

DOI: 10.1360/yc-007-0250

粟酒裂殖酵母全基因组中含信号肽蛋白质的研究

刘玉岭, 柳云帆, 谢建平

西南大学生命科学学院现代生物医药研究所, 重庆 400715

摘要: 对粟酒裂殖酵母全基因组 3 条染色体上的 4,997 个蛋白序列进行了全局性的分析, 利用 signalP3.0 软件分析这些蛋白的 N-末端信号肽序列, 预测有 N-末端分泌信号肽序列的蛋白 196 个; 利用 TMpred 软件分析跨膜结构, 预测跨膜蛋白 117 个; 使用 PrositeScan 程序分析膜脂蛋白的脂结合位点, 预测有膜脂结合蛋白 13 个, 进而预测分泌性蛋白序列 66 个。使用 Target P 分析 66 个分泌蛋白的蛋白序列, 研究这些蛋白在细胞中的定位。这些分泌蛋白的功能涉及粟酒裂殖酵母的营养、生殖、细胞间以及细胞与环境间的交流等许多方面, 对细胞的生存和繁殖有重要意义, 在系统生物学的研究中有重要参考价值。粟酒裂殖酵母分泌组的研究也将为粟酒裂殖酵母作为药物筛选模型以及开发为外源蛋白表达的宿主提供基础。

关键词: 粟酒裂殖酵母; 信号肽; 膜脂蛋白; 膜整合蛋白; 分泌蛋白

Characterization of *Schizosaccharomyces pombe* secreted proteins

LIU Yu-Ling, LIU Yun-Fan, XIE Jian-Ping

Institute of Modern Biopharmaceuticals, School of Life Sciences, Southwest University, Chongqing 400715, China

Abstract: Secreted proteins play a vital physiological role. *Schizosaccharomyces pombe* is an important model organism for cell cycle study and a potential useful drug screen model. Secreted proteins also initiate the mating. However, few global studies concerning the secreted proteins of *S. pombe* was reported. To address this issue, bioinformatics were used to reveal the global secreted proteins of *S. pombe*. The 4,997 proteins deduced from the *S. pombe* genome were analyzed by combined several programs. One hundred and sixty proteins were identified carrying an NH₂-terminal secretory signal peptide by signalP3.0. Among them, 117 proteins are integral membrane proteins(TMpred), 13 proteins are lipoproteins (PrositeScan), and 66 proteins are secreted proteins. The location of the secreted proteins was also predicted by Target P. Some of the secreted proteins are involved in the nutrition, reproduction, as well as the communication between cells and environment. The global information of the secreted proteins of *S. pombe* will benefit further studies in drug screening model and host searching for heterologous gene expression.

Keywords: *Schizosaccharomyces pombe*; signal peptides; lipoprotein; integral membrane protein; secreted protein

粟酒裂殖酵母(*Schizosaccharomyces pombe*)是研究细胞周期调控、DNA修复与重组和基因组稳定性的关卡控制的模式生物, 也是针对细胞周期紊乱

的肿瘤药物的理想细胞初筛模型, 其遗传背景清晰, 遗传操作方便。*S. pombe*的全基因组序列测定于 2002 年完成^[1], 基因组大小为 13.8 Mb, 分布在染色

收稿日期: 2006-06-14; 修回日期: 2006-09-06

基金项目: 教育部科学技术研究重点项目(编号: 105146)和西南大学 211 建设项目(微生物学)[Supported by Key Project of MOE S&T(No. 105146) and 211 Key Microbiology Project of Southwest University]

作者简介: 刘玉岭(1981—), 女, 山东泰安人, 硕士研究生, 研究方向: 粟酒裂殖酵母功能基因组学与药物筛选模型。Tel: 023-68253392; E-mail: lyling@swu.edu.cn

通讯作者: 谢建平(1971—), 男, 重庆人, 博士, 研究员, 研究方向: 工业(药用)微生物功能基因组学与新药筛选。Tel: 023-68254062; E-mail: georgex@swu.edu.cn

体 I (5.7 Mb), II (4.6 Mb), III (3.5 Mb) 和 20 kb 的线粒体基因组上。预测含有 4,997 个开放阅读框(open reading frame, ORF), 其中约有 1,200 个 ORFs 功能已知, 3,700 多个 ORFs 功能未知。

蛋白质分泌对于真核生物的活动非常重要。真核细胞的大多数蛋白质都是由细胞核内的基因组编码, 在细胞质内位于内质网上的核糖体中合成, 经过高尔基体翻译后修饰的根据其目的部位分拣(sorting)运输。依赖存在于 N-末端的靶序列, 许多蛋白质分选到达亚细胞组分^[2,3]。一般来说, 蛋白利用 N 末端序列穿过细胞质膜, 如果没有附加的跨膜片段和膜锚定序列, 蛋白质穿过质膜后, 信号肽酶切割去掉信号肽释放成熟蛋白。通过序列分析可以识别信号肽的保守特征(氨基酸的长度、组成和信号肽的切割位点), 进而预测这些蛋白质的亚细胞定位^[4-6]。

研究粟酒裂殖酵母的分泌蛋白和信号肽可能有助于更全面认识其生物学性状。细胞也可以通过这些蛋白的分泌, 实现与环境及其他细胞间的交流, 如营养的利用、细胞的接合生殖等。这不仅对其生长和繁殖有重要意义, 在系统生物学的研究中也有重要价值。

我们根据粟酒裂殖酵母的全基因组序列运用生物信息学的方法进行了全局性的分析, 寻找粟酒裂殖酵母中有信号肽的蛋白序列, 然后排除膜整合蛋白和膜脂蛋白, 得到分泌蛋白。这些分析为研究粟酒裂殖酵母分泌性蛋白的功能提供了较全面的资料。

1 材料和方法

1.1 粟酒裂殖酵母基因组序列及推测蛋白的氨基酸组成序列

从 Sanger 中心数据库(http://www.sanger.ac.uk/Projects/S_pombe/)中得到粟酒裂殖酵母的全基因组序列。

1.2 蛋白序列的生物信息学分析

使用 signal P3.0 软件(<http://www.cbs.dlu.dk/services/signalP/>)分析粟酒裂殖酵母基因组中所有蛋白的 ORFs 的 N 端氨基酸序列, 确定是否存在信号肽, 选项均按照默认值操作。

利用 TMpred (http://www.ch.embnet.org/software/TMPRED_form.html)分析跨膜结构。

使用 PrositeScan 程序(<http://myhits.isb-sib.ch/>

[cgi-bin/motif_scan](#))分析膜脂蛋白(membrane lipoprotein, LPP)的脂结合位点。

使用 targetP Server v1.01(<http://www.cbs.dlu.dk/services/targetP/>)分析有信号肽的 ORFs, 确定分泌蛋白的分泌途径和亚细胞位点。

1.3 统计分析

根据从 SignalP 中分析获取的 C_{max} 值(the cleavage site score)、 S_{mean} (the signal peptide score)和 HMM 值(hidden Markov model)判别信号肽序列。使用 TM_{pred} 值(Prediction of Transmembrane Regions and Orientation)分析跨膜序列, 根据 PrositeScan 程序分析膜脂蛋白的脂结合位点, 区分分泌蛋白, 整合膜蛋白和脂蛋白。利用 targetP Server v1.01 分析分泌蛋白的分泌途径和所到达的亚细胞位点。对膜脂蛋白以及分泌蛋白的信号肽序列进行统计分析。

2 结果

使用生物信息学的方法鉴别有分泌型信号肽且无附加的膜结合域的蛋白序列, 所使用的策略如图 1 所示。

2.1 粟酒裂殖酵母的蛋白序列

从 Sanger 中心数据库中得到粟酒裂殖酵母的全基因组序列, 预测蛋白的氨基酸序列为 4,997 个。

2.2 Signal P3.0 软件分析 N-末端信号肽

将粟酒裂殖酵母预测的 4,997 个蛋白序列使用 signal P3.0 软件分析, 获得每个 ORF 的 C_{max} 值(the cleavage site score)、 Y_{max} 值(the combination score)、 S_{max} 值(the signal peptide score)、 S_{mean} 值(the mean values of the signal peptide score)、D 值(the classification task)及其位点。根据软件的默认选择, 将 C_{max} 值 > 0.49 和 S_{mean} 值 > 0.5 的 ORFs 确定为具有 N-末端信号肽的 ORFs, 结果显示有 219 个 ORFs 有 N-末端信号肽序列。其中排除 HMM 值中的 Signal peptide probability < 0.5, Signal anchor probability > 0.5 的 ORFs 23 个, 得到 195 个有 N-末端信号肽的 ORFs。这些序列在 3 条染色体上的分布如表 1 所示。

2.3 TMpred 软件分析跨膜结构

只有缺少膜锚定序列的蛋白质穿过膜以后切割信号肽, 才能释放到外界环境。为了排除含有跨膜片断的蛋白质序列, 使用 TMpred (

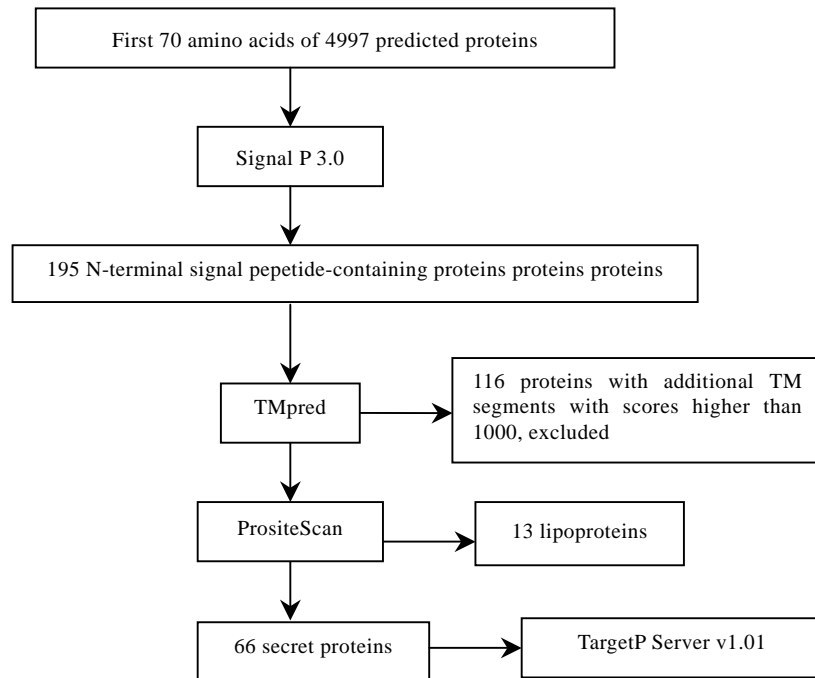


图 1 预测粟酒裂殖酵母分泌蛋白的策略

Fig. 1 Schematic representation of the strategy utilized for predicting *Schizosaccharomyces pombe* secreted proteins

embnet.org/software/TMPRED form.html)分析 195 个蛋白序列的跨膜结构,发现 116 个蛋白序列可能是整合膜蛋白,即成熟蛋白质的序列中含有得分高于 1,000 的跨膜片段(TMpred 排除跨膜片段的选择分数是 500)。

2.4 PrositeScan 程序分析脂结合蛋白

有 N-末端信号肽而无跨膜结构的 79 个蛋白质序列使用 PrositeScan 程序(<http://myhits.isb-sib>。

ch/cgi-bin/motif_scan)分析膜脂蛋白(membrane lipoprotein, LPP)的脂结合位点。三条染色体上共有 13 个蛋白序列有脂结合位点,推测这些序列可能为膜脂蛋白(表 2)。

2.5 TargetP 分析分泌蛋白的分泌途径和亚细胞定位

使用 targetP Server v1.01(<http://www.cbs.dtu.dk/services/targetP/>)软件分析预测的 66 个分泌蛋白序列,确定分泌蛋白的分泌途径和所达到的亚细胞位点。

表 1 粟酒裂殖酵母中不同类型的蛋白在三条染色体上的分布预测

Table 1 The distribution of the proteins of different groups in the three chromosomes of *Schizosaccharomyces pombe*

类别 Classification	染色体 I Chromosome I	染色体 II Chromosome II	染色体 III Chromosome III	粟酒裂殖酵母基因组 The genome of <i>S. pombe</i>
核酸序列大小(Mbp) Size(Mbp)	5.7	4.6	3.5	13.8
全基因组中的蛋白序列 Number of proteins	2255	1790	884	4929
SignalP 分析预测的有 N-末端信号肽的序列 Signal peptide sequences by SignalP 3.0	92	59	44	195
TMpred 分析预测的膜整合蛋白序列 Integral membrane protein sequences predicted by TMpred	50	32	34	116
PrositeScan 分析得到的膜脂蛋白序列 Lipoproteins predicted by PrositeScan	10	1	2	13
可能的分泌蛋白序列 Putative secreted proteins	32	26	8	66

表 2 预测的膜脂蛋白信号肽

Table 2 Predicted lipoprotein signal peptides

序号 No.	蛋白名称 Name	所在的 染色体 Chro- mosome	氨基酸 长度 Length of AA	Signal P 3.0 预测的信号肽序列 Signal P 3.0 predicted signal peptide sequence	信号肽 长度 Length of SP	signal P 预测 的信号肽 分值 Signal P score	蛋白相关的功能 或者注释 Relevant notes regarding function or identity
1	SPAC1F5.06		848	MKRSVLTIILFFSCQFWHAFA-SS	21	0.975	HSP 70 family protein
2	SPAC821.09		1016	MSSYLRSFIFGLLTISLAQC-SP	20	0.806	Endo-1,3-beta-glucanase Eng1
3	SPAC13F5.05		363	MLFRIPTLFTLFLACFSLVSGVFG-YS	24	0.964	Thioredoxin family
4*	SPAC19A8.14		205	MKVVPFVNFMISSFPAAVLVGA V VG-FM	24	0.853	Aminoacyl-tRNA hydrolase (predicted)
5*	SPAC4A8.04		467	MRIPYSNLFSAAAGLALFASTACA-AP	24	0.904	Serine protease
6*	SPACUNK4.16c		944	MGRILIAHLFLPSSV G -FS	16	0.731	Alpha, alpha-trehalose- phosphate synthase; glycosyl transferase family 20
7	SPAC2E1P3.05c		197	MLTQSLFTLTLALSLVSKTSA -SQ	23	0.949	Conserved fungal protein
8	SPAC27D7.09c		383	MKLTSIPIASTLLSLLAASGTLA -SP	23	0.916	Predicted N-terminal signal sequence
9	SPAC27D7.10c		383	MKLTSIPIASTLLSLLAASGTLA -SP	23	0.916	Possibly <i>S. pombe</i> specific
10	SPAC1B3.10c		680	MGVSVLTFHVSLFLKRLISIAFFLLSLTLRIVNA -QQ	36	0.948	SEL1 repeat protein
11	SPBPB2B2.06c		601	MKTASIHFWSTLVLLFSCIGSVIA -YS	24	0.956	Calcineurin-like phosphoesterase UDP-glucose by similarity
12	SPCC188.09c		609	MLFLRFFIFTFFTSIFTV V VS -EP	21	0.838	Glycoprotein (predicted)
13	SPCC1840.07c		332	MSSLGKLLKLTLLGILLSFSCKFVFG -YF	26	0.943	Calcineurin-like phosphoesterase (predicted)

: 实验验证(或者已发表的)序列; : 同源比对预测的蛋白质; : *S. pombe* 特有的家族; : 孤儿序列; * GPI 锚定蛋白。这些符号在本文所有图表中通用。

: Experimentally characterised (or published); : Protein inferred from homology alignment; : *Schizosaccharomyces pombe* specific families; orphan sequence; *: GPI Anchor. These symbols are used throughout the text.

根据 Target P 的结果分析, 分泌蛋白的分泌途径均为 S 型(即分泌到细胞质空间)。

2.6 对 3 条染色体上的分泌蛋白的信号肽序列分析

对 3 条染色体上的分泌蛋白的信号肽序列进行分析(表 3 ~ 表 5), 这些分泌蛋白的信号肽长度在 16 ~ 39 之间, H-domain 平均由 33 个氨基酸构成。C-domain(切割位点前 3 位的氨基酸)为 A-X-A(A 为丙氨酸, X 指任意一种氨基酸)的序列有 22 个, 占分泌蛋白序列总数的三分之一, 具有非常典型的分泌类信号肽的结构。

3 讨论

我们从基因组数据库中得到预测蛋白序列 4,997 个, 使用生物信息学方法分析这些蛋白序列, 推测粟酒裂殖酵母中有 66 个蛋白属于分泌蛋白, 这些蛋白组分涉及细胞与环境进行物质和信息交换, 使我们对可能发生在细胞膜外空间的反应和途径有

了全面的了解。虽然生物信息学得到的结果还需要实验验证, 但是进一步研究分泌蛋白对细胞的生存、生长和繁殖的影响, 将为粟酒裂殖酵母的研究和应用提供理论基础。

我们预测的 13 个膜脂蛋白中, 有 2 个是有实验证明的蛋白(SPAC821.09, SPAC4A8.04), 8 个是经序列同源性比对预测的蛋白, 另外 3 个是粟酒裂殖酵母特有的蛋白。SPAC4A8.04 是一个丝氨酸蛋白酶, 在氮饥饿时表达^[7], 它与粟酒裂殖酵母接合生殖中细胞融合过程有关, 该基因缺失将阻断粟酒裂殖酵母的接合和孢子的形成^[8]。SPAC821.09 是一个葡聚糖酶, 与粟酒裂殖酵母裂殖时膜分离有关^[9]。

13 个膜脂蛋白中有 3 个蛋白(SPAC4A8.04, SPAC19A8.14, SPACUNK4.16c)具有 GPI (glycosylphosphatidylinositol)锚定蛋白的结构。GPI 是哺乳动物、真菌、寄生虫细胞中非常普遍的表面蛋白锚定方式, 通过其羧基末端的糖基化磷脂酰肌醇结构锚定于真核细胞膜表面。细胞中存在很多 GPI 蛋白,

表 3 粟酒裂殖酵母染色体 I 中预测的分泌蛋白的信号肽分析

Table 3 The signal peptide of the predicted secrete proteins in chromosome of *Schizosaccharomyces pombe*

序号 No.	蛋白名称 Name	氨基酸 长度 Length of AA	Signal P 3.0 预测的信号肽序列 Signal P 3.0 predicted signal peptide sequence	信号肽的 长度 Length of SP	signal P 预测 的信号肽分 值 Signal P score	蛋白相关的功能或者注释 Relevant notes regarding function or identity	蛋白的定位 预测 Target
1	SPBC1348.06c	203	MFKNLIFLFFIGL <u>ATA</u> -IR	16	0.892	Hypothetical protein	S
2	SPAC977.05c	204	MIFKNLISLFFIGL <u>ATA</u> -IR	17	0.871	Hypothetical protein	S
3	NP_592789.1	117	MLLLLYICCLFLK <u>FILA</u> -NV	17	0.973	Hypothetical protein	S
4	SPAC1F8.06	385	MNSYISLIFTLLFFTS <u>AARS</u> -SS	20	0.819	Sim4 and Mal2 associated protein 5	S
5	SPAC11D3.09	394	MALQSLFLILLAGAAQL <u>AQA</u> -HP	20	0.918	Agmatinase (predicted)	S
6	SPAC1F5.02	492	MKISNLLAAFLAFSGGFCA <u>SA</u> -EV	22	0.858	Thioredoxin family	S
7	SPAC1751.02c	108	MQIFAQLTSLFTFHLMFWTSV <u>ARA</u> -RS	24	0.891	Mitochondrial ribosomal protein subunit S19	S
8	SPAC13C5.06c	177	MKGFVVISRFILTLFILITP <u>GLA</u> -GV	23	0.973	Sequence orphan	S
9	SPAC227.11c	322	MFPHLILPAIGSSKVRTMVLPPFAFVGGFFIP <u>ICLA</u> -SL	35	0.852	Sensor for misfolded ER glycoproteins Yos9 (predicted)	S
10	NP_593174.1	244	SKMDGIFTVSLFSTRLLSFYLFIEV <u>LLT</u> -KI	29	0.972	Probable histone acetyltransferase type b catalytic subunit	S
11	SPAC23C4.05c	431	MQPRFLLHGALLALGIQL <u>CLS</u> -IG	21	0.931	Hypothetical protein	S
12	NP_593216.1	385	MLLFKFNFTTAFILAF <u>AQA</u> -RS	22	0.702	Putative family 31 glucosidase c30d11.01c precursor	S
13	SPAC22A12.1 5c	663	MKKFQLFSILSYFVALFLLPMA <u>F</u> A-SG	24	0.958	BiP, heat shock protein 70 family involved in protein folding of core-glycosylated trimmed ligands, associated with newly synthesized glycoproteins	S
14	SPAC6F12.07	152	MRRSVIIGSLLATAAV <u>G</u> YA-IY	19	0.682	Mitochondrial TOM complex subunit Tom20	S
15	SPAC3C7.11c	560	MKYGKVSFLALLCSLYVRG <u>SLA</u> -DP	22	0.874	Calnexin	S
16	SPAC6C3.03c	100	MDVFFLGLLVIAIFT <u>N</u> A-KR	18	0.918	Sequence orphan	S
17	SPAC513.06c	368	MTSMMSGASPVIHWGFLGAGSIAA <u>V</u> FA-KD	26	0.500	Dihydrodiol dehydrogenase (predicted)	S
18	SPAPB2C8.01	1220	MAVSRLLILICLYSFVTF <u>A</u> YP-KV	21	0.938	Glycoprotein (predicted)	S
19	SPAPB24D3.0 7c	300	MKSFWVWTLGALSGLSLTT <u>A</u> YG-AN	22	0.921	Sequence orphan	S
20	SPAC652.01	108	MLCLRYFIPMLLV <u>A</u> NA-AP	16	0.968	Conserved eukaryotic protein	S
21	SPAC26A3.14 c	73	MPFLSRLFHYGVDLALVSTC <u>V</u> A-GI	22	0.610	Conserved fungal protein	S
22	SPAC16E8.05c	209	MLHATQLCYLLLFCLPIS <u>S</u> SA-VL	23	0.960	Sequence orphan	S
23	SPAC17G6.03	635	MKPAHTFSLIFSLFVINP <u>C</u> YS-AT	22	0.935	Calcineurin-like phosphoesterase (predicted)	S
24	SPAC19G12.1 0c	1002	MLMKQTFLYFLT <u>C</u> V <u>S</u> A-QF	18	0.946	Serine carboxypeptidase	S
25	SPAC26H5.08 c	321	MQFLSSFVFAALALLPLS <u>A</u> MA-VD	21	0.953	Glucan 1,3-beta-glucosidase	S
26	SPAC12B10.1 6c	509	MKYWQAILFFLFGI <u>A</u> FA-NN	17	0.957	Conserved protein	S
27	SPAPB8E5.04c	188	MRLTTFIYAITCLPIF <u>S</u> A-SS	19	0.916	Phosphatidylglycerol/phosphatidylinositol transfer protein (predicted)	S
28	SPAPJ760.03c	166	MFLRSIFQTLCAVSFLAGS <u>V</u> FA-DS	22	0.907	Glycoprotein (predicted)	S
29	SPAC1039.02	601	MKLSLPSGLGLASLLGLISS <u>A</u> TA-YS	24	0.885	Calcineurin-like phosphoesterase (predicted)	S
30	SPAC922.04	117	MKFFVWSLLGLLGL <u>S</u> TA-IP	18	0.972	Sequence orphan	S
31	SPAC869.07c	436	MISISFLNCFVLVFLFFSDV <u>H</u> G-SY	24	0.960	Alpha-galactosidase	S
32	SPAC212.08c	278	MSPLIVGTLHILLSSGL <u>A</u> TA-FY	20	0.970	<i>S.pombe</i> specific families	S

表 4 粟酒裂殖酵母染色体 II 中预测的分泌蛋白的信号肽分析

Table 4 The signal peptide of the predicted secrete proteins in chromosome II

序号 No.	蛋白名称 Name	氨基酸 长度 Length of AA	Signal P 3.0 预测的信号肽序列 Signal P 3.0 predicted signal peptide sequence	信号肽 长度 Length of SP	signal P 预 测的信号 肽分值 Signal P score	蛋白相关的功能或者注释 Relevant notes regarding function or identity	蛋白的定 位预测 Target
1●	SPBC359.04c	358	MNSLKSCLKCIVTLCLLVNAFA-FD	23	0.950	Glycoprotein (predicted)	S
2	NP_595181.1	463	MKLSGISLWLLAASIVHA-GK	18	0.923	Thiamine-repressible acid phosphatase	S
3	SPBC1685.12c	117	MVSNHVEWYLLLSIVSTA VES -LL	21	0.849	Dubious	S
4	NP_595239.1	412	MVKNTSVIIVGAGVFGLSAA-LE	20	0.740	Fructosyl amino acid oxidase	S
5●	SPBC947.04	973	MSLFPQILLRLLFLAFTLKST SNA -EK	24	0.831	Glycoprotein (predicted)	S
6	NP_595281.1	1448	MRWGFWFAIATLITIC CYA -AK	18	0.959	UDP-glucose-glycoprotein glucosyltransferase	S
7	NP_595346.1	577	MRPSVITVAVLVFQSTWA-SF	18	0.880	Vacuolar polyphosphatase	S
8◇	SPBC713.11c	50	MTQVIFAIILPPLGVFLERGCADVIINILLCCLGYVPG-II	39	0.974	UPF0057 family protein; possible stress response protein	S
9	NP_595512.1	66	SFPIHYHFSFLSPLLFLPHISFA-LP	23	0.849	Very hypothetical protein	S
10◆	SPBC3D6.02	390	MQLLSNSFLGFAASIAVLASSADA-AP	23	0.806	Neddlylation pathway protein But2	S
11◆	SPBC14C8.05c	450	MRTYWLFLLLGGV VSA -ES	16	0.903	Glucan-alpha-1,4-glucosidase	S
12◇	SPBC21H7.03c	463	MQLCISLWFLAAFI VNA -DN	18	0.955	Acid phosphatase (predicted)	S
13●	SPBC21D10.06c	948	MNSYAILLSFFSFERLTL ANA -NS	23	0.849	Glycoprotein (predicted)	S
14◇	SPBC2G5.01	374	MINKKLLFLVFALAKGV LA -DE	19	0.960	Conserved eukaryotic protein	S
15☆	SPBC685.03	452	MQFFGSLFVSLGAA GLANA -LP	20	0.924	Serine-rich protein	S
16◇	SPBC2A9.06c	158	MYDDIFFYALWVIQSV YG -AW	19	0.984	Di-trans,poly-cis-decaprenylcistrans-ferase(predicted)	S
17◇	SPBC30D10.09c	217	MQLLMLEQISLASSVVATTVLVAPV LS -TI	27	0.969	TB2/DP1 domain	S
18◇	SPBC4C3.09	376	MNFFKRLRLHTRLLRSKFVLISLILLNLGLLLG-IQ	35	0.826	Acetylglucosaminyltransferase	S
19◇	SPBC2G2.17c	319	MLFNNFLCAVSAIPL VSA -MP	19	0.808	Beta-glucosidase (predicted)	S
20◇	SPBC1105.05	407	MLSFTSVF SFFLHALLLKTAFS -YV	22	0.926	Beta-glucosidase (predicted)	S
21◇	SPBP8B7.25	201	MKLFYFSLFTLFFGL ISA -NR	19	0.960	Cyclophilin	S
22☆	SPBC21C3.14c	841	MKIERYFKAIARAFIITFLSLIQDNG VLA -RK	31	0.854	Sequence orphan	S
23●	SPBC1289.15	1283	MLVFTDLIFLFAIHGK ATA -KV	21	0.880	Glycoprotein (predicted)	S
24◇	SPBC8E4.03	413	MFTYKIIQLALLLSGVCG ALA -SI	22	0.949	Arginase family (predicted)	S
25*◆	SPBP4G3.02	453	MFLQNLFLGFLAVVC ANA -QF	18	0.950	Acid phosphatase	S
26◇	SPBPB2B2.15	203	MFKNLIFLFFIGL ATA -IR	16	0.892	Hypothetical protein;	S

表 5 粟酒裂殖酵母染色体 III 中预测的分泌蛋白的信号肽分析

Table 5 The analysis of the signal peptide of the predicted secrete proteins in chromosome III

蛋白名称 Name	氨基酸长度 Length of AA	Signal P 3.0 预测的信号肽序列 Signal P 3.0 predicted signal peptide sequence	信号肽的 长度 Length of SP	signal P 预 测的信号 肽分值 Signal P score	蛋白相关的功能或者注释 Relevant notes regarding function or identity	蛋白的定 位预测 Target	
1●	SPCP20C8.02c	111	MLLLFICCCVFIKL VLA -EV	17	0.975	<i>Schizosaccharomyces pombe</i> specific families	S
2◇	SPCC757.05c	400	MTMKISVSWLLIVIGYHLWMSPV VLA -GP	25	0.967	Peptidase family M20	S
3◇	SPCC613.03	189	MKFSTVGF FLFSTILFKSAFA -GW	20	0.842	EF hand	S
4☆	SPCC306.11	283	MFAFASFAISAIFFLCSFSY VSS -IK	23	0.915	Sequence orphan	S
5◆	SPCC1795.06	201	MKITAVIALLFSL AAA -SP	16	0.951	p-factor pheromone	S
6◇	SPCC825.02	506	MKFSQWYTLTAPLLISSLYT VNA -AN	23	0.916	Glucosidase II Gtb1 (predicted)	S
7◆	SPCC191.11	581	MFLKYILASGICLVSL SSTNA -AP	22	0.770	Beta-fructofuranosidase	S
8●	SPCC569.02c	113	MLLLFICCAFIKL VLA -EV	17	0.974	Hypothetical protein	S

包括细胞表面的酶、受体、粘附分子、特异性抗原等,功能涉及免疫识别,补体调节及跨膜信号转导等^[10]。我们得到的分泌蛋白中有一个蛋白具有GPI的

结构,这个蛋白位于染色体 ,是一个糖蛋白(SPBP4G3.02),具有酸性磷酸酶的功能,定位于细胞表面^[11,12]。

在预测的 66 个分泌蛋白中, 10 个蛋白有文献证明。其中染色体 中有 5 个蛋白(SPAC1F8.06, SPAC22A12.15c, SPAC3C7.11c, SPAC19G12.10c, SPAC869.07c), 染色体 中有 3 个蛋白(SPBC3D6.02, SPBC14C8.05c, SPBP4G3.02), 染色体 中有 2 个蛋白(SPCC1795.06, SPCC191.11)。这些蛋白在表 3, 4, 5 中使用 ‘ ’ 标记。这些蛋白的功能涉及动粒的结构, 糖蛋白的合成与折叠, 葡萄糖、半乳糖、氨基酸的分解, 以及磷酸酶和信息素等类别^[12-16]。

其他 56 个蛋白质是根据序列同源性比对预测得到, 它们的功能主要涉及细胞的营养和代谢, 以及细胞与细胞间的通讯和应答(表 2 ~ 表 4)。这些蛋白在细胞间以及细胞与环境间的交流中有重要作用。

利用软件分析为分泌蛋白的酸性磷酸酶, 经过我们采用 SDS-PAGE 分离、肽指纹分析确证分泌到胞外。虽然经此确证的分子不多, 但也说明了软件分析的价值。

粟酒裂殖酵母是研究细胞周期的模式生物, 也是初筛针对细胞周期紊乱的肿瘤药物的理想模型。研究其分泌蛋白和信号肽有助于更全面认识其生物学性状, 并为构建新的药物筛选模型提供基础。同时, *S.pombe* 作为表达系统也值得关注, 因为该菌的许多特性与高等真核细胞相似, 表达的外源基因产物具有较理想的天然构象和活性。分析粟酒裂殖酵母的分泌组将为进一步优化粟酒裂殖酵母表达系统提供参考信息。

参考文献(References):

- [1] Wood V, Gwilliam R. The genome sequence of *Schizosaccharomyces pombe*. *Nature*, 2002, 415(21): 871—880.[\[DOI\]](#)
- [2] Murphy CK, Beckwith Murphy J. Export of proteins to the cell envelope in *Escherichia coli* and *Salmonella typhimurium*: *Cellular and Molecular Biology*, 1987, 1: 967—978.
- [3] Gomez M, Johnson S, Gennaro LM. Identification of secreted proteins of *Mycobacterium tuberculosis* by a bioinformatics approach. *Infection and Immunity*, 2000, 68(4): 2323—2327.[\[DOI\]](#)
- [4] Watson MEE. Compilation of published signal sequences. *Nucleic Acids Res*, 1984, 12: 5145—5164.[\[DOI\]](#)
- [5] Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol*, 2000, 300: 1005—1016.[\[DOI\]](#)
- [6] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides-SignalP 3.0. *J Mol Biol*, 2004, 340: 783—795.[\[DOI\]](#)
- [7] Nakashima A, Yoshida M, Nakayama K, Kato-Furuno A, Ueno M, Ushimaru T, Uritani M. Genes for a nuclease and a protease are involved in the drastic decrease in cellular RNA amount in fission yeast cells during nitrogen starvation. *J Biochem (Tokyo)*, 2002, 131(3): 391—398.
- [8] Sato S, Suzuki H, Widyastuti U, Hotta Y, Tabata S. Identification and characterization of genes induced during sexual differentiation in *Schizosaccharomyces pombe*. *Curr Genet*, 1994, 26(1): 31—37.[\[DOI\]](#)
- [9] Martín-Cuadrado AB, Dueñas E, Sipiczki M, Carlos R, de Aldana V, del Rey F. The endo-b-1,3-glucanase *englp* is required for dissolution of the primary septum during cell separation in *Schizosaccharomyces pombe*. *Journal of Cell Science*, 2003, 116, 1689—1698.[\[DOI\]](#)
- [10] XIONG Mao-Lin. The structure, function and application of GPI anchor. *Foreign Medical Sciences*, 2003, 26(5): 230—233.
熊茂林. GPI 锚的结构、功能及 GPI 锚定 B7 分子在肿瘤免疫治疗中的应用. 国外医学免疫学分册, 2003, 26(5): 230—233.
- [11] Schwaninger R, Dumermuth E, Schweingruber ME. Effects of seven different mutations in the *pho1* gene on enzymatic activity, glycosylation and secretion of acid phosphatase in *Schizosaccharomyces pombe*. *Mol Gen Genet*, 1990, 221(3): 403—410.[\[DOI\]](#)
- [12] Maundrell K, Nurse P, Schonholzer F, Schweingruber ME. Cloning and characterization of two genes restoring acid phosphatase activity in *pho1*-mutants of *Schizosaccharomyces pombe*. *Gene*, 1985, 39(2-3): 223—230.[\[DOI\]](#)
- [13] Liu X, McLeod I, Anderson S, Yates JR, He X. Molecular analysis of kinetochore architecture in fission yeast. *The EMBO Journal*, 2005, 24: 2919—2930.[\[DOI\]](#)
- [14] Jannatipour M, Callejo M, Parodi AJ, Armstrong J, Rokeach LA. Calnexin and BiP interact with acid phosphatase independently of glucose trimming and reglucosylation in *Schizosaccharomyces pombe*. *Biochemistry*, 1998, 37(49): 17253—17256.[\[DOI\]](#)
- [15] Goddard A, Ladds G, Davey J. Development of a semi-quantitative plate-based alpha-galactosidase gene reporter for *Schizosaccharomyces pombe* and its use to isolate a constitutively active Mam2. *Yeast*, 2005, 22(1): 31—41.[\[DOI\]](#)
- [16] Tanaka N, Ohuchi N, Mukai Y, Osaka Y, Ohtani Y, Tabuchi M, Bhuiyan MS, Fukui H, Harashima S, Takegawa K. Isolation and characterization of an invertase and its repressor genes from *Schizosaccharomyces pombe*. *Biochem Biophys Res Commun*, 1998, 245(1): 246—253.[\[DOI\]](#)