网络优先数字出版时间:2014-10-24
网络优先数字出版地址:http://www.cnki.net/kcms/doi/10.13656/j.cnki.gxkx.20141024.002.html
广西科学 Guangxi Sciences 2014,21(6):656～663

Quantitative Prediction of Michaelis-Menten Constant for α-Amylase and Its Mutants during an Enzymatic Reaction*

定量预测 α-淀粉酶及其突变体的酶反应米氏常数

李检秀,龙思宇,严少敏**,吴　光,谢能中,黄艳燕,师德强,黄纪民,王何健
LI Jian-xiu, LONG Si-yu, YAN Shao-min, WU Guang, XIE Neng-zhong, HUANG Yan-yan, SHI De-qiang, HUANG Ji-min, WANG He-jian

(广西科学院,非粮生物质酶解国家重点实验室,国家非粮生物质能源工程技术研究中心,广西生物质产业化工程院,广西生物炼制重点实验室,广西南宁 530007)
(State Key Laboratory of Non-food Biomass Enzyme Technology, National Engineering Research Center for Non-food Biorefinery, Guangxi Biomass Industrialization Engineering Institute, Guangxi Key Laboratory of Biorefinery, Guangxi Academy of Sciences, Nanning, Guangxi, 530007, China)

Abstract:【Objective】We attempted to develop models to quantitatively predict the Michaelis-Menten constant *Km* with information about primary structure of α-amylase, which is a crucial enzyme for α-1-4 glucosidic linkages hydrolysis in starch, while *Km* is a very important parameter in enzymatic reactions. **【Methods】**By means of neural network, 535 properties of amino acids were used to quantitatively predict *Km* value of α-amylase Amy7C and its 52 mutants, which were divided into two datasets, 33 used for model training and the rest for model validation. The training and validation were conducted firstly by means of two-layer (20-1) feedforward back-propagation neural network, and then by multi-layer neural network models. **【Results】**Among 535 screened properties of amino acids, 109 properties can work as predictor and the dynamic properties give better results with 3 converged out of 4 in 20-1 neural network model. However, the best predicted results came from the amino acid properties with physicochemical property and second structure, of which nine predictors were conducted by seven multi-layer neural network models. The results showed that the increase in complexity of predictive models did not give too much improvement, indicating that the simpler 20-1 and 20-5-1 models should be the first choice. **【Conclusion】**The Michaelis-Menten constant *Km* of α-amylase can be quantitatively predicted by some amino acid properties through neural network, which paves the way for quantitatively predicting parameters in enzymatic reactions according to the information of primary structure of enzyme.

Key words: amino acid properties, α-amylase, *Km* value, quantitative prediction

摘要:【目的】α-淀粉酶是一种重要淀粉水解酶,而 *Km* 值是酶反应中重要的参数,尝试建立一种利用 α-淀粉酶初级结构定量预测米氏常数 *Km* 值的有效模型。**【方法】**通过神经网络模型,利用 535 种氨基酸属性定量预测 α-淀

收稿日期:2014-04-06
修回日期:2014-06-05

作者简介:李检秀(1985-),女,硕士,助理研究员,主要从事微生物遗传、育种及发酵研究。

*广西自然科学基金重点项目(2013GXNSFDA019007),广西科技创新能力与条件建设计划项目(桂科能 12237022)和广西人才小高地建设专项基金项目资助。

**通讯作者:严少敏(1958-),女,博士,研究员,主要从事生物信息学、计算生物学和定量诊断研究,E-mail:shaominyan@gxas.cn。

粉酶 Amy7C 及其 52 个突变体反应的 *Km* 值，其中 33 个酶用于模型训练，其余的用于模型验证。首先用双层的 20-1 前馈反向传播的神经网络进行预测，然后对多层神经网络模型进行筛选。【**结果**】535 种氨基酸属性中有 109 种属性可以用模型预测，其中动态属性拟合结果较好，4 个动态氨基酸属性中有 3 个属性可以用于模型预测，但拟合结果最好的氨基酸属性分别来自氨基酸理化性质和二级结构。对 9 种拟合和验证结果最好的氨基酸属性进行 7 种多层神经网络模型拟合，结果显示增加模型的复杂度并不能提高预测结果的精准度，表明较为简单的模型，如 20-1 或 20-5-1 是定量预测建模的首选。【**结论**】α-淀粉酶酶解反应的米氏常数 *Km* ，可以利用某些氨基酸属性通过神经网络模型进行定量预测。为今后利用酶的初级结构定量预测酶反应中各参数最适条件提供思路。

关键词：氨基酸属性　α-淀粉酶　*Km* 值　定量预测

中图分类号：Q556^{+}.2　　**文献标识码**：A　　**文章编号**：1005-9164(2014)06-0656-08

【**Research significance**】For substrantially economic development, the so-called green economy requires humans to pay great attention to their surrounding environments and to apply renewable resources to economic development. Under such context, the bio-fuel industry would play a more important role either in our daily life or in economic development, and consequently the enzymes would be crucial for bio-fuel industry. α-Amylase(α-1-4 D-glucan glucanohydrolase), which catalyzes endohydrolysis of α-1-4 glucosidic linkages in starch and any related oligosaccharides to make oligosaccharides and glucose[1], is widely used in modern industries. Recently much attention is paid to its application in bio-fuel industry[2~5]. Many chemicals produced by conventional chemical routs could be obtained from renewable resources by biotechnological processes[6]. Thus, availability of an inexpensive carbohydrate raw material is essential for developing an economical bioconversion process for the production of a desired compound, and various kinds of starch are considered to be ideal resources which need endohydrolysis before the bioconversion process. Because of this great prospective in environment-friendly economic development[7~9], many established and new laboratories are progressively moving their research focus on α-amylase[10]. However, the working conditions and enzymatic characteristics were poorly documented. It would be far more cost-effective to screen newly obtained enzymes based on their structure in order to have a concept on their activity, and then to conduct detailed experiments on selected candidates. In this regard, the prediction of parameters in enzymatic reactions is listed on agenda. 【**Achieved research progress**】 A possible way practiced by some researchers in building such an enzyme structure-function relationship is to correlate enzymatic functions with amino acid properties in the first place, which have been conducted in cellulases[11~15] but not in amylases. 【**Current entry point**】The Michaelis-Menten constant *Km* is a very important parameter in enzymatic reaction, which not only represents an enzymatic function[16~18], but also more exactly represents a quantitatively enzymatic function. Although the *Km* is important and useful, its determination is still conduced through experiments with respect to each particular substrate. The rapid increase of newly designed enzymes leads the determination of *Km* to be far behind the need. Therefore it would be considered important to find out which amino acid property would be useful to predict the *Km* of α-amylase. 【**Critical problem to be solved**】We attempt to develop models to quantitatively predict the *Km* of α-amylases with the information about their primary structure.

1　Materials and methods

1.1　*Km* Data

The *Km* data of α-amylase from *Bacillus subtilis* CN7 (Amy7C) and its 52 mutants were obtained from Dr. Cheng-hua Wang[19].

1.2　Predictors

The AAindex documented more than 540 amino acid properties[7], of which 531 properties were chosen including 39 amino acid compositions, 219 physicochemical properties, 273 second structures (Table 1). They are constants for each type of amino acid, and include amino acid compositions, spatial properties, hydrophobic properties, electronic properties, second structures and so on. On the contrary,

another 4 properties can be classiffied as dynamic properties, because their values for each type of amino acid are different one from another in different enzymes.

Table 2 shows the difference between constant amino acid property QIAN880130, which is an amino acid property related to second structure and describes the weights for coil at the window position of −3 (http://www.genome.jp/aaindex/), and 4 amino acid properties with changing values including amino acid number, current composition, future composition and distribution probability. As can be seen, the property QIAN880130 has a constant value (columns 4 and 5, Table 2) regardless of changed compositions (columns 2 and 3, Table 2). This would intuitively not reflect a different role that an amino acid could play at different position in an enzyme with different neighboring amino acids. To overcome this limitation, we weigh the amino acid properties by multiplying amino acid composition (columns 6 and 7, Table 2). The current composition (columns 8 and 9, Table 2) is calculated by the number of a type of amino acids divided by the total number of amino acids in an enzyme, and the future composition (columns 10 and 11, Table 2) is calculated according to the mutating probability[20] at http://www.nerc-nfb.ac.cn/calculation/fc.htm. Still, the values of amino acid distribution probability are different with different amino acid compositions and each amino acid located in different position, which are calculated according to the equation, $r!/(q_0! \times q_1! \times \cdots \times q_n!) \times r!/(r_1! \times r_2! \times \cdots \times r_n!) \times n^{-r}$, where ! is the factorial, r is the number of a type of amino acid, q is the number of partitions with the same number of amino acids and n is the number of partitions in the protein for a type of amino acid[4], and can be available at http://www.nerc-nfb.ac.cn/calculation/dp.htm.

1.3 Predictive model

As there may be various linear and nonlinear relationships between the amino acid property and *Km* value of α-amylase, we used a 20-1 feed forward backpropagation neural network (Figure 1) to account for these quantitative structure-function relationships[21]. In this model, the first layer contained 20 neurons corresponding to 20 inputs, which could be any piece of properties related to 20 types of amino acids in α-amylase. The second layer contained one neuron corresponding to one output that is *Km* value. The transfer functions were tan-sigmoid and linear for two layers, and log-sigmoid for output. The training algorithm was the resilient backpropagation, which was the fastest algorithm on pattern recognition in MatLab[9]. Actually, there are many variants for such a 20-input and 1-output feedforward back-propagation neural network, therefore we tested and compared with several more complicated and sophisticated neural network models including 20-5-1, 20-10-1, 20-10-5-1, 20-30-5-1, 20-30-10-1, 20-30-10-5-1 and 20-30-20-10-5-1 for 9 selected amino acid properties.

1.4 Training and validation

Of α-amylase Amy7C and its 52 mutants, 33 served as training group to generate the neural network model parameters, weights and biases, and 20 were used as validation group to validate the model with trained weights and biases. The above approaches were applied to each predictor listed in Table 1 in order to compare their predictions statistically.

Table 1 Amino acid properties and their screened results

Group	Number	State	Related properties	Webside available	Number of non-convergent properties	Number of convergent properties
Ⅰ	39	Constant	Composition	http://www.genome.jp/aaindex/	38	1
Ⅱ	219	Constant	Physicochemical property	http://www.genome.jp/aaindex/	176	43
Ⅲ	273	Constant	Second structure	http://www.genome.jp/aaindex/	211	62
Ⅳ	4	Dynamic	Composition or distribution	http://www.nerc-nfb.ac.cn/calculation/fc.htm http://www.nerc-nfb.ac.cn/calculation/dp.htm	1	3

Table 2 Comparison of constant property (QIAN880130) and dynamic properties (Amino acid Number, CC, FC and DP) in Amy7C and its A270T mutant

Amino acid 氨基酸	No.		QIAN880130		QIAN880130×No.		CC(%)		FC(%)		DP	
	Amy7C	A270T	Amy7C	A270T	Amy7C	A270T	Amy7C	A270T	Amy7C	A270T	Amy7C	A270T
A	37	36	−0.19	−0.19	−7.03	−6.84	8.62	8.39	7.28	7.23	0.0185	0.0201
R	18	18	−0.07	−0.07	−1.26	−1.26	4.20	4.20	7.33	7.34	0.0312	0.0312
N	39	39	0.17	0.17	6.63	6.63	9.09	9.09	4.63	4.64	0.0031	0.0031
D	29	29	−0.27	−0.27	−7.83	−7.83	6.76	6.76	4.78	4.77	0.0069	0.0069
C	1	1	0.42	0.42	0.42	0.42	0.23	0.23	2.78	2.78	1.0000	1.0000
E	18	18	−0.29	−0.29	−5.22	−5.22	4.20	4.20	4.21	4.20	0.0389	0.0389
Q	21	21	−0.22	−0.22	−4.62	−4.62	4.90	4.90	2.71	2.71	0.0062	0.0062
G	37	37	0.17	0.17	6.29	6.29	8.62	8.62	6.73	6.71	0.0056	0.0056
H	14	14	0.17	0.17	2.38	2.38	3.26	3.26	4.21	4.21	0.0010	0.0010
I	23	23	−0.34	−0.34	−7.82	−7.82	5.36	5.36	5.14	5.16	0.0112	0.0112
L	23	23	−0.22	−0.22	−5.06	−5.06	5.36	5.36	7.00	7.00	0.0460	0.0460
K	18	18	0	0	0	0	4.20	4.20	4.41	4.42	0.0831	0.0831
M	8	8	−0.53	−0.53	−4.24	−4.24	1.86	1.86	1.42	1.42	0.2243	0.2243
F	13	13	−0.31	−0.31	−4.03	−4.03	3.03	3.03	2.46	2.46	0.0463	0.0463
P	12	12	0.14	0.14	1.68	1.68	2.80	2.80	4.88	4.88	0.1241	0.1241
S	39	39	0.22	0.22	8.58	8.58	9.09	9.09	8.42	8.43	0.0064	0.0064
T	27	28	0.1	0.1	2.7	2.8	6.29	6.53	6.81	6.86	0.0161	0.0500
W	12	12	−0.15	−0.15	−1.8	−1.8	2.80	2.80	0.69	0.69	0.0310	0.0310
Y	19	19	−0.02	−0.02	−0.38	−0.38	4.43	4.43	3.32	3.32	0.0852	0.0852
V	21	21	−0.33	−0.33	−6.93	−6.93	4.90	4.90	6.50	6.48	0.0053	0.0053

QIAN880130 is an amino acid property related to second structure and describes the weights for coil at the window position of −3 (http://www.genome.jp/aaindex/). No. ,Number of amino acids;CC(%),Current composition of amino acids;FC(%),Future composition of amino acids;DP, Distribution probability of amino acids.

1.5 Statistics

For each predictor, one hundred trainings were conducted, and the obtained 100 sets of weights and biases were used to predict *Km* 100 times, and the results were presented as median with interquartile. The linear regression and Wilcoxon Signed Rank Test were used to compare predicted *Km* values with recorded ones, and Kruskal-Wallis One Way Analysis of Variance on Ranks was used to compare the predicted results among groups. $P<0.05$ is considered significant.

2 Results

Among 535 amino acid properties, 109 ones can converge during the fit, including 1 constant composition, 43 physicochemical properties, 67 second structures and 3 dynamic properties(Table 1). Thus these 109 properties can serve as predictors to quantitatively predict the *Km* of amylases, suggesting that the Michaelis-Menten kinetics in enzymatic reaction may be related to them, by which we can build a quantitative relationship. Also, the results in table 1 indicate that the dynamic properties give better consquences with 3 converged out of 4 in two-layer neural network model.

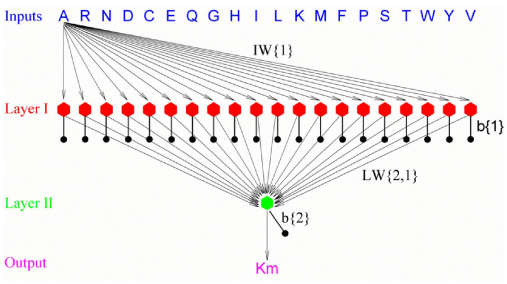

Fig. 1 20-1 feedforward backpropagation neural network to model the relationship between 20 pieces of information on primary structure of amylase

Amino acid propertees are labeled using the symbols of 20 types of amino acids its *Km*. Each hexagon represents a neuron. IW{1} is the input weights, LW{2,1} is the layer weights to the second layer from the first layer. b{1} and b{2} are the biases related to each neuron at the first and second layers.

Figure 2 demonstrates the predicted results for *Km* values of α-amylase and its mutants obtained in training (a) and validation (b) by 109 amino acid properties using 20-1 feedforward backpropagation neural network. We used four indicates to present the results: 1) *R* value is the correlation coefficient in regression between recorded and predicted *Km* values, which reflects the correlation tendency of predicted results; 2) *P* value is obtained from Wilcoxon Signed Rank Test, which reflects the difference

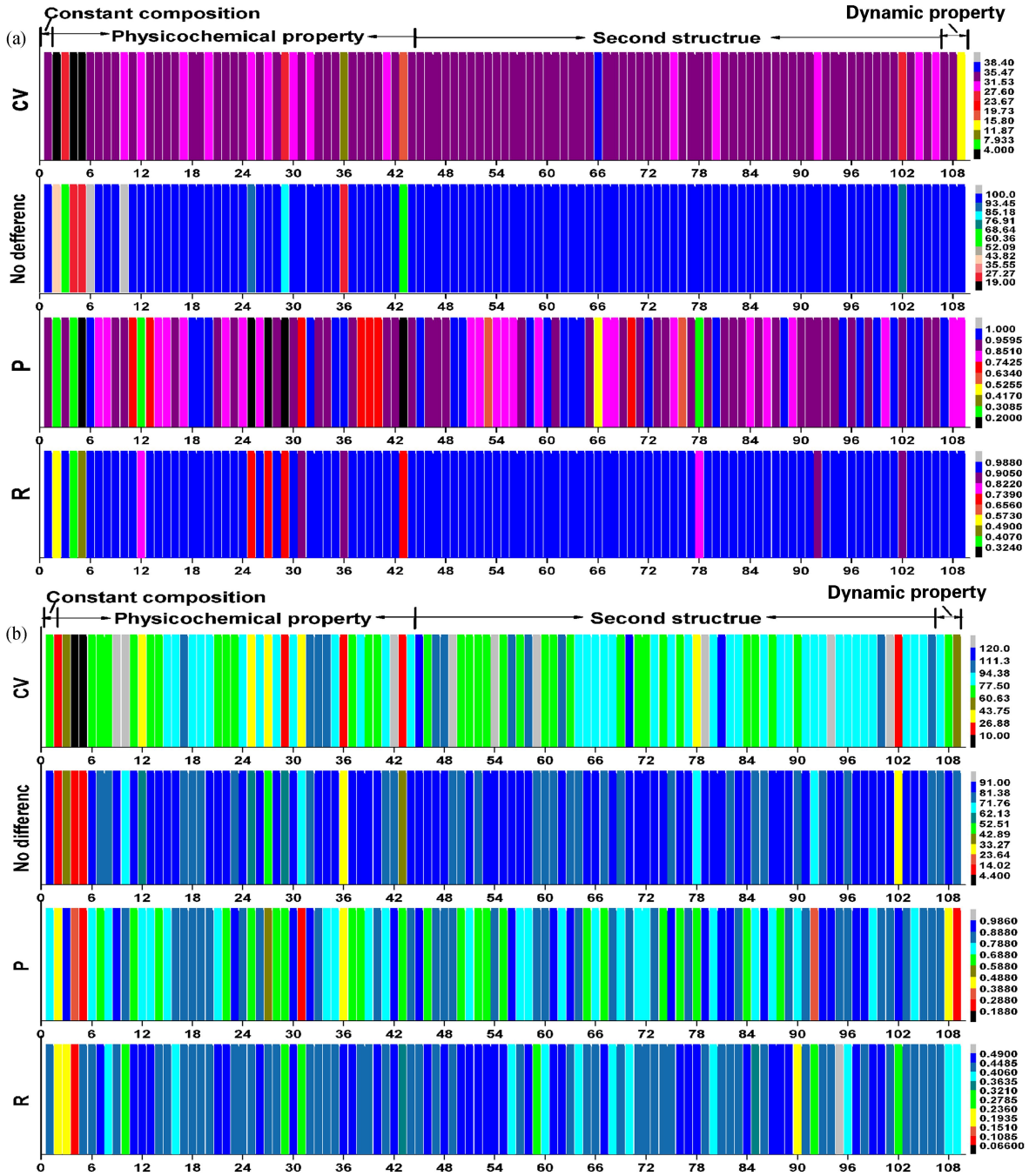

Figure 2 Predicted results for *Km* values of α-amylase and its mutants by 109 amino acid properties using 20-1 feedforward backpropagation neural network

(a) Training; (b) Validation; *R*, Correlation coefficient in regression between recorded and predicted *Km* values; *P*, *P* value from Wilcoxon Signed Rank Test; No difference, Percentage of amylase number to be predicted correctly; CV, Coefficient of variation.

between recorded and predicted *Km* values; 3) No defference is the percentage of amylases whose *Km* value is correctly predicted; 4) CV is the coefficient of variation, which reflects the degree of variation. As can be seen in Figure 2, different amino acid properties give different predicted results. Very good results were obtained from training, for example, the range is 0.325～0.988 for *R* value, 0.133～1.00 for *P* value, 18.18%～96.97% for the correctly predicted amylases and 0.34%～38.35% for the CV. In comparison, the validation results revealed not so good, especially for the *R* value.

Based on the above results, we further test the effect of different neural network models on the prediction of amylase *Km* . Figures 3 and 4 show two types of results: (i) how many layers and neurons work best with nine amino acid properties; and (ii) which amino acid property works best with eight neural network models. As can be seen from Figure 3, the increase in complexity of predictive models increased the *R* value ($P<0.001$) but did not affect the *P* value ($P=0.302$) for the training. On the contrary, the increase in complexity of predictive models did not affect the *R* value ($P=0.104$) but decreased the *P* value ($P<0.001$) for the validation. Figure 4 displays the comporison of predicted results from nine amino acid properties selected, of which significantly statistical diference was found in *P* values of the training ($P=0.046$) but not in *R* values of both training and validation ($P=0.834$ and $P=0.053$) as well as *P* values in validation ($P=0.245$).

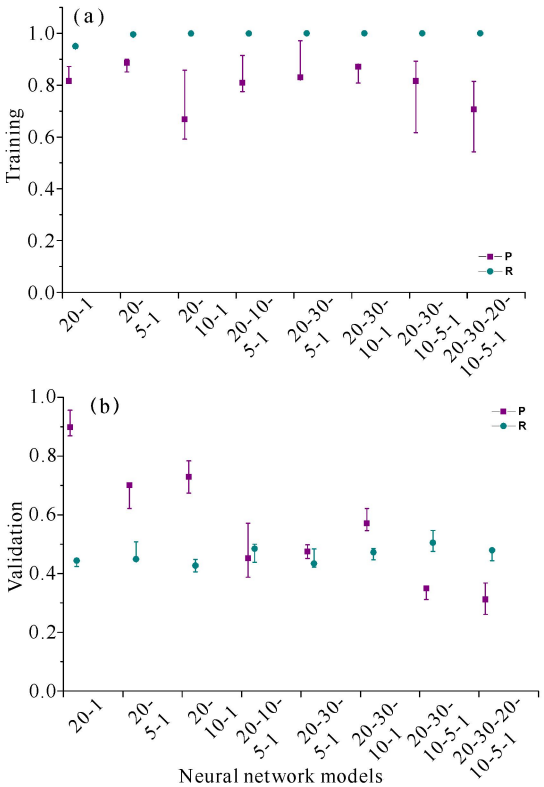

Figure 3 Predicted results for *Km* values of α-amylase and its mutants by eight different neural network models

(a) Training; (b) Validation; *R* , Correlation coefficient in regression between recorded and predicted *Km* values; *P* , *P* value from Wilcoxon Signed Rank Test.

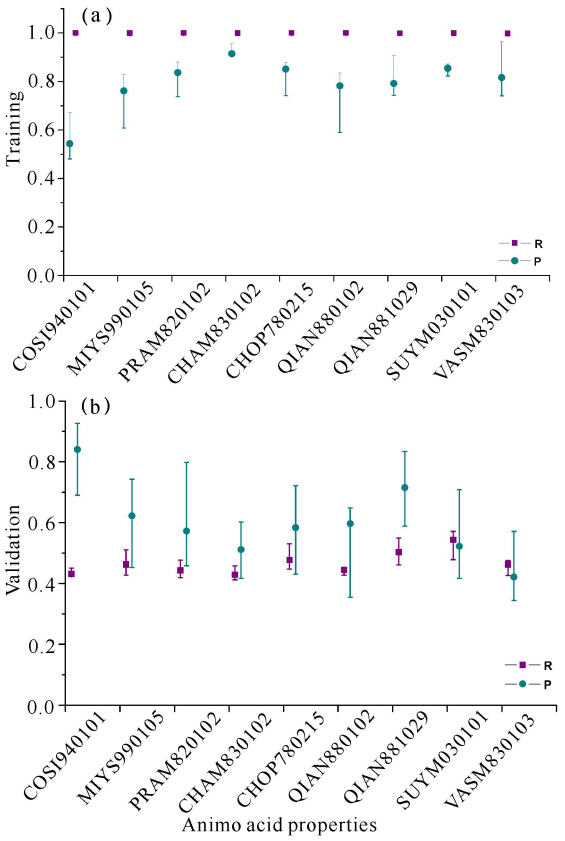

Figure 4 Predicted results for *Km* values of α-amylase and its mutants by nine animo acid properties in multi-layer neural network models

(a) Training; (b) Validation; *R* , Correlation coefficient in regression between recorded and predicted *km* values; *P* , *P* value from Wilcoxon Signed Rank Test.

3 Discussion

Currently, intriguring interests are focusing on thermostability and activity of a Ca-independent α-amylase from *Bacillus subtilis* CN7[22], whose mutants have been engineered by combinatorial coevolving-site saturation mutagenesis[23]. The feedforward backpropagation neural network is powerful model for modeling, by which 535 amino acid properties were used to quantitatively predict *Km* value of α-amylase Amy7C and its 52 mutants. The results show that different predictors give different results and that 109 properties screened can work as predictor. In general, the dynamic properties of amino acids give better results because three out of four can converge in the predictive models. This is reasonable for they are subject to the length of enzyme, position of each amino acid, and represent a dynamic aspect of amino acids in an enzyme[24].

However, the best predicted results came from the amino acid properties with physicochemical property and second structure, indicating that they have more impacts on enzymatic reaction. Moreover, our results show that the increase in complexity of predictive models does not give too much improvement, suggesting that the simpler 20-1 and 20-5-1 models should be the first choice in future to quantitatively predict α-amylase *Km* based on amino acid properties.

For an experimentalist, it would be easier to measure *Km* value than to predict. However, it is only the model that can provide the basis for generalization. Moreover, the model would provide the basis for simulation of catalytic reaction using computer. Actually, the prediction of optimal working condition for enzymes is an understudied area, thus it is important to develop methods along this line of studies. Experimentally and practically, it is important to develop methods to use as simple information as possible to predict the optimal working condition for enzymes. Unlike the modeling in bioinformatics where a considerable data are available, the modeling with parameters of enzymatic reactions always suffers from shortage of data. In fact, the prediction of parameters in enzymatic reactions is an under-studied field mainly due to the lack of data, which should be weighed by the fact that small dataset can reduce the chance that a small statistical difference can appear significant with large dataset. However, such studies should be conducted to catch up with the fast development in biotechnology and bio-fuel industry. Nevertheless, more studies are needed in order to better predict the optimal working conditions in different enzymes.

4 Conclusion

The Michaelis-Menten constant *Km* of α-amylase can be quantitatively predicted by some amino acid properties through neural network, which paves the way for predicting parameters in enzymatic reactions according to the information of primary structure of enzyme, like the amino acid properties.

Acknowledgements: The authors thank Dr. Cheng-hua Wang for providing the *Km* data of α-amylase Amy7C and its mutants.

References:

[1] Goswami A K, Jain K, Paul B, et al. α-and β-Amylases in seed germination[J]. Biologia Plantarum, 1977, 19(6): 469-471.

[2] Dill K A. Theory for the folding and stability of globular proteins[J]. Biochemistry, 1985, 24: 1501-1509.

[3] Dwyer D S. Electronic properties of amino acid side chains: Quantum mechanics calculation of substituent effects[J]. BMC Chem Biol, 2005, 5: 2.

[4] Feller W. An Introduction to Probability Theory and Its Applications[M]. 3rd Edition, Vol. I. NewYork: Wiley, 1968.

[5] Jeng W Y, Wang N C, Lin M H, et al. Structural and functional analysis of three beta-glucosidases from bacterium *Clostridiumcellulovorans*, fungus *Trichoderma reesei* and termite *Neotermes koshunensis* [J]. J Struc Biol, 2011, 173(1): 46-56.

[6] Ragauskas A J, Williams C K, Davison B H, et al. The path forward for biofuels and biomaterials[J]. Science, 2006, 311: 484-489.

[7] Kawashima S, Pokarowski P, Pokarowska M, et al. A Aindex: Amino acid index database, progress report 2008[J]. Nucleic Acids Res, 2008, 36: D202-205.

[8] Kyte J, Doolittle R F. A simple method for displaying the hydropathic character of a protein[J]. J Mol Biol, 1982, 157: 105-132.

[9] MathWorks Inc. MatLab—The language of technical computing (version 6. 1. 0. 450, release 12. 1), 1984-2001[CP]. 2001.

[10] Scheer M, Grote A, Chang A, et al. BRENDA, the enzyme information system in 2011[J]. Nucleic Acids Res, 2011, 39: 670-676.

[11] 严少敏, 师德强, 农浩, 等. 贝塔-葡萄糖苷酶催化反应的最适 pH 值和最适温度的同时预测[J]. 广西科学, 2011, 18(3): 253-260.
Yan S M, Shi D Q, Nong H, et al. Simultaneously predicting pH and temperature optimum in catalytic reaction of beta-glucosidase[J]. Guangxi Sci, 2011, 18(3): 253-260.

[12] Yan S M, Wu G. Prediction of Michaelis-Menten constant of beta-glucosidases using nitrophenyl-beta-D-glucopyranoside as substrate[J]. Protein Pept Lett, 2011, 18: 1053-1057.

[13] Yan S, Wu G. Searching of predictors to predict pH of cellulases[J]. Appl Biochem Biotechnol, 2011, 165: 856-869.

[14] Yan S, Wu G. Prediction of optimal pH and temperature of cellulases using neural network[J]. Protein Pept

Lett,2012,19:29-39.

[15] Yan S,Shi D Q,Nong H,et al. Predicting *Km* values of beta-glucosidases using cellobiose as substrate[J]. Interdiscip Sci:Comput Life Sci,2012,4(1):46-53.

[16] Rokitta D,Fuhr U. Comparison of enzyme kinetic parameters obtained *in vitro* for reactions mediated by human CYP2C enzymes including major CYP2C9 variants[J]. Curr Drug Metab,2010,11(2):153-161.

[17] Dong J Q,Salinger D H,Endres C J,et al. Quantitative prediction of human pharmacokinetics for monoclonal antibodies:Retrospective analysis of monkey as a single species for first-in-human prediction[J]. Clin Pharmacokinet,2011,50(2):131-142.

[18] Liu L,Halladay J S,Shin Y,et al. Significant species difference in amide hydrolysis of GDC-0834,a novel potent and selective Bruton's tyrosine kinase inhibitor [J]. Drug Metab Dispos,2011,39(10):1840-1849.

[19] 王成华.基于α-淀粉酶理性设计的酶分子改造方法的研究[D].南京:南京工业大学,2012.
Wang C H. Enzyme Engineering Methods Developed from Rational Design of Alpha Amylase[D]. Nanjing: Nanjing University of Technology,2012.

[20] Wu G,Yan S. Lecture Notes on Computational Mutation[M]. New York:Nova Science Publishers,2008.

[21] Demuth H,Beale M. Neural Network Toolbox for Use with MatLab—User's Guide[S]. Version 4. 2001.

[22] Wang C,Wang Q,Liao S,et al. Thermal stability and activity improvements of a Ca-independent α-amylase from *Bacillus subtilis* CN7 by C-terminal truncation and hexahistidine-tag fusion[J]. Biotechnol Appl Biochem,2014,61(2):93-100.

[23] Wang C,Huang R,He B,et al. Improving the thermostability of alpha-amylase by combinatorial coevolving-site saturation mutagenesis[J]. BMC Bioinformatics, 2014,13(1):263-269.

[24] Yan S,Wu G. Predictions of enzymatic parameters:a mini-review with focus on enzymes for biofuel[J]. Applied Biochemistry and Biotechnology Part A:Enzyme Engineering and Biotechnology, 2013, 171(3): 590-615.

(责任编辑:陈小玲)