

# $\alpha$ -淀粉酶 Amy7C 及其突变体催化常数的定量预测 \* Quantitative Predicting *Kcat* of $\alpha$ -amylase Amy7C and Its Mutants

黄纪民, 师德强, 严少敏 \*\*, 吴光, 谢能中, 龙思宇, 李检秀, 黄艳燕

HUANG Ji-min, SHI De-qiang, YAN Shao-min, WU Guang, XIE Neng-zhong, LONG Si-yu, LI Jian-xiu, HUANG Yan-yan

(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)

(Guangxi Academy of Sciences, State Key Laboratory of Non-food Biomass Enzyme Technology, National Engineering Research Center for Non-food Biorefinery, Biomass Industrialization Engineering Institute, Guangxi Key Laboratory of Biorefinery, Nanning, Guangxi, 530007, China)

**摘要:**【目的】利用  $\alpha$ -淀粉酶 Amy7c 及其突变体的氨基酸信息, 预测该酶的催化常数 (*Kcat*), 并筛选出能预测  $\alpha$ -淀粉酶 *Kcat* 最具效果的氨基酸属性。【方法】先以 20-1 前馈反向传播的神经网络为模型, 完成 535 种氨基酸属性对  $\alpha$ -淀粉酶 Amy7C 及其突变体催化常数的拟合。再将  $\alpha$ -淀粉酶 Amy7C 及其 54 个突变体的数据分为 2 组, 用 35 个酶作为训练组进行拟合, 20 个酶作为验证组进行检验。最后, 对 8 种不同层次及神经元个数的模型进行比较。【结果】110 个氨基酸属性可实现 20-1 神经网络模型收敛, 表明这些氨基酸属性可用于预测  $\alpha$ -淀粉酶的催化常数, 不同指标的预测效果不同。多模型的分析结果显示, 不同模型对训练组 *R* 值的结果具有显著性差异, 而对训练组 *P* 值、验证组 *R* 值和验证组 *P* 值结果无显著性差异。【结论】氨基酸分布概率等属性可以用于预测  $\alpha$ -淀粉酶催化常数。四层神经网络模型是预测  $\alpha$ -淀粉酶催化常数的相对理想的模型。

**关键词:** 氨基酸属性  $\alpha$ -淀粉酶 催化常数 预测

中图分类号: Q556<sup>+</sup>.2 文献标识码: A 文章编号: 1002-7378(2014)04-0294-05

**Abstract:**【Objective】The *Kcat* of  $\alpha$ -amylase Amy7C and its mutants was predicted using amino acid information, and the most suitable amino acid property for predicting *Kcat* of  $\alpha$ -amylase was selected.【Methods】20-1 feedforward backpropagation neural network was used to screen 535 amino acid properties as predictors to predict the *Kcat* of  $\alpha$ -amylase Amy7C and its 54 mutants, which were divided into two group, 35 of them served as training group for fitting, and the other 20 were treated as validation. Eight models for different layers and numbers of neurons were also compared.【Results】110 amino acid properties, which converged during fitting in the 20-1 neural network model, could be used to predict the *Kcat*. Different amino acid properties presented different predicting effect. The multi-model results

showed that there was significant difference between *R* values in training groups, but there was no significant difference between *P* values in training groups, as well as *R* and *P* values in validation groups.【Conclusion】Some amino acid properties such as distribution probability could be used to predict the *Kcat* of  $\alpha$ -amylase, to which four-layer neural network reveals the relative ideal model.

**Key words:** amino acid property,  $\alpha$ -amylase, *Kcat*, prediction

收稿日期: 2014-08-10

作者简介: 黄纪民(1976-), 男, 工程师, 主要从事生物质能源开发利用研究。

\* 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。

**【研究意义】** $\alpha$ -淀粉酶广泛分布于动物、植物和微生物中,能水解淀粉产生糊精、麦芽糖、低聚糖和葡萄糖等物质,是工业生产中应用最为广泛的酶制剂之一,已应用于变性淀粉及淀粉糖、焙烤工业、啤酒酿造、酒精工业、发酵以及纺织等行业<sup>[1]</sup>。目前,在生物质能源领域,淀粉质生产乙醇仍然是生物乙醇的主要生产方式,此过程中  $\alpha$ -淀粉酶不仅是参与者,而且对提高乙醇产率至关重要。酶活力是指酶催化某一化学反应的能力,然而,目前酶活力相关功能参数的测定需要通过实验得出,而实验方式比较耗费时间、财力和人力。因此,探寻一种利用酶的简单结构信息即可预测该酶的相关功能参数的方法很有必要。**【前人研究进展】**严少敏等运用神经网络模型对纤维素酶、 $\beta$ -葡糖苷酶和  $\beta$ -纤维乙糖苷酶的功能参数进行了预测,包括最适 pH 值<sup>[2~4]</sup>、最适温度<sup>[4~7]</sup>、米氏常数<sup>[8~11]</sup>和转换数<sup>[12]</sup>,发现某些氨基酸属性可以用来预测酶的功能参数<sup>[13]</sup>。**【本研究切**

**入点】**酶的催化常数 ( $K_{cat}$ ) 是表示酶的催化效率的参数,至今,未见有利用酶的简单结构信息预测酶催化常数的相关报道。**【拟解决的关键问题】**探索利用酶的简单结构信息预测  $\alpha$ -淀粉酶催化常数的可行性及方法,并筛选出预测  $\alpha$ -淀粉酶催化常数最具效果的氨基酸属性。

## 1 数据来源和研究方法

$\alpha$ -淀粉酶 Amy7C 及其 54 个突变体的催化常数相关数据由王成华博士<sup>[14]</sup>提供,所涉及的 535 种氨基酸属性分别为 39 个氨基酸构成属性,219 个理化性质属性,273 个二级结构属性和 4 个动态氨基酸属性。首先,用 535 种氨基酸属性分别对淀粉酶 Amy7C 及其 54 个变异体进行量化(表 1 以淀粉酶 Amy7C 及其 R172K 变异体为例列出了用静、动态氨基酸属性对 20 种氨基酸的具体量化结果)。

表 1 用静、动态氨基酸属性量化淀粉酶 Amy7C 及其 R172K 变异体\*

Table 1 Quantification of Amy7C and its R172K mutant using constant and dynamic properties of amino acids

氨基酸 AA	No.		SNEP660103		SNEP660103×No.		CC(%)		FC(%)		DP	
	Amy7C	R172K	Amy7C	R172K	Amy7C	R172K	Amy7C	R172K	Amy7C	R172K	Amy7C	R172K
A	37	37	-0.11	-0.11	-4.07	-4.07	8.62	8.62	7.28	7.28	0.0185	0.0185
R	18	17	0.079	0.079	1.422	1.343	4.20	3.96	7.33	7.28	0.0312	0.1098
N	39	39	-0.136	-0.136	-5.304	-5.304	9.09	9.09	4.63	4.68	0.0031	0.0031
D	29	29	-0.285	-0.285	-8.265	-8.265	6.76	6.76	4.78	4.78	0.0069	0.0069
C	1	1	-0.184	-0.184	-0.184	-0.184	0.23	0.23	2.78	2.78	1.0000	1.0000
E	18	18	-0.067	-0.067	-1.206	-1.206	4.20	4.20	4.21	4.23	0.0389	0.0389
Q	21	21	-0.246	-0.246	-5.166	-5.166	4.90	4.90	2.71	2.73	0.0062	0.0062
G	37	37	-0.073	-0.073	-2.701	-2.701	8.62	8.62	6.73	6.71	0.0056	0.0056
H	14	14	0.32	0.32	4.48	4.48	3.26	3.26	4.21	4.20	0.0010	0.0010
I	23	23	0.001	0.001	0.023	0.023	5.36	5.36	5.14	5.15	0.0112	0.0112
L	23	23	-0.008	-0.008	-0.184	-0.184	5.36	5.36	7.00	6.98	0.0460	0.0460
K	18	19	0.049	0.049	0.882	0.931	4.20	4.43	4.41	4.42	0.0831	0.0136
M	8	8	-0.041	-0.041	-0.328	-0.328	1.86	1.86	1.42	1.42	0.2243	0.2243
F	13	13	0.438	0.438	5.694	5.694	3.03	3.03	2.46	2.46	0.0463	0.0463
P	12	12	-0.016	-0.016	-0.192	-0.192	2.80	2.80	4.88	4.86	0.1241	0.1241
S	39	39	-0.153	-0.153	-5.967	-5.967	9.09	9.09	8.42	8.39	0.0064	0.0064
T	27	27	-0.208	-0.208	-5.616	-5.616	6.29	6.29	6.81	6.83	0.0161	0.0161
W	12	12	0.493	0.493	5.916	5.916	2.80	2.80	0.69	0.68	0.0310	0.0310
Y	19	19	0.381	0.381	7.239	7.239	4.43	4.43	3.32	3.32	0.0852	0.0852
V	21	21	-0.155	-0.155	-3.255	-3.255	4.90	4.90	6.50	6.50	0.0053	0.0053

\* 属性 SNEP660103 代表主成分 III 的氨基酸物理化学特性(<http://www.genome.jp/aaindex/>);No,氨基酸的数目;CC,氨基酸的现在构成(某种类型的氨基酸数目除于氨基酸总数);FC,根据变异概率计算出的氨基酸未来构成(<http://www.nerc-nfb.ac.cn/calculation/fc.htm>);DP,氨基酸的分布概率(<http://www.nerc-nfb.ac.cn/calculation/dp.htm>)。

\* SNEP660103 is a physicochemical property of amino acid and describes the principal component III (<http://www.genome.jp/aaindex/>). No., Number of amino acids;CC, Current composition of amino acids calculated by the number of a type of amino acids divided by the total number of amino acids in an enzyme;FC, Future composition of amino acids calculated according to the mutating probability (<http://www.nerc-nfb.ac.cn/calculation/fc.htm>);DP, Distribution probability of amino acids that can be available at <http://www.nerc-nfb.ac.cn/calculation/dp.htm>.

其次,用20-1前馈反向传播的神经网络模型,完成535种属性对催化常数的拟合和预测。将55个 $\alpha$ -淀粉酶 Amy7C 及其突变体的数据分为2组,用35个酶作为训练组进行拟合,20个酶作为验证组进行检验。再对8种不同层次及不同个数的神经网络模型(表2)的拟合、预测效果进行比较。最后,将拟合、预测结果与实验数据进行统计学检验。详细研究方法见文献[14]中1.2~1.4。

## 2 结果与分析

### 2.1 20-1神经网络模型的预测结果

用20-1神经网络模型对55个数值化的 $\alpha$ -淀粉酶 Amy7C 及其突变体和其所对应的催化常数进行模型分析,并对535种氨基酸属性进行筛选。结果表明,110种属性可实现模型的收敛,包括2个氨基酸组成属性,43个理化性质属性,62个二级结构属性和3个动态氨基酸属性等指标。

由图1可见,氨基酸属性对 $\alpha$ -淀粉酶 Amy7C 催化常数预测效果的 $R$ 值和 $P$ 值的优劣顺序依次为训练组 $R$ 值、训练组 $P$ 值、验证组 $P$ 值和验证组 $R$ 值。110种氨基酸属性的训练组 $R$ 值平均值为0.90,86.4%的氨基酸属性预测值高于均值,其中,氨基酸分布概率属性预测值为0.991,与 $\alpha$ -淀粉酶 Amy7C 催化常数的相关性最强;1种氨基酸疏水性指标(VENT840101)的预测 $R$ 值为0.438,与 $\alpha$ -淀粉酶 Amy7C 催化常数的相关性最弱。110种氨基酸属性的训练组 $P$ 值平均值为0.9,78.2%的氨基酸属性预测值高于均值,47种属性预测的 $P$ 值高于0.99,其中,属性 QIAN880116 代表 $\beta$ -折叠在-4窗口位置的权重,其预测 $P$ 值达到1,能够完全拟合 $\alpha$ -淀粉酶 Amy7C 的催化常数;而位阻参数(属性 CHAM810101)预测 $P$ 值仅为0.292,与 $\alpha$ -淀粉酶 Amy7C 催化常数的拟合效果较差。110种氨基酸

表2 8种不同神经网络模型的构成

模型 Model	神经元数 Number of neurons					
	第1层 Layer 1	第2层 Layer 2	第3层 Layer 3	第4层 Layer 4	第5层 Layer 5	第6层 Layer 6
20-1	20	—	—	—	—	—
20-5-1	20	5	1	—	—	—
20-10-1	20	10	1	—	—	—
20-10-5-1	20	10	5	1	—	—
20-30-5-1	20	30	5	1	—	—
20-30-10-1	20	30	10	1	—	—
20-30-10-5-1	20	30	10	5	1	—
20-30-20-10-5-1	20	30	20	10	5	1

属性的验证组 $P$ 值平均值为0.83,65种氨基酸属性的验证值高于均值,其中,表观分配能量(属性 GUYH850105)的验证 $P$ 值为0.994,对 $\alpha$ -淀粉酶 Amy7C 催化常数的预测最好;优化的相对分区能量(属性 MIYS990102)的验证 $P$ 值仅为0.22,对 $\alpha$ -淀粉酶 Amy7C 催化常数的预测最差。110种氨基酸属性的验证组 $R$ 值平均值为0.28,60种氨基酸属性的验证值高于均值,其中,一种电荷转移能力参数(属性 CHAM830108)的验证组 $R$ 值为0.482,与 $\alpha$ -淀粉酶 Amy7C 催化常数的相关性较强;而另一种电荷转移能力参数(属性 CHAM830107)的验证组 $R$ 值为0.096,与 $\alpha$ -淀粉酶 Amy7C 的催化常数无相关性。

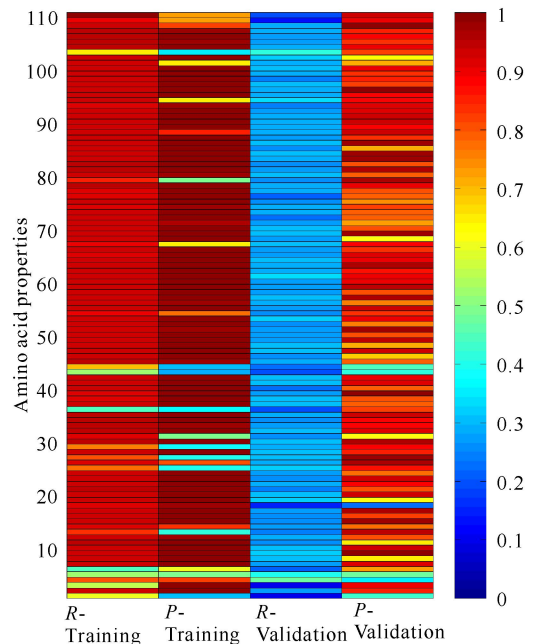


图1 用20-1神经网络模型预测淀粉酶 $K_{cat}$ 值的结果  
Fig.1 Results of predicting  $K_{cat}$  of amylases using 20-1 neural network

在使模型收敛的110种氨基酸属性中,属性 SNEP660103 代表主成分Ⅲ的化学结构,用于20-1

神经网络模型预测  $\alpha$ -淀粉酶 Amy7C 催化常数效果最好,其训练组与验证组  $R$  值和  $P$  值之和为 3.248;其次依次是属性 QIAN880104( $\alpha$ -螺旋在-3 窗口位置的权重)、属性 QIAN880106( $\alpha$ -螺旋在-1 窗口位置的权重)和属性 PUNT030101(以一维螺旋知识为基础的膜-倾向规模),训练组与验证组  $R$  值和  $P$  值之和分别为 3.211, 3.208 和 3.204。属性 CHAM810101(位阻参数)预测  $\alpha$ -淀粉酶 Amy7C 催化常数效果最差,其训练组与验证组  $R$  值和  $P$  值之和仅为 1.429,其次是氨基酸构成属性 GRAR740101 和平滑埃普西隆位阻参数(属性 FAUJ880102),训练组与验证组  $R$  值和  $P$  值之和分别为 1.445 和 1.673。氨基酸分布概率的训练组和验证组的  $P$  值和  $R$  值之和为 2.84,预测酶 Amy7C 的催化常数的效果较好。

## 2.2 不同神经网络模型的预测结果

图 2 为基于 2~6 层不同神经网络模型对  $\alpha$ -淀粉酶 Amy7C 酶催化常数的预测结果。可以看出,8 种属性的不同模型训练组  $R$  值结果具有显著性差异, $R$  值均值分布在 0.932~1,其中 20-30-5-1、20-30-10-1、20-30-10-5-1 和 20-30-20-10-5-1 模型预测的均值均为 1;20-1 模型预测值最低, $R$  值均值为 0.932。不同模型的训练组  $P$  值结果不具有显著性差异,20-1 模型预测的 8 种氨基酸属性的  $P$  值最高;20-30-5-1 模型预测的  $P$  值最低。不同模型的验证组  $R$  值结果也不具有显著性差异,20-10-5-1 模型预测的 8 种氨基酸属性的  $R$  值最高, $R$  值均值为 0.4;20-1 模型预测的  $R$  值最低, $R$  值均值为 0.297。不同模型的验证组  $P$  值结果具有显著性差异,20-1 模型预测的 8 种氨基酸属性的  $P$  值最高;20-30-10-5-1 模型预测的  $P$  值最低。

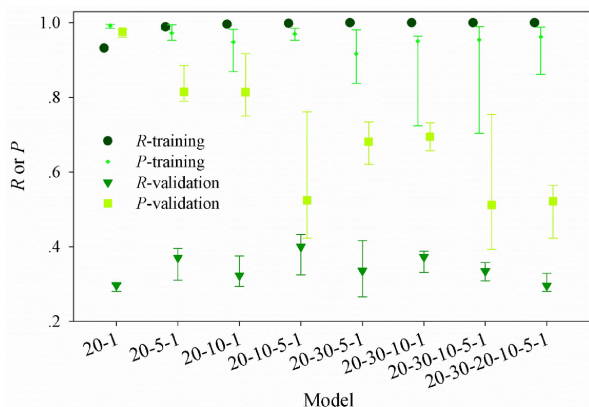


图 2 不同模型的预测结果

Fig. 2 Predicted results from different models

网络模型下预测  $\alpha$ -淀粉酶 Amy7C 催化常数的结果。可以看出,8 种属性的多模型分析的训练组  $R$  值、训练组  $P$  值、验证组  $R$  值和验证组  $P$  值结果均不具有显著性差异。属性 SNEP660103、VELV850101 和 QIAN880117 的训练组  $R$  值为 1,属性 QIAN880123 和现在构成的训练组  $R$  值为 0.99。属性 WERD780104 的训练组  $P$  值为 0.984,预测酶 Amy7C 的催化常数的效果最好;属性 VELV850101 的训练组  $P$  值为 0.776,预测酶 Amy7C 的催化常数的效果最差;现在构成的训练组  $P$  值为 0.968,预测酶 Amy7C 的催化常数的效果较好。8 种氨基酸属性多模型分析的验证组  $R$  值结果整体偏低,表明预测结果的相关性较低。属性 QIAN880117 的验证组  $P$  值为 0.772,预测酶 Amy7C 的催化常数的效果最好;现在构成属性的验证组  $P$  值为 0.595,预测酶 Amy7C 的催化常数的效果最差。

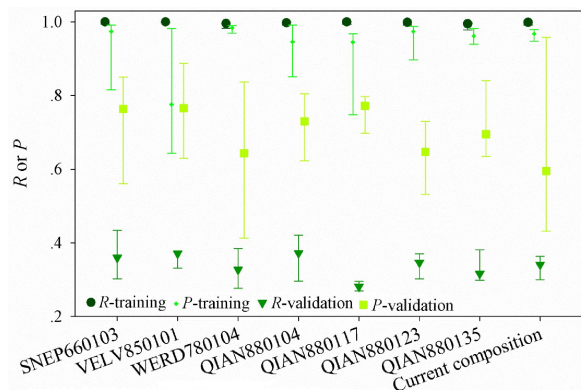


图 3 不同氨基酸属性的预测结果

Fig. 3 Predicted results from different amino acid properties

## 3 结论

本文利用 20-1 前馈反向传播的神经网络为模型,完成 535 种氨基酸属性对  $\alpha$ -淀粉酶 Amy7C 催化常数的拟合,并对 8 种不同层次及神经元个数的模型进行比较。结果:110 个氨基酸属性可实现 20-1 神经网络模型收敛,表明一些氨基酸属性可以用于预测  $\alpha$ -淀粉酶的催化常数;某些静态的氨基酸理化性质(如氨基酸主要成分)和二级结构属性(如  $\alpha$ -螺旋在不同窗口位置的权重)可影响  $\alpha$ -淀粉酶的酶促反应;而动态的氨基酸属性(如氨基酸的现在构成和分布概率)也可以预测  $\alpha$ -淀粉酶的催化常数,其中,计算变异学指标氨基酸分布概率在预测  $\alpha$ -淀粉酶的催化常数中显示出较好的效果,与以前的研究

图 3 为 8 种氨基酸属性分别在 8 种不同的神经

结果相一致<sup>[2~13]</sup>; 四层神经网络模型是预测  $\alpha$ -淀粉酶的催化常数相对理想的模型, 以 20-30-5-1 和 20-30-10-1 模型最佳。

#### 致谢:

感谢王成华博士提供  $\alpha$ -淀粉酶 Amy7C 及其突变体的催化常数数据。

#### 参考文献:

- [1] 罗志刚, 杨景峰, 罗发兴.  $\alpha$ -淀粉酶的性质及应用[J]. 食品研究与开发, 2007, 28(8): 163-167.  
Luo Z G, Yang J F, Luo F X. Properties and applications of  $\alpha$ -amylase[J]. Food Research and Development, 2007, 28(8): 163-167.
- [2] Yan S, Wu G. Searching of predictors to predict pH optimum of cellulases [J]. Appl Biochem Biotechnol, 2011, 165: 856-869.
- [3] Yan S, Wu G. Prediction of optimal pH in hydrolytic reaction of beta-glucosidase[J]. Appl Biochem Biotech A: Enzyme Eng Biotech, 2013, 169: 1884-94.
- [4] Yan S, Wu G. Prediction of optimal pH and temperature of cellulases using neural network [J]. Protein Pept Lett, 2012, 19: 29-39.
- [5] Yan S, Wu G. Predictors for predicting temperature optimum in beta-glucosidases[C]. Proceedings of the 6th International Symposium on Biocatalysis and Biosynthetic Engineering (iSBBE2013). November 11-13, 2013, Shanghai, China.
- [6] Yan S, Wu G. Exhausted jackknife validation exemplified by prediction of temperature optimum in enzymatic reaction of cellulases[J]. Appl Biochem Biotech A: Enzyme Eng Biotech, 2012, 166: 997-1107.
- [7] Yan S, Wu G. Prediction of temperature optimum in enzymatic reaction of beta-cellobiosidases with exhausted jackknife validation[J]. Life Sci J, 2013, 10(3): 2180-2189.
- [8] Yan S, 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.
- [9] Yan S, Wu G. Prediction of Michaelis-Menten constant in beta-cellobiosidase's reaction with lactoside as substrate[J]. Enzyme Eng, 2011, 1: 102. doi: 10.4172/eeg.1000102.
- [10] Yan S, Wu G. Prediction of Michaelis-Menten constant of cellulases using neural network[C]. Proceedings of the 2012 International Symposium on Advanced Biological Engineering (ISABE' 2012): 92-93. October 25~29, 2012, Guilin, China.
- [11] Yan S, Shi D, Nong H, et al. Predicting  $K_m$  values of beta-glucosidases using cellobiose as substrate[J]. Interdiscip Sci: Comput Life Sci, 2012, 4: 46-53.
- [12] Yan S, Wu G. Prediction of turnover number of cellulose 1, 4-beta-cellobiosidase[J]. Protein Pept Lett, 2013, 20: 255-64.
- [13] Yan S, Wu G. Predictions of enzymatic parameters: A mini-review with focus on enzymes for biofuel[J]. Appl Biochem Biotech A: Enzyme Eng Biotech, 2013, 171: 590-615.
- [14] 王成华. 基于  $\alpha$ -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.

(责任编辑: 尹 闯)