

· 工业生物技术 ·

毕赤酵母全蛋白质组信号肽及分泌蛋白分析

朱凌宣, 崔鑫, 胡善桐, 王诗卉, 张桂敏*

北京化工大学生命科学与技术学院, 北京 100029

朱凌宣, 崔鑫, 胡善桐, 王诗卉, 张桂敏. 毕赤酵母全蛋白质组信号肽及分泌蛋白分析[J]. 生物工程学报, 2024, 40(3): 834-846.

ZHU Lingxuan, CUI Xin, HU Shantong, WANG Shihui, ZHANG Guimin. Analysis of signal peptides and secreted proteins in the whole proteome of *Pichia pastoris*[J]. Chinese Journal of Biotechnology, 2024, 40(3): 834-846.

摘要: 信号肽是影响毕赤酵母异源蛋白分泌效率的关键因素之一。目前最常使用的信号肽是来自酿酒酵母的 α 交配因子(α -mating factor, MF α)的信号肽, 利用这个外源信号肽已经成功分泌表达了多种异源蛋白。然而, MF α 可能不适用于所有外源蛋白的分泌表达。毕赤酵母分泌蛋白中具有许多典型的信号肽序列, 这些内源信号肽可以提供除 MF α 信号肽之外更为广泛的信号肽选择。因此, 有必要分析预测所有内源信号肽引导外源蛋白分泌效果。本研究利用 SignalP、TMHMM、Phobius、WoLF PSORT 和 NetGPI 等生物信息工具对毕赤酵母 GS115 (ATCC 20864)的全蛋白质组进行内源蛋白信号肽及分泌特征的预测, 得到了信号肽分布、长度、氨基酸比例和保守性等规律, 同时筛选得到 69 个分泌蛋白及其信号肽, 并通过分泌蛋白组验证给出了 10 条可能用于外源蛋白表达的内源性信号肽。本研究获得的内源性信号肽或许为毕赤酵母表达外源蛋白提供了新的有力工具。

关键词: 毕赤酵母; 蛋白质组; 信号肽; 分泌蛋白; 生物信息学分析

资助项目: 国家重点研发计划(2021YFC2103901); 国家自然科学基金(31970059); 中央高校基本科研业务费专项资金(buctrc202131)

This work was supported by the National Key Research and Development Program of China (2021YFC2103901), the National Natural Science Foundation of China (31970059), and the Fundamental Research Funds for the Central Universities, China (buctrc202131).

*Corresponding author. E-mail: zhangguimin@mail.buct.edu.cn

Received: 2023-07-17; Accepted: 2023-11-22

Analysis of signal peptides and secreted proteins in the whole proteome of *Pichia pastoris*

ZHU Lingxuan, CUI Xin, HU Shantong, WANG Shihui, ZHANG Guimin*

College of Life Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China

Abstract: The signal peptide is a key factor that affects the efficiency of protein secretion in *Pichia pastoris*. Currently, the most used signal peptide is the α -mating factor (MF α) pre-pro leader from *Saccharomyces cerevisiae*. This exogenous signal peptide has been successfully utilized to express and secrete many heterologous proteins. However, MF α is not suitable for the secretory expression of all heterologous proteins. Many typical signal peptides are present in the secretory proteins of *P. pastoris*, which provides more options besides MF α . Therefore, it is necessary to analyze and identify more efficient endogenous signal peptides that can guide the secretion of heterologous proteins in *P. pastoris*. In this study, we employed bioinformatics tools such as SignalP, TMHMM, Phobius, WoLF PSORT, and NetGPI to predict endogenous signal peptides from the entire proteome of *P. pastoris* GS115 (ATCC 20864). Moreover, we analyzed the distribution, length, amino acid composition, and conservation of these signal peptides. Additionally, we screened 69 secreted proteins and their signal peptides, and through secretome validation, we identified 10 endogenous signal peptides that have potential to be used for exogenous protein expression. The endogenous signal peptides obtained in this study may serve as new valuable tools for the expression and secretion of heterologous proteins in *P. pastoris*.

Keywords: *Pichia pastoris*; proteome; signal peptide; secreted protein; bioinformatic analysis

毕赤酵母具有原核生物生长快、成本消耗低、可高密度发酵和基因操作简便等优点^[1], 同时具有典型的高等真核生物翻译后修饰的功能, 如信号肽(signal peptide, SP)加工、蛋白质折叠、二硫键形成、糖基化和酰化等^[2], 此外, 毕赤酵母能够将外源蛋白分泌到培养基中, 并且只有少量内源性分泌蛋白分泌, 使后期纯化过程较为简便^[3]。这些优势使毕赤酵母成为目前使用最广泛的蛋白表达系统之一, 如今, 利用该系统已生产了上千种蛋白。2009年, de Schutter等^[4]公布了毕赤酵母的全基因组, 其大小为9.43 Mbp, 并对其编码基因进行了人工注释。2011年, Küberl等^[5]结合454和Illumina测

序技术, 使毕赤酵母基因组序列精确到9.35 Mbp, 并第一次报道了该菌种的线粒体完整基因组序列。2016年Sturmberger等^[6]对毕赤酵母基因组的开放读框和染色体定位进行了更新, 使其更为精确。

信号肽是影响毕赤酵母异源蛋白分泌效率的关键因素之一。影响毕赤酵母高效表达重组蛋白的主要因素有基因拷贝数、密码子偏好、启动子、分子伴侣、信号肽、糖基化修饰和发酵工艺等^[7], 其中决定因素在于重组蛋白N端的信号肽, 其介导重组蛋白进入分泌途径。信号肽是一段位于蛋白质N端的3-30个氨基酸(amino acid, AA)的短肽, 广泛存在于真核和原

核生物中, 信号肽介导新生多肽跨膜转位到内质网腔, 在内质网中进行折叠和翻译后修饰, 随后被分配到目的地或胞外^[8]。SP 一般包括 3 个部分: (1) N 区: 带正电荷结构域; (2) H 区: 疏水核心; (3) C 区: 切割位点区^[9]。

目前在毕赤酵母系统中最常使用的信号肽是来自酿酒酵母的 α 交配因子(α -mating factor, MF α)的信号肽, 其结构和功能也得到了较为详细的研究。MF α 由 19 个氨基酸的 pre 区和 67 个氨基酸的 pro 区组成^[10]。利用 MF α 已成功表达了数百种蛋白, 如人中期因子^[11]、人溶解酵素^[12]、重组干扰素- α 2b^[13]等。然而, MF α 可能不适用于所有外源蛋白的分泌表达, 为解决这一问题, 研究人员对 MF α 进行了改造。Barrero 等^[14]通过结合 OST1 信号肽和 MF α 的 pro 区 Ser42 变体得到了新型信号肽, 显著提升了红色荧光蛋白在毕赤酵母中的分泌量。Ito 等^[15]对 MF α 进行了单点突变, 发现其中 8 个单点突变可以显著提高毕赤酵母中单链抗体的分泌量, 通过机制研究发现这些突变可以减少未折叠蛋白反应活化, 避免重组蛋白在分泌过程中进入细胞内蛋白质降解途径。

近年来, 越来越多的研究着眼于使用毕赤酵母内源信号肽表达外源蛋白。Liang 等^[16]通过 SignalP 4.0 预测了 3 个毕赤酵母内源信号肽(Scw、Dse 和 Exg), 利用这些信号肽有效表达了增强型绿色荧光蛋白(enhanced green fluorescent protein, EGFP)和南极假丝酵母脂肪酶 B (*Candida antarctica* lipase B, CALB), 除了 Scw-CALB 组, 其余组表达效果均优于 MF α 。Karaoglan 等^[17]分别将 MF α 、人血清白蛋白(human serum albumin, HSA)信号肽、毕赤酵母的酸性磷酸酶 PHO1 信号肽和毕赤酵母的内部重复蛋白 PIR1 信号肽用于黑曲霉木聚糖酶在毕赤酵母中的分泌表达, 结果显示内源信号肽 PIR1 有着最好的

分泌水平。Shen 等^[18]通过蛋白丰度与转录水平的比值分析毕赤酵母分泌蛋白的分泌强度, 筛选出比值较高的 *PAS_chr3_0030* 基因编码蛋白, 该蛋白的信号肽实现了脲化酶 NIT10 在毕赤酵母中的分泌, 但是 MF α 无法用于该酶的分泌。Duan 等^[19]将内源信号肽(Dan4、Gas1、Msb2 和 Fre2)用于毕赤酵母表达酵母增强绿色荧光蛋白、 β -半乳糖苷酶和头孢菌素 C 酰化酶, 其分泌效果均优于 MF α 。这些研究都表明毕赤酵母内源信号肽用于外源蛋白的表达有较好的效果。但是, 目前对于毕赤酵母 GS115 内源信号肽表达外源蛋白的研究大多通过实验筛选, 迄今该菌株还没有分析过信号肽的各种特征及含信号肽蛋白的跨膜区, 没有系统分析过能高效表达外源蛋白的内源信号肽。

本研究通过 SignalP、TMHMM、Phobius、WoLF PSORT 和 NetGPI 等多种生物信息工具对毕赤酵母 GS115 全基因组编码蛋白序列进行了生物信息学分析, 系统分析了该菌株信号肽的特征, 筛选得到了具有经典分泌特征的蛋白序列, 并进行了转录组分析与实验数据验证, 分析了可能用于高效表达的内源信号肽序列, 为后续利用内源信号肽表达外源蛋白提供了参考。

1 材料与方法

1.1 数据获取

利用 NCBI 获取毕赤酵母 GS115 全基因组(登录号: NC_012963.1、NC_012964.1、NC_012965.1、NC_012966.1)序列文件 sequence.txt, 将 sequence.txt 输入 SignalP 5.0 以获得成熟蛋白序列文件 output_mature.fasta 和蛋白类型文件 output_protein_type.txt, 输入 sequence.txt 到 SignalP 3.0 得到 N、H、C 区注释数据。将预测出含信号肽的蛋白序列输入 TMHMM 和 Phobius 得到跨膜结构域数量, 筛选含有跨膜区的序列,

将相同长度含信号肽的蛋白序列输入 Weblogo 获得氨基酸保守性图, 将删去含跨膜区和滞留信号的蛋白序列输入 WoLF PSORT 获得亚细胞定位数据, 然后将分泌到细胞外的蛋白序列输入到 NetGPI 1.1 获得糖基磷脂酰肌醇 (glycosylphosphatidylinositol, GPI) 锚定位点。各种预测工具所用网址见表 1。

1.2 分析方法

(1) 不同染色体分布分析: 将 SignalP 5.0 预测得到的成熟蛋白序列文件 output_mature.fasta 和编码蛋白序列的文件 sequence.txt 用 Python 代码转化为 xlsx 文件, 随后将 SignalP 5.0 预测得到的 output_protein_type.txt 文件用 Excel 打开并进行信号肽序列筛选, 得到含有信号肽的序列, 通过比对函数找到编码蛋白序列, 用替代函数将编码蛋白序列中的信号肽序列删去得到成熟蛋白序列, 并对含信号肽蛋白在 4 条染色体上的分布进行统计。

(2) N、H、C 区分析: 将得到的含有信号肽的蛋白列表通过 EmEditor 转化为 fasta 格式, 输入到 SignalP 3.0 进行预测, 将预测后的网页链接结果通过重新命名转成一系列 txt 文

件, 随后通过 VBA 批量转换为 xlsx 文件, 接着用 VBA 批量进行格式规范。通过比较 N、H、C 区的概率对信号肽进行区域划分, 最后通过合并一系列的 xlsx 文件得到整合后的结果。

(3) Weblogo 保守性分析: 将统计得到的信号肽按相同长度进行分组, 每组序列利用 Weblogo 进行保守性分析。Excel 合并各蛋白序列信号肽后 10 个氨基酸和成熟蛋白前 10 个氨基酸, 通过 Weblogo 分析信号肽酶切点前后共 20 个氨基酸保守性。通过分析结果画出序列标识图。

(4) 跨膜区和内质网滞留蛋白分析: 将 TMHMM2.0 预测得到的跨膜结构域数量 PredHel 整合到信号肽数据表格中, 筛选分析后将含 1 个跨膜区的蛋白序列用 Phobius 进行进一步验证, 筛选出不含跨膜区 TRANSMEM 的序列补充到无跨膜区蛋白序列中, 然后再筛选不含 C 端-DEL 滞留信号的蛋白序列。

(5) 亚细胞定位分析: 将 WoLF PSORT 预测得到的亚细胞定位数据整合到筛选跨膜区与内质网滞留蛋白后的表中, 筛选分析亚细胞定位 extr 可能性最高的蛋白序列。

表 1 数据获取网址及功能

Table 1 Data acquisition website and function

Name	Website	Function
SignalP 5.0	https://services.healthtech.dtu.dk/service.php?SignalP-5.0	Batch prediction of potential signal peptides in protein sequences
SignalP 3.0	https://services.healthtech.dtu.dk/service.php?SignalP-3.0	Batch annotation prediction of N, H, and C regions of Signal Peptides
TMHMM 2.0	https://services.healthtech.dtu.dk/service.php?TMHMM-2.0	Batch prediction of transmembrane helical domains of proteins
Phobius	https://phobius.sbc.su.se/index.html	Batch validation screened transmembrane regions
WoLF PSORT	https://wolffpsort.hgc.jp/	Batch prediction of subcellular localization
NetGPI 1.1	https://services.healthtech.dtu.dk/services/NetGPI-1.1/	Batch prediction of GPI anchor
Weblogo	https://weblogo.berkeley.edu/logo.cgi	Conservation analysis of signal peptide sequences with the same length

(6) GPI 锚定位点分析: 通过 NetGPI 1.1 预测得到含 GPI 锚定位点的蛋白序列, 将这些序列从亚细胞定位分析后的结果中排除, 得到分泌蛋白预测结果。

(7) 绘图: 通过 Origin 绘图软件对用 Excel 统计后的数据进行绘图。

2 结果与分析

2.1 毕赤酵母 GS115 菌株含信号肽蛋白分析和在不同染色体上的占比

毕赤酵母 GS115 菌株是使用最广泛的宿主, 其由 Cregg 等^[20]于 1985 年构建, 是一种组氨酸脱氢酶基因缺陷型菌株, 便于筛选阳性转化子。本研究利用信号肽预测工具 SignalP 5.0 对毕赤酵母 GS115 的基因组进行分析, 预测结果表明, 在总编码序列(coding sequence, CDS)编码的 5 040 个蛋白中, 有 248 个蛋白具有潜在的 SP, 同时对潜在信号肽蛋白在各自染色体上的比例进行了统计分析, 结果显示编码蛋白具有 SP 的基因在染色体上占比范围几乎一致, 一般在 4%–6% (图 1), 没有染色体偏爱性。

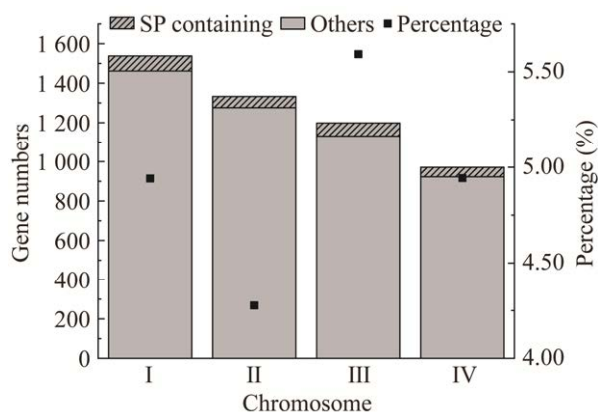


图 1 毕赤酵母 GS115 四条染色体上含信号肽蛋白编码基因比例

Figure 1 Proportion of genes encoding signal peptide proteins on the four chromosomes of *Pichia pastoris* GS115.

2.2 信号肽长度统计和 N、H、C 区特征分析

利用 Excel 统计 SP 长度, 结果显示这些 SP 大多具有 16–24 AA, 平均长度为 20 AA, 最长的 SP 为 37 AA, 最短的为 12 AA, 长度分布见图 2, Choo 等^[21]选取了 1 877 条来自不同真核生物的含信号肽序列进行统计, 发现真核生物 SP 的平均长度为 22 AA, 这与本研究预测酵母基因组中 SP 的平均长度一致。

通过 SignalP 3.0 注释预测得到的 248 个潜在的 SP, 统计得到其 N、H、C 区的长度(图 3)及氨基酸占比(图 4)。结果表明这些预测信号肽的 N 区长度一般为 3–6 AA, 也有一些极长的 N 区(23 AA)。H 区的长度为 8–11 AA, 最长的 H 区只有 13 AA, 氨基酸数量相对有限。C 区的长度相对一致, 其中 75.82% 的 C 区长度为 5 AA。

将氨基酸根据特性进行划分, 分析发现 N 区的氨基酸以带正电氨基酸和脂肪支链氨基酸为主, 所带正电氨基酸比例(23.99%)远远高于 H 区(0.89%)和 C 区(5.08%), 其中精氨酸和赖氨酸

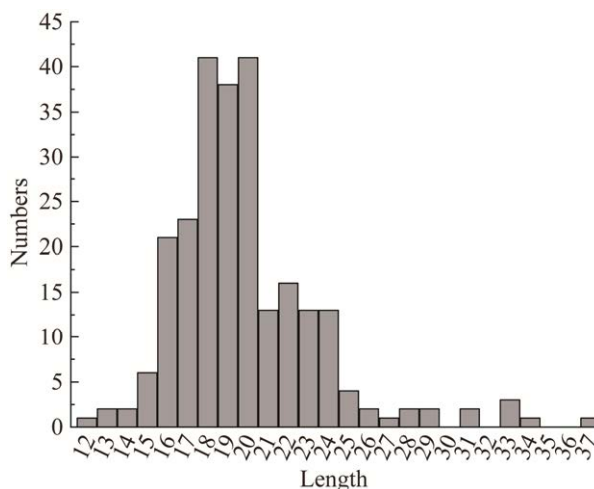


图 2 毕赤酵母 GS115 信号肽长度

Figure 2 Length of *Pichia pastoris* GS115 signal peptides.

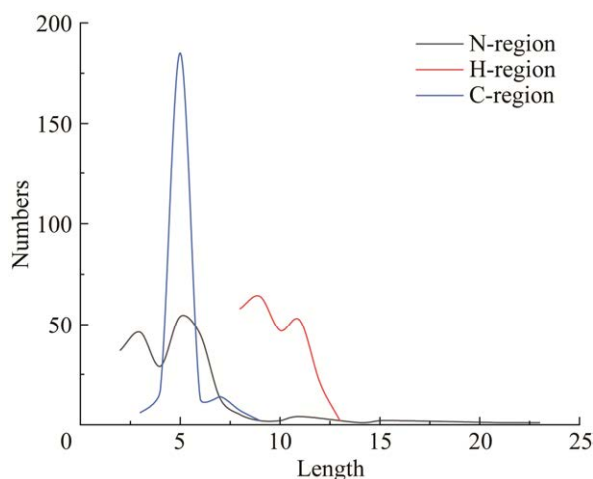


图3 毕赤酵母 GS115 预测信号肽 N、H、C 区长度统计

Figure 3 Statistics on the length of the N, H, and C regions of predicted signal peptide in *Pichia pastoris* GS115.

占比达到了 91.15%。H 区以疏水性氨基酸为主，其中亮氨酸占比达到 33.8%。相比于 N 区和 H 区，C 区由极性和不带电氨基酸组成，比如 C 区中丙氨酸占 22.5%，丝氨酸占 13.7%。

2.3 信号肽保守氨基酸分析

利用 WEBLOGO 网站对预测的信号肽进行保守性分析(图 5)，根据分析，毕赤酵母信号

肽 C 末端存在 21.37% 的 AXA 和 20.97% 的 VXA 的基序。同时在 H 区可以看到大量保守的亮氨酸位点，其中以 N 端第 8 位氨基酸最保守，该位置亮氨酸保守性为 39.75%，保守性第二的苯丙氨酸仅为 10.66%。

2.4 信号肽酶切点附近氨基酸占比分析

用 WEBLOGO 对切点前后 20 个氨基酸进行分析(图 6)，假设切割位点为 P_0 ，其 N 端方向为负，C 端方向为正， P_{-1} – P_{-3} 有 AXA/VXA 保守基序，其中 P_{-1} 位点丙氨酸出现比例最高 (58.06%)，同时， P_{-1} 位置也存在一定比例的甘氨酸(16.53%)和丝氨酸(14.11%)。 P_{-3} 位点中丙氨酸和缬氨酸出现频率相似，分别为 32.66% 和 35.48%。 P_{-6} – P_{-10} 位置中亮氨酸出现较多，达到了 33.39%。对 P_{-1} – P_{-10} 氨基酸比例的分析显示， P_0 后的成熟氨基酸并没有保守特征。

2.5 分泌蛋白的跨膜区和内质网滞留信号分析

有研究表明具有分泌型信号肽的蛋白质如果同时含有跨膜区，其可能是膜受体、锚定蛋白或者离子通道蛋白^[22]。为了排除本研究筛选内源性分泌信号肽中跨膜蛋白的干扰，本研究

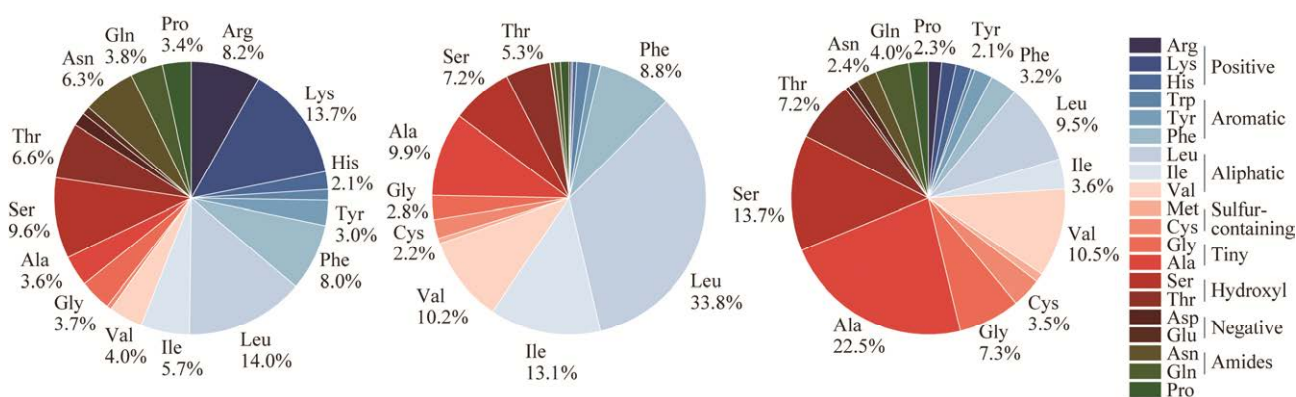


图4 信号肽 N、H、C 区(从左至右)氨基酸占比分析

Figure 4 Analysis of amino acid proportions in the N, H, and C regions of signal peptides (from left to right).

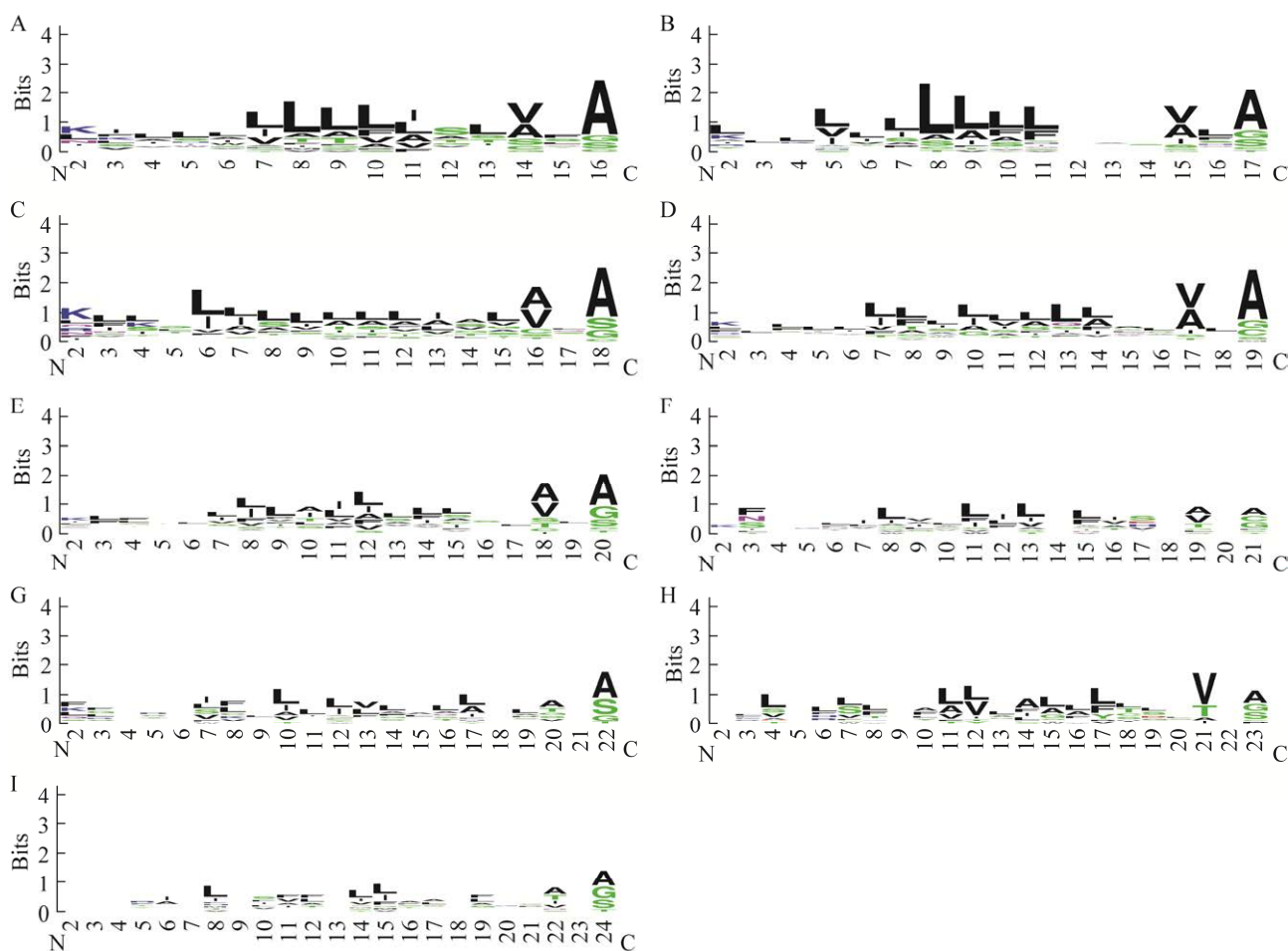


图 5 毕赤酵母 GS115 信号肽保守性分析(从 N 端第 2 个氨基酸开始标记) A-I: 16-24 AA

Figure 5 Conservation analysis of signal peptides in *Pichia pastoris* GS115 starts from the second amino acid at the N-terminal. A-I: 16-24 AA.

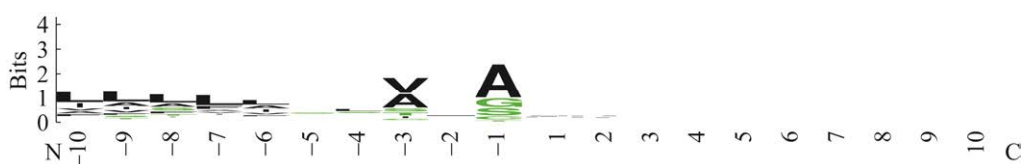


图 6 毕赤酵母 GS115 切点附近 P₋₁₀-P₁₀ 保守性分析

Figure 6 Conservation analysis of P₋₁₀-P₁₀ near the cleavage site of *Pichia pastoris* GS115.

利用 TMHMM 2.0 分析上述预测内源信号肽的跨膜区, 将预测得到的蛋白质跨膜螺旋数量与信号肽预测的数据进行整合, 结果显示在 248 个预测信号肽序列中有 47 个蛋白含有 2 个及以上跨膜结构域, 60 个蛋白含有 1 个跨膜结构

域, 141 个蛋白无跨膜结构域。此外, 由于 TMHMM 2.0 不能区分信号肽与跨膜结构域, 容易造成假阴性, 因此, 利用 Phobius 接着筛选只含 1 个跨膜结构域的 60 个蛋白序列, 结果显示其中有 21 个蛋白不含跨膜区, 最终本研究

得到 162 个无跨膜区且含有信号肽的蛋白序列。

酿酒酵母中内质网驻留蛋白的 C 端含有 -DEL 标记的滞留信号序列可以帮助蛋白滞留在内质网^[23]，筛选出来的 162 个蛋白序列中有 9 个蛋白序列含有 C 端滞留信号(表 2)。

2.6 亚细胞定位和 GPI 锚定分析

通过 WoLF PSORT 对具有信号肽的 153 个分泌型蛋白进行亚细胞定位预测分析，结果显示 119 条蛋白序列具有胞外定位特征，其余 34 条蛋白序列不分泌到胞外而是转移至各亚细胞器中，包括细胞质、内质网、高尔基体、线粒体、细胞核、过氧化物酶体和细胞膜。

GPI 锚定多见于真核生物转录后修饰，若含有分泌信号的蛋白中存在 GPI 锚定位点，其将不能分泌到细胞外^[24]。通过 NetGPI 1.1 对亚细胞定位分析后的蛋白序列进行 GPI 锚定位点分析，结果显示有 50 个蛋白序列含有 GPI 锚定位点，最终筛选得到 69 个具有经典分泌特征的蛋白。

2.7 实验数据验证

RNA-seq 基于第二代测序技术被用于揭示基因组中 RNA 特定时刻的存在和数量，是一种常用的转录组学研究方法^[25]。在 RNA-seq 分析

中，每千个碱基每百万次读取的读取数(reads per kilo bases per million reads, RPKM)是一种兼顾测序深度和基因长度影响的转录本丰度衡量参数，RPKM 越高，基因转录表达水平越高。RPKM 的衡量标准：总 RPKM 值>100 属于高表达；10-100 属于中度表达；<10 属于本底表达^[26]。Xu 等^[27]为了筛选毕赤酵母中高效启动子，基于 RNA-seq 计算了不同碳源下各基因的 RPKM。基于其数据，结合上述筛选的信号肽数据，分析了 69 个预测得到的分泌蛋白序列，发现以葡萄糖为碳源时，有 39 个高表达蛋白、26 个中度表达蛋白、4 个本底表达蛋白的序列；在以甘油为碳源时，有 33 个高表达蛋白、34 个中度表达蛋白、2 个本底表达蛋白的序列；以甲醇为碳源时，有 23 个高表达蛋白、45 个中度表达蛋白、1 个本底表达蛋白的序列。

目前有许多关于毕赤酵母内源信号肽介导外源蛋白高效表达的研究(表 3)，通过统计比较发现，在 21 个已报道的介导外源蛋白表达的内源信号肽中，有 12 个是本研究预测得到的分泌蛋白信号肽，且全部具有中转录和高转录表达水平。

表 2 含-DEL 滞留信号的蛋白

Table 2 Proteins with -DEL retention signal

Gene	C-terminal	Description
<i>PAS_chr1-3_0063</i>	HDEL	Molecular chaperone of the endoplasmic reticulum lumen
<i>PAS_chr1-3_0125</i>	HDEL	Member of the protein disulfide isomerase (PDI) family
<i>PAS_chr2-1_0140</i>	HDEL	ATPase involved in protein import into the ER, also acts as a chaperone to mediate protein folding in the ER and may play a role in ER export of soluble proteins
<i>PAS_chr3_0929</i>	HDEL	Protein required for β -1,6 glucan biosynthesis
<i>PAS_chr4_0844</i>	HDEL	Protein disulfide isomerase, multifunctional protein resident in the endoplasmic reticulum lumen
<i>PAS_chr1-1_0267</i>	KDEL	Peptidyl-prolyl <i>cis-trans</i> isomerase (cyclophilin) of the endoplasmic reticulum
<i>PAS_chr3_0179</i>	KDEL	Glucosidase II beta subunit
<i>PAS_chr1-3_0174</i>	RDEL	One of several homologs of bacterial chaperone DnaJ, located in the ER lumen
<i>PAS_chr1-1_0237</i>	SDEL	Nucleotide exchange factor for the endoplasmic reticulum (ER) luminal Hsp70 chaperone Kar2p

表3 介导外源蛋白在毕赤酵母中表达的内源信号肽

Table 3 Endogenous signal peptide mediating the expression of exogenous proteins in *Pichia pastoris*

Signal peptide	Gene	Reporter protein	RPKM _{glucose}	RPKM _{glycerin}	RPKM _{methanol}	Predicted as secreted protein here (Yes/No)	References
SP23	<i>PAS_chr1-1_0160</i>	Recombinant human	274.60	348.63	102.67	No	[28]
SP24	<i>PAS_c131_0001</i>	growth hormone	107.11	156.52	74.57	Yes	
SP26	<i>PAS_chr4_0040</i>		494.93	597.38	181.99	Yes	
SP34	<i>PAS_chr1-1_0293</i>		180.48	411.24	746.19	No	
Scw11	<i>PAS_chr2-1_0052</i>	Xylanase, mannanase	240.31	137.42	96.09	Yes	[29]
Dse4	<i>PAS_chr1-1_0130</i>		258.61	136.71	50.80	Yes	
Exg1	<i>PAS_chr2-1_0454</i>		544.71	355.22	257.89	Yes	
SCW10	<i>PAS_chr1-3_0229</i>	Laccase	568.35	511.69	273.78	Yes	[30]
FLO10	<i>PAS_chr1-4_0584</i>		55.73	54.80	46.00	Yes	
BGL2	<i>PAS_c0305hr1-4_0426</i>		806.37	1 151.42	498.21	Yes	
SIM1	<i>PAS_chr2-2_0064</i>		569.16	597.00	295.00	Yes	
PHO5	<i>PAS_chr2-1_0103</i>		197.80	74.19	38.51	Yes	
OST1	<i>PAS_chr3_0741</i>	Red fluorescent protein	154.00	267.92	53.72	No	[14]
CPR5	<i>PAS_chr1-1_0267</i>	Enhanced green fluorescent protein,	676.44	265.53	90.15	No	[31]
PRY2	<i>PAS_chr3_0076</i>	<i>Rhizomucor miehei</i>	111.18	102.11	94.34	Yes	
NUP145	<i>PAS_chr4_0643</i>	lipase	102.60	238.01	93.39	Yes	
FLO9	<i>PAS_chr3_1145</i>		1 657.42	693.84	142.22	No	
Gas1	<i>PAS_chr1-3_0227</i>	Yeast-enhanced green fluorescent protein,	626.24	430.65	372.82	No	[19]
Dan4	<i>PAS_chr2-1_0172</i>	β -galactosidase,	52.17	55.94	50.90	No	
Msb2	<i>PAS_chr1-3_0276</i>	cephalosporin C acylase	139.08	195.45	215.81	No	
Fre2	<i>PAS_chr4_0399</i>		22.91	47.14	111.16	No	

为了验证预测效果,基于田小梅等^[32]关于不同碳源下毕赤酵母 GS115 蛋白组学的分析数据,将筛选得到的 69 个分泌蛋白与其进行比较,结果显示在以葡萄糖或甘油为碳源的培养基中,胞外含量最高的前 35 个胞外蛋白有 10 个与预测结果一致(表 4),其中胞外含量最高的前 5 个含信号肽蛋白中有 4 个与预测结果一致,说明预测效果较好。表 4 中的 10 个信号肽已有 6 个被用于外源蛋白的表达,其中 Exg1 在表达南极念珠菌脂肪酶 B 时效果优于 MF α ^[16],因此,推测这 10 个信号肽可能是表达外源蛋白较好

的选择。

3 讨论

SP 是分泌途径中帮助蛋白质正确定位的主要成分之一,获得高效分泌效率的 SP 对于毕赤酵母蛋白表达系统的优化十分重要,近年来,很多研究提出使用内源性分泌 SP 更有利于分泌表达外源蛋白,本研究对毕赤酵母的 SP 进行了全基因组分析,揭示了其结构特征。在毕赤酵母 GS115 中,染色体上含有 SP 的基因分布在 4%–6%之间,大多数 SP 的长度为 16–24 AA,

表 4 毕赤酵母中可能用于表达外源蛋白的内源信号肽(N、H、C 区通过斜线划分)

Table 4 Endogenous signal peptides in *Pichia pastoris* that may be used to express exogenous proteins (N, H, and C regions are divided by slash)

Gene	Signal peptide
PAS_chr1-3_0229	MQVKS/IVNLLLAC/SLAVA
PAS_chr2-1_0454	MN/LYLITLLFA/SLCSA
PAS_chr1-4_0426	MIFNLK/TLAAVAISI/SQVSA
PAS_chr2-1_0887	MYRN/LIATAALT/CGAYS
PAS_chr2-1_0052	MLST/ILNIFILLFI/QASLQ
PAS_chr4_0046	MKS/QLIFMALAS/LVAS
PAS_chr2-2_0064	MKI/SALTACAVTLA/GLAIA
PAS_chr4_0305	MK/LAALSTIALTIL/PVALA
PAS_chr3_0960	MF/WLLVLSLI/SQALA
PAS_chr3_0337	MFNLK/TILISTLA/SIAVA

N-region: The positively charged N-terminal part of the signal peptide; H-region: The hydrophobic core; C-region: The cleavage site.

其 3 个区域具有不同的长度, SP 的 N 区含有丰富的带正电氨基酸和脂肪支链氨基酸, 这可以帮助蛋白质更好易位, 有研究表明 N 区缺少正电荷氨基酸会导致易位受损并对细胞有损伤^[33]。H 区长度为 8–11 AA, 以疏水氨基酸为主, 有较多的亮氨酸保守位点, 该区域的疏水性降低不利于蛋白加工和易位^[34], 过长的 H 区会影响信号肽酶切割效率^[35]。本研究结果也发现大部分 SP 的 C 区仅有 5 个 AA 组成, 这可能是信号肽酶处理的恰当距离, 多于或少于该长度会导致切割效率的降低^[36]。另外, 毕赤酵母的 SP 的 C 区出现了典型 AXA 基序, 该基序有利于信号肽和信号肽酶相互作用^[37]。同时也在毕赤酵母 SP 的 C 区中发现了 VXA 基序(20.97%)。毕赤酵母内源信号肽的染色体分布、长度、氨基酸分布和信号肽基序等特征与对酿酒酵母内源信号肽分析结果^[36]基本一致。对毕赤酵母 GS115 的分泌蛋白进行了预测分析, 本研究预测得到的分泌蛋白信号肽与已经报道的重组蛋白表达内源

信号肽有部分吻合, 且基本属于高表达和中度表达水平, 具有一定的参考价值, 预测结果可以为内源信号肽表达外源蛋白提供候选序列, 也发现有部分不具有分泌特征的蛋白信号肽被用于外源蛋白的表达, 说明重组蛋白分泌与介导的外源蛋白序列有一定关系。虽然分析结果显示信号肽酶切割位点之后的序列没有明显的特征, 但是 Tian 等^[38]通过机器学习对毕赤酵母中分泌蛋白进行研究, 结果发现蛋白 N 端序列相较于 C 端序列对于分泌影响更大, 原因可能是 N 端序列会影响分泌途径中 MF α 的识别^[39]。

4 结论

本研究通过 SignalP、TMHMM、Phobius、WoLF PSORT 和 NetGPI 等预测工具, 对毕赤酵母 GS115 全基因组编码蛋白序列进行了生物信息学分析, 得到了该菌株含有的信号肽的蛋白序列, 并分析这些信号肽的分布、长度、区域特征及切点附近氨基酸比例和保守位点等规律, 验证了诸多真菌信号肽的特征, 同时对含有信号肽的序列进行了跨膜区分析、内质网滞留蛋白的分析、亚细胞定位分析和 GPI 锚定位点分析, 共筛选得到了 69 个具有经典分泌特征的蛋白序列, 通过对比已报道的实验数据验证了预测结果, 选出了 10 个较优的内源信号肽。本研究为找到合适的内源性信号肽用于毕赤酵母表达外源蛋白提供了数据基础, 同时也有利于毕赤酵母分泌机制的研究。

REFERENCES

- [1] 吴静雯. α -factor 信号肽及其 N-糖基化修饰对毕赤酵母外源蛋白分泌的影响[D]. 福州: 福建师范大学硕士学位论文, 2019.
WU JW. Effects of α -factor signal peptides and its N-glycosylation on secretion of heterologous proteins in *Pichia pastoris*[D]. Fuzhou: Master's Thesis of

- Fujian Normal University, 2019 (in Chinese).
- [2] KARBALAEI M, REZAEI SA, FARSIANI H. *Pichia pastoris*: a highly successful expression system for optimal synthesis of heterologous proteins[J]. Journal of Cellular Physiology, 2020, 235(9): 5867-5881.
- [3] DAMASCENO LM, HUANG CJ, BATT CA. Protein secretion in *Pichia pastoris* and advances in protein production[J]. Applied Microbiology and Biotechnology, 2012, 93(1): 31-39.
- [4] de SCHUTTER K, LIN YC, TIELS P, van HECKE A, GLINKA S, WEBER-LEHMANN J, ROUZÉ P, van de PEER Y, CALLEWAERT N. Genome sequence of the recombinant protein production host *Pichia pastoris*[J]. Nature Biotechnology, 2009, 27(6): 561-566.
- [5] KÜBERL A, SCHNEIDER J, THALLINGER GG, ANDERL I, WIBBERG D, HAJEK T, JAENICKE S, BRINKROLF K, GOESMANN A, SZCZEPANOWSKI R, PÜHLER A, SCHWAB H, GLIEDER A, PICHLER H. High-quality genome sequence of *Pichia pastoris* CBS7435[J]. Journal of Biotechnology, 2011, 154(4): 312-320.
- [6] STURMBERGER L, CHAPPELL T, GEIER M, KRAINER F, DAY KJ, VIDE U, TRSTENJAK S, SCHIEFER A, RICHARDSON T, SORIAGA L, DARNHOFER B, BIRNER-GRUENBERGER R, GLICK BS, TOLSTORUKOV I, CREGG J, MADDEN K, GLIEDER A. Refined *Pichia pastoris* reference genome sequence[J]. Journal of Biotechnology, 2016, 235: 121-131.
- [7] 王庆华, 高丽丽, 梁会超, 巩婷, 杨金玲, 朱平. 影响毕赤酵母高效表达重组蛋白的主要因素及其研究进展[J]. 药学报, 2014, 49(12): 1644-1649.
WANG QH, GAO LL, LIANG HC, GONG T, YANG JL, ZHU P. Research advances of the influence factors of high level expression of recombinant protein in *Pichia pastoris*[J]. Acta Pharmaceutica Sinica, 2014, 49(12): 1644-1649 (in Chinese).
- [8] RUTKOWSKI DT, OTT CM, POLANSKY JR, LINGAPPA VR. Signal sequences initiate the pathway of maturation in the endoplasmic reticulum lumen[J]. Journal of Biological Chemistry, 2003, 278(32): 30365-30372.
- [9] LIACI AM, FÖRSTER F. Take me home, protein roads: structural insights into signal peptide interactions during ER translocation[J]. International Journal of Molecular Sciences, 2021, 22(21): 11871.
- [10] LIN-CEREHINO GP, STARK CM, KIM D, CHANG J, SHAHEEN N, POERWANTO H, AGARI K, MOUA P, LOW LK, TRAN N, HUANG AD, NATTESTAD M, OSHIRO KT, CHANG JW, CHAVAN A, TSAI JW, LIN-CEREHINO J. The effect of α -mating factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*[J]. Gene, 2013, 519(2): 311-317.
- [11] MURASUGI A, TOHMA-AIBA Y. Comparison of three signals for secretory expression of recombinant human midkine in *Pichia pastoris*[J]. Bioscience, Biotechnology, and Biochemistry, 2001, 65(10): 2291-2293.
- [12] OKA C, TANAKA M, MURAKI M, HARATA K, SUZUKI K, JIGAMI Y. Human lysozyme secretion increased by alpha-factor pro-sequence in *Pichia pastoris*[J]. Bioscience, Biotechnology, and Biochemistry, 1999, 63(11): 1977-1983.
- [13] GHOSALKAR A, SAHAI V, SRIVASTAVA A. Secretory expression of interferon-alpha 2b in recombinant *Pichia pastoris* using three different secretion signals[J]. Protein Expression and Purification, 2008, 60(2): 103-109.
- [14] BARRERO JJ, CASLER JC, VALERO F, FERRER P, GLICK BS. An improved secretion signal enhances the secretion of model proteins from *Pichia pastoris*[J]. Microbial Cell Factories, 2018, 17(1): 161.
- [15] ITO Y, ISHIGAMI M, HASHIBA N, NAKAMURA Y, TERAI G, HASUNUMA T, ISHII J, KONDO A. Avoiding entry into intracellular protein degradation pathways by signal mutations increases protein secretion in *Pichia pastoris*[J]. Microbial Biotechnology, 2022, 15(9): 2364-2378.
- [16] LIANG SL, LI C, YE YR, LIN Y. Endogenous signal peptides efficiently mediate the secretion of recombinant proteins in *Pichia pastoris*[J]. Biotechnology Letters, 2013, 35(1): 97-105.
- [17] KARAOGLAN M, YILDIZ H, INAN M. Screening of signal sequences for extracellular production of *Aspergillus niger* xylanase in *Pichia pastoris*[J]. Biochemical Engineering Journal, 2014, 92: 16-21.
- [18] SHEN Q, ZHOU XT, GUO Q, XUE YZ, XUE YP, ZHENG YG. Potential of the signal peptide derived from the *PAS_chr3_0030* gene product for secretory expression of valuable enzymes in *Pichia pastoris*[J]. Applied and Environmental Microbiology, 2022, 88(9):

- e0029622.
- [19] DUAN GD, DING LM, WEI DS, ZHOU HC, CHU J, ZHANG SL, QIAN JC. Screening endogenous signal peptides and protein folding factors to promote the secretory expression of heterologous proteins in *Pichia pastoris*[J]. Journal of Biotechnology, 2019, 306: 193-202.
- [20] CREGG JM, BARRINGER KJ, HESSLER AY, MADDEN KR. *Pichia pastoris* as a host system for transformations[J]. Molecular and Cellular Biology, 1985, 5(12): 3376-3385.
- [21] CHOO KH, RANGANATHAN S. Flanking signal and mature peptide residues influence signal peptide cleavage[J]. BMC Bioinformatics, 2008, 9: 1-11.
- [22] 张光荣, 袁再美, 孙丽媛, 邓茗月, 王荣蛟, 和世春, 王寒雪, 毛华明, 杨舒黎. 大额牛瘤胃细菌 CMC-1 蛋白结构和功能的生物信息学分析[J]. 黑龙江畜牧兽医, 2021(3): 16-21, 157.
- ZHANG GR, YUAN ZM, SUN LY, DENG MY, WANG RJ, HE SC, WANG HX, MAO HM, YANG SL. Bioinformatics analysis of protein structure and function of CMC-1 from rumen bacteria of Gayal (*Bos frontalis*)[J]. Heilongjiang Animal Science and Veterinary Medicine, 2021(3): 16-21, 157 (in Chinese).
- [23] MEI M, ZHAI C, LI XZ, ZHOU Y, PENG WF, MA LX, WANG QH, IVERSON BL, ZHANG GM, YI L. Characterization of aromatic residue-controlled protein retention in the endoplasmic reticulum of *Saccharomyces cerevisiae*[J]. Journal of Biological Chemistry, 2017, 292(50): 20707-20719.
- [24] 王建国, 赵南南, 何志云, 王泽. 白假丝酵母菌 (*Candida albicans*) WO-1 分泌蛋白质组的生物信息学分析[J]. 沈阳师范大学学报(自然科学版), 2020, 38(5): 463-467.
- WANG JG, ZHAO NN, HE ZY, WANG Z. Bioinformatics analysis of secretome in *Candida albicans* WO-1[J]. Journal of Shenyang Normal University (Natural Science Edition), 2020, 38(5): 463-467 (in Chinese).
- [25] CHU YJ, COREY DR. RNA sequencing: platform selection, experimental design, and data interpretation[J]. Nucleic Acid Therapeutics, 2012, 22(4): 271-274.
- [26] YUAN J, TAO WJ, CHENG YY, HUANG BF, WANG DS. Genome-wide identification, phylogeny, and gonadal expression of *fox* genes in Nile tilapia, *Oreochromis niloticus*[J]. Fish Physiology and Biochemistry, 2014, 40(4): 1239-1252.
- [27] XU N, ZHU JX, ZHU QY, XING YZ, CAI MH, JIANG TY, ZHOU M, ZHANG YX. Identification and characterization of novel promoters for recombinant protein production in yeast *Pichia pastoris*[J]. Yeast, 2018, 35(5): 379-385.
- [28] MASSAHI A, ÇALIK P. Endogenous signal peptides in recombinant protein production by *Pichia pastoris*: from *in-silico* analysis to fermentation[J]. Journal of Theoretical Biology, 2016, 408: 22-33.
- [29] 杨青, 汪斌, 王亚伟, 张华山, 熊海容, 张莉. 介导两种半纤维素酶分泌表达的信号肽比较[J]. 中国生物工程杂志, 2017, 37(8): 15-22.
- YANG Q, WANG B, WANG YW, ZHANG HS, XIONG HR, ZHANG L. Comparison of signal peptides for two hemicellulase secretory expression[J]. China Biotechnology, 2017, 37(8): 15-22 (in Chinese).
- [30] 石艺平, 周雪, 胡美荣, 林剑辉, 陶勇, 黄建忠. 不同信号肽对毕赤酵母表达漆酶的影响[J]. 微生物学报, 2014, 54(12): 1446-1452.
- SHI YP, ZHOU X, HU MR, LIN JH, TAO Y, HUANG JZ. Effect of signal peptides on the expression of laccase in *Pichia pastoris*[J]. Acta Microbiologica Sinica, 2014, 54(12): 1446-1452 (in Chinese).
- [31] 苗杨利, 李站胜, 韩双艳. 促进米黑根毛霉脂肪酶分泌的毕赤酵母内源信号肽的筛选[J]. 现代食品科技, 2019, 35(10): 155-163.
- MIAO YL, LI ZS, HAN SY. Screening of endogenous signal peptides from *Pichia pastoris* for promoting lipase secretion by *Rhizopus oryzae*[J]. Modern Food Science and Technology, 2019, 35(10): 155-163 (in Chinese).
- [32] 田小梅, 任建洪, 房聪. 不同碳源下毕赤酵母 GS115 蛋白组学分析[J]. 中国生物工程杂志, 2012, 32(1): 21-29.
- TIAN XM, REN JH, FANG C. Proteomic analysis of *Pichia pastoris* GS115 cultured in different carbon source mediums[J]. China Biotechnology, 2012, 32(1): 21-29 (in Chinese).
- [33] GUO H, SUN JH, LI X, XIONG Y, WANG HT, SHU H, ZHU RM, LIU Q, HUANG YM, MADLEY R, WANG YL, CUI JQ, ARVAN P, LIU M. Positive charge in the n-region of the signal peptide contributes to efficient post-translational translocation of small secretory preproteins[J]. Journal of Biological Chemistry, 2018, 293(6): 1899-1907.

- [34] OWJI H, NEZAFAT N, NEGAHDARIPOUR M, HAJIEBRAHIMI A, GHASEMI Y. A comprehensive review of signal peptides: structure, roles, and applications[J]. *European Journal of Cell Biology*, 2018, 97(6): 422-441.
- [35] NILSSON I, WHITLEY P, von HEIJNE G. The COOH-terminal ends of internal signal and signal-anchor sequences are positioned differently in the ER translocase[J]. *The Journal of Cell Biology*, 1994, 126(5): 1127-1132.
- [36] XUE SL, LIU XF, PAN YY, XIAO CF, FENG YZ, ZHENG L, ZHAO MM, HUANG MT. Comprehensive analysis of signal peptides in *Saccharomyces cerevisiae* reveals features for efficient secretion[J]. *Advanced Science* (Weinheim, Baden-Wurttemberg, Germany), 2023, 10(2): e2203433.
- [37] TING YT, HARRIS PWR, BATOT G, BRIMBLE MA, BAKER EN, YOUNG PG. Peptide binding to a bacterial signal peptidase visualized by peptide tethering and carrier-driven crystallization[J]. *IUCrJ*, 2016, 3(Pt 1): 10-19.
- [38] TIAN J, ZHANG YH, LIU B, ZUO DY, JIANG T, GUO J, ZHANG W, WU NF, FAN YL. Presep: predicting the propensity of a protein being secreted into the supernatant when expressed in *Pichia pastoris*[J]. *PLoS One*, 2013, 8(11): e79749.
- [39] WANG P, HUANG L, JIANG H, TIAN J, CHU XY, WU NF. Improving the secretion of a methyl parathion hydrolase in *Pichia pastoris* by modifying its N-terminal sequence[J]. *PLoS One*, 2014, 9(5): e96974.

(本文责编 郝丽芳)