
24.	LMW glutenin subunit precursor	356	33%
25.	gamma-gliadin B-I precursor	304	32%
26.	LMW glutenin subunit PTDUCD1 precursor	295	32%

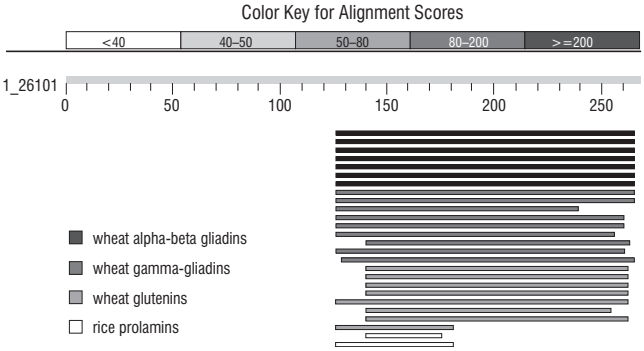

FIGURE 2. Graphical scheme of the structural similarity of cereal proteins.

ID 1421), the highest degree of similarity was demonstrated by wheat alpha-beta gliadins (74-80%), followed by rice prolamins (47%), oat avenin (40%), wheat gamma-gliadins (32-39%), barley hordeins (35%) and LMW glutenins (32%). Figure 2 presents a graphic distribution of similarity between protein sequences (similarity degree over 40%) according to the BLAST algorithm. The main fragment corresponding to points 0 – 260 is the sequence of wheat alpha-gliadin in respect of which the other sequences were grouped. Red-colored fragments correspond to sequences of wheat alpha/beta-gliadins; pink-colored ones to wheat gamma-gliadins; green-colored ones to wheat glutenins, and gray-colored ones to rice prolamins. The above hierarchy differs from that of sequences grouped according to decreasing values of the frequency of occurrence of toxic motifs in protein chains of the proteins. Hence, the value of parameter A calculated for avenin was higher than that computed for barley proteins and, obviously, for rice proteins – for which the value of parameter A was equal to 0. Wheat, rice and barley belong to the tribe Triticae or Hordeae. Proteins of oats belong to a different tribe. Still, a similarity exists between oat avenins and most wheat gliadins, resulting from a close affinity between the tribes [Kasarda, 2003]. Hence, these proteins are characterised by higher occurrence frequency of tetrapeptides, as compared to barley or rice (lack of tetramotifs).

Secondary structure of structural motifs

Table 6 contains sequences of extended structural motifs, contents of particular secondary structures, and the so-called "hydropathy index" according to Kyte & Doolitlle [1982]. Calculations were made automatically with PREDICT 7 software [Carménes *et al.,* 1989]. All extended fragments of A-gliadin were found to contain beta-turns and random coil. In one case, *i.e.* in the fragment with BIOPEP ID number 6996 (15 amino acid residues), a form of β-sheet occurred apart from the above-mentioned structures. In the analysed motifs, the contribution of beta-turn structure ranged from 14.3 to 64.3%, whereas that of the random coil – from 35.7 to 85.7%. In all of the eleven extended structural motifs the hydrophobicity was negative, which points to their hydrophilic character and indicates their easy accessibility for proteolytic enzymes during protein digestion [Dziuba *et al.,* 2004]. Short peptides formed as a result of wheat protein digestion provoke a toxic effect, hence, an understanding of issues linked with their structure is of key importance to a better understanding of the causes of celiac disease. For instance, the NPSQQQPQ motif (12–19 fragment of alpha-gliadin) as well as the fragment with the sequence QQPYPQPQ (position 77-84) are acknowledged as the most active. In the case of the former, the hydrophilic character of asparagine and lateral serine chains conjugated through beta-turns, which stems from the presence of proline residues, is assumed to be the key factor in celiac disease induction [Cornell & Wills-Johnson, 2001].

Tetrapeptides contained in the extended fragments of alpha-gliadin sequence occurred in the motifs with a structure of beta-turn or random coil. Cornell & Wills-Johnson [2001] postulated that beta-turn were a dominant structural trait of peptides originated from gliadins and containing proline residues. This trait affords the possibility of the formation of spatial systems that are likely to be "recognizable" to such immunological reactions as stimulation of lymphocytes. The folding of lateral chains caused by the presence of proline residues breaking hydrogen bonds may be of significance to the "orientation" of tyrosine residues as a characteristic trait of celiac-toxic peptides. In addition, tyrosine is recognized as having amino acid generating opioid properties and demonstrating a capacity for the production of gamma-interferone [Cornell & Wills-Johnson, 2001].

TABLE 6. Structural characteristics of extended motifs of A-gliadin of wheat *(Triticum aestivum)* – PREDICT 7 program.

No.	Sequence	Percentage of:				Hydrophobicity
		α-helix	β-turn	β-sheet	Random coil	
1.	PFPPQQPYPQPQPF	0	14.3	0	85.7	-1.624
2.	PFRPQQPYPQPQPQ	0	28.6	0	71.4	-2.248
3.	QPFRPQQPYPQPQP	0	28.6	0	71.4	-2.027
4.	QPFRPSQQPYLQLQP	0	28.6	28.6	42.8	-0.985
5.	QPFPPQQPYPQPQP	0	14.3	0	85.7	-1.938
6.	LQPQNPSQQQPQEQ	0	28.6	0	71.4	-2.569
7.	QPQPFPSQQQNPQAQ	0	26.7	0	73.3	-1.982
8.	QQQKQQQQPSSQVS	0	42.9	0	57.1	-2.184
9.	QQFLGQQQPFPPQQ	0	64.3	0	35.7	-1.284
10.	PQNPSQQQPQEQVP	0	28.6	0	71.4	-2.186
11.	QPQPFPSQQNPQAQ	0	21.4	0	78.6	-1.874

TABLE 7. Profile of the potential biological activity of alpha gliadin (ID 1420). A BIOPEP database report.

ID of protein:	1420	Name:	Alpha-gliadin, wheat (*Triticum aestivum*)
Protein sequence:	MVRVPVPQLQPQNPSQQQPQEQVPLVQQQQFPGQQQPFPPQQPYPQPQPFPSQQPYLQLQPFPQPQLPYPQPQL PYPQPQPFRPQQPYPQSQPQYSQPQQPISQQQQQQQQQQQQKQQQQQQQQILQQILQQQLIPCRDVVLQQHSIAYGSSQVL QQSTYQLVQQLCCQQLWQIPEQSRCQAIHNVVHAIILHQQQQQQQQQQQQPLSQVSFQQPQQQYPSGQGSFQPSQQNPQA QGSVQPQQLPQFEEIRNLALETLPAMCNVYIPPYCTIAPVGIFGTNYR		

ID	Name of peptide	Activity	Number	Sequence	Location
3460		antiamnestic	1	PG	[32-33]
3258	b-lactokinin	antihypertensive	1	IR	[257-258]
3341	ACE inhibitor	antihypertensive	1	FQP	[233-235]
3372		antihypertensive	5	PYP	[43-45],[68-70],[75-77],[82--84],[94-96]
3389	ACE inhibitor	antihypertensive	1	LW	[178-179]
3492	ACE inhibitor from sake	antihypertensive	1	VY	[271-272]
3502	ACE inhibitor (BSA fr. 221-222)	antihypertensive	4	FP	[31-32],[38-39],[50-51],[62-63]
3522	ACE inhibitor	antihypertensive	1	IPP	[273-275]
3542	ACE inhibitor	antihypertensive	2	LQP	[9-11],[59-61]
3550	ACE inhibitor	antihypertensive	1	YL	[56-57]
3553	ACE inhibitor	antihypertensive	1	YG	[156-157]
3563		antihypertensive	1	AY	[155-156]
3666		antihypertensive	6	YP	[44-45],[69-70],[76-77],[83-84],[95-96],[226-227]
3714		antihypertensive	4	LQQ	[133-135],[137-139],[149-151],[162-164]
2578		toxic	1	PSQQQP	[14-19]
2797	peptide toxic from alpha-gliadin	toxic	4	YPQPQ	[44-48],[69-73],[76-80],[83-87]
2798	peptide toxic from alpha-gliadin	toxic	2	YPQPQPF	[44-50],[83-89]
2800	peptide toxic from alpha-gliadin	toxic	2	QQPYPQ	[41-46],[92-97]
6986	celiac toxic peptide	toxic	3	QQQP	[16-19],[34-37],[210-213]
6987	celiac toxic peptide	toxic	3	PSQQ	[14-17],[51-54],[235-238]
6988	celiac toxic peptide	toxic	3	QQPY	[41-44],[53-56],[92-95]
6989	celiac toxic peptide	toxic	2	QPYP	[42-45],[93-96]
6990	potential celiac toxic peptide (extended motif)	toxic	1	PQNPSQQQPQEQVP	[11-24]
6993	potential celiac toxic peptide (extended motif)	toxic	1	LQPQNPSQQQPQEQ	[9-22]
6995	potential celiac toxic peptide (extended motif)	toxic	1	QPFPPQQPYPQPQP	[36-49]
6996	potential celiac toxic peptide (extended motif)	toxic	1	QPFPSQQPYLQLQP	[48-61]
6998	potential celiac toxic peptide (extended motif)	toxic	1	PFPPQQPYPQPQPF	[37-50]
7001	synthetic peptide from alpha-gliadin	toxic	1	QLQPQNPSQQQP	[8-19]
7002	synthetic peptide from alpha-gliadin	toxic	1	PSQQQPQEQVPL	[14-25]
7003	synthetic peptide from alpha-gliadin	toxic	1	QEQVPLVQQQQF	[20-31]
7006	synthetic peptide from alpha-gliadin	toxic	1	FPPQQPYPQPQP	[38-49]
7007	synthetic peptide from alpha-gliadin	toxic	1	YPQPQPFPSQQP	[44-55]
7008	synthetic peptide from alpha-gliadin	toxic	1	FPSQQPYLQLQP	[50-61]
7009	synthetic peptide from alpha-gliadin	toxic	1	YLQLQPFPQPQL	[56-67]
7014	synthetic peptide from alpha-gliadin	toxic	1	PQYSQPQQPISQ	[100-111]
7016	synthetic peptide from alpha-gliadin	toxic	1	QILQQILQQQLI	[131-142]
7026	synthetic peptide from alpha-gliadin	toxic	1	HNVVHAIILHQQ	[191-202]
7033	synthetic peptide from alpha-gliadin	toxic	1	QQNPQAQGSVQP	[237-248]
7034	synthetic peptide from alpha-gliadin	toxic	1	QGSVQPQQLPQF	[243-254]
7035	synthetic peptide from alpha-gliadin	toxic	1	QQLPQFEEIRNL	[249-260]
7038	synthetic peptide from alpha-gliadin	toxic	1	AMCNVYIPPYCT	[267-278]
3285		antithrombotic	1	PG	[32-33]
3770		immunomodulating	1	YG	[156-157]
2882	opioid fragment of alpha lactorphin fr:50-51	opioid	1	YG	[156-157]
3474	alpha lactorphin fr:91-92	opioid	1	YL	[56-57]
3305		antioxidative	1	LH	[199-200]
2890		inhibitor	2	GQ	[33-34],[229-230]
3166	diprotin B	inhibitor	1	VPL	[23-25]
3170	dipeptidyl peptidase IV inhibitor	inhibitor	2	PP	[39-40],[274-275]
3175	dipeptidyl-aminopeptidase IV inhibitor	inhibitor	1	LA	[260-261]
3177	dipeptidyl-aminopeptidase IV inhibitor	inhibitor	1	AP	[280-281]
3178	dipeptidyl-aminopeptidase IV inhibitor	inhibitor	4	FP	[31-32],[38-39],[50-51],[62-63]
3179	dipeptidyl-aminopeptidase IV inhibitor	inhibitor	1	PA	[266-267]
3180	dipeptidyl-aminopeptidase IV inhibitor	inhibitor	5	LP	[67-68],[74-75],[81-82],[251--252], [265-266]
3181	dipeptidyl-aminopeptidase IV inhibitor	inhibitor	3	VP	[4-5],[6-7],[23-24]
3183	dipeptidyl-aminopeptidase IV inhibitor	inhibitor	2	VV	[147-148],[193-194]
3184	dipeptidyl-aminopeptidase IV inhibitor	inhibitor	1	HA	[195-196]
4006		activating ubiquitin-mediated proteolysis	1	LA	[260-261]

***In silico* proteolysis of toxic peptide precursors**

The amino acid sequences of wheat, oat and barely proteins in the BIOPEP database were characterised (as potential precursors of celiac-toxic peptides) by determining profiles of their potential biological activity and calculating parameter A. An example profile of the potential biological activity of alpha gliadin (ID 1420) is presented in Table 7. The amino acid sequence of this protein contains eighty fragments corresponding to peptides with different activity, including as many as thirty-nine which are toxic to celiac disease patients. Among the toxic peptides are both short tetra/pentapeptides containing mainly proline (P) and glutamine (Q) residues, as well as longer, extended motifs containing up to fourteen amino acids.

The stage of *in silico* proteolysis was carried out on twenty-eight wheat proteins selected for the highest value of parameter A calculated for tetrapeptides and extended structural motifs ($A_{tetrapep.}$, $A_{extended}$). Two sequences of oat proteins and one sequence of barley protein were selected out of the other proteins. A computer-aided simulation of proteolysis with twenty-four enzymes (occurring in the BIOPEP database) was carried out. On the basis of previous studies [Dziuba *et al.,* 1999], the assumption was made that the highest value of A factor and consequently higher number of toxic motifs encrypted in a protein sequence (richer profile of potential celiac toxic activity of crop protein), provides a higher probability of the release of these fragments *via* enzymatic hydrolysis.

The *in silico* proteolysis run demonstrated the possibility of releasing tetrapeptides or toxic peptides (extended motifs) from twenty-eight out of the thirty-one proteins analysed. Only such proteolytic enzymes as thermolysin, K proteinase and prolyl oligopeptidase were found to release celiac-toxic peptides. Thermolysin released peptides from twenty-five of the proteins examined. These were extended fragments with the following sequences: LQPQNPSQQQPQEQ and FPPQQPYPQPQP. Proteinase K and prolyl oligopeptidase released QQQP tetrapeptides only from four proteins analysed. The highest number of QQQP fragments (four/five) was obtained from gamma-hordein (*Hordeum vulgare*) – ID 1602, using proteinase K or proline oligopeptidase, respectively. The following proteins: alpha-gliadin, wheat *(Triticum aestivum)*– ID 111304 and avenin precursors of oat *(Avena sativa)* – ID 1452 and 1458 failed to provide fragments with such an activity. It should be mentioned that the presence of four amino acid motifs gives rise to toxic effect when flanked by other specific amino acid residues. Prolyl endopeptidase [EC 3.4.21.26] can be recommended as an enzyme for destroying celiac toxic peptides as was confirmed by Shan *et al.* [2002, 2004]. Proteinase K [3.4.21.14] is an enzyme with broad spectrum specificity and may be a potential alternative for prolyl endopeptidase [Dziuba *et al.,* 1995]. Proline residue specific enzymes can be recommended for releasing many peptides with different activities [Dziuba *et al.,* 2004].

The assumption made in the *in silico* research is rather a simplification and does not cover the impact of proper physicochemical conditions necessary for carrying out efficient hydrolysis nor the physiological aspects determining the release and assimilation of bioactive peptides in the body [Adler-Nissen, 1986; Friedman, 1996; Vorob'ev & Goncharova, 1998]. However, it takes into account the location of active fragments in the amino acid sequence of a protein, its structure and specificity of the activity of enzymes applied. In many cases this is sufficient to evaluate the possibilities of bioactive peptide release from their protein precursors [Dziuba *et al.,* 2004].

CONCLUSIONS

1. Tetrapeptides (QQQP, QQPY, PSQQ, QPYP) regarded as responsible for the induction of celiac disease occurred in sequences of proteins of wheat (gliadins), barley (hordeins), and oat (avenins). In the case of wheat proteins, the enumerated motifs constituted so-called "extended sequential motifs".

2. Neither extended motifs nor tetrapeptides were detected in sequences of rice-, pea- or buckwheat proteins, thaumatins or inhibitors of trypsin and alpha-amylases. This may result from weak evolutionary bonds between these proteins and wheat gliadins, or from a complete lack of such bonds.

3. The highest value of parameter A for wheat proteins confirms that they may be the main source of celiac-toxic peptides. The relatively low values of parameter A reported for oat avenins confirmed that they are evolutionary-related with alpha- and gamma-gliadins of wheat – which is likely to affect their toxicity.

4. The presence of tetrapeptides inducing celiac disease is not a prerequisite for protein toxicity evaluation. The similarity between proteins of different species within longer fragments of sequences might also determine their toxicity.

5. A dominant feature of celiac disease-inducing motifs is the occurrence of a beta-turn structure, as well as fragments with a sequence hydrophilic in character. This may be an important guideline for the design of forms recognizable to systems indispensable for celiac disease generation.

6. Out of the 24 enzymes gathered in the BIOPEP database, only thermolysin, proteinase K and proline oligopeptidase released celiac-toxic peptides. Food design for the population affected by this disease should entail the elimination of the above enzymes from the production process. However, the results obtained require experimental validation.

ACKNOWLEDGEMENTS

This work was supported by project No. PBZ-KBN-097/PZ06/2003/1.1.

REFERENCES

1. Adler-Nissen J., Enzymic Hydrolysis of Food Proteins. 1986, Elsevier, Applied Science Publishers, Copenhagen, pp. 57–109.
2. Apweiler R., Bairoch A., Wu C.H., Protein sequence databases. Curr. Opin. Chem. Biol., 2004, 8, 76–80.
3. Breteneder H., Plant-food and seafood allergens – an overview, Allergy, 1998, 53 (Suppl. 46), 31–4.
4. Bruijzeel-Koomen C., Ortolani C., Aas K., Adverse reactions to food: position paper. Allergy, 1995, 50, 623–35.
5. Carménes R.S., Freije J.P., Molina M.M., Martin J.M., Predict 7, a program for protein structure prediction. Biochem. Biophys. Res. Commun., 1989, 2, 687–693.
6. Cornell H.J., Coeliac disease: a review of the causative agents and their possibly mechanisms of action. Amino Acids, 1996, 10, 1–19.

7. Cornell H.J., Mothes T., The activity of wheat gliadin peptides in *in vitro* assays for coeliac disease. Biochim. Biophys. Acta, 1993, 1181, 169–173.
8. Cornell H.J., Wills-Johnson G., Structure-activity relationships in coeliac-toxic gliadin peptides. Amino Acids, 2001, 21, 243–253.
9. De Ritis G., Auricchio S., Jones H.W., Lew E.J-L., Bernardin J.E., Kasarda D.D., *In vitro* (organ culture) studies of the toxicity of specific A-gliadin peptides in celiac disease. Gastroenterology, 1988, 94, 41–9.
10. Dewar D., Pereira S. P., Ciclitira P. J., The pathogenesis of coeliac disease. Int. J. Biochem. Cell Biol., 2004, 36, 17–24.
11. Dziuba J., Iwaniak A., Database of bioactive peptide sequences**.** 2006, *in:* Nutraceutical Proteins and Peptides in Health and Disease (eds. Y. Mine, F. Shahidi). CRC Press Inc., Boca Raton, London, New York, pp. 543–564.
12. Dziuba J., Minkiewicz P., Nałęcz D., Iwaniak A., Database of biologically active peptide sequences. Nahrung, 1999, 43, 190–195.
13. Dziuba J., Minkiewicz P., Puszka K., Wheat gliadins as potential precursors of bioactive peptides. Acta Acad. Agricult. Tech. Olst., Technologia Alimentorum, 1995, 28, 3–16.
14. Dziuba J., Niklewicz M., Iwaniak A**.**, Darewicz M., Minkiewicz P., Bioinformatic-aided prediction for release possibilities of bioactive peptides from plant proteins. Acta Alim. Hungarica, 2004, 33, 227–235.
15. Egan C.A., Smith E.P., Taylor T.B., Eyer L.J., Samowitz W.S., Zone J.J., Linear IgA bullous dermatosis responsive to a gluten-free diet. Am. J. Gastroenterol., 2001, 96, 1927–29.
16. Farell R.J., Kelly C.P., Celiac sprue. N. Engl. J. Med., 2002, 346, 180–88.
17. Friedman M., Nutritional value of proteins from different food sources. A review. J. Agric. Food Chem., 1996, 44, 6–29.
18. Kasarda D.D., 2003, [http://www.pw.usda.gov/ggpages/topics/celiac.vs.garins.html].
19. Kasarda D.D., 2002, [http://www.wheat.pw.usda.gov/ggpages/topics/celiac.html].
20. Korf I., Serial BLAST searching. Bioinformatics, 2003, 19, 1492–1496.
21. Kyte J., Doolittle R.F., A simple method for displaying the hydropathic character of a protein. J. Mol. Biol., 1982, 157, 105–132.
22. McLachlan A., Gullis P.G., Cornell H.J., The use of extended amino acid motifs for focussing on toxic peptides in coeliac disease. J. Biochem. Mol. Biol. Biophys., 2002, 6, 319–324.
23. Nieuwenhuizen W.F., Pieters R.H.H., Knippels L.M.J., Jansen M.C.J.F., Koppelman S.J., Is Candida albicans a trigger in the onset of coeliac disease? The Lancet, 2003, 361, 2152–2154.
24. Rocher A., Calero M., Soriano F., Mèndez E., Identification of major rye secalins as coeliac immunoreactive proteins. Biochim. Biophys. Acta, 1995, 1295, 13–22.
25. Sampson H.A., Anderson J.A., Summary and recommendations: classification of gastrointestinal manifestations due to immunologic reactions to foods in infants and young children. J. Pediatr. Gastroenterol. Nutr., 2000, 30, S87–94.
26. Sampson H.A., Update on food allergy. J. Allergy Clin. Immunol., 2004, 113, 805–19.
27. Schuppan D., Ciccocioppo R., Coeliac disease and secondary autoimmunity. Digest Liver Dis, 2002, 34, 13–15.
28. Shan L., Martin T., Sollid L. M., Gray G. M., Khosla C., Comparative biochemical analysis of three bacterial prolyl endopeptidases: implications for coeliac sprue. Biochem. J., 2004, 383, 311–318.
29. Shan L., Quiao S.-W., Arentz-Hansen H., Molberg Ø., Gray G. M., Sollid L.M., Khosla C., Identification and analysis of multivalent proteolytically resistant peptides from gluten: implications for celiac sprue. J. Proteome Res., 2005, 4, 1732–1741.
30. Shan. L., Molberg Ø., Parrot I., Hausch F., Filiz F., Gray G.M., Sollid L.M., Khosla C., Structural basis for gluten intolerance in celiac sprue. Science, 2002, 297, 2275–2279.
31. Sicherer S.H., Food alergy. Lancet, 2002, 360, 701–710.
32. Sollid L.M., Coeliac disease: dissecting a complex inflammatory disorder. Nature Rev. Immunol., 2002, 2, 647–655.
33. Vader L.W., de Ru A., van der Wal I., Kooy Y.M.C., Benckhuijsen W., Mearin M.L., Drijfhout J.W., van Veelen P., Koning F., Specificity of tissue transglutaminase explains cereal toxicity in celiac disease. J. Exp. Med., 2002a, 5, 643–649.
34. Vader L.W., Kooy Y., van Veelen P., de Ru A., Harris D., Benckhuijsen W., Pena S., Mearin L., Drijfhout J.W., Koning F., The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. Gastroenterology, 2002b, 122, 1729–1737.
35. van Belzen M.J., Mulder C.J.J., Pearson P.L., Houwen R.H.J., Wijmenga C., The tissue transglutaminase gene is not primary factor predisposing to celiac disease. Am. J. Gastroenterol., 2001, 96, 3337–3340.
36. Vorob'ev M.M., Goncharova I.A., Computer simulation of proteolysis. Peptic hydrolysis of partially demasked β-lactoglobulin. Nahrung, 1998, 42, 61–67.

Received September 2006. Revision received and accepted January 2007.

BIOINFORMATYCZNA CHARAKTERYSTYKA STRUKTURALNYCH MOTYWÓW WYBRANYCH, POTENCJALNIE CELIAKO-TOKSYCZNYCH BIAŁEK ZBÓŻ I ROŚLIN STRĄCZKOWYCH

Marta Dziuba, Jerzy Dziuba, Anna Iwaniak

Katedra Biochemii Żywności, Uniwersytet Warmińsko-Mazurski w Olsztynie, Olsztyn

Produkty pochodzenia roślinnego takie jak białka niektórych zbóż czy nasion, mogą wywoływać celiakię in. nietolerancję glutenu. W pracy przeprowadzono analizę strukturalną peptydów odpowiedzialnych za wywoływanie celiakii oraz fragmentów struktury, w których się one znajdują czyli tzw. rozszerzonych motywów strukturalnych. Motywy te pochodziły z A-gliadyny pszenicy i stanowiły wzorzec do analizy pozostałych białek. Badania przeprowadzono w oparciu o metody *in silico* wykorzystując do tego celu bazę danych sekwencji białek i bioaktywnych peptydów – BIOPEP [http://www.uwm.edu.pl/biochemia], bazę danych sekwencji białek SWISS-PROT [http://www.expasy.org], algorytm BLAST [http://www.ncbi.nlm.nih.gov/blast] oraz program PREDICT 7 przeznaczony do przewidywania struktury drugorzędowej.

Analizie poddano ogółem 403 sekwencje białek roślinnych. U 155 z nich stwierdzono obecność tetrapeptydów odpowiedzialnych za celiakię a mianowicie: QQQP, QQPY, PSQQ, QPYP, w tym u 29 białek wymienione tetrapeptydy wchodziły w skład motywu strukturalnego o dłuższej sekwencji. Były to białka pszenicy, jęczmienia i owsa. Wymienione białka należą do rodzin o zbliżonym pokrewieństwie ewolucyjnym i wykazują różny stopień homologii, co zostało potwierdzone za pomocą programu BLAST. Wszystkie motywy rozszerzone znajdowały się w hydrofilowym otoczeniu, jak również przyjmowały strukturę beta-skrętu i nieuporządkowaną, co jest charakterystyczne dla fragmentów zawierających w swoim składzie reszty tyrozyny, proliny czy glutaminy. Przeprowadzono również komputerowe projektowanie procesu proteolizy wybranych białek pszenicy, owsa i jęczmienia (najbogatsza rodzina białek, prekursorów toksycznych peptydów) w aspekcie uwalniania peptydów toksycznych dla osób chorych na celiakię. Aktywne fragmenty uwalniane były z dwudziestu ośmiu białek przez termolizynę, proteinazę K i oligopeptydazę prolinową.