Journal of Molecular Graphics and Modelling 66 (2016) 26–40

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

Ole e 13 is the unique food allergen in olive: Structure-functional, substrates docking, and molecular allergenicity comparative analysis

J.C. Jimenez-Lopez[a,b,*], P. Robles-Bolivar[a], F.J. Lopez-Valverde[a], E. Lima-Cabello[a], S.O. Kotchoni[c,d], J.D. Alché[a]

[a] Plant Reproductive Biology Laboratory; Department of Biochemistry, Cell & Molecular Biology of Plants; Estación Experimental del Zaidín; Spanish National Research Council (CSIC), Granada 18008, Spain
[b] The UWA Institute of Agriculture, The University of Western Australia, Crawley, Perth 6009, Australia
[c] Department of Biology; Rutgers, The State University of New Jersey; Camden, NJ 08102, USA
[d] Center for Computational and Integrative Biology; Rutgers, The State University of New Jersey; Camden, NJ 08102, USA

ARTICLE INFO

Article history:
Received 19 December 2015
Received in revised form 9 February 2016
Accepted 13 March 2016
Available online 17 March 2016

Keywords:
Catalytic cleft
Electrostatic potential
Food allergy
Homology modeling
IgE-binding and T-cell epitopes
Molecular docking
Olea europaea L.
Osmotin
PR5 family

ABSTRACT

Thaumatin-like proteins (TLPs) are enzymes with important functions in pathogens defense and in the response to biotic and abiotic stresses. Last identified olive allergen (Ole e 13) is a TLP, which may also importantly contribute to food allergy and cross-allergenicity to pollen allergen proteins. The goals of this study are the characterization of the structural-functionality of Ole e 13 with a focus in its catalytic mechanism, and its molecular allergenicity by extensive analysis using different molecular computer-aided approaches covering a) functional-regulatory motifs, b) comparative study of linear sequence, 2-D and 3D structural homology modeling, c) molecular docking with two different β-D-glucans, d) conservational and evolutionary analysis, e) catalytic mechanism modeling, and f) IgE-binding, B- and T-cell epitopes identification and comparison to other allergenic TLPs.

Sequence comparison, structure-based features, and phylogenetic analysis identified Ole e 13 as a thaumatin-like protein. 3D structural characterization revealed a conserved overall folding among plants TLPs, with mayor differences in the acidic (catalytic) cleft. Molecular docking analysis using two β-(1,3)-glucans allowed to identify fundamental residues involved in the *endo*-1,3-β-glucanase activity, and defining E84 as one of the conserved residues of the TLPs responsible of the nucleophilic attack to initiate the enzymatic reaction and D107 as proton donor, thus proposing a catalytic mechanism for Ole e 13. Identification of IgE-binding, B- and T-cell epitopes may help designing strategies to improve diagnosis and immunotherapy to food allergy and cross-allergenic pollen TLPs.

1. Introduction

Thaumatin-like proteins (TLPs) are structurally quite similar, with conserved domains in several regions of the protein [1]. They are proteins highly synthetized in response to biotic and environmental stresses [2], and in some cases giving sweet taste to foods, particularly fruits [3].

TLPs belong to the pathogenesis-related (PR) protein family 5 (PR5) (one of 17 distinct PR protein families) based in their sequence similarity [4,5], which represent a divergent group of protein families involved in plant defense functions and different kinds of responses to stresses, *i.e.* the osmotin (cumulates in response to both biotic and abiotic stresses–antifreeze activity–and exhibits antifungal activities), permeatins (present in cereal seeds), including zeamatin from *Zea mays,* hordomatin from *Hordeum vulgare*, and avematin from *Avena sativa* [6,7]. TLPs are classified into three groups: those produced in response to (i) pathogen infection, (ii) osmotic stress (also called osmotins), and (iii) antifungal proteins present in cereal seeds. At present, the basic mechanism of this antifungal activity is still not fully understood, although it may be related to a membrane-permeabilizing activity of the pathogens [8], and involves the typical structural feature such as an acidic cleft in thaumatin-like proteins [9]. The predicted mechanism throughout which the PR-5 proteins exhibit antifungal activity is probably the disruption of the proper assembly of the fungal cell wall during hypha elongation by binding and degrada-

* Corresponding author at: Plant Reproductive Biology Laboratory; Department of Biochemistry, Cell & Molecular Biology of Plants; Estación Experimental del Zaidín; Spanish National Research Council (CSIC), Granada 18008, Spain.
E-mail addresses: josecarlos.jimenez@eez.csic.es, jcjimenezl75@gmail.com (J.C. Jimenez-Lopez).

http://dx.doi.org/10.1016/j.jmgm.2016.03.003
1093-3263/

tion of nascent β-(1,3)-D-glucan molecules [10]. The affinity of PR-5 proteins for carbohydrates varies among proteins [11,12], thus, these differences might explain differential antifungal specificities among fungal species [10].

The availability of information concerning sequenced genomes is making possible the identification of the TLP gene superfamily in *Oryza sativa*, *Arabidopsis thaliana*, *Picea glauca*, *Pinus monticola*, and mosses. However, no TLP gene has been retrieved from the green alga *Chlamydomonas reinhardtii* [2].

TLPs have been identified in several organisms, being universal in plants including both gymnosperms and angiosperms [2,5], and bryophytes like the moss *Physcomitrella patens* subsp. Patens. In some species of plants, TLPs are constitutively expressed in flowers and fruits, where they perform defense functions against infection [11]. TLPs are also induced in response to wounding and insect feeding [12]. Some TLPs with antifungal activity act by permeabilizing fungal membranes [13]. Other TLPs appear to function by inhibiting fungal xylanases [14]; potentially as α-amylase and trypsin inhibitors [15]; or by binding and hydrolyzing β-(1,3)-glucans [16] throughout unclear mechanisms of binding and catalytic activity. The hydrolytic activity of complex sugars (*i.e.* β-glucan oligomers with different degrees of polymerization) seems to elicit and induce various defense responses of the plants, such as the activation of chitinase activity [18], and other defense responses in tobacco [17]; the specific degradation of cell wall β-(1,3)-D-glucans [18]; or a contribution to the softening of the pulp or flesh of fruits as grape and cherry [19] in ripening fruits.

On the other hand, pathogenesis-related proteins are reported to be allergens [20], particularly those of the PR-5 family. A limited number of TLPs have been identified from pollens and plant foods as allergens in susceptible atopic individuals [20], due to lack of studies involving molecular allergy. Among these TLPs from pollen, the following could be highlighted: Jun a 3 [21], Cup a 3 [22], and Cry j 3 [23] allergens. Among foods, the following allergenic proteins are included: Lyc e NP24 [24], Pru av 2 [25], Act c 2 [26], Mal d 2 [27], Cap a 1 [28], Vit v TLP [29], and Mus a 4 [30]. Although there is no currently experimental evidence, their wide distribution (ubiquitous location) in plant species, and their allergenic properties, (*i.e.* cross-reactivity between pollen and fruit as PR-5 allergens, likely due to sharing similar structures), strongly indicate that this family of proteins might be proposed as a panallergen, similarly to profilin, LTP [31].

To date, twelve allergens have been identified and characterized in olive tree pollen. However, Ole e 13 is the only food allergen identified, until now, in the olive fruit. In the current study, we performed an extensive structure-functional, molecular interaction and conservational analysis of Ole e 13 allergen protein in comparison to other allergenic TLPs by using homology modeling and protein-ligand docking methods, defining the catalytic-binding cleft for different substrates. Furthermore, the potential key residues for enzyme activity and substrate specificity were highlighted, and a putative catalytic mechanism was proposed for this protein. We also identified the sequences of the IgE, B- and T-cell epitopes of Ole e 13, and perform a molecular analysis of possible cross-reactivity between food and/or pollen allergens. Structural knowledge of the epitopes responsible for allergy of these proteins is essential to design therapeutic tools (*i.e.* immunotherapy) to tackle allergy.

2. Material and methods

2.1. Thaumatin-like protein sequences retrieval

Olive TLP (Ole e 13 allergen) sequence (NCBI accession number EU927297) was used as query to search TLPs against publicly available sequence databases Swiss-Prot/TrEMBL (Uniprot) (http://www.uniprot.org/), and NCBI (http://www.ncbi.nlm.nih.gov/), by using BLASTX, BLASTN and BLAST (low complexity filter, Blosum62 substitution matrix) (http://blast.ncbi.nlm.nih.gov/Blast.cgi/). Others databases were searched looking for other allergenic TLPs (Allergome, http://www.allergome.org/); SDAP, https://fermi.utmb.edu/), Allergen online (http://www.allergenonline.com/); and FARRP databases (http://farrp.unl.edu/resources/farrp-databases).

TLPs characteristic patterns, as well as functional (biologically meaningful) motifs were analyzed for each sequence by using the PROSITE database (http://prosite.expasy.org/).

2.2. Phylogenetic analysis of thaumatin-like proteins

Amino acid sequences retrieved from 56 TLPs were used to make multiple alignments by using ClustalW tools (http://www.ebi.ac.uk/Tools/clustalw/). These alignments were made using the Gonnet protein weight matrix, multiple alignment gap opening/extension penalties of 10/0.5 and pairwise gap opening/extension penalties of 10/0.1. The outputs were manually checked to optimize the alignment by using Bioedit v7.0.5.3 (www.mbio.ncsu.edu/bioedit/bioedit.html). Phylogenetic trees were generated by the neighbor-joining method (NJ), and the branches were tested with 1000 bootstrap replicates. Trees were visualized by using Treedyn (www.treedyn.org).

2.3. Template searching for thaumatin-like protein models building

Protein Data Bank (PDB) was searched for Ole e 13 protein homology. The best homologous templates for Ole e 13 and TLPs from other plant species were selected by BLAST server (http://ncbi.nlm.nih.gov/). The BioInfoBank Metaserver (http://meta.bioinfo.pl/) and Swiss-model server for template identification (swiss-model.expasy.org) were also used for templates selection. The best templates (1Z3Q, 1AUN, 1PCV, and 1DU5) were retrieved from PDB database and used for homology modeling.

2.4. Homology modeling, structural comparisons and evolutionary conservation analysis of thaumatin-like proteins

TLPs protein retrieved sequences were used to build a protein model by using the closer PDB template structures by SWISS-MODEL (swissmodel.expasy.org). An initial structural model was analyzed for recognition of errors in 3D structure by using ProSA (prosa.services.came.sbg.ac.at/prosa.php), and for a first overall quality estimation of the model with QMEAN (swissmodel.expasy.org/qmean/cgi/index.cgi). Final TLP structures were subjected to energy minimization with GROMOS96 force field energy implemented in DeepView/Swiss-PDBViewer v3.7 (spdbv.vitalit.ch) to improve the van der Waals contacts and correct the stereochemistry of the model. The final quality of the models was assessed by QMEAN, the stereology with PROCHECK (www.ebi.ac.uk/thornton-srv/software/PROCHECK), and ProSA (prosa.services.came.sbg.ac.at/prosa.php) programs, and the protein energy with ANOLEA (protein.bio.puc.cl/cardex/servers/anolea). The Ramachandran plot statistics for the models were calculated to show the number of protein residues in the favored regions.

Superimposition method of the structural Cα carbons of the final built models was used for structural similarity comparisons. The different protein structural features were visualized in PyMol software (http://www.pymol.org/). Conservation scores were obtained by ConSurf server (consurf.tau.ac.il). Therefore, identification of ligand-binding domains in the thaumatin-like protein structures was performed using Cofactor software (zhanglab.