Computational Biology and Chemistry 70 (2017) 21–30

Contents lists available at ScienceDirect

Computational Biology and Chemistry

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

Research Article

# In silico locating the immune-reactive segments of *Lepidium draba* peroxidase and designing a less immune-reactive enzyme derivative

Yaser Fattahian[a], Ali Riahi-Madvar[a,*], Reza Mirzaee[b], Gholamreza Asadikaram[c], Mohammad Reza Rahbar[d]

[a] Department of Biotechnology, Institute of Science and High Technology and Environmental Sciences, Graduate University of Advanced Technology, Kerman, Iran
[b] Jiroft University of Medical Sciences, Jiroft, Iran
[c] Endocrinology and Metabolism Research Center, Institute of Basic and Clinical Physiology Sciences, and Department of Biochemistry, School of Medicine, Kerman University of Medical Sciences, Kerman, Iran
[d] Pharmaceutical Sciences Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

ARTICLE INFO

*Article history:*
Received 5 March 2017
Received in revised form 14 June 2017
Accepted 12 July 2017
Available online 17 July 2017

*Keywords:*
*Lepidium draba*
Peroxidase
Bioinformatics
Immune-dominant
Epitope

ABSTRACT

Peroxidases have broad applications in industry, environmental as well as pharmaceutical and diagnosis. Recently applicability of peroxidases in cancer therapy was mentioned. In the present study, a horseradish peroxidase homologue from *Lepidium draba* was subjected to in silico analyzes aiming at identifying and locating immune-reactive regions. A derivative sequence with decreased immunogenicity and increased stability also suggested. The tertiary structure of the enzyme was predicted. The functional and structural importance of residues was annotated as well as the conservatory status of each residue. The immune-dominant regions of protein were predicted with various software. N-terminal 4 residues, NFSHTGL (186–192), PRNGN (210–214), PLVRAYADGTQKFFN (261–275), and last 4 residues in C-terminal were predicted to be the consensus immunogenic segments of *L. draba* peroxidase. The modifications were applied to wild type sequence in order to mitigate its immune-reactiveness. The modifications were based on predicted energetic status of residues and naturally occurred amino acids in each position of the enzyme sequence, extracted from alignment file of 150 homologous peroxidases. The new enzyme derivative is predicted to be less immune-reactive and more stable. Thus the sequence is better suited to therapeutic applications.



## 1. Introduction

Horseradish peroxidase is the most well known example of peroxidases (Veitch, 2004). Although more than two centuries have been passed from the first appearance of this enzyme in scientific literature and extensive researches have been completed on it, studies are ongoing until now with new descriptions and somehow novel insights into the enzyme applications (see the review article in reference (Krainer and Glieder, 2015)).

Peroxidases have vast applications in industry, environmental and health sectors as well as pharmaceutical and diagnosis (Azevedo et al., 2003). Increasing the number of publications containing the explanation of the structure and even novel applications of the enzyme emphasize the new interests in the study of the enzyme.

Recently the application of enzyme in cancer therapy has been addressed in a number of studies (Kim et al., 2004; Huang et al., 2005; Jeong et al., 2010). The induction of apoptosis toward a combination of horseradish peroxidase with other compounds such as indole-3-acetic acid (a plant hormone (Basse et al., 1996)) has been shown in several in vitro studies (Veitch, 2004). The resulting free radicals are involved in IAA/HRP- induced apoptosis via two pathways: death receptor-mediated and mitochondrial apoptotic pathways. Cancer is one of the major mortality causes worldwide. Millions of new cases and tumor-related deaths are counted annually (Stewart and Wild, 2016). Targeted delivery of the toxic agent to the tumor is a problematic issue of antitumor therapy. Since most therapeutic approaches harm healthy tissue and cause unwanted side effects. Two potantial solutions are gene-directed enzyme prodrug therapy (GDEPT) and antibody-directed enzyme prodrug therapy (ADEPT) which allow the selective release

* Corresponding author at: Department of Biotechnology, Institute of Science and High Technology and Environmental Sciences, Graduate University of Advanced Technology, Highway Haft Bagh Alavi, P.O. Box 76315-117, Kerman, Iran.
*E-mail address:* Riahi.ali@gmail.com (A. Riahi-Madvar).

http://dx.doi.org/10.1016/j.compbiolchem.2017.07.003
1476-9271/

of a cytotoxic agent from a non-toxic prodrug at the site of tumor (Bonifert et al., 2016).

Many peroxidase derivatives are introduced formerly for example Mogharrab et al. chemically modified Lysine residues and successfully improved the activity and stability of enzyme (Mogharrab et al., 2007) but in therapeutic studies in which horseradish peroxidase was used, in silico studies and immunoreactiveness of the enzyme were not addressed. Immunogenic response to therapeutic molecules can generate anti-drug antibodies (ADAs), which can be either neutralizing or non-neutralizing. Neutralizing antibodies (NAbs) bind to sites in therapeutic proteins in such a way that they directly impair or repeal the biological functions of therapeutic proteins (Kuriakose et al., 2016). The elucidation of the IgG and IgE binding epitopes in HRP and defining the critical residues within these epitopes provide useful information for the molecular design necessary to alter the protein to broaden the applicability of HRP in therapeutic purposes. Usefulness of these information has been mentioned in previous researches (Mine and Zhang, 2002).

Helpfulness and application of bioinformatics tools in designing new protein drugs and protein engineering as well as defining epitopes or immunoreactive regions were confirmed in numerous studies (Chou, 2004; Lahti et al., 2012; Khalili et al., 2015; Gill et al., 2016; Shahbazi et al., 2016).

Previously we have cloned and characterized a peroxidase isoenzyme from *Lepidium draba* (LDP) (the sequence has been deposited in the GenBank with accession No. AIJ01351) which is more than 87% identical to HRP. In the present study, we introduced the immunogenic determinants of LDP and made alterations to the sequence of our recombinant enzyme using bioinformatics tools in order to mitigate the immunogenic and allergenic potential of the enzyme. We then predicted that these alterations did not affect the tertiary structure of enzyme thus we expect a non-allergenic and non-antigenic enzyme, applicable in therapeutics.

## 2. Methods

### 2.1. Sequences and primary analyses

The amino acid sequence of *L. draba* peroxidase was retrieved from GenBank with accession number AIJ01351, and named hereon as LDP. Peroxidase C1A sequence (UniProt Accession No. P00433-1) from *Armoracia rusticana* (Horseradish) was obtained from UniProtKB (Boutet et al., 2016) and used as reference sequence. The fasta formatted sequences were saved and used in all assessments. Primary analyzes were included the evaluation of physicochemical properties of sequences (using ProtParam (Gasteiger et al., 2005)). The critical residues of reference sequence were compared to LDP one by one (i.e. the annotations correspond to each functional or structural residue as provided by UniProtKB experimentally at protein level).

### 2.2. Building the 3D models

A homology modeling approach was employed with Modeller (Sali and Blundell, 1994) stand-alone software to generate the 3D structure of LDP sequence. The method started from a sequence alignment of the target sequence with five template sequences for which the structures were known. Structure templates for homology modeling were identified by LOMETS (Wu and Zhang, 2007) from the PDB library (Rose et al., 2015) employing threading approach containing multiple threading programs. Each threading program can generate tens of thousands of template alignments. Five best models were selected to serve as templates for Modeller. It is necessary to explain that LOMETS combines seven state-of-the-art programs to find the best templates; the Z-score scale is different in any threading program, which renders the comparison of Z-scores among different threading algorithms based on their absolute values meaningless. We used I-Tasser (Roy et al., 2010) server for finding the best templates and normalizing the LOMETS Z-score. A normalized Z-score is the Z-score of the alignment divided by a program-specific cutoff $Z_0$, which has been calculated based on large-scale threading benchmark tests for finding the best templates. Template with a Z-score greater than $Z_0$ usually implies that the alignment corresponds to a correct fold. A confident alignment has a Z-score $> 1$. An alignment with these homologous templates of known structure was used to directly predict the main chain coordinates by mapping the LDP sequence into the template main chain structure with Modeller.

ProsaWeb server (Wiederstein and Sippl, 2007) was used to estimate the overall quality of all models.

### 2.3. Structural refinement

Initial structures undergo refinement employing Kobamin software (Rodrigues et al., 2012), 3drefine (Bhattacharya et al., 2016), and ModRefiner (Xu and Zhang, 2011) separately; the overall quality of all refined models was evaluated, and the best model was selected for further analyses.

### 2.4. Estimating the evolutionary conservation of amino acids in each position

The evolutionary conservation of amino acid positions in protein sequences was estimated using ConSurf protocol (Ashkenazy et al., 2016) based on the phylogenetic relations between homologous sequences. Briefly, the evolutionary rate is assessed based on the evolutionary relatedness between the protein and its homologues. The software parameters were set as follows:

The amino acid sequence was extracted from the 3D structure; ConSurf carries out a search for close homologous sequences using CSI-BLAST (Biegert and Söding, 2009) against UniRef90 database. The sequences were clustered, and highly similar sequences were removed using CD-HIT (Li and Godzik, 2006). Homologous sequences were aligned using T-COFFEE (Di Tommaso et al., 2011). The multiple sequence alignment is then used to build a phylogenetic tree using the neighbor-joining algorithm as implemented in the Rate4Site program (Pupko et al., 2002). Position-specific conservation scores are computed using the empirical Bayesian method (Mayrose et al., 2004). The number of iterations was set as 5-iteration of CSI-BLAST, with a maximum of 150 homologs to collect and E-value cutoff of 0.0001.

### 2.5. Epitope prediction

The existence of any B-cell epitope or allergen determinants (IgE epitopes) within the sequence of LDP was assessed using multiple software. The overlapping of highly scored regions or residues was considered as potential epitopes.

Conformational B cell epitopes were predicted by:

Ellipro (Ponomarenko et al., 2008) based on the geometrical properties of protein structure.

Discotope (Kringelum et al., 2012); The final scores of software were intended by combining the propensity scores of residues in spatial proximity and the contact numbers.

Pepito (or BEpro) (Sweredoski and Baldi, 2008) and BCEP were also used for prediction of conformational B-cell epitopes. BCEP generated a PDB file with surface residue only. All amino acid residue are replaced by their CB atom, in the case of GLY Cα atom is used.