http://www.scirp.org/journal/jbise

J. Biomedical Science and Engineering, 2019, Vol. 12, (No. 8), pp: 414-426

Predictors for Predicting Temperature Optimum in Beta-Glucosidases

Shaomin Yan, Guang Wu

State Key Laboratory of Non-Food Biomass and Enzyme Technology, Guangxi Key Laboratory of Bio-Refinery, Guangxi Biomass Engineering Technology Research Center, National Engineering Research Center for Non-Food Biorefinery, Guangxi Academy of Sciences, Nanning, China

Correspondence to: Guang Wu, Hongguanglishibahao@gxas.cn

Keywords: Beta-Glucosidase, Enzyme, Temperature Optimum, Prediction

Received: July 16, 2019 **Accepted:** August 24, 2019 **Published:** August 27, 2019

ABSTRACT

This is the continuation of our studies on beta-glucosidase, which plays an important role in biological processes and recently strong interests focus on their potential role in biofeul production. In order to develop simple methods to predict the optimal working condition for beta-glucosidase, we used a 20-1 feedforward backpropagation neural network to screen possible predictors to predict the temperature optimum of beta-glucosidase from 25 amino-acid properties related to the primary structure of beta-glucosidases. The results show that the normalized polarizability index and amino-acid distribution probability can predict the temperature optimum of beta-glucosidase, which highlights a cost-effective way to predict various enzymatic parameters of beta-glucosidase.

1. INTRODUCTION

The β-glucosidase (EC 3.2.1.21) plays an important role in biological processes because it cuts the β-bond linkage into glucose molecules [1]. For example, mutations in the gene of lysosomal enzyme acid beta-glucosidase can lead to human metabolic disorder Gaucher disease characterized by deficient activity of the enzyme [2, 3]. β-glucosidase can deglycosylase isoflavones to their aglycone forms, which provides wide applications in food and pharmaceutical industries [4]. Recently, more and more interest on its potential role on biofeul production because cellulose is a linear biopolymer of glucose molecules connected by β-1,4-glycosidic bonds, of which enzymatic hydrolysis requires mixtures of hydrolytic enzymes including endoglucanases, exoglucanases (cellobiohydrolases), and β-glucosidases [5]. Therefore, great efforts have been made to develop renewable biofuel by enzymatically hydrolyzing carbohydrate polymers in biomass to sugars and fermenting them to ethanol [6].

Generally speaking, the optimal working conditions for enzymes are determined through the experi-

mental approaches, which are costly and time-consuming. Nowadays, the experimental speed apparently lags the speed of increase of enzymes in database because in 2002 there were only 789 enzymes documented in the Comprehensive Enzyme Information System BRENDA [7, 8]. However, there are enzymes from 33,721 organisms currently. In this situation, it is easily found that many enzymes have their sequence information but lack their optimal working conditions. Thus it is intriguing to develop methods to predict the optimal working conditions of enzymes based on their primary structure, and recently we have conducted several studies on predicting functional parameters of enzymes using amino acid properties, including pH optimum [9-12], temperature optimum [11-15], Michaelis-Menten constant [16-18] and turnover number [19]. However, more studies are needed in order to get solid conclusions. The aim of this study is to find out the predictors that are useful to predict the temperature optimum of β-glucosidase.

2. MATERIALS AND METHODS

2.1. Data

From the Comprehensive Enzyme Information System BRENDA, 37 β-glucosidases (EC 3.2.1.21) have their sequence information under the category of temperature optimum, of which one β-glucosidase was documented with its mutant [20, 21]. Also, two temperature values are documented in the β-glucosidases B5TWK3 at 22˚C and 37˚C [22] and Q12715 [23] at 65˚C and 70˚C. In total, this databank provides 40 matched sequences and temperature values of β-glucosidases. The amino-acid sequences of β-glucosidases are obtained from the Universal Protein Resource (UniProt) [24].

2.2. Possible Predictors

Table 1 lists the amino acid properties to be scanned, which involve the characteristics of charge, hydrophilicity or hydrophobicity, size and functional groups, and they are crucial for protein structure and protein–protein interactions [25]. Some properties are related to primary structure of enzymes and include the spatial properties [26, 27] listed in rows 2 - 5 in **Table 1**; hydrophobic properties [28-30] listed in rows 6 - 10 in **Table 1**; electronic properties [31] listed in rows 11 - 17 in **Table 1**, and the secondary structure predictions [32] listed in rows 18 - 24 in **Table 1**. All of these properties have a particular number to a certain amino acid in proteins, thus each amino acid has a fixed value, which surely cannot represent different β-glucosidases. Because each β-glucosidase has its own amino-acid composition, we multiply the values listed in **Table 1** by their amino-acid composition for each β-glucosidase.

Based on occupancy of subpopulations and partitions [33], we have developed a measure to calculate amino acid distribution probability according to the following equation:

$$r!/\left(q_0!\times q_1!\times\cdots\times q_n!\right)\times r!/\left(r_1!\times r_2!\times\cdots\times r_n!\right)\times n^{-r}$$

where ! is the factorial function, 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. And its calculation can be available at http://www.gxas.cn/dp.htm. Each type of amino acids has its distribution probability as example shown in **Table 2**. However, the same type of amino acids can have different values in different proteins according to their real distribution pattern along protein sequence [34-38].

2.3. Predictive Model

In order to find out possible predictors to predict the temperature optimum of β-glucosidases, a 20-1 feedforward backpropagation neural network was used as predictive model [39], whose structure is shown in **Figure 1**. In this model, the first layer contains 20 neurons corresponding to 20 inputs (or 20 elements of input in neural network terminology), which can be any measure related to 20 types of amino acids. The second layer contains a single neuron corresponding to the single output, temperature optimum. The transfer functions are tan-sigmoid and linear for two layers. The training algorithm is the resilient backpropagation, which is the fastest algorithm on pattern recognition in MatLab [40].

Table 1. Features of amino acids used as predictors. A, alanine; R, arginine; N, asparagine; D, aspartic acid; C, cysteine; E, glutamic acid; Q, glutamine; G, glycine; H, histidine; I, isoleucine; L, leucine; K, lysine; M, methionine; F, phenylalanine; P, proline; S, serine; T, threonine; W, tryptophan; Y, tyrosine; V, valine. σ_I: Inductive effect scale; $H_M\Delta PH$: Normalized Mulliken population data for the amino-acid side chains in the context of phenol; σ_R: Resonance effect scale; σ_α: Normalized polarizability index; σ_F: Field effect index; A_I: Additional scale; f(i): Frequency of the 1st residue in turn; f(i + 1): Frequency of the 2nd residue in turn; f(i + 2): Frequency of the 3rd residue in turn; f(i + 3): Frequency of the 4th residue in turn.

Amino acid	A	R	N	D	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V
Mass, Dalton	71.09	156.19	114.11	115.09	103.15	129.12	128.14	57.05	137.14	113.16	113.16	128.17	131.19	147.18	97.12	87.08	101.11	186.12	163.18	99.14
Surface Area, Å^2	115	225	160	150	135	190	180	75	195	175	170	200	185	210	145	115	140	255	230	155
Residue Volume, Å^3	88.6	173.4	114.1	111.1	108.5	138.4	143.8	60.1	153.2	166.7	166.7	168.6	162.9	189.9	112.7	89.0	116.1	227.8	193.6	140.0
van der Waals volume, Å^3	67	148	96	91	86	114	109	48	118	124	124	135	124	135	90	73	93	163	141	105
Residue Non-polar Surface Area, Å^2	86	89	42	45	48	69	66	47	129	155	122	164	137	194	124	56	90	236	154	135
Residue Burial, kcal/mol	2.15	2.23	1.05	1.13	1.20	1.73	1.65	1.18	2.45	3.88	3.05	4.10	3.43	3.46	3.10	1.40	2.25	4.11	2.81	3.38
Side Chain Burial, kcal/mol	1.0	1.1	−0.1	−0.1	0.0	0.5	0.5	0.0	1.3	2.7	1.9	2.9	2.3	2.3	1.9	0.2	1.1	2.9	1.6	2.2
Hydropathy index	4.5	4.2	−0.8	−0.9	−3.5	−0.7	−1.6	1.8	−3.9	−3.5	−1.3	2.5	−0.4	−3.2	−3.5	2.8	1.9	4.5	3.8	−3.5
Ranking of amino acid polarities	9	15	16	19	7	18	17	11	10	1	3	20	5	2	13	14	12	6	8	4
pK_a	9.69	9.04	8.80	9.60	10.28	9.67	9.13	9.60	9.17	9.68	9.60	8.95	9.21	9.13	10.60	9.15	9.10	9.39	9.11	9.62
σ_I	0.05	−0.26	−0.14	0.51	−0.01	0.68	−0.10	0.00	−0.01	0.06	0.02	−0.16	0.08	0.04	0.00	−0.03	−0.05	0.06	0.05	0.01
$H_M\Delta PH$	0.05	−0.75	−0.20	1.80	−0.01	1.25	−0.07	0.00	0.21	0.08	0.07	−1.11	−0.04	0.06	0.10	−0.05	−0.03	0.15	0.02	0.09
σ_R	0.00	−0.49	−0.06	1.29	0.01	0.57	0.03	0.00	0.22	0.02	0.05	−0.95	−0.12	0.02	0.10	−0.02	0.02	0.09	−0.03	0.08
σ_α	−0.01	−0.08	−0.04	−0.03	−0.03	−0.04	−0.05	0.00	−0.06	−0.04	−0.04	−0.05	−0.05	−0.08	−0.04	−0.02	−0.03	−0.12	−0.09	−0.03
σ_F	0.05	0.27	−0.56	−1.77	0.06	−1.14	−0.35	0.00	−0.58	0.04	−0.03	0.51	−0.30	−0.45	0.02	−0.38	−0.44	−0.24	−0.42	−0.04
A_I	0.05	0.26	0.24	0.51	0.01	0.68	0.10	0.00	0.01	0.06	0.02	0.16	0.08	0.04	0.00	0.03	0.05	0.06	0.05	0.01
P(α − helix)	142	98	101	67	70	151	111	57	100	108	121	114	145	113	57	77	83	108	69	106
P(β − sheet)	83	93	54	89	119	37	110	75	87	160	130	74	105	138	55	75	119	137	147	170
P(turn)	66	95	146	156	119	74	98	156	95	47	59	101	60	60	152	143	96	96	114	50
f(i)	0.060	0.070	0.147	0.161	0.149	0.056	0.074	0.102	0.140	0.043	0.061	0.055	0.068	0.059	0.102	0.120	0.086	0.077	0.082	0.062
f(i + 1)	0.076	0.106	0.110	0.083	0.050	0.060	0.098	0.085	0.047	0.034	0.025	0.115	0.082	0.041	0.301	0.139	0.108	0.013	0.065	0.048
f(i + 2)	0.035	0.099	0.179	0.191	0.117	0.077	0.037	0.190	0.093	0.013	0.036	0.072	0.014	0.065	0.034	0.125	0.065	0.064	0.114	0.028
f(i + 3)	0.058	0.085	0.081	0.091	0.128	0.064	0.098	0.152	0.054	0.056	0.070	0.095	0.055	0.065	0.068	0.106	0.079	0.167	0.125	0.053

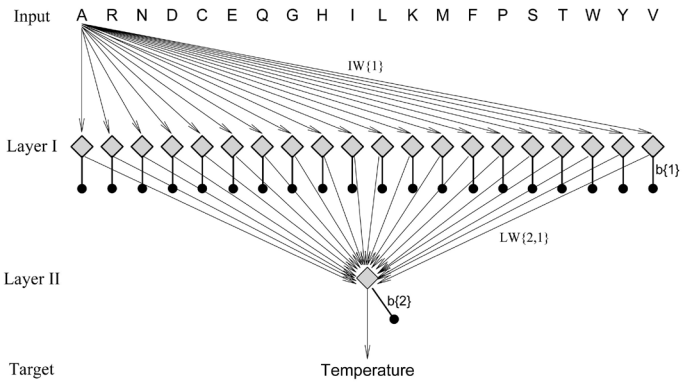

Figure 1. 20-1 feedforward backpropagation neural network to model the relationship between 20 pieces of information on primary structure of β-glucosidase, which are labeled using the symbols of 20 types of amino acids, and its temperature. Each diamond presents a neuron. IW{1} is the input weights, 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.

Table 2. Difference between normalized polarizability index ($\sigma\alpha$) and amino acid distribution probability in β-glucosidases A9UIG0 and Q4U4W7.

Amino	Number		$\sigma\alpha$		$\sigma\alpha \times$ Number		Distribution probability	
Acid	A9UIG0	Q4U4W7	A9UIG0	Q4U4W7	A9UIG0	Q4U4W7	A9UIG0	Q4U4W7
A	82	74	−0.01	−0.01	−0.82	−0.74	0.0021	0.0015
R	29	36	−0.08	−0.08	−2.32	−2.88	0.0043	0.0012
N	56	46	−0.04	−0.04	−2.24	−1.84	0.0202	0.0174
D	50	50	−0.03	−0.03	−1.5	−1.5	0.0039	0.0180
C	8	8	−0.03	−0.03	−0.24	−0.24	0.1682	0.0841
E	35	36	−0.04	−0.04	−1.4	−1.44	0.0218	0.0224
Q	28	25	−0.05	−0.05	−1.4	−1.25	0.0642	0.0051
G	94	86	0	0	0	0	0.0006	0.0027
H	16	11	−0.06	−0.06	−0.96	−0.66	0.0715	0.0808
I	35	43	−0.04	−0.04	−1.4	−1.72	0.0194	0.0240
L	58	65	−0.04	−0.04	−2.32	−2.6	0.0002	0.0054
K	29	29	−0.05	−0.05	−1.45	−1.45	0.0317	0.0317
M	11	19	−0.05	−0.05	−0.55	−0.95	0.0404	0.0138
F	34	33	−0.08	−0.08	−2.72	−2.64	0.0285	0.0193
P	53	65	−0.04	−0.04	−2.12	−2.6	0.0058	0.0010
S	62	61	−0.02	−0.02	−1.24	−1.22	0.0029	0.0112
T	57	56	−0.03	−0.03	−1.71	−1.68	0.0023	0.0005
W	18	16	−0.12	−0.12	−2.16	−1.92	0.0023	0.1362
Y	41	33	−0.09	−0.09	−3.69	−2.97	0.0142	0.0174
V	70	74	−0.03	−0.03	−2.1	−2.22	0.0067	0.0008

2.4. Validation of Predictions

Each predictor went through this predictive model with same procedures in order to compare its output statistically. Table 3 lists a total of 40 β-glucosidases to be analyzed, of which 25 were used to generate the weights and biases in neural network as training group, and 15 were used to validate the neural network with trained weights and biases as validation group. This is a traditional way used in neural network. Then, the delete-1 observation jackknife was used and each time one observation was left out from the sample set for validation, because it is most effective in comparison with independent dataset test and subsampling test, and is widely used [41]. Finally, cross-validation was used, and the data were split into 10 or 4 subsets, which had 4 or 10 cases and was held out in turn as the validation set [42].

2.5. Statistics

One hundred trainings were conducted for each predictor in the predictive model, and their weights and biases were used to predict the temperature optimum 100 times. The mean and standard deviation of predicted values were compared with the recorded temperature optimum for each β-glucosidase [43], and linear regression was also used to evaluate the predicted temperature values with their recorded ones.

3. RESULTS AND DISCUSSION

Theoretically, the neural network displayed in Figure 1 can account for various linear and nonlinear relationships between amino acid properties of primary structure and temperature optimum of β-glucosidases, which can guarantee the screening of various predictors, no matter whether the relationship between predictors and temperature is linear or nonlinear [39].

Technically, the initialization of weights and biases and number of training epochs govern whether the neural network can converge during training process, for which the weights and biases were initialized by random initialization function, and 250 training epochs were conducted. Only 4 out of 25 amino acid properties can be converged and shown in Figure 2, where each line represents that a training process contains random initialization of weights and biases with 250 training epochs. As seen, the convergence can be reached within 250 training epochs with any random initialization, which lays the foundation to guarantee the training process, indicating that these 4 properties can be served as predictors to predict the temperature optimum of β-glucosidases. However, it can be found that different predictors have different profiles of their convergence and the convergence of profiles of amino-acid distribution probability (bottom panel) reached narrower than others.

Table 3 demonstrates the comparison of recorded temperature optimum with predicted temperature optimum for 40 β-glucosidases. If there is no statistical difference between recorded and predicted temperature optimum, a predictor would be considered workable. Accordingly, if no statistical difference was found between recorded and predicted temperature optimum, the predicted temperature optimum is marked with asterisk. The last row in Table 3 summarizes the overall performance, where it can be seen that the normalized polarizability index ($\sigma\alpha$) and amino-acid distribution probability works better than the other two.

Figure 3 displays the percentage of β-glucosidases with correctly predicted temperature during the training process. As can be seen, the amino-acid distribution probability worked best in training group, which resulted that the temperature optimum of all β-glucosidases was correctly predicted, and followed by the normalized polarizability index (88%), whereas only the normalized polarizability index reached 60% of correctly prediction in validation group. Figure 4 visualized the regression between recorded and predicted temperature optimum by using these four amino-acid properties as predictors. Figure 5 shows the results of delete-1, delete-4 and delete-10 jackknife validations, where it can be seen that both normalized polarizability index and amino-acid distribution probability gave better performance and that there was generally no significant difference between different deletions.

In conclusion, many studies have been focused on revealing the structure-function relationship of enzymes [44-46]. This study is consistent with our previous studies [9-19], demonstrating that some predictors

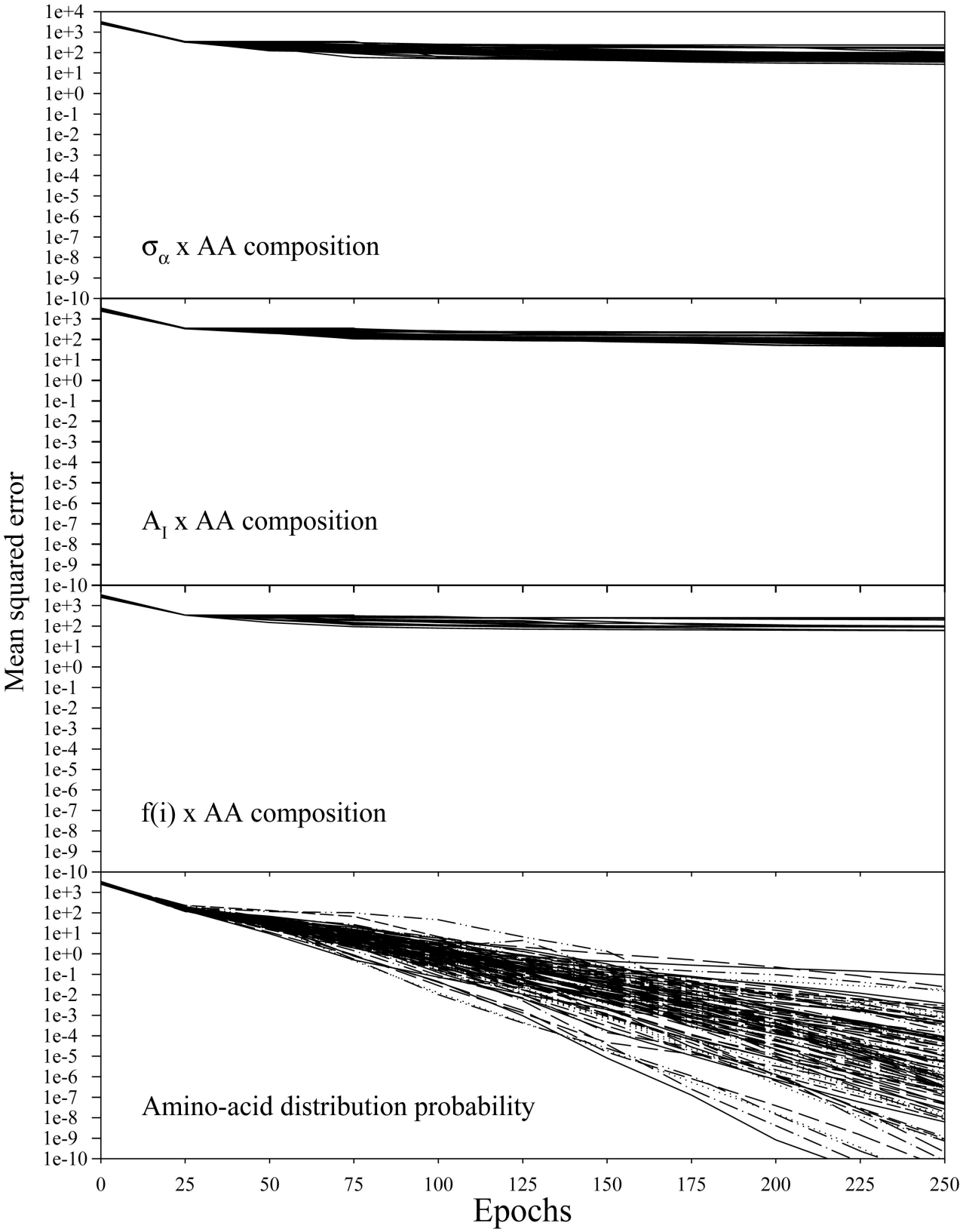

Figure 2. Convergence of mean squared error performance function with 100 different initial weights and biases generated by random initialization function.

Table 3. Comparison between recorded and predicted temperature optimum in 40 of β-glucosidases. The predicted temperature optimum was presented as mean ± SD of 100 predictions. AA, the amino-acid composition; AA DP, amino-acid distribution probability. *, no statistical difference with the recorded temperature optimum.

Group	Accession Number	Recorded temperature	Temperature optimum predicted by			
			$\sigma_\alpha \times$ AA	$A_I \times$ AA	I(i) × AA	AA DP
Training	Q47RE2	25.0	43.5 ± 7.1	48.4 ± 6.0	49.2 ± 4.8	25.0 ± 0.0*
	Q25BW5	30.0	43.9 ± 6.6	48.3 ± 5.2	49.2 ± 4.4	30.0 ± 0.0*
	P15885	30.0	42.7 ± 7.8*	48.2 ± 6.0	48.3 ± 5.3	30.0 ± 0.0*
	Q59976	30.0	43.5 ± 7.1*	48.4 ± 5.5	49.1 ± 4.8	30.0 ± 0.0*
	Q08IT7	30.0	44.1 ± 6.1	48.8 ± 4.7	48.6 ± 4.6	30.0 ± 0.0*
	Q86D78	35.0	42.7 ± 7.2*	48.1 ± 5.4	48.4 ± 5.1	35.0 ± 0.0*
	Q2WGB4	37.0	44.2 ± 6.1*	48.7 ± 4.1	48.6 ± 4.6	37.0 ± 0.0*
	A1C3J9	40.0	46.0 ± 4.6*	49.5 ± 3.3	49.1 ± 3.6	40.0 ± 0.0*
	Q9SLA0	40.0	43.6 ± 6.5*	48.2 ± 5.2*	48.5 ± 4.9*	40.0 ± 0.0*
	Q875K3	40.0	44.1 ± 6.1*	48.6 ± 4.6*	48.5 ± 5.0*	40.0 ± 0.0*
	Q6QGY5	40.0	43.2 ± 6.7*	48.2 ± 5.2*	48.5 ± 5.0*	40.0 ± 0.0*
	P94248	45.0	42.7 ± 7.4*	48.1 ± 5.5*	48.5 ± 5.0*	45.0 ± 0.0*
	Q9AT27	50.0	44.0 ± 6.7*	48.4 ± 4.9*	48.9 ± 4.6*	50.0 ± 0.0*
	Q8T0W7	50.0	52.4 ± 3.2*	52.7 ± 5.0*	51.2 ± 1.9*	50.0 ± 0.1*
	Q4U4W7	50.0	59.3 ± 10.2*	51.8 ± 4.4*	53.6 ± 7.4*	50.0 ± 0.0*
	P49235	50.0	47.5 ± 4.4*	48.7 ± 4.3*	49.6 ± 3.1*	50.0 ± 0.0*
	Q9H227	50.0	43.9 ± 6.4*	48.3 ± 5.0*	48.8 ± 4.5*	50.0 ± 0.0*
	O08331	65.0	59.3 ± 8.9*	53.4 ± 5.7	53.1 ± 6.0*	65.0 ± 0.0*
	Q12715	65.0	59.7 ± 8.9*	54.1 ± 7.4*	53.5 ± 6.5*	65.0 ± 0.1*
	A9UIG0	70.0	60.6 ± 9.8*	54.2 ± 7.6	53.9 ± 7.4	70.0 ± 0.1*
	Q9P8F4	70.0	62.5 ± 11.7*	55.2 ± 9.7*	54.1 ± 8.0*	70.0 ± 0.1*
	Q7Z9M5	70.0	61.8 ± 10.7*	54.2 ± 7.7	54.0 ± 7.6	70.0 ± 0.1*
	Q8TGI8	71.5	58.6 ± 7.9*	52.9 ± 5.3	53.3 ± 6.1	71.5 ± 0.1*
	Q746L1	88.0	64.9 ± 14.3*	55.2 ± 9.9	54.0 ± 8.5	88.0 ± 0.0*
	B9K7M5	95.0	66.5 ± 16.0*	55.0 ± 9.5	53.8 ± 7.9	95.0 ± 0.0*

Continued

Validation	B5TWK3	22.0	63.4 ± 13.0	55.8 ± 11.8	54.4 ± 9.0	60.7 ± 6.7
	B6ZKM3	30.0	52.0 ± 8.8	50.5 ± 3.6	51.0 ± 3.5	51.6 ± 13.1*
	O61594	30.0	44.5 ± 5.9	49.3 ± 3.8	49.3 ± 4.0	62.5 ± 10.4
	Q12601	35.0	47.0 ± 7.5*	49.8 ± 4.2	50.6 ± 4.9	59.3 ± 9.4
	P96316	35.0	65.4 ± 16.7*	55.6 ± 12.3*	54.5 ± 9.3	65.9 ± 9.8
	B5TWK3	37.0	63.4 ± 13.0	55.8 ± 11.8*	54.4 ± 9.0*	60.7 ± 6.7
	P96316	45.0	65.4 ± 16.7*	55.6 ± 12.3*	54.5 ± 9.3*	65.9 ± 9.8
	Q9H227 V168Y	50.0	43.8 ± 6.4*	48.3 ± 5.0*	48.7 ± 4.5*	46.9 ± 3.8*
	Q9SPK3	50.0	43.0 ± 7.1*	48.2 ± 5.2*	48.4 ± 5.2*	63.5 ± 9.9*
	Q9C3Z9	50.0	64.3 ± 16.3*	55.1 ± 10.8*	53.7 ± 7.8*	74.4 ± 10.3
	Q2UUD6	60.0	58.9 ± 9.9*	54.2 ± 7.8*	53.4 ± 6.6*	62.9 ± 3.7*
	Q12715	70.0	59.7 ± 8.9*	54.1 ± 7.4	53.5 ± 6.5	65.0 ± 0.1
	P26208	65.0	48.3 ± 8.0	48.7 ± 4.5	49.6 ± 3.8	57.7 ± 14.8*
	P10482	80.0	51.4 ± 6.7	50.6 ± 3.1	50.9 ± 2.8	54.3 ± 14.2*
	Q08638	85.0	67.3 ± 17.4*	55.0 ± 9.8	54.1 ± 8.7	68.7 ± 14.5*
Total			31	18	18	32

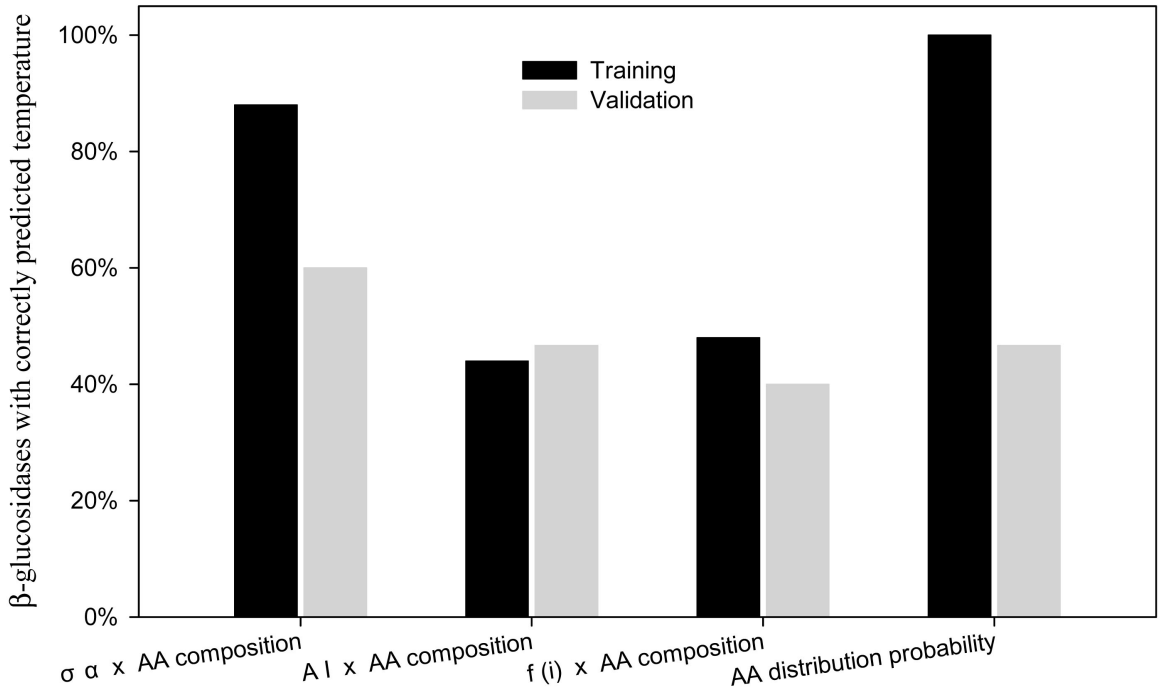

Figure 3. **Percentage of β-glucosidases with correctly predicted pH. The training and validation groups contained 25 and 15 β-glucosidases.**

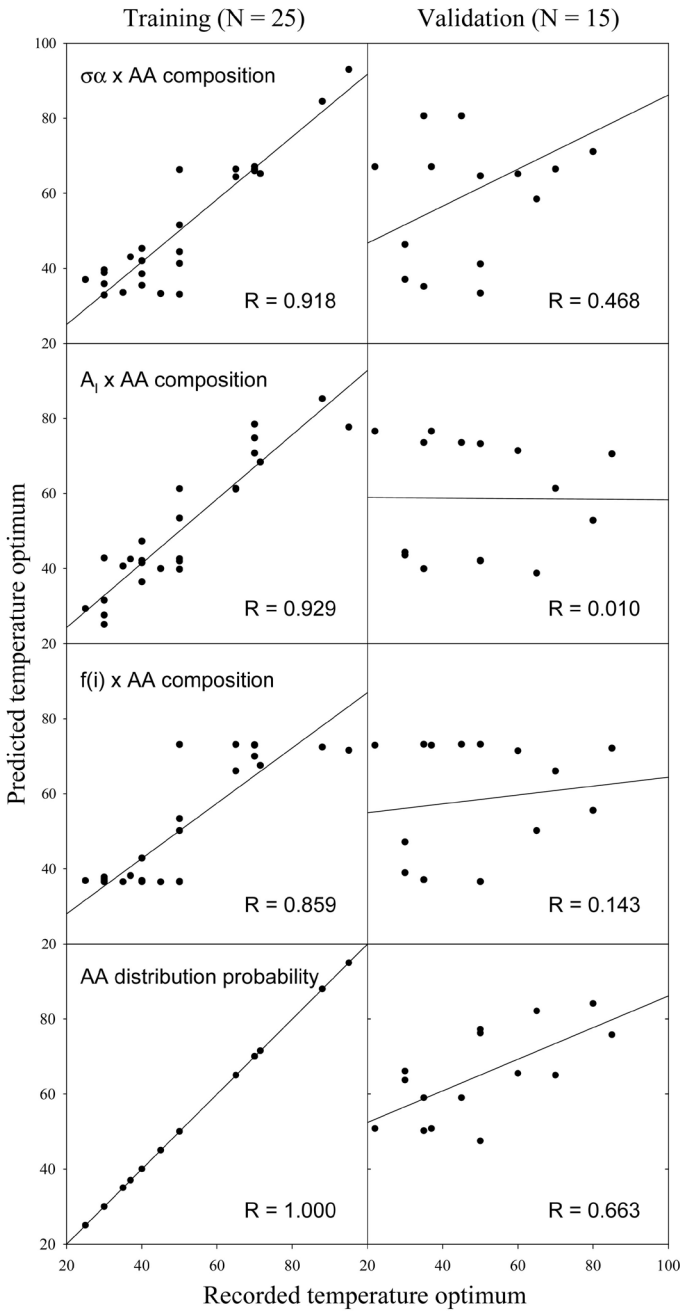

Figure 4. Linear regression between recorded and predicted temperature optimum in training and validation groups, respectively. Linear regressions for training groups are: (1) Temperature Optimum = 8.2944 × ($\sigma\alpha$ × AA composition) + 0.8352, $P < 0.0001$; (2) Temperature Optimum = 7.1604 × (A_I × AA composition) + 0.8563, $P < 0.0001$; (3) Temperature Optimum = 13.3125 × (f(i) × AA composition) + 0.7368, $P < 0.0001$; (4) Temperature Optimum = 0.0166 × AA distribution probability + 0.9997, $P < 0.0001$. Linear regressions for validation groups are: (1) Temperature Optimum = 0.4935 × ($\sigma\alpha$ × AA composition) + 36.8632, $P = 0.0783$; (2) Temperature Optimum = −0.0079 × (A_I × AA composition) + 59.0869, $P = 0.9726$; (3) Temperature Optimum = 0.1182 × (f(i) × AA composition) + 52.6179, $P = 0.6118$; (4) Temperature Optimum = 0.4216 × AA distribution probability + 43.9512; $P = 0.0071$.

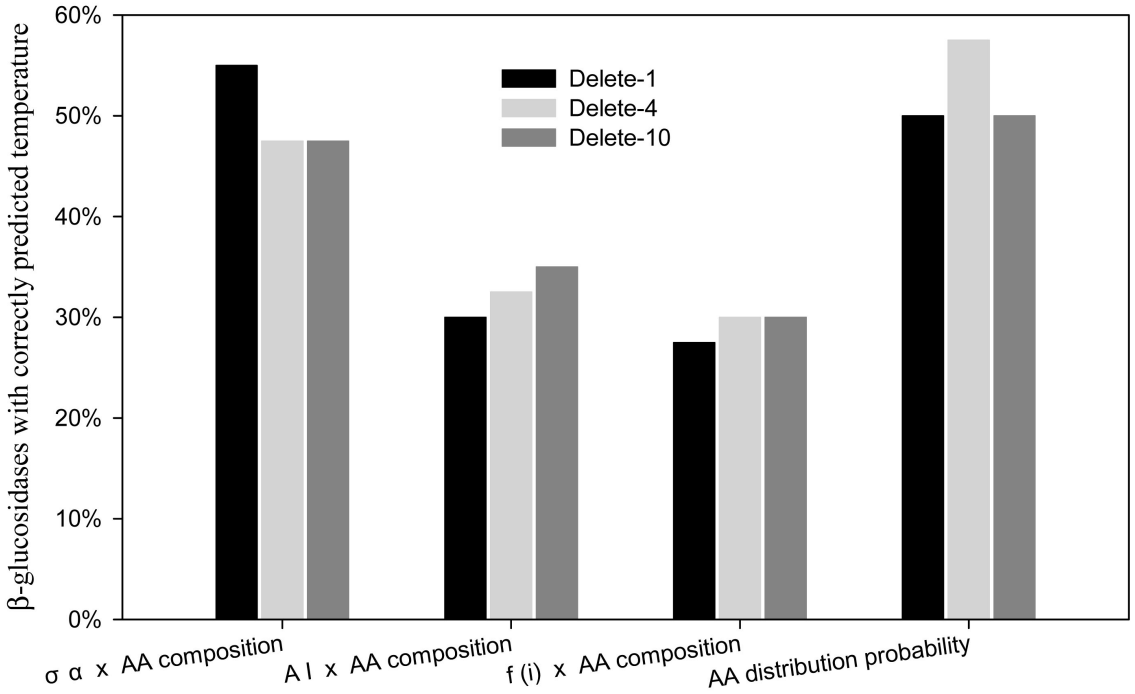

Figure 5. Percentage of β-glucosidases with correctly predicted temperature. The validation among 40 β-glucosidases was conducted using MatLab by means of delete-1, delete-4 and delete-10 jackknifing. AA, amino-acid.

do have a promising prospective to predict the enzymatic optimal working conditions based on the information related to enzyme primary structure. Surely, further efforts are needed to explore a cost-effective way to predict various enzymatic parameters of β-glucosidases.

FUND

This study was supported by National Natural Science Foundation of China (31560315), and Key Project of Guangxi Scientific Research and Technology Development Plan (AB17190534).

CONFLICTS OF INTEREST

The authors declare no conflicts of interest regarding the publication of this paper.

REFERENCES

1. Jeng, W.Y., Wang, N.C., Lin, M.H., Lin, C.T., Liaw, Y.C., Chang, W.J., Liu, C.I., Liang, P.H. and Wang, A.H. (2011) Structural and Functional Analysis of Three β-Glucosidases from Bacterium *Clostridium cellulovorans*, Fungus *Trichoderma reesei* and Termite *Neotermes koshunensis*. *Journal of Structural Biology*, **173**, 46-56.
2. Kacher, Y., Brumshtein, B., Boldin-Adamsky, S., Toker, L., Shainskaya, A., Silman, I., Sussman, J.L. and Futerman, A.H. (2008) Acid β-Glucosidase: Insights from Structural Analysis and Relevance to Gaucher Disease Therapy. *Biological Chemistry*, **389**, 1361-1369. https://doi.org/10.1515/BC.2008.163
3. Granovsky-Grisaru, S., Belmatoug, N., vom Dahl, S., Mengel, E., Morris, E. and Zimran, A. (2011) The Management of Pregnancy in Gaucher Disease. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, **156**, 3-8. https://doi.org/10.1016/j.ejogrb.2010.12.024
4. Chen, K.I., Erh, M.H., Su, N.W., Liu, W.H., Chou, C.C. and Cheng, K.C. (2012) Soyfoods and Soybean Prod-

ucts: from Traditional Use to Modern Applications. *Applied Microbiology & Biotechnology*, **96**, 9-22. https://doi.org/10.1007/s00253-012-4330-7

5. Dashtban, M., Maki, M., Leung, K.T., Mao, C. and Qin, W. (2010) Cellulase Activities in Biomass Conversion: Measurement Methods and Comparison. *Critical Reviews in Biotechnology*, **30**, 302-309.
6. Wilson, D.B. (2009) Cellulases and Biofuels. *Current Opinions in Biotechnology*, **20**, 295-299. https://doi.org/10.1016/j.copbio.2009.05.007
7. Schomburg, I., Chang, A., Hofmann, O., Ebeling, C., Ehrentreich, F. and Schomburg, D. (2002) BRENDA: A Resource for Enzyme Data and Metabolic Information. *Trends in Biochemical Sciences*, **27**, 54-56. https://doi.org/10.1016/S0968-0004(01)02027-8
8. Placzek, S., Schomburg, I., Chang, A., Jeske, L., Ulbrich, M., Tillack, J. and Schomburg, D. (2017) BRENDA in 2017: New Perspectives and New Tools in BRENDA. *Nucleic Acids Research*, **45**, D380-D388. https://doi.org/10.1093/nar/gkw952
9. Yan, S. and Wu, G. (2011) Searching of Predictors to Predict pH of Cellulases. *Applied Biochemistry and Biotechnology*, **165**, 856-869. https://doi.org/10.1007/s12010-011-9303-2
10. Yan, S. and Wu, G. (2013) Prediction of Optimal pH in Hydrolytic Reaction of Beta-Glucosidase. *Applied Biochemistry and Biotechnology*, **169**, 1884-1894. https://doi.org/10.1007/s12010-013-0103-8
11. Yan, S. and Wu, G. (2012) Prediction of Optimal pH and Temperature of Cellulases Using Neural Network. *Protein & Peptide Letters*, **19**, 29-39. https://doi.org/10.2174/092986612798472794
12. Yan, S., Shi, D., Nong, H. and Wu, G. (2011) Simultaneously Predicting pH and Temperature Optimum in Catalytic Reaction of Beta-Glucosidase. *Guangxi Sciences*, **18**, 253-260.
13. Yan, S. and Wu, G. (2019) Predicting pH Optimum for Activity of Beta-Glucosidases. *Journal of Biomedical Science and Engineering*, **12**, 354-367. https://doi.org/10.4236/jbise.2019.127027
14. Yan, S. and Wu, G. (2012) Exhausted Jackknife Validation Exemplified by Prediction of Temperature Optimum in Enzymatic Reaction of Cellulases. *Applied Biochemistry and Biotechnology*, **166**, 997-1107. https://doi.org/10.1007/s12010-011-9487-5
15. Yan, S. and Wu, G. (2013) Prediction of Temperature Optimum in Enzymatic Reaction of Beta-Cellobiosidases with Exhausted Jackknife Validation. *Life Science Journal*, **10**, 1673-1678.
16. Yan, S. and Wu, G. (2011) Prediction of Michaelis-Menten Constant in Beta-Cellobiosidase's Reaction with Lactoside as Substrate. *Enzyme Engineering*, **1**, 102.
17. Yan, S. and Wu, G. (2011) Prediction of Michaelis-Menten Constant of Beta-Glucosidases Using Nitrophenyl-Beta-D-Glucopyranoside as Substrate. *Protein & Peptide Letters*, **18**, 1053-1057. https://doi.org/10.2174/092986611796378747
18. Yan, S., Shi, D., Nong, H. and Wu, G. (2012) Predicting *Km* Values of Beta-Glucosidases Using Cellobiose as Substrate. *Interdisciplinary Sciences*: *Computational Life Sciences*, **4**, 46-53. https://doi.org/10.1007/s12539-012-0115-z
19. Yan, S. and Wu, G. (2013) Prediction of Turnover Number of Cellulose 1,4-Beta-Cellobiosidase. *Protein & Peptide Letters*, **20**, 255-264.
20. Berrin, J.G., Czjzek, M., Kroon, P.A., McLauchlan, W.R., Puigserver, A., Williamson, G. and Juge, N. (2003) Substrate (aglycone) Specificity of Human Cytosolic Beta-Glucosidase. *Biochemical Journal*, **373**, 41-48. https://doi.org/10.1042/bj20021876
21. Tsukada, T., Igarashi, K., Fushinobu, S. and Samejima, M. (2008) Role of Subsite +1 Residues in Temperature Dependence and Catalytic Activity of the Glycoside Hydrolase Family 1 Beta-Glucosidase BGL1A from the Ba-

sidiomycete *Phanerochaete chrysosporium*. *Biotechnology and Bioengineering*, **99**, 1295-1302. https://doi.org/10.1002/bit.21717

22. Gundllapalli, S.B., Pretorius, I.S. and Cordero Otero, R.R. (2007) Effect of the Cellulose-Binding Domain on the Catalytic Activity of a Beta-Glucosidase from *Saccharomycopsis fibuligera*. *Journal of Industrial Microbiology & Biotechnology*, **34**, 413-421. https://doi.org/10.1007/s10295-007-0213-9
23. Chen, H., Hayn, M. and Esterbauer, H. (1992) Purification and Characterization of Two Extracellular Beta-Glucosidases from *Trichoderma reesei*. *Biochimica et Biophysica Acta*, **1121**, 54-60. https://doi.org/10.1016/0167-4838(92)90336-C
24. UniProt Consortium (2019) UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Research*, **47**, D506-D515. https://doi.org/10.1093/nar/gky1049
25. Burlingame, A.L. and Carr, S.A. (1996) Mass Spectrometry in the Biological Sciences. Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-4612-0229-5
26. Zamyatin, A.A. (1972) Protein Volume in Solution. *Progress in Biophysics & Molecular Biology*, **24**, 107-123. https://doi.org/10.1016/0079-6107(72)90005-3
27. Darby, N.J. and Creighton, T.E. (1993) Dissecting the Disulphide-Coupled Folding Pathway of Bovine Pancreatic Trypsin Inhibitor. Forming the First Disulphide Bonds in Analogues of the Reduced Protein. *Journal of Molecular Biology*, **232**, 873-896. https://doi.org/10.1006/jmbi.1993.1437
28. Kyte, J. and Doolittle, R.F. (1982) A Simple Method for Displaying the Hydropathic Character of a Protein. *Journal of Molecular Biology*, **157**, 105-132. https://doi.org/10.1016/0022-2836(82)90515-0
29. Trinquier, G., Sanejouand, Y.H. and Hausman, R.E. (1998) Which Effective Property of Amino Acids is Best Preserved by the Genetic Code? *Protein Engineering, Design and Selection*, **11**, 153-169. https://doi.org/10.1093/protein/11.3.153
30. Cooper, G.M. (2004) The Cell: A Molecular Approach. ASM Press, Washington DC, 51.
31. Dwyer, D.S. (2005) Electronic Properties of Amino Acid Side Chains: Quantum Mechanics Calculation of Substituent Effects. *BMC Chemical Biology*, **5**, 2. https://doi.org/10.1186/1472-6769-5-2
32. Chou, P.Y. and Fasman, G.D. (1978) Prediction of Secondary Structure of Proteins from Amino Acid Sequence. *Advances in Enzymology and Related Subjects of Biochemistry*, **47**, 45-148. https://doi.org/10.1002/9780470122921.ch2
33. Feller, W. (1968) An Introduction to Probability Theory and Its Applications. 3rd Edition, Wiley, New York.
34. Wu, G. and Yan, S. (2008) Prediction of Mutations Engineered by Randomness in H5N1 Hemagglutinins of Influenza A Virus. *Amino Acids*, **35**, 365-373. https://doi.org/10.1007/s00726-007-0602-4
35. Wu, G. and Yan, S. (2008) Lecture Notes on Computational Mutation. Nova Science Publishers, New York.
36. Yan, S. and Wu, G. (2009) Descriptively Quantitative Relationship between Mutated *N*-Acetylgalactosamine-6-Sulfatase and Mucopolysaccharidosis IVA. *Peptide Science*, **92**, 399-404. https://doi.org/10.1002/bip.21205
37. Yan, S. and Wu, G. (2010) Prediction of Mutation Positions in H5N1 Neuraminidases by Means of Neural Network. *Annals of Biomedical Engineering*, **38**, 984-992. https://doi.org/10.1007/s10439-010-9907-7
38. Yan, S. and Wu, G. (2010) Linking Mutated Structure of Adrenoleukodystrophy Protein with X-Linked Adrenoleukodystrophy. *Computer Methods in Biomechanics and Biomedical Engineering*, **13**, 403-411. https://doi.org/10.1080/10255840903279974
39. Demuth, H. and Beale, M. (2001) Neural Network Toolbox for Use with MatLab. User's Guide, Version 4, MathWorks Inc., Natick, MA.
40. MathWorks Inc (1984-2001) MatLab-The Language of Technical Computing (Version 6.1.0.450, Release 12.1).

MathWorks Inc., Natick, MA.

41. Chou, K.C. and Zhang, C.T. (1995) Prediction of Protein Structural Classes. *Critical Reviews in Biochemistry and Molecular Biology*, **30**, 275-349. https://doi.org/10.3109/10409239509083488
42. Chou, K.C. and Shen, H.B. (2010) Plant-mPLoc: A Top-Down Strategy to Augment the Power for Predicting Plant Protein Subcellular Localization. *PLoS One*, **5**, e11335. https://doi.org/10.1371/journal.pone.0011335
43. Sokal, R.R. and Rohlf, F.J. (1995) Biometry: The Principles and Practices of Statistics in Biological Research. 3rd Edition, W. H. Freeman, New York, 203-218.
44. Campbell, R.L. and Davies, P.L. (2012) Structure-Function Relationships in Calpains. *Biochemical Journal*, **447**, 335-351. https://doi.org/10.1042/BJ20120921
45. Sacchi, S., Caldinelli, L., Cappelletti, P., Pollegioni, L. and Molla, G. (2012) Structure-Function Relationships in Human D-Amino Acid Oxidase. *Amino Acids*, **43**, 1833-1850. https://doi.org/10.1007/s00726-012-1345-4
46. Silavi, R., Divsalar, A. and Saboury, A.A. (2012) A Short Review on the Structure-Function Relationship of Artificial Catecholase/Tyrosinase and Nuclease Activities of Cu-Complexes. *Journal of Biomolecular Structure and Dynamics*, **30**, 752-772. https://doi.org/10.1080/07391102.2012.689704