J Comput Aided Mol Des (2013) 27:873–895
DOI 10.1007/s10822-013-9686-y

Structural functionality, catalytic mechanism modeling and molecular allergenicity of phenylcoumaran benzylic ether reductase, an olive pollen (Ole e 12) allergen

Jose C. Jimenez-Lopez · Simeon O. Kotchoni · Maria C. Hernandez-Soriano · Emma W. Gachomo · Juan D. Alché

Received: 16 June 2013 / Accepted: 16 October 2013 / Published online: 24 October 2013

Abstract Isoflavone reductase-like proteins (IRLs) are enzymes with key roles in the metabolism of diverse flavonoids. Last identified olive pollen allergen (Ole e 12) is an IRL relevant for allergy amelioration, since it exhibits high prevalence among atopic patients. The goals of this study are the characterization of (A) the structural-functionality of Ole e 12 with a focus in its catalytic mechanism, and (B) its molecular allergenicity by extensive analysis using different molecular computer-aided approaches covering (1) physicochemical properties and functional-regulatory motifs, (2) sequence analysis, 2-D and 3D structural homology modeling comparative study and molecular docking, (3) conservational and evolutionary analysis, (4) catalytic mechanism modeling, and (5) sequence, structure-docking based B-cell epitopes prediction, while T-cell epitopes were predicted by inhibitory concentration and binding score methods. Structural-based detailed features, phylogenetic and sequences analysis have identified Ole e 12 as phenylcoumaran benzylic ether reductase. A catalytic mechanism has been proposed for Ole e 12 which display Lys133 as one of the conserved residues of the IRLs catalytic tetrad (Asn-Ser-Tyr-Lys). Structure characterization revealed a conserved protein folding among plants IRLs. However, sequence polymorphism significantly affected residues involved in the catalytic pocket structure and environment (cofactor and substrate interaction-recognition). It might also be responsible for IRLs isoforms functionality and regulation, since micro-heterogeneities affected physicochemical and post-translational motifs. This polymorphism might have large implications for molecular differences in B- and T-cells epitopes of Ole e 12, and its identification may help designing strategies to improve the component-resolving diagnosis and immunotherapy of pollen and food allergy through development of molecular tools.

Keywords Allergen · B- and T-cell epitopes · Electrostatic potential · Flavonoids · Homology modeling · Isoflavone reductase · Molecular docking ·

Electronic supplementary material The online version of this article (doi:10.1007/s10822-013-9686-y) contains supplementary material, which is available to authorized users.

J. C. Jimenez-Lopez (✉) · J. D. Alché
Department of Biochemistry, Cell and Molecular Biology of Plants, Estación Experimental del Zaidín, Spanish National Research Council (CSIC), Profesor Albareda 1, 18008 Granada, Spain
e-mail: josecarlos.jimenez@eez.csic.es; jcjimenezl75@gmail.com; jose.jimenez-lopez@uwa.edu.au

J. C. Jimenez-Lopez
The UWA Institute of Agriculture, The University of Western Australia, 35 Stirling Highway, Crawley, Perth, WA 6009, Australia

S. O. Kotchoni · E. W. Gachomo
Department of Biology, Rutgers University, 315 Penn Street, Camden, NJ 08102, USA

S. O. Kotchoni · E. W. Gachomo
Center for Computational and Integrative Biology (CCIB), Rutgers University, 315 Penn Street, Camden, NJ 08102, USA

M. C. Hernandez-Soriano
School of Agriculture and Food Sciences, Faculty of Science, The University of Queensland, St Lucia, QLD 4072, Australia

M. C. Hernandez-Soriano
Department of Earth and Environmental Sciences, Katholieke Universiteit Leuven (KULeuven), Kasteelpark Arenberg 20, Box 2459, 3001 Heverlee, Belgium

NAD(P)H · Olive · Pollen · Polyphenols · Sequence polymorphism

Introduction

Lignans and isoflavonoids synthesized through the central phenylpropanoid pathway are essential secondary metabolism macromolecules of vascular plants. They are synthesized by a large family of IRLs [1]. They are involved in human health protection because of their estrogenic, antiangiogenic, antioxidant, and anticancer activities [2]. These metabolites are widely spread throughout the plant kingdom from mosses to angiosperms, where they are involved in a wide range of functions such as defense against microbial, fungal and herbivores [3, 4], heartwood development, flower and fruit pigmentation (anthocyanins), UV radiation protection, signal and defense compounds in their interactions with beneficial and pathogenic microorganisms [5, 6].

Several genes with significant homology to IRLs family members have been isolated from various plant species [7–10]. Isoflavone reductases, which belong to the IRLs, are specific to legumes. The functions of many IRLs in non-legume plants are not yet understood, i.e. birch protein Bet v 6 [11]. In pollen, IRLs seems to be involved in the germination process, pollen tube growth, and the possible implication in pollen-stigma recognition since IRL NAD(P)H-dependent oxidoreductase gene expression increased after interaction between pistil and pollen tubes in *Solanum tuberosum* [7].

The Expression analysis of IRLs at different developmental stages of plants, and the high number of sequences that exist, i.e. more than 13,000 entries in CATH database (http://www.cathdb.info/), and more than 170 IRL plant protein sequences in Uniprot database (http://www.uniprot.org/), suggests that these enzymes play important roles in a wide range of physiological processes.

At present, studies dealing with the synthesis pathway of various important bioactive lignans [12] and isoflavonoids [4] have resulted in the isolation and characterization of various phylogenetically related NADPH-dependent aromatic alcohol reductases, exhibiting high homology in gene sequence, and are named pinoresinol-lariciresinol reductase (PLR) [13, 14], isoflavone reductase (IFR) [15–17], and phenylcoumaran benzylic ether reductase (PCBER). PLRs and IFRs catalyze the more specific enantiospecific conversions with high specificity for isoflavones than PCBER that catalyze reductive processes in 8–5′ linked lignans, and it only shows regiospecific discrimination, reducing both (α)- and (β)-enantiomeric forms. This is the reason why PCBER is likely to be the progenitor of PLR and IFR [10]. To date, only the three-dimensional x-ray crystal structures of IFR [18] from *Medicago sativa* (PDB accession number 2gas), PLR [19] from *Forsythia intermedia* (1qyd) and PCBER [19] from *Pinus taeda* (1qyc) have been solved and refined, showing high structural similarity.

The sequence of a new IRL protein has been isolated from olive pollen. This sequence has been identified as a new allergen (putatively recognized as IgE-inducers of allergy symptoms) named Ole e 12 (http://www.allergome.org/ and/or https://fermi.utmb.edu/SDAP/). The functional diversity of proteins in olive (*Olea europaea* L.) pollen is one of the main causes of respiratory Type-I allergy, affecting more than 25 % of the human population in developed countries [20], especially in Mediterranean areas, America, and Australia. Up until now, twelve allergens (Ole e 1 to Ole e 12) have been identified in olive pollen [21] with allergic activity, but only the first eleven have been at least partially characterized. Among them, few have been identified as high polymorphic allergens such as the major allergen of olive pollen, Ole e 1 [22], the Pan-allergen Ole e 2 or profilin [23, 24], and Ole e 11 or pectin methylesterase [25, 26].

Additionally, other members of the IRL family of proteins have been identified as allergens, e.g., in pollen of birch (Bet v 6), Japanese cedar (Cry j IFR), Ash (Fra e 12), and vegetative tissues of sweet orange (Cit s IFR), carrot (Dau c 5), sharon fruit (Dio k IFR), latex (Hev b IFR), lichee (Lit c IFR), and pear (Pyr c 5) (http://www.allergome.org/).

The sequence of Ole e 12 (NCBI accession number EU927297) was used to perform an initial functional assessment, searching for characteristic architectural protein domains followed by sequence comparison using BLAST to other well identified and characterized proteins from the same family in multiple databases. From this initial approach, Ole e 12 was identified as putative isoflavone reductase-like protein.

In the present study, we performed an extensive structure-functional and conservational analysis of Ole e 12 protein by homology modeling and docking methods, defining the catalytic-binding pockets for co-factor and substrate. In addition, the potential key residues for enzyme activity and substrate specificity were highlighted, and a specific catalytic mechanism was proposed for this protein. Moreover, we identified the sequences of the B- and T-cell epitopes following different in silico approaches, and performed their molecular analysis and comparison with other allergen proteins previously identified.

Materials and methods

Isoflavone reductase-like sequence database search

Olive IRL (Ole e 12 allergen) sequence (NCBI accession number EU927297) was used as queries to search pollen and vegetative IRLs 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/).

Domain architecture analyses

In order to investigate the possible domains contained in Ole e 12 protein isoforms, characteristic motifs and patterns were additionally queried using Pfam v25.0 (http://pfam.sanger.ac.uk/), Prosite (http://prosite.expasy.org/scanprosite/), SMART v6.0 (http://smart.embl-heidelberg.de/), Conserved Domain Database (CDD) v3.02, CDART (Conserved Domain Architecture Retrieval Tool) and CD-Search tools (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml/), InterPRO v35.0 (http://www.ebi.ac.uk/interpro/), ProDom release 2010.1 (http://prodom.prabi.fr/prodom/current/html/home.php/), CATH v3.4 (http://www.cathdb.info/), Superfamily v1.75 (http://supfam.cs.bris.ac.uk/SUPERFAMILY/), PIRSF (http://pir.georgetown.edu/pirwww/dbinfo/pirsf.shtml/), and functional search using PANTHER (http://www.pantherdb.org/).

Phylogenetic analysis of isoflavone reductase-like proteins

Amino acid sequences retrieved of 136 IRLs were used to make alignments by using ClustalW multiple sequence alignment tools (http://www.ebi.ac.uk/Tools/clustalw/). The alignments data were used for homology modeling of Ole e 12.

These alignments were created 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). Portions of sequences that could not be reliably aligned were eliminated. Phylogenetic trees were generated by the neighbor-joining method (NJ), and the branches were tested with 1,000 bootstrap replicates. Trees were visualized by using Treedyn (www.treedyn.org).

Physicochemical properties and post-translational patterns/motifs

Physicochemical properties of the profilin sequences were analyzed by using the Expert Protein Analysis System (ExPASy) Proteomics Server (expasy.org). The ProtParam tool was implemented to calculate the MW/pI of the different IRLs isoforms, as well as instability index, aliphatic index and grand average of hydropathicity (GRAVY).

IRLs characteristic patterns were checked for each original sequence and further analysis were performed to highlight the presence of functional motifs by using the PROSITE database (http://prosite.expasy.org/). Biologically meaningful motifs, susceptible of posttranslational modifications were derived from multiple alignments and the ScanProsite tool (http://prosite.expasy.org/scanprosite/), from the ExPASy proteomics server of the Swiss Institute of Bioinformatics (expasy.org), as well as phosphorilation motives with more that 80 % of probability of occurrence was analyzed by using NETPhos v1.0 and NETPhosK v1.0 (www.cbs.dtu.dk).

Secondary structure prediction

Ole e 12 secondary structural elements recognition was assessed by Segmer algorithm [27], which threads sequence segments through Protein Data Bank (PDB) library (http://www.pdb.org/) to identify conserved substructures. Furthermore, the secondary structure elements were also identified and compared with the results obtained with other different approaches: SSpro8 (Scratch Protein Predictor), which adopts the full DSSP 8-class output classification (scratch.proteomics.ics.uci.edu), PredictProtein (www.predictprotein.org), NetSurfP ver. 1.1 (www.cbs.dtu.dk), and PSIPRED (bioinf.cs.ucl.ac.uk/psipred) fold servers.

Structure template searching

The Ole e 12 protein sequence was searched for homology in the Protein Data Bank (PDB). The homologous templates suitable for Ole e 12, as well as for others plant species were selected by BLAST server (http://ncbi.nlm.nih.gov/). The BioInfoBank Metaserver (http://meta.bioinfo.pl/) which employs fold recognition for homology search was also used for templates selection. Furthermore, the results obtained by previous methods were also compared with these obtained by Swiss-model server for template identification (swissmodel.expasy.org). The best templates (2gas, 1qyc, 1qyd) were retrieved from PDB database and used for homology modeling.

Homology modeling

IRL protein sequences were retrieved and used to build a protein model by using the top PDB closed template structures by SWISS-MODEL, a protein structure homology-modeling server, via the ExPASy web server (swissmodel.expasy.org).

An initial structural model was generated and checked for recognition of errors in 3D structure by using ProSA (prosa.services.came.sbg.ac.at/prosa.php), and also for a

first overall quality estimation of the model with QMEAN (swissmodel.expasy.org/qmean/cgi/index.cgi). Final structures of Ole e 12 and other IRLs were subjected to energy minimization with GROMOS96 force field energy implemented in DeepView/Swiss-PDBViewer v3.7 (spdbv.vital-it.ch) to improve the van der Waals contacts and correct the stereochemistry of the model. For each sequence analyzed, the quality of the model was assessed by QMEAN, checking proteins stereology with PROCHECK (www.ebi.ac.uk/thornton-srv/software/PROCHECK), and ProSA (prosa.services.came.sbg.ac.at/prosa.php) programs, as well as the protein energy with ANOLEA (protein.bio.puc.cl/cardex/servers/anolea). The Ramachandran plot statistics for the models were also calculated to show the number of protein residues in the favored regions.

Ligand-binding domains and functional annotation based on 3D protein structures

Prediction of the best identified ligand-binding domains in the Ole e 12 structure was made through structure-based approaches to protein functional inference and ligand screening, using a threading algorithm based on binding site conservation across evolutionary distant proteins for ligand binding site prediction, ligand screening and molecular functional prediction.

Cofactor software (zhanglab.ccmb.med.umich.edu/COFACTOR), a structure-based method for biological function annotation of proteins was used after providing a 3D-structural model of the protein of interest to identify functional homology. Functional insights, including ligand-binding site, and enzyme classification, were derived from the best functional homology template. The identification of functional analogs of the query proteins was possible by Gene Ontology (GO) terms (The Gene Ontology project), based on built 3D models, and describing the molecular function and biological processes in which proteins are implicated (www.geneontology.org).

Structural comparisons and evolutionary conservation analysis

Structural comparisons between Ole e 12 and other plant IRLs proteins were performed by superimposition of the structural Cα carbons, aiming to calculate the average distance between their Cα backbones. The 2-D protein structural analysis, protein superimpositions and surface protein contours analysis were performed and visualized in PyMol software (http://www.pymol.org/).

Built protein model was submitted to ConSurf server (consurf.tau.ac.il) in order to generate evolutionary related conservation scores helping to identify functional regions in the proteins. Functional and structural key residues in Ole e 12 sequence were confirmed by ConSeq server (conseq.tau.ac.il).

Poisson–Boltzmann electrostatic potential

The Electrostatic Poisson-Boltzmann (PB) potentials for all the structures were obtained using APBS (DeLano Scientific LLC) molecular modeling software implemented in PyMol 0.99 (www.pymol.org) with AMBER99 to assign the charges and radii to all of the atoms (including hydrogens), which were added and optimized with the Python software package PDB2PQR. Fine grid spaces of 0.35 Å were used to solve the linearized PB equation in sequential focusing multigrid calculations in a mesh of 130 points per dimension at 310.00 K. The dielectric constants were 2.0 for the protein and 80.00 for water. The output mesh was processed in the scalar OpenDX format to render isocontours and maps onto the surfaces with PyMOL 0.99. Potential values are given in units of kT per unit charge (k Boltzmann's constant; T temperature).

Protein docking process and clustering analysis

We performed the docking analysis of Ole e 12 protein. We considered backbone flexibility by using rigid-body ensemble docking with multiple structures derived from NMR. The Fast Fourier Transform (FFT) correlation approach to protein–protein docking evaluated the energies of billions of docked conformations on a grid. To obtain decoys by rigid-body docking, we used the option ZDOCK for sampling at 6-degree rotational steps in CLUSpro server [28]. Using Fast Fourier Transform, ZDOCK searches for all possible binding orientations of a ligand along the surface of a receptor protein, optimizing desolvation, shape complementarily and electrostatics. The top 2,000 structures, along with their ZDOCK scores, were used as candidates of near-native structures. The docking score was calculated by considering several interaction properties, e.g., shape complementarity, desolvation, and electrostatics potential.

For each docked conformation, we computed the residues of the ligand within 10 Å of its receptor. This RMSD computing analysis was done for all the 2,000 ligands in this study. After clustering, the ranked complexes are subjected to a straightforward (300 steps and fixed backbone) van der Waals minimization using CHARMM to remove potential side chain clashes. Best scoring was chosen as higher possible model for proteins interaction.

Allergenicity profile assessment

Allergenicity of Ole e 12 protein was checked by a full FASTA alignment in the Structural Database of Allergenic

Proteins (SDAP) (fermi.utmb.edu/SDAP). Allergenicity profile was assessed by combination of different parameters: hydrophobicity [29–31], antigenicity [32], and Solvent Accessible Surface Areas (SASA) of all residues in the generated models were calculated by using GETAREA (curie.utmb.edu/getarea.html), and compared to absolute surface area (ASA) of each residue calculated by DSSP program (swift.cmbi.ru.nl/gv/dssp). These values were transformed to relative values of ASA and visualized by ASAView (http://www.netasa.org/asaview/).

B-cell epitopes identification

For determination of linear (continuous) epitopes, the protein sequence of Ole e 12 was submitted to ABCpred (www.imtech.res.in/raghava), BepiPred 1.0b (www.cbs.dtu.dk), BCPREDS (ailab.cs.iastate.edu/bcpreds), Bcepred (www.imtech.res.in/raghava), Ellipro (tools.immuneepitope.org), and COBEpro (scratch.proteomics.ics.uci.edu) web servers.

For prediction of discontinuous (conformational) epitopes, the best structures generated by homology modeling were submitted to Discotope 1.2 server (www.cbs.dtu.dk), BEpro (pepito.proteomics.ics.uci.edu/index.html), and PPI-PRED (bmbpcu36.leeds.ac.uk/ppi_pred). The methods were selected on the basis of performance measures that include both threshold dependent (sensitivity, specificity, accuracy) and independent parameters. While selecting consensus epitopes, a preference was given to results predicted from tools tested on larger datasets and having an area under receiver operating characteristic curve (AUC) greater than 0.70. Tools based solely on physic-chemical parameters were given the least preference. Default settings were applied to all the tools used. The regions recognized frequently by four or more tools were selected.

T-cell epitopes identification

Stabilization matrix alignment methods allowing direct prediction of peptide MHC binding sequences and affinities were used for Ole e 12 protein: TEPITOPE (www.bioinformation.net/ted), Propred (www.imtech.res.in/raghava/propred) (using quantitative matrices), NetMHCII (www.cbs.dtu.dk), Multipred (antigen.i2r.a-star.edu.sg/multipred), CTLpred (www.imtech.res.in/raghava/ctlpred) (artificial neural network approach), and RANKPEP (Position-specific scoring matrix, bio.dfci.harvard.edu/Tools/rankpep.html) which employ binding status scoring qualitative prediction methods and were used for predicting T-cell epitopes. HLA class II binding peptides of Ole e 12 were also predicted by inhibitory concentration (IC50) value based in quantitative prediction methods MHCpred (www.ddg-pharmfac.net/mhcpred/MHCPred), SVRMHC (SVRMHC.umn.edu/SVRMHCdb), ARB matrix methods (www.epitope.liai.org:8080/matrix), SMM-Align in MetaMHC (www.biokdd.fudan.edu.cn/Service/MetaMHCII/server.html). Promiscuous peptides binding to multiple HLA class II molecules were selected.

B- and T-cell epitopes identified by computational tools were mapped onto linear sequence and on the three dimensional model of Ole e 12 to determine their position and secondary structure elements involved.

Results

Ole e 12 sequence analysis

Ole e 12 has been identified as an IRL protein [33]. It belongs to the short-chain dehydrogenase/reductase (SDR) superfamily, NAD(P)-binding Rossmann-like Domain Superfamily (CATH Superfamily 3.40.50.720; InterProt number IPR016040), and NmrA-like protein family (InterProt number IPR008030; Pfam number PF05368; Gene Ontology number, GO: 0000166).

The screening of characteristic motifs/patterns (http://prosite.expasy.org) indicated that Ole e 12 exhibit two defined domains integrated by three patters, a NmrA-like family pattern defined by the entire sequence (position 7–298), a second domain NAD(P)-binding Rossmann-like patter (positions N-t to 153, 190–215, and 278–298), which also integrated the active (co-enzyme/ligand-binding) domain of the protein, and a third pattern corresponding to UDP-galatose 4-epimerase motif (position 154–189, 217–279, and 295–308), which integrated the polymerization domain of the protein (Fig. 1).

The exploration of allergy databases (Allergome, http://www.allergome.org/; and SDAP, http://fermi.utmb.edu/SDAP/) for plant IRLs already identified as allergens, in both vegetative and reproductive tissues, include sequences of Ole e 12 (*O. europaea* L.), Bet v 6 (*Betula verrucosa*), Cit s IFR (*Citrus sinensis*), Cry j IFR (*Cryptomeria japonica*), Dau c 5 (*Daucus carota*), Dio k IFR (*Diospyros kaki*), Fra e 12 (*Fraxinus excelsior*), Hev b IFR (*Hevea brasiliensis*), Lit c IFR (*Litchi chinensis*), Pyr c 5 (*Pyrus communis*).

Sequence alignment and identity comparison with a representative number of other plant IRLs has shown a wide range (40–97 %) of identity among them. Particularly, low identity was found when we compared olive to grapevine (43.4 %), lupin (44.4 %), orange (45 %), rice (54.5 %) or pea (54.9 %), but olive showed higher identity with IRLs from *F. excelsior* (96.4 %), or *Betula pendula* (81.1 %) pollen IRLs.

Previous comparison of several matures IRLs [18], have shown few strictly conserved residues: Met1 or M^1 in Ole e

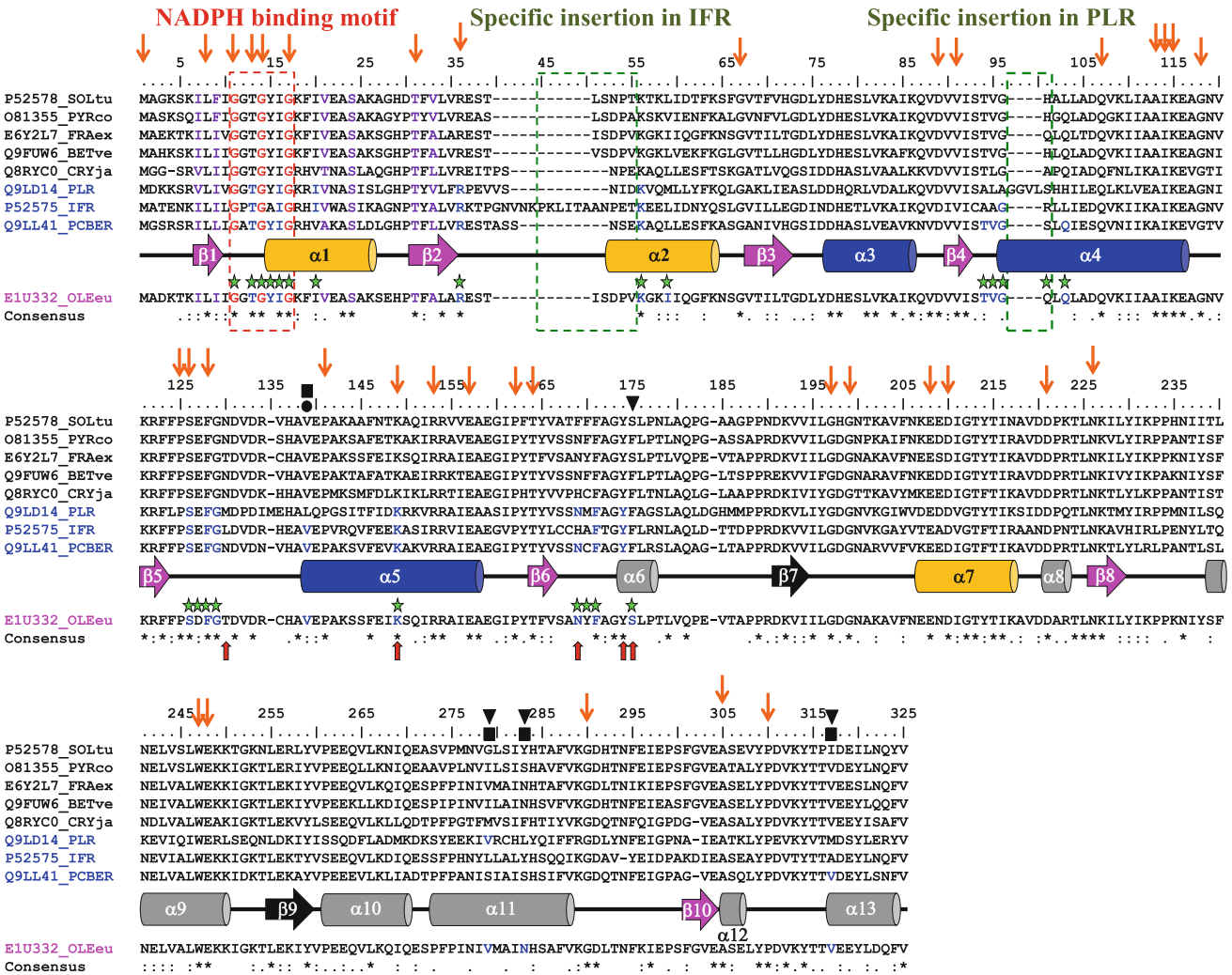

Fig. 1 Sequences and 2-D structure analysis of isoflavone reductase—like proteins from different plant species. Sequence comparison of different plant IRLs (potato, P52578; pear, O81355; ash, E6Y2L7; birch, Q9FUW6; Japanese cedar, Q8RYC0; giant arborvitae, Q9LD14; alfalfa, P52575; loblolly pine, Q9LL41; and olive, E1U332) was performed by multiple alignment. Conserved residues among the nine species were highlighted with *asterisks*, while the highly conserved among plant species are highlighted with *orange arrows*. Residues implicated in cofactor and substrate binding were pointed out with *green stars*, while those implicated in the catalytic activity were pointed out with red *filled arrows*. 2-D elements were depicted as barrels (α-helices) and *arrows* (β-sheets). 2-D elements implicated in the Rossmann-fold were highlighted with *blue* and *yellow barrels* and *pink arrows*. NADPH binding motif is highlighted by discontinue *red line*. Specific amino acids insertions in isoflavone reductases and pinoresinol-lariciresinol reductase are highlighted by discontinue *green line*. *Black dots*, *invert triangles* and *squares* indicate amino acid located only near to the catalytic center of PLRs, amino acids involved in stereo-specificity, and conserved amino acids in PLRs with different stereochemistry, respectively. Conserved amino acid in giant arborvitae, alfalfa, loblolly pine, and olive are highlighted in *blue color*

12 corresponding to the alignment of the Fig. 1, Gly^{11} (G^{11}), Tyr^{13} (T^{13}), Gly^{14} (G^{14}), Gly^{17} (G^{17}), Gly ^{28} (E^{28}), Tyr^{31} (T^{31}), Arg^{36} (R^{36}), and Gly^{67} (G^{67}) also conserved in Ole e 12, as well as several others highly conserved residues. Furthermore, these alignments using amino acid sequences had shown three characteristics molecular patters of IRLs with functional importance [34]. Only one appeared to be conserved in all of the IRLs, corresponding to the NADPH binding (cofactor) domain, which sequence is G^{11}XXG^{14}XXG^{17}, where "X" could be any amino acid. The residues of this motif are known to stabilize both the adenine and nicotinamide moieties of the cofactor in the NADPH-bound form of quinone oxidoreductase [35]. On the other hand, the remaining part of the nucleotide-binding domain has a low level of sequence similarity to other IRLs and participates mainly in subunit interactions.

The other two patterns, one specific for isoflavone reductase proteins is located between the positions 45–55, and the second specific for pinoresinol-lariciresinol reductase proteins pattern is located between the positions 95–100 (Fig. 1). Comparison of different IRLs has revealed that at least 36 positions are highly conserved,

four positions are involved in stereo-specificity, one of them is involved directly with Ole e 12 catalytic mechanism (S^{159}) (Fig. 1), and no position out of the four involved is in PLRs stereochemistry. Olive Ole e 12 had shown only 1 cysteine (Fig. 1) located internally at the catalytic center and near the ligand-binding cleft, showing no possibility to form disulfide bridges, either internally or externally with other proteins.

Analysis of physicochemical parameters of allergenic IRLs displayed a molecular weight range from 33,573.6–34,190.3 Da, and an isoelectric point with acidic character ranging 5.51–6.73 (Table S1). All sequences exhibited hydrophilic character, as indicated by the negative average value (-0.0718 ± 0.001) of the Grand index (GRAVY), and may be considered as stable proteins (average aliphatic index value of 97.422 ± 2.87; average stability index of 30.034 ± 7.177) (Table S1).

Post-translational motifs showed high degree of polymorphism (Table S2). More than 19 potential N-myristoylation motifs, ranging 3–4 per sequence, and two differential glycosylation motifs (0–2 sites per sequence).

Phosphorilation motifs were found the most variable feature. Ten different types of kinases were predicted to be phosphorylated the IRLs sequences (Table S3), with a variable number of multi-optional phosphorilation sites in residues such as serine (3–4 residues), threonine (2–4 residues) and tyrosine (2–3 residues) (Table S3).

Phylogenetic analysis

Phylogenetic analysis was performed in order to determine the relationships between IRL sequences in different species, including those from allergenic pollen, aiming to gain insights into different functional groups (Fig. 2). A hundred and thirty six sequences of a wide representation of IRLs were aligned and cluster-analyzed. 18 groups were established. The group that contained olive IRL sequence, and defined as group 1, was integrated by sequences from the allergenic species *F. excelsior*, in addition to sequences from *Striga asiatica*, *F. intermedia*, and *Pinus tadea*. Both birch allergenic protein sequences were clustered in a small group (group 18) together with sequences from poplar and eucalyptus PCBER proteins. Other IRL allergenic sequences, such as those from pear, potato and carrots were clustered in the group 3, which was mostly integrated by IRLs, while the sequence from Japanese cedar was clustered in the group 6, which was integrated almost exclusively by PCBER sequences.

Searching for templates for Ole e 12

Searching for proteins with known tertiary structure in the Protein Data Bank (PDB), *P. taeda* phenylcoumaran benzylic ether reductase PT1 (Uniprot accession number Q9LL41), *M. sativa* isoflavone reductase (P52575), and *Thuja plicata* pinoresinol-lariciresinol reductase (Q9LD14), PDB accession number 1qyc, 2gas, and 1qyd, respectively, showed highest sequence identity of 67, 58, 48 % with Ole e 12, respectively. The suitability of the selected model was evaluated by BioInfoBank Metaserver which returned a 3D Jury score (J-score) of 96.85 (pine), 92.68 (alfalfa), and 85.60 (green giant) for Ole e 12, respectively. The crystal structure of pine, green giant and alfalfa were retrieved from PDB based in these above results. We also used the Swiss-model server to identify the best possible template to built the Ole e 12 structure, finding a high score of 393, 341, and 279, respectively, and very low E-value of e-109, 5e-94 and 2e-75, respectively for the templates 1qyc, 2gas and 1qyd retrieved from PDB database and used for homology modeling.

Structural assessment of the Ole e 12 model

Different tools and parameters have been used to assess the quality of the models built for this study and summarized in the Table 1.

I-TASSER analysis returned parameters scores, which values were between the limits of high structural accuracy on the basis of C-score (2<C-score>-5), TM-score, large number of decoys performed, and cluster density for Ole e 12. The main chain conformations of the Ole e 12 model were located in the acceptable regions of the Ramachandran plot (Fig. S1a) reflected in the Procheck analysis parameters (Table 1), displaying most of the plot of $\times 1$ versus $\times 2$ torsion angles for each residue in low energy regions, and main-chain and side-chain parameters in the best region. The goodness factor (G-factor) based on the observed distribution of stereo-chemical parameters (main chain bond angles, bond length and phi–psi torsion angles) returned accurate values for a reliable model (Fig. S1a, Table 1) [36], in comparison with the templates 1qyc, 2gas, and 1qyd (Table 1).

ProSa analysis returned Z-scores within the range usually found for native proteins of similar size such as 1qyc, 2gas, and 1qyd (Table 1), and negative interaction and lowest energy ranges as showed in ProSAweb energy plots (Fig. S1b, Fig. S1d) for all residues of Ole e 12 structural model.

QMEAN analysis returned a Q value for Ole e 12 structure in concordance and within the range of the values of the templates (Table 1).

Root mean square deviation (RMSD) between Ole e 12 structural model and the crystal templates Cα backbones of the closed templates were 0.072, 0.953, and 0.943 Å for 1qyc, 2gas, and 1qyd, respectively (Table 1).

The parameters selected assessing the quality of the olive IRL structure model were similarly evaluated for other IRLs structures built, showing comparable values.

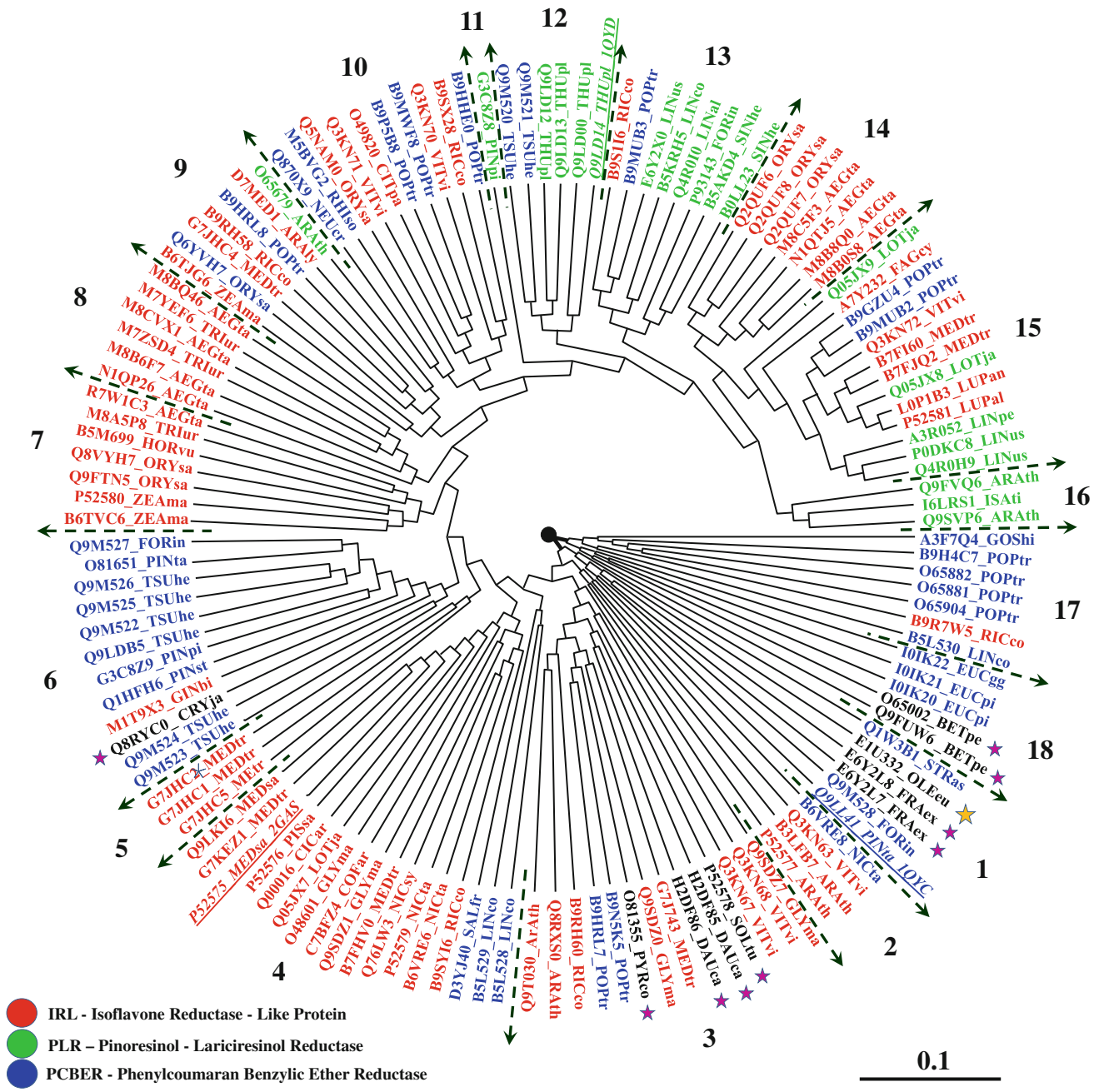

Fig. 2 Phylogenetic analysis of isoflavone reductase—like proteins. Neighbor-joining (NJ) method was used to perform a phylogenetic analysis of 136 IRLs from 52 species such as *Aegilops tauschii* (AEGta), *Arabidopsis lirata* (ARALi), *Arabidopsis thaliana* (ARAth), *B. pendula* (BETpn), *Citrus paradisi* (CITpa), *Cicer arietinum* (CICar), *C. paradisi* (CITpa), *Coffea arabica* (COFar), *C. japonica* (CRYja), *D. carota* (DAUca), *Desmodium uncinatum* (DESun), *Eucalyptus globulus* (EUCgl), *Eucalyptus pilularis* (EUCpi), *Fagopyrum cymosum* (FAGcy), *F. intermedia* (FORin), *F. intermedia* (1QYD), *F. excelsior* (FRAex), *Ginkgo biloba* (GINBI), *Glycine max* (GLYma), *Gossypium hirsutum* (GOShi), *Hordeum vulgare* (HORvu), *Isatis tinctoria* (ISAti), *Linum album* (LINal), *Linum corymbulosum* (LINco), *Linum perenne* (LINpe), *Linum usitatissimum* (LINus), *Lotus japonicus* (LOTja), *Lupinus albus* (LUPal), *Lupinus angustifolius* (LUPan), *M. sativa* (2gas model template), *M. sativa* (MEDsa), *Medicago truncatula* (MEDTR), *Neurospora crassa* (NEUCr), *Nicotiana sylvestris* (NICsy), *Nicotiana tabacum* (NICta), *O. europaea* L. (OLEeu), *Oryza japonica* (ORYja), *Oryza sativa* (ORYsa), *Pinus pinaster* (PINpi), *P. taeda* (PINta), *P. taeda* (1qyc model template), *Pisum sativum* (PISsa), *Pinus strobus* (PINst), *P. communis* (PYRco), *Rhizoctonia solani* (RIZso), *Ricinus communis* (RICco), *Salvia fruticosa* (SALfr), *Sinopodophyllum hexandrum* (SINhe), *S. tuberosum* (SOLtu), *S. asiatica* (STRas), *T. plicata* (1qyd model template), *T. plicata* (THUpl), *Tsuga heterophylla* (TSUhe), *Triticum urartu* (TRIur), *Vitis vinifera* (VITvi), *Zea mays* (ZEAma). Eighteen defined groups were identified and highlighted by *different numbers* and *green arrows*. Sequences already identified as IRL, PLR and/or PCBER were highlighted by *red*, *green* and *blue colors*. IRLs identified as allergen are highlighted by *purple stars*, and Ole e 12 with a *yellow star*

Table 1 Parameters used for proteins structural assessment

Protein	Structural assessment methods	Scores	Ramachandran plot (%)	G-factor	MCBL (%)	CBA (%)	Z-score	Q-value	RMSD (Å)
Ole e 12	I-TASSER analysis	1.26[A]							
		0.89 ± 0.07[B]							
		9952[C]							
		1.2500[D]							
	Procheck analysis		88.1[E]	−0.15[I]	99.4	96.6			
			10.8[F]	0.19[J]					
			0.4[G]	0.01[K]					
			0.7[H]						
	ProSa						−9.87		
	QMEAN							0.729	
	RMSD (Å)								0[L]
1QYC	I-TASSER analysis	0.79[A]							
		0.82 ± 0.08[B]							
		9144[C]							
		0.6667[D]							
	Procheck analysis		90.1[E]	−0.15[I]	99.6	89.7			
			8.4[F]	−0.15[J]					
			1.3[G]	−0.11[K]					
			0.2[H]						
	ProSa						−10.3		
	QMEAN							0.778	
	RMSD (Å)								0.072[L]
2GAS	I-TASSER analysis	1.53[A]							
		0.93 ± 0.06[B]							
		10200[C]							
		1.2500[D]							
	Procheck analysis		91.6[E]	0.23[I]	100	99.1			
			7.8[F]	0.52[J]					
			0.5[G]	0.36[K]					
			0[H]						
	ProSa						−9.91		
	QMEAN							0.775	
	RMSD (Å)								0.953[L]
1QYD	I-TASSER analysis	1.38[A]							
		0.91 ± 0.06[B]							
		9024[C]							
		1.2500[D]							
	Procheck analysis		77[E]	−0.51[I]	98.9	80.4			
			19[F]	−0.60[J]					
			2.7[G]	−0.53[K]					
			1.3[H]						
	ProSa						−8.95		
	QMEAN							0.722	
	RMSD (Å)								0.943[L]

MCBL distribution of the main chain bond lengths, *CBA* distribution of the covalent bond angles

[A] C-score; [B] TM-score; [C] Decoy N°; [D] cluster density; [E] Residues in favorable regions; [F] residues in allowed regions; [G] residues in generally allowed regions; [H] residues in disallowed regions; [I] G-factor score of the dihedral bonds; [J] G-factor score of the covalent bonds; [K] overall G-factor score; [L] root mean square deviation between Cα Ole e 12 structure and each template

Structure analysis of Ole e 12

In the present work, we have built the best model for Ole e 12, based in homology modeling. The overall structure of Ole e 12 is composed of two domains (N- and C-terminal domains) separated by a cleft (Fig. 3a). The N-terminal domain contains residues 1–153, 190–215, and 278–298. It adopts a Rossmann-fold (3α-8β-3α) motif: eight parallel β-strands forming a central pleated sheet surrounded by six α-helices (Fig. 3b). All but the eighth β-strand belong to the N-terminal amino acids. The smaller C-terminal domain showed homology with UDP-galatose 4-epimerase, which is supposed to be substrate specific in SDRs, contains four α-helices, two helices 3_{10}, and 1 β-pleated sheet with two parallel strands (Fig. 3a). The binding sites for the cofactor NADPH and the substrates are located in a large cleft formed between the N- and C-terminal domains of enzymes (Fig. 3a). The NADPH may bind near the conserved fingerprint region (e.g., G^{11}-X-X-G^{14}-X-X-G^{17} region in IFR-like proteins) in the N-terminus.

Secondary structure prediction with SSpro8 server identified 12 α-helices and 10 β-sheets in Ole e 12, whereas PredictProtein, and NetSurfP v1.1 servers predicted 13 α-helices and 10 β-sheets for Ole e 12. The 13 α-helical regions and 10 β-sheets were recognized by PSIPRED, which were superimposed on 3D structure of Ole e 12 (Fig. 3c). PHD predicted 14 α-helices and 10 β-sheets for Ole e 12. These predictions made by different servers correlated well with the modeled 3D structure built.

A 3D search alignment, showed that the best superimpositions are obtained with three proteins of the SDR superfamily: IFR from *M. sativa* (PDB code: 2gas), PLR

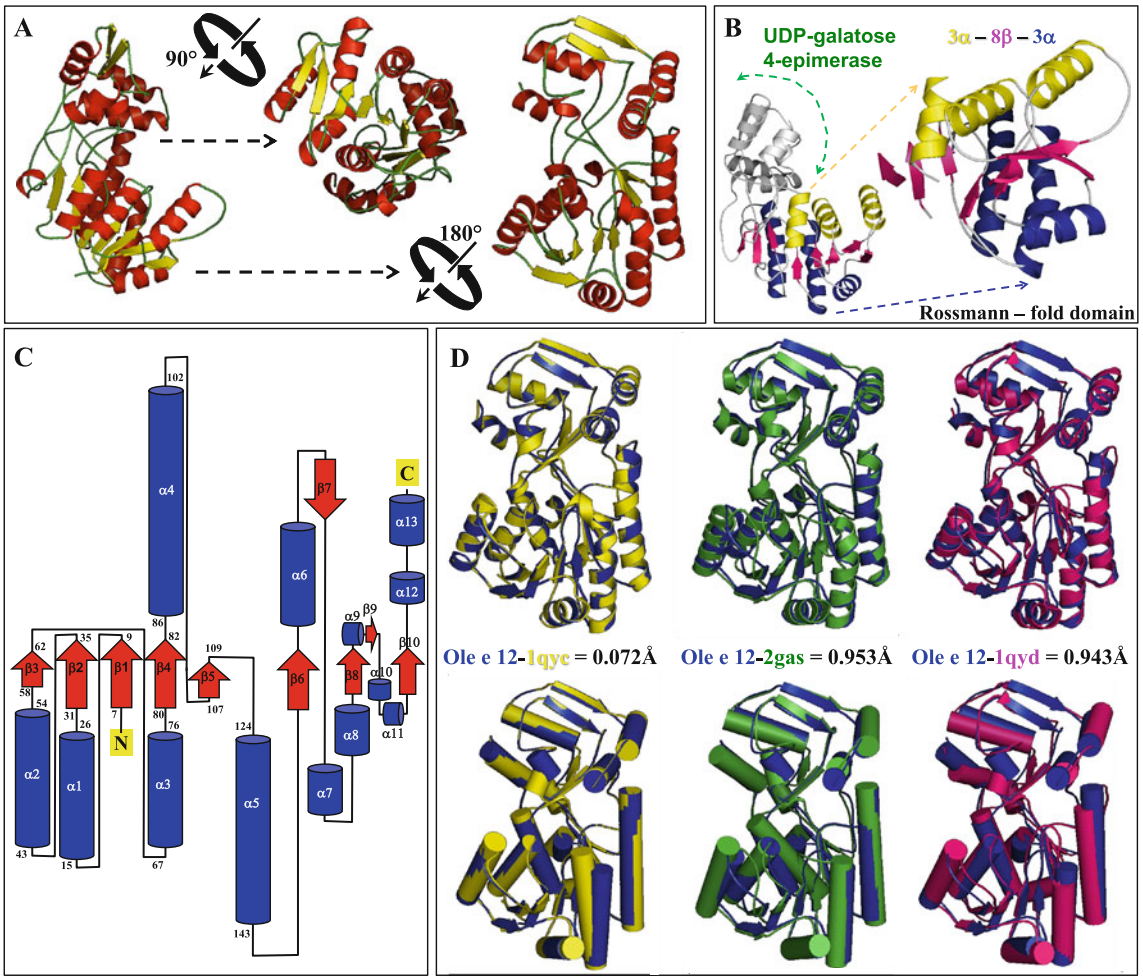

Fig. 3 Structural analysis of Ole e 12. **a** Three-dimensional structure of olive pollen Ole e 12 (Uniprot accession number E1U332). Structures were depicted as a cartoon diagram. α-helices, β-sheets and coils are depicted in *red*, *yellow* and *green* respectively. Three views (rotated 90 around the x-axis, one looking down from the N-terminal side) are provided for Ole e 12. **b** Three-dimensional structure of Ole e 12 is depicted highlighting different domains of this protein (UDP-galatose 4-epimerase and Rossmann-fold domains). A detailed view of the Rossmann-fold (3α-8β-3α) domain is depicted. **c** Secondary structure of Ole e 12 elements, α-helices and β-sheets, are depicted in *blue* and *red colored*, respectively. **d** Superimposition between olive pollen Ole e 12 and IFR (2gas), PLR (1qyd), and PCBER (1qyc) are depicted in *blue* (Ole e 12), *yellow*, *green* and *pink* (PCBER, IFR and PLR) *color*

from *F. intermedia* (PDB code: 1qyd) and PCBER from *P. taeda* (PDB code: 1qyc) (Fig. 3d).

When Ole e 12 was superimposed on PCBER, IFR and PLP, RMSD values for superimposed Cα of 0.072, 0.953, and 0.943 Å were respectively obtained (Fig. 3d), having a continuous alpha/beta NADPH-binding domain and a smaller substrate-binding domain. They are the only three-dimensional X-ray crystal structures that have been solved and refined, showing relatively low identity among amino acids sequences, but a high structural similarity (Table S4), especially between IFRs and PCBERs. This general fold was also quite similar when comparisons were established among other IRLs from allergenic (Table S4a), and non-allergenic plant species (Table S4b). Moreover, a relatively good conservation of the overall fold (Cα carbon chain) of IRLs was found among plant species (Table S4b).

Individual comparisons between Ole e 12 and other allergenic IRLs showed small structural differences among them; maize PCBER structure was the closest one with a RMSD of 0.043Å among the 31 non allergenic plant structures analyzed (Table S4b). In addition, comparison between allergenic pollen IRLs and the three protein templates models through structures superimposition had showed comparatively lower RMSD differences, with the lowest PCBER value being 0.072Å when compared to Ole e 12 (Table S4a).

Overall, Ole e 12 is quite similar in polypeptide-chain fold to three other IRLs family proteins structurally characterized, PLR [19], IFR [18], and PCBER [19]. Structurally, Ole e 12 and the other three members of the PIP-family proteins organize around an N-terminal, Rossmann-fold domain, containing a core, typically between six to eight-stranded parallel β-sheets flanked on each face by an α-helical layer (Fig. 3). One edge of the core β-sheet provides the extended binding surface for the NADPH cofactor. The C-terminal polypeptide-chain segment of the PIP-family proteins forms a predominantly α-helical domain, and this C-terminal segment also contributes three additional β-strands to the Rossmann-fold domain. The C-terminal domain may function in substrate binding. Indeed, this domain together with the last α-helix of the Rossmann-fold domain is part of the cavity located immediately adjacent to the nicotinamide ring of the NADPH cofactor. Within the IRL family, the substrate-binding domains appear more structurally divergent than the nicotinamide-cofactor binding domains. From comparisons of polypeptide-chain backbones (Fig. 3c), Ole e 12 differs from PCBER, IFR and PLR by 0.070, 0.798 and 0.0.706 Å for the Rossmann-fold domain, respectively.

Structural comparisons between Ole e 12 and IRLs (IFR, PCBER, and PLR) also showed that the largest differences are present in the loop β2-α2 region (Fig. 3d), where a specific IFR amino acids motif insertion is located (Fig. 1). A second region showing large differences among IFR, PCBER and PLR is also close to this loop region, a loop between strand β4 and the helix α4 (Fig. 3d), where a specific PLR amino acids motif insertion is located (Fig. 1) and where PLR is four residues longer. All these flexible regions are near the putative NADPH-binding pocket and may play roles in the binding of NADPH. Since the NADPH molecule is relatively large and flexible compared to the substrate (flavonoids), flexibility of the corresponding binding regions, which may assume different conformations in different enzyme molecules, would be necessary. These structural differences between IRLs, including Ole e 12, would define the different topologies of substrate-binding pockets in these different enzymes, and also may reflect the differences in their binding pockets for co-factor NADPH.

Molecule surface, electrostatic potential, conservational and docking analysis of Ole e 12

Surface of Ole e 12 (rotated 180°) showing the secondary structure elements is depicted in the Fig. 4a. The morphology of the cavity that accommodates the NADPH cofactor of the IRLs is integrated by a solvent-accessible cleft in between catalytic and oligomerization (UDP-galatose 4-epimerase) domain of the protein, formed by the external loops (Figs. 3a, 4a). The central part of this cleft is lined by several polar (positive) residues (i.e. Lys^{45}, Arg^{36}, Gln^{86}) (Figs. 4a, 5), as is characteristic for accommodating the negative charges of the cofactor phosphates. The putative active (ligand-binding) site is located near the cofactor, on the same cavity, and contains the amino acids Gln^{86}, Gly^{114}, Thr^{115}, Phe^{130}, Lys^{133}, Asn^{153}, Tyr^{158}, Ser^{159}, Ile^{261}, Ile^{265}, several of them well conserved (Gly^{114}, Lys^{133}, Asn^{153}, Tyr^{158}, Ser^{159}) (Fig. 1).

Surface electrostatic potential analysis, reveals several prominent charged residues, with more than half of the side exhibiting large negative values, red regions (Fig. 4a). By assigning a value of +1 to basic residues (Arg, Lys) and -1 to the acidic residues (Asp, Glu), net charge of protein was calculated to be −5 for Ole e 12.

Consurf conservational analysis has shown that IRLs are not well conserved protein, especially residues located in the surface of the protein (Fig. 4b). This is in agreement with the relatively low identity of the amino acid sequence calculated, as well as with multiple alignment comparison between plant species. The only region of the protein being better conserved is the one integrating Rossmann-fold domain (Fig. 4b), in addition to others amino acids with a major role in the maintenance of the protein structure. A clear distribution of the conserved (purple) and variable (turquoise) residues can be appreciated in the core of this

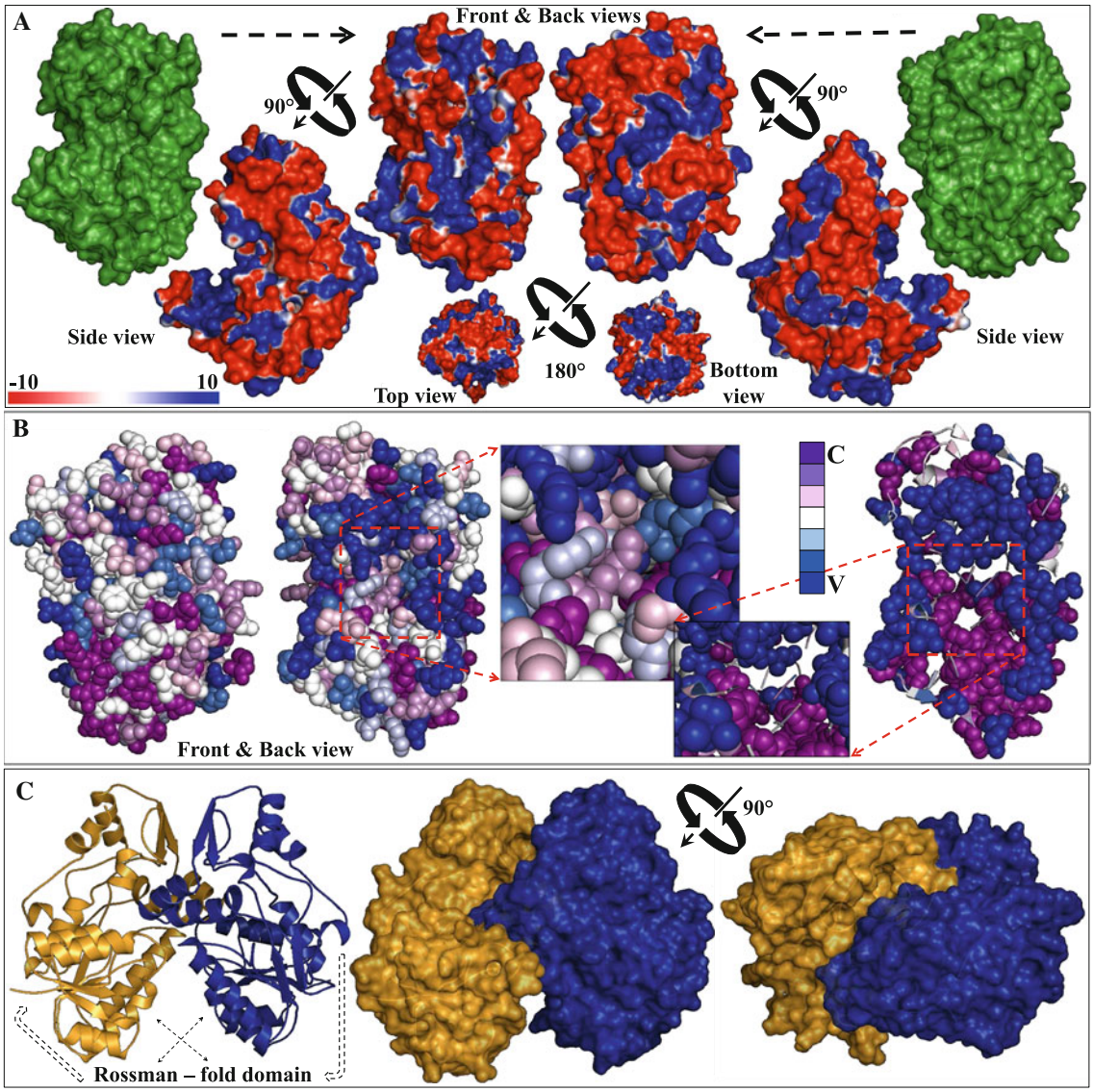

Fig. 4 Olive Ole e 12 surface structure, electrostatic potential, conservational and molecular docking analysis. **a** Surface representation views of Ole e 12 rotated 180° (*green color*), showing the NADPH and substrate binding cleft. 90 and 180° rotated views of the electrostatic potential representation on the Ole e 12 protein surfaces. The surface colors are clamped at *red* (−10) or *blue* (+10). **b** Consurf-conservational analysis of Ole e 12 protein showed in two individual views rotated 180°, respectively. The conserved and variable residues are presented as space-filled models and colored according to the conservation scores. A detailed view of the cavity with active center is shown in high magnification. Additionally, a detailed view of Ole e 12 protein showing the distribution of strictly conserved (*purple*) and variable (*blue*) residues is depicted. The active center is shown in high magnification. *C* conserved, *V* variable. **c** Docking representation of Ole e 12 dimmer. One monomer is shown in *blue* and the other in *gold*. Two molecules in the dimmer are surface represented in views of Ole e 12 rotated 90°

domain, although the surface of the protein seems to be more dominated by variable residues (Fig. 4b).

Docking analysis revealed the way of dimerization of Ole e 12. It forms twofold rotationally symmetric homodimer, with a defined homodimeric association (Fig. 4c). The surface of interaction between both subunits is relatively large (>3,000 Å^2), limiting the accessibility to cofactor and substrate to the binding pockets. The electrostatic potential in the dimer interaction interface shows good charge/chemical complementarily (Fig. 4a).

The co-factor and substrate-binding pockets of the two molecules in the dimer face form a very narrow channel, thereby hindering their accessibility. Between these two molecules, there is extensive hydrogen bonding interactions between these two molecules. These contacts are near loops β2-α2, β4-α4, β7-β8, and α8-α9.

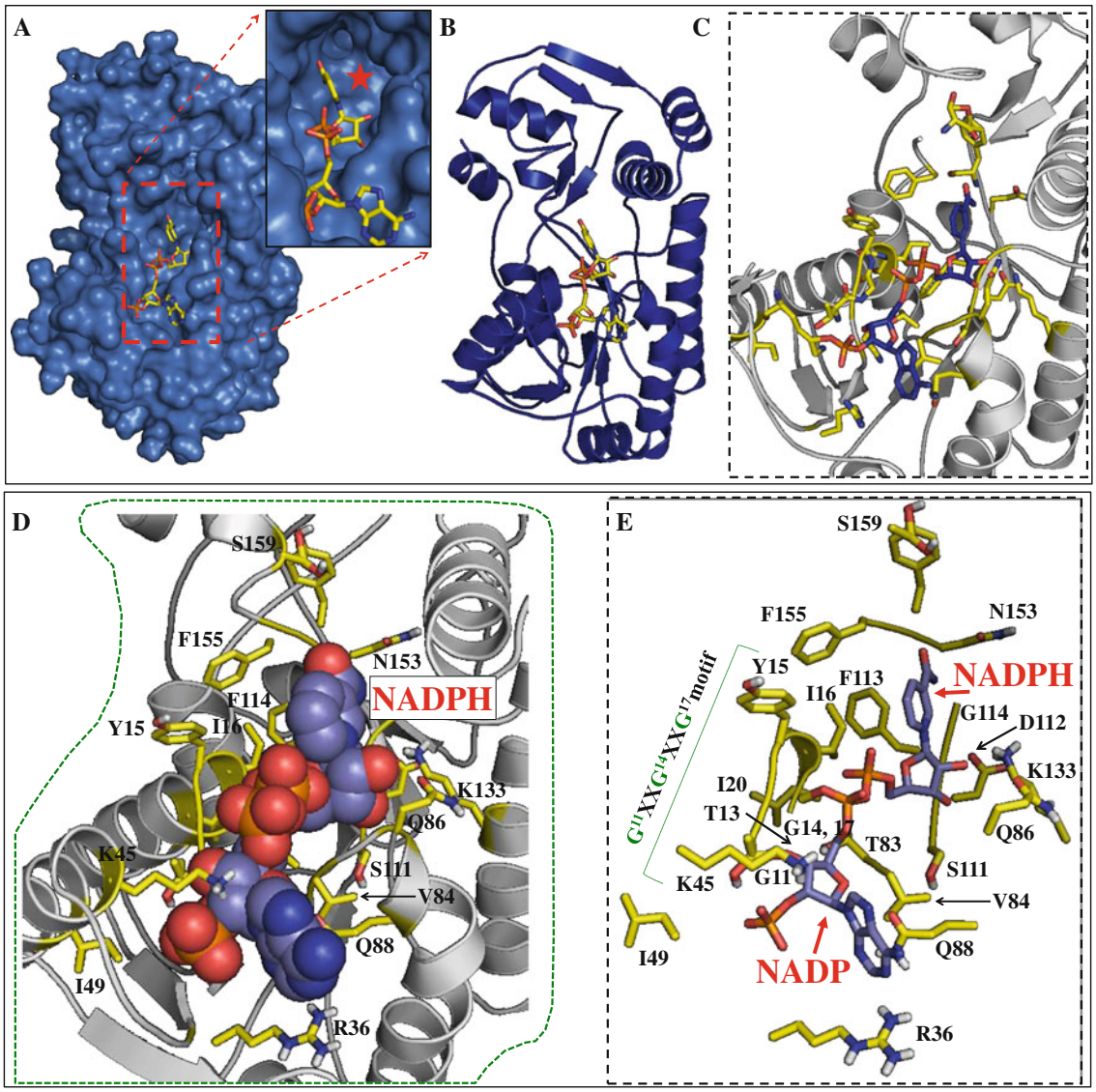

Fig. 5 Ligand-binding domain analysis of Ole e 12 protein. **a** Surface representation of the cofactor (NADPH) and substrate binding cleft in *blue color* showing the NADPH chain lining on the cleft binding surface. A detailed view is shown, and the substrate location is highlighted with a *red star*. **b** *Blue cartoon* representation of Ole e 12 showing the cofactor (NADPH) binding domain. **c** Distribution of the NADPH cofactor and the spatial distribution of the residues that integrate the cofactor-substrate binding cleft. Residues are depicted as *stick* and *colored* according with atoms. **d** Detailed view of the active site with the cofactor depicted as *spheres*, and the interacting residues of Ole e 12 structure is depicted as *stick* and *colored* according with the atoms. **e** Detailed representation of the amino acids interacting with NADPH and stabilizing the cofactor in the catalytic domain. This view includes a detailed representation of the key GxxGxxG motif of NADPH interaction

Ole e 12 co-factor (NADPH) binding site

A conserved and canonical sequence-motif for NAD(P) binding [34], G^{11}-X-X-G^{14}-X-X-G^{17} fingerprint region, is present in the N terminus (loop β1-α1) of Ole e 12 protein. The cavity holding the cofactor is partially open to the soluble phase. The NADPH pyrophosphate group fit between loop β1-α1 (in proximity to Arg^{36}) and loop β4-α4 (Fig. 5a, b) and the cofactor adopts an extended conformation. The pyrophosphate group of NADPH is located between these two loops and likely forms hydrogen bonds with main-chain atoms. The adenine and phosphate group of NADPH may fit into a small pocket formed by loops β1-α1, β2-α2, β3-α4, and β4-α5 (Figs. 3, 5b, c).

The nicotinamide ribose appears in a C2′ endo conformation and interacts with Ser^{111} and Lys^{133}, via its hydroxyl groups. The nicotinamide ring adopts an anti-conformation (pro-R). Its A-face is oriented towards the substrate binding pocket, whereas its B face stacks against hydrophobic residues Gly^{11} and Gly^{14}, Thr^{13}, and Phe^{113} side chains. Thus, Ole e 12 can be considered as a pro-R (A-type) reductase.

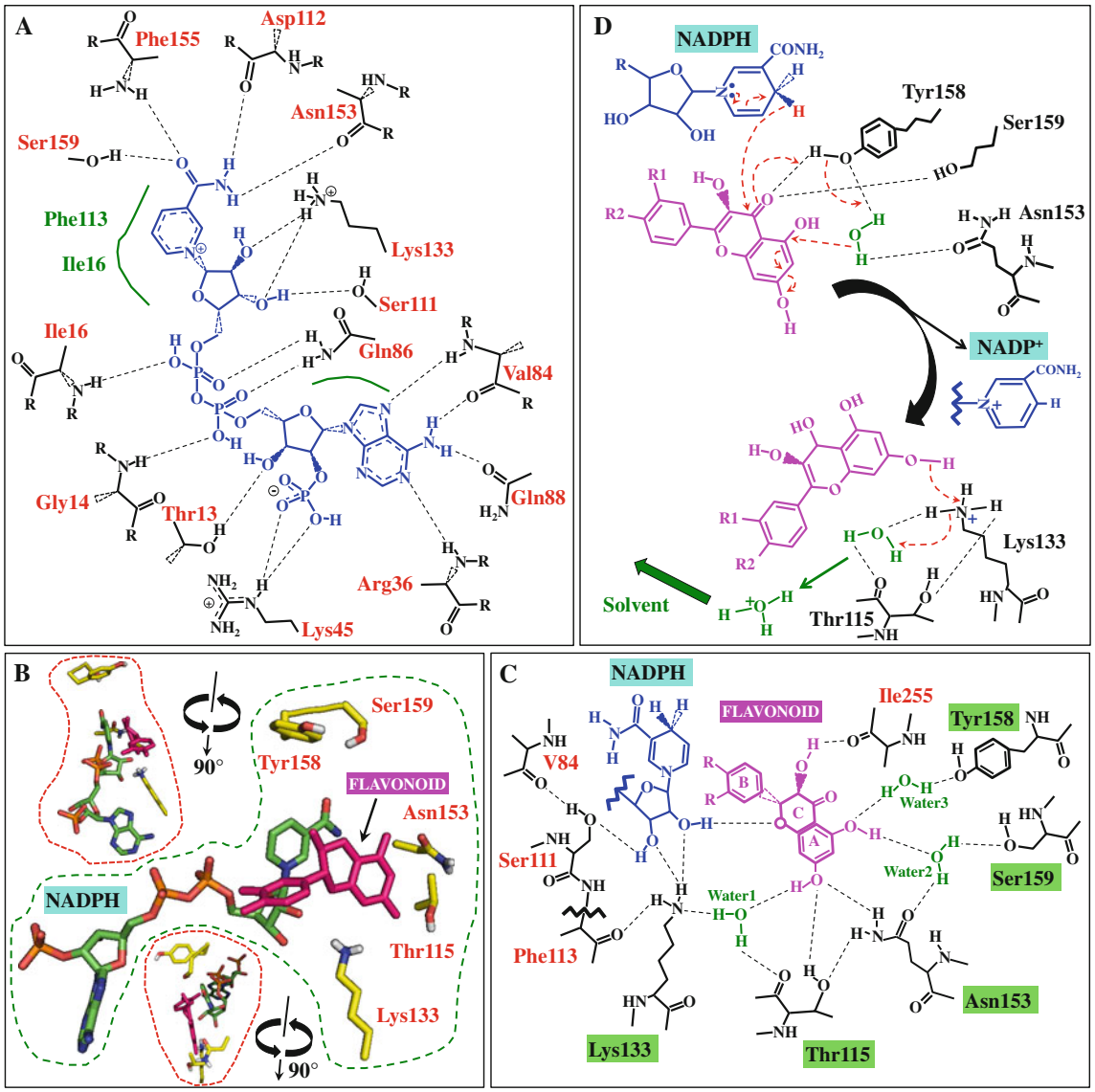

Fig. 6 Hydrogen-bonding interactions in the olive Ole e 12 active-site and proposed reaction mechanism. **a** Ole e 12 active site and its interaction with NADPH. Hydrogen bonds are shown in *broken lines*; NADPH is represented in *blue color*. **b** Detailed view of the cofactor and substrate (*pink color*) depicted as *stick*, and the relative position in the active site. Amino acids involved in the catalytic mechanism are also depicted as *sticks*. Two 90 rotated views are also depicted. **c** Ole e 12 active site and its interaction with NADPH (*blue color*) and the substrate (*pink color*). Hydrogen bonds are shown in *broken lines*. Amino acids involved in the catalytic mechanism are highlighted in *green* background. **d** Catalytic mechanism proposed for Ole e 12

The coenzyme interacts with the protein through a network of H-bonds, electrostatic interactions, and hydrophobic contacts (Figs. 5d, e, 6a). These interactions mostly involve residues of the Rossmann-fold. Two basic residues, Lys^{45} and Arg^{36}, may play important roles in interacting with adenine and 2′-phosphate groups, potentially forming a salt-bridge with the 2′-phosphate group and stacking with the adenine ring (Figs. 5c, d, 6a).

The adenine ring adopts the anti conformation and is sandwiched between the d-guanido group of Arg^{36} and the carboxamide group of Gln^{88}. The adenine-ribose lies in the C39-endo conformation. The ribose ring is packed against the α-carbons of both Thr^{13} and Gly^{14}, and the central diphosphate group forms hydrogen bonds with the backbone amide-nitrogens of residues Ile^{16} and Gln^{86} (Figs. 5e, 6a).

Other residues contribute to the stability of the cofactor, i.e. Lys^{45}, Val^{84}, and Ser^{111} (Fig. 6a), and the cofactor is bound through a number of other polar and non-polar interactions. The 2′-phosphate is sequestered by a short loop formed by residues 35–43, and is hydrogen bonded to the side chains of Thr^{13}, Arg^{36} and Lys^{45}. Thr^{13} from the Glycine-rich loop is also hydrogen bonds to the 2′-phosphate, as well as the adjacent 3′-hydroxyl group.

The nicotinamide-ribose adopts a C29—endo conformation, and its hydroxyl groups are involved in hydrogen bonds with residues Asp^{112} and Asn^{153} carbonyl oxygen, Ser^{111} side chain hydroxyl moiety, and the side-chain amino group of Lys^{133}, and hydrogen bonds to backbone atoms of Phe^{155}, and Ser^{159} (Fig. 6a).

Ole e 12 substrate binding site

Polyphenols are abundant in pollen from different plant species, and exhibit potent antioxidant activity [37]. Flavonoids constitute the most important class, represented by a plethora (>5,000) of different compounds already described and based in the flavonoid skeleton (Fig. S2a). Pollen flavonoids share a rings structure with different combinations of radicals in the positions C3, C7, C8, C3′ and C4′ (Fig. S2b), for which the combined substitution produce a variety of structures, and some examples are depicted in the figure S2c. Quercetin in one of the most abundant polyphenols in pollen, together with Kaempferol and Apigenin [38].

Notably, stacking of a substrate aromatic-ring against the NAD(P) nicotinamide ring is a common feature of the binding modes of SDRs [39]. Below-further the nicotinamide ring of the docked NADPH, there is a pocket formed by loops β4-α4 and β5-α5, and helices α5 and α9, and segment β6-β8 with a basic residue, Lys^{133}, located in the bottom (Figs. 5, 6b), further confirming the role of Lys^{133} as a catalytic residue.

The substrate quercetin fits well into this pocket, close to the NZ atom of Lys^{133}. The aromatic ring C of the substrate may stack with the nicotinamide ring of NADPH. An aromatic residue, Tyr^{158}, might stack/interact with the nicotinamide ring of NADPH from the other side (Fig. 6b, c). Lys^{133} is conserved in all IRLs. In the model docked with substrate (Fig. 6b, c), this conserved lysine residue may act as a general proton donor and is essential for enzyme catalysis and functioning. Various hydrophobic residues (Thr^{115}, Ala^{126}, Phe^{130}, Phe^{155}, Ala^{156}, Ile^{265}) and loop β4-α4 (Thr^{83}, Val^{84}, Leu^{87}) may provide a hydrophobic environment or structural support for the substrate. Otherwise, several residues identified from the docking study for possible interaction with substrate, such as Thr^{115}, Lys^{133}, Asn^{153}, Tyr^{158}, Ser^{159}, being Lys^{133} and Tyr^{158} are conserved among pollen allergens IRLs (PLR, IFR and PCBER), including Ole e 12 proteins. They may be related to the substrate recognition and specificity. Other residues have been described as involved in stereo-specificity and stereochemistry. Hydrophobic amino acid Leu^{128} in PLR (1qyd model) involved in stereochemistry is conserved in Ole e 12 as Val^{123}, and Val^{134} for IFR (2gas) and Val^{124} in PCBER (1qyc) (Fig. 1). Ser^{159} in Ole e 12 is substituted by corresponding Phe^{170} in the other three IRLs. Val^{262} in Ole e 12 is only conserved in PLR (1qyd), and Asn^{266} is not conserved in the others IRLs (IFR, PCBER and PLR). Finally, Val^{300} in Ole e 12 is conserved only in PCBER. According to the structural location of these 5 amino acids, Ser^{159} and Asn^{266} are likely to be implicated in the specificity of substrate in Ole e 12.

Catalytic mechanism of Ole e 12

Ole e 12 has been assumed to be a member of an SDR NADPH-dependent reductase family. Structural analysis of the Ole e 12 active site has shown to have an Asn-Ser-Tyr-Lys consensus group of amino acids proposed to be involved in the catalytic mechanisms, with Tyr^{158} as an intermediate electron carrier [40], instead of a conserved catalytic triad. The location of these catalytically essential amino acids differs from others SDR proteins. Lys^{133} lies in a position that enables its ε-amino group to interact with the nicotinamide ribose hydroxyl groups and with the quercetin C7-OH group via a water molecule (Wat1) (Fig. 6c). Ser^{159} hydroxyl groups are hydrogen-bonded to the hydroxyl group in C5 of quercetin through a second water molecule (Wat2), helped by Asn^{153} (Fig. 6c), and by Tyr^{158} hydroxyl group through a third water molecule (Wat3).

Ser^{159} together with Tyr^{158} and Lys^{133} may act as the catalytic triad. These residues are conserved in the most of the IRL homologs, and may be essential for enzyme catalytic activity and functioning. The pocket near the catalytic triad is presumably the substrate-binding site. Moreover, Ole e 12 has a feature not common to most of SDRs family members, which is the lack of the consensus sequence Ser…TyrXXXLys. However, Ole e 12 conserve four amino acids, Lys^{133}-Asn^{153}-Tyr^{158}-Ser^{159}, which have been proposed in this study to undertake the catalytic activity of this protein. Lys133 is most likely a triad member that has been conserved, which suggest that the Lys residue has a generic function in both SDRs enzymes family.

Thus, the side chains of those four amino acids could be involved in the catalytic mechanism (Fig. 6d): Asn^{153}, Tyr^{158}, Lys^{133}, and Ser^{159}. Tyr^{158} is the only one in direct contact with the substrate. Lys133 may be deprotonated in the ternary complex, whereas its protonated state in the binary complexes may be acting as acceptor of the electron from the substrate. Tyr^{158} occupies an ideal structural position for proton capture and as an electron carrier between a water molecule and the C4 of the substrate, and Ser^{159} may form a hydrogen bond with this oxygen atom to assist in the reaction. Finally the Lys^{133} with the assistance of the ammonium form of Lys^{133} may act as a general acid–base catalyst through hydrogen-bonded Wat1 to deprotonate and transfer a H^+ and an electron to the water

molecule. Lys^{133} may interact with the ribose of NADPH and form hydrogen bonds to its 2′ and/or 3′-hydroxyl groups, to stabilize the position of the nicotinamide moiety. According to this mechanism, the major role of Tyr^{158} is to contribute to an optimal electron withdrawing H-bond to O7. The Tyr^{158} hydroxyl group orients its phenolic substrate in such a way that the C5-OH hydrogen remains in optimal interaction with this water molecule, which is at the same time interacting with Asn^{153}, and the catalytic lysine is hydrogen-bonded to the substrate via a water molecule. Ole e 12 adds a hydride anion to the electrophilic carbon of a C=O double bond, with the concomitant extrusion of a leaving group (H_3O^+), and the cetonic group of quercetin is reduced to a hydroxyl group in association with the oxidation of one NADPH molecule.

Identification of high antigenicity regions of olive Ole e 12 sequence

Parameters such as hydrophilicity, flexibility, accessibility, turns, exposed surface, polarity and antigenic propensity of polypeptides chains have been correlated with the location of continuous epitopes. Hydrophobicity (or hydrophilicity) plots were designed to display the distribution of polar and apolar residues along a protein sequence. Protein antigenicity is a surface property, and in our study, antigenicity determinants have been targeted by locating the positive peaks in hydrophilicity plots, thus identifying the regions of maximum potential antigenicity (Fig. S3). Kyte-Doolitte scale [29] was used to search hydrophobic regions in the proteins. Hopp-Woods scale [32] was used for predicting potential antigenic sites of protein which is essentially a hydrophilic index, with apolar residues assigned negative values. Welling antigenicity plot [30] gives antigenicity value as the log of the quotient between percentage in a sample of known antigenic regions and percentage in average proteins. Parker [31] antigenicity method was also analyzed. We identified in the sequences of Ole e 12, up to 8 regions with high potential of antigenicity (Fig. S3), which correlated well with the B- and T-cell epitopes predicted by using different methodologies.

T-cell epitopes prediction of olive Ole e 12 sequence

Different publicly available tools were employed for determination of HLA class II binders. Four amino acid sequences (P6: 7–21, P7: 92–110, P8: 255–265, P9: 292–305) (Fig. 7a, Table S5) were identified as promiscuous binders by comparison of prediction data. One of these epitopes was located in the N-terminal region. The epitopes were located in highly antigenic regions previously identified in the Fig. S3. The predicted epitopes comprised of 64 residues with a high frequency of occurrence for Ile (17.19 %) and Val or Ala or Gly (14.07 %). 59.38 % of the total residues were non polar whereas 18.75 % were polar non-charged. Charged residues (21.86 %) were least preferred in the T-cell epitopes. All four T-cell epitopes are located in areas of the protein, in which most of the amino acids integrating the epitopes displayed a solvent accessibility greater than 30 and 50 %. All four T-cell epitopes are primarily composed of parts of α-helices and loops. Surface distribution of T-cell epitopes were superimposed in the structure Ole e 12 and depicted in the Fig. 7a.

B-cell epitopes prediction of olive Ole e 12 sequence

Five antigenic regions were predicted most common in Ole e 12 by several prediction tools used in the present study. These seven regions are P1: 25–35, P2: 37–55, P3: 114–130, P4: 160–175, and P5: 195–210 (Fig. 7b, Table S5) were identified as promiscuous binders by comparison of prediction data. The predicted epitopes comprised of 83 residues with a high frequency of occurrence for Ala or Lys (9.64 %); Val or Phe (8.4 %) were the most common residues in epitopes. Hydrophobic content was analyzed to identify regions with a higher probability of interaction with immunoglobulin. Around 48.19 % of the 83 residues were non polar, 28.92 % were charged residues and 22.89 % were recorded as polar non-charged residues for the predicted epitopes. The relative surface accessibility values were greater than 30 % for more than 40 % of the residues whereas more than 25 % residues presented a solvent accessibility above 50 %. Surface distribution of B-cell epitopes were superimposed in the structure Ole e 12 and depicted in the Fig. 7b.

Discussion

Sequence and structural analysis of Ole e 12 revealed high polymorphism

The variability of IRL sequences rising among plant species as consequence of single nucleotide polymorphisms (SNP) and post-translational modifications [22–24, 26] contribute to the generation of several protein isoforms that display common and/or different functional and regulatory properties implicated in different metabolic pathways [1]. This variability could even be increased, particularly for species as *O. europaea* L., where different changes (micro-heterogeneities) might contribute to generate functional variability. In addition, it might contribute to a differential regulation among isoforms present in the olive pollen from the wide different genetic background (olive cultivars), with a germplasm that includes more than 2,000 cultivars around

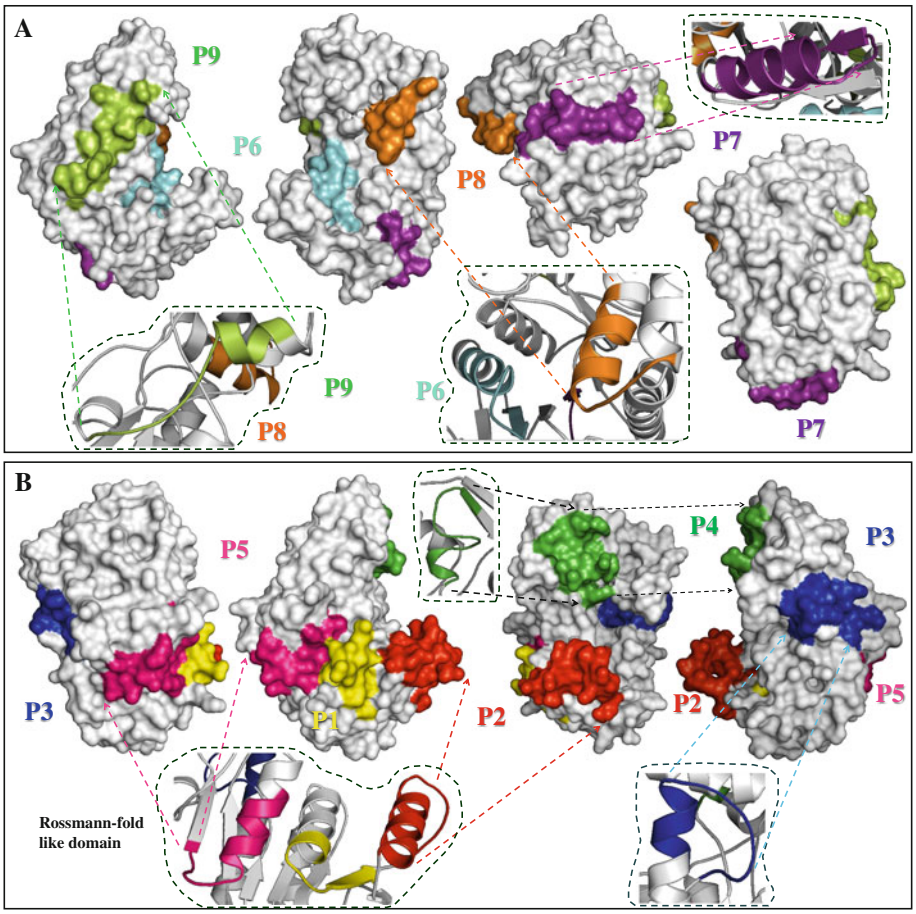

Fig. 7 B- and T-cell epitopes superimposition on the surface of Ole e 12 allergen structure. **a** B-cell epitopes depicted on the Ole e 12 protein surface in four views rotated 90°, respectively. All epitopes are integrated by final part of α-helice and its corresponding flanking loop. B1 (P1), B2 (P2), B3 (P3), B4 (P4) and B5 (P5) epitopes are depicted in *yellow*, *red*, *blue*, *green*, and *pink colors*. A detailed view in *cartoon* representation of every epitope is depicted showing the localization of the five epitopes (B1–B5) in the 2-D structural elements of the protein. **b** T-cell epitopes depicted on the Ole e 12 protein surface in four views rotated 90, 180° and one 90° looking down from the N-terminal side, respectively. All epitopes are integrated by final part of α-helice and its corresponding flanking loop. T1 (P6), T2 (P7), T3 (P8), and T4 (P9) epitopes are depicted in *light blue*, *purple*, *orange* and *light green*. A detailed view in cartoon representation of every epitope is depicted showing the localization of the four epitopes (T1–T4) in the 2-D structural elements of the protein

the world (World Catalogue of Olive Varieties—International Olive Council, http://www.internationaloliveoil.org/store/view/10-world-catalogue-of-olive-varieties). This has been previously demonstrated for the major olive pollen allergen Ole e 1 [22], the pan-allergen Ole e 2 [23, 24], and olive pollen allergen Ole e 11 [26].

Our results showed that this variability is particularly affecting physicochemical parameters analyzed in olive Ole e 12, and pollen IRL sequences from other species. Furthermore, differential post-translational sites and functional motifs, i.e., N-myristoylation, number and type of N-glycosylation sites, and variability of Ser/Thr or Tyr phosphorylation sites implicated in IRLs functional regulation have been observed, which supports the polymorphism as a possible mechanism to generate isoforms variability. Furthermore, micro-heterogeneities in sequences resulting in post-translational modifications and variability may have substantial influence in partners-interacting properties, and likely play the role as an additional factor affecting regulatory pathways and generating multifunctionality [22–24, 26].

These differential features in IRL isoforms might also be reflected in the clustering analysis, where sequences sharing common and/or different functional-regulatory motifs are closely grouped [41].

Ole e 12 belongs to the same clade as PCBER from *F. intermedia*, *P. taeda*, *S. asiatica*, and other pollen allergen proteins from *F. excelsior*. The clustering analysis indicates a close relationship for the members of this clade and the PCBER neighboring species, indicating that Ole e 12 may share close-relationship to PCBER and IFR in comparison to PLR.

Sequence analysis of olive and other IRLs showed numerous conserved residues and several micro-

heterogeneities in different regions of the gene, affecting different areas of the protein, i.e., 2-D structure elements, ligand-binding region (co-factor and substrate). In addition, few studies described so far variable residues in IRLs with regard to catalytic properties (enantio- and region- specificity) [18], and mechanism and substrate specificity [19]. Ole e 12 conservational analysis further confirmed a high variability in the residues surrounding the catalytic pocket (Fig. 4) when compared to other plant species [18, 19]. This feature might be an additional mechanism to generate functional variability, i.e. additional catalytic mechanisms, enantio-specificity conversions and region-specificity [42]. These changes in substrate recognition residues along with the binding of the pockets for several members of the IRLs [40] might partially explain the large number of variable substrates that can be accommodated in the active center [43, 44]. However, glycine rich co-factor NAD(P)H binding motif (*GxxGxxG* fingerprint region), which might play a significant role in the recognition and binding of the substrate but mostly of the co-factor, has been shown to be highly conserved in the PIP-family of proteins (including Ole e 12) [45].

Overall, the 3D structure analysis of Ole e 12 suggests that the protein fold has been strongly conserved among SDR protein family, having significant implications for both functional and regulatory properties. Ole e 12 along with other PIP-family members form two-fold rotationally symmetric homodimers, but the various homodimeric associations are distinct in each case. Ole e 12 homodimers are closely associated, showing a small enzymatic cavity, restricting the entrance to co-factors and substrates. This feature may be a regulatory mechanism in olive enzymatic activity [46–48]. Comparison of Ole e 12 with other plant IRLs revealed a clearly close similarity to PCBER. In our analysis, RMSD showed the lowest values with several PCBERs (Table S4) from angiosperms, i.e. *A. thaliana* (0.031 Å) and gymnosperms, i.e. *Zea mays* (0.043 Å), *P. taeda* (0.072 Å) plant species.

Similar to UDP-galactose 4-epimerase, PCBER and Ole e 12 contains two continue domains, a catalytic domain integrated by a classic Rossmann-fold structure (3α-8β-3α), and a domain UDP-galactose 4-epimerases, with structural similarity to others IRLs [49]. Compared with single domain dehydrogenase enzymes, and having only a co-factor and substrates-binding pocket widely open [50], the additional domain of Ole e 12 would provides a extra structure, which makes the substrate-binding pocket considerably smaller and restricted, dramatically enhancing the substrate specificity. The single-domain enzymes were probably the evolutionary progenitors, and the second domains were likely produced for more specific catalysis on divergent substrates, contributing to the remarkably diverse array of natural products synthesized in plants.

Generation of multiple epitopes by sequence polymorphism may be involved in both specific and wide cross-reactivity among IRL isoforms in plant species

The sequence polymorphism implicated in the generation of epitopes that are recognized by B- and T-cells in IRLs is critical for allergy response, i.e. IgE recognition and binding capacity, as well as IgE-cross-reactivity among cultivars and different plant species. A birch pollen allergen, Bet v 6, has been previously described as IRLs [8], which may also represent a group of cross-reactive allergens in plant foods, i.e. allergen Pyr c 5 from pear fruit. Bet v 6 and Pyr c 5 have been identified as belonging to the PCBER family of plant proteins based on demonstrated catalytic activity [51], and both Pyr c 5 and Bet v 6 allergens had similar IgE binding characteristics. However, the regions on the proteins responsible for these allergenic responses have not been characterized yet.

In the present study, potential IgE and HLA Class I and II binding regions of Ole e 12 have been identified by in silico methods through high quality, reliable computational tools for epitopes predictions [52], using numerous databases of experimentally derived epitopes [53–59]. Numerous micro-heterogeneities of the Ole e 12 sequences in comparison with other species have been located in B-cell epitopes, which demonstrate high variability of Ole e 12 and homologues in plant species. In addition, they have been determined to be the sequence features of IgE binding sites, i.e. Gly and Lys have a key role in the IgE binding allergenic epitopes [60]. Moreover, other studies demonstrated that allergen epitopes comprise a high proportion of hydrophobic amino acids [61]. The most common residues in the B-cell epitopes of Ole e 12 were Ala, Lys, and Thr, and followed by Pro, Val, and Glu. Furthermore, approximately half of the total residues lying in B-cell epitopes were hydrophobic. Besides, IgE binding requires partial accessibility for solvent [62]. Ole e 12 showed that more than half of residues have relative solvent accessibility of $\geq$30 %, indicating a possibility for immunoglobulin binding. Additional solvent accessible charged residues as Arg, Glu or Asp may interact with several amino acid residues facilitating subunits interaction. Moreover, the electrostatic interactions determine orientation of the molecules and stabilize antigen–antibody complexes [63].

In the present study, the epitopes P2, P3, P5 and P7 showed a strong negative potential, although scattered blue regions are observed due to Lys residues (Table S5). Thus, epitopes P2, P3, P5 and P7 with a strong negative potential and low hydrophobicity about 40 % were predicted as a high affinity binder. These four epitopes have been found to be conserved in the allergens Bet v 6 (birch), Dau c 5

(carrot), and Pyr c 5 (pear). Bet v 6 displayed a degree of amino acid sequence identity up to 81 % with PCBER [13, 15], as well as identities of 60 and 51 % with IFR and PLR [13], respectively.

PCBER and IFR may represent a new family of pollen-related food allergens that occur in olive, birch and birch-pollen-related food allergens, but also may be present in many allergenic fruits and vegetables such as apple, peach, orange, lychee fruit, strawberry, persimmon, zucchini, and carrot [51], where protein extracts from several of them exhibited Bet v 6 cross-reactive epitopes against two Bet v 6-specific monoclonal antibodies [64]. Thus, Ole e 12 may be also implicated in pollen-plant food allergy cross-reactivity (at different extension) and sensitization, throughout common-shared epitopes, i.e. P2, P3, P5 and P7, and may also integrate a group of cross-reactive allergens in plant foods together with Dau c 5 and Pyr c 5, and pollen Bet v 6 allergens.

However, characteristic differences between PCBER and IFR appear to be in 10 amino-acid specific insertion that is not present in PCBER or PLR (Fig. 1). This sequence insertion is only partially present in the sequence of the epitope P2 in Ole e 12 allergen. Additionally, P7 epitope displays a differential feature between PLR and IFL and PCBER, since only part of the sequence is present in Ole e 12 or in PCBER (1qyc). These differences may play an important role in the extent of cross-reactivity between pollen-food IRLs related allergens.

Allergy to peanut and other legumes is becoming a significant health issue [65]. Legume-induced allergic reactions can be potentially severe and sometimes fatal [66]. Ara h 8 peanut allergen belongs to the pathogenesis-related protein family (PR)-10 and it is relevant for peanut-allergic patients with birch pollen allergy because of the cross-reactivity to the homologous Bet v 1 allergen [67] and 40 % of sensitization in peanut atopic patients [68]. Our study has showed that Ole e 12 displayed others common epitopes (P3) differentially distributed in allergens such as Ara h 8 (peanut) (fermi.utmb.edu/SDAP). However, implication of Ole e 12 in cross-reactivity to Ara h 8 may not be relevant for sensitization and cross-reactivity with peanut allergens, since previous studies suggesting that more than 99 % of atopic patients studied, exclusive IgE reactivity to Ara h 8 is completely harmless, and its sensitization seems to indicate peanut tolerance instead of one of the allergens responsible for sensitization to peanut [69].

Otherwise, latex proteins allergy constitute a significant occupational hazard for health care worker, since natural rubber latex (NRL) is widely used in the manufacturing of medical devices and a number of domestic and industrial utility articles. The development of clinical latex allergy in individuals involves a series of interactions between T- and B-lymphocytes. Several proteins are considered to be responsible for this latex allergy [70], and many of these proteins appeared to be defense-related proteins involved in plant defense mechanisms [71].

We have identified the epitope P5 in Ole e 12 as commonly shared with Hev b 14 allergen of rubber, a Chitinase and one of the several proteins responsible of latex allergy. Besides, the epitope P7 is another part of the Ole e 12 allergen sequence commonly shared with the major latex Hev b 1 allergen or rubber elongation factor, with a high prevalence among the population [up to 80 % (67 ± 81 %)] of Spine bifida patients sensitized to latex [72]. It has been previously identified and demonstrated that two immunodominant IgE epitope regions near to the C-terminal protein domein (residue 90–108) for Hev b 1 [73] have relevance in triggering allergy reactions. This region is one of the commonly shared with Ole e 12 (epitope P7). Overall, the epitopes P5 and P7 of Ole e 12 may represent two molecular features implicated in cross-reactivity to latex [74], as well as a mechanism for latex sensitization to Ole e 12 atopic patients and vice versa. In addition, P5 and P7, and in some extent epitope P4 (commonly-shared with Hev b 7 rubber allergen) may be responsible of cross-reactions and implicated in latex-fruit syndrome, since patients who are sensitive to natural rubber products often cross-react to various fruits such as banana, avocado, fig, and passion fruit [75].

Furthermore, the epitope P7 of Ole e 12 is also present in Amb a 1, the major allergen of short ragweed (*Ambrosia artemisiifolia*) pollen [76], the single most problematic seasonal allergen in North America and Europe, affecting up to 36 million individuals. Epitope 7 of Ole e 12 shared a high percentage of the residues (67 %) with Amb a 1. This may also represent a molecular link for cross-reactivity of pollen allergens like Ole e 12, Amb a 1 and others pollen allergens sharing epitopes like P7 and implicated in pollen cross-reactions [77].

The systematic design of peptides representing antibody binding regions of allergens and evaluation of epitopes recognized by IgE for other allergens, such as various pollens, seeds/nuts and fruits is very valuable in diagnosis and allergy immunotherapy, but further investigations will be required to evaluate the clinical significance of the new allergen Ole e 12 as a source of pollen-related food and/or latex allergies, and to highlight whether related IRLs enzyme families are similarly involved in the elicitation of allergy reactions to allergens of different pollen species and fruits.

Ole e 12 exhibits a close structural relationship with PCBER, with a specific catalytic mechanism

Ole e 12 protein may be classified as a member of SDR NADPH-dependent reductase family (http://www.cathdb.

info/), showing a similar polypeptide-chain fold to PCBER based on sequence and structural data, expectedly with the most significant differences in substrate- and cofactor-binding sites. It can be assumed that this may have evolved to accommodate different reactions among IRLs [44, 78–80].

Overall, a frequent feature of several PIP members is the catalysis of the reductive cleavage of a carbon–oxygen bond that belongs to a para-positioned phenol-ring substituent to a phenolic hydroxyl, presumably resulting in the formation of a quinone methide intermediate [34, 81]. Similarly to many other members of SDRs, Ole e 12 does not share this catalytic mechanism (Fig. 6). Instead, Ole e 12 adds a hydronium cation to the electrophilic carbon of a C=O double bond. However, like other members of PIP proteins, the Ole e 12 catalytic reaction may be accompanied by the extrusion to the solvent of a leaving group H_3O^+ cation.

NADPH is the first (co)substrate to bind to the Ole e 12, as it is often the case in SDRs [82]. The nicotinamide ring adopts the anti-conformer consistent with the class-A reductase (donation of the pro-R hydride) activity of the PIP-family enzymes. The orientation of the nicotinamide ring in Ole e 12 may be influenced by interactions of the carboxamide group with the polypeptide-chain backbone, feature that has been previously observed for the SDRs [82]. The Ole e 12 residues that are involved in interactions with the NADP(H) cofactor (Figs. 1, 5, 6) are also highly conserved in the other IRLs, and thus these enzymes can be expected to maintain a cofactor-binding site highly similar to that observed in Ole e 12.

Like other typical SDRs, Ole e 12 has a glycine-rich motif, G*XX*G*XX*G, at the first α-β-α unit (^{11}GGTGYIG^{17}), which is known to participate in binding of pyrophosphate group of NADP through a helical dipole of αA [45]. However, Ole e 12 has another common feature with other members of PIP family, which is the lack of the consensus sequence Ser…TyrXXXLys containing the catalytic triad Ser/Tyr/Lys as in most SDRs. Overall, IRLs have a conserved catalytic lysine (e.g., Lys^{134} for PCBER, and Lys^{138} for PLR), for which the substitution for Alanine led to a complete loss of enzymatic activity, confirming its proposed essential role [18, 40]. Ole e 12 exhibits a Lys^{133} in an equivalent position to PCBER and PLR. However there was no serine and tyrosine in the other two equivalent positions, suggesting that Ole e 12 may have a different catalytic mechanism. Alternatively, and given its position in the primary sequence and Ole e 12 structural catalytic pocket, Tyr^{158} and Ser^{159} may play an important role in the development of the catalytic reaction. Overall, residues seldom undergo electron transfer in enzymes catalyzing red-ox processes; and this task is usually undertaken by cofactors. Somewhat more frequently, residues act as intermediates through which electrons pass when transferring between red-ox centres by means of quantum tunneling.

A corresponding Ser and Tyr residues has been found to play an essential role for epimerase and carbonyl reductase activities [39, 83]. Ole e 12 Tyr^{158} and Ser^{159} cannot be a remnant of the catalytic triad of SDRs, whereas Lys^{133} is a triad member that has been conserved [82]. These observations suggest that the Lys residue has a generic function as a proton donor in Ole e 12 and in general for both SDRs and PIP enzymes.

Accordingly to the catalytic mechanisms that have been described for SDRs, this lysine residue has often been assigned an important role in the general acid–base catalytic mechanism. Lys^{133} in Ole e 12 most likely acts as a proton donor transferring an H^+ and e^- to a water molecule, with the concomitant extrusion of the cation hydronium to the solvent.

According to the catalytic model proposed in this study, Ole e 12 has a catalytic tetrad (Lys^{133}-Tyr^{158}-Ser^{159}-Asn^{153}) previously found in other enzymes [50], where an additional Asn153 may stabilize the water molecule implicated in the electron-transfer bridge between Tyr^{158} and the substrate (Fig. 6). Tyr^{158} and specifically Ser^{159} may have an important role in defining the substrate stereo-specificity in Ole e 12, since this residue is only present in olive Ole e 12 (Fig. 1).

Polyphenolic compounds, particularly flavonoids, represent one of the most numerous, structurally diverse, and widely distributed groups of substances in the plant kingdom [84]. The biggest obstacle to increase diversity of new flavonoids activities and applications is preserving the functionality of the enzymes involved in their biosynthetic pathways. In addition, modifications such as glycosylation, methylation and prenylation are necessary to preserve the stability, the cellular distribution, the bioactivity and the bioavailability of these enzymes. However, those modifications can also be detrimental for their biological activities [85]. Parallelly, the bottleneck in the in vivo approach lies in streamlining the intrinsic pathways in order to maximize yield. Indeed, the yield of flavonoids produced by recombinant microorganisms can be increased by means of genetic and process engineering modifying enzymes and corresponding genes [86].

However, application of computational techniques for predicting optimal protein functionality throughout structural features characterization [87], i.e. as result of the present study for Ole e 12, provide a better understanding of the enzymatic functional-regulation, with a focus in the molecular substrate-binding pocket environment, and key residues of the active center for cofactor and substrate recognition and interaction (mainly in specialized secondary metabolism of the phenylpropanoid and isoflavonoids

pathways). Moreover, characterization of catalytic mechanisms of enzymes such as Ole e 12, novel pathways, retrosynthesis of metabolic compounds and mining of 'omics' data [88–90] represent promising approaches to overcome the limitations listed above. Thus, a better alignment of the data and communication between genomics, proteomics, metabolomics, etc., will further assist in engineering biosynthetic tailoring enzyme-catalyzed reactions and assembling putative enzymatic mechanisms from different pathways/organisms [91]. Furthermore, the combination of multiple tailoring enzymes and generation of diversified libraries of new molecules will provide high throughput screening to identify active molecules before scaling-up the reactions and isolate the individual molecules for characterization [92, 93]. This approach would significantly increase the variety of chemical structures that could be tested as pharmaceutical candidates for a broad number of diseases.

The characterization of biosynthetic pathways can increase our understanding of the array of enzymes required for synthesis and diversification of natural and non-natural compounds. Combination between metabolic engineering and in vitro bio-catalysis can also facilitate the generation of non-natural analogs of plant natural products. Non-natural analogs of flavonoids using precursors with specific functional groups can further be diversified using tailoring enzymes. Moreover, combination of more than one tailored enzyme can also improve their physicochemical properties, as well as their biological activity [94].

In summary, the main remaining issue to address is the structural characterization of enzymes and identification of their functions, the production and isolation of active enzymes, as well as the identification and viability of co-factors needed for their catalytic activity. Such knowledge might allow successfully modifying flavonoids content and composition. Although there has been significant progress in metabolic engineering of the flavonoids biosynthesis, the process remains a challenge. Better understandings of the proteins features and processes described above by using computational approaches are necessary for a more precise control of modifications in the enzymes for flavonoids compounds modifications, which will allow searching for new activities and applications, and essential to allow engineering other pathways that are less well understood.

Conclusions

Computer-aided molecular design has been extensively proved as an fundamental methodological approach for modeling and deriving specific proteins structural features. Moreover, it has been successfully used to analyze molecular mechanisms and the functional conservation of proteins within the same family, since knowledge of the spatial 3D conformation and/or domain organization(s) of proteins is a prerequisite to better understand protein functionality and regulation.

Rational studies of structural and functional features of enzymes involved in the flavonoids biosynthetic pathways provide a critical insight into the diverse functions and regulation of the SDR complex family in various organisms such as vascular plants. Those enzymes display very distinct substrate versatilities and catalytic mechanisms. The detailed structural knowledge (residues involved in the catalytic pocket structure and environment, cofactor and substrate recognition and interaction) provided by this study might allow targeting and engineering specific enzyme(s) with novel activity and specificity for producing flavonoids derivatives with valuable new pharmacological activities. Furthermore, it can facilitate metabolic engineering to prevent plant disease and improve crop plants as sources of food and medicinal compounds.

Outcomes of the current work provide a comprehensive understanding of (1) the structure of Ole e 12 protein, characterization and its particular close-relationship with PCBER; (2) the structural similarities and differences of Ole e 12 to others IRLs; (3) the implications of the polymorphism generating IRLs functional variability and differential regulations; (4) the molecular environment of the substrate-binding pocket of the Ole e 12 for cofactor and substrate recognition and interaction. This is markedly important in specialized (secondary) metabolism where even small variations in the sequence can lead to profound alterations in the specificities of the mechanism and binding properties for various substrates; (5) a specific modeling of the catalytic mechanism for Ole e 12, which is centered in a conserved catalytic tetrad (Asn-Ser-Tyr-Lys); (6) the structural and molecular basis for engineering biosynthetic tailoring enzyme reactions for generating new active compounds; and (7) the implications of the molecular variability of epitopes, which may significantly contribute in designing strategies for allergy symptoms amelioration by developing diagnosis and therapeutic tools.

Acknowledgments This study was supported by the following European Regional Development Fund co-financed Grants: MCINN BFU 2004-00601/BFI, BFU 2008-00629, BFU2011-22779, CICE (Junta de Andalucía) P2010-CVI15767, P2010-AGR6274 and P2011-CVI-7487, and by the coordinated project Spain/Germany MEC HA2004-0094. JCJ-L thanks Spanish CSIC and the European Marie Curie research program (FP7-PEOPLE-2011-IOF) for his I3P-BPD-CSIC and PIOF-GA-2011-301550 Grants, respectively.

Glossary

Å	Armstrong
ANOLEA	Atomic non-local environment assessment
ASA	Absolute surface area
Cα	Carbon alpha
CATH	Protein structure classification
CDART	Conserved domain architecture retrieval tool
CDD	Conserved domain database
ExPASy	Expert protein analysis system
FASTA	Fast alignment
FFT	Fast fourier transform
G-factor	Goodness factor
GETAREA	Solvent accessible surface area or solvation energy
GO terms	The gene ontology project
GRAVY	Grand average of hydropathicity
GROMOS96	Force field for molecular dynamics simulation
HLA	Human leukocyte antigen
IC50	Inhibitory concentration
IFR	Isoflavone reductase
IRLs	Isoflavone reductase-like proteins
Jscore	Jury score
MHC	Major histocompatibility complex
NCBI	National center for biotechnology information
NJ	Neighbor-joining
NMR	Nuclear magnetic resonance
PB	Electrostatic Poisson-Boltzmann
PCBER	Phenylcoumaran benzylic ether reductase
PDB	Protein data bank
PIP-Family	PCBER, IFR and PRL family
PIRSF	Family classification system at the protein information resource
PLR	Pinoresinol-lariciresinol reductase
ProSa	Protein structure analysis
PROSITE	Database of protein domains, families and functional sites
PROCHECK	Protein structure validation server
QMEAN	Protein model quality estimation server
SASA	Solvent accessible surface areas
SDR	Short-chain dehydrogenase/reductase
SDAP	Structural database of allergenic proteins
SMART	Simple modular architecture research tool
RMSD	Root mean square deviation

References

1. Moummou H, Kallberg Y, Tonfack LB, Persson B, van der Rest B (2012) BMC Plant Biol 12:219
2. Banerjee S, Li Y, Wang Z, Sarkar FH (2008) Cancer Lett 269:226–242
3. Lewis NG, Davin LB (1999) In: Barton DHR, Nakanishi K, Meth-Cohn O (eds) Comprehensive natural products chemistry. Elsevier, London
4. Dixon RA (1999) In: Sankawa U (ed) Comprehensive natural products chemistry. Elsevier, Oxford
5. Taylor LP, Grotewold E (2005) Curr Opin Plant Biol 8:317–323
6. Buer CS, Imin N, Djordjevic MA (2010) J Integr Plant Biol 52:98–111
7. van Eldik GJ, Ruiter RK, Colla PH, van Herpen MM, Schrauwen JA, Wullems GJ (1997) Plant Mol Biol 33:923–929
8. Vieths S, Frank E, Scheurer S, Meyer HE, Hrazdina G, Haustein D (1998) Scand J Immunol 47:263–267
9. Lers A, Burd S, Lomaniec E, Droby S, Chalutz E (1998) Plant Mol Biol 36:847–856
10. Gang DR, Kasahara H, Xia ZQ, Mijnsbrugge KV, Bauw G, Boerjan W, Van Montagu M, Davin LB, Lewis NG (1999) J Biol Chem 274:7516–7527
11. Karamloo F, Schmitz N, Scheurer S, Foetisch K, Hoffmann A, Haustein D, Vieths S (1999) J Allergy Clin Immunol 104(5):991–999
12. Suzuki S, Umezawa T (2007) J Wood Sci 53:273–284
13. Dinkova-Kostova AT, Gang DR, Davin LB, Bedger DL, Chu A, Lewis NG (1996) J Biol Chem 271:29473–29482
14. Fujita M, Gang DR, Davin LB, Lewis NG (1999) J Biol Chem 274:618–627
15. Paiva NL, Edwards R, Sun Y, Hrazdina G, Dixon RA (1991) Plant Mol Biol 17:653–667
16. Tiemann K, Inzé D, Van Montagu M, Barz W (1991) Eur J Biochem 200:751–757
17. Paiva NL, Sun Y, Dixon RA, Van Etten HD, Hrazdina G (1994) Arch Biochem Biophys 312:501–510
18. Wang X, He X, Lin J, Shao H, Chang Z, Dixon RA (2006) J Mol Biol 358:1341–1352
19. Min T, Kasahara H, Bedgar DL, Youn B, Lawrence PK, Gang DR, Halls SC, Park H, Hilsenbeck JL, Davin LB, Lewis NG, Kang C (2003) J Biol Chem 278:50714–50723
20. Will-Karp M, Santeliz J, Karp CL (2001) Nat Rev Immunol 1:69–75
21. Blanca M, Boulton P, Brostoff J, González-Reguera I (1983) Clin Allergy 13:473–478
22. Hamman-Khalifa A, Castro AJ, Jimenez-Lopez JC, Rodríguez-García MI, Alché JD (2008) BMC Plant Biol 8:10
23. Jimenez-Lopez JC, Morales S, Castro AJ, Volkmann D, Rodríguez-García MI, Alché JD (2012) PLoS ONE 7(2):e30878
24. Jimenez-Lopez JC, Rodríguez-García MI, Alché JD (2013) PLoS ONE 8(10):e76066. doi:10.1371/journal.pone.0076066
25. Salamanca G, Rodriguez R, Quiralte J, Moreno C, Pascual CY, Barber D, Villalba M (2010) FEBS J 277:2729–2739
26. Jimenez-Lopez JC, Kotchoni SO, Rodríguez-García MI, Alché JD (2012) J Mol Model 18:4965–4984
27. Wu S, Zhang Y (2010) Structure 18:858–867
28. Kozakov D, Hall DR, Beglov D, Brenke R, Comeau SR, Shen Y, Li K, Zheng J, Vakili P, Paschalidis IC, Vajda S (2010) Proteins 78:3124–3130
29. Kyte J, Doolittle RF (1982) J Mol Biol 157:105–132
30. Welling GW, Weijer WJ, van der Zee R, Welling-Wester S (1985) FEBS Lett 188:215–218
31. Parker JMR, Guo D, Hodges RS (1986) Biochemistry 25:5425–5431
32. Hopp TP, Woods KR (1981) Proc Natl Acad Sci USA 78:3824–3828
33. Esteve C, Montealegre C, Marina ML, Garcia MC (2012) Talanta 92:1–14
34. Louie GV, Baiga TJ, Bowman ME, Koeduka T, Taylor JH et al (2007) PLoS ONE 2:e993

35. Levin I, Schwarzenbacher R, McMullan D, Abdubek P, Ambing E et al (2004) Proteins 56:629–633
36. Arcangeli C, Cantale C, Galeffi P, Rosato V (2008) J Struct Biol 164:119–133
37. Bovy A, de Vos R, Kemper M, Schijlen E, Almenar M, Muir S, Collins G, Robinson S, Verhoeyen M, Hughes S, Santos C, van Tunen A (2002) Plant Cell 14:2509–2526
38. Kim SS, Grienenberger E, Lallemand B, Colpitts CC, Kim SY, Souza A, Geoffroy P, Heintz D, Krahn D, Kaiser M, Kombrink E, Heitz T, Suh DY, Legrand M, Douglas CJ (2010) Plant Cell 22:4045–4066
39. Tanaka N, Nonaka T, Nakanishi M, Deyashiki Y, Hara A, Mitsui Y (1996) Structure 4:33–45
40. Shao H, Dixon RA, Wang X (2007) J Mol Biol 369:265–276
41. Shimada N, Sato S, Akashi T, Nakamura Y, Tabata S, Ayabe S-I, Aoki T (2007) DNA Res 14:25–36
42. Gutierrez-Gonzalez JJ, Wu X, Gillman JD, Lee JD, Zhong R, Yu O, Shannon G, Ellersieck M, Nguyen HT, Sleper DA (2010) BMC Plant Biol 10:105
43. Jornvall H, Persson B, Krook M, Atrian S, Gonzalez-Duarte R, Jeffery J, Ghosh D (1995) Biochem 34:6003–6013
44. Oppermann U, Filling C, Hult M, Shafqat N, Wu X, Lindh M, Shafqat J, Nordling E, Kallberg Y, Persson B, Jornvall H (2003) Chem Biol Interact 143–144:247–253
45. Lesk AM (1995) Curr Opin Struct Biol 5:775–783
46. Grimm C, Maser E, Mobus E, Klebe G, Reuter K, Ficner R (2000) J Biol Chem 275:41333–41339
47. Ghosh D, Sawicki M, Pletnev V, Erman M, Ohno S, Nakajin S, Duax WL (2001) J Biol Chem 276:18457–18463
48. Zubieta C, Ross JR, Koscheski P, Yang Y, Pichersky E, Noel JP (2003) Plant Cell 15:1704–1716
49. Sakuraba H, Kawai T, Yoneda K, Ohshima T (2011) Arch Biochem Biophys 512(2):126–134
50. Kavanagh KL, Jörnvall H, Persson B, Oppermann U (2008) Cell Mol Life Sci 65(24):3895–3906
51. Karamloo F, Wangorsch A, Kasahara H, Davin LB, Haustein D, Lewis NG, Vieths S (2001) Eur J Biochem 268:5310–5320
52. Wang P, Sidney J, Dow C, Mothe B, Sette A, Peters B (2008) PLoS Comput Biol 4:e1000048
53. Dall'Antonia F, Gieras A, Devanaboyina SC, Valenta R, Keller W (2011) J Allergy Clin Immunol 128(4):872–879
54. García-Casado G, Pacios LF, Díaz-Perales A, Sánchez-Monge R, Lombardero M, García-Selles FJ, Polo F, Barber D, Salcedo G (2003) J Allergy Clin Immunol 112(3):599–605
55. Ball T, Fuchs T, Sperr WR, Valent P, Vangelista L, Kraft D, Valenta R (1999) FASEB J 13:1277–1290
56. Midoro-Horiuti T, Schein CH, Mathura V, Braun W, Czerwinski EW, Togawa A, Kondo Y, Oka T, Watanabe M, Goldbluma RM (2006) Mol Immunol 43(6):509–518
57. Jimenez-Lopez JC, Gachomo WE, Ariyo O, Baba-Moussa L, Kotchoni SO (2012) Mol Biol Rep 39(1):123–130
58. Radauer C, Willerroiderw M, Fuchs H, Hoffmann K, Thalhamerw J, Ferreira F, Scheiner O, Breiteneder H (2006) Clin Exp Allergy 36:920–929
59. Jahn-Schmid B, Kelemen P, Himly M, Bohle B, Fischer G, Ferreira F, Ebner C (2002) J Immunol 169:6005–6011
60. Oezguen N, Zhou B, Negi SS, Ivanciuc O, Schein CH, Labesse G, Braun W (2008) Mol Immunol 45:3740–3747
61. Wolff N, Yannai S, Karin N, Levy Y, Reifen R, Dalal I, Cogan U (2004) J Allergy Clin Immunol 114:1151–1158
62. Brinda KV, Vishveshwara S (2005) BMC Bioinformatics 6:296
63. Sinha N, Mohan S, Lipschultz CA, Smith-Gill SJ (2002) Biophys J 83:2946–2968
64. Karamloo F, Schmitz N, Scheurer S, Foetisch K, Hoffmann A, Haustein D, Vieths S (1999) J Allergy Clin Immunol 104:991–999
65. Custovic A, Nicolaou N (2011) J Allergy Clin Immunol 127:631–632
66. Bock SA, Munoz-Furlong A, Sampson HA (2007) J Allergy Clin Immunol 119:1016–1018
67. Mittag D, Akkerdaas J, Ballmer-Weber BK, Vogel L, Wensing M, Becker WM, Koppelman SJ, Knulst AC, Helbling A, Hefle SL, Van Ree R, Vieths S (2004) J Allergy Clin Immunol 114:1410–1417
68. Nicolaou N, Custovic A (2011) Curr Opin Allergy Clin Immunol 11:222–228
69. Asarnoj A, Nilsson C, Lidholm J, Glaumann S, Ostblom E, Hedlin G, van Hage M, Lilja G, Wickman M (2012) J Allergy Clin Immunol 130(2):468–472
70. Levy DA, Charpin D, Pecquet C, Leynadier F, Vervtoet D (1992) Allergy 47:579–587
71. Düring K (1993) Plant Mol Biol 23:209–214
72. Chen Z, Cremer R, Posch A, Raulf-Heimsoth M, Rihs HP, Baur X (1997) J Allergy Clin Immunol 100:684–693
73. Chen Z, van Kampen V, Raulf-Heimsoth M, Baur X (1996) Clin Exp Allergy 26:406–415
74. Schuler S, Ferrari G, Schmid-Grendelmeier P, Harr T (2013) Clin Transl Allergy 3(1):11
75. Wagner S, Breiteneder H (2002) Biochem Soc Trans 30(6):935–940
76. Wopfner N, Gadermaier G, Egger M, Asero R, Ebner C, Jahn-Schmid B, Ferreira F (2005) Int Arch Allergy Immunol 138(4):337–346
77. Jahn-Schmid B, Hauser M, Wopfner N, Briza P, Berger UE, Asero R, Ebner C, Ferreira F, Bohle B (2012) J Immunol 188(3):1559–1567
78. Jornvall H, Persson M, Jeffery J (1981) Proc Natl Acad SciUSA 78:4226–4230
79. Hoffmann F, Maser E (2007) Drug Metab Rev 39:87–144
80. Kallberg Y, Oppermann U, Jornvall H, Persson B (2002) Protein Sci 11:636–641
81. Mauge C, Granier T, Langlois d'Estaintot B, Gargouri B, Manigand C, Schmitter JM, Chaudière J, Gallois B (2010) J Mol Biol 397:1079–1091
82. Filling C, Berndt KD, Benach J, Knapp S, Prozorovski T, Nordling E, Ladenstein R, Jornvall H, Oppermann U (2002) J Biol Chem 277(28):25677–25684
83. Swanson BA, Frey PA (1993) Biochem 32:13231–13236
84. Tohge T, Watanabe M, Hoefgen R, Fernie AR (2013) Crit Rev Biochem Mol Biol 48(2):123–152
85. Limem I, Guedon E, Hehn A, Bourgaud F, Chekir Ghedira L, Engasser JM, Ghoul M (2008) Process Biochem 43:463–479
86. Liu CJ, Blount JW, Steele CL, Dixon RA (2002) Proc Natl Acad Sci USA 99:14578–14583
87. Julien Racle JO, Hatzimanikatis V (2012) Biotechnol Bioeng 109:2127–2133
88. Brunk E, Neri M, Tavernelli I, Hatzimanikatis V, Rothlisberger U (2012) Biotechnol Bioeng 109:572–582
89. Soh KC, Hatzimanikatis V (2010) Trends Biotechnol 28:501–508
90. Miskovic L, Hatzimanikatis V (2010) Trends Biotechnol 28:391–397
91. Xu P, Bhan N, Koffas MAG (2013) Curr Opin Biotechnol 24(2):291–299
92. Kim MI, Kwon SJ, Dordick JS (2009) Org Lett 11:3806–3809
93. Kwon SJ, Lee M-Y, Ku B, Sherman DH, Dordick JS (2007) ACS Chem Biol 2:419–425
94. Bhan N, Xu P, Koffas MAG (2013) Curr Opin Biotechnol. doi:10.1016/j.copbio.2013.02.019