Prediction of Michaelis-Menten Constant in Beta-Cellobiosidase’s Reaction with Lactoside as Substrate

Shaomin Yan1 and Guang Wu1,2*

1State Key Laboratory of Non-Food Biomass Enzyme Technology, National Engineering Research Center for Non-Food Biorefinery, Guangxi Key Laboratory of Biorefinery, Guangxi Academy of Sciences, 98 Daling Road, Nanning, Guangxi, 530007, China
2DreamSciTech Consulting, 301, Building 12, Nanyou A-zone, Jiannan Road, Shenzhen, Guangdong, 518054, China

Abstract

The Michaelis-Menten constant, $K_m$, is important to understand the characteristics of enzyme and its relationship with substrates and numerous conditions in biochemical reactions. Although the fast development is evidenced in enzymatic research, the $K_m$ value in each enzyme under various conditions still needs to be measured individually. On the other hand, the modern computational techniques and bioinformatics provide the opportunity to theoretically predict $K_m$ in enzyme with different substrates under various conditions. Cellobiose 1,4-beta-cellobiosidase is an enzyme used in cellulose hydrolysis for bio-fuel industry, and huge efforts are made to enhance its efficiency through searching for new strains of beta-cellobiosidase as well as enzymatic engineering. Therefore, it is considered important to develop methods to predict the $K_m$ value in beta-cellobiosidase’s reaction. In this study, the information of amino acid properties in beta-cellobiosidase, pH and temperature in reaction, and lactoside as substrate were chosen as predictors to predict the $K_m$ values by feedforward backpropagation neural networks, and the delete-1 jackknife was used to validate the predictive model. The results show that 11 of 25 scanned amino acid properties could act as predictors, and that the amino-acid distribution probability appeared the best predictor. The two-layer structure of neural network configuration was sufficient for initial scanning. In consistent with previous studies, the $K_m$ value of enzymatic reactions was predictable using enzyme sequence information and reaction conditions with neural network models.

Keywords: Beta-cellobiosidase; Lactoside; $K_m$ value; Prediction

Introduction

In biochemical reactions, an important measure related to enzymatic function is the Michaelis-Menten constant, $K_m$, because other measures such as pH, temperature, substrate concentrations, are mainly related to reactive conditions. Thus, the $K_m$ value is important to understand the characteristics of enzyme and its relationship with substrates and numerous conditions in biochemical reactions [1,2].

Actually, $K_m$ is the only parameter, from which the enzymatic kinetics as well as stimulators and inhibitors to enzymatic reactions can be formulated [3,4]. Therefore, it is important not only from a practical viewpoint but also from a theoretical viewpoint. However, the measurement of $K_m$ is performed case by case [5], and the measured value is difficult to extrapolate to the enzymes under same category. Yet, $K_m$ is also important to describe the absorption process, i.e. active absorption process [6], for which the $K_m$ value is measured individually too. Technically, $K_m$ is directly related to the affinity of enzyme to a certain substrate.

Without costly and time-consuming measurement, no $K_m$ values are available for newly found and designed enzymes. Therefore, it is necessary to develop methods to predict $K_m$ based on simple information for each enzyme before conducting costly experiments. Along this research line, several studies have been carried out very recently [7-11]; however, more studies are needed in order to systematically approach this issue from various angles.

Cellobiose 1,4-beta-cellobiosidase (EC 3.2.1.91) hydrolyzes 1,4-beta-D-glucosidic linkages in cellulose and cellobiose, and then releases cellobiose from the non-reducing ends of the chains. Recently, a new interest was directed to cellobiosidase because of its potential role in bio-fuel industry, meanwhile lactoside is a major substrate in biochemical reaction of beta-cellobiosidase. In this study, the information of amino acid properties in beta-cellobiosidase, pH and temperature in reaction, and lactoside as substrate were chosen as predictors to predict the $K_m$ values by means of neural network in order to develop the predictive model.

Materials and Methods

Data

The $K_m$ values related to cellulose 1,4-beta-cellobiosidases (EC 3.2.1.91) with lactoside as substrate were found in the Comprehensive Enzyme Information System BREnda [13]. Up to May 2011, 5 beta-cellobiosidases had their sequence information under the category of $K_m$ value as functional parameter, of which beta-cellobiosidases P62694 and Q8J0K6 were documented with their mutants [14-18]. Still, each cellulose 1,4-beta-cellobiosidase could have different $K_m$ values regarding different catalytic conditions, such as pH, temperature and substrate [7-12]. In total, this databank provided 38 matched sequences and $K_m$ values of beta-cellobiosidases (Supplementary Data). The amino-acid sequences of beta-cellobiosidases were obtained from the UniProt [19].

Predictors

The information of amino acid properties was mainly obtained from AAindex [20], which contains 540-plus amino-acid properties
with redundancy [21,22], so not all documented amino acid properties were used in this study, but the ones screened in the previous studies [7-12]. Those amino acid properties included amino acid charge, hydrophilicity or hydrophobicity, size and functional groups [23], such as the spatial properties [24,25], the hydrophobic properties [26-28], the electronic properties [29], and the secondary structure predictions [30]. Each of those amino acid properties had numerically constant value for a type of amino acid Supplementary Data; therefore they were not sensible to amino aid composition, location in enzyme, etc.

The predictors of pH and temperature were measured values in database [13], and the predictor of substrate, lactoside, was different with respect to its substituent groups. So there were 23 predictors for predicting the $K_m$ values.

### Predictive Model

The previous studies showed that the neural network could be the best model for the prediction [7-12], because the relationship between predictors and $K_m$ value is not readily known, while the neural network can theoretically model either cause-consequence relationship or phenomenological relationship. However, the neural network model used in previous studies appeared very complicated, thus the application of simple neural network model without compromising its predictive ability was the main task, which meant to determine how many layers and how many neurons work better.

The developed predictive model was validated using the delete-1 jackknife validation as used in previous studies [7-12] because it was considered very powerful for this type of studies [33].

### Results and Discussion

Using a 12-1 feedforward backpropagation neural network as an example, Figure 1 showed the predictive model for predicting the $K_m$ value. This predictive model was different from current models in bioinformatic studies that included as many predictors as possible.
Theoretically, a predictive model would be chosen with as fewer predictors as possible [34,35], which was the approach used in this study.

Figure 2 showed the convergence with each predictor. This was an important step during model development, because Figure 2 gave a clear picture of which predictor could not converge, which were predictors I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XVIII, XIX, and XX (Supplementary Data). Nevertheless, the predictors that could not converge would not be useful for the prediction. On the other hand, whether a predictor converged guaranteed whether correct model parameters could be obtained [36,37]. Therefore Figure 2 played a role to select workable predictors.

Figure 3 showed the performance of predictions obtained by using the selected predictors from Figure 2 in terms of $P$ value, which was the statistical difference between predicted and measured $K_m$ values, and $R^2$ value, which was the squared correlation coefficient between predicted and measured $K_m$ values. Actually, Figure 3 played role to determine number of neurons ranged from 1-1 to 20-1 for the first layer in two-layer neural networks. In general, the $P$ value and $R^2$ increased in predictions using 1-1 to 3-1 neurons, and some predictors got the highest values in predictions using 5-1 neurons. Thereafter, the $R^2$ values were stable while the $P$ values changed as the neuron increased. Among 11 predictors, the last one (XXV) that was the amino-acid distribution probability provided better predicting results than others.

Figure 4 showed the performance of predictions obtained from 11 selected predictors in different 3-1 models for fitting and delete-1 jackknife validation. Black bars, $P$ value obtained from Wilcoxon signed rank test; gray bars, squared correlation coefficient in regression.

The predictions of 11 selected predictors were conducted in different 12-1 models for fitting (Figure 5) and validation (Figure 6), to further evaluate the influence of multi-layer models on the prediction performance, and to answer whether a very sophisticated model could improve the predictions. Generally speaking, the $P$ values were higher obtained from fitting and the $R^2$ values higher obtained from validation, but no clear feature could be drawn as the model layers increased.
Currently, considerable data are available for various bioinformatic models, but unfortunately few data are available related to parameters of enzymatic reactions. Therefore a small dataset was used in this study although they were all the data available in literature. This is the pressing point that the methods for the prediction of enzyme function parameters should be developed.

This study advanced our knowledge on the prediction of Km values not only in view of the amino acid properties in enzymes as predictors, but also in view of pH, temperature and substrate in enzymatic reaction as predictors, whereas previous studies included only the amino acid properties in enzymes as predictors, among which the amino-acid distribution probability appeared the best predictor, and the two-layer structure of neural network models with their sequence information and reaction conditions. Eleven of 25 scanned amino acid properties could act as predictors, whereas previous studies included only the amino acid properties as predictors [8,10].

In conclusion, the results demonstrated that the Km value of cellulose 1,4-beta-cellobiosidases could be predicted using the neural network models with their sequence information and reaction conditions. Eleven of 25 scanned amino acid properties could act as the predictors, among which the amino-acid distribution probability appeared the best predictor, and the two-layer structure of neural network configuration was sufficient for initial scanning.

Acknowledgment

This study was partly supported by Guangxi Science Foundation (0991006Z, 0991013, 10-046-06, 2010GXNSFA013003 and 2010GXNSFA013046).

References


Figure 5: Comparison between recorded and mean predicted Km in different 12-1 models for fitting. Black bars, P value obtained from Wilcoxon signed rank test; gray bars, squared correlation coefficient in regression.

Figure 6: Comparison between recorded and mean predicted Km in different 12-1 models for delete-1 jackknife validation. Black bars, P value obtained from Wilcoxon signed rank test; gray bars, squared correlation coefficient in regression.


35. Akaike information criterion
