

罗非鱼源无乳链球菌 *cpsE* 基因的克隆和分子特性分析

汪开毓^{1,2*}, 付希^{1,2}, 肖丹³, 黄锦炉^{1,2}, 王均^{1,2}, 王浩丞^{1,2}

(1. 四川农业大学鱼病研究中心, 四川 雅安 625014;

2. 四川农业大学动物疫病与人类健康四川省重点实验室, 四川 雅安 625014;

3. 通威股份有限公司, 四川 成都 610041)

摘要: 利用特异性引物, 采用 PCR 扩增出分离自患病罗非鱼无乳链球菌强毒株的 *cpsE* 基因, 将其克隆到 pMD19-T 载体上, 通过菌落 PCR 鉴定和利用限制性内切酶 *Bam*H I 和 *Hind* III 对重组质粒进行双酶切鉴定之后送测序公司测序, 并利用生物信息学软件 Clustal X 2.0、MEGA 4.1、Bioedit 7.0、TMHMM、NetPhos 2.0、NetNGlyc 1.0、SignalP 3.0 Server、PSIpred、SAM_T08 以及 CUSP 等分析 *cpsE* 基因的分子特性。结果显示, 罗非鱼源无乳链球菌 *cpsE* 编码氨基酸序列具有高度保守性, 与人源、动物源无乳链球菌亲缘性达 100%, 具有 1 个参与催化糖基元转运的 Glycosyltransferases 超级家族结构域, 具有与蛋白翻译后修饰功能相关的磷酸化位点 3 个, 编码多肽链中亲水区大于疏水区, 且存在跨膜区。密码子偏爱性分析表明, 罗非鱼源无乳链球菌 *cpsE* 基因密码子使用频率差异较大, 密码子偏爱性更接近真核生物。

关键词: 无乳链球菌; *cpsE* 基因; 克隆; 分子特性

中图分类号: Q 786; S 917

文献标识码: A

无乳链球菌 (*Streptococcus agalactiae*), 也称 B 族链球菌 (group B streptococci, GBS), 是兼性的革兰氏阳性球菌^[1]。根据荚膜多糖 (capsular polysaccharide, CP_s) 抗原不同将无乳链球菌分为 I a、I b、II - VIII 和 IX 10 种血清型^[2-4]。目前, 无乳链球菌仍然是引发新生儿败血症和脑膜炎的常见病原^[5]; 对未孕成年人主要引起尿脓毒症、肺炎、软组织及皮肤感染^[6]; 该菌也是牛乳腺炎的常见病原^[7], 同时也可感染多种水产动物, 引起罗非鱼、虹鳟、海鲤和鲢等鱼类败血症和脑膜炎^[8-9], 具有较高的致病率和致死率^[10-11]。

无乳链球菌荚膜的产生受荚膜多糖合成 (capsular polysaccharide synthesis, *cps*) 基因座控制^[2]。尽管各血清型无乳链球菌荚膜多糖的生物特性及其重复单位不同, 但对编码多糖相关基因的序列分析表明, 在所有血清型无乳链球菌荚膜多糖合成基因座中存在 5 个高度保守的基因,

即 *cpsA-E*^[12], 其中 *cpsABCD* 控制多糖链长度和转运多糖^[12-14]。RUBENS 等^[15]通过突变体实验推定 III 血清型无乳链球菌 *cpsD* 基因编码产物具有半乳糖基转移酶功能; YAMAMOTO 等^[16]将 I a 血清型无乳链球菌 *cpsE* 基因表达于大肠杆菌 JM109 菌株及糖基转移酶测定试验, 推定 *cpsE* 基因编码产物执行葡萄糖基转移酶功能, 且已推知 *cps I aE* 基因产物与 III 型无乳链球菌 CpsD 表现高度同源性 (98.9%)。唾液酸化的荚膜多糖是无乳链球菌主要的毒力因子^[12], 对协助该菌逃避宿主防御机制至关重要^[17]。LANCEFIELD 等^[18-19]利用特异荚膜多糖兔免疫血清证实其对无乳链球菌感染的老鼠具有免疫保护作用, 表明荚膜多糖是一种重要的保护性抗原物质。目前, 国内外没有报道过鱼源无乳链球菌 *cpsE* 基因的相关信息。

本文通过对分离自海南文昌患病罗非鱼的无乳链球菌强毒株进行基因组 DNA 提取, 结合

收稿日期: 2011-01-11 修回日期: 2011-02-16

资助项目: 教育部《长江学者和创新团队发展计划》创新团队项目 (IRT0848); 通威股份有限公司重点资助项目 (2006-2009)

通讯作者: 汪开毓, E-mail: kywang@sicau.edu.cn

NCBI 的 ORF Finder 和 BLAST 工具分析得到了编码该菌 *cpsE* 基因的 ORF, PCR 扩增 *cpsE* 基因并将其克隆到 pMD19-T 载体上,经 PCR 和双酶切鉴定及序列的进化关系比对,验证扩增基因即为 *cpsE* 基因,随后应用生物信息学分析工具对该基因进行了分子特性分析,以期为进一步研究罗非鱼源无乳链球菌 *cpsE* 基因的功能和作用机理提供分子生物学依据。

1 材料与方法

1.1 菌株

共 8 株无乳链球菌菌株分离自海南文昌患病罗非鱼,经毒性回归实验选取强毒株,菌株编号为 HN_08; *Escherichia coli* DH5 α 菌株购于天根公司。

1.2 主要试剂

pMD19-T (Simple) 克隆载体、高保真酶 PrimeSTAR[®] HS (Premix)、T₄ DNA 连接酶、限制性内切酶 *Bam*H I 和 *Hind* III 以及 DNA Marker、DNA 凝胶回收试剂盒和质粒 DNA 小量提取试剂盒均购自大连 TaKaRa 公司。细菌基因组 DNA 提取试剂盒 (DP302-02) 购自天根公司。

1.3 引物设计与合成

根据 GenBank 中无乳链球菌 *cpsE* 基因全序列 (GenBank 登录号为 AB017355),运用生物软件 Oligo 6.0 和 Primer 5.0 设计 *cpsE* 全基因序列的引物 1 对,由上海生工生物工程有限公司合成。引物序列分别为:上游 5'-CGGGATCCATGAAAATT-TGTCTGGTTGG-3',划线部分为添加的 *Bam*H I 酶切位点;下游 5'-CCAAGCTTTTAAAAAATTCCTCCTAAATT-3',划线部分为添加的 *Hind* III 酶切位点。

1.4 *cpsE* 基因的 PCR 扩增

BHI 肉汤 37 °C 恒温培养 24 h 收集菌体,依照天根细菌基因组 DNA 提取试剂盒 (DP302-02) 操作说明提取菌体基因组 DNA,以此 DNA 为模板,按常规克隆方法对 *cpsE* 基因进行 PCR 扩增。采用 25 μ L PCR 反应体系:PrimeSTAR[®] HS (Premix) 12.5 μ L, DNA 模板 1.0 μ L,上、下游引物 (20 pmol) 各 0.5 μ L,无菌超纯水 10.5 μ L。反应条件:94 °C 预变性 5 min;94 °C 1 min,55 °C 1 min,72 °C 40 s,共 30 个循环;最后 72 °C 延伸 10 min。取 5 μ L 扩增产物用 1% 琼脂糖凝胶电泳观察扩增结果。

1.5 *cpsE* 基因的 T 克隆、测序及序列分析

将 PCR 产物用 DNA 凝胶回收试剂盒回收,与 pMD19-T 载体连接过夜,构建 pMD19-T-*cpsE* 重组质粒,按常规方法转化 DH5 α 大肠杆菌感受态细胞,筛选阳性克隆并进行 PCR 和酶切鉴定。鉴定后的阳性质粒 pMD19-T-*cpsE* 送上海生工生物工程有限公司测序。

1.6 核酸序列特征分析

利用 NCBI 的 ORF Finder 工具分析开放阅读框架^[20],并利用 NCBI 的 BLASTN 工具进行序列的相似性搜索^[21]。

1.7 蛋白质序列特征分析

运用瑞士蛋白质专家网在线翻译工具将 *cpsE* 基因序列翻译成氨基酸序列^[22],NCBI Conserved Domains 查找工具分析蛋白质结构域^[23];利用 NCBI 的 BLASTP 程序进行氨基酸序列同源性检索^[24],利用 Clustal X 2.0 和 MEGA 4.1 软件^[25-26]进行氨基酸序列同源性及系统进化树分析;应用 Bioedit 7.0 软件进行疏水性分析^[27],TMHMM 进行跨膜区进行预测^[28];SignalP 3.0 Server 在线搜索查找信号肽序列^[29],NetPhos 2.0 及 NetNGlyc 1.0 在线程序分析磷酸化位点和糖基化位点^[30-31];PSIpred 程序预测蛋白质二级结构^[32],SAM_T08 自动蛋白同源建模数据库进行 3D 结构预测^[33],最后利用在线预测程序 <http://www.biotech.ou.edu/>进行 *cpsE* 编码蛋白的溶解度预测。

1.8 密码子偏爱性分析

利用 EMBOSS 的 2 个在线密码子偏爱性分析程序——CHIPS 和 CUSP 程序对蛋白核酸序列进行密码子偏爱性计算及分析,并将结果与大肠杆菌、酵母和人的密码子偏爱性进行比较计算 (本文中各密码子偏爱性数据来自 Condon usage database)^[34]。

2 结果

2.1 无乳链球菌 *cpsE* 基因的扩增和 T 克隆鉴定

以罗非鱼源无乳链球菌基因组 DNA 为模板,根据设计的 PCR 引物,经高保真 DNA 聚合酶特异性的扩增出一段约为 470 bp 的条带,与预计片段大小 466 bp 一致 (图 1)。PCR 产物经胶回收纯化后,与 pMD19-T 载体连接获得重组质粒 pMD19-T-*cpsE*,重组质粒经 *Bam*H I/*Hind* III 以

双酶切鉴定,结果显示两条明显的条带,一条大小约为470 bp,其大小与目的片段大小相同;另一条带约为2 500 bp,与载体 pMD19-T 大小相同(图2)。送上海生工生物工程有限公司测序,结果与目的基因 *cpsE* 碱基序列一致,本文基因登录号为 HQ888682。

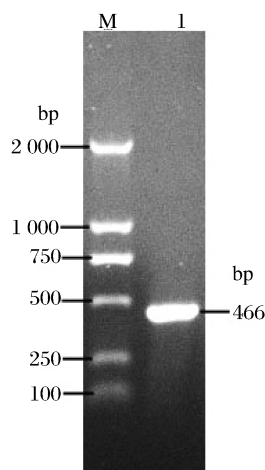


图1 罗非鱼源无乳链球菌 *cpsE* 基因的 PCR 扩增结果鉴定

M. DL2000Marker; 1道. *cpsE* 基因扩增产物。

Fig.1 PCR amplification result of *cpsE* gene of tilapia *S. agalactiae*

M. DNA marker-DL 2000; Lane1. PCR product of *cpsE* gene.

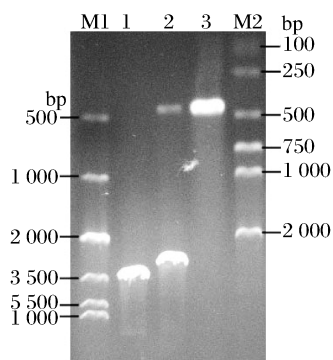


图2 重组质粒 pMD19-T-*cpsE* 的酶切和 PCR 鉴定

M1. Marker DL7000; M2: Marker DL2000; 1道. 重组克隆质粒经 *Bam*H I 单酶切产物; 2道. 重组克隆质粒经 *Hind* III 和 *Bam*H I 双酶切产物; 3道. 重组质粒的 PCR 产物。

Fig.2 Identification of recombinant plasmid of pMD19-T-*cpsE* by restriction enzyme digestion and PCR
M1. DNA marker-DL7000M2: Marker-DL2000; Lane1. pMD19-T-*cpsE*/*Bam*H I; Lane2. pMD19-T-*cpsE*/*Bam*H I + *Hind* III; Lane3. PCR product of recombinant plasmid.

2.2 核苷酸序列特征分析

对重组质粒 pMD19-T-*cpsE* 测序结果表明,

罗非鱼源无乳链球菌 *cpsE* 基因是一个由 450 个碱基组成的完整开放性阅读框 (ORF), 全序列中 GC 含量为 33.3%。

2.3 蛋白序列特征分析

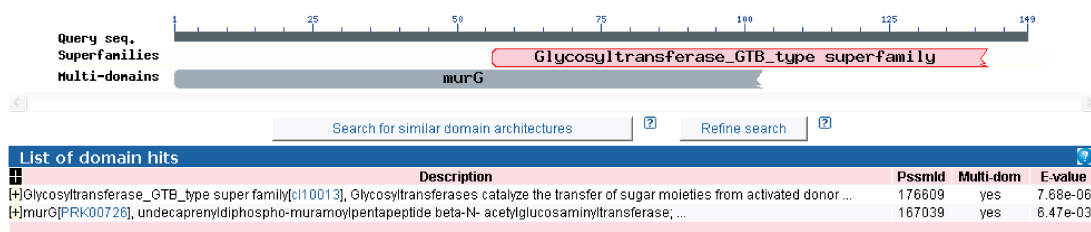
瑞士蛋白质分析专家系统 protparam 工具预测表明,罗非鱼源无乳链球菌 *cpsE* ORF 编码一条由 149 个氨基酸残基组成的多肽, 相对分子量为 17.2 ku, 等电点理论值为 9.36, 分子式为 $C_{802}H_{1254}N_{202}O_{206}S_5$ 。 *cpsE* 推导氨基酸序列成分分析表明 (表 1), 蛋白质中含量较高的氨基酸为 Lys (10.7%) 和 Val (10.1%), 含量较少的氨基酸为 Gln (0.7%) 和 Met (1.3%), 而 CpsE 氨基酸组成中不含有 Pyl 和 Sec。

表 1 罗非鱼源无乳链球菌 *cpsE* 基因序列推导肽链的氨基酸组分分析

Tab.1 Composition of amino acid deduced by *cpsE* gene of tilapia *S. agalactiae*

类型 type	比例/(%) fract	氨基酸 amino acid	数量 no.	比率/(%) ratio
亲水氨基酸 hydrophilic amino acid	18.2	Asn	6	4.0
		Cys	3	2.0
		Gln	1	0.7
		Ser	5	3.4
		Thr	7	4.7
		Tyr	5	3.4
	疏水氨基酸 hydrophobic amino acid	53.7	Ala	7
		Gly	9	6.0
		Ile	13	8.7
		Leu	13	8.7
		Met	2	1.3
		Phe	11	7.4
		Pro	7	4.7
酸性氨基酸 acidic amino acid	10.7	Glu	9	6.0
		Asp	7	4.7
	碱性氨基酸 basic amino acid	17.4	Lys	16
		Arg	7	4.7
		His	3	2.0

通过 NCBI 的 CDD (conserved domain) 工具分析发现,罗非鱼源无乳链球菌 CpsE 氨基酸序列第 56 和第 142 个氨基酸残基之间含有 1 个糖基转移酶 (Glycosyltransferases) 超家族的保守结构域 (图 3)。经一次迭代分析显示,该结构域序列与链球菌属的其他成员菌,如嗜热链球菌 (*S. thermophilus*)、肺炎链球菌 (*S. pneumoniae*)、链球菌 (*S. bovis*) 具有极高相似性。

图3 罗非鱼源无乳链球菌 *cpsE* 结构域预测结果Fig. 3 Prediction result of conserved domain in *cpsE* of tilapia *S. agalactiae*

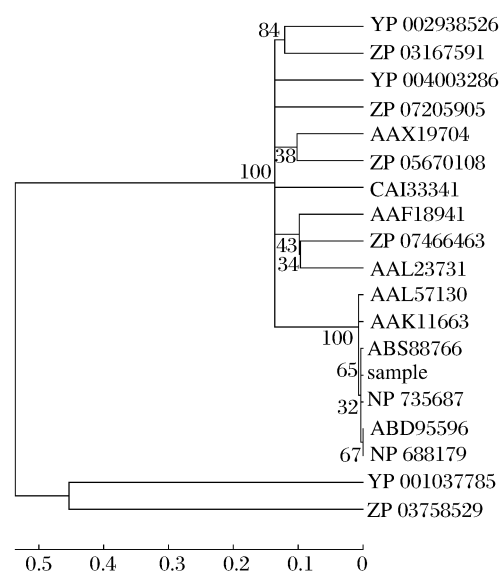
利用 Clustal X 2.0 软件对 18 个相关菌株 CpsE 氨基酸序列(表 2)运行 Multiple Sequence Alignment 程序和 MEGA 4.1 程序构建系统发生树系图(图 4),发现罗非鱼源无乳链球菌 *cpsE* 基因推导氨基酸序列与血清型 I b、III、V、VI 以及未知血清型无乳链球菌 CpsF 的同源性达 100%;与同属链球菌属的未知血清型肺炎链球菌、嗜热链球菌和牛链球菌以及血清 I 型猪链球菌(*S. suis*)的亲缘关系较远,而与直肠真杆菌(*Eubacterium rectale*)和酸奶瘤胃球菌(*Ruminococcus lactaris*)等的亲缘关系最远。

表 2 相关菌株的 *cpsE* 氨基酸序列Tab. 2 The *cpsE* amino acid sequence of related strains

名称 name	细菌 bacteria	GenBank 登录号 accession no.	血清型 serotype
CpsF	<i>S. agalactiae</i>	NP_735687	NT
CpsF	<i>S. agalactiae</i>	AAK11663	VI
CpsF	<i>S. agalactiae</i>	NP_688179	V
CpsF	<i>S. agalactiae</i>	ABS88766	III
CpsF	<i>S. agalactiae</i>	AAL57130	NT
CpsF	<i>S. agalactiae</i>	ABD95596	Ib
Eps3G	<i>S. thermophilus</i>	AAL23731	NT
EpsE	<i>L. lactis</i>	AAX19704	NT
CpsF	<i>S. pneumoniae</i>	CAI33341	NT
Cps2F	<i>S. suis</i>	AAF18941	I
CpsF	<i>S. bovis</i>	ZP_07466463	NT
Alg14	<i>E. faecium</i>	ZP_05670108	NT
EpsE	<i>E. rectale</i>	YP_002938526	NT
hypothetical protein RUMLAC	<i>R. lactaris</i>	ZP_03167591	NT
Alg14	<i>C. owensensis</i>	YP_004003286	NT
CpsE	<i>C. thermocellum</i>	YP_001037785	NT
biofilm regulatory protein A	<i>L. salivarius</i>	ZP_07205905	NT
hypothetical protein	<i>C. asparagiforme</i>	ZP_03758529	NT

注:血清型一栏中,“NT”表示对应菌株的血清型为未知。

Note: In the column of serotype, “NT” means the serotypes of corresponding strains were unknown.

图4 罗非鱼源无乳链球菌 *cpsE* 与

参考菌株同源基因氨基酸序列进化树分析

“sample”代表罗非鱼源无乳链球菌 *cpsE* 编码氨基酸序列。

Fig. 4 The phylogenetic tree analysis between the *S. agalactiae cpsE* and its homologous of the reference strains

“sample” represents the amino acid sequence of tilapia *S. agalactiae cpsE*.

亲/疏水性预测分析表明,罗非鱼源无乳链球菌 CpsE 多肽链有 5 个主要的疏水区,分别位于氨基酸序列的 20-43 位、52-62 位、72-79 位、108-116 位和 129-141 位,整个蛋白质疏水性最大值为 2.31,最小值为 -2.13,与疏水区相比,亲水区占据该蛋白多肽链的区域大于疏水区域,表明该蛋白可能为亲水蛋白;NetPhos2.0 程序分析表明,当阈值为 0.5 时,该序列共有 3 个潜在的磷酸化位点,包含 1 个丝氨酸(Ser)磷酸化位点、1 个酪氨酸(Tyr)磷酸化位点和 1 个苏氨酸(Thr)磷酸化位点;NetNGlyc1.0 程序预测表明,在该序列中没有潜在的 N-糖基化位点;SignalP V3.0 程序的 SignalP-NN 和 SignalP-NNSIP-HMM 两个

模型预测显示,该序列没有信号肽切割位点,成熟蛋白即为149aa。在线预测工具 TMHMM Server v. 2.0 跨膜区预测结果表明该多肽链存在跨膜区,位于78-100aa。重组蛋白可溶性预测结果显示,当选择大肠杆菌(*E. coli*)为表达宿主时,该重组蛋白的不溶解度达64.1%。

PSIpred 程序预测结果表明,罗非鱼源无乳链球菌 CpsE 二级结构组成中, α -螺旋(Alpha-Helix)占的比例最大,达40.94%, β -折叠(Beta-Strand)为25.50%, β -转角仅为5.37%,无规则卷曲(Coil)为28.19%;SAM_T08 自动蛋白同源建模数据库进行3D结构预测,结果找到与 *cpsE* 基因较高同源性的已知3D结构模型,其三维构象主要包含 α -螺旋、 β -折叠和无规则卷曲,三维构象预测见图5。

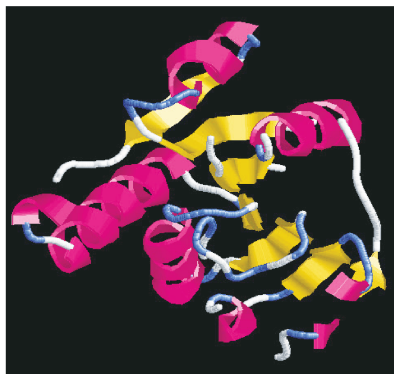


图5 罗非鱼源无乳链球菌 *cpsE* 三级结构预测图
 α -螺旋标记为红色, β -折叠标记为黄色,无规则卷曲为蓝色,其他为灰色。

Fig. 5 The tertiary structure of tilapia *S. agalactiae cpsE*

Helix is colored in red, the β strands are in yellow, and the loops structures are in blue, grey for other.

2.4 密码子偏爱性分析

CHIPS 程序分析结果表明,罗非鱼源无乳链球菌 *cpsE* 的 Nc 值为 48.730,说明其部分密码子的使用频率出现一定的差异,因此有必要进行 CUSPS 程序分析以检查该蛋白密码子使用的详细情况。经过 CUSPS 程序分析可知,该基因序列中编码同一氨基酸的不同密码子使用频率具有不同程度的差异,如编码 Ala 的 4 种密码子偏爱 GCA 和 GCT,编码 Gln 的 2 种密码子仅使用 CAG,编码 Tyr 的 2 种密码子仅使用 TAT, Cys 偏爱 TGT, Asp 偏爱 GAT 等。序列中包括 GCG、

GGC、GGG、CTC、CCG 等在内的 16 个密码子出现的频率为零,终止密码子仅出现 TAA。另外,通过将无乳链球菌 *cpsE* 的密码子与大肠杆菌、酵母及人的密码子偏爱性进行比较,分析结果显示,罗非鱼源无乳链球菌 *cpsE* 基因与大肠杆菌等 3 种宿主密码子使用频率差值较大的在大肠杆菌有 20 个,酵母有 18 个,人有 24 个。

3 讨论

无乳链球菌是感染人和动物的重要致病菌,荚膜多糖作为该菌关键的毒力因子,经唾液酸修饰后可钝化宿主旁路补体激活途径进而产生致病作用^[35]。目前,无乳链球菌已发现十种荚膜多糖血清型^[2-4],其中 I a、III 和 V 是已知对人类具有致病力的最常见血清型^[3],而 I a、I b 和 III 也已证明能引起鱼类感染发病^[36-37]。荚膜多糖合成基因座中的 *cpsE* 基因位于荚膜多糖合成操纵子的中央区,作为操纵子中亚基装配与寡糖聚合区的第一个基因,*cpsE* 是存在于所有血清型无乳链球菌荚膜多糖合成基因座中的高度保守序列,已被证实能启动多糖重复单位的合成作用,并与 *cpsFG*, *cpsI* 和 *cpsJ* 协同编码糖基转移酶^[12,16],是荚膜多糖合成及聚合的关键,这就提示 *cpsE* 基因在该菌感染宿主过程中发挥重要的作用。因此,进行 *cpsE* 基因缺失菌株的构建和筛选对无乳链球菌基因工程疫苗的研究值得关注。

宿主免疫系统的选择压力可促使病原菌抗原性差异荚膜类型的进化(变异)^[38],病原菌特异性毒力基因可能在其逃避宿主免疫防御作用下发生变异。通过进化关系分析,发现罗非鱼源无乳链球菌 *cpsE* 基因推导氨基酸序列与人源、牛源 Ib、III、V、VI 血清型无乳链球菌 CpsF 表现显著的同源性(100%)。目前,无乳链球菌 Cps_{III}F 功能未知,但该蛋白的肺炎链球菌 14 型同系物 Cps14F 已被证实可增强 Cps14G 活性,后者执行 β -1,4-半乳糖基转移酶功能,由此推断罗非鱼源无乳链球菌 *cpsE* 基因编码蛋白可能与增强 β -1,4-半乳糖基转移酶活性相关。

疏水性预测表明,该多肽链有 5 个主要的疏水区,其亲水区占据区域大于疏水区域,且该多肽链含跨膜区,提示该蛋白可能存在较多的 B 细胞抗原表位。蛋白翻译后修饰(post-translational modifications, PTMs)(如磷酸化)是一种几乎存

在于所有蛋白质形成的过程,修饰后的蛋白常能较强烈地影响蛋白功能的发挥,该过程主要通过蛋白水解作用裂解或者向一个或多个氨基酸添加一个修饰基团而改变蛋白的特性^[30,39]。本文研究的 *cpsE* 推导氨基酸序列中存在 3 个的磷酸化位点,这些潜在的磷酸化位点可能在蛋白翻译后得到一定程度的修饰,并且可能在 CpsE 发挥其生物学功能中起重要调控作用。同时分析出罗非鱼源无乳链球菌 CpsE 重组蛋白不溶解度理论预测值为 64.1%,表明在选择大肠杆菌作为表达宿主构建重组质粒时,在 *cpsE* 基因两端选择加入合适的融合蛋白标签非常重要,也更有利于提高重组蛋白的可溶解度及其纯化水平^[40]。

关于鱼源无乳链球菌 *cpsE* 在密码子使用上的偏爱性问题,目前国内外未有报道。作者对临床分离到的罗非鱼源无乳链球菌 *cpsE* 基因编码的密码子偏爱性进行了分析,结果表明该蛋白中编码相同氨基酸密码子的偏爱性存在较大的差异。通过与大肠杆菌、酵母和人的密码子偏爱性进一步比较可以看出,*cpsE* 的密码子与大肠杆菌和人的相差较大,其表达系统选择更适合在真核系统(如酵母)中高效作用。当然,外源基因的表达受到多种因素的共同作用,如要表达蛋白的种类、是否需糖基化或其它修饰、是否是毒素、是否需大量表达、如何纯化等,而密码子偏爱性的影响只是其中一种因素。通过密码子偏爱性分析不仅可为目的基因表达选择合适的表达系统,同时也可为改造密码子从而提高基因的表达量提供依据。

参考文献:

- [1] WAGNER B, WAGNER W, KUBIN V. Immunoelectron microscopic study of the location of group-specific and protein type-specific antigens of group B streptococci[J]. J Gen Microbiol,1980,118(1):95-105.
- [2] CHAFFIN D O, BERES S B, YIM H H, et al. The serotype of type I a and III group B streptococci is determined by the polymerase gene within the polycistronic capsule operon[J]. J Bacteriol,2000,182:4466-4477.
- [3] PERSSON E, BERG S, TROLLFORS B, et al. Serotypes and clinical manifestations of invasive group B streptococcal infections in western Sweden 1998-2001[J]. Clin Microbiol Infect,2004,10:791-796.
- [4] SLOTVED H C, KONG F, LAMBERTSEN L, et al. Serotype IX, a Proposed New *Streptococcus agalactiae* Serotype[J]. J Clin Microbiol,2007,45:2929-2936.
- [5] SCHRAG S J. A population-based comparison of strategies to prevent earlyonset group B streptococcal disease in neonates[J]. N Engl J Med,2002,347:233-239.
- [6] SKOFF T H, FARLEY M M, PETIT S, et al. Increasing burden of invasive group B streptococcal disease in nonpregnant adults,1990-2007[J]. Clin Infect Dis,2009,49(1):85-92.
- [7] TRIGO G, FERREIRA P, RIBEIRO N. Identification of immunoreactive extracellular proteins of *Streptococcus agalactiae* in bovine mastitis[J]. Can J Microbiol,2008,54(11):899-905.
- [8] ELDAR A, BEJERANO Y H B. *Streptococcus shiloi* and *Streptococcus difficile*: two new streptococcal species causing a meningoencephalitis in fish[J]. Current Microbiology,1994,28(3):139-143.
- [9] EVANS J J, KLESZIUS P H, GILBERT P M. Characterization of β -haemolytic Group B *Streptococcus agalactiae* in cultured seabream, *Sparus auratus* L., and wild mullet, *Liza klunzingeri* (Day), in Kuwait[J]. Journal of Fish Diseases,2002,25(9):505-513.
- [10] SALVADOR R, MULLER E E, LEONHARDT J H. Isolamento de *Streptococcus* spp. detilápia do Nilo (*Oreochromis niloticus*) equalidade da água de tanques rede na Região Norte do Estado do Paraná, Brasil[J]. Semina: Ciências Agrárias,2003,24(1):35-42.
- [11] SALVADOR R, MULLER E E, FREITAS J C. Isolation and characterization of *Streptococcus* spp. group B in Nile tilapia (*Oreochromis niloticus*) reared in hapas nets and earth nurseries in the northern region of Parana State, Brazil[J]. Ciência Rural,2005,35(6):1374-1378.
- [12] CIESLEWICZ M J, KASPER D L, WANG Y. Functional analysis in type I a group B *Streptococcus* of a cluster of genes involved in extracellular polysaccharide production by diverse species of streptococci[J]. J Biol Chem,2001,276:139-146.
- [13] BENDER M H, CARTEE R T, YOTHER J. Positive correlation between tyrosine phosphorylation of CpsD and capsular polysaccharide production in *Streptococcus pneumoniae* [J]. J Bacteriol,2003,

- 185;6057 – 6066.
- [14] BENDER M H, YOTHER J. CpsB is a modulator of capsule-associated tyrosine kinase activity in *Streptococcus pneumoniae* [J]. *J Biol Chem*, 2001, 276;47966 – 47974.
- [15] RUBENS C E, HEGGEN L M, HAFT R F, *et al.* Identification of cpsD, a gene essential for type III capsule expression in group B streptococci [J]. *Molecular Microbiology*, 1993, 8(5):843 – 855.
- [16] YAMAMOTO S, MIYAKE K, KOIKE Y, *et al.* Molecular characterization of type-specific capsular polysaccharide biosynthesis genes of *Streptococcus agalactiae* type I a [J]. *J Bacteriol*, 1999, 181: 5176 – 5184.
- [17] CHAFFIN D O, MCKINNON K, RUBENS C E. CpsK of *Streptococcus agalactiae* exhibits $\alpha 2, 3$ -sialyltransferase activity in *Haemophilus ducreyi* [J]. *Molecular Microbiology*, 2002, 45(1):109 – 122.
- [18] LANCEFIELD R C. A serological differentiation of specific types of bovine hemolytic streptococci (group B) [J]. *J Exp Med*, 1934, 59:441 – 458.
- [19] LANCEFIELD R C. Two serological types of group B hemolytic streptococci with related, but not identical, typespecific substances [J]. *J Exp Med*, 1938, 67:25 – 40.
- [20] ROMBEL I T, SYKES K F, RAYNER S. ORF-FINDER: a vector for high-throughput gene identification [J]. *Gene*, 2002, 282(1 – 2):33 – 41.
- [21] EKLUND A C, FRIIS P, WEMERSSON R. Optimization of the BLASTN substitution matrix for prediction of non-specific DNA microarray hybridization [J]. *Nucleic Acids Res*, 2010, 38(4):e27.
- [22] WILKINS M R, GASTEIGER E, BAIROCH A. Protein identification and analysis tools in the ExPASy server [J]. *Methods Mol Biol*, 1999, 112: 531 – 552.
- [23] MARCHLER-BAUER A, ANDERSON J B, CHERUKURI P F. CDD: a Conserved domain database for protein classification [J]. *Nucleic Acids Res*, 2005, 33 (Database issue):192 – 196.
- [24] JACOB A, LANCASTER J, BUHLER J. Mercury BLASTP: accelerating protein sequence alignment [J]. *Association for Computing Machinery*, 2008.
- [25] TAMURA K, DUDLEY J. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0 [J]. *Mol Biol Evol*, 2007, 24(8):1596 – 1599.
- [26] THOMPSON J D, GIBSON T J, PLEWNIAK F, *et al.* The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools [J]. *Nucleic Acids Res*, 1997, 25(24):4876 – 4882.
- [27] KYTE J, DOOLITTLE R F. A simple method for displaying the hydropathic character of a protein [J]. *J Mol Biol*, 1982, 157(1):105 – 132.
- [28] MOLLER S, CRONING M D, APWEILER R. Evaluation of methods for the prediction of membrane spanning region [J]. *Bioinformatics*, 2001, 17(7):646 – 653.
- [29] BENDTSEN J D, NIELSEN H, von HEIJNE G. Improved prediction of signal peptides: SignalP 3.0 [J]. *J Mol Biol*, 2004, 340(4):783 – 795.
- [30] NIKOLAJ B, THOMAS S P, RAMNEEK G, *et al.* Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence [J]. *Proteomics*, 2004, 4(6):1633 – 1649.
- [31] BLOM N, GAMMELTOFT S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites [J]. *J Mol Biol*, 1999, 294(5): 1351 – 1362.
- [32] JONES D T. Protein secondary structure prediction based on position-specific scoring matrices [J]. *J Mol Biol*, 1999, 292(2):195 – 202.
- [33] KARPLUS K. SAM-T08, HMM-based protein structure prediction [J]. *Nucleic Acids Res*, 2009, 37 (Web Server issue):492 – 497.
- [34] NAKAMURA Y, GOJOBORI T, IKENURA T. Condon usage tabulate form the international DNA sequence databases; status for the year 2000 [J]. *Nucl Acids Res*, 2000, 28(1):292.
- [35] CHAFFIN D O, MENTELE L M, RUBENS C E. Sialylation of group B streptococcal capsular polysaccharide is mediated by *cpsK* and is required for optimal capsule polymerization and expression [J]. *J Bacteriol*, 2005, 187(13):4615 – 4626.
- [36] SUANYUK N, KONG F, KO D, *et al.* Occurrence of rare genotypes of *Streptococcus agalactiae* in cultured red tilapia *Oreochromis* sp. and Nile tilapia *O. niloticus* in Thailand—Relationship to human isolates? [J]. *Aquaculture*, 2008(284):35 – 40.
- [37] VANDAMME P, DEVRIESES L A, POT B. *Streptococcus difficile* is a nonhemolytic group B, type I b *Streptococcus* [J]. *Int J Syst Bacteriol*, 1997, 47(1):81 – 85.
- [38] LIPSITCH M. Bacterial vaccines and serotype

- replacement: lessons from *Haemophilus influenzae* and prospects for *Streptococcus pneumoniae* [J]. *Emerg Infect Dis*, 1999, 3: 336 – 345.
- [39] MANN M, JENSEN O N. Proteomic analysis of post-translational modifications [J]. *Nat Biotechnol*, 2003, 21 (3): 255 – 261.
- [40] WILKINSON D L, HAMISON R G. Predicting the solubility of recombinant proteins in *Escherichia coli* [J]. *Biotechnology (New York)*, 1991, 9 (5): 443 – 448.

Cloning and molecular characterization of *cpsE* gene of *Streptococcus agalactiae* isolated from tilapia

WANG Kai-yu^{1,2*}, FU Xi^{1,2}, XIAO Dan³, HUANG Jin-lu^{1,2},
WANG Jun^{1,2}, WANG Hao-cheng^{1,2}

(1. Fisheries Department of Sichuan Agricultural University, Ya'an 625014, China;

2. Key Laboratory of Animal Disease and Human Health of Sichuan Province,
Sichuan Agricultural University, Ya'an 625014, China;

3. Tongwei Co. Ltd., Chengdu 610041, China)

Abstract: The *cpsE* gene of *Streptococcus agalactiae* virulent strain isolated from tilapia was amplified by PCR with specific primers and cloned into pMD19-T vector. The recombinant plasmid was strongly confirmed by PCR combined with restriction enzyme digestion (*Bam*H I and *Hind* III). Molecular characterization analysis of *cpsE* gene was performed by bioinformatics tools like Clustal X 2.0, MEGA 4.1, Bioedit 7.0, TMHMM, NetPhos 2.0, NetNGlyc 1.0, SignalP 3.0 Server, PSIPred, SAM_T08 and CUSP. Results indicated that amino acid sequence deduced by *cpsE* of tilapia *S. agalactiae* was highly conserved and has revealed a surprising degree (100%) of homology among strains isolated from human and other mammals. Polypeptide analyzed in this study contained a Glycosyltransferases superfamily conserved domain that functioned as enzyme that catalyzed the transfer of sugar moieties from activated donor molecules to specific acceptor molecules. Moreover, the polypeptide possessed 3 phosphorylation sites which related to post-translational modification. The hydrophilic regions were larger than hydrophobic regions and transmembrane domain was predicted to exist in polypeptide chain. The analysis of codon bias demonstrated the codon usage frequency of *cpsE* of tilapia *S. agalactiae* was distinctly different and it preferred to perform in eucaryote.

Key words: *Streptococcus agalactiae*; *cpsE* gene; cloning; molecular characterization

Corresponding author: WANG Kai-yu. E-mail: kywang@sicau.edu.cn