

Analysis of ingredient lists of commercially available gluten-free and gluten-containing food products using the text mining technique

AMANDA BAGOLIN DO NASCIMENTO¹, GIOVANNA MEDEIROS RATAICHESCK FIATES²,
ADILSON DOS ANJOS³, & EVANILDA TEIXEIRA¹

¹Programa de Pós-Graduação em Ciência dos Alimentos, Centro de Ciências Agrárias, Universidade Federal de Santa Catarina, Florianópolis, Brazil, ²Programa de Pós-Graduação em Nutrição, Centro de Ciências da Saúde, Universidade Federal de Santa Catarina, Florianópolis, Brazil, and ³Programa de Pós-Graduação em Engenharia de Produção, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, Brazil

Abstract

Ingredients mentioned on the labels of commercially available packaged gluten-free and similar gluten-containing food products were analyzed and compared, using the text mining technique. A total of 324 products' labels were analyzed for content (162 from gluten-free products), and ingredient diversity in gluten-free products was 28% lower. Raw materials used as ingredients of gluten-free products were limited to five varieties: rice, cassava, corn, soy, and potato. Sugar was the most frequently mentioned ingredient on both types of products' labels. Salt and sodium also were among these ingredients. Presence of hydrocolloids, enzymes or raw materials of high nutritional content such as pseudocereals, suggested by academic studies as alternatives to improve nutritional and sensorial quality of gluten-free food products, was not identified in the present study. Nutritional quality of gluten-free diets and health of celiac patients may be compromised.

Keywords: *nutritional content, raw material, celiac disease, pseudocereals, labels*

Introduction

Celiac Disease is an immune-mediated systemic disorder elicited by gluten and related prolamines in genetically susceptible individuals and characterized by the presence of a variable combination of gluten-dependent clinical manifestations, CD-specific antibodies, HLA-DQ2 or HLA-DQ8 haplotypes, and enteropathy. CD-specific antibodies comprise autoantibodies against TG2, including endomysial antibodies, and antibodies against deamidated forms of gliadin peptides (Husby et al. 2012).

Several longitudinal studies have brought to light an increase in the occurrence of such disease, estimating that approximately 1% of world population is afflicted by it, including also the population of developing countries (Lohi et al. 2007; Fasano et al. 2008; Ludvigsson and Green 2011).

A gluten-free diet is the only possible treatment for celiac disease. Gluten restriction will improve the symptoms and intestinal tissue abnormalities, enhance quality of life of celiac people and likely reduce mortality (Ludvigsson and Green 2011). Due to the need of gluten restrictions in celiac diet, since 2003 the Brazilian government has enacted a bill mandating that all industrialized foods add to their labels the information 'contains gluten' or 'gluten-free', according to each case (Brasil 2003).

As the disease occurrences rise, research interests in developing new technologies and new gluten-free food products have also become increasingly frequent, with the common goal of answering the demands of an emerging market (Mariotti et al. 2009; Han et al. 2010; Mezaize et al. 2009). Due to the fact that gluten

Correspondence: Amanda Bagolin do Nascimento, Programa de Pós-Graduação em Ciência dos Alimentos, Centro de Ciências Agrárias, Universidade Federal de Santa Catarina, Rodovia Admar Gonzaga 1346, Florianópolis, SC 88034-001, Brazil. Tel: + 55 48 3721 5384. Fax: + 55 48 3236 1436. E-mail: amandabagolin@hotmail.com

is a structural protein with exclusive viscoelastic properties, its substitution in baking goods poses an enormous technological challenge. Gluten is responsible for the elasticity and viscosity in dough, promoting gas retention and providing a structured final product after baking (Gallagher et al. 2004; Sciarini et al. 2010a). Consequently, its removal may result in products with low palatability and acceptance (Lerner 2010). In this context, researchers have been trying to overcome such limitations employing alternative ingredients that could mimic gluten properties, such as hydrocolloid agents (e.g. hydroxypropyl methylcellulose, xanthan gum, guar gum and locust bean gum) and enzymes such as transglutaminases and cyclodextrin glycosyl transferase (Yalcin and Basman 2008; Ronda et al. 2009; Sciarini et al. 2010a, 2010b; Sumnu et al. 2010).

Regarding raw materials employed in the elaboration of gluten-free food products, different authors have suggested the use of pseudocereals such as buckwheat, amaranth and quinoa; they represent a safe option concerning gluten absence while improving product's nutritional content, since they are rich in dietary fibre, protein and unsaturated fatty acids (Alvarez-Jubete et al. 2009; Saturni et al. 2010).

The processing of gluten-free food products demands a number of modifications in their ingredient composition, but studies investigating the nutritional quality of such products are scarce (Thompson 1999, 2000; Caponio et al. 2008; Segura and Rosell 2011). Furthermore, although there appears to exist an academic concern about improving gluten-free food product characteristics (Ronda et al. 2009; Sciarini et al. 2010b; Sumnu et al. 2010), studies on celiac consumers have consistently found a low degree of satisfaction with products currently available in the market (Zarkadas et al. 2006; Araújo and Araújo 2011). This may indicate that scientific innovation has not yet reached the final consumer.

Knowing the exact ingredients of gluten-free food products available to celiac consumers is extremely important when planning a diet for such population. Also, comparison with similar gluten-containing products could evidence if similarities and differences between the diet composition of celiac and non-celiac persons are restricted to the mere presence/absence of gluten. In this sense, the objective of the present study was to conduct a descriptive analysis of the ingredients contained in gluten-free food products available in the capital of a Brazilian Province, Santa Catarina, and compare them with the ingredients of similar gluten-containing products.

Method

Selection and content analysis of listed ingredients on gluten-free and gluten-containing food products' labels.

Florianópolis, the city where the study was conducted, is the capital of Santa Catarina Province. The city hosts a regional branch of the Brazilian Celiac Association (Associação dos Celíacos do Brasil – ACELBRA-SC). The association regularly publishes, online, lists of trustworthy retailer establishments of gluten-free food products. All nine retailers listed by the Association in Florianópolis were sent presentation letters (three supermarkets and six health-food stores). Only the establishments whose managers signed consent forms were surveyed.

In Brazil, since January 1998, packaged food products' ingredient labelling is mandatory. Ingredients must be listed on package labels in decreasing quantity order (Brasil 1998). This information was retrieved from all identified gluten-free products in the participating establishments. Products considered in the present study were not the ones naturally gluten-free (such as fruits, eggs, potatoes), but the ones where wheat was replaced by substitute raw materials (such as corn, rice, cassava, amaranth, quinoa).

In order to compare the ingredients used in gluten-free food products with the ones used in gluten-containing products, a systematic selection of information on ingredient lists of packaged gluten-containing products was performed.

Statistical analysis

Analysis of the ingredient lists from labelled packaged products was performed through the text mining (tm) technique with the tm package (Feinerer 2011) of the R open source software (R Development Core Team 2011).

Tm can be broadly defined as knowledge-intensive process in which a user interacts with a document collection over time by using a suite of analysis tools. Tm seeks to extract useful information from data sources through the identification and exploration of interesting patterns. The data sources are document collections, and interesting patterns are found not among formalized database records but in the unstructured textual data in the documents in these collections (Feldman and Sanger 2007).

Data collected from products' labels (ingredient lists) were processed prior to analysis with the removal of numbers, punctuation, removal of cap-sized letters and blank spaces. Compound nouns were rewritten as a single word, e.g: 'soy flour' became 'soyflour'. For both product types (gluten-containing and gluten-free), association degrees between terms and most frequent ingredients were obtained.

Results

Data collection took place between February and May 2011. Eight out of nine invited retail stores agreed to participate. Among them, 162 gluten-free food products were identified, such as bread, cake mix,

snacks, biscuits, energy bars, granola, chocolate, pasta and cake. For each gluten-free product, a gluten-containing counterpart had its ingredient list analyzed, adding up to a total of 324 products' labels from both gluten-containing and gluten-free products.

Regarding the variety of terms corresponding to ingredients used in gluten-free and gluten-containing product types, the number found in gluten-free products was 28% lower. Text analysis identified the 20 most frequent terms on gluten-containing and gluten-free products' labels (Table I). Twelve of those terms corresponded to ingredients found in both product types. Sugar was the most frequently found ingredient in both product types. Other frequently found terms/ingredients in both types of product were *bicarbonate*, *milk*, *salt*, *soy lecithin*, *whole-grain*, *flavour*, *yeast*, *cocoa*, *emulsifier* and *chemical*. Most frequently found terms/ingredients only on gluten-free products were *rice flour*, *egg*, *cassava fecula*, *natural corn starch*, *soy flour*, *rice* and *vanilla*. Most frequent terms/ingredients which have appeared exclusively on gluten-containing products were *wheat flour*, *iron*, *folic acid*, *hydrogenated vegetable fat*, *ammonium*, *stabilizers* and *wehey*.

The way in which the 10 most frequent terms present in gluten-free and gluten-containing products are associated between themselves, considering a correlation higher than 0.2, is presented in Figures 1

Table I. Twenty most frequently occurring terms in gluten-free and gluten-containing food products.

Term/ingredient	Gluten-free products frequency	Gluten-containing products frequency
Folic acid	0	108
Sugar	113	184
Corn starch	39	0
Ammonium	0	51
Flavour	39	72
Rice	27	0
Vanilla	24	0
Bicarbonate	52	125
Cocoa	51	58
Emulsifier	34	57
Stabilizer	0	49
Rice flour	62	0
Soy flour	34	0
Wheat flour	0	139
Cassava fecula	49	0
Yeast	48	64
Iron	0	109
Fortified	0	106
Hydrogenated vegetable fat	0	85
Whole-grain	53	76
Lecithin	47	95
Milk	35	126
Natural	45	0
Egg	57	0
Chemical	32	45
Salt	45	101
Sodium	51	122
Whey	0	45

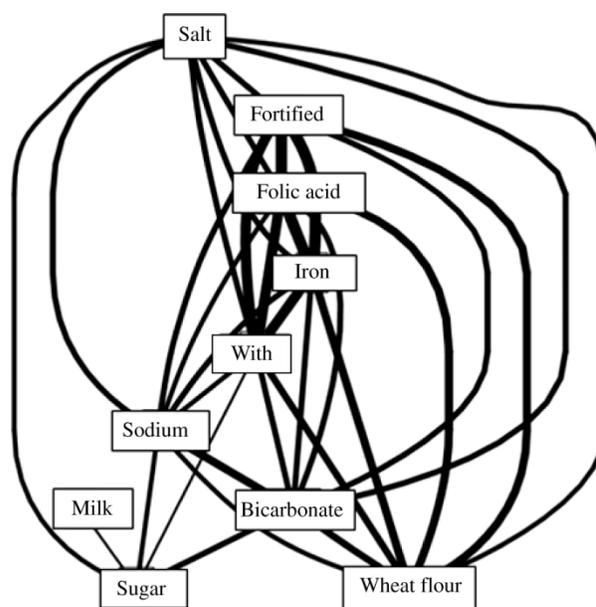


Figure 1. Association among the 10 most frequent terms with correlations higher than 0.2 in gluten-containing products.

and 2. Lines between words (terms) represent the existence of association between them – thicker lines correspond to higher association scores.

When only the ingredients present on at least 25% of analyzed products' labels were considered, it was observed that among gluten-free food products terms/ingredients found were *sugar*, *bicarbonate*, *cocoa*, *rice flour*, *cassava fecula*, *yeast*, *whole-grain*, *lecithin*, *natural*, *egg*, *salt* and *sodium*. Among gluten-containing counterparts, the same terms/ingredients were found, plus one: *whey*.

When the analysis of ingredients present in at least 50% of the gluten-free products was performed, only one ingredient was identified, *sugar*. But among gluten-containing products, the following terms/ingredients were found: *sugar*, *wheat flour*, *fortified*, *folic acid*, *iron*, *bicarbonate*, *hydrogenated vegetable oil*, *lecithin*, *milk*, *salt* and *sodium*.

Associations between terms were also identified and are presented in Table II.

Regarding the raw materials used in gluten-free products, only five varieties were identified: rice, cassava, corn, soy and potato. Rice flour was the most frequently occurring ingredient, followed by cassava fecula, corn starch, soy flour, rice, potato fecula and corn flour.

Discussion

This study proposed an innovative approach of comparing the composition of commercially available packaged gluten-free food products with gluten-containing counterparts by using the tm technique to analyze the terms/ingredients mentioned on labels.

Historically, nutritional counselling for celiac patients has been focused on what should be avoided

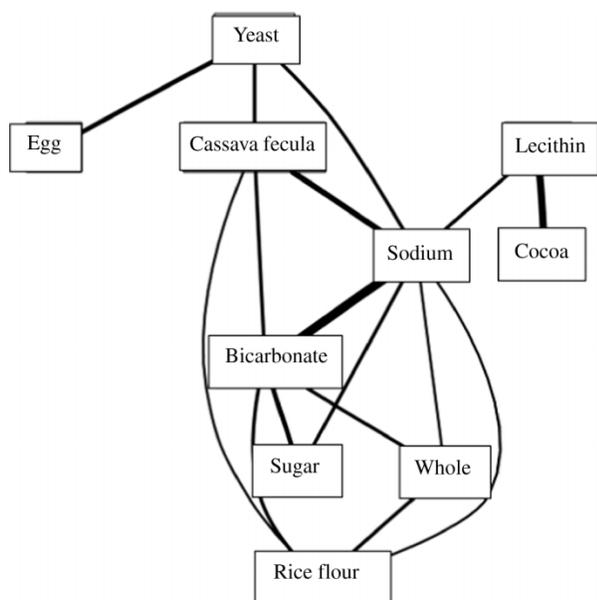


Figure 2. Association among the 10 most frequent terms with correlations higher than 0.2 in gluten-free products.

in their diets, and little importance has been devoted to what was actually being consumed by those patients as long as it was gluten-free. Consequently, studies that discuss the nutritional content of gluten-free foods are scarce (Thompson 1999, 2000; Caponio et al. 2008; Segura and Rosell 2011).

Tm analysis identified that the number of terms/ingredients in gluten-free food products was 28% lower than in gluten-containing products. This may indicate that a lower diversity of ingredients is being employed in the formulation of gluten-free food products, when compared with their gluten-containing counterparts.

Tm has evidenced *sugar* as the most frequently occurring ingredient in both product types (296 times on the 324 analyzed labels). Sodium and salt also were among the most frequently occurring ingredients. Several authors have pointed out the presence of high

amounts of sugar, salt and sodium in industrialized products (Sarno et al. 2009; Powell et al. 2011), associating the habitual intake of such foods to the development of chronic diseases (Malik et al. 2006; Lawrance et al. 2011). For this reason the Global strategy on diet, physical activity and health (2004) recommends the implementation of governmental actions to encourage the reduction of such ingredients in foods and drinks, as well as the cooperation of the food industry to achieve such a goal (World Health Organization 2004).

Although sodium and salt were among the 20 most frequent ingredients on gluten-free food products' labels, and also in at least 25% of gluten-free food products, their frequency was 60% lower than in gluten-containing counterparts. Therefore, gluten-free products appear to contain such ingredients in smaller amounts in their composition when compared with gluten-containing products.

Another ingredient that stood out among the 20 most frequent terms in both product types was milk. Considering that a great deal of the population inflicted with celiac disease is also lactose intolerant (Ojetti et al. 2005), it would be desirable that such an ingredient was not part of the formulation of these products. Even if the presence of milk was not observed among the most frequent ingredients of the first quartile (25%) of gluten-free food products, its presence on such products may limit even more food choices of those celiac patients who are also lactose intolerant.

The remaining most frequently observed terms/ingredients present in both product types are considered by Brazilian regulations as food additives (with the exception of cocoa). In other words, they are ingredients added to food with no nutritional purpose in order to modify the physical, chemical, biological or sensorial characteristics of food (Brasil 1998).

Analyzing the five raw materials used in substitution to wheat flour in the gluten-free food products investigated, results are in accordance with studies

Table II. Identified associations among terms, for gluten-containing and gluten-free food products.

Product	Term	Terms with strong or moderate association				
Gluten-free food products	Whole-grain	Soy flour			Milk	
		0.43			0.33	
	Natural	Food coloring	Flavour	Sweetener	Refined salt	Corn
		0.52	0.50	0.48	0.39	0.37
		Soy extract	Butter	Cocoa mass	Polydextrose	Palm oil
		0.34	0.34	0.34	0.34	0.33
	Bicarbonate	Sodium			Ammonium	
		0.79			0.63	
Gluten-containing food products	Whole-grain	Cocoa mass		Dehydrated	Milk	
		0.35		0.35	0.34	
	Bicarbonate	Ammonium	Sodium	Yeast	Chemical	Wheat flour
		0.86	0.71	0.64	0.64	0.56
		Stabilizer	Iron	Corn starch	Fortified	Folic acid
		0.54	0.54	0.52	0.52	0.51

Note: 1.0–0.69 strong association; 0.69–0.3 moderate association.

which mention that gluten-free food products are elaborated with non-fortified refined raw materials (Gallagher et al. 2004; Niewinski 2008). This practice results in products with low contents of B-complex vitamins, iron, folate and dietary fibre when compared with their gluten-containing counterparts (Thompson 2000; Yazynina et al. 2008).

In Brazil, fortification of wheat and corn flours with iron and folic acid is mandatory, as a strategy of reducing high population rates of anaemia and diseases related to folic acid deficiency (Brasil 2002). However, the mandatory fortification is not extended to flours usually utilized in the formulation of gluten-free products, leaving this population unprotected.

In spite of the advantages mentioned in the literature regarding the use of pseudocereals such as amaranth, quinoa, or buckwheat (Alvarez-Jubete et al. 2009; Saturni et al. 2010), gluten-free food products containing such ingredients were not found in the present research.

Terms frequently found both among gluten-containing and gluten-free food products suggest that both product types follow similar manufacturing patterns, and the concern with quality and safety of gluten-free food products is limited to the fact that raw materials must be gluten-free.

Conclusions

High frequencies of sugar, salt and sodium were identified in both types of products. Identification of the most frequent terms/ingredients present in gluten-containing and gluten-free products suggests that the concern with quality and safety of gluten-free food products is limited to use of gluten-free raw materials.

Improvements in the nutritional and sensorial quality of gluten-free food products proposed by academic researchers with the use of raw materials of high nutritional content such as, for example, pseudocereals, as well as the utilization of hydrocolloids and enzymes, were not identified. Evaluated food products were elaborated with traditional, refined, non-fortified five raw materials, a situation that may compromise dietary intake of fibre and other important nutrients for the maintenance of good health among celiac patients.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper. AB Nascimento receives a scholarship from FAPESC – Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina.

References

Alvarez-Jubete L, Arendt EK, Gallagher E. 2009. Nutritive value and chemical composition of pseudocereals as gluten-free ingredients. *Int J Food Sci Nutr* 60:240–257.

- Araújo HMC, Araújo WMC. 2011. Coeliac disease. Following the diet and eating habits of participating individuals in the Federal District, Brazil. *Appetite* 57:105–109.
- Brasil. Ministério da Saúde. Agência Nacional de Vigilância Sanitária. 1998. Aprova o regulamento técnico para rotulagem de alimentos embalados. *Diário Oficial da União, Brasília*. Available at: http://www.anvisa.gov.br/legis/portarias/42_98.htm. Accessed on 24 March 2012.
- Brasil. Ministério da Saúde. Agência Nacional de Vigilância Sanitária. 2002. Aprova regulamento técnico para a fortificação das farinhas de trigo e das farinhas de milho com ferro e ácido fólico. *Diário Oficial da União, Brasília*. Available at: http://www.anvisa.gov.br/legis/resol/2002/344_02rdc.htm. Accessed on 03 March 2012.
- Brasil. Lei n. 10.674, de 16 de maio de 2003. 2003. Obriga a que os produtos alimentícios comercializados informem sobre a presença de glúten, como medida preventiva e de controle da doença celíaca. *Diário Oficial da União, Brasília*. Available at: <http://e-legis.anvisa.gov.br/leisref/public/showAct.php?id=5854>. Accessed on 24 March 2012.
- Caponio F, Summo C, Clodoveo ML, Pasqualone A. 2008. Evaluation of the nutritional quality of the lipid fraction of gluten-free biscuits. *Eur Food Res Technol* 227:135–139.
- Fasano A, Araya M, Bhatnagar S, Cameron D, Catassi C, Dirks M, et al. 2008. Federation of international societies of pediatric gastroenterology, hepatology, and nutrition consensus report on celiac disease. *J Pediatr Gastroenterol Nutr* 47:214–219.
- Feinerer I. 2011. Tm: text mining package, R package version 0.5-6.
- Feldman R, Sanger J. 2007. *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge: Cambridge University Press.
- Gallagher E, Gormley TR, Arendt EK. 2004. Recent advances in the formulation of gluten-free cereal-based products. *Trends Food Sci Technol* 15:143–152.
- Han JJ, Janz JAM, Gerlat M. 2010. Development of gluten-free cracker snacks using pulse flours and fractions. *Food Res Int* 43: 627–633.
- Husby S, Koletzko S, Korponay-Szabo IR, Mearin ML, Phillips A, Shamir R, et al. 2012. European society for pediatric gastroenterology, hepatology, and nutrition guidelines for the diagnosis of coeliac disease. *J Pediatr Gastroenterol Nutr* 54: 136–160.
- Lawrance JA, Fröhlich ED, Hall JE, Pearson TA, Sacco RL, Seals DR, et al. 2011. The importance of population wide sodium reduction as a mean to prevent cardiovascular disease and stroke: a call to action from the American Heart Association. *Circulation* 123:1138–1143.
- Lerner A. 2010. New therapeutic strategies for celiac disease. *Autoimmun Rev* 9:144–147.
- Lohi S, Mustalahti K, Kaukinen K, Laurila K, Collin P, Rissanen H, et al. 2007. Increasing prevalence of coeliac disease over time. *Aliment Pharmacol Ther* 25:1217–1225.
- Ludvigsson JF, Green PH. 2011. Clinical management of coeliac disease. *J Intern Med* 269:560–571.
- Malik VS, Shulze MB, Hu FB. 2006. Intake of sugar-sweetened beverages and weight gain: a systematic review. *Am J Clin Nutr* 84:274–288.
- Mariotti M, Lucisano M, Pagani MA, Ng PKW. 2009. The role of corn starch, amaranth flour, pea isolate, and Psyllium flour on the rheological properties and the ultrastructure of gluten-free doughs. *Food Res Int* 42:963–975.
- Mezaize S, Chevallier S, Le-Bail A, Lamballerie M. 2009. Optimization of gluten-free formulations for French-style breads. *J Food Sci* 74:E140–E147.
- Niewinski MM. 2008. Advances in celiac disease and gluten-free diet. *J Am Diet Assoc* 108:661–672.
- Ojetti V, Nucera G, Migneco A, Gabriellia M, Lauritano C, Danese S, et al. 2005. High prevalence of celiac disease in patients with lactose intolerance. *Digestion* 71:106–110.

- Powell LM, Schermbeck RM, Szczypka G, Chaloupka FJ, Braunschweig CL. 2011. Trends in the nutritional content of television food advertisements seen by children in the United States. *Arch Pediatr Adolesc Med* 165:1078–1086.
- R Development Core Team. 2011. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available at: <http://www.R-project.org/>. Accessed on 24 March 2012.
- Ronda F, Gómez M, Caballero PA, Oliete B, Blanco CA. 2009. Improvement of quality of gluten-free layer cakes. *Food Sci Technol Int* 15:193–202.
- Sarno F, Claro RM, Levy RB, Bandoni DH, Ferreira SRG, Monteiro CA. 2009. Estimativa de consumo de sódio pela população brasileira, 2002–2003. *Rev Saúde Públ* 43:219–225.
- Saturni L, Ferretti G, Bacchetti T. 2010. The gluten-free diet: safety and nutritional quality. *Nutrients* 2:16–34.
- Sciarini LS, Ribotta PD, León AE, Pérez GT. 2010a. Influence of gluten-free flours and their mixtures on batter properties and bread quality. *Food Bioprocess Technol* 3:577–585.
- Sciarini LS, Ribotta PD, León AE, Pérez GT. 2010b. Effect of hydrocolloids on gluten-free batter properties and bread quality. *Int J Food Sci Technol* 45:2306–2312.
- Segura MEM, Rosell CM. 2011. Chemical composition and starch digestibility of different gluten-free breads. *Plant Food Hum Nutr* 66:224–230.
- Sumnu G, Koksel F, Sahin S, Basman A, Meda V. 2010. The effects of xanthan and guar gums on staling of gluten-free rice cakes baked in different ovens. *Int J Food Sci Technol* 45:87–93.
- Thompson T. 1999. Thiamin, riboflavin, and niacin contents of the gluten-free diet: is there cause for concern? *J Am Diet Assoc* 99: 858–862.
- Thompson T. 2000. Folate, iron, and dietary fiber contents of the gluten-free diet. *J Am Diet Assoc* 100:1389–1396.
- World Health Organization. 2004. Global strategy on diet, physical activity and health. Fifty-seventh world health assembly. Available at: http://www.who.int/dietphysicalactivity/strategy/eb11344/strategy_english_web.pdf. Accessed on 04 March 2012.
- Yalcin S, Basman A. 2008. Effects of gelatinisation level, gum and transglutaminase on the quality characteristics of rice noodle. *Int J Food Sci Technol* 43:1637–1644.
- Yazynina E, Johansson M, Jägerstad M, Jastrebovaet J. 2008. Low folate content in gluten-free cereal products and their main ingredients. *Food Chem* 111:236–242.
- Zarkadas M, Cranney A, Case S, Molloy M, Switzer C, Graham ID, et al. 2006. The impact of a gluten-free diet on adults with coeliac disease: results of a national survey. *J Hum Nutr Diet* 19:41–49.