Computational Biology and Chemistry 60 (2016) 32–42

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Research article

In silico allergenicity prediction of several lipid transfer proteins

Cristiano Garino*, Jean Daniel Coïsson, Marco Arlorio

Dipartimento di Scienze del Farmaco & Drug and Food Biotechnology (DFB) Center, Università del Piemonte Orientale "A. Avogadro", largo Donegani 2, 28100 Novara, Italy

ARTICLE INFO

Article history:
Received 7 May 2015
Received in revised form 4 November 2015
Accepted 10 November 2015
Available online 28 November 2015

Keywords:
LTP
Allergenicity prediction
In silico analysis
Cross-reactivity

ABSTRACT

Non-specific lipid transfer proteins (nsLTPs) are common allergens and they are particularly widespread within the plant kingdom. They have a highly conserved three-dimensional structure that generate a strong cross-reactivity among the members of this family. In the last years several web tools for the prediction of allergenicity of new molecules based on their homology with known allergens have been released, and guidelines to assess potential allergenicity of proteins through bioinformatics have been established. Even if such tools are only partially reliable yet, they can provide important indications when other kinds of molecular characterization are lacking. The potential allergenicity of 28 amino acid sequences of LTPs homologs, either retrieved from the UniProt database or *in silico* deduced from the corresponding EST coding sequence, was predicted using 7 publicly available web tools. Moreover, their similarity degree to their closest known LTP allergens was calculated, in order to evaluate their potential cross-reactivity. Finally, all sequences were studied for their identity degree with the peach allergen Pru p 3, considering the regions involved in the formation of its known conformational IgE-binding epitope. Most of the analyzed sequences displayed a high probability to be allergenic according to all the software employed. The analyzed LTPs from bell pepper, cassava, mango, mungbean and soybean showed high homology (>70%) with some known allergenic LTPs, suggesting a potential risk of cross-reactivity for sensitized individuals. Other LTPs, like for example those from canola, cassava, mango, mungbean, papaya or persimmon, displayed a high degree of identity with Pru p 3 within the consensus sequence responsible for the formation, at three-dimensional level, of its major conformational epitope. Since recent studies highlighted how in patients mono-sensitized to peach LTP the levels of IgE seem directly proportional to the chance of developing cross-reactivity to LTPs from non-Rosaceae foods, and these chances increase the more similar the protein is to Pru p 3, these proteins should be taken into special account for future studies aimed at evaluating the risk of cross-allergenicity in highly sensitized individuals.

1. Introduction

Non-specific lipid transfer proteins are low-molecular-weight proteins ubiquitously present throughout the plant kingdom, and they are currently believed to play a primary role in fighting plant stress, by inhibiting fungal and bacterial pathogens (Egger et al., 2010). This defensive action seems to be confirmed by the fact that their genetic expression is increased when plants are exposed to pathogen attack (Salcedo et al., 2007). Their evolutionary conservation seems to suggest that this class of proteins, belonging to the prolamin superfamily, might play a vital role in the survival of plants. Unfortunately, nsLTPs are also potent allergens. Although the sensitization via the respiratory route is believed to account for the majority of LTP allergy (pollen food syndrome), and even skin contact has been proposed as a route of exposure, the ingestion of LTP-containing fruits or vegetables seems to be responsible for the primary sensitization. Primary food allergy can be accompanied by the possibility of severe allergic symptoms, and anaphylaxis has been associated with LTPs in various plant-derived foods (Van Winkle and Chang, 2014).

Due to the severity of their induced symptoms and their large presence within all known cultivated plants, LTPs have been, and currently are, extensively studied. By 2011 the Allergome database reported 63 allergenic LTPs, and the number raised to 80 by the end of 2014 (www.allergome.org). The level of investigation can vary a lot within this group of allergens, going from the basic identification at genetic level to more advanced studies focused on their physical and biochemical properties. The most studied one is undoubtedly the peach LTP, Pru p 3, believed to be the primary

* Corresponding author.
E-mail addresses: garino@uniupo.it (C. Garino), coisson@uniupo.it (J.D. Coïsson), arlorio@uniupo.it (M. Arlorio).

http://dx.doi.org/10.1016/j.compbiolchem.2015.11.006
1476-9271/

source of sensitization for LTP allergic individuals in the Mediterranean area (Fernandez-Rivas et al., 2006). The Allergome database reports around 280 original papers regarding various aspects of Pru p 3 allergen written before the end of 2013.

The ability of a given protein to induce humoral and cellular Th2 immune responses, resulting in the formation of allergen-specific IgE and Th2 cytokines and subsequently in the induction of clinical symptoms, is a measure of its allergenicity (Scheurer et al., 2015). Such property is usually assessed by *in vitro* and *in vivo* tests, and no protein can be classified as allergenic without at least the evidence of the ability to bind *in vitro* IgEs from sera of sensitized individuals. However, the probability that a newly discovered protein has an allergenic potential can be calculated *in silico* using specific algorithms. The need to assess the allergenic potential primarily arises from the current large employ of biotechnology-derived products, such as genetically modified crops, vaccines and therapeutics, which need to undergo an extensive safety evaluation prior to commercialization (Ladics et al., 2011), as well as it is important to identify new allergens from sequenced genomes. Although no single factor is currently recognized as a reliable marker for protein allergenicity, *in silico* screening of genetic modifications has become mandatory: this kind of test may help to identify novel proteins possibly cross-reacting with known allergens, and therefore warranting subsequent experimental tests for confirmation (Mirsky et al., 2013).

The FAO/WHO and the European Food Safety Authority (EFSA) have established guidelines to assess potential allergenicity of proteins using bioinformatics (FAO/WHO, 2001; EFSA, 2004). However, these guidelines focus on sensitivity to prevent potential new allergens entering the food market. Therefore, these criteria lack accurate prediction and generate a high number of false positives (Ladics et al., 2011). Over the last decade, a variety of internet-based bioinformatics testing tools for protein potential allergenicity have been developed, traditionally based on a search for short linear epitopes through a comparison of primary amino acid sequences (Cressman and Ladics, 2009). There are fundamentally two types of approaches for allergen prediction: the first one follows FAO/WHO guidelines seeking for sequence similarity, while the second approach is based on identification of patterns for allergenicity, called motifs. Most of the methods that follow this approach are based on supervised machine learning.

The use of prediction tools in the determination of the allergenic power of newly identified 'natural' proteins is very recent, and few papers have been issued on the matter. Recently, two papers focused on the *in silico* determination of the potential allergenicity features of 7 known proteins from chickpea (Kulkarni et al., 2013) and a thaumatin-like protein from sapodilla (*Manilkara zapota*), a tropical fruit recently associated to isolated cases of oral allergy syndrome (Ashok Kumar and Venkatesh, 2014), were published. A third one, issued this year, employed bioinformatics to predict the allergenicity of a newly characterized LTP isoform from almond (Buhler et al., 2015).

In this work we performed an *in silico* analysis on 28 amino acid sequences homolog to LTPs. Even though so far none of the foods considered in this analysis, with the exception of asparagus, has ever been reported as a threat to LTP allergic patients, the expression of this class of proteins has been proven at least at genetic level. The objective of our study was to employ bioinformatics-based tools to predict the allergenic potential of several lipid transfer proteins currently not classified as allergens. Moreover, their similarity with known allergens, with particular attention toward Pru p 3, was evaluated, in order to provide new insights into possible cross-reactions.

2. Materials and methods

2.1. Data retrieval

All databases and software used in this study are publicly available on the world-wide web. All information about currently known plant LTPs were retrieved from the Allergome database. The UniProt database (www.uniprot.org) was employed to collect primary sequences of both known allergens and known lipid transfer proteins not yet characterized as allergens.

2.2. Identification of EST sequences potentially coding for allergenic LTPs

The BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to individuate the mRNA sequences coding for the putative lipid transfer proteins: firstly, the tBLASTn suite, which searches translated nucleotide databases using a protein query, was employed to screen the various cDNA libraries present within the NCBI database. Secondly, EST sequences showing the highest *E*-value were selected and analyzed by the BLASTx suite, which searches protein databases using a translated nucleotide query. The CDD tool (Marchler-Bauer et al., 2013) was employed to recognize the nsLTP1 conserved domain on the translated EST sequence.

2.3. Prediction analysis for potential allergenicity of proteins

For prediction of the potential allergenicity of all the amino acid sequences (retrieved from the databases or deduced) according to the current recommendations of the FAO/WHO Expert Consultation, as outlined in the Codex Alimentarius Commission (2003), the web-based tool Allermatch™ (www.allermatch.org) was employed. Three different algorithms were used: algorithm 1 carries out an 80 amino acids sliding window alignment using a 35% of identity cutoff. Algorithm 2 looks for a small exact wordmatch: for our analysis we tried with a 6, 8 and 13 wordlength. Finally, algorithm 3 performs a full FASTA alignment, providing the *E*-values of each query.

The sequence-based prediction analysis was compared to the motif-based one by employing 6 different tools, AlgPred™ (www.imtech.res.in/raghava/algpred/submission.html), Allerdictor (http://allerdictor.vbi.vt.edu), AllerHunter (http://tiger.dbs.nus.edu.sg/AllerHunter/running.html), AllerTOP (www.pharmfac.net/allertop), EVALLER™ (www.slv.se/en-gb/Group1/Food-Safety/e-Testing-of-protein-allergenicity/e-Test-allergenicity) and proAp (gmobl.sjtu.edu.cn/proAP/prediction.html). The proAP analysis was performed by selecting the SVM-AAC method.

Finally, based on the known information on the conformational epitopes of peach LTP, the analyzed amino acid sequences were aligned to the Pru p 3 sequence Q9LED1, using the free software LALIGN (http://www.ch.embnet.org/software/LALIGN_form.html), in order to evaluate sequence identity degree and common residues.

3. Results and discussion

3.1. Non-specific LTPs retrieval from available databases

Information on all the allergens reported as members of the nsLTP family were collected starting from the most updated database, Allergome (www.allergome.org). The list was generated by setting the following options in the advanced search tool: in the first archive we set 'biological functions' and looked for 'LTP', while in the second archive we set 'source' and looked for 'plant'. Eighty molecules resulted from the search: of these, 44 allergens are officially approved by the WHO/IUIS Allergen Nomenclature Subcommittee, while 36 have been added to the database based on