Journal of Theoretical Biology 365 (2015) 96–103

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

Prediction of β-lactamase and its class by Chou's pseudo-amino acid composition and support vector machine

Ravindra Kumar [1], Abhishikha Srivastava [1], Bandana Kumari, Manish Kumar *

Department of Biophysics, University of Delhi South Campus, New Delhi, India

HIGHLIGHTS

- We developed a two-level support vector machine based prediction system, PredLactamase, for β-lactamase protein and its class prediction.
- At first level PredLactamase identifies β-lactamase from non-β-lactamase and at second level it classifies predicted lactamase into one of the 4 classes.
- Performance of class B is higher among all classes, which indicates that metallo-lactamase is different from remaining three classes.
- We also developed a user-friendly web-server and standalone, which can be downloaded from http://14.139.227.92/mkumar/predlactamase

ARTICLE INFO

Article history:
Received 5 August 2014
Received in revised form
1 October 2014
Accepted 6 October 2014
Available online 22 October 2014

Keywords:
β-Lactamase protein
Antibiotic resistance
Leave-one-out cross-validation

ABSTRACT

β-Lactam class of antibiotics is used as major therapeutic agent against a number of pathogenic microbes. The widespread and indiscriminate use of antibiotics to treat bacterial infection has prompted evolution of several evading mechanisms from the lethal effect of antibiotics. β-Lactamases are endogenously produced enzyme that makes bacteria resistant against β-lactam antibiotics by cleaving the β-lactam ring. On the basis of primary structures, β-lactamase family of enzymes is divided into four classes namely A, B, C and D. Class B are metallo-enzymes while A, C and D does not need any metal in the enzyme catalysis. In the present study we developed a SVM based two level β-lactamases protein prediction method, which differentiate β-lactamases from non-β-lactamases at first level and then classify predicted β-lactamases into different classes at second level. We evaluated performance of different input vectors namely simple amino acid composition, Type-1 and Type-2 Chou's pseudo amino acid compositions. Comparative performances indicated that SVM model trained on Type-1 pseudo amino acid composition has the best performance. At first level we were able to classify β-lactamases from non-β-lactamases with 90.63% accuracy. At second level we found maximum accuracy of 61.82%, 89.09%, 70.91% and 70.91% of class A, class B, class C and class D, respectively. A web-server as well as standalone, PredLactamase, is also developed to make the method available to the scientific community, which can be accessed at http://14.139.227.92/mkumar/predlactamase.

1. Introduction

Antibiotics are chemical compounds used to treat bacterial infections. Constant exposure of antibiotics prompted bacteria to develop resistance against them. There are different ways which bacteria uses to evade antibiotics like (i) inactivation or modification of the antibiotic; (Ho et al., 2000; Bonomo and Gill, 2005; Poirel et al., 2001) (ii) alteration of the drug target site; (Franco et al., 2009; McManus, 1997) and (iii) reduced intracellular antibiotic accumulation by decreasing permeability and/or increasing active efflux of the antibiotic (Jacoby and Archer, 1991; Katrijn and Arthur, 2009; McKeegan et al., 2002). β-Lactam group of antibiotics is the most important and the most widely used group of antibiotics (Davies and Davies, 2010; Fisher et al., 2005). It is being used to treat infections caused by both Gram-negative and Gram-positive pathogens (Demain and Sanchez, 2009; Kong et al., 2010; Nikaido, 2009). β-Lactam antibiotics hinder the bacterial cell wall biosynthesis causing failure in formation of protective barrier essential to maintain the cellular rigidity, resist the internal osmotic pressure and participate in cell division (Nanninga, 1998). Pathogen resistance against the β-lactam antibiotics produces an enzyme known as β-lactamase that cleaves the amide bond of β-lactam ring thus resulting in inactivation of antibiotics (McKeegan et al., 2002; Petrosino et al., 1998; Zervosen et al., 2012). β-Lactamase is a group of highly diversified super-family of

* Correspondence to: Department of Biophysics, University of Delhi South Campus, Benito Juarez Road, New Delhi 110021, India. Tel.: +91 11 24114111.
E-mail address: manish@south.du.ac.in (M. Kumar).
[1] Equal contribution.

http://dx.doi.org/10.1016/j.jtbi.2014.10.008
0022-5193/

enzymes, which is classified into four different classes: A, B, C and D on the basis of their amino acid similarity (Ambler, 1980; Hall and Barlow, 2005). The three classes A, C and D are serine β-lactamases, which employ an active site serine to catalyze hydrolysis, whereas class B β-lactamases are zinc-dependent metallo-enzymes. Issue of drug resistance spreading very fast and is regarded as a therapeutic challenge, as increasingly greater number of resistant infection began to reported. Resistance to β-lactam antibiotics is an especially severe threat because β-lactam antibiotics are effective against a broad spectrum of pathogen and have very low toxicity to humans (Livermore, 1996). Recent advances in genome sequencing technology have provided us with a lot of data about both pathogenic and non-pathogenic microbial genomes, and recent advances in phylogenetic methods have given us practical tools to analyze and infer their functional relevance. Therefore an automated prediction method that uses only the primary amino acid sequence for identification, classification and annotation of β-lactamase can be a great asset in understanding of molecular details of β-lactam resistance. Recently we have tried to predict the family of β-lactamase proteins on the basis of presence of a unique family specific motif called as fingerprint (Srivastava et al., 2014). The motifs were extracted using the MEME–MAST (Bailey et al., 2009). But this method has limited utility since it searches for presence of a conserved motif. Considering the rapidly evolving nature of β-lactamase, we need a more robust method that can work even in absence of sequence similarity. This can be achieved by development of machine learning based prediction method. To the best of our knowledge, at present there was no machine learning based prediction method exists which can discriminate β-lactamase from non-lactamase proteins and subsequently predicts its class also, using only the primary amino acid sequence. Further there is a need to develop a webserver also for the experimental biologist, which can distinguish β-lactamase from non-β-lactamase and also classify the predicted β-lactamase into its different classes with very high accuracy. It will enable us to quick and accurate identification of the functional types of β-lactamase proteins.

As demonstrated by a series of recent publications (Chen et al., 2014; Ding et al., 2014; Guo et al., 2014; Liu et al., 2014a, 2014b; Qiu et al., 2014; Xu et al., 2014a, 2014b) in response to the suggestion by Chou (2011), to develop a really useful statistical predictor for a biological system, we need to consider the following procedures: (i) construct or select a valid benchmark dataset to train and test the predictor; (ii) formulate the biological samples with an effective mathematical expression that can truly reflect their intrinsic correlation with the target to be predicted; (iii) introduce or develop a powerful algorithm (or engine) to operate the prediction; (iv) properly perform cross-validation tests to objectively evaluate the anticipated accuracy of the predictor; (v) establish a user-friendly web-server for the predictor that is accessible to the public. Below, we have described how these guidelines were implemented to develop an efficient predictor of β-lactamase proteins. In this study, first time we report a two-level pseudo amino acid composition (PseAAC) and support vector machine (SVM) based prediction method to differentiate β-lactamase from non-lactamase proteins and assign the predicted β-lactamase to its respective class. At first level the predictor will predict whether the query protein sequence is β-lactamase or non-β-lactamase and if it will turn out to be the former, the second level will predict the class to which it might belong. We also developed a user-friendly web-server and standalone, which can be downloaded from http://14.139.227.92/mkumar/predlactamase. The overall prediction schema is shown in Fig. 1.

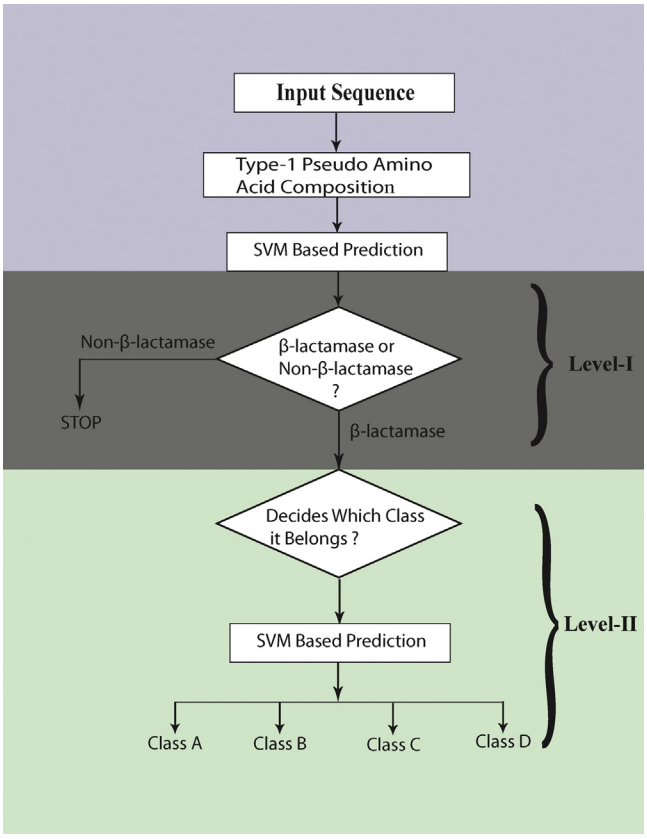

Fig. 1. Prediction schema of PredLactamase web-server. The overall schema is divided into two levels. At first level, it discriminates β-lactamase from non-β-lactamase and at second level it predicts the different classes of β-lactamase.

2. Materials and methods

2.1. Datasets

We retrieved the β-lactamase proteins from Uniprot database (Magrane and UniProt Consortium, 2011). To obtain high quality β-lactamase protein sequences we used 'beta-lactamase' as keyword and manually selected only those sequences which have following attributes: (a) should not annotated as 'by similarity', 'probable' or 'potential'; (b) protein sequences must be a full length sequence meaning not annotated as "fragment". Using above criteria, we retrieved total 2033 sequences of which 979 belongs to class A, 341 to class B, 394 to class C and 319 to class D (Table 1). In order to remove redundancy, CD-HIT (Li and Godzik, 2006) was used with identity threshold 40% to remove those sequences that have $\geq$40% pairwise sequence identity to any

Table 1
Number of sequences in different class of β-lactamase before and after redundancy reduction.

Class	Number of sequences		$Data^{MAIN}$	$Data^{IND}$
	Before redundancy reduction	After redundancy reduction		
A	979	15	11	4
B	341	24	18	6
C	394	21	15	6
D	319	15	11	4
Total	2033	75	55	20

$Data^{MAIN}$ and $Data^{IND}$ are the datasets used for training and independent evaluation, respectively.